

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Mauricio Quinones Vega

Aplicação de técnicas proteômicas à identificação de proteínas
sem evidência de existência e estudo de placentas ZIKV
positivas com fetos microcefálicos

RIO DE JANEIRO
2021

Mauricio Quinones Vega

Aplicação de técnicas proteômicas à identificação de proteínas sem evidência de existência e estudo de placentas ZIKV positivas com fetos microcefálicos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciências (Bioquímica).

Orientador: Prof. Gilberto B. Domont
Co-orientador: Prof. Fábio C.S. Nogueira

RIO DE JANEIRO
2021

Agradecimentos

Agradeço em primeiro lugar aos meus pais, Maribel e Yair por todo o amor e carinho que me deram que me mantêm forte a cada dia, por acreditar em mim e me encorajar a amar a ciência.

Agradeço especialmente ao amor da minha vida, Patricia por estar comigo todos os dias me apoiando e celebrando todas as minhas conquistas.

Aos meus orientadores, Fábio Nogueira e Gilberto Domont, por acreditarem no meu potencial e pelas oportunidades que me fizeram crescer como pessoa e profissional;

Aos meus amigos e companheiros de trabalho da Unidade Proteômica e LabProt pelo incentivo e amizade essenciais no dia-a-dia; eles são minha família brasileira.

Agradeço à Coordenação de Formação de Pessoal de Nível Superior (CAPES), à Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e em especial ao Instituto de Química e à Universidade Federal do Rio de Janeiro pela infraestrutura de alta qualidade oferecida para poder desenvolver meu projeto nos últimos anos

Resumo

Quinones-Vega, Mauricio. **Aplicação de técnicas proteômicas à identificação de proteínas sem evidência de existência e estudo de placentas ZIKV positivas com fetos microcefálicos.** Rio de Janeiro, 2021. Dissertação (Mestrado em Ciências – Bioquímica) - Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

As abordagens proteômicas têm se mostrado úteis para detectar e caracterizar as proteínas como produto de cada gene humano, além de medir alterações quantitativas frente a um estresse biótico, como uma infecção viral. Ainda existem proteínas que carecem de evidência de existência (PE, Protein Existence), que são classificadas de acordo com o banco de dados NextProt de PE1 à PE5, sendo PE1 para proteínas com evidência confirmada ao nível proteico, PE2 para aquelas com evidência transcricional, PE3 com evidência por homologia, PE4 com existência por predição de modelos de genes e PE5 para proteínas com existência incerta. A descoberta experimental dessas proteínas é uma tarefa desafiadora para a proteômica dada a grande complexidade temporal e espacial do proteoma humano. Neste trabalho foram aplicadas diferentes técnicas proteômicas, para detectar proteínas PE2-5 e, posteriormente, comparar o proteoma de tecidos placentários de mães infectadas com ZIKV. Na primeira parte, foram aplicadas diferentes técnicas proteômicas centradas na extração, imunopurificação e o fracionamento de proteínas e peptídeos para detectar proteínas PE2-5. Para isso, foram realizadas três extrações de proteínas com três tampões diferentes: (1) Triton X-100; (2) Ureia/Tiourea/SDC e (3) SDS/TEAB a partir dos tecidos tireoidianos e testiculares. Através da separação das proteínas por SDS-PAGE e sua detecção por proteômica alvo-direcionada (PRM), a proteína P0DSN7 (PE5) foi confirmada a partir de 3 peptídeos únicos. A proteína Q8NHW5 (PE5) foi detectada com um único peptídeo sequenciado nas amostras de placenta e tireoide por digestão em solução e por fracionamento no nível de peptídeos, respectivamente. A integração de todas as metodologias foi capaz de detectar a proteína P01893 (PE5) com 4 peptídeos exclusivos de amostras de placenta, tireoide e testículos. Na segunda parte, aplicamos uma abordagem de quantificação label-free para comparar o proteoma de tecidos placentários de mães infectadas com ZIKV durante a gravidez e que desenvolveram fetos microcefálicos (MC+) com placentas controles saudáveis (Ctr) e infectadas pelo ZIKV sem microcefalia nos recém nascidos (Z+). Na comparação do grupo Z+ com o grupo Ctr encontramos proteínas associadas a danos ao DNA e inibição da expressão gênica reguladas positivamente. A organização dos filamentos de actina e a resposta imune também foram encontradas desreguladas no grupo Z+. Uma comparação entre os grupos MC+ e Z+ mostrou uma maior abundância de proteínas relacionadas à endocitose e autofagia em MC+, sugerindo uma maior transcitose de vesículas com partículas virais na interface materno-fetal. Nossos resultados sugerem que a maior expressão de integrinas em MC+ pode estar associada à alta internalização do vírus, uma vez que essas proteínas são conhecidas como receptores de vírus. Da mesma forma, um aumento da resposta imunológica na placenta e uma maior infiltração do vírus no feto podem contribuir para malformações neurológicas associadas à síndrome congênita da Zika. As estratégias proteômicas utilizadas neste estudo permitiram a detecção de proteínas sem previa evidência experimental além da comparação quantitativa entre placentas ZIKV-positivas portadoras ou não de fetos microcefálicos.

Palavras-chave: Proteomica, Espectrometria de massas, Missing Proteins, Placenta, ZIKV, Microcefalia.

Abstract

Quinones-Vega, Mauricio. **Application of proteomic techniques for the identification of missing proteins and the study of ZIKV positive placentas of microcephalic fetuses.** Rio de Janeiro, 2021. Dissertation (Masters in Science – Biochemistry) - Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, 2021.

Proteomic approaches have been shown to be useful to detect and characterize proteins as a product of each human gene, in addition to measuring quantitative changes in the face of biotic stress, such as a viral infection. There are still proteins that lack evidence of existence (PE, Protein Existence), which are classified according to the NextProt database from PE1 to PE5, being PE1 for proteins with confirmed evidence at the protein level, PE2 for those with transcriptional evidence, PE3 with homology based evidence, PE4 with gene model prediction and PE5 for proteins with uncertain existence. The experimental discovery of these proteins is a challenging task for proteomics given the great temporal and spatial complexity of the human proteome. In this work, different proteomic techniques were applied to detect PE2-5 proteins and, subsequently, to compare the proteome of placental tissues from ZIKV-infected mothers. In the first part, different proteomic techniques centered on the extraction, immunopurification and fractionation of proteins and peptides were applied to detect PE2-5 proteins. For this, three protein extractions were performed with three different buffers: (1) Triton X-100; (2) Urea/Thiourea/SDC and (3) SDS/TEAB from thyroid and testicular tissues. Through protein separation by SDS-PAGE and detection by target-directed proteomics (PRM), the protein P0DSN7 (PE5) was confirmed from 3 unique peptides. Protein Q8NHW5 (PE5) was detected with a single sequenced peptide in placental and thyroid samples by solution digestion and by fractionation at the peptide level, respectively. The integration of all methodologies was able to detect protein P01893 (PE5) with 4 unique peptides from placental, thyroid and testis samples. In the second part, we apply a label-free quantification approach to compare the proteome of placental tissues from mothers infected with ZIKV during pregnancy and who developed microcephalic fetuses (MC+) with healthy control placentas (Ctr) and ZIKV infected placentas without microcephaly in newborns (Z+). Comparing the Z+ group with the Ctr group, we found positively regulated proteins associated with DNA damage and inhibition of gene expression. The organization of actin filaments and the immune response were also found to be dysregulated in the Z+ group. A comparison between the MC+ and Z+ groups showed a greater abundance of proteins related to endocytosis and autophagy in MC+, suggesting a greater transcytosis of vesicles with viral particles at the maternal-fetal interface. Our results suggest that the higher expression of integrins in MC+ may be associated with high virus internalization, since these proteins are known as virus receptors. Likewise, an increased immune response in the placenta and increased virus infiltration into the fetus may contribute to neurological malformations associated with congenital Zika syndrome. The proteomic strategies used in this study allowed the detection of proteins without previous experimental evidence and the quantitative comparison between ZIKV-positive placentas with or without microcephalic fetuses.

Keywords: Proteomics, Mass Spectrometry, Missing Proteins, Placenta, ZIKV, Microcephaly.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1: Categorização dos 606 genes do Cr15 de acordo com seu nível de evidência segundo neXtProt (Revisado junho 2021)..... | 24 |
| Figura 2: Dados de sequenciamento de RNA de 28 MP com nível PE2 do Cromossomo 15 em 53 amostras de tecidos humanos | 26 |
| Figura 3. Níveis de expressão de RNA dos genes de cromossomo 15, de acordo com as anotações no HPA versão 20.1 | 27 |
| Figura 4: Fluxo de trabalho geral do processamento das amostras para a identificação de proteínas sem evidência de existencia..... | 28 |
| Figura 5. Comparação de diferentes cromatogramas e espectros MS1 dos peptídeos das MP modelos identificados nos controles da primeira abordagem de imunoprecipitação. | 30 |
| Figura 6. Sequenciamento dos peptídeos controles das proteínas O60320 e A6NFU8 a partir dos espectros MS2 obtidos na análise por nanoLC-MS/MS da imunoprecipitação. | 32 |
| Figura 7. SDS-PAGE realizada a partir de extrações de proteínas em tecidos da tireoide e testículo usando diferentes tampões | 33 |
| Figura 8. Análise do peptídeo RVAVEGVDFIF da proteína A6NFU8 (PE2) detectado no controle e na amostra de testículo. | 36 |
| Figura 9. Análise das proteínas identificadas nas regiões de massa de 56 kDa e 21kDa a partir da digestão em gel | 37 |
| Figura 10. Resultados da análise por PRM das bandas de gel. | 39 |
| Figura 11. Cromatogramas do fracionamento em RP-HPLC em pH básico de peptídeos trípticos obtidos a partir de extratos proteicos de tireoide. | 41 |
| Figura 12. Análise proteômica do fracionamento por RP-HPLC em pH básico de extratos de proteínas procedentes de tecidos de tireoide. | 42 |
| Figura 13. Espectro de fragmentação do peptídeo DLISHNEMFSDIYK correspondente à proteína Q9HAU6. | 44 |
| Figura 14. Espectro de fragmentação dos peptídeos GYEQHAYDGKDYIALNEDLR e YTCHVQHEGLPEPLTLR correspondente à proteína P01893. | 45 |
| Figura 15. Espectros de fragmentação dos peptídeos YTCHVQHEGLPEPLTLR e GYEQHAYDGKDYIALNEDLR da proteína P01893, obtidos em diferentes abordagens... .. | 48 |

| | |
|--|----|
| Figura 16. Espectros de fragmentação do peptídeo SVETDYTFPLAENVK correspondente à proteína Q8NHW5 detectadas nas placentas controles..... | 50 |
| Figura 17. Espectros de fragmentação dos peptídeos DIQLTQSPSSLSASVGDR, ASNLQSGVPSR, e LLIYSASNLSQSGVPSR da proteína P0DSN7. | 51 |
| Figura 16. Análise proteômica global das placentas MC+, Z+, Ctr..... | 53 |
| Figura 17. Processos biológicos do GO enriquecidos na lista de proteínas reguladas positivamente e negativamente obtidos pela análise na ferramenta bioinformática DAVID. | 54 |
| Figura 18. Anotações e vias biológicas de GO mais representadas entre as proteínas reguladas positivamente do grupo MC+ em comparação com os grupos Ctr e Z+. | 56 |
| Figura 19. Anotações dos processos biológicos de GO mais representadas entre as proteínas reguladas negativamente do grupo MC+ em comparação com o grupo Ctr e Z+. | 57 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela I. Lista de inclusão dos peptídeos FAM189A1 e PGPEP1L..... | 20 |
| Tabela II. Lista de inclusão de peptídeos para a análise por PRM | 21 |
| Tabela III. Número de proteínas identificadas na primeira abordagem de imunoprecipitação de MP..... | 29 |
| Tabela V. Peptídeos do PrEST identificados na abordagem sequencial de imunoprecipitação de MP acoplada a espectrometria de massas. | 31 |
| Tabela VI. Número de proteínas identificadas na segunda abordagem de imunoprecipitação de MP..... | 34 |
| Tabela VII. Resultado da identificação por espectrometria de massas das <i>Missing Proteins</i> FAM198A1 e A6NFU8 usando uma abordagem de processamento em paralela. | 34 |
| Tabela VIII. Proteínas com nível de existência PE4-5 identificadas nas análises proteômicas das bandas de SDS-PAGE. | 38 |
| Tabela IX. Proteínas sem prévia evidência experimental identificadas na extração com o tampão de SDS/TEAB 100 mM, processadas por S-trap e fracionadas por RP-HPLC em pH básico..... | 42 |
| Tabela X. Lista de proteínas identificadas sem evidência experimental por MS usando a extração com ureia 7M/tiourea 2M/SDC 10%, e o fracionamento de peptídeos por RP-HPLC em pH básico. | 43 |
| Tabela XI. Proteínas sem prévia evidência experimental por espectrometria de massas detectadas no estudo de placentas de grávidas infectadas com ZIKV. | 46 |
| Tabela XII. Peptídeos únicos correspondentes à proteína P01893 detectados neste trabalho. | 47 |
| Tabela XIII. Peptídeos únicos correspondentes à proteína Q8NHW5 detectados neste trabalho | 49 |

SUMÁRIO

| | | |
|--------|---|----|
| 1. | Introdução..... | 1 |
| 1.1. | Projeto do Proteoma Humano..... | 1 |
| 1.1.1. | Iniciativas do HPP..... | 3 |
| 1.2. | A definição e desafio das Missing Proteins..... | 4 |
| 1.3. | Metodologias para identificar MPs..... | 6 |
| 1.4. | Tecidos importantes na pesquisa de MPs..... | 9 |
| 1.4.1. | Estudos em testículo..... | 9 |
| 1.4.2. | Estudos em tireoide..... | 10 |
| 1.4.3. | Estudos em placenta..... | 10 |
| 1.4.4. | MPs no contexto da doença..... | 11 |
| 1.5. | Zika Vírus..... | 11 |
| 1.5.1. | Síndrome Congênita da Zika..... | 12 |
| 1.5.2. | Proteômica no estudo do Zika..... | 13 |
| 2. | Objetivos..... | 14 |
| 2.1. | Objetivos geral..... | 14 |
| 2.2. | Objetivos específicos..... | 14 |
| 3. | Materiais e Métodos..... | 15 |
| 3.2. | Coleção de amostras..... | 15 |
| 3.3. | Extração de proteínas..... | 15 |
| 3.4. | Imunopurificação de Missing Proteins..... | 16 |
| 3.4.1. | Imunopurificação com esferas Dynabeads M-270 Epoxy (abordagem 1)..... | 16 |
| 3.4.2. | Imunopurificação com esferas Dynabeads M-280 Sheep anti-Rabbit IgG (abordagem 2)..... | 16 |
| 3.5. | Eletroforese..... | 17 |
| 3.6. | Digestão e limpeza de amostra..... | 17 |
| 3.6.1. | Quantificação das amostras em Qubit..... | 17 |
| 3.6.2. | Digestão em solução..... | 18 |
| 3.6.3. | Digestão em gel..... | 18 |
| 3.6.4. | Limpeza dos peptídeos em ponteiras com matriz POROS R2..... | 18 |
| 3.6.5. | Digestão e Limpeza em S-Trap..... | 18 |
| 3.7. | Fracionamento off-line em cromatografia de Fase Reversa de pH básico..... | 19 |
| 3.8. | Análise por espectrometria de massas – nanoLC-MS/MS..... | 19 |
| 3.8.1. | Aquisição Dependente de Dados..... | 19 |
| 3.8.2. | Análise por PRM..... | 21 |

| | |
|--|----|
| 3.9. Análise de dados proteômicos | 22 |
| 3.9.1. Análise funcional das proteínas | 23 |
| 4. Resultados e Discussão | 24 |
| Metodologias para identificação de <i>Missing Proteins</i> | 24 |
| 4.1. Estudo de expressão das Missing Proteins do Cr 15 e estratégia geral | 24 |
| 4.2. Imunopurificação de 2 Missing Proteins modelos | 28 |
| 4.2.1. Abordagem 1 de imunopurificação de Missing Proteins Modelos | 29 |
| 4.2.2. Abordagem 2 de imunopurificação de Missing Proteins Modelos | 33 |
| 4.3. Fracionamento de proteínas para identificar Missing Proteins | 36 |
| 4.3.1. Digestão de bandas de MP | 36 |
| 4.3.2. Fracionamento em RP-HPLC em pH básico | 40 |
| 4.3.3. Identificação de Missing em tecido de placenta infectada com Zika | 45 |
| 4.4. Integração dos dados proteômicos para a detecção de proteínas sem evidência ao nível de espectrometria de massas | 46 |
| Análise proteômica de Placenta infectada pelo ZIKV | 52 |
| 4.5. Proteômica de placentas infectadas pelo Zika | 52 |
| 4.6. Análise funcional das proteínas expressas diferencialmente | 54 |
| 5. Conclusões | 59 |
| 6. Perspectivas | 61 |
| 7. Referências Bibliográficas | 62 |

1. Introdução

1.1. Projeto do Proteoma Humano

Estimulada pelo sucesso no sequenciamento completo do genoma humano, em fevereiro de 2001 foi lançada a Organização do Proteoma Humano (*Human Proteome Organization*, HUPO). Fundada para agrupar cientistas dos setores acadêmico, industrial e governamental e atrair jovens cientistas do mundo todo, a organização promove e estimula a pesquisa e desenvolvimento das técnicas proteômicas e tem um papel ativo na coordenação de iniciativas científicas (Omenn, 2021) aplicadas ao estudo de proteínas humanas e importantes para a espécie humana. As principais missões são consolidar uma organização proteômica mundial; envolver-se em atividades científicas e educacionais para disseminar tecnologias e conhecimentos relativos ao proteoma humano e de organismos modelos; e auxiliar na coordenação de iniciativas públicas de proteoma (Hanash and Celis, 2002).

Uma década depois do surgimento da HUPO e junto com os avanços em técnicas capazes de identificar um grande número de proteínas, em setembro de 2010, durante o 9º Congresso Mundial Anual da HUPO em Sydney foi proposto o lançamento do Projeto do Proteoma Humano (*Human Proteome Project*, HPP) coordenado pelo Dr. Gil Omenn (Legrain *et al.*, 2011). O HPP começou, e assim continua, sem financiamento de longo prazo. O apoio financeiro ao longo da década veio por meio de projetos individuais, infraestrutura institucional/instalações básicas e doações filantrópicas. O HPP estimula a colaboração internacional entre cientistas, médicos, indústria, instituições e parceiros de base de conhecimento para criar uma estrutura que garanta o compartilhamento de dados de qualidade, visando descobrir todo o complemento de proteínas codificadas pelo genoma humano. A principal missão do HPP é mapear todo o proteoma humano usando técnicas atualmente disponíveis e emergentes. A análise sistemática dos dados oferecidos pela comunidade permite a melhor compreensão molecular da natureza dinâmica do proteoma, suas modificações e relações com a biologia humana e as doenças estabelecendo uma base para o desenvolvimento de aplicações médicas diagnósticas, prognósticas, terapêuticas e preventivas. Isso se alinha de perto com o objetivo do HUPO de "traduzir o código da vida", fornecendo informações bioquímicas e biológicas celulares cruciais que a genômica por si só não pode fornecer. Desde o seu início, o HPP afirmou dois objetivos estratégicos:

1. Catalogar com rigor o proteoma humano e descrever sua complexidade, incluindo modificações pós-traducionais (PTMs), variantes de *splicing*, interações e funções, ao:
 - a. Estabelecer padrões acordados, rigorosos e confiáveis.
 - b. Identificar pelo menos um produto proteico de cada gene codificador no genoma.
2. Tornar a proteômica um componente integrado de estudos multi-ômicos para o avanço das ciências da vida, ciências biomédicas e medicina de precisão.

No início o HPP foi organizado em uma matriz estratégica sustentada por três pilares principais: O pilar de espectrometria de massa (MS); de anticorpo (AB); de banco de conhecimentos (de dados, (KB). Recentemente, foi adicionado um quarto pilar relacionado a patologias (P). Esses pilares asseguram a geração, integração e implementação eficiente de dados, além de estabelecerem métricas e diretrizes assim como plataformas tecnológicas, bem como o desenvolvimento de reagentes e o uso ideal de fluxos de dados existentes e emergentes.

O pilar de MS informa a comunidade sobre os avanços em tecnologia e estratégias associadas a MS, padrões de alto rigor apropriados e contatos com a indústria em relação ao desenvolvimento de instrumentos, sempre levando a maior exatidão e precisão na identificação, quantificação e modificação de proteoma. Ele acompanha o desenvolvimento de ensaios baseados no Monitoramento de Reações Seleccionadas (*Selected Reaction Monitoring, SRM*) de peptídeos proteotípicos, incluindo peptídeos marcados com isótopos estáveis, que permitem medições quantitativas e alvo-direcionadas de proteína com alto rendimento e alto grau de reprodutibilidade (Legrain *et al.*, 2011). Para isso foi gerado um banco de dados especializado a partir de espectros de íons fragmentos para vários peptídeos proteotípicos de quase todas as proteínas do proteoma humano, fornecendo as informações para a identificação definitiva e quantificação precisa de cada proteína em amostras biológicas (www.SRMAtlas.org) (Kusebauch *et al.*, 2016). Hoje é possível a quantificação de 99,7 % das proteínas humanas anotadas com mais de 158,000 ensaios de SRM (Elguoshy *et al.*, 2019).

O pilar do AB é liderado pelo Atlas de Proteína Humano (*Human Protein Atlas, HPA*)(Thul and Lindskog, 2018), desenvolvido na Suécia no ano 2003, que promove o uso de estratégias baseadas em anticorpos para relacionar a identificação de uma proteína com sua localização no tecido ou na célula em tempo real, melhorando o entendimento das funções biológicas e moleculares das proteínas. Para construir este atlas de proteínas humanas, foram usados 48 tecidos humanos normais e 20 tipos diferentes de câncer (Uhlén *et al.*, 2005). O HPA tem disponível cerca de 31.000 anticorpos direcionados a mais de 18.000 proteínas para imuno-histoquímica e imunofluorescência que, juntamente com a transcriptômica, constituem um atlas de localização de proteínas em tecidos, células, patologia, cérebro, metabolismo, sangue e secretoma (Omenn, 2021).

O pilar do KB coleta, agrupa, analisa e redistribui todos os dados de proteoma humano, o que o torna uma estrutura útil para colaboração internacional, compartilhamento e reanálise de dados, garantia de qualidade, diretrizes, comunicação e aceleração do progresso na construção e utilização dos conhecimentos proteômicos. Em 2013, o repositório neXtProt (Gaudet *et al.*, 2013; Zahn-Zabal *et al.*, 2020) foi oficialmente designado como o banco de dados de referência do HPP (Gaudet *et al.*, 2013; Lane *et al.*, 2012). Anualmente, uma versão do neXtProt é designada como a 'versão HPP' e isso serve como base para subseqüentes análises, planejamento e relatório de progresso de futuras versões. Ele recebe e seleciona dados de

UniProtKB/SwissProt, adicionando evidências de MS do PeptideAtlas (Desiere et al., 2006) e desde 2019 do MassIVE. NeXtProt também coleta dados baseados em AB, genoma, transcriptômica, informações estruturais 3D (*Protein Data Bank*, PDB) e modificações pós-traducionais da literatura com curadoria manual. De acordo com a versão v2.35.0 (2021-02-18) a plataforma conta com 20.379 entradas de proteínas e 42.368 isoformas produzidas por “splicing”. O PeptideAtlas (<http://www.peptideatlas.org>) usa motores de busca de sequência como o Comet (Eng, Jahan, and Hoopmann, 2013), X!Tandem (Bjornson *et al.*, 2008) e SpectraST (Lam *et al.*, 2007) para reprocessar espectros MS/MS carregados publicamente e depositados através do ProteomeXchange (PX). Os dados são agregados usando critérios rigorosos, incluindo correspondência espectral dos peptídeos com FDR ~ 0,0009% para atingir um FDR ≤1% para proteínas (Adhikari *et al.*, 2020).

Desde 2018, no congresso mundial da HUPO em Orlando, USA foi criado um novo pilar de Patologia devido ao papel fundamental que essa área tem na sua tradução em aplicações clínicas de dados e métodos proteômicos. O pilar coordena a identificação das principais necessidades não atendidas na área clínica e estimula diretrizes e padrões para o desenvolvimento de ensaios clínicos validados e adequados assim como promove a conscientização sobre o uso das melhores práticas e o acesso a biorecursos de amostras clínicas com seus dados associados. Também ajuda no estabelecimento de parcerias da HUPO com organizações internacionais de patologia, indústrias de diagnóstico e agências regulatórias.

O projeto tem alcançado muitas conquistas desde o seu início. Por exemplo, a criação de uma estrutura para o mapeamento comunitário do proteoma humano. Para garantir que a confirmação de uma proteína a partir de dados obtidos em instrumentos de massa de alta precisão sejam de alta qualidade são aplicadas análises e critérios cada vez mais rigorosos organizados nas diretrizes mais recentes da HPP MS v3.0 (Deutsch *et al.*, 2019). Estas diretrizes para a interpretação dos dados proteômicos estão projetadas para tornar os dados espectrais localizáveis, acessíveis, interoperáveis e reutilizáveis. As diretrizes promovem a aplicação de análises padronizadas dos dados para completar progressivamente a lista de partes do proteoma humano. Outra conquista do projeto foi o aumento do perfil e a visibilidade da proteômica, como um componente essencial das ciências da vida e da pesquisa biomédica, promovendo o desenvolvimento de instrumentação e métodos para análise de proteoformas, bem como atividade/função das proteínas que não pode ser abordada pela genômica. O projeto iniciou um programa - μ PE1 - para determinar a função biológica de proteínas com identificações confiáveis que atualmente carecem de anotação funcional.

1.1.1. Iniciativas do HPP

Desde o início o HPP estabeleceu várias iniciativas como o objetivo de utilizar tecnologias proteômicas avançadas para complementar as informações genômicas. Essas iniciativas multinacionais começaram com os projetos de proteoma de plasma (Omenn *et al.*, 2005), fígado (He, 2005) e cérebro (Meyer, Klose, and Hamacher, 2003), depois rim/urina (Yamamoto, Langham, Ronco, Knepper, and Thongboonkerd, 2008) e

cardiovascular (Ping *et al.*, 2005), juntamente com a Iniciativa de Padrões de Proteína coordenada pelo Instituto Europeu de Bioinformática (Orchard, Hermjakob, and Apweiler, 2003). Mais tarde essas iniciativas da HUPO formaram um consórcio, a saber, o Projeto do Proteoma Humano impulsionado por Biologia/Doença (*Biology/Disease-driven Human Proteome Project*, B/D-HPP) (Aebersold *et al.*, 2014).

Em 2011, a HUPO anunciou oficialmente outro consórcio, o Projeto do Proteoma Humano centrado em Cromossomo (*Chromosome-centric Human Proteome Project*, C-HPP), para mapear e anotar um conjunto completo de proteínas codificadas pelo genoma, cromossomo por cromossomo. Liderado por Young-Ki Paik e William Hancock, e mais tarde por Chris Overall e Lydie Lane, o C-HPP reúne equipes focadas em cada um dos 24 cromossomos e o mitocondrial. Esta estratégia é análoga ao projeto do genoma humano com divisão de trabalho entre 25 equipes de pesquisadores de proteômica em muitas nações ou regiões ao redor do mundo. As equipes do C-HPP desempenham um papel fundamental no estabelecimento dos principais acontecimentos do projeto em seis áreas de cooperação com as equipes de bioinformática e investigadores individuais. As áreas de cooperação são: (i) o sistema de 'Métricas' para atualizar o progresso anual na anotação de proteínas (Omenn *et al.*, 2020), (ii) a regra de envio e depósito de dados ao PX, que foi um primeiro passo para o compartilhamento de dados em toda a comunidade, (iii) as diretrizes de interpretação de dados de MS v3.0 (Deutsch *et al.*, 2019), (iv) ferramentas de bioinformática de gerenciamento de dados, (v) colaboração para as publicações de edição especial da JPR e (vi) utilização de amostra rara para detecção de proteínas sem evidências experimentais ao nível proteico. Essas realizações do C-HPP teriam sido impossíveis sem o apoio e cooperação de toda a comunidade da HUPO (Y. K. Paik, Omenn, Hancock, Lane, and Overall, 2017).

1.2. A definição e desafio das Missing Proteins

Em um esforço comum para definir o proteoma humano, os dados dos laboratórios participantes em todo o mundo são submetidos a repositórios do PX, principalmente no PRIDE (*Proteomics Identification Database*) (Perez-Riverol *et al.*, 2019) e, em seguida, os dados são reanalisados de acordo com procedimentos padrão e enviados para o neXtProt (Gaudet *et al.*, 2013). Combinando essas informações, o neXtProt classifica cerca de 20.000 entradas do UniProtKB/Swiss-Prot em cinco categorias com base na evidência de existência da proteína (*Protein Existence*, PE). O nível PE1 refere-se a proteínas cuja detecção é fortemente apoiada por evidências experimentais, notadamente, espectrometria de massa. O nível PE2 representa proteínas com expressão transcrita, mas sem evidência de proteína suficiente. As proteínas PE3 são aquelas com expressão homóloga em espécies não humanas, mas sem evidência da presença de transcritos de RNA ou proteína em humanos. O nível PE4 denota produtos traduzidos com base em previsões de modelagem de genes. Finalmente, o nível PE5 inclui entradas incertas ou duvidosas com baixa probabilidade de tradução. As proteínas dentro dos níveis de PE2, PE3 e PE4 são aquelas consideradas como "*Missing Proteins*" (MP) (Marko-Varga, Omenn, Paik, and Hancock, 2013). O último lançamento de referência do neXtProt HPP (<https://www.>

Nextprot.org/about/statistics, data: 2021-02-18) designa evidência PE1 confiável para 92,8% do proteoma humano (18.357 PE1s de 19.778 entradas PE1-4, excluindo as 577 entradas PE5). Isso deixa apenas 1.421 entradas para serem identificadas por MS, o que representa 7,2% do total de entradas PE1-4. Para a identificação confiável das MPs o projeto publicou diretrizes para interpretação dos dados por MS (Deutsch *et al.*, 2019) que têm como objetivo principal melhorar a qualidade da análise e minimizar a probabilidade de atribuições de falsos positivos levando a identificações errôneas de MP. Portanto, é agora acordado que a detecção de proteínas deve ser baseada em dois peptídeos únicos com 9 ou mais resíduos de comprimento e FDR <1% ao nível da proteína como um padrão geral dentro do HPP. As possíveis exceções devem ser apoiadas por evidências experimentais adicionais.

Vários motivos podem explicar a falta de evidências experimentais das MP. Primeiro, as limitações dos métodos proteômicos atuais para detectar proteínas de baixa abundância em amostras biológicas; a detecção dessas espécies de proteínas requer a redução da complexidade da amostra pré-fracionando os extratos/misturas de proteínas/peptídeos antes da análise por cromatografia líquida acoplada à espectrometria de massa (nanoLC-MS/MS) ou enriquecendo proteomas subcelulares. A identificação de MP melhora consideravelmente usando uma combinação de bibliotecas espectrais múltiplas durante a identificação. Este método fornece sensibilidade aprimorada e taxa de erro mais baixa do que a pesquisa de banco de dados de sequência convencional (J. Y. Cho *et al.*, 2015). O segundo fator que explica as dificuldades para detectar as MP são suas propriedades físico-químicas, como o tamanho reduzido ou sequências de aminoácidos que provavelmente não produzem peptídeos trípticos únicos ou adequados para a análise por MS. Foi relatado que 135 MP não geram peptídeos trípticos únicos após a digestão de tripsina (Garin-Muga *et al.*, 2016) e, portanto, fluxos de trabalho alternativos devem ser implementados para sua detecção por espectrometria de massa. Outro dos problemas das MP é sua hidrofobicidade, cerca de 34% das proteínas PE2-4 são proteínas de membrana com domínios hidrofóbicos transmembrana. Da mesma forma, também foi demonstrado que as proteínas PE5, se expressas, pertencem preferencialmente a transportadores e receptores localizados principalmente no compartimento da membrana plasmática (Dong, Menon, Omenn, and Zhang, 2015). Além disso, a expressão de MP pode ocorrer apenas em tecidos ou tipos celulares específicos, em resposta a certos estímulos ou em certos estágios de desenvolvimento, conforme sugerido por estudos evolutivos (Xu *et al.*, 2015).

A busca por MP em amostras biológicas adequadas é então de extrema importância e, portanto, a identificação de tecidos/células e condições que aumentam a probabilidade de detecção de MP é obrigatória. A análise dos dados obtidos pela transcriptômica, pode ajudar na otimização das estratégias destinadas a identificar MP. De acordo com essa ideia, foi demonstrado que o nível de expressão do gene codificador correspondente e as propriedades das proteínas forneceram pistas para entender as razões para a falta de evidências de MPs codificadas pelo cromossomo 1 (C. Zhang *et al.*, 2014). Em relação à expressão específica para tecidos, Pinto *et al.* examinaram 30

tecidos e células humanos não patológicos e relataram 47 MPs codificadas por genes do cromossomo 22 (cr22); além disso, a análise proteogenômica permitiu a detecção de novas regiões codificantes neste cromossomo (Pinto *et al.*, 2014). Finalmente, a falta de expressão, anotações errôneas do genoma que resulta em uma predição falsa da proteína, proteínas altamente homólogas ou proteínas com grande variabilidade de sequência que não podem ser diferenciadas com base em dados de MS/MS também podem fornecer explicação para a falta de evidência experimental das MP. Nesse sentido, diferentes esforços têm sido feitos para melhorar a anotação das MPs, com o objetivo de melhorar a nossa compreensão de suas características biológicas o que beneficiaria o desenvolvimento de estratégias mais eficientes para sua detecção.

Para acelerar e promover a identificação de MPs, após o Congresso da HUPO em Taipei no ano 2016, o C-HPP iniciou o “Next-MP50”, um desafio de 2 anos onde as 24 equipes do projeto trabalharam juntos para encontrar 50 MPs por cada cromossomo como um passo para completar a lista de partes do proteoma humano. Essa meta de 50 MP foi atingida para os cromossomos 19, 1, 5 e 17. Os cromossomos 2, 3, 10 e X têm 43 a 47 MPs a menos após o Congresso da HUPO em 2016. A partir de março do 2018, o C-HPP iniciou um projeto piloto para expandir sua missão e incluir a caracterização funcional das proteínas que tem identificações confiáveis mas que não têm função anotada no neXtProt (uPE1), referidas como proteoma escuro (Melaine *et al.*, 2018; Y.-K. K. Paik *et al.*, 2018). O projeto foi denominado neXt-CP50 (Y.-K. K. Paik *et al.*, 2018), onde CP significa "caracterização de proteínas" e visa determinar a função de até 50 proteínas uPE1 em três anos. Das equipes internacionais do consórcio C-HPP, 14 de 11 países se juntaram a este projeto: Cr 2 (Suíça), Cr 3 (Japão), Cr 4 (Taiwan), Cr 9, 11, 13 (Coreia), Cr 10, 17 (EUA), Cr 15 (Brasil), Cr 16 (Espanha), Cr 18 (Rússia), Cr 19 (México), Cr 20 (China) e Cr Y (Irã). Enquanto o progresso tem sido rápido na localização de MPs até agora, a taxa de novas identificações está diminuindo à medida que os frutos mais fáceis são colhidos. O número limitado de MPs encontrados a cada ano pelas equipes C-HPP e pela comunidade proteômica global reflete a crescente dificuldade em conceber e executar a descoberta profunda de MPs no proteoma humano. Os últimos 5% para a conclusão do projeto sempre são os mais difíceis de atingir. Embora vários cromossomos tenham menos de 50 MPs, alguns cromossomos têm mais de 200 MPs. Métodos mais sensíveis (incluindo técnicas de enriquecimento) e análises direcionadas de tecidos e células pouco estudados são necessários.

1.3. Metodologias para identificar MPs

Desde 2012, laboratórios em todo o mundo estão trabalhando com estratégias complementares para superar as dificuldades inerentes ao mapeamento das MPs como uns dos objetivos prioritários do C-HPP. Vários artigos desenvolveram novos fluxos de trabalho específicos que usam tecidos pouco estudados (rim, espermatozoides, testículos), preparação de amostras otimizadas, busca de dados combinados com validação por comparação com bibliotecas espectrais do SRMAtlas ou técnicas não baseadas na MS para detectar candidatos a MP.

Várias novas metodologias de processamento de amostras foram desenvolvidas com o objetivo de melhorar a eficiência na extração de proteínas, simplificar a complexidade biológica do material inicial e aumentar a cobertura do proteoma identificado. Por exemplo, Li *et al.* (S. Li *et al.*, 2017) usaram uma extração com Triton X-100/TBS junto com pérolas cobertas por hexapeptídeos (Proteominer (Hartwig *et al.*, 2009)) como estratégia para o enriquecimento/equalização de proteínas de baixa abundância, reduzindo a concentração de proteínas de alta abundância em quatro tecidos cancerosos (rins, bexiga, fígado e colorretal). Entre 20 candidatos de MP com dois peptídeos únicos com sequências não sobrepostas de ≥ 9 aa, 15 MP correspondem bem com os espectros de fragmentação e características cromatográficas de seus peptídeos sintéticos por PRM. Além disso, eles descobriram que uma alta porcentagem do total de proteínas identificadas com 1% FDR (nível de proteína) tem ≥ 2 peptídeos únicos de ≥ 9 aa.

Também, como parte do desafio neXt-MP50 Carapito *et al.* (Carapito *et al.*, 2017) usou uma estratégia baseada na solubilização e insolubilização das proteínas na solução de 1% de Triton X-100. Ambas as frações foram separadas por eletroforese e digeridas em gel. Para detecção das MPs usaram uma abordagem baseada em ensaios de PRM (com peptídeos sintéticos marcados) visando 92 peptídeos proteotípicos de 38 MPs dos Cr 2 e 14. Das 38 MPs 12 foram identificados com dois ou mais peptídeos e 3 com um peptídeo. A caracterização adicional por imuno-histoquímica em seções de tecido de testículos humanos e citoquímica em esperma foi realizada para 8 dessas MPs com anticorpos disponíveis no HPA (Carapito *et al.*, 2017).

Wang *et al.* (Y. Wang *et al.*, 2017) usaram uma estratégia multiprotease para a detecção de candidatos a MP de suas amostras de testículo previamente analisadas. Treze misturas de multiproteases (tripsina, lisargiNase, GluC) melhoraram a diversidade de peptídeos e a cobertura de sequência, em comparação com o uso de apenas tripsina. O maior número de candidatos a MP foi proveniente do uso de lisargiNase, que gerou peptídeos únicos e espectros ricos em íons b para 25 do total de 30 candidatos de MPs. No entanto, este trabalho reflete as dificuldades na pesquisa de MPs, pois de 7.838 proteínas, os autores não foram capazes de identificar nenhum candidato a MP que cumprisse totalmente as Diretrizes HPP.

Outros autores trabalharam com o fosfoproteoma de câncer de rim e o tecido não tumoral adjacente para detectar 75 candidatos a MP entre 8.962 proteínas e 6.415 fosfoproteínas identificadas. Após rigorosa triagem e verificação manual, 9 dessas 75 detecções de candidatos a MP foram submetidas à validação contra peptídeos sintéticos e filtragem de correspondências com variantes de aminoácidos individuais (*Single amino acid variants*, SAAVs); infelizmente, apenas uma MP (P81133, Cr 6q), foi confirmada (Peng *et al.*, 2017).

Waldemariam *et al.*, 2018 aplicaram uma análise proteômica de frações subcelulares de células-tronco embrionárias humanas (hESCs) para identificar as proteínas nucleares, citoplasmáticas e de membrana nestas células e para procurar as MPs de membrana. Usando essa metodologia o estudo conseguiu o maior conjunto de

dados com identificações confiáveis de 11.970 proteínas (1% de FDR para peptídeo, proteína e PSM) em hESCs, incluindo a descrição mais abrangente de 6.138 proteínas de membrana anotadas (Weldemariam et al., 2018). De acordo com as diretrizes da HPP, foram identificadas 26 MPs e 87 potenciais candidatos de MPs com um único peptídeo; dos quais 69 eram proteínas de membrana. Das MPs detectadas em 21 foi verificada a sua expressão por SRM ou por correspondência com peptídeos sintéticos no banco de dados de PeptideAtlas (Desiere et al., 2006).

Em 2016, haviam 1.860 proteínas PE1 com evidências não baseadas em MS; a partir da liberação de neXtProt 2020-01, muitas foram confirmadas por MS e esse número foi reduzido para 950. Destes, 73 são baseados no sequenciamento de Edman, 122 em mutações de doenças, 35 em estruturas 3D, 342 em interações proteína-proteína, 49 em técnicas baseadas em anticorpos, 127 em PTMs e 202 de estudos bioquímicos (Omenn, 2021). Uma das abordagens não baseada em MS usadas para caracterizar as MPs foi implementada por Meyfour *et al.* que usaram uma metodologia para a detecção da proteína TBL1Y, medindo a expressão diferencial de genes na região masculina do Cr Y (MSY) e suas contrapartes no Cr X durante a diferenciação cardíaca de hESCs. (Meyfour *et al.*, 2017). Foi usado anticorpos para as distinções das duas proteínas homólogas, que diferem apenas em um peptídeo tríptico de ≥ 9 aa. A expressão de TBL1Y mostrou um aumento durante a diferenciação, enquanto TBL1X diminuiu. A validação foi alcançada em nível celular usando ensaios de *knockdown* de gene durante a cardiogênese. Assim, a TBL1Y foi proposta para consideração pelo neXtProt como PE1.

O desenvolvimento de ferramentas bioinformáticas e a reanálise de dados proteômicos nos repositórios é também outra estratégia que contribui na identificação de MPs. Usando uma abordagem integrativa, Guruceaga *et al.* identificaram 5 candidatos a MP (FREM3, Cr 4; LAMB4, Cr 7; MYEOV, Cr 11; RAD21L1, Cr 20; e TLDC2, Cr 20) combinando os resultados de vários motores de busca (Comet, Mascot, OMSSA e X!Tandem) a partir de dados proteômicos de linhagens de células NCI60 (Guruceaga *et al.*, 2017). Este fluxo de trabalho se beneficiou da complementaridade dessas máquinas de busca para aumentar a cobertura do proteoma. Experimentos de biologia celular revelaram o envolvimento desses candidatos a MP em vários processos celulares nos tecidos onde são expressos (sistema nervoso, imunológico, muscular, secretor, reprodutivo e interno).

Choong *et al.* (Choong, Lih, Chen, and Sung, 2017) conduziram uma extensa análise comparativa de conjuntos de dados no “proteoma de pesquisa humana integrada em camadas do PeptideAtlas” (*Tiered Human Integrated Search Proteome*, THISP, versão 2017-03), para investigar sistematicamente a possibilidade de peptídeos únicos em MPs, peptídeos únicos em proteínas duvidosas e variantes de peptídeos afetados por 11 tipos de substituições isobáricas, trazendo resultados de identificações duvidosas. Eles concluíram que <5% dos peptídeos únicos de proteínas ausentes e <6% das variantes de peptídeos encontravam-se compartilhados com peptídeos de proteínas PE1 após substituições isobáricas.

O uso de ENCODE ou GENCODE no mapeamento do proteoma permite uma melhor compreensão da função do gene (Y. K. Paik and Hancock, 2012). Hwang *et al.* (Hwang *et al.*, 2017) usaram um fluxo de trabalho simplificado composto de um neXtProt personalizado, nomeado neXtPP e GENCODE para identificar MPs e variantes de *splicing* alternativo (*alternative splicing variants*, ASVs) com uma filtragem de alto rigor de FDR (0,1% no nível do peptídeo e 1% no nível da proteína) a partir de um tecido raramente estudado, o hipocampo humano (MSV000081385). A triagem dos conjuntos de dados relatados anteriormente foi depositada em PX (PXD007166). Usando esses recursos combinados, eles reivindicaram dois candidatos a MP (ARC, Cr 8; GRIK5, Cr 19) mais sete ASVs.

Cho *et al.* (J.-Y. Cho, Lee, Jeong, and Paik, 2017) criaram o Epsilon-Q, uma interface de análise automatizada para pesquisa usando bibliotecas espectrais e quantificação *label-free* de proteína. Esta ferramenta é uma versão atualizada do Combo-Spec Search (J. Y. Cho *et al.*, 2015) que melhora as funções desse algoritmo e da quantificação de proteína por *label-free*. O Epsilon-Q usa várias bibliotecas espectrais, com controle específico de FDR e integração dos resultados da identificação e quantificação *label-free*.

Encontrar o restante 7,2% das proteínas de PE2,3,4 exigirá um esforço futuro excepcional, incluindo amostragem cuidadosa de células/tecidos raros (Bell, Solis, Kizhakkedathu, Matthew, and Overall, 2019) combinados com melhor fracionamento de amostra e limites de detecção aprimorados. Proteínas de baixa abundância podem ser enriquecidas usando anticorpos antes da MS.

1.4. Tecidos importantes na pesquisa de MPs

Desde o início da HUPO a iniciativa de investigar o proteoma de vários tecidos do corpo humano foi um assunto importante. Um dos dois principais programas do Projeto Proteoma Humano é o B/D-HPP, seus principais componentes incluem plasma, rim/urina, cromatina epigenética, cardiovascular, fígado, cérebro, mitocondrial, glicoproteômica, células-tronco, organismos modelo, anticorpos humanos, câncer, diabetes, doenças infecciosas, olho e de padrões de proteínas. A análise transcriptômica e o mapeamento baseado em anticorpos de vários tecidos (por exemplo, fígado (Kampf *et al.*, 2014), cérebro (Sjöstedt *et al.*, 2015), rim (Habuka *et al.*, 2014), pâncreas (Danielsson *et al.*, 2014), pele (Edqvist *et al.*, 2015), tecido adiposo (Mardinoglu *et al.*, 2014), etc.) tem sido fundamental na descoberta de MPs.

1.4.1. Estudos em testículo

A análise da expressão específica de tecido é um recurso importante para a pesquisa de MP. Curiosamente, a análise proteogenômica do tecido de testículo mostrou que ele possui muitos genes específicos do tecido e abundantes transcrições de MPs (Yao Zhang *et al.*, 2015). As proteínas extraídas e separadas por tricina-SDS-PAGE e glicina-SDS-PAGE foram analisadas por nanoLC-MS/MS. Um total de 9.597 proteínas foram identificadas; destas, 166 foram listadas como MP (Yao Zhang *et al.*, 2015).

Usando a mesma metodologia de processamento da amostra Wei *et al.* (Wei *et al.*, 2016) compararam os dados obtidos em 2 espectrômetros diferentes conseguindo identificar um total de 8.526 e 7.537 proteínas no Q-Exactive HF e no LTQ-Orbitrap, respectivamente; destas, as proteínas de baixa abundância foram detectadas exclusivamente usando dados de Q-Exactive HF. Do total de 81 MPs identificadas, 74 e 39 proteínas foram listadas como MPs em conjuntos de dados de HF e Velos, respectivamente. Entre as MPs acima, 47 proteínas (43 PE2 e 4 PE3) classificadas como MPs confirmadas após verificação rigorosa com a correspondência de espectros e filtragem de variantes isobáricas. Além disso, e como continuação do trabalho de Wang (Y. Wang *et al.*, 2017) mencionado anteriormente, Sun *et al.* (Sun *et al.*, 2018) usaram a separação das proteínas em SDS-PAGE e fracionamento dos peptídeos por RP-HPLC em pH elevado e, para aumentar a cobertura de proteína, também aplicaram uma estratégia multiproteases. Um total de 11.558 proteínas foram identificadas, que é o maior conjunto de dados de proteoma para uma única amostra de tecido humano até agora. Com base neste conjunto de dados em grande escala, foram identificadas 14 MPs (PE2) pelo neXtProt (2018-01) após análise de qualidade de espectro, modificações pós-traducionais, filtragem de SAAV e correspondência de peptídeo sintetizado. A análise de expressão de tecido mostrou que 3 de 14 MPs eram proteínas específicas de testículo (Sun *et al.*, 2018).

1.4.2. Estudos em tireoide

Embora muitos artigos descrevam o tecido de tireoide como um dos candidatos para a descoberta de MPs, não há muitos trabalhos usando esse tecido para essa finalidade. De acordo com o HPA (Thul and Lindskog, 2018), ainda existem 29 MPs que são altamente expressas na tireoide. Para entender melhor o mecanismo de regulação da expressão gênica de MPs em diferentes órgãos ou tecidos, foi comparada a distribuição da abundância de mRNAs em 27 tecidos. Apesar do tecido testicular ter a maior expressão de MPs, o tecido de tireoide também pode servir como uma amostra importante para a busca de MPs (Yao Zhang *et al.*, 2015). Duek *et al.* (Duek, Bairoch, Gateau, Vandenbrouck, and Lane, 2016) verificaram se as MPs dos Cr 2 e 14 carecem de peptídeos trípticos únicos e detectáveis. Para isso digeriram teoricamente cada entrada de MP no neXtProt e filtraram para peptídeos entre 9 e 50 aminoácidos. Os autores descobriram que a proteína OTOS tem apenas um peptídeo tríptico único e pode ser detectada na tireoide e no cérebro de acordo com as informações de sequenciamento de RNA. A validação desta e de outras proteínas como esta exigiria, portanto, protocolos específicos e desafiadores usando outras proteases.

1.4.3. Estudos em placenta

É conhecido que 69% de todas as proteínas humanas são expressas na placenta e que 356 desses genes apresentam expressão elevada na placenta em comparação com outros tipos de tecido (Yu *et al.*, 2015). Em uma pesquisa no banco de dados PRIDE foi escolhido o tecido de placenta para selecionar um conjunto de dados potencialmente enriquecidos em MPs (Garin-Muga *et al.*, 2016). GenomewidePDB é um banco de dados

proteômico que integra dados de proteínas codificadas por cromossomos com dados transcriptômicos e outras informações de bancos de dados públicos. Como um trabalho piloto, foi construído uma base de dados com as 3.615 proteínas identificadas no tecido normal da placenta humana e destas, 20 foram classificadas como MPs (Jeong *et al.*, 2013). Usando uma estratégia de separação de peptídeos por precipitação em 98% de acetonitrila antes da RP-HPLC em pH básico foram identificadas 30 MPs com FDR<1% em várias amostras biológicas, incluindo 13 MPs em preparações de placenta humana (Yuanliang Zhang *et al.*, 2018). Em outro estudo, um extrato de proteínas de placenta foi passado através de uma coluna C18 SPE e as retidas foram posteriormente separadas por SDS-PAGE ou filtro de 50 kDa. Para a identificação de proteínas foram usados 2 modos de digestão diferentes: tripsina total e semi-tripsina. A estratégia resultou na identificação de 4 MPs com 8 peptídeos únicos. A identificação de 6 desses peptídeos únicos foram ainda apoiados por Monitoramento de Reações Paralelas (*Parallel Reaction Monitoring*, PRM), uma técnica de proteômica alvo-direcionada, confirmando a presença de 3 MPs (Q6NT89, Cr1; A0A183, Cr1; e Q6UWQ7, Cr19) (Lin *et al.*, 2019).

1.4.4. MPs no contexto da doença

Além do aperfeiçoamento de metodologias para a busca de MPs, a seleção de amostras alternativas como tecidos tumorais, infectados ou após uma lesão, pode ser chave neste sentido. A maioria dos estudos conduzidos para identificar MPs foi realizada em amostras de células/tecidos de indivíduos sãos, limitando a capacidade de detectar proteínas envolvidas em alguma doença (Segura, Garin-Muga, Guruceaga, and Corrales, 2017). Da mesma forma, as respostas do tecido a infecções bacterianas, virais e parasitárias podem aportar pistas sobre MPs associadas a regeneração ou reparo desses tecidos específicos. A integração do C-HPP com as equipes B/D-HPP nesta tarefa serão benéficas para ambos os consórcios porque o C-HPP pode fornecer novos ensaios para MPs ou isoformas proteicas cujo papel e função podem ser imediatamente estudados por equipes do B/D-HPP no contexto da saúde e doenças. Em um estudo proteogenômico realizado no tecido de testículo foi reportado que de 182 MPs confirmadas, 72% tem correlação com a ocorrência ou desenvolvimento de doenças tumorais (Yao Zhang *et al.*, 2015). Por exemplo, as proteínas de membrana ADORA3 e HLA-C, confirmadas no estudo anterior, foram sugeridas como biomarcadores para a detecção precoce de vários tipos de doenças graves, incluindo câncer e doenças infecciosas (Jenner and Rose-Meyer, 2006; Khakoo *et al.*, 2004). Em resumo, a proteômica desempenha papéis cada vez mais importantes na compreensão da biologia do surto viral, diagnóstico preciso e tratamento eficaz.

1.5. Zika Vírus

O vírus Zika é um arbovírus, pertencente ao gênero *Flavivirus* e à família *Flaviviridae* (Westaway *et al.*, 1985), primeiramente isolado em 1947 em macacos *rhesus* na Floresta Zika em Uganda (Dick, 1952). A infecção humana foi descrita pela primeira vez na Nigéria em 1954 (MacNamara, 1954); até a década de 1950 havia pouco menos de 20 relatos em humanos (Faye *et al.*, 2014). O vírus é transmitido por

mosquito, vetor também responsável pela transmissão de outras arboviroses como febre amarela, dengue e chikungunya (Kenney and Brault, 2014). *Aedes aegypti* e *Aedes albopictus* são relacionados aos casos de surto (Grard *et al.*, 2014) apesar de outras formas de transmissão, não mediadas por vetores, como a transmissão sexual (Foy *et al.*, 2011; Didier Musso *et al.*, 2015), transfusão sanguínea (D. Musso, Nilles, and Cao-Lormeau, 2014) e transmissão vertical (mãe para filho) (Besnard, Lastère, Teissier, Cao-Lormeau, and Musso, 2014) serem consideradas.

O vírus possui um genoma de RNA de fita simples não segmentado com sentido positivo com cerca de 11 Kb de comprimento. O genoma contém duas regiões não codificantes (5' e 3') e um único quadro de leitura aberta (ORF) que codifica para uma poliproteína 5'-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3' que é clivada em sete proteínas não estruturais (NS) e três proteínas estruturais: capsídeo (C), envelope (E) e precursor de membrana (prM) (A. Wang, Thurmond, Islas, Hui, and Hai, 2017). A proteína C é uma proteína de 13 KDa altamente básica que forma um complexo com o RNA viral no nucleocapsídeo. Uma protease semelhante à furina da célula cliva a proteína prM no complexo de Golgi à medida que a partícula viral sai. A superfície da membrana viral exibe as proteínas M e E, mas antes desta etapa, foi levantada a hipótese de que a proteína prM está envolvida no enovelamento da proteína E2. As proteínas NS estão relacionadas à evasão da resposta imune inata do hospedeiro e replicação viral (A. Wang *et al.*, 2017), entretanto, sua função específica ainda não foi elucidada (Zhu *et al.*, 2016).

1.5.1. Síndrome Congênita da Zika

A infecção por vírus Zika tem sido associada a algumas complicações neurológicas como à síndrome de Guillain-Barré e mais recentemente, como detectado no Brasil, à microcefalia (Oliveira Melo *et al.*, 2016; Schuler-Faccini *et al.*, 2016). Os primeiros estudos e casos de infecção pelo vírus em fetos e neonatos demonstraram consequências devastadoras incluindo lisencefalia, hidrocefalia, necrose, calcificação periventricular e cortical, além de microcefalia severa (Mlakar *et al.*, 2016; Schuler-Faccini *et al.*, 2016). Além das informações epidemiológicas, esses estudos indicam a capacidade neuro-invasiva do vírus Zika, além da capacidade de atravessar a barreira placentária e afetar o desenvolvimento neural intrauterino.

Os dados que associam microcefalia à incidência do vírus Zika foram reforçados pelo sequenciamento do genoma do vírus no líquido amniótico em dois casos de má formação fetal, com microcefalia, durante a gravidez por Calvet *et al.* (Calvet *et al.*, 2016). Os autores também demonstraram que o vírus compartilha filogeneticamente 97-100% de identidade genômica com vírus isolados durante o surto na Polinésia Francesa, indicando uma origem comum (Calvet *et al.*, 2016). Entretanto, não há na literatura científica nenhuma evidência causal confirmando que o vírus Zika é o único agente etiológico responsável pelo desenvolvimento de má-formações cerebrais em fetos humanos durante a gravidez. Dentro deste quadro, as estratégias proteômicas do tipo *discovery-driven* aparecem como fundamentais para elucidar as bases moleculares

envolvidas nas infecções pelo vírus Zika e identificar proteínas com potencial diagnóstico ou candidatos vacinais.

1.5.2. Proteômica no estudo do Zika

A proteômica é uma ferramenta poderosa para estudar a patogenicidade do vírus. O desenvolvimento de novas técnicas de fracionamento de amostra, métodos baseados em MS de alto rendimento e análises bioinformáticas dos dados tornaram possível a investigação sistemática das interações vírus-hospedeiro. A proteômica permite avaliar alterações dinâmicas na abundância de proteínas e modificações pós-traducionais que ocorrem em células infectadas (Macedo-da-silva, Marinho, Palmisano, and Rosa-fernandes, 2020). Usando células progenitoras neurais e a linha celular neuronal SK-N-BE2, Scaturro *et al.*, 2018 caracterizaram a resposta celular ante uma infecção viral da Zika por proteômica e fosfoproteômica. Nesse trabalho foi usada uma abordagem de purificação de afinidade acoplada a espectrometria de massa (*affinity purification coupled to mass spectrometry*, AP-MS) para identificar as proteínas do hospedeiro alvo do vírus. Foram identificadas 1.216 locais de fosforilação que são desregulados após a infecção pelo vírus zika (ZIKV), o que indica uma profunda modulação das vias de sinalização fundamentais, como AKT, MAPK – ERK e ATM – ATR e fornece informações sobre o mecanismo de interrupção da proliferação provocada pela infecção. Nesse estudo foram identificadas também 386 proteínas que interagem com o ZIKV e fatores do hospedeiro com funções conhecidas no desenvolvimento neuronal, defeitos retiniais e infertilidade (Scaturro *et al.*, 2018). Os experimentos de proteoma global também são usados para avaliar proteínas expressas diferencialmente após a infecção por ZIKV. Garcez *et al.* (Garcez *et al.*, 2017) infectou neuroesferas humanas derivadas de células-tronco pluripotentes para observar as alterações produzidas pelo vírus durante a infecção. Combinando a proteômica e a transcriptômica, mais de 500 proteínas e genes diferencialmente expressos foram associados à interrupção da progressão do ciclo celular, replicação viral e diferenciação neuronal (Garcez *et al.*, 2017). Aguiar *et al.* (Aguiar *et al.*, 2020). usaram uma abordagem transcriptômica, genômica e proteômica integrada em cérebros neonatais com síndrome do Zika congênita post-mortem (CZS). O grupo evidenciou que colágenos e seus mRNAs foram significativamente reduzidos assim como os produtos gênicos proteicos de adesão celular relacionados à orientação do axônio e à extensão da neurite foram regulados positivamente em cérebros com CZS (Aguiar *et al.*, 2020).

O presente trabalho apresenta a aplicação de diferentes abordagens proteômicas na busca de MPs em tecidos com alta expressão dessas proteínas e a análise diferencial do proteoma da placenta de mulheres infectadas com o ZIKV que desenvolveram fetos com microcefalia. O trabalho demonstra a utilidade da espectrometria de massas na caracterização de proteínas humanas e na compreensão dos processos moleculares alterados após a infecção pelo ZIKV.

2. Objetivos

2.1. Objetivos geral

2.1.1. Desenvolver metodologias proteômicas baseadas na extração, imunopurificação e fracionamento de proteínas para detectar *Missing Proteins* por espectrometria de massas

2.1.2. Identificar, quantificar e comparar as proteínas diferencialmente expressas nas placentas de mães infectadas pelo ZIKV sem ou com fetos microcefálicos empregando técnicas proteômicas.

2.2. Objetivos específicos

2.2.1. Imunopurificar duas MPs modelo (FAM189A1 e PGPEP1L) usando IgG anti-MP a partir dos tecidos de tireoide e testículo.

2.2.2. Identificar por espectrometria de massa e confirmar a presença de MPs a partir da separação de proteínas por SDS-PAGE.

2.2.3. Identificar MPs a partir de extratos com detergentes e reagentes caotrópicos e fracionados por RP-HPLC em pH básico

2.2.4. Proteômica comparativa de placentas controle com placentas Zika positivos de grávidas que deram à luz fetos sem ou com microcefalia.

2.2.5. Em função dos resultados sugerir e propor mecanismos de transmissão transplacentária do vírus e mecanismos que induzem a microcefalia nos fetos.

3. Materiais e Métodos

3.2. Coleção de amostras

As amostras de tecido de tireoides foram coletadas de pacientes a partir de biópsia, sob consentimento escrito, no Hospital Universitário Clementino Fraga Filho. A aprovação ética foi dada pelo CAEE 63219916.6.0000.5257; Número do Parecer: 1.885.591, Plataforma Brasil. Um total de 19 amostras foi coletado, abrangendo lesões benignas (hiperplasia nodular) e lesões malignas (carcinoma papilífero). As amostras de testículo e as amostras de tireoide foram armazenadas a -80 ° C até o processamento. Foram coletadas também três grupos de placentas: um contendo quatro amostras controle ZIKV negativo e sem microcefalia (Ctr); outro de duas placentas que foram positivas para ZIKV, mas o feto não tinha microcefalia (Z+); e um terceiro com duas amostras de placenta que eram positivas para ZIKV e os fetos tinham microcefalia (MC+). A infecção por ZIKV ocorreu entre as 12 e 18 semanas de gestação, para as amostras do grupo ZIKV+MC+ e a coleta da placenta ocorreu no nascimento após 32 semanas de gestação para ambas as amostras. Para detectar o RNA viral, o PCR foi realizado para ZIKV e CHIKV. Este último foi negativo em todas as amostras. Além disso, o ZIKV também foi detectado no cérebro, líquido amniótico e sangue. Os cérebros dos fetos ZIKV+MC+ também fazem parte de outra análise proteômica publicada anteriormente (Aguiar *et al.*, 2020). As amostras imediatamente armazenadas a -80 ° C foram fornecidas pela Dra. Adriana Melo, IPESQ, Campina Grande, Paraíba, Brasil. Este estudo (CAAE 52888616.4.0000.5693) foi aprovado pelo Comitê de Ética da FURNE com parecer favorável em 04.04.2019 pelo CEP-CONEP, Brasil.

3.3. Extração de proteínas

Foram feitas 3 extrações de proteínas usando 3 estratégias diferentes de extração: Triton X-100 0.5%/PBS 1X (tampão A); 7 M ureia/2 M tioureia/desoxicolato de sódio 2% (SDC)/100 mM TEAB pH 8,5 (tampão B); e SDS 10%/100 mM TEAB pH 8,5 (tampão C). Foram adicionados inibidores de protease e fosfatase (complete, Roche) a todas as soluções de acordo com as especificações do fabricante. Pesou-se 100 mg de tecido de testículo e tireoide. Após a pesagem do material foi adicionado 1mL de tampão de extração e as amostras foram homogeneizadas com o homogeneizador ULTRA TURRAX T10 (IKA, Alemanha). Foram feitos 3 ciclos de homogeneização de 30 segundos *on* e 1 minuto *off*. Antes do último ciclo foi adicionado mais 1 mL de tampão de extração. As amostras foram incubadas em banho de ultrassom em 10 ciclos de 2 min *on* e 2 min *off* e seguidamente foram centrifugadas a 14.000 g por 30 minutos. As proteínas nos sobrenadantes de cada extração foram quantificadas pelo método do Qubit descrito no (capítulo 1.5.1). As amostras foram armazenadas até seu uso a -80 °C.

O tecido da placenta (100 mg) foi homogeneizado usando ULTRA TURRAX T10 (IKA, Alemanha) em tampão B. As amostras foram centrifugadas a 10.000 g por 30 minutos a 4 °C; as proteínas no sobrenadante foram precipitadas com quatro volumes

de acetona fria por 16 horas a -20 °C e centrifugadas a 10.000 g. O sedimento foi lavado duas vezes com 400 µL de acetona fria a 100% e centrifugado durante 10 minutos. O precipitado foi deixado secar a 25 °C e ressuspensão em 7M ureia/2M tiourea/100 mM TEAB pH 8,5. A concentração de proteína foi quantificada usando o ensaio fluorimétrico Qubit de acordo com as instruções do fabricante (vide 1.5.1).

3.4. Imunopurificação de Missing Proteins

Duas abordagens com pérolas magnéticas foram usadas para a imunopurificação das MPs FAM189A1 e PGPEP1L nos tecidos de testículo, cérebro e tireoide.

3.4.1. Imunopurificação com esferas Dynabeads M-270 Epoxy (abordagem 1)

Foram preparados 2 tubos cada um com 4mg de pérolas magnéticas Dynabeads M-270 Epoxy (2.7×10^8 beads) (Invitrogen, USA). As pérolas foram lavadas 2 vezes com 1 mL de solução de lavagem (100 mM de fosfato de sódio pH 7.4) e agitadas vigorosamente por 1 minuto, seguido de agitação leve por 15 minutos. Seguidamente foi adicionado 1 µg de IgG anti-FAM189A1 e IgG anti-PGPEP1L, respectivamente em cada tubo e acrescentado com sulfato de amônio até uma concentração de 1 M. Após uma incubação de 72 horas a temperatura ambiente sob leve agitação, as pérolas foram lavadas com as seguintes soluções, PBS 1X, ácido acético 2%, 100 mM Tris pH 8.89 e finalmente com ácido acético 2% (Branco). Antes de adicionar as amostras, as pérolas foram condicionadas duas vezes com 100 µL de PBS 1X

Foram adicionadas sequencialmente 100 µg de proteínas totais de testículo e tireoide extraídas com o tampão A (vide 1.2) e 16 µg do PrEST (controle positivo). Os antígenos PrEST™ são fragmentos de proteína recombinante usados para a produção dos anticorpos incluídos como controles de ensaio positivo juntamente com o anticorpo correspondente. Para cada interação das proteínas com as pérolas, as amostras foram incubadas por 2 horas a 37 °C sob agitação. Após a incubação as pérolas foram lavadas 3 vezes com solução de lavagem e as proteínas que ficaram retidas pelos anticorpos eluíram-se com 100 µL de ácido acético 2%. As eluições foram armazenadas a -20 °C até análise por MS. Para condicionar as pérolas para as amostras seguintes, adicionou-se 100 µL de ácido acético 2%. Após o descarte do sobrenadante, lavaram-se 3 vezes com 1 mL solução de lavagem e 2 vezes com 100 µL de PBS 1X.

3.4.2. Imunopurificação com esferas Dynabeads M-280 Sheep anti-Rabbit IgG (abordagem 2)

Dez tubos com 500µg de pérolas Dynabeads M-280 Sheep anti-Rabbit IgG ($6-7 \times 10^8$ beads/mL) (Invitrogen, USA) foram separados para a imunopurificação das MPs. Seguidamente as pérolas foram lavadas 3 vezes com 1mL de PBS 1X e com agitação vigorosa por 1 minuto. O anticorpo IgG anti-FAM189A1 (4µg) foi adicionado a 5 tubos com as pérolas lavadas e o anticorpo IgG anti-PGPEP1L (4µg) foi adicionado aos restantes 5 tubos em um volume final de 250µL de PBS 1X. Os anticorpos anti-MP e as

pérolas foram incubados 12 horas a 4°C. Depois da incubação as pérolas foram lavadas 3 vezes com 1mL de PBS 1X. Antes da interação das amostras com os anticorpos, as proteínas extraídas com o tampão B (vide 1.2) foram precipitadas com 4 volumes de acetona e ressuspensas em 100mM de bicarbonato de amônio (BCA) e quantificado no Qubit (capítulo 1.5.1).

A imunopurificação foi feita paralelamente em todas as amostras. Para cada um dos 5 tubos com anticorpo anti-MP foi adicionado 100 µg de proteínas totais de tireoide, cérebro e testículo extraídas com tampão B, 16 µg de PrEST (controle positivo) e 200µL de PBS (branco) respectivamente. Seguidamente foram incubadas por 2 horas a 37 °C. Após a incubação do anticorpo com as amostras, as pérolas lavaram-se 3 vezes com 1 mL de PBS 1X. Para a eluição das proteínas que interagem com o anticorpo adicionou-se 100 µL de ácido acético 2%. As eluições foram armazenadas a -20 °C até análise por MS.

3.5. Eletroforese

Para a análise das amostras por eletroforese, utilizou-se um gel de poliacrilamida separador de 12,5% e um gel concentrador 4% de acordo com as condições padrão descritas por Laemmli (Laemmli, 1970). Como marcadores de massa molecular usou-se uma mistura de proteínas padrão (GE Healthcare, USA) composta por: fosforilase B (97,4 kDa), albumina de soro bovino (66,2 kDa), ovalbumina (45,0 kDa), anidrase carbônica (31,0 kDa), inibidor de tripsina (21,0 kDa) e lisozima (14,4 kDa). Para aplicação na eletroforese, 30 µg de proteínas totais de tireoide e testículo extraídas com o tampão A e B foram reduzidas com 10 mM de ditioneitol (DTT) por 1h a 30 °C e alquilados com 40mM de iodoacetamida (IAA) por 45 minutos ao escuro.

A eletroforese foi executada em sistema vertical mini-Protean II (Bio-Rad, USA) conectada a uma fonte de energia (Bio-Rad, USA). O experimento durou aproximadamente 45 min, foi realizado a 20 °C e uma corrente constante de 30 mA. A corrida foi interrompida quando o bromofenol atingiu a extremidade inferior do gel. Terminada a corrida, o gel foi colocado em solução de fixação etanol 40%/ácido acético 1%, durante duas horas, corado com solução de Coomassie R 350 (PhastGel Blue R) durante quatro horas, descorado com solução de metanol 30%/ácido acético 1% em duas lavagens de uma hora e armazenados em solução de 5% ácido acético.

3.6. Digestão e limpeza de amostra

3.6.1. Quantificação das amostras em Qubit

A quantificação de proteínas ou peptídeos, foi realizada pelo ensaio fluorimétrico Qubit® 2.0 (Invitrogen, USA), de acordo com as instruções do fabricante. Resumidamente, para cada microtubo, é adicionado 1 µL de cada amostra e mais 199 µL de solução de trabalho do Qubit contendo o reagente fluoróforo. Estes são agitados e incubados a temperatura ambiente por 15 minutos; e em seguida, a leitura da absorbância das amostras é realizada. A curva padrão foi obtida através da leitura de

três padrões contidos no kit, preparados também de forma semelhante às amostras, modificando-se apenas os volumes do padrão (10 µL) e da solução de trabalho do Qubit (190 µL).

3.6.2. Digestão em solução

As proteínas eluídas da imunopurificação foram secas em concentrador sob vácuo Speed-vac SAVANT (Thermo Scientific, USA) e ressuspensa com 20 mM BCA pH 8.0. Por outro lado, 300 µg de proteínas totais do tecido de tireoide procedentes da extração com tampão B, foram separadas para fracionamento off-line dos peptídeos trípticos. As proteínas totais e eluídas foram reduzidas com DTT 10 mM por 1 hora a 30 °C e alquiladas com IAA 40 mM a temperatura ambiente por 30 minutos ao abrigo da luz. As proteínas totais foram diluídas 1/10 com 50 mM TEAB pH 8.5 e a hidrólise das duas amostras foi feita adicionando tripsina (Promega, USA) na proporção de 1:50 (tripsina:proteína) e mantendo em incubação a 35 °C por 18 h. Os peptídeos trípticos foram secos no Speed-vac e estocados a -20 °C até a análise por nanoLC-MS/MS.

3.6.3. Digestão em gel

Para a digestão em gel as bandas na faixa de massa de 56 e 21 kDa, correspondente à massa molecular das MPs FAM189A1 e PGPEP1L respetivamente, foram cortadas do gel. Após a descoloração, as bandas foram incubadas com 1 ml de H₂O por 5 minutos, a solução foi removida e as bandas foram cortadas em cubos de aproximadamente 1 mm³. O gel foi desidratado com 1 mL de 30% acetonitrila (ACN)/20 mM BCA. Os tubos foram secos em centrífuga de evaporação a vácuo Speed-vac SAVANT (Thermo Scientific, USA) para remover a ACN residual. Os géis foram subsequentemente reidratados com uma solução de tripsina 12,5 ng/µL em 50 mM BCA e incubados durante 16 horas a 37 °C. Para a difusão e coleta dos peptídeos, foram adicionados 40 µl de 5% ácido fórmico (FA) e incubados com agitação por 45 min.

3.6.4. Limpeza dos peptídeos em ponteiras com matriz POROS R2

Para dessalinização e concentração, os peptídeos foram aplicados em minicolunas *stage-tips* confeccionadas manualmente usando resina Poros R2 (Applied Biosystems, USA). Inicialmente foi feita lavagem da coluna em acetonitrila 100%, seguida de equilíbrio em 0,1% ácido trifluoroacético (TFA), adição dos peptídeos a coluna pré-equilibrada, adição de 0,1% TFA e por fim eluição por meio de dois passos de ACN (50 e 70%). Os peptídeos eluídos foram colocados no concentrador a vácuo SAVANT (Thermo Scientific, USA) por aproximadamente 3 horas a 40 °C até que estivessem totalmente secos. As amostras foram ressuspensas em 20 µL de FA 0,1% e quantificadas com Qubit.

3.6.5. Digestão e Limpeza em S-Trap

1,2 mg de proteínas totais de tireoide extraídas com tampão C reduziram-se com 20mM DTT por 1h a 30°C e alquilaram-se com 40mM IAA por 45 minutos a temperatura ambiente e ao abrigo da luz. Adicionou-se ácido fosfórico em uma concentração final de

1,2% e 7 volumes de 90% metanol/100mM TEAB (tampão de ligação). As amostras foram colocadas em colunas de spin S-trap MIDI (PROTFI, USA) e centrifugadas a 4.000 g por 30 segundos. Depois de descartar a solução que passou pela coluna, as amostras lavaram-se 3 vezes com 9 volumes de tampão de ligação para remover o SDS. Para a digestão das proteínas incubou-se 24 µg de tripsina em um volume final de 350 µL de tampão TEAB 50 mM por 18 horas. Os peptídeos foram eluídos em 3 passos por adição de 3 soluções (250 µL de TEAB 50mM, 500 µL de 0,1% AF e 500 µL de 50% ACN/0,1% AF) seguido de centrifugação a 4.000 g por 60 segundos. No final as 3 eluições juntaram-se e concentraram-se em Speed-vac SAVANT (Thermo Scientific, USA) para o fracionamento em RP-HPLC em pH básico.

3.7. Fracionamento off-line em cromatografia de Fase Reversa de pH básico

As proteínas de tireoide procedentes das extrações com tampão B e tampão C foram hidrolisadas em solução segundo 1.5.2 e em S-trap segundo 1.5.5, respectivamente. Os peptídeos trípticos das duas extrações foram separados e fracionados em cromatografia de fase reversa em pH básico. A mistura dos peptídeos trípticos da extração com tampão B adicionou-se 10% ácido trifluoroacético (TFA) para precipitar o SDC do tampão. A amostra foi centrifugada a 10.000 g por 10 minutos e o sobrenadante coletado foi concentrado 5 vezes em Speed-vac SAVANT (Thermo Scientific, USA). 100 µg de cada amostra foi injetado no cromatógrafo HPLC UltiMate™ 3000 (Thermo Scientific, USA) utilizando uma coluna Gemini C18, 250 x 4,6 mm; 5 µm (Phenomenex, USA), fluxo de 0.5 ml/min; fases móveis A: 20mM BCA pH 8,5 e B: 90% ACN/20mM BCA pH 8,5; gradiente: 5%B durante 10 minutos; 5% a 50%B em 45 minutos; 50% a 95%B em 10 minutos; 95%B por 5 minutos. Para cada conjunto de misturas, coletaram-se 22 frações de 1.500 µL a partir do 2º minuto de corrida. As frações foram secas em Speed-vac e ressuspensas em FA 0.1% para análise por nanoLC-MS/MS.

3.8. Análise por espectrometria de massas – nanoLC-MS/MS

3.8.1. Aquisição Dependente de Dados

As amostras da imunopurificação e das digestões em gel foram analisadas por nanocromatografia líquida (nLC) no instrumento Easy-nLC 1000 (Thermo Scientific) acoplado a um espectrômetro de massas Q-Exactive Plus (Thermo Scientific). Foi carregado na coluna guarda (ReposilPur C18, 3 cm x 150 µm ID, x 5 µm) um volume de amostra de 20 µL com 250 bar como limite de pressão. Os peptídeos foram separados na coluna analítica (ReposilPur C18, 15 cm x 75 µm ID, x 1.7 µm). Para as amostras imunopurificadas usou-se um fluxo constante de 200 nL/min, fase móvel A: 5% ACN/0,1% FA (Solvente A); 95% ACN/0,1% FA (Solvente B) e um gradiente linear de 5-40% de solvente B em 100 minutos; de 40-100%B em 13 minutos e 7 min em 100%B. Para a análise da digestão em gel utilizou-se um fluxo constante de 300 nL/min e um gradiente linear de 5-35% de solvente B em 60 minutos; de 35-90%B em 5 minutos e 20 min em 100%B. Foi usado também uma lista de inclusão com 6 peptídeos da proteína PGPEP1L e 3 peptídeos da proteína FAM189A1 (Tabela I). Para ambos os tipos de

amostras a fonte de ionização por ESI foi configurada com uma voltagem de 2.6 kV e 250°C no aquecimento do capilar. O espectrômetro Q-Exactive Plus foi operado no modo de aquisição dependente de dados (*Data Dependent Acquisition*, Full-MS/DDA-MS2), com um tempo de aquisição de 120 minutos para a amostra imunopurificada e 85 min para a amostra digerida em gel; o espectro MS1 na faixa de 200–2000 m/z; resolução de 70.000 (200 m/z); tempo máximo de injeção de 100 ms; fragmentação dos 20 íons mais intensos no modo HCD com energia de colisão normalizada (NCE) de 30; resolução de 17.500 (m/z 200) e limite mínimo de intensidade de 8.0E4 na aquisição dos espectros MS2. Espécies carregadas com +1 ou com carga não determinadas foram excluídas da análise MS2.

Tabela I Lista de inclusão dos peptídeos FAM189A1 e PGPEP1L

| [m/z] | Carga | Proteína | Sequência |
|---------|-------|----------|----------------------------|
| 479.782 | 2 | PGPEP1L | AIILEQSGK |
| 788.782 | 2 | PGPEP1L | AIILEQSGKNQGYR |
| 526.287 | 3 | PGPEP1L | AIILEQSGKNQGYR |
| 716.378 | 3 | PGPEP1L | AIILEQSGKNQGYRDADIR |
| 870.069 | 3 | PGPEP1L | SFWPEGGVC*LPGSPDVLESGVC*MK |
| 1304.6 | 2 | PGPEP1L | SFWPEGGVC*LPGSPDVLESGVC*MK |
| 557.949 | 3 | FAM189A1 | TDQRPTPEFPATSK |
| 693.34 | 2 | FAM189A1 | AQLSPAGDPDTWK |
| 836.419 | 2 | FAM189A1 | TDQRPTPEFPATSK |

Para a análise das frações coletadas os peptídeos foram carregados em uma coluna de carga Acclaim Pepmap C18 2 cm x 75 µm ID x 3 µm (Thermo Scientific, USA) e separados em uma coluna EASY-Spray PepMap C18 50 cm x 75 µm ID x 2 µm (Thermo Scientific, USA), com 60°C na coluna. Foram aplicados 4 µL de amostra e carregado um volume de 20 µL com 600 bar como limite de pressão. Utilizou-se um fluxo de 300 nL/min com um gradiente de 10-34 %B em 70 min; de 34-45 %B em 10 min; de 45-95 %B em 5 min e o gradiente permaneceu em 100%B por 10 min. Para a ionização por ESI usaram-se uma voltagem de 1.9 kV e 250°C no aquecimento do capilar. A aquisição dos dados foi no modo Full-MS/DDA-MS2 por um tempo de 90 minutos. A faixa de massa do espectro MS1 foi de 350-2000 m/z centroide com resolução de 70.000 (200 m/z) e tempo máximo de injeção de 100 ms. Para fragmentação foram selecionados os 20 íons mais intensos no modo HCD com NCE de 30, resolução de 17.000 (m/z 200) e limite mínimo de intensidade de 8.0E4 na aquisição dos espectros MS2. Espécies carregadas com +1 ou sem assinalação de cargas foram excluídas da análise MS2.

Para a análise proteômica de placenta 2 µg de peptídeos foram analisados em três replicatas técnicas. As amostras foram carregadas em uma pré-coluna e separadas em outra embalada no laboratório com as mesmas características das usadas nas amostras imunopurificadas. Usou-se um gradiente de 5-25% B por 115 min, 25-32% B por 10 min, 32-45% por 7 min, 45-90 por 8 min e mantidos em 90% por 5 min, em 250 nL/min de fluxo. Os modos Full-MS/DDA-MS2 e polaridade positiva foram empregados

para análise de espectrometria de massa. Uma lista de exclusão dinâmica de 45 s e voltagem de pulverização de 2,6 kV foi definida para todas as análises. A resolução de 70.000 (200m/z), tempo máximo de injeção de 100 ms e a faixa de massa adquirida foi 375-1650 m/z. Foram escolhidos para fragmentação os 20 íons principais por HCD. Os parâmetros dos espectros MS2 foram 17.500 de resolução (200 m/z), com NCE de 28, limite mínimo de intensidade de 1,6E5.

3.8.2. Análise por PRM

Os peptídeos foram separados por nanocromatografia líquida com o mesmo método cromatográfico usado para as amostras imunopurificadas e digeridas em gel. Para a aquisição dos dados foi utilizada uma lista de 5 peptídeos da proteína FAM189A1 e 5 peptídeos da proteína PGPEP1L. Foram incluídos também, 40 peptídeos de 13 proteínas PE2-5 identificadas na análise Full-MS/DDA-MS2 das amostras digeridas em gel (Tabela II). A voltagem de *spray* e a temperatura do tubo de transferência de íons usados foram 2,6 kV e 250 ° C, respectivamente. Os parâmetros PRM foram: modo de aquisição positivo, com resolução 17.500 (200 m/z), janela de isolamento de 2 m/z e com NCE 27.

Tabela II. Lista de inclusão de peptídeos para a análise por PRM

| Massa [m/z] | carga [z] | NCE | Sequência |
|-------------|-----------|-----|-----------------|
| 349.1722 | 3 | 27 | QRQEEEAR |
| 391.7025 | 2 | 27 | MPPAGGPR |
| 393.2315 | 3 | 27 | SPTGGIKKPHR |
| 394.7425 | 2 | 27 | KLPFQR |
| 424.2191 | 2 | 27 | HYVDSVK |
| 460.7272 | 2 | 27 | THYADSVK |
| 462.7323 | 2 | 27 | WTFGQGTK |
| 476.2483 | 2 | 27 | FLSASVGDR |
| 479.782 | 2 | 27 | AIILEQSGK |
| 497.29 | 2 | 27 | LLIYGASTR |
| 526.2862 | 3 | 27 | AIILEQSGKNQGYR |
| 557.9494 | 3 | 27 | TDQRPTPEPFATSK |
| 558.2938 | 2 | 27 | ASNLQSGVPSR |
| 611.3091 | 2 | 27 | DYIALNEDLR |
| 645.8563 | 2 | 27 | VAVEGVDVIFSR |
| 648.3331 | 2 | 27 | SLVDSKAYADAR |
| 649.875 | 2 | 27 | LGLGNETVVQLR |
| 673.8059 | 2 | 27 | AEDTAVYYC*VR |
| 690.3586 | 2 | 27 | NTLYLQMNNLR |
| 693.3384 | 2 | 27 | AQLSPAGDPDTWK |
| 767.3759 | 2 | 27 | GLEWVSGVSWNGSR |
| 767.3759 | 2 | 27 | GLEWVSGVSWNGSR |
| 788.9257 | 2 | 27 | AIILEQSGKNQGYR |
| 817.4014 | 2 | 27 | LEEEAVFGADVLDQV |

| | | | |
|----------|---|----|---------------------------------|
| 823.4147 | 2 | 27 | FISVGYVDDTQFVR |
| 828.3992 | 2 | 27 | FQSAAIGALQEASEAY |
| 836.4205 | 2 | 27 | TDQRPTPEPFATSK |
| 852.9676 | 2 | 27 | LLIYSASNLQSGVPSR |
| 870.0675 | 3 | 27 | SFWPEGGVC*LPGSPDVLESGVC*MK |
| 930.9685 | 2 | 27 | DIQLTQSPSSLSASVGDR |
| 969.9649 | 2 | 27 | DIQMTQSPSFLSASVGDR |
| 971.5011 | 2 | 27 | EIVMTQSPPTLSLSPGER |
| 974.9786 | 2 | 27 | VNVDIINFGEDEVNTEK |
| 1010.978 | 2 | 27 | TLDNDC*EVLTTLPDTGR |
| 1023.557 | 3 | 27 | FLAGILKGVQC*EVQLVESGGGLVQPGGSLR |
| 1185.553 | 2 | 27 | GYEQHAYDGDYIALNEDLR |
| 1304.598 | 2 | 27 | SFWPEGGVC*LPGSPDVLESGVC*MK |

3.9. Análise de dados proteômicos

A identificação da proteína foi realizada usando o software Proteome Discoverer 2.4 (PD2.4, Thermo Scientific). Usou-se um banco de dados FASTA da UniProt (24/06/2019, sequências 96026) e o algoritmo SequestHT para identificação de proteínas. Como parâmetros de pesquisa definiram-se três clivagens perdidas e peptídeos semi-trípticos. Como modificação fixa foi incluída a carbamidometilação dos resíduos de cisteína (C) e para a modificação variável, aceitamos a oxidação da metionina, a acetilação da proteína N-terminal e desamidação da asparagina (N) e glutamina (Q). Para a análise do fracionamento foi incluído também uma busca com biblioteca espectral (696.692 precursores) disponível em <https://chemdata.nist.gov>. A tolerância de massa do precursor foi de 10 ppm e para o fragmento de íon de 0,1 Da. Para validação de proteínas e peptídeos, a taxa de descoberta falsa (*False Discovery Rate*, FDR) foi menor que 1% para peptídeos e 1% para proteínas; as proteínas foram agrupadas em proteínas principais usando o princípio de máxima parcimônia. Para o estudo das mudanças no proteoma da placenta infetadas pelo ZIKV as proteínas foram quantificadas relativamente pelo método livre de marcação (*label-free*). A quantificação das proteínas foi realizada usando o Cromatograma de Íons Extraídos (XIC) de acordo com a área de pico dos três peptídeos distintos mais abundantes de cada proteína. Foi usado o nó de Correspondência entre Corridas (*Match Between Runs*, MBR) com uma mudança de tempo de retenção de 5 min. Para verificar se o peptídeo único identificado corresponderia a variantes de aminoácidos únicos (*Single Amino Acid Variants*, SAAVs) de qualquer proteína ou substituições isobáricas foi usada a ferramenta *Peptide Uniqueness Checker* (Schaeffer et al., 2017).

Para a estratégia de PRM foi construída uma biblioteca espectral utilizando os espectros MS2 do Full-MS/DDA-MS2 dos protocolos de digestão em gel e imunopurificação (precursores 295 280), combinada com uma biblioteca espectral disponível em <https://massive.ucsd.edu/ProteoSAFe/static/massive-kb-libraries.jsp> derivada de mais de 30 TB de dados proteômicos humanos e incluindo peptídeos

sintéticos (2.154.269 precursores) no Skyline 20.2. A confirmação da presença do peptídeo alvo na amostra foi realizada através da comparação da relação entre as intensidades das transições selecionadas com o seu espectro de fragmentação presente nas bibliotecas espectrais (capítulo 1.7.2). Esta relação entre o padrão de fragmentação foi avaliada a partir dos valores de *dotp* fornecidos pelo software Skyline v.20.2. Este parâmetro correlaciona a intensidade das transições dos peptídeos da biblioteca espectral de referência e os peptídeos endógenos correspondentes identificados. O *dotp* foi calculado utilizando a correlação das 5 transições mais intensas.

3.9.1. Análise funcional das proteínas

A lista de proteínas “Master” identificadas no estudo da placenta, foi importada para o software Perseus (versão 1.6.12.0) para a análise dos dados. Foram utilizadas proteínas com FDR menor que 1% ao nível de proteínas. As proteínas contaminantes e as identificações com menos de 2 PSMs foram removidas. Apenas proteínas detectadas em pelo menos duas réplicas biológicas de cada grupo foram utilizadas para a análise. A média da abundância de proteínas foi convertida para log₂ e normalizada subtraindo a mediana da distribuição da amostra. Três grupos de *fold-change* normalizados (FC) foram criados para a análise de proteínas diferencialmente abundantes: Z+ vs Ctr, MC+ vs Ctr e MC+ vs Z+. As proteínas com uma alteração na sua abundância relativa inferior ou superior a 2 desvios padrões da população média foram consideradas diferencialmente abundantes. Para o enriquecimento dos processos biológicos do Gene Ontology (GO) nos três grupos foi utilizado o “DAVID Bioinformatic Resources 6.8” (<https://david.ncifcrf.gov>). Para complementar a análise de enriquecimento foi usado também o STRING (<https://string-db.org>). As vias metabólicas enriquecidas nas proteínas diferencialmente expressas foram obtidas nos bancos de dados Reactome Pathway (<https://reactome.org>) e o KEGG. Redes de anotações biológicas enriquecidas foram obtidas no software Metascape (<https://metascape.org/>) com um valor $p < 0,01$ e um fator de enriquecimento $> 1,5$ e Cytoscape 3.8 foi usado para visualização das redes de interação.

4. Resultados e Discussão

Metodologias para identificação de *Missing Proteins*

4.1. Estudo de expressão das Missing Proteins do Cr 15 e estratégia geral

De acordo com o neXtProt, o cromossomo 15 possui ainda 40 proteínas com evidência ao nível de transcrito (PE2) e 31 com expressão incerta (PE5) o que soma 71 “proteínas perdidas” (*Missing Proteins*, MP) potencialmente detectáveis por MS (Figura 1A) (Tabela complementar I). A maioria dessas proteínas tem um ou dois peptídeos tripticos únicos com mais de 9 aminoácidos e são proteínas de membranas, segundo as anotações do Uniprot, o que torna difícil a sua extração e detecção confiável por espectrometria de massas (Figura 1B e C).

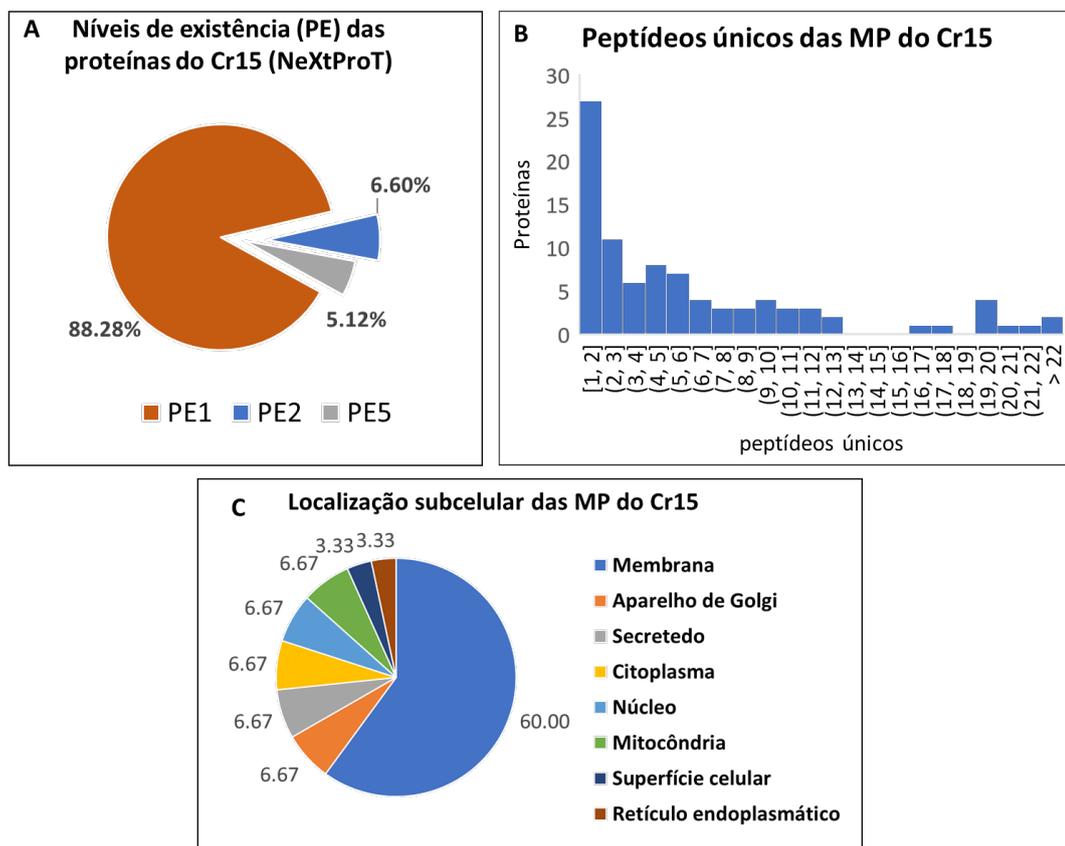


Figura 1: Categorização dos 606 genes do Cr15 de acordo com seu nível de evidência segundo neXtProt (revisado em junho de 2021) (A). Número de peptídeos únicos por *Missing Protein* (MP) do cromossomo 15 (B). Localização subcelular das 30 MP com anotações no banco de dados UniProt (C).

O estudo da localização e abundância de transcritos celulares pode fornecer pistas da expressão gênica das *Missing Proteins*. De acordo com dados de

sequenciamento de RNA do projeto de expressão genótipo-tecido (Genotype-Tissue Expression, GTEx), 21 MP do cromossomo 15 são altamente expressas em tecidos de testículo, 13 na glândula tireoide, e 16 em tecido de cérebro (Figura 2). Uma análise similar no Atlas de Proteínas Humano (Human Protein Atlas, HPA) mostra os dados transcriptômicos de expressão de 38 genes de MP do Cr15. O tecido com maior expressão de MP é o testículo e cérebro junto com a retina, a próstata, os gânglios basais e a glândula tireoide (Figura 3A).

A análise proteômica em profundidade desses tecidos especiais é um estudo crucial para a identificação de MP. Os genes do cromossomo 15 apresentam expressão enriquecida nos tecidos de testículo, cérebro e retina (Figura 3) segundo a classificação de Uhlén *et al.* previamente desenvolvida para perfis de mRNA (Uhlén *et al.*, 2015). De fato, os tecidos de testículo e cérebro expressam a maior quantidade de genes enriquecidos em tecidos (Djureinovic *et al.*, 2014; Uhlén *et al.*, 2015; D. Wang *et al.*, 2019). Muitos transcritos de mRNAs altamente expressos no tecido de testículo não puderam ser detectados ao nível de proteína, e nenhum outro tecido contém tantos mRNAs altamente expressos sem evidência de proteína (D. Wang *et al.*, 2019).

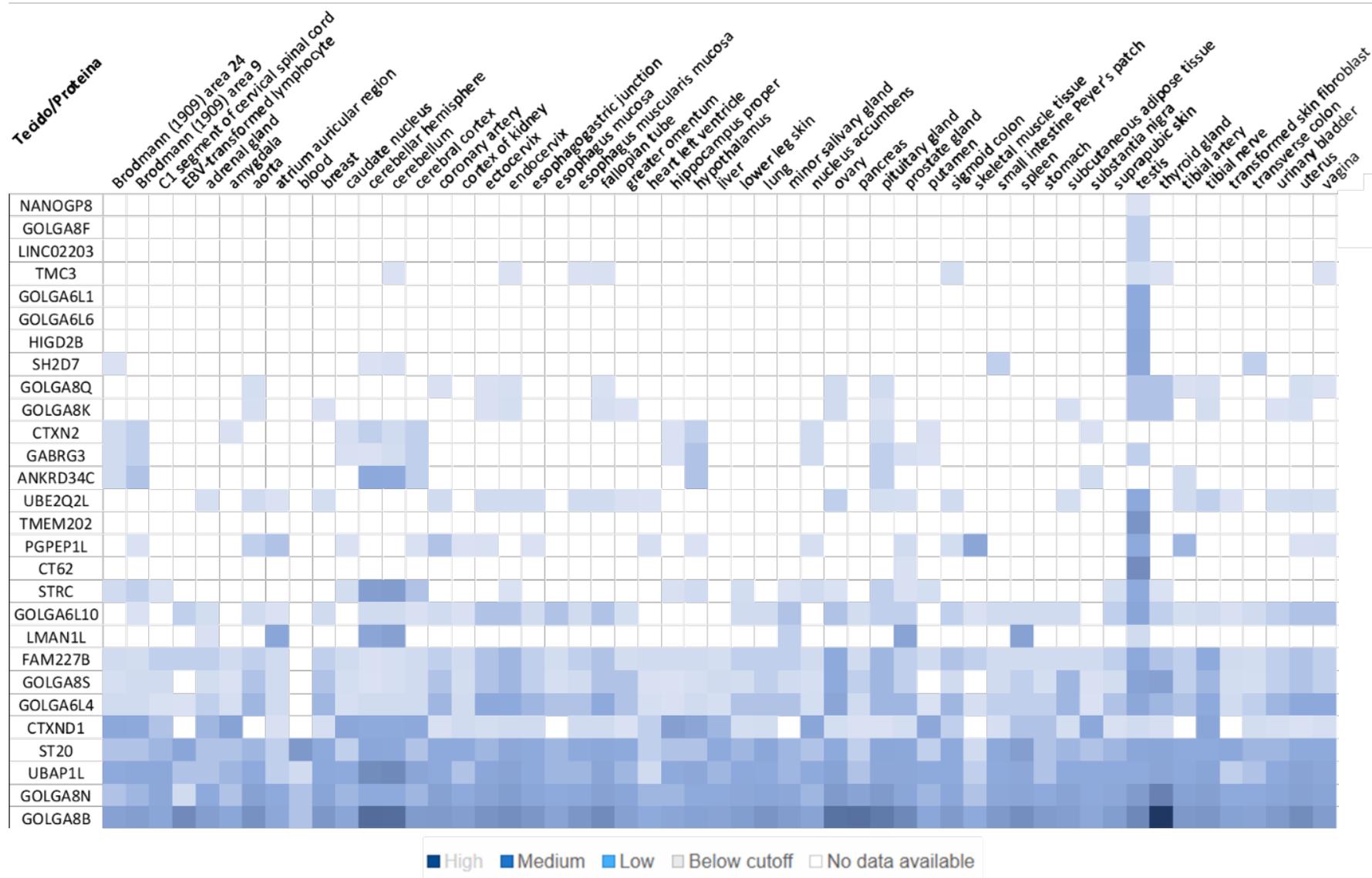


Figura 2: Dados de sequenciamento de RNA de 28 MP com nível PE2 do Cromossomo 15 expressados como Transcritos por Milhão (TPM) em 53 amostras de tecidos humanos oferecido pelo projeto de expressão genótipo-tecido (*Genotype-Tissue Expression project, GTEx*).

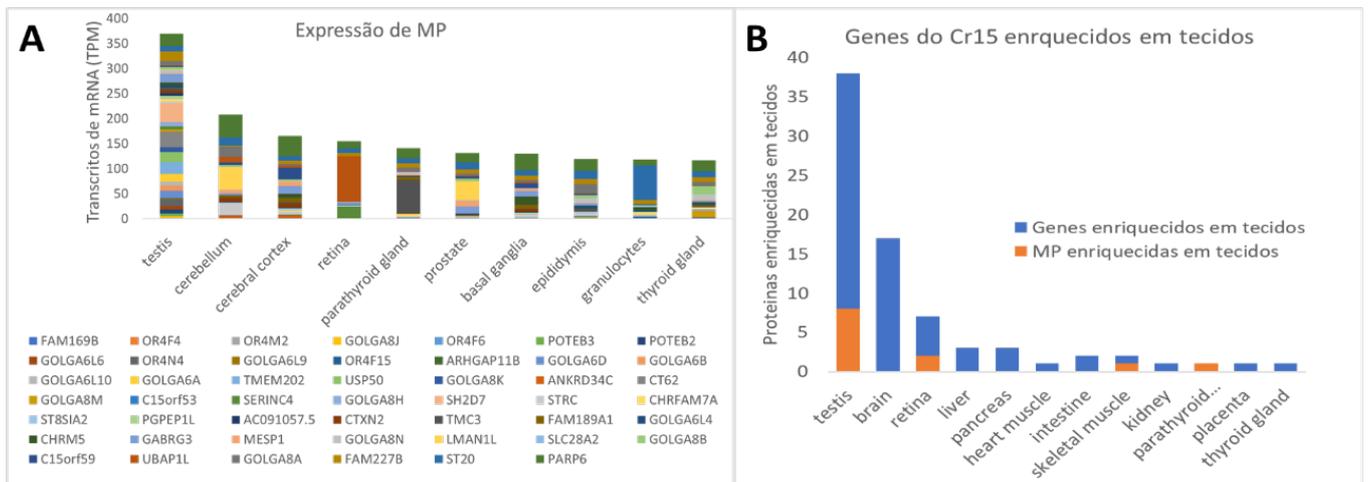


Figura 3. Níveis de expressão de RNA dos genes de cromossomo 15, de acordo com as anotações no HPA versão 20.1. (A) Os 10 tecidos com maior expressão de MP do cromossomo 15. (B) Os 10 tecidos com maior expressão de genes do cromossomo 15.

Com base nesses dados nós visamos criar estratégias de processamento que permitissem a detecção por espectrometria de massas de MPs focando a análise nas proteínas do cromossomo 15. Partindo de tecidos de tireoide e testículo foram feitas 3 extrações de proteínas com diferentes tampões.

Na primeira estratégia de extração usou-se o detergente Triton X-100 para a solubilização das proteínas hidrofóbicas. Duas MPs modelos do cromossomo 15, a proteína O60320 (proteína FAM189A1) e a proteína A6NFU8 (Proteína semelhante a piroglutamil-peptidase 1) foram imunoprecipitadas a partir dos extratos proteicos obtidos com Triton X-100. Foram usadas pérolas magnéticas Dynabeads M-270 Epoxy para a ligação covalente dos anticorpos anti-MP. As amostras de testículo, tireoides, antígenos PrEST (controle positivo) e uma solução de PBS 1X (Branco) adicionaram-se sequencialmente após realizar as respectivas etapas de eluição e condicionamento da matriz. As eluições foram analisadas por nanoLC-MS/MS no modo DDA. Além disso, as proteínas totais dos tecidos de tireoide e testículo extraídas como o Triton X-100 foram separadas em uma SDS-PAGE e as bandas na região de massa das 2 MP (21 kDa para A6NFU8 e 56 kDa para A6NFU8) foram digeridas em gel e analisadas por DDA-nanoLC-MS/MS e PRM-nanoLC-MS/MS (Figura 4).

Para a segunda estratégia de extração, uma mistura de ureia, tiourea e o detergente desoxicolato de sódio (SDC) foi usada como reagentes solubilizantes de proteínas de membranas. A mistura de proteínas obtida com este tampão foi processada usando diferentes abordagens. Em primeiro lugar, as proteínas totais de cada tecido foram precipitadas e ressuspensas em bicarbonato de amônio para imunopurificação das MP O60320 e A6NFU8. Neste caso, foram preparados separadamente 10 tubos diferentes (5 tubos por cada anticorpo anti-MP) com 500µg de pérolas Dynabeads M-280 *Sheep anti-Rabbit* IgG. Após o acondicionamento com os anticorpos, foi adicionado a cada tubo uma amostra diferente, de forma que a

imunopurificação realizou-se em paralelo. Usou-se os extratos de testículo, tireoide, uma solução de PBS 1X (branco) e um antígeno PrEST (controle positivo). Também as proteínas totais foram analisadas por SDS-PAGE e a banda na faixa de massa das MP de interesse foram digeridas e analisadas por nanoLC-MS/MS em modo DDA e as MP detectadas foram confirmadas por PRM. Além disso, as proteínas extraídas com o tampão ureia/tioureia/SDC foram hidrolisadas em solução com tripsina e a mistura de peptídeos foi separada em uma cromatografia de alto desempenho em fase reversa (*reverse phase high-performance liquid chromatography*, RP-HPLC) em pH básico. As 22 frações coletadas foram analisadas por espectrometria de massas em modo DDA.

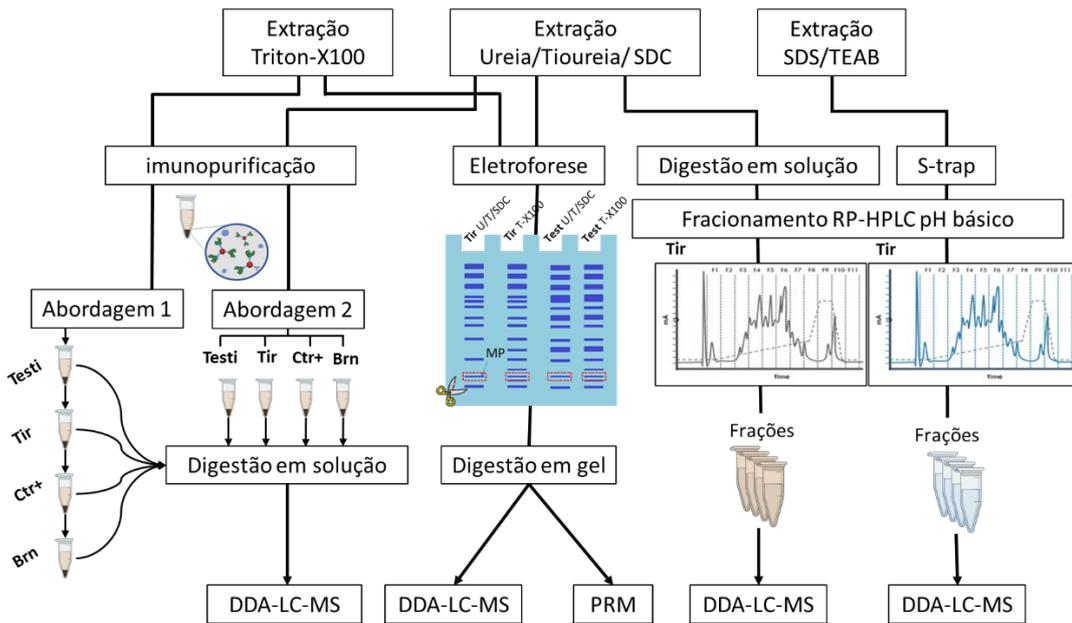


Figura 4: Fluxo de trabalho geral para a imunopurificação ou enriquecimento por SDS ou RP-HPLC de MP, começando com extrações de tecido do cérebro, testículo e tireoide usando 3 estratégias diferentes: Triton X-100, ureia/tioureia/SDC e SDS/TEAB.

A terceira estratégia de extração foi realizada aproveitando as características do SDS na solubilização das proteínas hidrofóbicas. Recentemente, um método de captura em suspensão de proteínas (*suspension trapping*, S-Trap) foi descrito por Zougman *et al.* (2014)(Zougman, Selby, and Banks, 2014). O protocolo permite a preparação de lisados de proteína contendo SDS em um curto período de tempo. As proteínas ficam presas em um material de filtração e o SDS residual é lavado em uma curta etapa de lavagem. As amostras são hidrolisadas no filtro usando a tripsina e os peptídeos eluídos foram fracionados por RP-HPLC em pH básico e analisados via nanoLC-MS/MS.

4.2. Imunopurificação de 2 Missing Proteins modelos

A espectrometria de massa fornece a plataforma ideal para a identificação e quantificação confiável de proteínas em uma amostra biológica complexa, no entanto a técnica apresenta uma desvantagem na detecção de proteínas pouco abundantes pois durante a aquisição dos dados, existe uma tendência para fragmentar os peptídeos de proteínas altamente expressas. Devido a que muitas MP podem estar em baixas

quantidades nas amostras por causa de uma baixa expressão ou por extração ineficiente (proteínas de membranas) a combinação da sensibilidade da detecção baseada em anticorpos com a especificidade da análise por espectrometria de massa pode ser uma estratégia viável na busca de MP.

4.2.1. Abordagem 1 de imunopurificação de Missing Proteins Modelos

Em uma primeira abordagem de imunoprecipitação, dois tubos contendo as esferas super-paramagnéticas Dynabeads® M-270 Epoxy (Invitrogen) foram incubados com 1µg IgG de coelho anti-O60320 e anti-A6NFU8 em cada um. Na segunda etapa, a suspensão contendo anticorpos acoplados covalentemente às pérolas (IgG-Dynabeads) foram incubados com solução de PBS 1X como amostra branco. Após a lavagem e a eluição de possíveis impurezas, as IgG-Dynabeads foram acondicionadas novamente e incubadas com 100 µg de proteínas totais de tireoide e testículo extraídas com Triton X-100. Esse detergente é um dos surfactantes não iônicos amplamente usado em protocolos de técnicas imuno-histoquímicas por sua capacidade de preservar a estrutura secundária da proteína minimizando os efeitos das interações hidrofóbicas não específicas durante a incubação do anticorpo com seu antígeno. Após a incubação das amostras com as IgG-Dynabeads para capturar as MP de interesse, as proteínas retidas foram eluídas e coletadas. Posteriormente, as IgG-Dynabeads acondicionaram-se novamente e foi adicionado um controle positivo PrEST. Os antígenos PrEST são fragmentos de proteínas recombinantes de 50 a 150 aminoácidos de comprimento, selecionados por conter epítopos únicos presentes na proteína nativa adequados para desencadear a geração de anticorpos de alta especificidade. As eluições foram secas e ressuspensas em bicarbonato de amônio 20 mM pH 8.0. A seguir, as proteínas foram digeridas com tripsina e analisadas por nanoLC-MS/MS no modo DDA.

Tabela III. Número de proteínas identificadas na primeira abordagem de imunoprecipitação de MP

| | Proteína O60320 | Proteína A6NFU8 |
|--------------|------------------------|------------------------|
| IP-Controle | 43 | 39 |
| IP-Branco | 17 | 52 |
| IP-Tireoide | 91 | 143 |
| IP-Testículo | 95 | 50 |
| total | 136 | 159 |

Foram identificadas 189 proteínas a partir das eluições da imunoprecipitação (tabela complementar II). A Tabela III mostra o número de proteínas identificadas em cada amostra analisada.

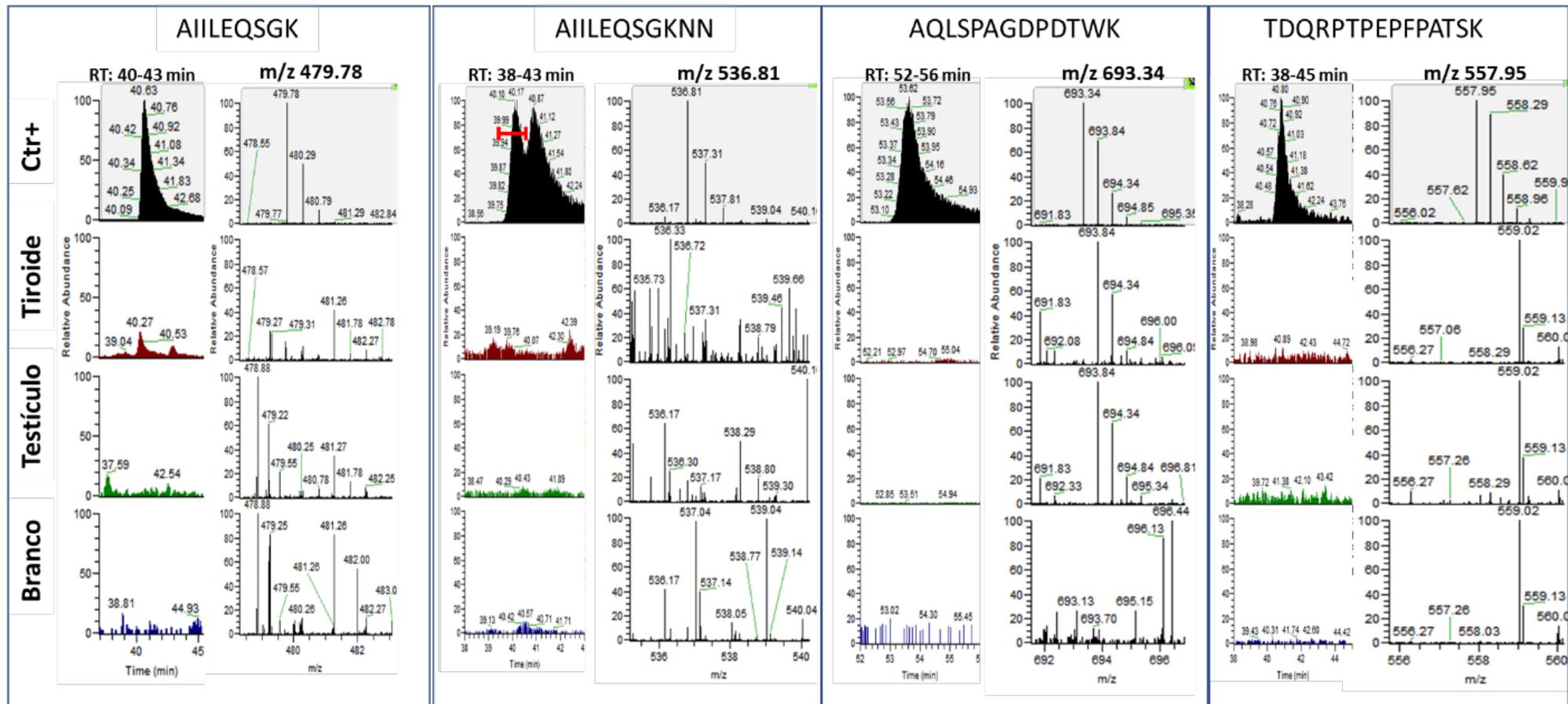


Figura 5. Comparação de diferentes cromatogramas e espectros MS1 dos peptídeos das MP modelos identificados nos controles da primeira abordagem de imunoprecipitação.

Usando esta abordagem não conseguimos detectar as proteínas O60320 e A6NFU8 nas amostras biológicas (Figura 5). A análise no *Peptide uniqueness Checker* mostrou que as proteínas O60320 e A6NFU8 produzem 16 e 11 peptídeos únicos, respectivamente. Como as duas MP contêm peptídeos exclusivos, a falta de detecção nas amostras de tecidos pode acontecer por não haver suficiente quantidade nas amostras ou a não ser expressas nesses tecidos. A captura das MP de interesse requer, então, a presença em níveis suficientes para detecção por espectrometria de massas o que depende, em parte, dos métodos de extração solubilizarem eficientemente as MP.

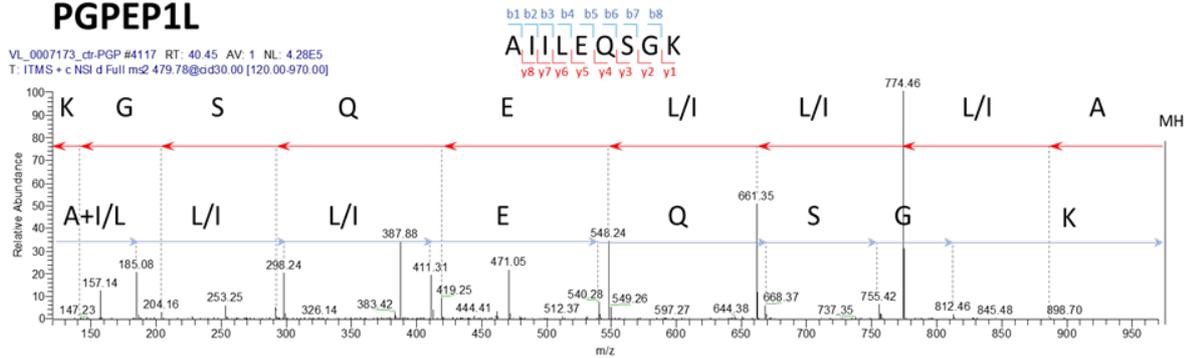
Tabela IV. Peptídeos do PrEST identificados na abordagem sequencial de imunoprecipitação de MP acoplada a espectrometria de massas.

| Proteína | Sequencia | M+H exp. | M+H Teo. | Erro (PPM) | amostra | Pept. Único | comprimento |
|----------|--------------------|----------|----------|------------|---------|-------------|-------------|
| O60320 | K.TDQRPTPEFPATSK.E | 1671.834 | 1671.834 | 5.57 | Ctr | TRUE | 15 |
| | R.AQLSPAGDPDTWK.T | 1385.673 | 1385.667 | 4.87 | Ctr | TRUE | 13 |
| A6NFU8 | K.AIILEQSGKN.Q | 1072.605 | 1072.6 | 4.97 | Ctr | TRUE | 10 |
| | K.AIILEQSGK.N | 958.5618 | 958.5568 | 5.26 | Ctr | TRUE | 9 |

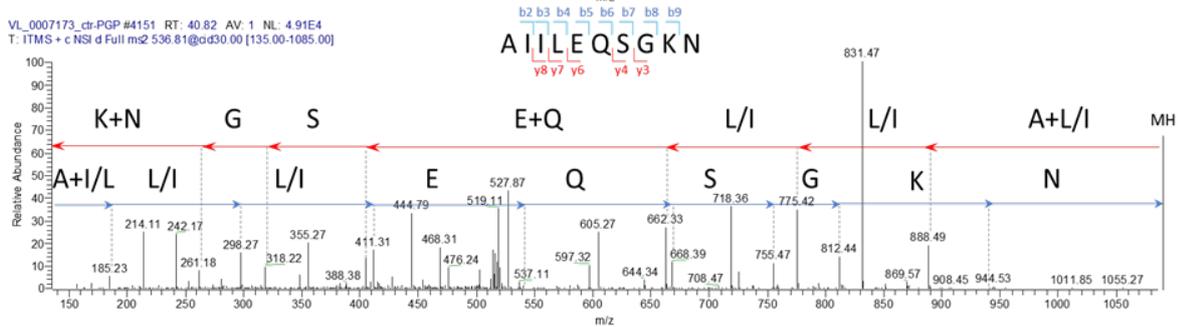
No entanto, peptídeos únicos foram sequenciados a partir dos PrESTs usados nos controles positivos (Figura 5 e Tabela IV), o que indica que a metodologia é capaz de enriquecer especificamente as regiões imunogênicas das proteínas. Para comprovar que a sequência dos peptídeos detectados é realmente única usou-se a ferramenta do neXtProt *Peptide uniqueness Checker* (Schaeffer *et al.*, 2017). A ferramenta leva em conta substituições isobáricas de isoleucina por leucina nos peptídeos identificados que não podem ser distinguidas pela espectrometria de massas e muitas vezes levam a identificações erradas, além disso a ferramenta também considera possíveis variações de sequência decorrentes de variantes de aminoácidos individuais (*Single Amino Acid Variants, SAAVs*) e *splicing* alternativo. Em todos os casos, os peptídeos detectados nos controles são únicos sem variantes de aminoácidos. Os espectros de fragmentação (espectros MS2) identificados analisaram-se manualmente para confirmar a correspondência da sequência detectada com a sequência teórica anotadas nos bancos de dados (Figura 6). Esses resultados foram utilizados para criar uma biblioteca espectral para a busca com outras metodologias.

PGPEP1L

VL_0007173_ctf-PGP #4117 RT: 40.45 AV: 1 NL: 4.28E5
T: ITMS + c NSI d Full ms2 479.78@cd30.00 [120.00-970.00]

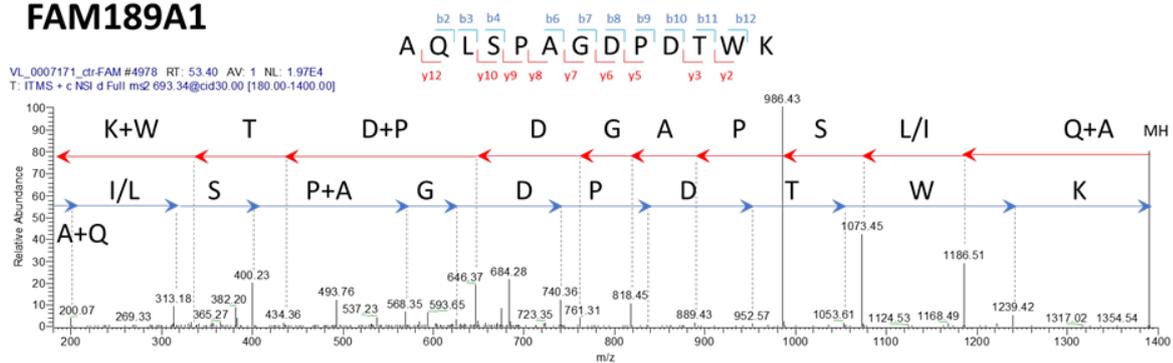


VL_0007173_ctf-PGP #4151 RT: 40.82 AV: 1 NL: 4.91E4
T: ITMS + c NSI d Full ms2 536.81@cd30.00 [135.00-1085.00]



FAM189A1

VL_0007171_ctf-FAM #4978 RT: 53.40 AV: 1 NL: 1.97E4
T: ITMS + c NSI d Full ms2 693.34@cd30.00 [180.00-1400.00]



VL_0007171_ctf-FAM #3836 RT: 40.53 AV: 1 NL: 2.80E4
T: ITMS + c NSI d Full ms2 557.95@cd30.00 [140.00-1685.00]

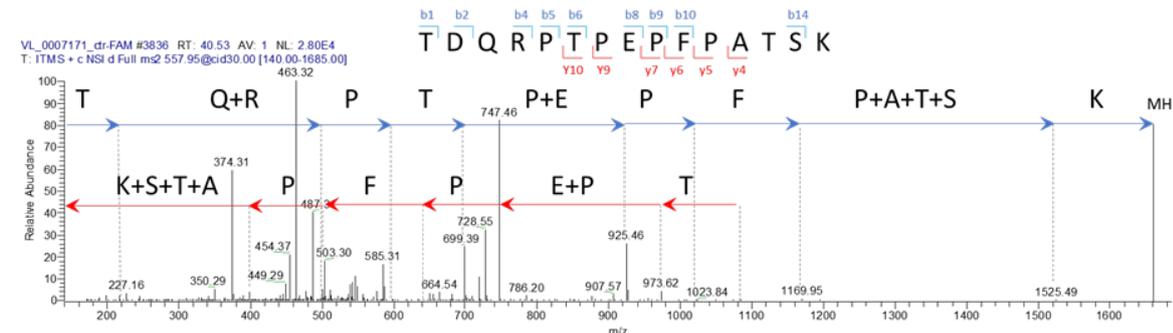


Figura 6. Sequenciamento dos peptídeos controles das proteínas O60320 e A6NFU8 a partir dos espectros MS2 obtidos na análise por nanoLC-MS/MS da imunoprecipitação.

Visando otimizar o protocolo de imunopurificação foram modificados alguns parâmetros para melhorar a extração das MP nas amostras biológicas e seu enriquecimento com os anticorpos correspondentes. Em um segundo processamento as proteínas foram extraídas com outro tampão e incubadas com maior quantidade de anticorpo, usando matrizes independentes para evitar a degradação dos anticorpos.

4.2.2. Abordagem 2 de imunopurificação de Missing Proteins Modelos

Em uma segunda abordagem foi usado um novo tampão de 7M ureia/2M tiourea/2% SDC para a solubilização das proteínas. O SDC é um detergente iônico solúvel em água comumente usados nos tampões de lise celular, e na preparação de lipossomas, isolamento de proteínas de membrana e lipídios e para evitar ligações não específicas em cromatografia de afinidade. Em comparação com o Triton X-100, o SDC desnatura as proteínas e liga-se às proteínas de membrana (hidrofóbica) e solúveis. Os reagentes caotrópicos como a ureia e tiourea interrompem as ligações de hidrogênio e as interações hidrofóbicas entre e dentro das proteínas levando ao desdobramento da estrutura secundária o que também contribui a solubilização. Para comparar o perfil de proteínas extraídas foi realizada uma eletroforese (Figura 7). É possível observar maior número de bandas de proteínas nas amostras extraídas com ureia/tiourea/SDC (faixas 2,3,5 e 6) comparado com o Triton X-100 (faixas 4 e 7), o que sugere que a combinação de ureia, tiourea e SDC consegue extrair mais proteínas que a solução com Triton X-100.

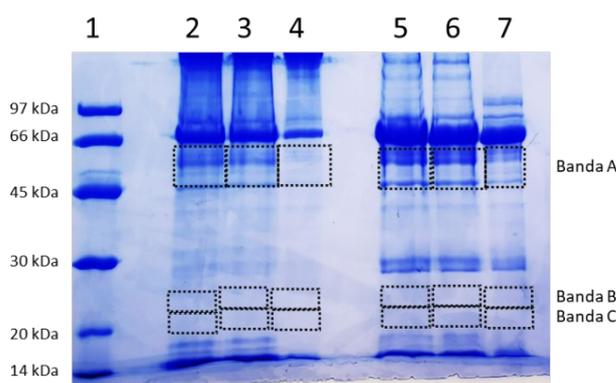


Figura 7. SDS-PAGE realizada a partir de extrações de proteínas em tecidos da tireoide e testículo usando tampões com ureia/tiourea/SDC e Triton X-100. 1: padrão de peso molecular; 2 e 3: extração da tireoide com ureia/tiourea/SDC, replicatas 1 e 2, respectivamente; 4: extração da tireoide com Triton X-100; 5 e 6: extração de testículo com ureia/tiourea/SDC, repetições 1 e 2, respectivamente; 7: extração em testículo com Triton X-100. as bandas A, B e C indicadas foram cortadas e digeridas em gel para análise por espectrometria de massas

Neste caso foram usadas pérolas DynaBeads M-280 *Sheep anti-Rabbit IgG* para a imunopurificação. A matriz possui anticorpos anti-IgG de ovelha covalentemente ligados à sua superfície. Esses anticorpos têm afinidade para todas as subclasses de IgG de coelho. A nossa estratégia consistiu em utilizar as pérolas como suporte para a ligação indireta das IgG anti-MP. Embora a interação com a matriz do anticorpo secundário anti-MP seja mais fraca do que na abordagem 1, a ligação específica do anticorpo pela região constante orienta a região variável em um arranjo adequado para a melhor interação com os antígenos (as duas MP de interesse). Cinco tubos com matriz foram incubados com cada IgG anti-MP. Seguidamente foram adicionados a cada tubo, as proteínas totais de testículo, tireoide e foram incluídos também uma amostra controle (PrEST) e uma solução de PBS 1X como branco. O processamento das amostras em paralelo evita a

reutilização da matriz, diminuindo a possibilidade de desnaturação, degradação ou perda do anticorpo. Além disso, uma maior quantidade de antígeno foi utilizada para reter mais proteínas de interesse. Depois da lavagem, as proteínas retidas foram eluídas com ácido acético 2%. As amostras foram secas e ressuspensas para hidrólise com tripsina e posterior análise por espectrometria de massa.

Identificamos 279 proteínas entre todas as amostras (Tabela complementar III). A Tabela mostra as identificações nas amostras analisadas depois de filtrar os principais contaminantes e identificações no banco de dados reverso (*decoy*). A maioria das proteínas detectadas nas amostras controles e branco são anticorpos liberados durante a etapa de eluição. O maior número de proteínas detectadas nas amostras de tecido se deve ao fato de haver mais interações inespecíficas ou indiretas de outras proteínas nas amostras com as MP alvos.

Tabela V. Número de proteínas identificadas na segunda abordagem de imunoprecipitação de MP.

| | Proteína O60320 | Proteína A6NFU8 |
|--------------|------------------------|------------------------|
| IP-Controle | 10 | 49 |
| IP-Branco | 26 | 43 |
| IP-Tireoide | 56 | 86 |
| IP-Testículo | 176 | 152 |
| total | 229 | 205 |

Foram detectados 12 e 3 peptídeos únicos a partir dos PrEST de A6NFU8 e O60320 respectivamente, nos controles (Tabela VI). Não foram observados peptídeos da proteína O60320 nas amostras biológicas, no entanto para A6NFU8 foram detectados 2 peptídeos na amostra testículo e um na amostra de tireoide. Curiosamente o peptídeo AIILEQSGK também foi identificado nos brancos, sugerindo que talvez os anticorpos utilizados estejam contaminados com os PrEST, apesar de lavar a matriz varias a matriz depois de ligado o anticorpo. Os anticorpos anti-MP usados neste trabalho são anticorpos policlonais de coelho desenvolvidos dentro do projeto HPA para o mapeamento do proteoma humano. Para o processo de purificação dos anticorpos são usados os antígenos PrEST alvos como ligantes de afinidade, o que pode trazer a co-eluição dos antígenos ou fragmentos deles junto com os anticorpos.

Tabela VI Resultado do segundo abordagem de imunopurificação das *Missing Proteins* FAM198A1 e A6NFU8.

| Proteína | Sequencia | M+H exp. | M+H Teo. | Erro (PPM) | amostra | Pept. Único | comprimento |
|-----------------|------------------|-----------------|-----------------|-------------------|----------------|--------------------|--------------------|
| A6NFU8 | IILEQSGK | 887.522 | 887.519 | 3.12 | Ctr | TRUE | 8 |
| | KAIILEQSGK | 1086.656 | 1086.651 | 4.13 | Ctr | TRUE | 10 |
| | VLESGVCMK | 1022.498 | 1022.5 | -1.91 | Ctr | TRUE | 9 |
| | IILEQSGK | 774.436 | 774.436 | 0.76 | Ctr | TRUE | 7 |
| | AIILEQSGK | 958.56 | 958.557 | 4.01 | Tireoide | TRUE | 9 |
| 958.558 | | 958.557 | 1.97 | Ctr | TRUE | 9 | |

| | | | | | | | |
|--------|---------------------------|----------|----------|-------|-----------|------|----|
| | | 958.559 | 958.557 | 2.89 | branco | TRUE | 9 |
| | | 958.559 | 958.557 | 3.26 | Testículo | TRUE | 9 |
| | | 958.559 | 958.557 | 2.96 | Branco | TRUE | 9 |
| | AIILEQSGKN | 1072.6 | 1072.599 | 1.07 | Ctr | TRUE | 10 |
| | AIILEQSGKNQGYR | 1576.854 | 1576.844 | 6.49 | Ctr | TRUE | 14 |
| | RVAVEGVDVIF | 1203.677 | 1203.673 | 3.93 | Testículo | TRUE | 11 |
| | RVAVEGVDVIF | 1203.677 | 1203.673 | 3.93 | Ctr | TRUE | 11 |
| | QLVVHVGMDTAAK | 1368.737 | 1368.73 | 4.96 | Ctr | TRUE | 13 |
| | SFWPEGGVCPLPGSPD | 1604.712 | 1604.704 | 4.65 | Ctr | TRUE | 15 |
| | SFWPEGGVCPLPGSPDVLESGVCMK | 2608.207 | 2608.188 | 7.45 | Ctr | TRUE | 24 |
| | VAVEGVDVIF | 1047.577 | 1047.572 | 5.12 | Ctr | TRUE | 10 |
| O60320 | AQLSPAGDPDTWK | 1385.67 | 1385.673 | 2.69 | Ctr | TRUE | 13 |
| | AYADARV | 765.389 | 765.3888 | -0.17 | Ctr | TRUE | 7 |
| | TDQRPTPEFPATSK | 1671.834 | 1671.834 | 0.09 | Ctr | TRUE | 15 |

Apesar de ter peptídeos do PrEST providos dos anticorpos contaminando as amostras de tecido, detectamos outro peptídeo (RVAVEGVDVIF) da proteína A6NFU8 que não foi identificado nas amostras branco e sim no controle (Figura 8A). O espectro de fragmentação do peptídeo no controle, serviu para confirmar a sequência de aminoácidos e para atribuir os sinais mais intensos detectados no espectro MS2 da amostra de testículo. Foram detectados no controle todos os íons fragmentos da série “b” e os íons “y2” e “y1” (Figura 8B, painel superior). Por outro lado, na amostra de testículo, foi encontrado o espectro MS2 com a mesma m/z e tempo de retenção detectados no controle (Figura 8B, painel inferior). Neste caso foram encontrados os íons “b6”, “b7”, “b8”, “b9” e “y1”. A detecção dos outros sinais não foi possível devido à baixa abundância do peptídeo na amostra. Durante a análise, foi descartada a possibilidade de detectar este peptídeo por *carry over* da amostra controle, uma vez que esta foi aplicado por última.

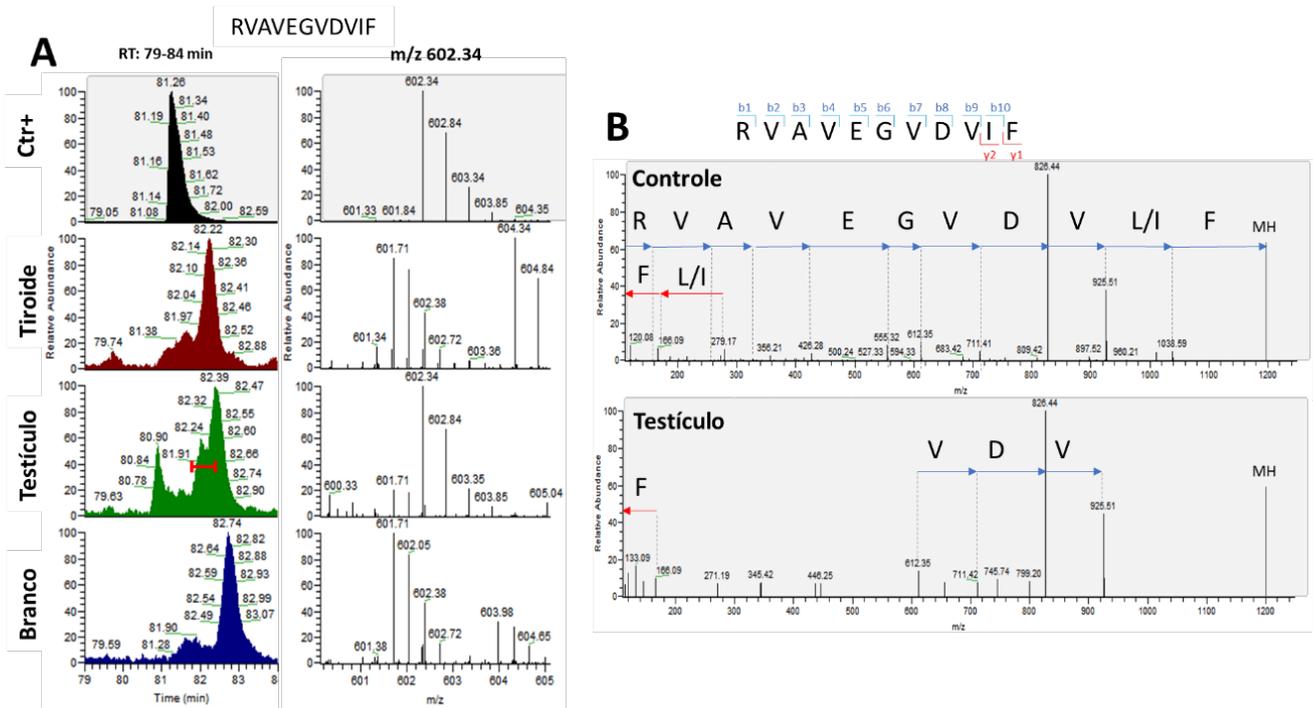


Figura 8. Análise do peptídeo RVAVEGVDF. (A) comparação dos cromatogramas de íon extraído e dos espectros MS1 da m/z 602,34. (B) comparação dos espectros de fragmentação detectados na amostra controle e na amostra de testículo.

4.3. Fracionamento de proteínas para identificar Missing Proteins

As estratégias projetadas para fracionar e simplificar a complexidade da amostra demonstram um aumento no número de proteínas identificadas. A probabilidade de detectar MP em amostras fracionadas é maior em comparação com a análise direta de extratos de proteína sem processamento.

4.3.1. Digestão de bandas de MP

Como estratégia ortogonal para o método de imunopurificação de MP, decidimos isolar as MP por SDS-PAGE, visando também identificar outras prováveis MP presentes nas amostras. A eletroforese desnaturante em géis de poliacrilamida é uma técnica ideal para o fracionamento de amostras biológicas devido ao seu alto poder de resolução. A SDS-PAGE separa as proteínas com base em seu volume molecular. O detergente SDS (dodecilsulfato de sódio) liga-se às proteínas e cobre-as com uma carga negativa. Em geral, uma molécula de SDS pode se ligar a dois aminoácidos. Após a exposição ao SDS, a carga negativa do detergente supera qualquer carga intrínseca que a proteína originalmente tinha, desdobrando sua estrutura tridimensional. A mistura de proteínas é desnaturada também pela adição de beta-mercaptoetanol, usado para quebrar as ligações dissulfeto. Em nosso caso, 30 µg de proteínas totais das amostras de tireoide e testículo extraídas com Triton X-100 e ureia/tioureia/SDC foram aplicadas em um gel de poliacrilamida (Figura 7). Terminada a corrida o gel foi corado usando o Coomassie R 350. Após a visualização das bandas do padrão de peso molecular, foram

cortadas 3 bandas na faixa de massa das MP, uma banda na altura de 56 kDa (O60320) e dois na altura de 21-25kDa (A6NFU8). As bandas foram hidrolisadas no gel usando a tripsina como enzima proteolítica. Para a aquisição dos dados no modo DDA usou-se uma lista de inclusão dos 3 peptídeos mais comuns das MP de interesse, identificados nos controles positivos da imunopurificação.

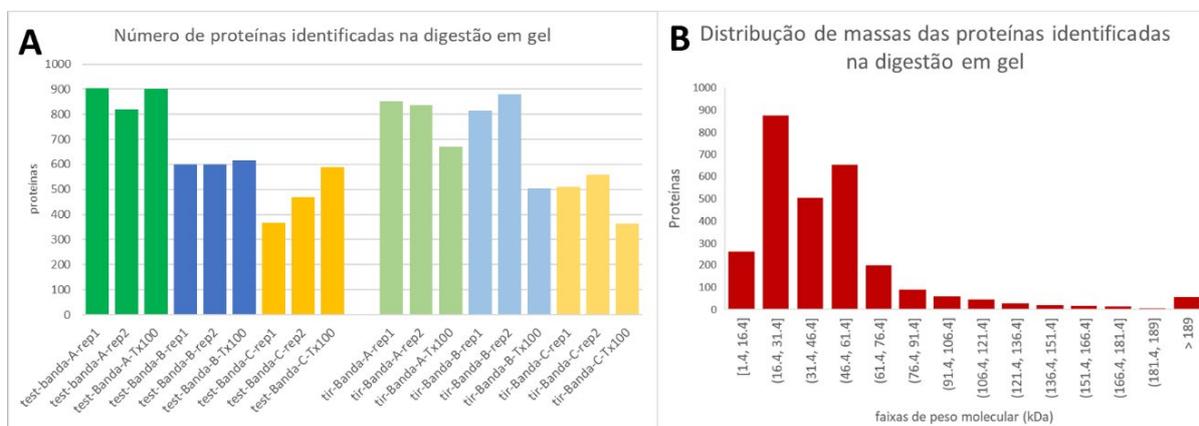


Figura 9. Análise das proteínas identificadas nas regiões de massa de 56 kDa e 21kDa a partir de uma digestão em gel. (A) Numero de proteínas identificadas em testículo (cores claras) e tireoides (cores pastéis) nas bandas A, B e C representadas pelas cores verde, azul e amarelo. (B) distribuição dos pesos moleculares das proteínas identificadas na hidrólise em gel.

A análise proteômica da digestão em gel das bandas de SDS-PAGE é mostrada na Figura 9 e a lista de proteínas identificadas encontra-se na tabela complementar IV Tabela II. De forma geral na banda A foi identificado maior número de proteínas do que nas bandas B e C, o que é consistente com a análise por SDS-PAGE. A coloração por Coomassie mostrou um maior número de proteínas na faixa de massa de 50 kDa em comparação com as regiões de baixa massa (Figura 7). Outra análise que mostra coerência entre a análise proteômica e a eletroforese é a distribuição da massa molecular das proteínas identificadas. Na Figura 9B, pode-se observar como a maioria das proteínas detectadas tem massa entre 57 e 21 kDa, que é justamente a faixa de massa cortada no SDS-PAGE.

Usando o fracionamento SDS-PAGE, as MPs de interesse não foram detectadas. Durante a hidrólise em gel ou nos passos de coloração, as proteínas podem difundir para fora do gel, principalmente as menos abundantes, o que pode afetar sua recuperação em etapas posteriores da análise. Também, peptídeos trípticos podem ser adsorvidos no gel e ficar presos em seu interior, afetando sua recuperação. Apesar de não termos identificado as duas MP modelos, conseguimos detectar 20 proteínas que não tinham evidência experimental por espectrometria de massas (PE4-5) a partir de 99 peptídeos, dos quais 43 foram considerados peptídeos únicos pelo Proteome Discoverer. Os peptídeos usados para identificação foram importados na ferramenta *Peptide Uniqueness Checker* (Schaeffer *et al.*, 2017) do neXtProt para garantir que eram exclusivos das suas respectivas proteínas. A análise com esta ferramenta resultou em 10 proteínas PE4-5 detectadas a partir de 16 peptídeos únicos (Tabela VII).

Tabela VII: Proteínas com nível de existência PE4-5 identificadas nas análises proteômicas das bandas de SDS-PAGE.

| Código de acesso | Nome da proteína | Nível de existência | # Pept únicos | # pept | # PSMs | cromossomo |
|------------------|--|---------------------|---------------|--------|--------|------------|
| A0A075B6H7 | Probable non-functional immunoglobulin kappa variable 3-7 isoform Iso 1 | PE5 | 1 | 4 | 21 | 2p11.2 |
| A0A075B6H8 | Probable non-functional immunoglobulin kappa variable 1D-42 isoform Iso 1 | PE5 | 1 | 1 | 4 | 2p11.2 |
| A0A0A0MT89 | Immunoglobulin kappa joining 1 isoform Iso 1 | PE4 | 1 | 1 | 2 | 2p11.2 |
| A0A0C4DH30 | Probable non-functional immunoglobulin heavy variable 3-16 isoform Iso 1 | PE5 | 3 | 6 | 79 | 14q32.33 |
| A0A0C4DH35 | Probable non-functional immunoglobulin heavy variable 3-35 isoform Iso 1 | PE5 | 2 | 5 | 77 | 14q32.33 |
| A0A0C4DH36 | Probable non-functional immunoglobulin heavy variable 3-38 isoform Iso 1 | PE5 | 1 | 1 | 8 | 14q32.33 |
| A2A3N6 | Putative PIP5K1A and PSMD4-like protein isoform Iso 1 | PE5 | 1 | 5 | 13 | 10q23.33 |
| P01893 | Putative HLA class I histocompatibility antigen, alpha chain H isoform Iso 1 | PE5 | 1 | 6 | 49 | 6p22.1 |
| P0DPK5 | Histone H3.X isoform Iso 1 | PE5 | 2 | 2 | 21 | 5p15.1 |
| P0DSN7 | Probable non-functional immunoglobulin kappa variable 1D-37 isoform Iso 1 | PE5 | 3 | 8 | 34 | 2p11.2 |

Por ser um método sensível e específico, o PRM foi adotado para confirmar os peptídeos identificados. Na técnica de PRM um íon precursor predefinido é selecionado no quadrupolo e transferido via C-trap para a célula HCD para fragmentação. O C-trap pode ser preenchido com íons por mais tempo, aumentando a relação sinal-ruído dos íons medidos no Orbitrap. Da célula HCD, os íons fragmentos são transferidos de volta para o C-trap e, eventualmente, injetados e analisados no analisador de massa Orbitrap. Como os espectros de fragmentação dos peptídeos selecionados são adquiridos com alta resolução e alta precisão de massa, os métodos alvo-direcionados baseados em PRM são altamente seletivos e específicos. Os ensaios PRM modernos permitem que várias proteínas sejam medidas simultaneamente. Idealmente a confirmação das proteínas de interesse precisa a utilização dos correspondentes peptídeos sintetizados quimicamente, o que permite a correspondência entre as intensidades de cada íon fragmento do peptídeo alvo com as intensidades dos íons fragmentos do peptídeo sintético. Neste trabalho, porém, usamos apenas os dados disponíveis nos resultados gerados das análises por DDA para selecionar as transições mais intensas. Para confirmar a presença das proteínas PE4-5 por PRM, foi elaborada uma lista dos peptídeos detectados no estudo por DDA (Tabela II). Peptídeos únicos foram incluídos na lista junto com 5 peptídeos ≥ 9 aminoácidos das 2 MPs modelos do cromossomo 15. A lista foi importada no programa Skyline para estabelecer o melhor método de PRM a ser usado. A partir das análises de imunopurificação e digestão em gel, foi construída

uma biblioteca espectral, contendo 295.280 precursores. Outra biblioteca espectral com 2 154 269 precursores procedente de peptídeos sintéticos, disponíveis na plataforma MassIVE do consórcio "Proteome Xchange" também foi usada. Foram selecionadas entre 5 e 7 transições tipo "b" ou "y" mais intensas desses peptídeos de acordo com seus espectros MS2 nas bibliotecas espectrais. Uma vez que o método foi otimizado, os géis digeridos foram analisados por PRM usando a lista de precursores exportada do Skyline. Neste caso, os sinais precursores foram adquiridos durante toda a corrida de forma que alterações no tempo de retenção dos peptídeos não afetem sua detecção no espectrômetro de massa.

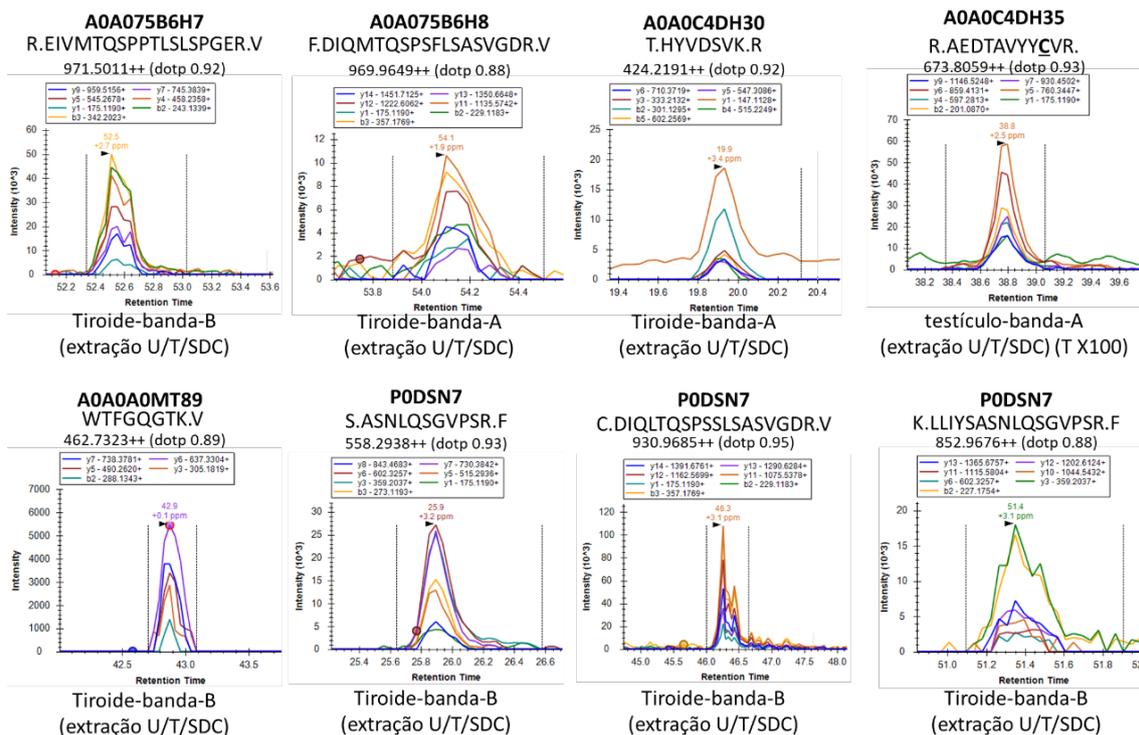


Figura 10. Resultados da análise por PRM das bandas de gel. Foram detectados 8 peptídeos únicos correspondentes a 6 proteínas com nível PE4-5.

A presença em tecidos testiculares e tireoidianos de 6 proteínas sem evidência experimental prévia ao nível de espectrometria de massa foi confirmada por PRM, a partir de 8 peptídeos únicos. A Figura 10 mostra as transições rastreadas por PRM e as amostras aonde foram detectadas. Em todos os casos seguiram-se 7 transições para cada peptídeo, com exceção do peptídeo WTFGQGTK que foi rastreado com 5 transições. Foram confirmados três peptídeos únicos com mais de 9 aminoácidos cada um para a proteína P0DSN7, cujo nível de existência era incerto (PE5). A maioria das proteínas foram detectadas nas amostras de tireoide extraídas com o tampão ureia/tioureia/SDC, o que novamente sugere que este tampão extrai maior número de proteínas do que a solução com Triton X-100. A separação na SDS-PAGE das proteínas extraídas com ureia/tioureia/SDC permitiu a detecção de 5 proteínas que não tinham sido detectadas por espectrometria de massas previamente.

4.3.2. Fracionamento em RP-HPLC em pH básico

Outra metodologia que permite a simplificação de amostras biológicas complexas é a cromatografia líquida de alta eficiência (HPLC). Esta técnica provou ser extremamente versátil para o isolamento e purificação de peptídeos. O HPLC de fase reversa (*Reverse Phase-HPLC*, RP-HPLC) continua sendo o modo mais amplamente usado de HPLC para separações e fracionamento de peptídeos. A disponibilidade de fases móveis voláteis a torna ideal para separações analíticas e preparativas. O RP-HPLC de pH básico é um método robusto de fracionamento que separa os peptídeos por hidrofobicidade e fornece excelente ortogonalidade com os gradientes de nanoLC-MS/MS de fase reversa de pH ácido. Em contraste com o fracionamento de troca catiônica forte (SCX), as frações de fase reversa de alto pH não requerem uma etapa de dessalinização adicional antes da análise de nanoLC-MS/MS. No nosso caso, o fracionamento ao nível do peptídeo permite uma maior aplicação da amostra inicial na coluna (100 µg em comparação com 30 µg aplicado no SDS-PAGE) e melhor recuperação dos peptídeos eluídos, o que aumenta a possibilidade de detecção de proteínas que estão em níveis baixos. Com base nos resultados da análise de PRM, extratos de tecido tireoidiano foram utilizados para o fracionamento RP-HPLC, uma vez que encontramos mais evidências de proteínas PE2-5 nessas amostras. Trabalhos futuros com tecidos de testículo podem ser desenvolvidos com base nas metodologias aqui descritas.

Com a finalidade de identificar MPs adicionais presentes na preparação de tireoide, os peptídeos trípticos gerados a partir da digestão em solução foram aplicados em um RP-HPLC de alto pH para fracionamento de peptídeos. Também comparamos os resultados desta análise com outro fracionamento a partir de uma extração usando SDS 10%/TEAB 100mM como tampão para solubilização de proteínas e uma coluna S-trap que captura as proteínas e permite a digestão e limpeza da amostra. O SDS é um surfactante aniônico e o detergente mais amplamente utilizado para solubilização de proteínas hidrofóbicas. No entanto, para a análise por MS é preciso adicionar uma etapa para removê-lo da amostra, pois não é compatível com o fracionamento por RP-HPLC ou a espectrometria de massas. A metodologia de S-trap consiste na criação de uma suspensão de partículas finas a partir de uma solução de proteína e SDS pela acidificação e adição de uma solução de metanol a pH neutro. As partículas de proteína são capturadas por um filtro que retém as proteínas, permitindo a passagem de moléculas de SDS que permanecem solúveis nessas condições. As proteínas são então lavadas e digeridas com tripsina e os peptídeos são transferidos para uma matriz C18, onde são limpos e concentrados para o fracionamento por RP-HPLC em pH básico (Figura 11A). Além da extração com SDS, foi fracionado também o hidrolisado das proteínas extraídas com o tampão de ureia/tiourea/SDC (Figura 11B). Antes da separação cromatográfica o detergente SDC foi removido da amostra por acidificação com ácido trifluoroacético 10%. Nos 2 casos foram coletadas 22 frações ao longo de uma corrida de 70 min. Cada fração foi seca e ressuspensa em ácido fórmico 1% para a análise por nanoLC-MS/MS no modo DDA.

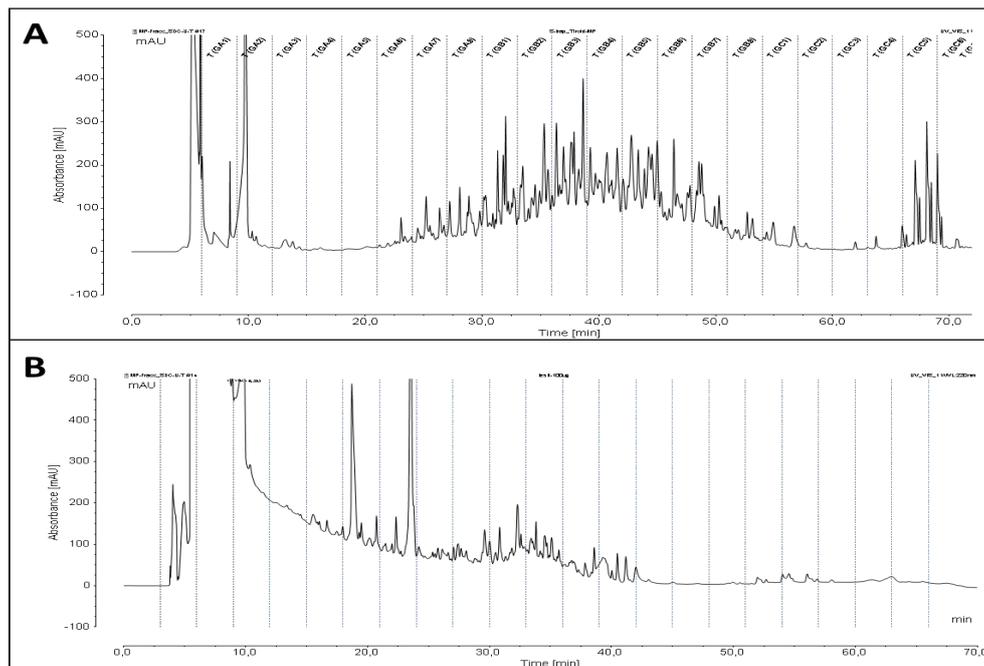


Figura 11. Cromatogramas do fracionamento em RP-HPLC em pH básico de peptídeos trípticos obtidos a partir de extratos proteicos de tireoide usando tampões com 10% SDS/TEAB 100 mM pH 8,5 (A) ou 7 M ureia/2 M tioureaia/2% SDC /100 mM TEAB pH 8,5 (B).

A análise proteômica relatou um total de 4356 proteínas, para o método de extração com ureia/tioureaia/SDC e digestão em solução (Tabela complementar V). Para o método de extração com SDS e digestão em colunas S-trap identificaram-se 4460 proteínas (Tabela complementar VI). Em ambas análises usou-se um FDR menor de 5% ao nível proteico. Em uma análise global das proteínas, é possível observar uma semelhança na distribuição dos pesos moleculares das proteínas identificadas nos dois métodos, com um pequeno aumento no número de proteínas com massas entre 10-20 kDa pelo método de S-trap (Figura 12A). No entanto, a maioria das proteínas (52,8%) foram detectadas por ambas as metodologias (Figura 12B). Para analisar a eficiência da digestão, comparamos o número de peptídeos com cortes inespecíficos da tripsina (Figura 12C). Usando o tampão ureia/tioureaia/SDC, foram identificados mais peptídeos deste tipo, provavelmente porque o método S-trap envolve um passo de limpeza e dessalinização dos peptídeos por C18. Durante esta etapa, muitos peptídeos grandes com clivagens não específicas são retidos na matriz e não são subsequentemente identificados por nanoLC-MS/MS. Finalmente, analisamos a eficiência da extração avaliando a localização subcelular das proteínas identificadas (Figura 12D). Ambas as metodologias apresentam um percentual semelhante de proteínas no mesmo compartimento celular.

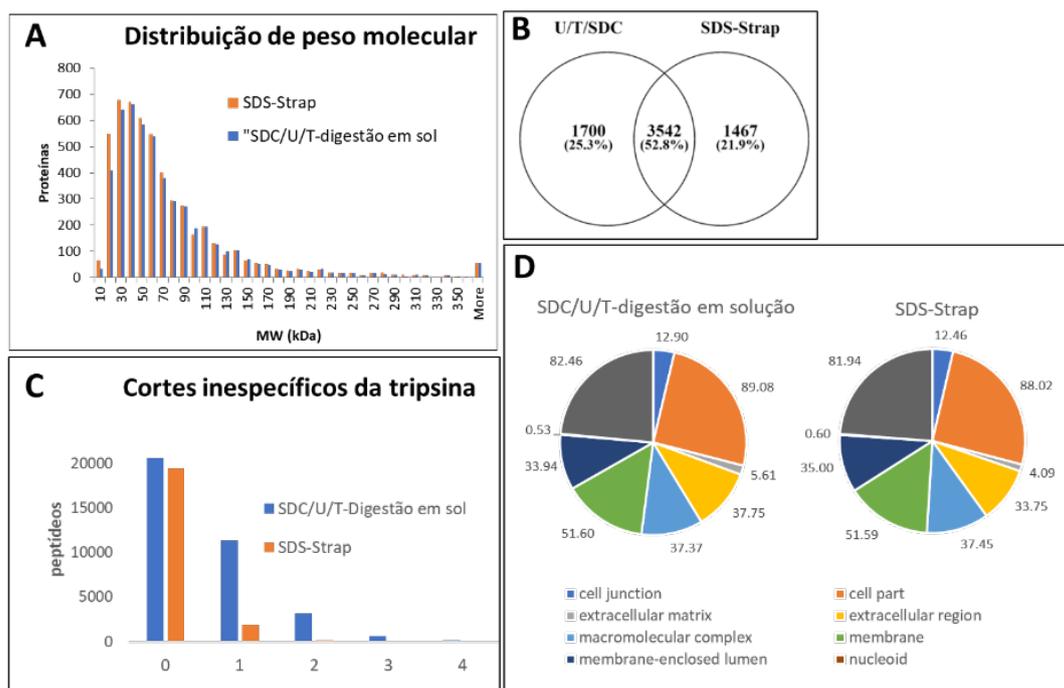


Figura 12. Análise proteômica do fracionamento por RP-HPLC em pH básico de extratos de proteínas procedentes de tecidos de tireoide. (A) Histograma do peso molecular das proteínas identificadas por 2 metodologias de extração y digestão: usando um tampão com ureia/tioureia/SDC e digestão em solução ou usando um tampão com SDS e digestão em S-trap. (B) Diagrama de Venn das proteínas identificadas entre os dois métodos. (C) Número de peptídeos identificados com cortes inespecíficos da tripsina entre as duas metodologias. (D) Comparação entre a localização subcelular das proteínas de tireoide identificadas por as duas metodologias.

As listas de peptídeos detectados com a metodologia de S-trap foram filtradas pelo "*peptide uniqueness checker*", para comprovar a detecção de alguma MP. Foram detectadas 5 proteínas com nível PE2-5 no procedimento S-trap, respectivamente (Tabela VIII). A partir do processamento com a coluna de S-trap e o fracionamento em RP-HPLC conseguimos identificar os peptídeos "S.CQPVLNSYQFHIVNGAVVCGCTLGPGASCHRLK.A"; "V.SLFYGAIMLMYVSPTPGNSVAMHK.L" e "N.GSTQTHFILLGFSDRPHLER.I" que mapeiam exclusivamente para as MPs Q5R387, Q8NGC8 e Q7Z3T1 de nível PE2 respectivamente. Todos os peptídeos apresentam uma sequência maior de >9 aminoácidos e vários íons fragmentos no espectro MS2.

Tabela VIII. Proteínas sem prévia evidência experimental identificadas na extração com o tampão de SDS/TEAB 100 mM, processadas por S-trap e fracionadas por RP-HPLC em PH básico.

| Código de acesso | Nome da proteína | Nível de existência | # Pept únicos | # pept | # PSMs | Cromossomo |
|------------------|--|---------------------|---------------|--------|--------|------------|
| P01893 | Putative HLA class I histocompatibility antigen, alpha chain H | PE5 | 1 | 7 | 14 | 6p22.1 |
| Q5R387 | Putative inactive group IIC secretory phospholipase A2 | PE2 | 1 | 1 | 1 | 1p36.12 |
| Q7Z3T1 | Olfactory receptor 2W3 isoform Iso 1 | PE2 | 1 | 1 | 1 | 1q44 |

| | | | | | | |
|--------|---|-----|---|---|---|---------|
| Q8NGC8 | Olfactory receptor 11H7 isoform Iso 1 | PE2 | 1 | 1 | 1 | 14q11.2 |
| Q9HAU6 | Putative translationally-controlled tumor protein-like protein TPT1P8 isoform Iso 1 | PE5 | 1 | 1 | 1 | 8p11.23 |

A análise das proteínas PE2-5 detectadas no protocolo com ureia/tioureia/SDC e fracionamento em RP-HPLC básico, mostrou maior número de identificações de MP comparado com o método de S-trap. Após a filtragem pelo "*Peptide uniqueness Checker*" foram detectadas 9 proteínas sem previa evidência experimental (PE2-5) (Tabela IX). As MP Q7Z572-2 e Q9H195 com nível de existência PE2 foram identificadas a partir dos peptídeos "R.MAQP GAVEHACHPSTLGGQGGGSLEEVRDIGE.R" e "TALSNSPHSRYLAEWPQGVPQLASPAPGHRENAPMTLTTSPH.D" respectivamente.

Tabela IX. Lista de proteínas identificadas sem evidência experimental por MS usando a extração com ureia 7M/tioureia 2M/SDC 10%, e o fracionamento de peptídeos por RP-HPLC em pH básico.

| Código de acesso | Nome da proteína | Nível de existência | # Pept únicos | # pept | # PSMs | Cromossomo |
|------------------|---|---------------------|---------------|--------|--------|------------|
| P01893 | Putative HLA class I histocompatibility antigen, alpha chain H | PE5 | 2 | 9 | 16 | 6p22.1 |
| P0C7V6 | Putative transcription factor SPT20 homolog-like 2 | PE5 | 1 | 1 | 2 | Xp22.11 |
| Q13434 | Putative E3 ubiquitin-protein ligase makorin-4 | PE5 | 1 | 1 | 1 | Xp11.4 |
| Q58FF6 | Putative heat shock protein HSP 90-beta 4 | PE5 | 1 | 5 | 20 | 15q21.3 |
| Q7Z572-2 | Isoform 2 of Spermatogenesis-associated protein 21 | PE2 | 1 | 1 | 3 | 1p36.13 |
| Q86T23 | Putative ciliary rootlet coiled-coil protein-like 1 protein | PE5 | 1 | 1 | 1 | 1p36.13 |
| Q8NHW5 | 60S acidic ribosomal protein P0-like | PE5 | 1 | 9 | 30 | 2p22.1 |
| Q9H195 | Mucin-3B (Fragments) OS=Homo sapiens | PE2 | 1 | 1 | 1 | 7q22 |
| Q9HAU6 | Putative translationally-controlled tumor protein-like protein TPT1P8 | PE5 | 1 | 1 | 1 | 8p11.23 |

As proteínas com níveis PE2-5 foram analisadas de acordo com suas anotações no Gene Ontology. As proteínas P01893, Q7Z3T1, Q8NGC8 e Q9H195 estão localizadas na membrana e a proteína Q5R387 é secretada ao exterior da célula. Curiosamente foi identificada a proteína Q9HAU6 por dois metodologias diferentes. O mesmo peptídeo de esta proteína foi sequenciado tanto na extração com SDS/TEAB e a digestão com S-trap como com a extração com ureia/tioureia/SDC e digestão em solução (Figura 13).

sequência: DLISHNEMFSDIYK, Charge: +2, m/z: 856.89874 Da, MH+: 1712.79021 Da, RT: 36.8623 min,
 Identificado por: Sequest HT (v1.17); XCorr:5.09,
 Tolerância de fragmento usada para pesquisa : 0.1 Da
 Fragmentos usados para pesquisa: -H₂O; y; -NH₃; y; b; b; -H₂O; y

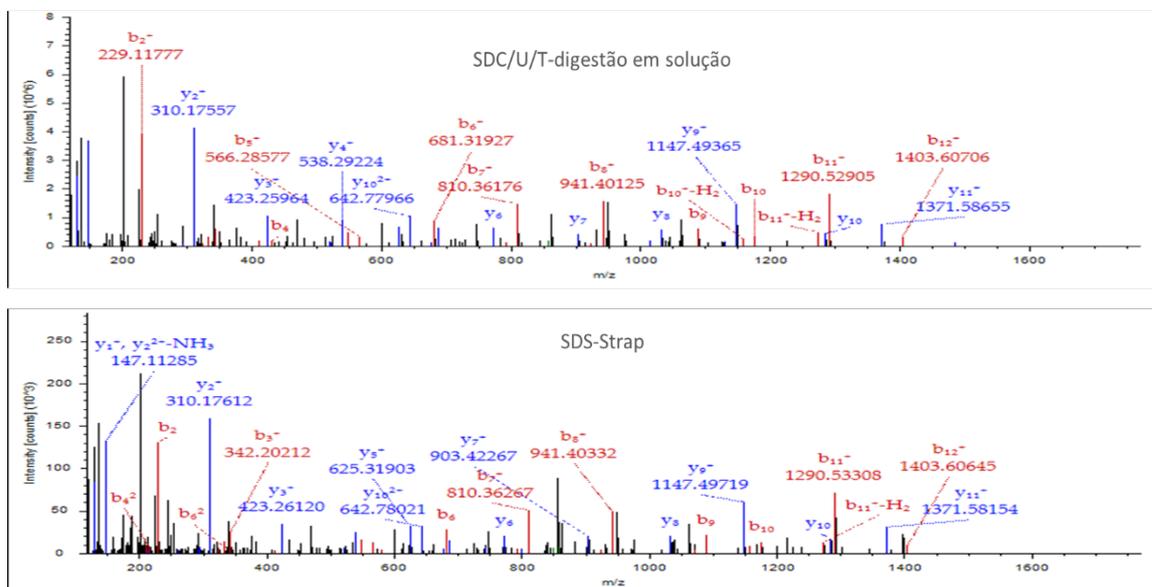
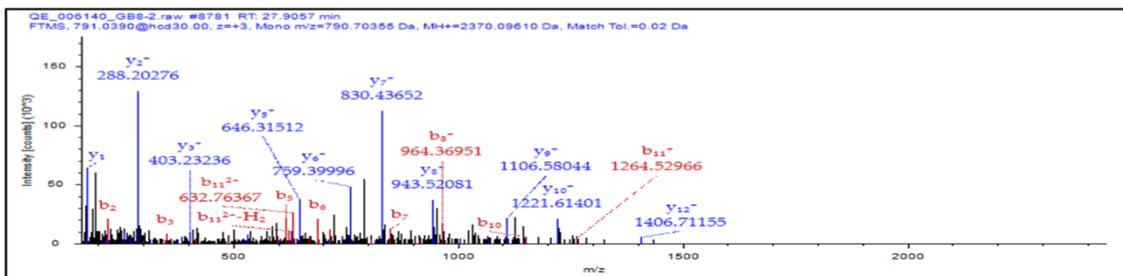


Figura 13. Espectro de fragmentação do peptídeo DLISHNEMFSDIYK correspondente à proteína Q9HAU6, detectado tanto pelo processamento com ureia/tioureia/SDC e digestão em solução (painel superior) como pela extração com SDS/TEAB e digestão em S-trap (painel inferior).

De acordo com as diretrizes de interpretação de dados de espectrometria de massa do projeto proteoma humano (Deutsch *et al.*, 2019), para afirmar novas detecções de proteína PE1 é preciso a identificação de pelo menos 2 peptídeos exclusivos das MP. No nosso trabalho foram identificados para a proteína P01893 (PE5), dois peptídeos exclusivos com mais de 9 aminoácido, cujas sequências não são sobrepostas uma com outra. A avaliação dos espectros de fragmentação dos 2 peptídeos que mapeiam para a proteína P01893 mostrou vários íons fragmentos da série “b” e “y”, o que confirma a presença de essa proteína nas amostras de tireoide (Figura 14).

sequência: GYEQHAYDGKDYIALNEDLR, Charge: +3, Monoisotopic m/z: 790.70355 Da, MH+: 2370.09610 Da, RT: 27.9057 min, Identificado por : Sequest HT (v1.17); XCorr:4.80, Tolerância de fragmento usada para pesquisa : 0.02 Da Fragmentos usados para pesquisa : -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y



sequência : YTC*HVQHEGLPEPLTLR, Charge: +4, Monoisotopic m/z: 513.25751, MH+: 2050.00820 Da, RT: 22.4365 min, Identificado por : Sequest HT (v1.17); XCorr:4.19, Tolerância de fragmento usada para pesquisa : 0.02 Da Fragmentos usados para pesquisa : -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y

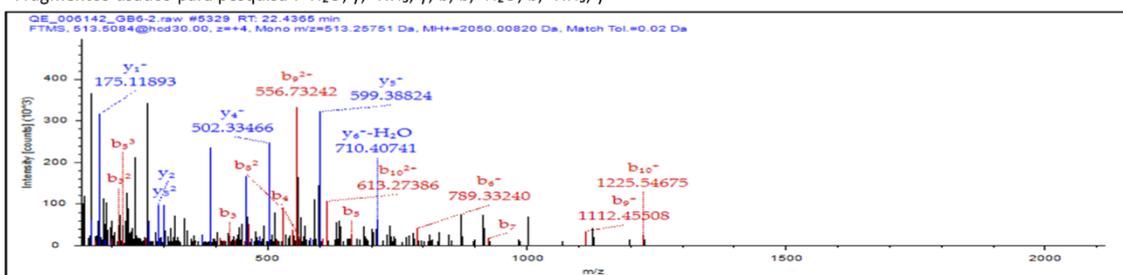


Figura 14. Espectro de fragmentação dos peptídeos GYEQHAYDGKDYIALNEDLR (painel superior) e YTC*HVQHEGLPEPLTLR (painel inferior) correspondente à proteína P01893 detectada no protocolo com ureia/tioureia/SDC e fracionamento em RP-HPLC em pH básico dos peptídeos tripticos.

Para validar estas proteínas ainda é preciso confirmar as identificações por técnicas alternativas como a proteômica alvo-direcionada usando peptídeos sintéticos. A proteína Q58FF6 com expressão incerta do cromossomo 15 também foi detectada pela digestão do extrato com ureia/tioureia/SDC e o fracionamento por RP-HPLC.

4.3.3. Identificação de Missing em tecido de placenta infectada com Zika

Para a análise proteômica do tecido da placenta as amostras foram processadas usando o tampão de ureia/tioureia/SDC. Neste caso as proteínas foram precipitadas com acetona para remover o detergente e seguidamente dissolveram-se em uma solução contendo 7M urea/ 2M thiourea/ 100 mM of TEAB pH 8.0. Após a digestão com tripsina os peptídeos foram carregados em um sistema nanoLC-MS/MS e analisados por espectrometria de massas no modo DDA.

As proteínas detectadas no estudo proteômico de placentas infectadas com ZIKV foram filtradas de acordo com seu nível de existência de acordo com o banco de dados neXtProt. A Tabela X mostra as proteínas com níveis de evidência de PE5 detectadas. quatro proteínas foram identificadas e seus peptídeos foram importados na ferramenta "Peptide uniqueness Checker"(Schaeffer *et al.*, 2017) para descobrir se eles realmente

mapeiam exclusivamente para as proteínas detectadas. Cada proteína tem pelo menos 1 peptídeo sequenciado único.

Tabela X. Proteínas sem previa evidência experimental ao nível de espectrometria de massas detectadas no estudo de placentas de mulheres grávidas infectadas com ZIKV durante a gravidez.

| Código de acesso | Nome da proteína | Nível de existência | # Pept únicos | # pept | # PSMs | Cromossomo |
|------------------|---|---------------------|---------------|--------|--------|------------|
| A6NMY6 | Putative annexin A2-like protein | PE5 | 1 | 33 | 1010 | 9p13.3 |
| P01893 | Putative HLA class I histocompatibility antigen, alpha chain H | PE5 | 1 | 8 | 66 | 6p22.1 |
| P59074 | Putative charged multivesicular body protein 4B-like protein CHMP4BP1 | PE5 | 1 | 1 | 4 | 14q22.3 |
| Q8NHW5 | 60S acidic ribosomal protein P0-like | PE5 | 1 | 12 | 204 | 2p22.1 |

4.4. Integração dos dados proteômicos para a detecção de proteínas sem evidência ao nível de espectrometria de massas

Em 2013, o HPP excluiu a categoria PE5 do denominador de todas as proteínas previstas, devido ao status dessas sequências como genes “duvidosos ou incertos”. Estudos relatam que, se expressas, essas proteínas seriam encontradas principalmente na membrana celular como transportadores e receptores (Dong et al., 2015). Muitas proteínas PE5 são pseudogenes que perderam sua capacidade de produzir proteínas funcionais devido à falta de características essenciais para transcrição e/ou mutações de sequência (Lane et al., 2014). Embora não exista a certeza da expressão das proteínas PE5, há relatos na literatura de casos de proteínas PE5 que foram confirmadas como PE1 (Dong et al., 2015; Kumar, Jain, and Dash, 2015). Esses dados sugerem que, além das características genômicas desfavoráveis, a não detecção dessas proteínas PE5 pode ser atribuída à dificuldade experimental em identificar proteínas e peptídeos ligados à membrana. A utilização de técnicas de enriquecimento com melhor solubilização e maior sensibilidade para proteínas da membrana pode ser necessária para a caracterização desse "matéria escura" no proteoma humano.

A detecção de proteínas de nível PE2-4 é a cada ano uma tarefa mais complicada e desafiadora, pois mais e mais proteínas estão sendo confirmadas por MS. Das 5587 proteínas PE2 anotadas em 2011, hoje existem apenas 1265 entradas a serem detectadas (neXtProt 2021-02-15), ou seja, há cada vez menos MPs a serem detectadas. No entanto, o número de proteínas PE5 tem se mantido relativamente constante, sendo 614 em 2011 para 601 em 2021, levando em consideração que esse número de proteínas é algo dinâmico e que algumas podem ter sido adicionadas à lista ou mudado de categoria. Vários fatores têm sido associados à falta de identificação das MPs restantes que ainda permanecem sem evidências, por exemplo, falta de peptídeos trípticos proteotípicos, alta homologia de sequência com outras proteínas PE1, níveis de transcritos indetectáveis ou expressão em certas condições (infecção, gravidez, etc.) ou tecidos específicos.

A análise conjunta de todos os dados obtidos neste trabalho confirmou a presença da proteína P01893 com um nível PE5 nas bases de dados neXtProt a partir de 4 peptídeos únicos em diferentes amostras e com vários protocolos. A Tabela XI mostra os dados para esses peptídeos, bem como as amostras onde foram sequenciados.

Tabela XI. Peptídeos únicos correspondentes à proteína P01893 detectados neste trabalho.

| Proteína | Confiança de FDR | Motor de Busca | Sequencia | M+H exp. | M+H Teo. | Erro (PPM) | amostra | metodologia |
|----------|------------------|----------------|------------------------|----------|----------|------------|--------------|--|
| P01893 | High | Sequest HT | YTCHVQHEGLPEPLTLR | 2050.013 | 2050.017 | -2.03 | placenta-Ctr | Digestão em solução |
| | High | MSPepSearch | GYEQHAYDGKDYIALNEDLR | 2370.095 | 2370.100 | -1.95 | Testículo | Digestão em gel |
| | High | Sequest HT | GGSYSQAASSNSAQGSDVSL | 1873.804 | 1873.805 | -0.59 | Tireoide | Digestão em coluna de S-trap e Fracionamento |
| | High | Sequest HT | GGSYSQAASSNSAQGSDVSLTA | 2045.891 | 2045.889 | 0.85 | Tireoide | Digestão em solução e fracionamento |
| | High | Sequest HT | GYEQHAYDGKDYIALNEDLR | 2370.096 | 2370.100 | -1.51 | Tireoide | Digestão em solução e fracionamento |
| | High | Sequest HT | YTCHVQHEGLPEPLTLR | 2050.008 | 2050.017 | -4.52 | Tireoide | Digestão em solução e fracionamento |

O peptídeo GYEQHAYDGKDYIALNEDLR foi identificado pelo mecanismo de busca Sequest HT e por comparação dos espectros de fragmentação da biblioteca espectral disponível em <http://www.proteometools.org/index.php?id=53> usando o algoritmo de MSPepSearch. Além disso, 2 peptídeos desta proteína foram detectados em diferentes amostras (Figura 15). O peptídeo YTCHVQHEGLPEPLTLR foi sequenciado em amostras de placentas controle e nas amostras de tireoide hidrolisadas em solução e fracionadas por RP-HPLC; o peptídeo GYEQHAYDGKDYIALNEDLR foi detectado na digestão em gel de amostras de tireoide processadas com Triton X-100 e no mesmo tecido processado com ureia/tioureaia/SDC, digerido em solução e fracionado por RP-HPLC.

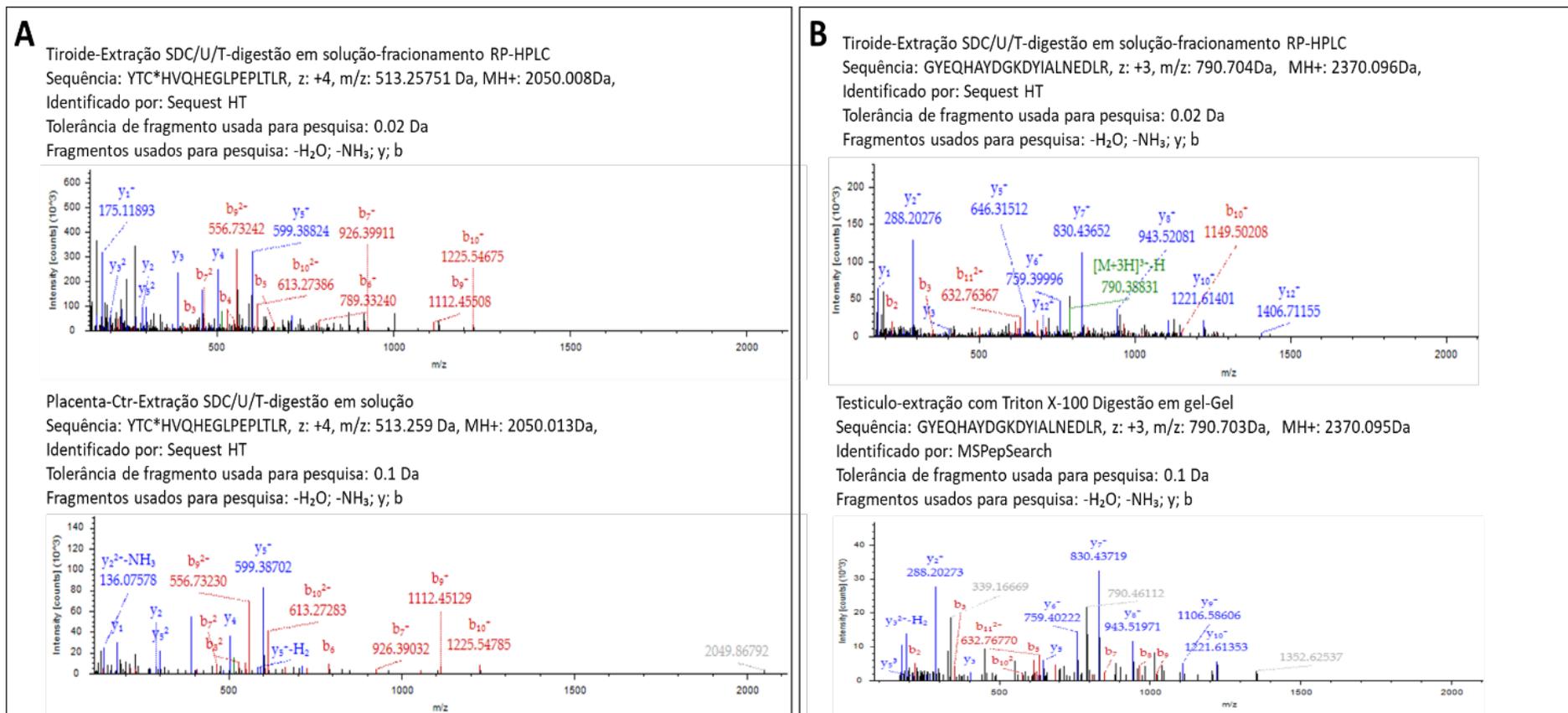


Figura 15. Espectros de fragmentação dos peptídeos YTC*HVQHEGLPEPLTLR (painel A) e GYEQHAYDGKDYIALNEDLR (painel B) exclusivos da proteína P01893, obtidos em diferentes abordagens proteômicas de este trabalho.

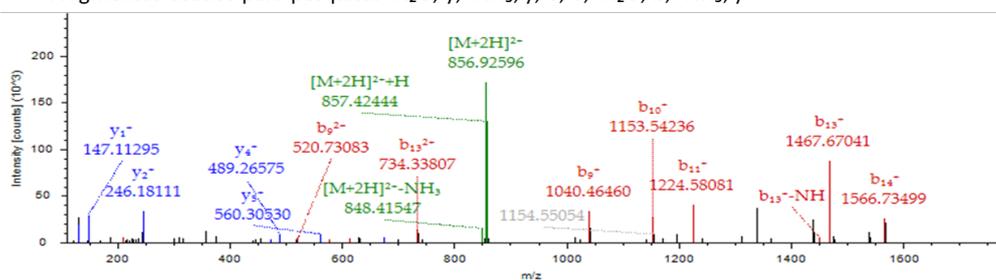
A proteína P01893 (cadeia alfa H do antígeno de histocompatibilidade HLA putativa de classe I) pertence à família MHC de classe I e está envolvida na apresentação de antígenos estranhos ao sistema imunológico. O principal complexo de histocompatibilidade humano contém aproximadamente 20 genes de classe I, pseudogenes e fragmentos de genes. Isso inclui os genes para os três principais antígenos de transplante, HLA-A, HLA-B e HLA-C, bem como vários outros genes ou pseudogenes de significado biológico desconhecido. A maioria dos últimos têm sequências ricas em C+G em suas extremidades 5' que são não-metiladas na linha de células B-linfoblastóide (Chorney *et al.*, 1990). De acordo com as anotações do UniProt, a análise da sequência da proteína P01893 mostra um domínio extracelular (de posição 25-308), um domínio helicoidal transmembrana (de comprimento 309-332); e um domínio citoplasmático (de comprimento 333-362). O gene da proteína, localizado no cromossomo 6 (locus 6p22.1), foi definido como um gene não funcional, ou pseudogene, principalmente devido à sua deleção genética em uma fração significativa da população e porque codifica uma proteína truncada potencialmente não funcional (Geraghty, Koller, Pei, and Hansen, 1992). No entanto, diferentes estudos indicaram uma possível atividade transcricional do gene HLA-H (Chorney *et al.*, 1990; Grimsley, Mather, and Ober, 1998; Jordier *et al.*, 2020). Estes estudos sugerem que o HLA-H deriva de um *locus* de apresentação de antígeno que se tornou inativado, possivelmente como resultado da seleção natural positiva e às mudanças nas demandas da imunidade das células T. Assim, a ausência de diversidade pode às vezes se correlacionar com a perda em vez da preservação da função nos genes do MHC de classe I (Zemmour *et al.*, 1990).

Tabela XII. Peptídeos únicos correspondentes à proteína Q8NHW5 detectados em este trabalho

| Proteína | Confiança de FDR | Motor de Busca | Sequencia | M+H exp. | M+H Teo. | Erro (PPM) | amostra | metodologia |
|----------|------------------|----------------|-----------------|----------|----------|------------|--------------|-------------------------------------|
| Q8NHW5 | High | Sequest HT | SVETDYTFPLAENVK | 1712.847 | 1712.838 | 5.31 | tireoide | Digestão em solução e fracionamento |
| | High | Sequest HT | SVETDYTFPLAENVK | 1712.845 | 1712.838 | 4.09 | placenta ctr | Digestão em solução |

Outra proteína sem evidências experimentais anteriores detectadas por diferentes técnicas proteômicas aplicadas neste trabalho foi a proteína Q8NHW5 (proteína ribossômica ácida 60S semelhante a P0) a partir do peptídeo SVETDYTFPLAENVK, identificado nas amostras de placentas controle e no fracionamento das amostras da tireoide extraídas com ureia/tiourea/SDC e fracionadas por RP-HPLC em pH básico (Tabela XII). Os espectros de fragmentação do peptídeo mostraram similaridade nos sinais principais (Figura 16).

Placenta-Ctr-Extração SDC/U/T-digestão em solução
 Sequência: SVETDYTFPLAENVK, z: +2, m/z: 856.92602 Da, MH+: 1712.845Da,
 Identificado por: Sequest HT
 Tolerância de fragmento usada para pesquisa: 0.1 Da
 Fragmentos usados para pesquisa: -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y



Tiroide-Extração SDC/U/T-digestão em solução-fracionamento RP-HPLC
 Sequência: SVETDYTFPLAENVK, z:+2, m/z: 856.927Da, MH+: 1712.847 Da,
 Identificado por: Sequest HT
 Tolerância de fragmento usada para pesquisa: 0.02 Da
 Fragmentos usados para pesquisa: -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y

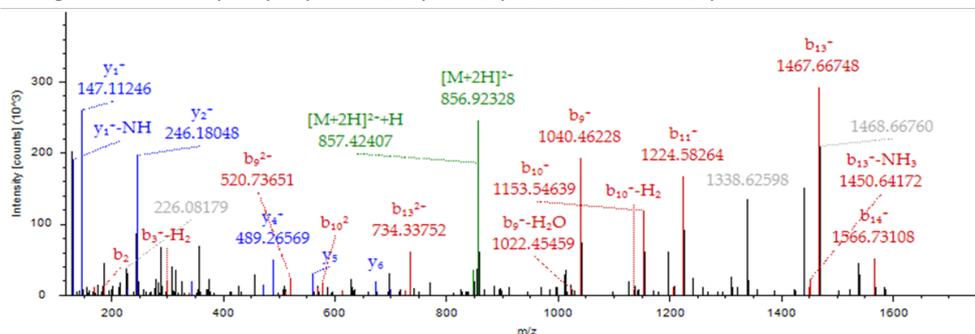
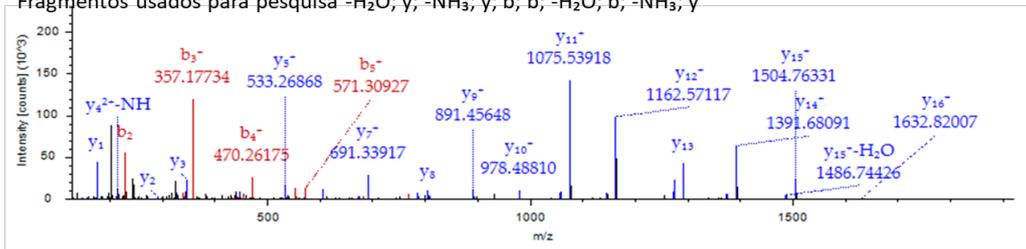


Figura 16. Espectros de fragmentação do peptídeo SVETDYTFPLAENVK correspondente à proteína Q8NHW5 detectadas nas placentas controles por digestão em solução e nas amostras de tireoide processadas com o tampão de ureia/tiourea/SDC e fracionadas por RP-HPLC em pH básico.

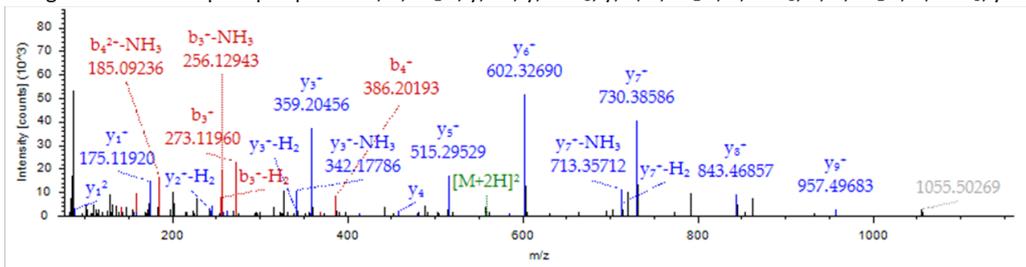
De acordo com o Uniprot, o gene dessa proteína pode ser um pseudogene que é duplicado por retrotransposição e pode se transformar em um gene não funcional. Se a proteína for expressa, ela estaria localizada no citosol como um componente da grande subunidade do ribossomo e sua função principal está associada à tradução da proteína no citoplasma.

Finalmente, a detecção específica da proteína PODSN7 (Provável imunoglobulina região kappa variável 1D-37) por proteômica *shotgun* e por proteômica direcionada confirma sua presença no tecido tireoidiano. Embora uma validação mais confiável exija a síntese dos peptídeos únicos para comparar o padrão de fragmentação com os padrões detectados na amostra biológica a detecção específica de 3 peptídeos únicos dessa proteína usando a técnica PRM é uma forte evidência da existência dessa proteína na amostra biológica (Figura 10). Os espectros de fragmentação desses peptídeos mostraram boa qualidade e intensidade e vários íons fragmentos correspondem com suas massas teóricas esperadas. (Figura 17).

Sequência : DIQLTQSPSSLSASVGDR, z: +2, m/z: 930.96814, MH+: 1860.92900 Da,
 Identificado por : Sequest HT
 Tolerância de fragmento usada para pesquisa : 0.1 Da
 Fragmentos usados para pesquisa -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y



Sequência: ASNLQSGVPSR, z: +2, m/z: 558.294Da, MH+: 1115.58143 Da
 Identificado por: MSPepSearch
 Tolerância de fragmento usada para pesquisa: 0.1 Da
 Fragmentos usados para pesquisa: +H; a; -H₂O; y; -H; y; -NH₃; y; a; a; -H₂O; a; -NH₃; b; b; -H₂O; b; -NH₃; y



Sequência : LLIYSASNLQSGVPSR, z: +2, m/z: 852.96806 Da, MH+: 1704.92884 Da,
 Identificado por : Sequest HT
 Tolerância de fragmento usada para pesquisa : 0.1 Da
 Fragmentos usados para pesquisa : -H₂O; y; -NH₃; y; b; b; -H₂O; b; -NH₃; y

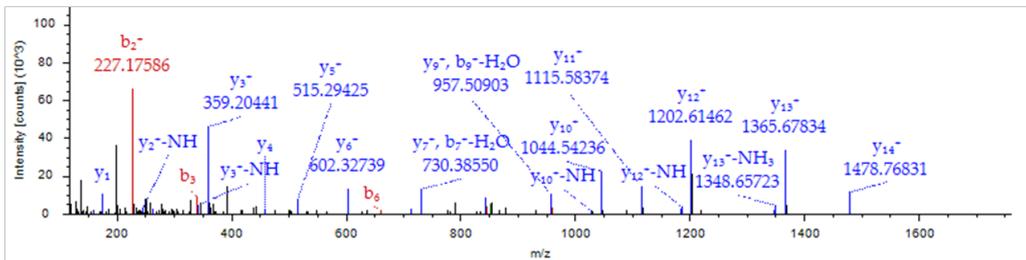


Figura 17. Espectros de fragmentação dos peptídeos DIQLTQSPSSLSASVGDR, ASNLQSGVPSR, e LLIYSASNLQSGVPSR da proteína PODSN7 detectada na digestão da banda B das amostras de tireoide processadas com os tampões de Triton X-100 e ureia/tiourea/SDC.

A proteína PODSN7 participa na resposta imune adaptativa, embora provavelmente seja um marco de leitura aberta não funcional que não participa na síntese de uma cadeia de imunoglobulina produtiva devido a uma mutação na posição 110, correspondendo à segunda cisteína da ponte dissulfeto, potencialmente levando a dobramento incorreto. Esta alteração pode ser produto de recombinação de troca e/ou local de *splicing* (mRNA) e/ou mudança de aminoácido conservada (proteína). Segundo o UniProt pode ser expressa como proteína de membrana ou secretada ao meio extracelular.

Análise proteômica de Placenta infectada pelo ZIKV

A aplicação da proteômica não está limitada apenas à caracterização de proteínas desconhecidas. As técnicas proteômicas podem ser usadas também para estudar as mudanças nos processos biológicos e vias metabólicas em uma doença e investigar as interações patógeno-hospedeiro para identificar biomarcadores potenciais e alvos terapêuticos, assim como entender todo o sistema de ação, toxicidade, resistência e eficácia do fármaco. Como outro objetivo deste trabalho, também usamos técnicas proteômicas para comparar os perfis proteômicos diferenciais entre placentas controles e placentas de mulheres infectadas com ZIKV durante a gravidez e que desenvolveram fetos microcefálicos.

4.5. Proteômica de placentas infectadas pelo Zika

Após a digestão das proteínas extraídas e a análise por MS, identificamos 2.895 proteínas nas amostras de placenta (Tabela complementar VII). Para o grupo controle saudável (Ctr), grupo infectado por ZIKV sem microcefalia (Z+) e grupo afetado por microcefalia e ZIKV-positivo (MC+), foram detectadas 2.886, 2.847 e 2.877 proteínas totais respectivamente.

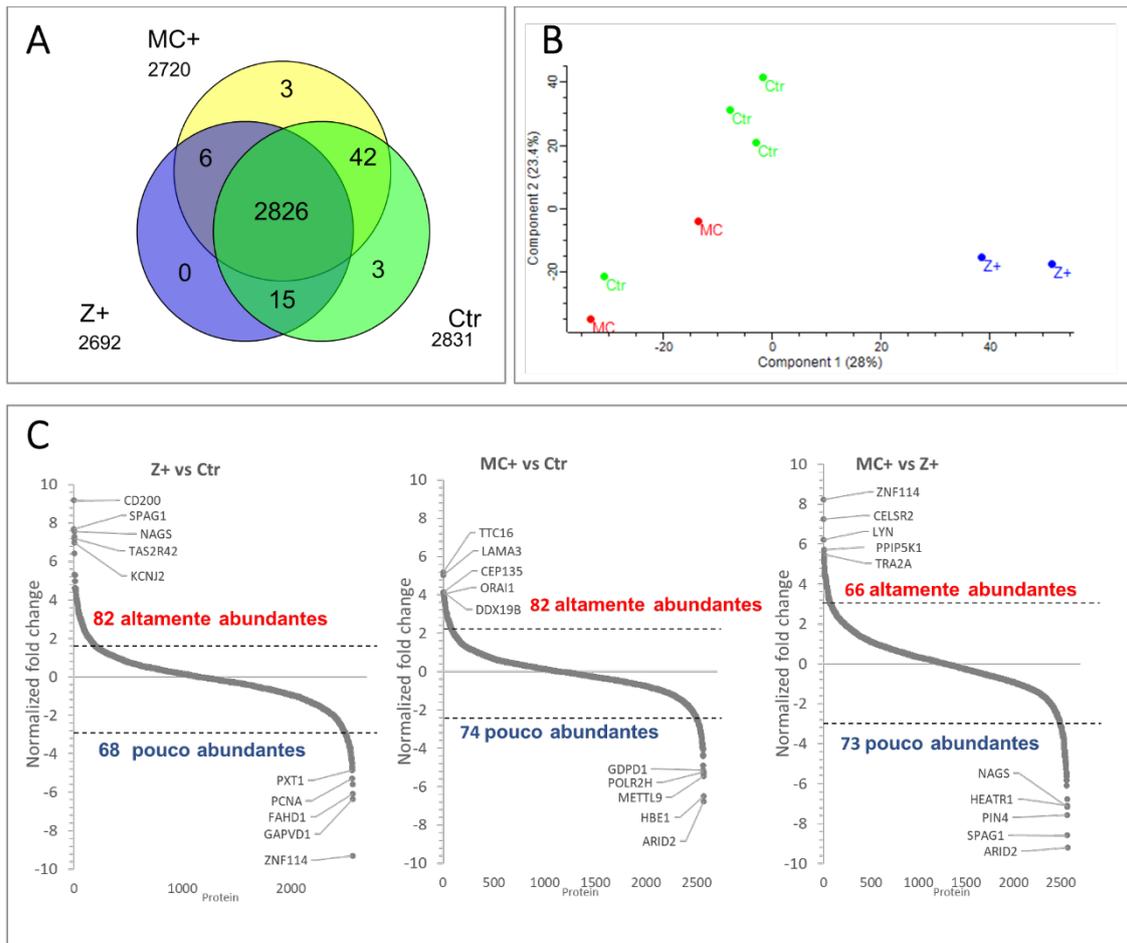


Figura 18. Análise proteômica global das placentas infectadas com ZIKV que carregavam fetos com microcefalia (MC+), placentas infectadas com ZIKV, mas geram recém-nascidos típicos (Z+) e um grupo de controle de placentas saudáveis (Ctr). **A)** Diagrama de Venn representando as proteínas identificadas no MC+, Z+ e Ctr, em pelo menos 2 réplicas. **B)** Análise de componentes principais (PCA) dos perfis de expressão de proteínas do tecido placentário dos 3 grupos de estudo (Z+, MC+ e Ctr). A PCA foi realizada em 2.438 proteínas identificadas em todas as réplicas biológicas em cada grupo. **C)** Distribuição das alterações de dobramento (FC) da expressão da proteína no Z + vs Ctr; Grupos MC+ vs Ctr e MC+ vs Z+. A linha pontilhada representa o limite para determinar as proteínas diferencialmente abundantes. Proteínas com FC maior ou menor que 2 desvios padrão da população média foram consideradas diferencialmente abundantes.

Além disso, 2.826 proteínas foram identificadas em comum entre os três grupos (Figura 18A). Uma análise de componente principal (*Principal Component Analysis*, PCA) dos dados foi realizada para gerar uma visão global do perfil de amostras obtido (Figura 18B). O perfil de proteínas das amostras separou-se de acordo com os 3 grupos principais da análise. Para realizar a quantificação *label-free*, usamos a abundância relativa das proteínas detectadas em pelo menos duas réplicas biológicas em cada grupo. A análise diferencial foi alcançada comparando o *fold-change* (FC) das proteínas entre os diferentes grupos (Z+ vs Ctr; MC+ vs Ctr e MC+ vs Z+). As proteínas cujos valores de FC eram maiores ou menores do que dois desvios padrão da população média foram consideradas diferencialmente expressas. Estas proteínas são representadas acima ou abaixo das linhas pontilhadas na Figura 18C.

4.6. Análise funcional das proteínas expressas diferencialmente

Para avaliar as alterações do proteoma em placentas infectadas pelo ZIKV com ou sem neonatos microcefálicos, realizamos uma análise de enriquecimento usando as anotações do GO e Kegg e consideramos como anotações enriquecidas aquelas com P-valor de <0,05.

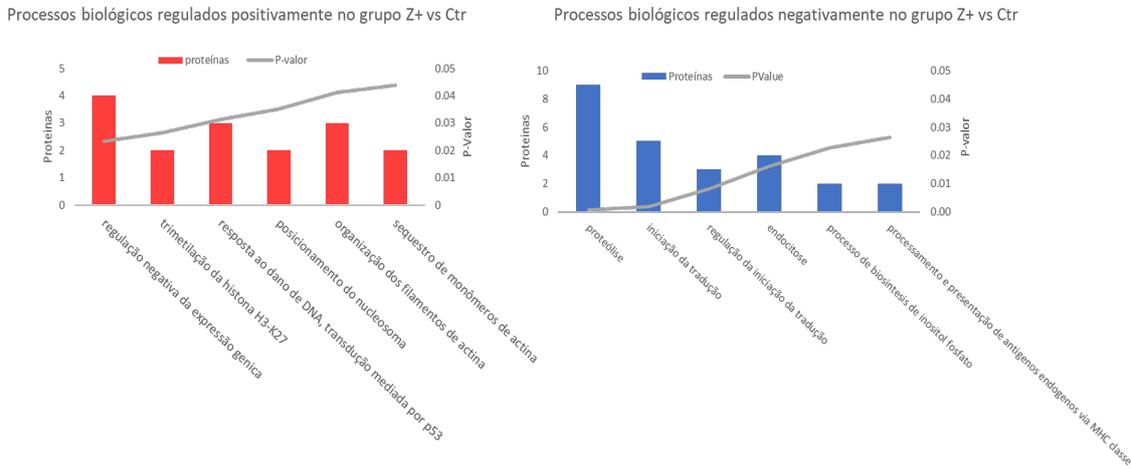


Figura 19. Processos biológicos do GO enriquecidos na lista de proteínas reguladas positivamente (vermelho) e negativamente (azul) obtidos pela análise na ferramenta bioinformática DAVID.

Na análise do Z+ vs Ctr, todas as proteínas com FC maior que 2,69 ou menor que -2,91 foram consideradas diferencialmente abundantes. Encontramos 82 e 68 proteínas com abundâncias aumentada e diminuída em relação ao grupo controle, respectivamente (Figura 18C, painel esquerdo) (Tabela complementar VIII). Entre as proteínas mais abundantes, os processos mais enriquecidos foram associados à regulação negativa da expressão gênica e resposta a danos no DNA (Figura 19). Muitos vírus de RNA, como flavivírus, podem induzir danos significativos ao DNA, apesar de seu ciclo de replicação ocorrer exclusivamente no citoplasma (Ryan, Hollingworth, and Grand, 2016). Hammack *et al.* relatam que o ZIKV pode quebrar o DNA de fita dupla e ativar a via de ATM/Chk2 para reparar o DNA em células progenitoras neuronais (Hammack *et al.*, 2019). Este evento interrompe a progressão do ciclo celular e aumenta o recrutamento de várias proteínas de reparo de DNA, como a Proteína 1 de ligação a p53 (Q12888) (Huyen *et al.*, 2004; Panier and Boulton, 2014). Além disso, Ghouzzi *et al.* correlacionou o efeito genotóxico do ZIKV em células progenitoras neurais humanas com a ativação de genes conhecidos na via de P53 (El Ghouzzi *et al.*, 2016). Os níveis aumentados de efetores da cascata de p53 são responsáveis pela apoptose e perda neuronal em muitos casos de microcefalias congênitas. De forma interessante, o dano ao DNA em resposta ao sinal de p53 foi um dos processos representados no grupo Z+ em comparação ao grupo controle segundo o enriquecimento dos processos biológicos na ferramenta DAVID.

As proteínas que diminuíram sua abundância no grupo Z+ em relação ao controle foram relacionadas ao processamento e apresentação de antígenos peptídicos (Figura

19). O processo de iniciação da tradução e a via de transporte de RNA também foram processos sub-regulados no grupo infectado. Curiosamente, a proteína RBL2, potente repressor da transcrição, aumentou seus níveis no grupo Z+. A infecção pelo ZIKV altera a transcrição na microglia, fibroblasto, rim embrionário e linhas celulares de macrófagos derivadas de monócitos (Tiwari *et al.*, 2017). A regulação negativa da expressão gênica pode estar associada a uma regulação negativa da tradução do mRNA, que foi observada em ambos os grupos infectados (Z+ e MC+) em comparação ao controle. Lum *et al.* relataram que a infecção por ZIKV, no disco da placenta, pode perturbar a expressão de genes relacionados à sinalização eIF2, que regula o início da tradução e está relacionada a respostas de estresse celular, eritropoiese e respostas imunes à infecção viral (Lum *et al.*, 2019). Outro dos processos regulados negativamente no grupo Z+ comparado com o grupo Ctr, foi a organização do filamento de actina e a regulação do citoesqueleto. A superfície dos sinciciotrofoblastos (STBs) fetais, forma uma barreira multinuclear que está em contato direto com o sangue materno e é resistente à infecção pelo ZIKV. O complexo e denso citoesqueleto de actina dessas células serve como uma restrição biofísica que impede a passagem de vários patógenos (Robbins, Skrzypczynska, Zeldovich, Kapidzic, and Bakardjiev, 2010; Zeldovich *et al.*, 2013). A diminuição na abundância de proteínas relacionadas a estes processos sugere um possível mecanismo da transmissão através da barreira sincicial do ZIKV. Embora os STBs sejam altamente resistentes a muitas infecções, eles também podem funcionar como uma possível via de transcitose para vírus. Vários estudos relatam mecanismos de endocitose potencializados por anticorpos e secreção de vesículas com partículas virais no centro das vilosidades coriônicas como estratégias para cruzar a interface materno-fetal (M.-L. Li *et al.*, 2020; Teixeira, Pietrobon, Oliveira, Oliveira, and Sato, 2020).

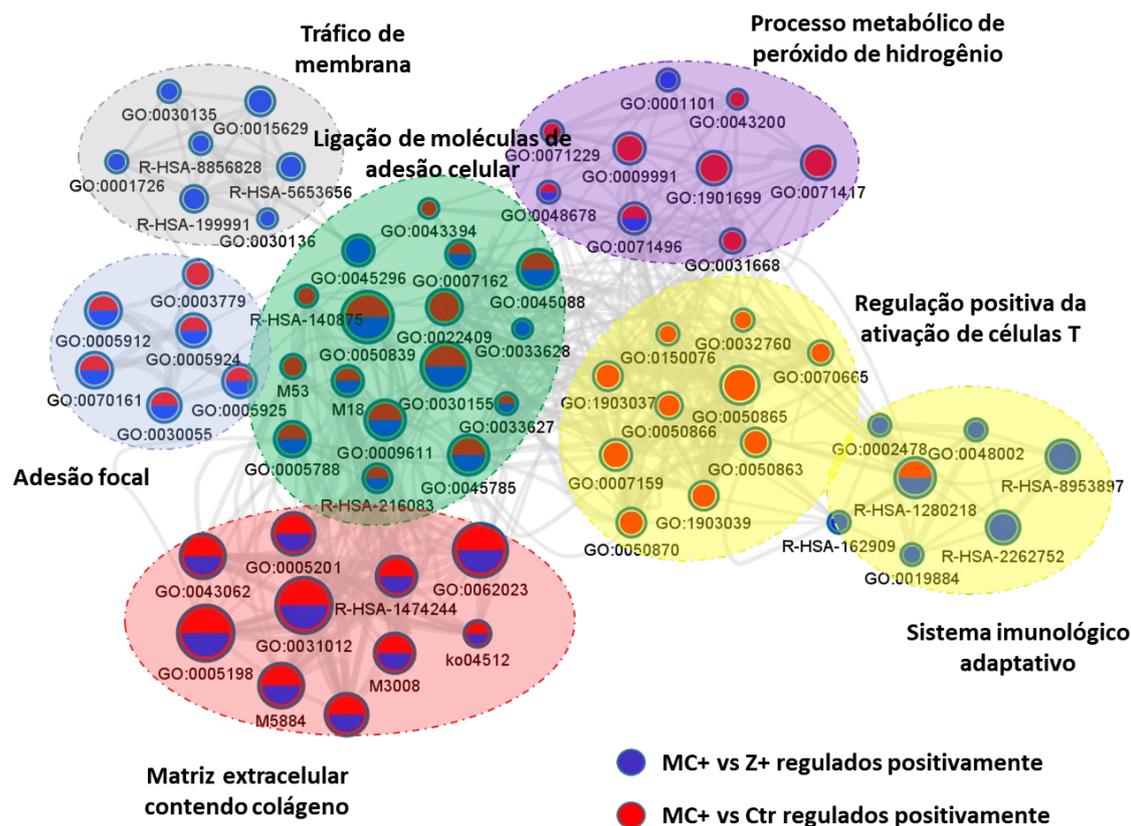


Figura 20. Anotações e vias biológicas de GO mais representadas entre as proteínas reguladas positivamente do grupo MC+ em comparação com os grupos Ctr (vermelho) e Z+ (azul). O tamanho do nó é proporcional ao número de proteínas encontradas em cada processo.

Para a análise de enriquecimento no MC+ vs Ctr, encontramos 82 proteínas de alta abundância e 68 de baixa abundância com um FC maior que 2,23 e menor que -2,41, respectivamente (Figura 18C, painel central) (Tabela Complementar VIII). As proteínas altamente abundantes foram relacionadas à fibrinólise e à regulação positiva da ativação de células T. Anfasa *et al.* determinaram que a infecção pelo ZIKV em células endoteliais da veia umbilical humana induz uma elevação do Fator Tecidual (FT), que tem um papel importante na ativação dos mecanismos de coagulação, formação de coágulos e sua subsequente dissolução por fibrinólise (Anfasa *et al.*, 2019). Além disso, a ativação da coagulação e da fibrinólise é observada em pacientes com dengue com uma expressão aumentada de FT (De Azeredo, Monteiro, and De-Oliveira Pinto, 2015). Também descobrimos que as proteínas de adesão focal reguladas positivamente aumentaram a abundância em comparação com o grupo controle. Durante o desenvolvimento normal da placenta, o citotrofoblasto (CTB) indiferenciado adquire um fenótipo invasivo que ancora as vilosidades flutuantes na membrana basal e na matriz extracelular da decídua materna. As moléculas de adesão celular e a exposição de proteínas da matriz extracelular alteram o padrão de expressão gênica desses trofoblastos extravilosos (TEVs). Estudos anteriores descrevem as integrinas, proteínas de adesão celular, como um receptor de entrada às células (Stewart and Nemerow, 2007). O ZIKV pode usar integrina $\alpha\beta 5$ para ser internalizado nas células-tronco neurais (S. Wang *et al.*, 2020). Curiosamente, Amaral *et al.* relataram que a infecção de

trofoblasto por ZIKV de gêmeos afetados pela síndrome congênita da Zika (*Congenital Zika syndrome, CZS*) diminuiu a expressão das proteínas da matriz extracelular (Amaral et al., 2020). No entanto, diferentes cepas do ZIKV podem causar diferentes padrões de expressão de proteínas de adesão (Tabata et al., 2018). Além disso, a infecção por ZIKV em monócitos primários humanos induz uma expressão significativa de proteínas associadas à adesão focal e transporte vesicular, aumentando a transmigração de monócitos através da parede do vaso e disseminação para células neuronais no feto (Ayala-Nunez et al., 2019). Uma análise de rede dos processos mais representados também mostra a ligação de adesão da molécula celular como um *cluster* enriquecido (Figura 20) (Tabela complementar IX).

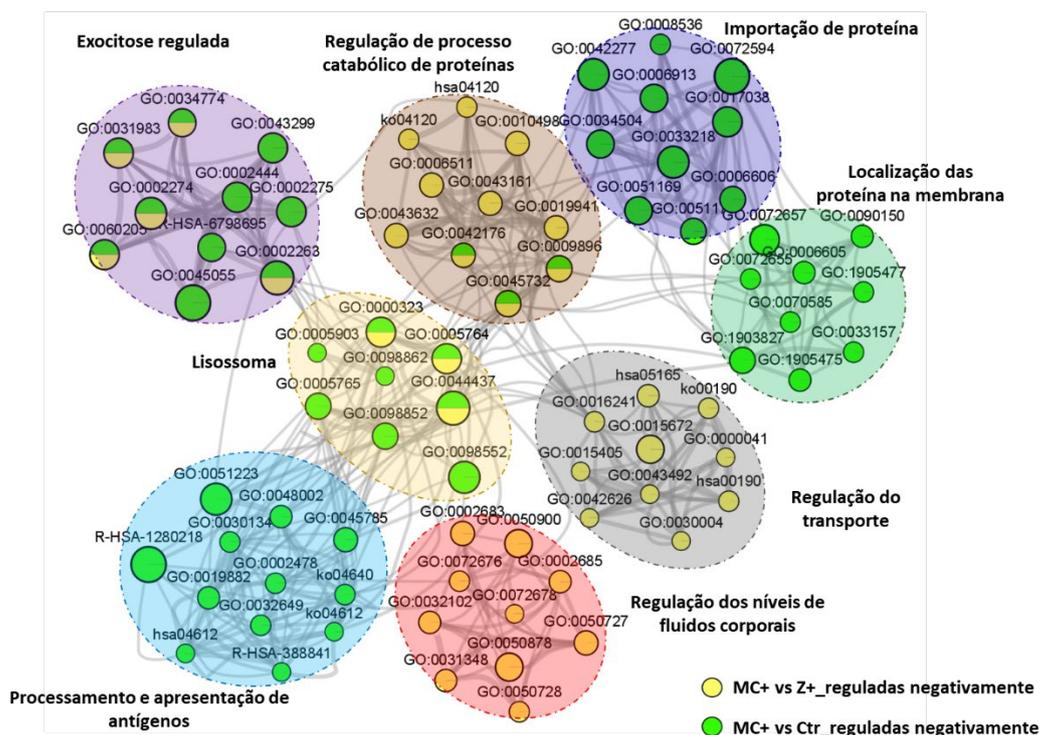


Figura 21. Anotações dos processos biológicos e as vias metabólicas de GO mais representadas entre as proteínas reguladas negativamente do grupo MC+ em comparação com o grupo Ctr (verde) e Z+ (amarelo). O tamanho do nó é proporcional ao número de proteínas em cada processo.

Na lista de proteínas reguladas negativamente no grupo MC+ vs Ctr, foram encontradas ABCF1, AKT1, NOLC1, PIN4, RPL10L, RPL37A, RPL3L, RPS29 relacionadas à tradução e processamento de RNA. O enriquecimento da via também revelou as proteínas RPL10L, RPL37A, RPL3L, RPS29 associadas à tradução do mRNA, o que pode estar associado à regulação negativa da expressão genica observada na comparação das proteínas diferencialmente expressas no grupo Z+ vs Ctr. O processamento e a apresentação do antígeno, entre outros processos do sistema imunológico, foram regulados negativamente nesta comparação (Figura 21, 19). A rede de termos enriquecidos revelou vários processos relacionados ao transporte e localização de proteínas, bem como à regulação da exocitose como regulada negativamente (Figura 21) (Tabela Complementar X).

O estudo do MC+ vs Z+ revelou 66 proteínas abundantes com um FC superior a 2,23 e 73 proteínas reguladas negativamente com FC inferior a -2,41 (Figura 18C) (Tabela complementar VIII). Detectamos proteínas reguladas positivamente relacionadas à endocitose, replicação viral e ativação do complemento mediada por anticorpo. Curiosamente, as proteínas envolvidas na organização da matriz extracelular e na adesão celular foram encontradas mais abundantes no grupo MC+ em comparação com Z+.

A análise da rede dos termos mais representados na lista de proteínas reguladas negativamente mostra-se na Figura 21 (Tabela complementar X). Os processos mais representados na lista de proteínas que diminuíram sua abundância no grupo MC+ comparado com o grupo Z+, foram relacionados à regulação da macroautofagia e organização dos filamentos de actina. Uma das estratégias que a placenta usa para proteger o feto dos agentes microbianos é a autofagia. Este mecanismo ocorre nos trofoblastos das vilosidades placentárias em desenvolvimento e tem como alvo os componentes intracelulares para degradação lisossomal. No entanto, vários patógenos podem evadir, inibir ou sequestrar essa via para facilitar a infecção em diferentes estágios do ciclo de vida viral. Os níveis reduzidos de proteínas que controlam a autofagia detectados nas placentas do grupo MC+ em relação ao Ctr, sugere que o ZIKV pode desregular a autofagia para sua vantagem replicativa. A inibição genética ou farmacológica da autofagia na interface materno-fetal limita a infecção pelo ZIKV (Delorme-Axford *et al.*, 2013). A autofagia secretora parece ser um mecanismo importante de transferência através da barreira placentária. Na autofagia secretora, as partículas virais dentro dos exossomos são secretadas ao invés de fundidas com o lisossoma como na autofagia degradativa (Singh Chahar, Bao, and Casola, 2015; Z. W. Zhang, Li, and Yuan, 2017). Para complementar este resultado, também observamos um aumento na abundância de proteínas relacionadas à endocitose no grupo MC+ e desregulação das proteínas de processos de exocitose, bem como níveis aumentados de proteínas de tráfego de membrana (Figuras 17 e 18). Esse mecanismo está intimamente relacionado à via da autofagia secretora. A endocitose é responsável pela transcitose de partículas pela placenta (Papa *et al.*, 2017). Os STBs expressam em sua superfície o receptor Fc neonatal que transporta a IgG materna para o feto através do sincício. O ZIKV pode-se ligar aos anticorpos não neutralizantes para tirar vantagem desse transporte e infectar os CTBs abaixo da camada sincicial (M.-L. Li *et al.*, 2020). Este mecanismo facilita a transmissão do vírus através da placenta e pode influenciar a incidência de microcefalia (Z. W. Zhang *et al.*, 2017). A regulação do processo catabólico da proteína também foi negativa. A análise da rede dos termos mais representados mostra a regulação do transporte e exocitose como enriquecida na lista de proteínas de baixa abundância (Fig. 18) (Tabela complementar X).

A infecção pelo ZIKV na placenta provavelmente leva a alterações na resposta imune materna. Várias células imunológicas nas vilosidades coriônicas, bem como TEVs, são suscetíveis à infecção pelo ZIKV. TEVs expressam em sua superfície o Complexo Principal de Histocompatibilidade (MHC) classe I. Detectamos em ambos os grupos infectados (Z+ e MC+) uma diminuição na abundância de proteínas relacionadas ao

processamento e apresentação de antígenos (Figura 21). Uma das estratégias utilizadas pelos vírus para evadir a resposta imune é a regulação negativa do processamento e apresentação do antígeno pelo MHC classe I. O Citomegalovírus humano prejudica a expressão do HLA-G, uma molécula não clássica do MHC-I, responsável por manter tolerância materna ao feto semialogênico (Fisher, Genbacev, Maidji, and Pereira, 2000). Além disso, o vírus herpes simplex (HSV) anula a expressão de moléculas HLA de classe I em células de coriocarcinoma (Schust, Hill, and Ploegh, 1996). O processamento e a apresentação do antígeno mediado pelo MHC de classe I também foram desregulados em cérebros post-mortem de neonatos com CZS (Aguiar *et al.*, 2020). Uma diminuição na apresentação e processamento de antígenos provavelmente desencadeia uma resposta imune materna que promove a ativação de células T, que é um processo aumentado no grupo MC+ vs Ctr (Figura 20). A redução nos níveis de HLA-E e HLA-G ativa as células CD8⁺ e CD4⁺ (Sargent, Borzychowski, and Redman, 2006) e combinada com os altos níveis de proteínas de adesão e matriz extracelular poderia aumentar o recrutamento de células NK e TEVs que são direcionados por essas proteínas e células musculares lisas. Todas essas alterações poderiam criar uma resposta imune descontrolada nessa região que também pode desencadear uma resposta inflamatória fetal caracterizada pela secreção em altas concentrações de citocinas, como IL-1, IL-6, IL-8 e TNF- α (Davies *et al.*, 2000; Madsen-Bouterse *et al.*, 2010; Salaun, Romero, and Lebecque, 2007). Essas citocinas podem afetar o SNC e os sistemas circulatórios e induzir malformações fetais em modelos animais. Esta Síndrome de Resposta Inflamatória Fetal pode causar malformações morfológicas do cérebro fetal, bem como aumento do risco de diagnóstico de autismo, esquizofrenia, déficits neurosensoriais e psicose mais tarde na vida [54]

5. Conclusões

A partir dos resultados obtidos neste trabalho, pode-se concluir que as estratégias de imunopurificação demonstraram sua utilidade para enriquecer especificamente proteínas de interesse, porém a combinação desta metodologia com a espectrometria de massa ainda precisa otimizar algumas condições. Dado que os anticorpos usados para a captura das MP são purificados com os mesmos antígenos usados nos controles, é necessário incluir uma etapa para eliminar impurezas nos anticorpos anti-MP, desde que o anticorpo não perca sua afinidade pela proteína de interesse.

A simplificação de amostras pelo fracionamento de proteínas por eletroforese e a análise por espectrometria de massa detectou 11 proteínas sem previa evidência proteica, demonstrando a importância do fracionamento da amostra para a identificação de MP. A confirmação de 5 destas proteínas logrou-se pela técnica de proteômica-alvo direcionada PRM. a proteína P0DSN7 cujo nível de expressão era incerto de acordo com o banco de dados neXtProt, detectou-se por PRM com 3 peptídeos únicos com mais de 9 aminoácidos cujas sequencias não são sobrepostas. Embora ainda necessite de

validação mais rigorosa, as evidências apresentadas por esta metodologia podem ser usadas para categorizar a proteína P0DSN7 como PE1.

O fracionamento de misturas de peptídeo por RP-HPLC em pH básico combinado com espectrometria de massa também se mostrou uma metodologia eficiente para detecção de MP. Novamente o tampão com ureia, tiourea e o detergente SDC conseguiu ter melhores resultados, pois foram detectadas 9 proteínas com níveis PE4-5 a partir de 10 peptídeos únicos em comparação com as 5 proteínas com níveis PE4-5 detectadas usando S-trap que utiliza um tampão de SDS/TEAB para a extração. A proteína P01893 (PE5) identificou-se com duas sequências peptídicas distintas não sobrepostas de mapeamento exclusivo e de comprimento ≥ 9 aminoácidos, o qual é a principal condição para reivindicações de novas detecções de proteína PE1, de acordo com as diretrizes da HUPO. Estes peptídeos também foram detectados em outras análises junto com outras duas sequências exclusivas somando 4 peptídeos únicos para a proteína P01893.

A análise proteômica das placentas infectadas com ZIKV que desenvolvem ou não fetos com características microcefálicas permitiu estudar os mecanismos de transmissão transplacentária do vírus, bem como possíveis alterações na placenta que podem influenciar malformações nos neonatos. A infecção pelo Zika vírus mostrou alteradas as proteínas relacionadas ao dano ao DNA e à tradução do mRNA na placenta, reduzindo assim a expressão gênica. Processos relacionados à transcrição viral que podem estar associados a uma possível rota de transmissão vertical também foram detectados em placentas infectadas com ZIKV. A comparação das placentas ZIKV positivas com neonatos microcefálicos com as placentas infectadas pelo ZIKV sem fetos microcefálicos mostrou que os processos mais desregulados estão relacionados à adesão celular e matriz extracelular, sugerindo uma invasão de trofoblastos extravilosos da placenta em direção à decídua materna. Detectamos também, nessa comparação, desregulação de proteínas relacionadas à resposta imune que pode indicar um rompimento da tolerância materna ao feto, desencadeando malformações morfológicas no cérebro do feto.

6. Perspectivas

Neste trabalho, várias metodologias baseadas em espectrometria de massa foram desenvolvidas a fim de identificar proteínas codificadas no genoma humano que não apresentavam evidências experimental suficiente ao nível de proteína. Para o enriquecimento específico de 2 proteínas modelo deste tipo, foi aplicado um método de imunopurificação utilizando anticorpos contra regiões recombinantes dessas proteínas (PrEST). Embora 2 peptídeos de uma das proteínas tenham sido detectados em amostras biológicas, a presença de peptídeos no branco implica uma provável contaminação dos anticorpos com o fragmento PrEST utilizado para a sua purificação. Portanto, etapas de lavagem mais rigorosas são necessárias para remover prováveis impurezas remanescentes na preparação do anticorpo, tendo o cuidado de preservar a integridade dos anticorpos retidos na matriz.

Outra das técnicas utilizadas para a identificação das proteínas PE2-5 foi a separação das proteínas por SDS-PAGE e a digestão em gel das bandas em seu peso molecular. Utilizando essa estratégia, foram detectadas 11 proteínas com nível de PE4-5, por isso, como perspectiva de nosso trabalho, resta a análise completa da separação eletroforética, visando detectar outras MPs. O fracionamento da amostra ao nível do peptídeo permitiu detectar 14 proteínas sem evidências experimentais prévias, o que mostra sua utilidade para caracterizar proteomas em grande profundidade. A combinação de protocolos que permitem a simplificação da amostra pode ser uma estratégia útil para detectar ainda mais proteínas perdidas. O acoplamento da separação de proteínas por SDS-PAGE com o fracionamento em nível de peptídeo por RP-HPLC reduz bastante a complexidade da amostra, aumentando a probabilidade de detecção de proteínas com baixa abundância. Porém, é preciso chegar a um balance, pois, o extenso processamento e fracionamento de uma amostra inicial pode modificar as proteínas e diminuir a recuperação ao final do processo. Em geral, para confirmar a presença de proteínas inexistentes detectadas neste trabalho, é necessária uma validação por técnicas proteômicas alvo-direcionadas utilizando peptídeos marcados isotopicamente. Além disso, o estudo em outros tecidos e amostras interessantes também permanece para trabalhos futuros.

O estudo proteômico de placentas infectadas com ZIKV contribui para o entendimento da fisiopatologia da infecção, além de fornecer pistas sobre como o vírus pode ser transmitido da mãe para o feto através da barreira placentária. A verificação dos mecanismos alterados durante a infecção que foram encontrados neste trabalho pode ser realizada utilizando linhagens celulares infectadas pelo ZIKV e ainda é possível estudar se a inibição ou ativação de algumas das vias e processos desregulados podem influenciar a replicação viral. Nesse sentido, o estudo em modelos animais pode ser uma ferramenta útil para verificar a importância desses processos no desenvolvimento embrionário e na incidência de microcefalia após a infecção pelo ZIKV.

7. Referências Bibliográficas

- Adhikari, S., Nice, E. C., Deutsch, E. W., Lane, L., Omenn, G. S., Pennington, S. R., ... Baker, M. S. (2020). A high-stringency blueprint of the human proteome. *Nature Communications*, *11*(1), 5301. <https://doi.org/10.1038/s41467-020-19045-9>
- Aebersold, R., Bader, G. D., Edwards, A. M., van Eyk, J., Kussman, M., Qin, J., and Omenn, G. S. (2014). Highlights of B/D-HPP and HPP Resource Pillar Workshops at 12th Annual HUPO World Congress of Proteomics: September 14-18, 2013, Yokohama, Japan R. Aebersold et al. *Proteomics*, *14*(9), 975–988. <https://doi.org/10.1002/pmic.201400041>
- Aguiar, R. S., Pohl, F., Morais, G. L., Nogueira, F. C. S., Carvalho, J. B., Guida, L., ... Nakaya, H. I. (2020). Molecular alterations in the extracellular matrix in the brains of newborns with congenital Zika syndrome. *Science Signaling*, *13*(635), 1–13. <https://doi.org/10.1126/scisignal.aay6736>
- Amaral, M. S., Goulart, E., Caires-Júnior, L. C., Morales-Vicente, D. A., Soares-Schanoski, A., Gomes, R. P., ... Verjovski-Almeida, S. (2020). Differential gene expression elicited by zikv infection in trophoblasts from congenital zika syndrome discordant twins. *PLoS Neglected Tropical Diseases*, *14*(8), 1–24. <https://doi.org/10.1371/journal.pntd.0008424>
- Anfasa, F., Goeijenbier, M., Widagdo, W., Siegers, J. Y., Mumtaz, N., Okba, N., ... Martina, B. E. E. (2019). Zika virus infection induces elevation of tissue factor production and apoptosis on human umbilical vein endothelial cells. *Frontiers in Microbiology*, *10*, 1–11. <https://doi.org/10.3389/fmicb.2019.00817>
- Ayala-Nunez, N. V., Follain, G., Delalande, F., Hirschler, A., Partiot, E., Hale, G. L., ... Gaudin, R. (2019). Zika virus enhances monocyte adhesion and transmigration favoring viral dissemination to neural cells. *Nature Communications*, *10*(1), 1–16. <https://doi.org/10.1038/s41467-019-12408-x>
- Bell, P. A., Solis, N., Kizhakkedathu, J. N., Matthew, I., and Overall, C. M. (2019). *Proteomic and N-Terminomic TAILS Analyses of Human Alveolar Bone Proteins: Improved Protein Extraction Methodology and LysargiNase Digestion Strategies Increase Proteome Coverage and Missing Protein Identification*. <https://doi.org/10.1021/acs.jproteome.9b00445>
- Besnard, M., Lastère, S., Teissier, A., Cao-Lormeau, V. M., and Musso, D. (2014). Evidence of perinatal transmission of zika virus, French Polynesia, December 2013 and February 2014. *Eurosurveillance*, *19*(13), 20751. <https://doi.org/10.2807/1560-7917.ES2014.19.13.20751>
- Bjornson, R. D., Carriero, N. J., Colangelo, C., Shifman, M., Cheung, K. H., Miller, P. L., and Williams, K. (2008). X!!Tandem, an improved method for running X!Tandem in parallel on collections of commodity computers. *Journal of Proteome Research*, *7*(1), 293–299. <https://doi.org/10.1021/pr0701198>
- Calvet, G., Aguiar, R. S., Melo, A. S. O. O., Sampaio, S. A., de Filippis, I., Fabri, A., ... de Filippis, A. M. B. B. (2016). Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *The Lancet Infectious*

Diseases, 16(6), 653–660. [https://doi.org/10.1016/S1473-3099\(16\)00095-5](https://doi.org/10.1016/S1473-3099(16)00095-5)

- Carapito, C., Duek, P., Macron, C., Seffals, M., Rondel, K., Delalande, F., ... Pineau, C. (2017). Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *Journal of Proteome Research*, 16(12), 4340–4351. <https://doi.org/10.1021/acs.jproteome.7b00374>
- Cho, J.-Y., Lee, H.-J., Jeong, S.-K., and Paik, Y.-K. (2017). Epsilon-Q: An Automated Analyzer Interface for Mass Spectral Library Search and Label-Free Protein Quantification. *J. Proteome Res*, 16, 4435–4445. <https://doi.org/10.1021/acs.jproteome.6b01019>
- Cho, J. Y., Lee, H. J., Jeong, S. K., Kim, K. Y., Kwon, K. H., Yoo, J. S., ... Paik, Y. K. (2015). Combination of Multiple Spectral Libraries Improves the Current Search Methods Used to Identify Missing Proteins in the Chromosome-Centric Human Proteome Project. *Journal of Proteome Research*, 14(12), 4959–4966. <https://doi.org/10.1021/acs.jproteome.5b00578>
- Choong, W. K., Lih, T. S. M., Chen, Y. J., and Sung, T. Y. (2017). Decoding the Effect of Isobaric Substitutions on Identifying Missing Proteins and Variant Peptides in Human Proteome. *Journal of Proteome Research*, 16(12), 4415–4424. <https://doi.org/10.1021/acs.jproteome.7b00342>
- Chorney, M. J., Sawada, I., Gillespie, G. A., Srivastava, R., Pan, J., and Weissman, S. M. (1990). Transcription analysis, physical mapping, and molecular characterization of a nonclassical human leukocyte antigen class I gene. *Molecular and Cellular Biology*, 10(1), 243–253. <https://doi.org/10.1128/mcb.10.1.243-253.1990>
- Danielsson, A., Pontén, F., Fagerberg, L., Hallström, B. M., Schwenk, J. M., Uhlén, M., ... Lindskog, C. (2014). The human pancreas proteome defined by transcriptomics and antibody-based profiling. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0115421>
- Davies, J. K., Shikes, R. H., Sze, C. I., Leslie, K. K., McDuffie, R. S., Romero, R., and Gibbs, R. S. (2000). Histologic inflammation in the maternal and fetal compartments in a rabbit model of acute intra-amniotic infection. *American Journal of Obstetrics and Gynecology*, 183(5), 1088–1093. Mosby Inc. <https://doi.org/10.1067/mob.2000.108888>
- De Azeredo, E. L., Monteiro, R. Q., and De-Oliveira Pinto, L. M. (2015). Thrombocytopenia in dengue: Interrelationship between virus and the imbalance between coagulation and fibrinolysis and inflammatory mediators. *Mediators of Inflammation*, 2015, 1–16. <https://doi.org/10.1155/2015/313842>
- Delorme-Axford, E., Donker, R. B., Mouillet, J. F., Chu, T., Bayer, A., Ouyang, Y., ... Coyne, C. B. (2013). Human placental trophoblasts confer viral resistance to recipient cells. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29), 12048–12053. <https://doi.org/10.1073/pnas.1304718110>
- Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., ... Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Research*, 34(Database issue). <https://doi.org/10.1093/nar/gkj040>

- Deutsch, E. W., Lane, L., Overall, C. M., Bandeira, N., Baker, M. S., Pineau, C., ... Omenn, G. S. (2019). Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *Journal of Proteome Research*, 18(12), 4108–4116. <https://doi.org/10.1021/acs.jproteome.9b00542>
- Dick, G. W. A. (1952). Zika virus (II). Pathogenicity and physical properties. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 46(5), 521–534. [https://doi.org/10.1016/0035-9203\(52\)90043-6](https://doi.org/10.1016/0035-9203(52)90043-6)
- Djureinovic, D., Fagerberg, L., Hallström, B., Danielsson, A., Lindskog, C., Uhlén, M., and Pontén, F. (2014). The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Molecular Human Reproduction*, 20(6), 476–488. <https://doi.org/10.1093/MOLEHR/GAU018>
- Dong, Q., Menon, R., Omenn, G. S., and Zhang, Y. (2015). Structural bioinformatics inspection of neXtProt PE5 proteins in the human proteome. *Journal of Proteome Research*, 14(9), 3750–3761. <https://doi.org/10.1021/acs.jproteome.5b00516>
- Duek, P., Bairoch, A., Gateau, A., Vandenbrouck, Y., and Lane, L. (2016). Missing Protein Landscape of Human Chromosomes 2 and 14: Progress and Current Status. *Journal of Proteome Research*, 15(11), 3971–3978. <https://doi.org/10.1021/acs.jproteome.6b00443>
- Edqvist, P. H. D., Fagerberg, L., Hallström, B. M., Danielsson, A., Edlund, K., Uhlén, M., and Pontén, F. (2015). Expression of Human Skin-Specific Genes Defined by Transcriptomics and Antibody-Based Profiling. *Journal of Histochemistry and Cytochemistry*, 63(2), 129–141. <https://doi.org/10.1369/0022155414562646>
- El Ghouzzi, V., Bianchi, F. T., Molineris, I., Mounce, B. C., Berto, G. E., Rak, M., ... Di Cunto, F. (2016). ZIKA virus elicits P53 activation and genotoxic stress in human neural progenitors similar to mutations involved in severe forms of genetic microcephaly and p53. *Cell Death and Disease*, 7(10), e2440–e2440. <https://doi.org/10.1038/cddis.2016.266>
- Elguoshy, A., Hirao, Y., Yamamoto, K., Xu, B., Kinoshita, N., Mitsui, T., and Yamamoto, T. (2019). Utilization of the Proteome Data Deposited in SRMATlas for Validating the Existence of the Human Missing Proteins in GPM. *Journal of Proteome Research*, 18(12), 4197–4205. <https://doi.org/10.1021/acs.jproteome.9b00355>
- Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics*, 13(1), 22–24. <https://doi.org/10.1002/pmic.201200439>
- Faye, O., Freire, C. C. M., Iamarino, A., Faye, O., de Oliveira, J. V. C., Diallo, M., ... Sall, A. A. (2014). Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Neglected Tropical Diseases*, 8(1), 36. <https://doi.org/10.1371/journal.pntd.0002636>
- Fisher, S., Genbacev, O., Maidji, E., and Pereira, L. (2000). Human Cytomegalovirus Infection of Placental Cytotrophoblasts In Vitro and In Utero: Implications for Transmission and Pathogenesis. *Journal of Virology*, 74(15), 6808–6820. <https://doi.org/10.1128/jvi.74.15.6808-6820.2000>

- Foy, B. D., Kobylinski, K. C., Foy, J. L. C., Blitvich, B. J., da Rosa, A. T., Haddow, A. D., ... Tesh, R. B. (2011). Probable Non-Vector-borne Transmission of Zika Virus, Colorado, USA. *Emerging Infectious Diseases*, 17(5), 880–882. <https://doi.org/10.3201/eid1705.101939>
- Garcez, P. P., Nascimento, J. M., De Vasconcelos, J. M., Madeiro Da Costa, R., Delvecchio, R., Trindade, P., ... Rehen, S. K. (2017). Zika virus disrupts molecular fingerprinting of human neurospheres. *Scientific Reports*, 7, 1–10. <https://doi.org/10.1038/srep40780>
- Garin-Muga, A., Odriozola, L., Martínez-Val, A., Del Toro, N., Martínez, R., Molina, M., ... Segura, V. (2016). Detection of Missing Proteins Using the PRIDE Database as a Source of Mass Spectrometry Evidence. *Journal of Proteome Research*, 15(11), 4101–4115. <https://doi.org/10.1021/acs.jproteome.6b00437>
- Gaudet, P., Argoud-Puy, G., Cusin, I., Duek, P., Evalet, O., Gateau, A., ... Lane, L. (2013). NeXtProt: Organizing protein knowledge in the context of human proteome projects. *Journal of Proteome Research*, 12(1), 293–298. <https://doi.org/10.1021/pr300830v>
- Geraghty, D. E., Koller, B. H., Pei, J., and Hansen, J. A. (1992). Examination of four HLA class I pseudogenes: Common events in the evolution of HLA genes and pseudogenes. *Journal of Immunology (Baltimore, Md. : 1950)*, 149(6), 1947–19456.
- Grard, G., Caron, M., Mombo, I. M., Nkoghe, D., Mboui Ondo, S., Jiolle, D., ... Leroy, E. M. (2014). Zika Virus in Gabon (Central Africa) - 2007: A New Threat from *Aedes albopictus*? *PLoS Neglected Tropical Diseases*, 8(2). <https://doi.org/10.1371/journal.pntd.0002681>
- Grimsley, C., Mather, K. A., and Ober, C. (1998). HLA-H: A pseudogene with increased variation due to balancing selection at neighboring loci. *Molecular Biology and Evolution*, 15(12), 1581–1588. <https://doi.org/10.1093/oxfordjournals.molbev.a025886>
- Guruceaga, E., Garin-Muga, A., Prieto, G., Bejarano, B., Marcilla, M., Marín-Vicente, C., ... Segura, V. (2017). Enhanced Missing Proteins Detection in NCI60 Cell Lines Using an Integrative Search Engine Approach. *Journal of Proteome Research*, 16(12), 4374–4390. <https://doi.org/10.1021/acs.jproteome.7b00388>
- Habuka, M., Fagerberg, L., Hallström, B. M., Kampf, C., Edlund, K., Sivertsson, A., ... Odeberg, J. (2014). The kidney transcriptome and proteome defined by transcriptomics and antibody-based profiling. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0116125>
- Hammack, C., Ogden, S. C., Madden, J. C., Medina, A., Xu, C., Phillips, E., ... Tang, H. (2019). Zika Virus Infection Induces DNA Damage Response in Human Neural Progenitors That Enhances Viral Replication. *Journal of Virology*, 93(20), 638–657. <https://doi.org/10.1128/jvi.00638-19>
- Hanash, S., and Celis, J. E. (2002, June 1). The Human Proteome Organization: a mission to advance proteome knowledge. *Molecular & Cellular Proteomics : MCP*, Vol. 1, pp. 413–414. Elsevier. <https://doi.org/10.1074/mcp.R200002-MCP200>

- Hartwig, S., Czibere, A., Kotzka, J., Paßlack, W., Haas, R., Eckel, J., and Lehr, S. (2009). Combinatorial hexapeptide ligand libraries (ProteoMiner™): An innovative fractionation tool for differential quantitative clinical proteomics. *Archives of Physiology and Biochemistry*, 115(3), 155–160. <https://doi.org/10.1080/13813450903154224>
- He, F. (2005). Human liver proteome project: Plan, progress, and perspectives. *Molecular and Cellular Proteomics*, 4(12), 1841–1848. <https://doi.org/10.1074/mcp.R500013-MCP200>
- Huyen, Y., Zgheib, O., DiTullio, R. A., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., ... Halazonetis, T. D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature*, 432(7015), 406–411. <https://doi.org/10.1038/nature03114>
- Hwang, H., Park, G. W., Park, J. Y., Lee, H. K., Lee, J. Y., Jeong, J. E., ... Yoo, J. S. (2017). Next Generation Proteomic Pipeline for Chromosome-Based Proteomic Research Using NeXtProt and GENCODE Databases. <https://doi.org/10.1021/acs.jproteome.7b00223>
- Jenner, T. L., and Rose'Meyer, R. B. (2006). Adenosine A3 receptor mediated coronary vasodilation in the rat heart: Changes that occur with maturation. *Mechanisms of Ageing and Development*, 127(3), 264–273. <https://doi.org/10.1016/j.mad.2005.10.005>
- Jeong, S. K., Lee, H. J., Na, K., Cho, J. Y., Lee, M. J., Kwon, J. Y., ... Paik, Y. K. (2013). GenomewidePDB, a proteomic database exploring the comprehensive protein parts list and transcriptome landscape in human chromosomes. *Journal of Proteome Research*, 12(1), 106–111. <https://doi.org/10.1021/pr3009447>
- Jordier, F., Gras, D., De Grandis, M., D'Journo, X. B., Thomas, P. A., Chanez, P., ... Di Cristofaro, J. (2020). HLA-H: Transcriptional Activity and HLA-E Mobilization. *Frontiers in Immunology*, 10, 2986. <https://doi.org/10.3389/fimmu.2019.02986>
- Kampf, C., Mardinoglu, A., Fagerberg, L., Hallström, B. M., Edlund, K., Lundberg, E., ... Uhlen, M. (2014). The human liver-specific proteome defined by transcriptomics and antibody-based profiling. *FASEB Journal*, 28(7), 2901–2914. <https://doi.org/10.1096/fj.14-250555>
- Kenney, J. L., and Brault, A. C. (2014). The role of environmental, virological and vector interactions in dictating biological transmission of arthropod-borne viruses by mosquitoes. In *Advances in Virus Research* (Vol. 89, pp. 39–83). Academic Press Inc. <https://doi.org/10.1016/B978-0-12-800172-1.00002-1>
- Khakoo, S. I., Thio, C. L., Martin, M. P., Brooks, C. R., Gao, X., Astemborski, J., ... Carrington, M. (2004). HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science*, 305(5685), 872–874. <https://doi.org/10.1126/science.1097670>
- Kumar, D., Jain, A., and Dash, D. (2015). Probing the Missing Human Proteome: A Computational Perspective. *Journal of Proteome Research*, 14(12), 4949–4958. <https://doi.org/10.1021/acs.jproteome.5b00728>

- Kusebauch, U., Campbell, D. S., Deutsch, E. W., Chu, C. S., Spicer, D. A., Brusniak, M. Y., ... Moritz, R. L. (2016). Human SRMATlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell*, 166(3), 766–778. <https://doi.org/10.1016/j.cell.2016.06.041>
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), 680–685. <https://doi.org/10.1038/227680a0>
- Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R. (2007). Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*, 7(5), 655–667. <https://doi.org/10.1002/pmic.200600625>
- Lane, L., Argoud-Puy, G., Britan, A., Cusin, I., Duek, P. D., Evalet, O., ... Bairoch, A. (2012). NeXtProt: A knowledge platform for human proteins. *Nucleic Acids Research*, 40(D1). <https://doi.org/10.1093/nar/gkr1179>
- Lane, L., Bairoch, A., Beavis, R. C., Deutsch, E. W., Gaudet, P., Lundberg, E., and Omenn, G. S. (2014). Metrics for the human proteome project 2013-2014 and strategies for finding missing proteins. *Journal of Proteome Research*, 13(1), 15–20. <https://doi.org/10.1021/pr401144x>
- Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., ... Omenn, G. S. (2011). The human proteome project: Current state and future direction. *Molecular and Cellular Proteomics*, 10(7). <https://doi.org/10.1074/mcp.M111.009993>
- Li, M.-L., Chiu, C.-F. F., Chu, L.-W. W., Liao, I.-C. C., Simanjuntak, Y., Lin, Y.-L. L., ... Ping, Y.-H. H. (2020). The Mechanism of the Zika Virus Crossing the Placental Barrier and the Blood-Brain Barrier. *Frontiers in Microbiology*, 11, 1–15. <https://doi.org/10.3389/fmicb.2020.00214>
- Li, S., He, Y., Lin, Z., Xu, S., Zhou, R., Liang, F., ... Ren, Y. (2017). Digging More Missing Proteins Using an Enrichment Approach with ProteoMiner. *Journal of Proteome Research*, 16(12), 4330–4339. <https://doi.org/10.1021/acs.jproteome.7b00353>
- Lin, Z., Zhang, Y., Pan, H., Hao, P., Li, S., He, Y., ... Watson, J. D. (2019). *Alternative Strategy To Explore Missing Proteins with Low Molecular Weight*. <https://doi.org/10.1021/acs.jproteome.9b00353>
- Lum, F. M., Narang, V., Hue, S., Chen, J., McGovern, N., Rajarethinam, R., ... Ng, L. F. P. (2019). Immunological observations and transcriptomic analysis of trimester-specific full-term placentas from three Zika virus-infected women. *Clinical and Translational Immunology*, 8(11), 22. <https://doi.org/10.1002/cti2.1082>
- Macedo-da-silva, J., Marinho, C. R. F., Palmisano, G., and Rosa-fernandes, L. (2020, August 1). Lights and shadows of torch infection proteomics. *Genes*, Vol. 11, pp. 1–33. MDPI AG. <https://doi.org/10.3390/genes11080894>
- MacNamara, F. N. (1954). Zika virus: A report on three cases of human infection during an epidemic of jaundice in Nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 48(2), 139–145. [67](https://doi.org/10.1016/0035-</p>
</div>
<div data-bbox=)

- Madsen-Bouterse, S. A., Romero, R., Tarca, A. L., Kusanovic, J. P., Espinoza, J., Kim, C. J., ... Draghici, S. (2010). The transcriptome of the fetal inflammatory response syndrome. *American Journal of Reproductive Immunology*, 63(1), 73–92. <https://doi.org/10.1111/j.1600-0897.2009.00791.x>
- Mardinoglu, A., Kampf, C., Asplund, A., Fagerberg, L., Hallström, B. M., Edlund, K., ... Nielsen, J. (2014). Defining the human adipose tissue proteome to reveal metabolic alterations in obesity. *Journal of Proteome Research*, 13(11), 5106–5119. <https://doi.org/10.1021/pr500586e>
- Marko-Varga, G., Omenn, G. S., Paik, Y. K., and Hancock, W. S. (2013, January 4). A first step toward completion of a genome-wide characterization of the human proteome. *Journal of Proteome Research*, Vol. 12, pp. 1–5. American Chemical Society. <https://doi.org/10.1021/pr301183a>
- Melaine, N., Com, E., Bellaud, P., Guillot, L., Lagarrigue, M., Morrice, N. A., ... Pineau, C. (2018). Deciphering the Dark Proteome: Use of the Testis and Characterization of Two Dark Proteins. *Journal of Proteome Research*, 17(12), 4197–4210. <https://doi.org/10.1021/acs.jproteome.8b00387>
- Meyer, H. E., Klose, J., and Hamacher, M. (2003, November 1). HBPP and the pursuit of standardisation. *Lancet Neurology*, Vol. 2, pp. 657–658. Elsevier. [https://doi.org/10.1016/S1474-4422\(03\)00580-5](https://doi.org/10.1016/S1474-4422(03)00580-5)
- Meyfour, A., Ansari, H., Pahlavan, S., Mirshahvaladi, S., Rezaei-Tavirani, M., Gourabi, H., ... Salekdeh, G. H. (2017). Y Chromosome Missing Protein, TBL1Y, May Play an Important Role in Cardiac Differentiation. *Journal of Proteome Research*, 16(12), 4391–4402. <https://doi.org/10.1021/acs.jproteome.7b00391>
- Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., ... Zupanc, T. A. (2016). Zika virus associated with microcephaly. *New England Journal of Medicine*, 374(10), 951–958. <https://doi.org/10.1056/NEJMoa1600651>
- Musso, D., Nilles, E. J., and Cao-Lormeau, V. M. (2014). Rapid spread of emerging Zika virus in the Pacific area. *Clinical Microbiology and Infection*, 20(10), O595–O596. <https://doi.org/10.1111/1469-0691.12707>
- Musso, Didier, Roche, C., Robin, E., Nhan, T., Teissier, A., and Cao-Lormeau, V. M. (2015). Potential sexual transmission of zika virus. *Emerging Infectious Diseases*, 21(2), 359–361. <https://doi.org/10.3201/eid2102.141363>
- Oliveira Melo, A. S., Malinger, G., Ximenes, R., Szejnfeld, P. O., Alves Sampaio, S., and Bispo De Filippis, A. M. (2016). Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: Tip of the iceberg? *Ultrasound in Obstetrics and Gynecology*, 47(1), 6–7. <https://doi.org/10.1002/uog.15831>
- Omenn, G. S. (2021). Reflections on the HUPO Human Proteome Project, the Flagship Project of the Human Proteome Organization, at 10 Years. *Molecular & Cellular Proteomics*, 20, 100062. <https://doi.org/10.1016/j.mcpro.2021.100062>
- Omenn, G. S., Lane, L., Overall, C. M., Cristea, I. M., Corrales, F. J., Lindskog, C., ... Snyder,

- M. P. (2020). Research on the Human Proteome Reaches a Major Milestone: >90% of Predicted Human Proteins Now Credibly Detected, According to the HUPO Human Proteome Project HHS Public Access. *J Proteome Res*, 19(12), 4735–4746. <https://doi.org/10.1021/acs.jproteome.0c00485>
- Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., ... Hanash, S. M. Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. , 5 *Proteomics* § (2005). John Wiley & Sons, Ltd.
- Orchard, S., Hermjakob, H., and Apweiler, R. (2003). The Proteomics Standards Initiative. *Proteomics*, 3(7), 1374–1376. John Wiley & Sons, Ltd. <https://doi.org/10.1002/pmic.200300496>
- Paik, Y.-K. K., Lane, L., Kawamura, T., Chen, Y.-J. J., Cho, J. Y. J. Y. J. Y. J.-Y. J.-Y., Labaer, J., ... Overall, C. M. Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. , 17 *Journal of Proteome Research* § (2018).
- Paik, Y. K., and Hancock, W. S. (2012). Uniting ENCODE with genome-wide proteomics. *Nature Biotechnology*, 30(11), 1065–1067. <https://doi.org/10.1038/nbt.2416>
- Paik, Y. K., Omenn, G. S., Hancock, W. S., Lane, L., and Overall, C. M. (2017). Advances in the Chromosome-Centric Human Proteome Project: looking to the future. *Expert Review of Proteomics*, 14(12), 1059–1071. <https://doi.org/10.1080/14789450.2017.1394189>
- Panier, S., and Boulton, S. J. (2014, January 11). Double-strand break repair: 53BP1 comes into focus. *Nature Reviews Molecular Cell Biology*, Vol. 15, pp. 7–18. Nature Publishing Group. <https://doi.org/10.1038/nrm3719>
- Papa, M. P., Meuren, L. M., Coelho, S. V. A., de Oliveira Lucas, C. G., Mustafá, Y. M., Matassoli, F. L., ... de Arruda, L. B. (2017). Zika virus infects, activates, and crosses brain microvascular endothelial cells, without barrier disruption. *Frontiers in Microbiology*, 8, 1–17. <https://doi.org/10.3389/fmicb.2017.02557>
- Peng, X., Xu, F., Liu, S., Li, S., Huang, Q., Chang, L., ... Xu, P. (2017). Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer. *Journal of Proteome Research*, 16(12), 4364–4373. <https://doi.org/10.1021/acs.jproteome.7b00332>
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., ... Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research*, 47(D1), D442–D450. <https://doi.org/10.1093/nar/gky1106>
- Ping, P., Vondriska, T. M., Creighton, C. J., Gandhi, T. K. B., Yang, Z., Menon, R., ... Pandey, A. (2005). A functional annotation of subproteomes in human plasma. *Proteomics*, 5(13), 3506–3519. <https://doi.org/10.1002/pmic.200500140>
- Pinto, S. M., Manda, S. S., Kim, M. S., Taylor, K., Selvan, L. D. N., Balakrishnan, L., ... Pandey, A. (2014). Functional annotation of proteome encoded by human

- chromosome 22. *Journal of Proteome Research*, 13(6), 2749–2760. <https://doi.org/10.1021/pr401169d>
- Robbins, J. R., Skrzypczynska, K. M., Zeldovich, V. B., Kapidzic, M., and Bakardjiev, A. I. (2010). Placental Syncytiotrophoblast Constitutes a Major Barrier to Vertical Transmission of *Listeria monocytogenes*. *PLoS Pathogens*, 6(1), e1000732. <https://doi.org/10.1371/journal.ppat.1000732>
- Ryan, E. L., Hollingworth, R., and Grand, R. J. (2016). Activation of the DNA damage response by RNA viruses. *Biomolecules*, 6(1), 2–24. <https://doi.org/10.3390/biom6010002>
- Salaun, B., Romero, P., and Lebecque, S. (2007, December 1). Toll-like receptor's two-edged sword: When immunity meets apoptosis. *European Journal of Immunology*, Vol. 37, pp. 3311–3318. John Wiley & Sons, Ltd. <https://doi.org/10.1002/eji.200737744>
- Sargent, I. L., Borzychowski, A. M., and Redman, C. W. G. (2006). NK cells and human pregnancy - an inflammatory view. *Trends in Immunology*, 27(9), 399–404. <https://doi.org/10.1016/j.it.2006.06.009>
- Scaturro, P., Stukalov, A., Haas, D. A., Cortese, M., Draganova, K., Płaszczycza, A., ... Pichlmair, A. (2018). An orthogonal proteomic survey uncovers novel Zika virus host factors. *Nature*, 561(7722), 253–257. <https://doi.org/10.1038/s41586-018-0484-5>
- Schaeffer, M., Gateau, A., Teixeira, D., Michel, P.-A., Zahn-Zabal, M., and Lane, L. (2017). The neXtProt peptide uniqueness checker: a tool for the proteomics community. *Bioinformatics*, 33(21), 3471. <https://doi.org/10.1093/BIOINFORMATICS/BTX318>
- Schuler-Faccini, L., Ribeiro, E. M., Feitosa, I. M. L., Horovitz, D. D. G., Cavalcanti, D. P., Pessoa, A., ... Sanseverino, M. T. V. (2016). Possible association between Zika virus infection and microcephaly — Brazil, 2015. *Morbidity and Mortality Weekly Report*, 65(3), 59–62. <https://doi.org/10.15585/mmwr.mm6503e2>
- Schust, D. J., Hill, A. B., and Ploegh, H. L. (1996). Herpes Simplex Virus Blocks Intracellular Transport of HLA-G in Placentally Derived Human Cells. *The Journal of Immunology*, 157(8), 3375–3380.
- Segura, V., Garin-Muga, A., Guruceaga, E., and Corrales, F. J. (2017, January 2). Progress and pitfalls in finding the ‘missing proteins’ from the human proteome map. *Expert Review of Proteomics*, Vol. 14, pp. 9–14. Taylor and Francis Ltd. <https://doi.org/10.1080/14789450.2017.1265450>
- Singh Chahar, H., Bao, X., and Casola, A. (2015). Exosomes and Their Role in the Life Cycle and Pathogenesis of RNA Viruses. *Viruses*, 7, 3204–3225. <https://doi.org/10.3390/v7062770>
- Sjöstedt, E., Fagerberg, L., Hallström, B. M., Häggmark, A., Mitsios, N., Nilsson, P., ... Mulder, J. (2015). Defining the human brain proteome using transcriptomics and antibody-based profiling with a focus on the cerebral cortex. *PLoS ONE*, 10(6). <https://doi.org/10.1371/journal.pone.0130028>
- Stewart, P. L., and Nemerow, G. R. (2007, November 1). Cell integrins: commonly used

- receptors for diverse viral pathogens. *Trends in Microbiology*, Vol. 15, pp. 500–507. Elsevier. <https://doi.org/10.1016/j.tim.2007.10.001>
- Sun, J., Shi, J., Wang, Y., Chen, Y., Li, Y., Kong, D., ... Xu, P. (2018). Multiproteases Combined with High-pH Reverse-Phase Separation Strategy Verified Fourteen Missing Proteins in Human Testis Tissue [Research-article]. *Journal of Proteome Research*, 17(12), 4171–4177. <https://doi.org/10.1021/acs.jproteome.8b00397>
- Tabata, T., Petitt, M., Puerta-Guardo, H., Michlmayr, D., Harris, E., and Pereira, L. (2018). Zika Virus Replicates in Proliferating Cells in Explants From First-Trimester Human Placentas, Potential Sites for Dissemination of Infection. *The Journal of Infectious Diseases*®, 1202, 1202–1215. <https://doi.org/10.1093/infdis/jix552>
- Teixeira, F. M. E., Pietrobon, A. J., Oliveira, L. de M., Oliveira, L. M. da S., and Sato, M. N. (2020, February 14). Maternal-Fetal Interplay in Zika Virus Infection and Adverse Perinatal Outcomes. *Frontiers in Immunology*, Vol. 11, p. 175. Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2020.00175>
- Thul, P. J., and Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Science*, 27(1), 233–244. <https://doi.org/10.1002/pro.3307>
- Tiwari, S. K., Dang, J., Qin, Y., Lichinchi, G., Bansal, V., and Rana, T. M. (2017). Zika virus infection reprograms global transcription of host cells to allow sustained infection. *Emerging Microbes and Infections*, 6(4), 1–10. <https://doi.org/10.1038/emi.2017.9>
- Uhlén, M., Björling, E., Agaton, C., Szigartyo, C. A. K., Amini, B., Andersen, E., ... Pontén, F. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular and Cellular Proteomics*, 4(12), 1920–1932. <https://doi.org/10.1074/mcp.M500279-MCP200>
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419–1260419. <https://doi.org/10.1126/science.1260419>
- Wang, A., Thurmond, S., Islas, L., Hui, K., and Hai, R. (2017). Zika virus genome biology and molecular pathogenesis. *Emerging Microbes and Infections*, 6(3), 1–6. <https://doi.org/10.1038/emi.2016.141>
- Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., ... Kuster, B. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology*, 15(2). <https://doi.org/10.15252/msb.20188503>
- Wang, S., Zhang, Q., Tiwari, S. K., Lichinchi, G., Yau, E. H., Hui, H., ... Rana, T. M. (2020). Integrin $\alpha\beta 5$ Internalizes Zika Virus during Neural Stem Cells Infection and Provides a Promising Target for Antiviral Therapy. *Cell Reports*, 30(4), 969–983.e4. <https://doi.org/10.1016/j.celrep.2019.11.020>
- Wang, Y., Chen, Y., Zhang, Y., Wei, W., Li, Y., Zhang, T., ... Xu, P. (2017). Multi-Protease Strategy Identifies Three PE2 Missing Proteins in Human Testis Tissue. *Journal of Proteome Research*, 16(12), 4352–4363. <https://doi.org/10.1021/acs.jproteome.7b00340>
- Wei, W., Luo, W., Wu, F., Peng, X., Zhang, Y. Y., Zhang, M., ... Xu, P. (2016). Deep

- Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *Journal of Proteome Research*, 15(11), 3988–3997. <https://doi.org/10.1021/acs.jproteome.6b00390>
- Weldemariam, M. M., Han, C. L., Shekari, F., Kitata, R. B., Chuang, C. Y., Hsu, W. T., ... Chen, Y. J. (2018, December 7). Subcellular Proteome Landscape of Human Embryonic Stem Cells Revealed Missing Membrane Proteins [Research-article]. <https://doi.org/10.1021/acs.jproteome.8b00407>
- Westaway, E. G., Brinton, M. A., Gaidamovich, S. Y. A., Horzinek, M. C., Igarashi, A., Kääriäinen, L., ... Trent, D. W. (1985). Flaviviridae. *Intervirology*, 24(4), 183–192. <https://doi.org/10.1159/000149642>
- Xu, A., Li, G., Yang, D., Wu, S., Ouyang, H., Xu, P., and He, F. (2015). Evolutionary Characteristics of Missing Proteins: Insights into the Evolution of Human Chromosomes Related to Missing-Protein-Encoding Genes. *Journal of Proteome Research*, 14(12), 4985–4994. <https://doi.org/10.1021/acs.jproteome.5b00450>
- Yamamoto, T., Langham, R. G., Ronco, P., Knepper, M. A., and Thongboonkerd, V. (2008). Towards standard protocols and guidelines for urine proteomics: A report on the Human Kidney and Urine Proteome Project (HKUPP) symposium and workshop - 6 October 2007, Seoul, Korea and 1 November 2007, San Francisco, CA, USA. *Proteomics*, 8(11), 2156–2159. John Wiley & Sons, Ltd. <https://doi.org/10.1002/pmic.200800138>
- Yu, N. Y. L., Hallström, B. M., Fagerberg, L., Ponten, F., Kawaji, H., Carninci, P., ... Daub, C. O. (2015). Complementing tissue characterization by integrating transcriptome profiling from the human protein atlas and from the FANTOM5 consortium. *Nucleic Acids Research*, 43(14), 6787–6798. <https://doi.org/10.1093/nar/gkv608>
- Zahn-Zabal, M., Andr´, P.-A., Michel, A., Gateau, A., Fréd´, F., Nikitin, F., ... Lane, L. (2020). The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research*, 48. <https://doi.org/10.1093/nar/gkz995>
- Zeldovich, V. B., Clausen, C. H., Bradford, E., Fletcher, D. A., Maltepe, E., Robbins, J. R., and Bakardjiev, A. I. (2013). Placental Syncytium Forms a Biophysical Barrier against Pathogen Invasion. *PLoS Pathogens*, 9(12), e1003821. <https://doi.org/10.1371/journal.ppat.1003821>
- Zemmour, J., Koller, B. H., Ennis, P. D., Geraghty, D. E., Lawlor, D. A., Orr, H. T., and Parham, P. (1990). HLA-AR, an inactivated antigen-presenting locus related to HLA-A. Implications for the evolution of the MHC. *The Journal of Immunology*, 144(9), 3619–3629.
- Zhang, C., Li, N., Zhai, L., Xu, S., Liu, X., Cui, Y., ... Xu, P. (2014). Systematic analysis of missing proteins provides clues to help define all of the protein-coding genes on human chromosome 1. *Journal of Proteome Research*, 13(1), 114–125. <https://doi.org/10.1021/pr400900j>
- Zhang, Yao, Li, Q., Wu, F., Zhou, R., Qi, Y., Su, N., ... Liu, S. (2015). Tissue-based proteogenomics reveals that human testis endows plentiful missing proteins. *Journal of Proteome Research*, 14(9), 3583–3594.

<https://doi.org/10.1021/acs.jproteome.5b00435>

Zhang, Yuanliang, Lin, Z., Hao, P., Hou, K., Sui, Y., Zhang, K., ... Ren, Y. (2018). Improvement of Peptide Separation for Exploring the Missing Proteins Localized on Membranes. *Journal of Proteome Research*, 17(12), 4152–4159. <https://doi.org/10.1021/acs.jproteome.8b00409>

Zhang, Z. W., Li, Z. L., and Yuan, S. (2017). The role of secretory autophagy in Zika virus transfer through the placental barrier. *Frontiers in Cellular and Infection Microbiology*, 6, 1–6. <https://doi.org/10.3389/fcimb.2016.00206>

Zhu, Z., Chan, J. F. W., Tee, K. M., Choi, G. K. Y., Lau, S. K. P., Woo, P. C. Y., ... Yuen, K. Y. (2016). Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. *Emerging Microbes and Infections*, 5(3), 1–11. <https://doi.org/10.1038/emi.2016.48>

Zougman, A., Selby, P. J., and Banks, R. E. (2014). Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics*, 14(9), 1006–1000. <https://doi.org/10.1002/pmic.201300553>