



**UNIVERSIDADE FEDERAL DO AMAZONAS**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**PROSPECÇÃO DE CANDIDATOS A BIOMARCADORES ONCOLÓGICOS DE CÂNCERES  
GÁSTRICOS POR ESPECTROMETRIA DE MASSAS**

**PRISCILA FERREIRA DE AQUINO**

**MANAUS, 2012**

**UNIVERSIDADE FEDERAL DO AMAZONAS**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**PRISCILA FERREIRA DE AQUINO\***

**PROSPECÇÃO DE CANDIDATOS A BIOMARCADORES ONCOLÓGICOS DE CÂNCERES  
GÁSTRICOS POR ESPECTROMETRIA DE MASSAS**

Dissertação apresentada à  
Coordenação do Programa de Pós-  
- Graduação em Química, da  
Universidade Federal do  
Amazonas, com parte dos  
requisitos para obtenção do Título  
de Mestre em Química. Área de  
atuação: Química Orgânica.

Orientador: Dr. Afonso Duarte Leão de Souza

Co-orientador: Dr. Paulo Costa Carvalho

\*Bolsista CAPES

**MANAUS, 2012**

A657p	<p>Aquino, Priscila Ferreira de</p> <p>Prospecção de candidatos a biomarcadores oncológicos de cânceres gástricos por espectrometria de massas / Priscila Ferreira de Aquino. - Manaus: UFAM, 2012.</p> <p>174 f.; il. color.</p> <p>Dissertação (Mestrado em Química Orgânica) — Universidade Federal do Amazonas, 2012.</p> <p>Orientador: Prof. Dr. Afonso Duarte Leão de Souza</p> <p>Co-orientador: Prof. Dr. Paulo Costa Carvalho</p> <p>1. Oncologia 2. Espectrometria 3. Câncer - Tratamento I. Souza, Afonso Duarte Leão de (Orient.) II. Carvalho, Paulo Costa (co-orient.) III. Universidade Federal do Amazonas IV. Título</p> <p>CDU 616.33-066.6 (043.3)</p>
-------	--

**UNIVERSIDADE FEDERAL DO AMAZONAS**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

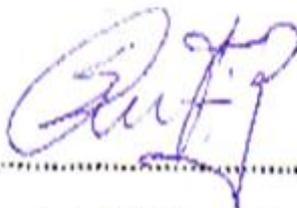
Esta dissertação foi apresentada como parte dos requisitos necessários para obtenção do grau de Mestre em Química Orgânica, outorgado pela Universidade Federal do Amazonas, e em cuja Biblioteca Central encontra-se à disposição de quem interessar.

**PRISCILA FERREIRA DE AQUINO**

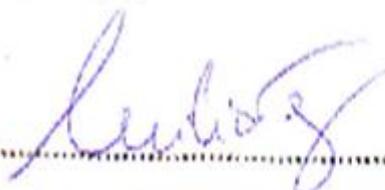
Banca Examinadora



Prof. Dr. Afonso Duarte Leão de Souza – *Presidente*



Prof. Dr. Spartaco Astolfi Filho – *Membro (UFAM)*



Prof<sup>a</sup> Dr<sup>a</sup> Cecília Verônica Nunez - *Membro (DQ- INPA)*

**“Você não pode provar uma definição. O que você pode fazer é mostrar que ela faz sentido”.**

(Albert Einstein)

## **AGRADECIMENTOS**

Em primeiro lugar a Deus, pela minha vida, saúde, força a cada dia, guiando meus passos e me proporcionando grandes oportunidades de aprendizado.

À minha família, em especial aos meus pais Manuel e Cirene, e a minha irmã Monalisa, pela compreensão nos momentos difíceis, amor, educação e apoio dedicado.

Aos meus orientadores Afonso Duarte Leão de Souza e Paulo Costa Carvalho pela orientação, pelos conselhos, apoio, conhecimentos transmitidos, confiança, paciência, compreensão, amizade e dedicação ao longo desses últimos dois anos.

A CAPES, pela bolsa de estudos concedida durante os dois anos de mestrado e aos órgãos financeiros que apoiaram este projeto.

Aos médicos Sidney Raimundo Silva Chalub e Lourenço Candido Neves, a aluna de medicina Jucilana Viana e a equipe de residência do centro cirúrgico da Fundação Centro de Controle em Oncologia do Estado do Amazonas – FCECON, pelo auxílio na coleta das amostras utilizadas nesse projeto.

Aos pacientes da FCECON que aceitaram voluntariamente em participar desse projeto.

À Dr<sup>a</sup>. Maria da Glória da Costa Carvalho e ao Dr. José Francisco de Oliveira Carvalho, por terem me recebido em sua casa com a maior disposição e alegria, pelo apoio, ensinamentos e orientações nos trabalhos de biologia molecular.

À Dr<sup>a</sup>. Ana Gisele da Costa Neves e Dr<sup>a</sup> Juliana Saldanha da Gama Fischer pela orientação, apoio, ensinamentos teóricos e experimentais em bioquímica e proteômica.

Ao Dr. Jonas Perales e Dr. Gilberto Barbosa Domont, por terem me recebido em seus laboratórios de pesquisa, Toxinologia e Química de Proteínas, respectivamente, proporcionando a execução dos trabalhos proteômicos desse projeto assim como o apoio, amizade, discussão científica, confiança e a utilização do espectrômetro de massas.

À Dr<sup>a</sup> Antonia Queiroz Lima de Souza pelo apoio e discussão científica em torno desse projeto.

Ao Dr. José Raimundo Gomes da Silva do departamento de estatística da UFAM no auxílio e orientação do tratamento estatístico dos dados de metabolômicas.

Ao meus amigos, em especial à Aimêe Oliveira, Luciana Freire, Adriana Silva, Laís Menezes, Mayane Souza e Elzalina Soares pelo apoio, compreensão e amizade.

Aos meus colegas do grupo GEMMA pelo aprendizado mútuo.

Aos meus colegas do laboratório de espectrometria de massas - UFAM, especialmente Felipe Moura e Richardson Alves pela amizade, idéias, apoio e auxílio nos experimentos de metabolômica.

Aos meus colegas de laboratório de Toxinologia (Fiocruz) e de Química de proteínas (UFRJ) pela compreensão e auxílio nos trabalhos de proteômicas.

Aos colegas do laboratório de controle da Expressão Gênica do Instituto de Biofísica Carlos Chagas Filho (UFRJ) pelo auxílio nos trabalhos de biologia molecular.

Por fim, a todos que de alguma forma contribuíram para a realização deste trabalho.

## RESUMO

O câncer é uma doença em expansão, cuja mortalidade é elevada no mundo, fazendo com que a comunidade científica busque cada vez mais diagnosticar precocemente, visto que as chances de curá-lo neste caso são maiores. O câncer gástrico é de especial interesse devido ser o quarto em malignidade e o segundo que mais causa mortes no mundo. Neste trabalho, amostras de tecidos tumorais e respectivas margens de ressecção de pacientes com câncer gástrico, e de biópsia do estômago de indivíduos controles (sem essa doença), foram investigadas através de abordagens proteômica, metabolômica e por biologia molecular. A análise proteômica utilizando uma abordagem da proteômica *shotgun*, o MudPIT *off-line* e a quantificação *label-free* pelo cromatogramas de íons extraídos (XIC<sub>S</sub>) demonstrou que cada tipo de tecido é bem distinto. A margem de ressecção apresentou várias proteínas previamente correlacionadas ao câncer, assim como algumas proteínas superexpressadas que podem estar relacionadas ao crescimento tumoral e a metástase. Alguns exemplos são: colágeno alfa-1, ceruloplasmina, calpastatina e E-caderina. A análise multivariada, segundo o método de Ward e o agrupamento de k-Means, dos dados de metabolômica revelou que foi capaz de agrupar os diferentes tipos de amostras de acordo com seu tipo de tecido. De maneira que os resultados da análise metabolômica estão alinhados com os da proteômica em que os tipos de tecidos são muito distintos, particularmente a margem de ressecção, o qual pode ser resultado de uma contínua batalha bioquímica entre o corpo e as células cancerígenas. Além disso, uma investigação viral das amostras mediante a técnica de PCR (*Polymerase Chain Reaction*) evidenciou que o vírus *Epstein-Barr* (EBV) deve ser considerado um fator de risco para o câncer gástrico em Manaus. Os resultados proteômicos deste trabalho indicam que a margem de ressecção pode desempenhar um papel fundamental para a nutrição das células cancerígenas. O presente trabalho corrobora com a idéia de que é necessário um maior entendimento da doença em nível molecular, o que pode proporcionar tratamentos mais efetivos, e que o EBV deve ser levado em consideração para a incidência do câncer gástrico.

**Palavras-chave:** Câncer gástrico, espectrometria de massas, biomarcadores.

## ABSTRACT

Cancer is a disease in expansion, whose mortality is one of the highest in the world, motivating the scientific community to increasingly seek early diagnosis, as the chances of cure are higher in this case. Gastric cancer is of particular interest as it is the fourth most common in malignancy and the second leading cause of death worldwide. In this work, we compared tissue samples from control subjects (without the disease) and patients with gastric cancer and their corresponding resection margin by using proteomics, metabolomics, and molecular biology. The proteomic analysis comprised an offline MudPIT shotgun proteomic approach with label-free quantitation by extracted ion chromatograms (XICs); it revealed that each of these tissue-types is very distinct. The resection margin presented several proteins previously correlated with cancer, but also several other proteins that could be related to tumor nourishment and metastasis. Some examples are: collagen alpha-1, ceruloplasmin, calpastatin, and E-cadherin. The multivariate analysis, according to Ward's method and k-means clustering, of our metabolomics data revealed that was able to group the different samples according to their tissue type. Our metabolomic analysis results are aligned with those from the proteomics in that these tissue-types are very distinct, particularly the resection margin which may be the result of an ongoing biochemical battle between the body and the cancer cells. Furthermore, a viral investigation of the samples by PCR (Polymerase Chain Reaction) showed that the Epstein-Barr virus (EBV) should be considered as a risk factor for gastric cancer in Manaus. In all, the proteomic results poses the resection margin as playing a key role in the nutrition of the cancer cells. Regardless, we still need a better understanding of this disease at the molecular level to provide more effective treatments.

**Keywords:** gastric cancer, mass spectrometry, biomarkers.

## SUMÁRIO

RESUMO.....	VIII
ABSTRACT .....	IX
LISTA DE SIMBOLOS E ABREVIACOES.....	XIV
LISTA DE FIGURAS.....	XVIII
LISTA DE GRFICOS.....	XXI
LISTA DE TABELAS .....	XXII
1. INTRODUO.....	23
1.1 Cncer .....	23
1.2 Cncer gstrico .....	26
1.3 Origem do cncer gstrico .....	29
1.4 Biomarcadores .....	33
1.5 Biologia molecular – Reao em cadeia pela polimerase (PCR- Polymerase chain reaction).....	35
1.6 Protemica .....	36
1.7 Separao de Protenas .....	38
1.7.1 Descrio das tcnicas utilizadas nesse trabalho:.....	39
1.8 Identificao de Protenas.....	43
1.8.1 Espectrometria de massas .....	44
1.9 Marcadores Proticos Aplicveis na Deteco de Processos Cancergenos.....	51
1.10 Metabolmica.....	54

2. OBJETIVOS .....	59
2.1 Objetivo geral .....	59
2.2 Objetivos específicos .....	59
3. MATERIAIS E MÉTODOS.....	60
3.1 Seleção de pacientes com câncer e indivíduos controle.....	60
3.2 Coleta e preservação das amostras.....	61
3.3 Caracterização dos tumores .....	62
3.4 Estudo da presença da infecção pelo vírus Epstein-Barr (EBV) em tecidos de cânceres gástricos através da biologia molecular.....	62
3.4.1 A Extração de DNA.....	63
3.4.2 A Análise da presença viral por Reação em cadeia pela polimerase (PCR – “ <i>Polimerase Chain Reaction</i> ”).....	63
3.4.3 A visualização dos produtos amplificados por eletroforese em gel de poli-acrilamida.....	65
3.5 Estudo exploratório comparativo de candidatos a biomarcadores protéicos dos cânceres gástricos mediante a técnica de MudPIT off-line.....	66
3.5.1 B Pulverização do tecido .....	66
3.5.2 B Dosagem de Proteínas.....	67
3.5.3 B Eletroforese unidimensional .....	68
3.5.4 B Coloração com Coomassie R-250.....	68
3.5.5 B Digestão trípica em gel de poli-acrilamida .....	69
3.5.6 B Digestão trípica de proteínas em solução .....	70

3.5.7 B Preparação de mini-colunas (Poros R2).....	71
3.5.8 B Cromatografia de troca iônica forte (SCX).....	72
3.5.9 B Aquisição de dados no Orbitrap Velos (Thermo, San Jose).....	73
3.5.10 B Análise de dados pela proteômica <i>Shotgun</i> .....	74
3.6 Estudo exploratório comparativo de perfis metabolômicos de cânceres gástricos.....	78
3.6.1 C Preparação da amostra .....	78
3.6.2 C Aquisição de dados no LCQ <i>Fleet</i> (Thermo-Scientific). .....	79
3.6.3 C Agrupamento não supervisionado dos perfis de metabólitos .....	79
4. RESULTADOS E DISCUSSÃO .....	81
4 A: Estudo da presença da infecção pelo vírus Epstein-Barr (EBV) em tecidos de cânceres gástricos através da biologia molecular .....	81
4 B: Estudo exploratório comparativo de candidatos a biomarcadores protéicos de cânceres gástricos mediante à técnica de MudPIT off-line.....	85
4.1 B Análise do gel unidimensional – 1 D.....	90
4.2 B Análise dos dados obtidos com a tecnologia multidimensional de identificação de proteínas (MudPIT) <i>off-line</i> .....	95
4 C: Estudo exploratório comparativo de perfis metabolômicos de cânceres gástricos.....	105
5. CONSIDERAÇÕES FINAIS .....	120
6. REFERÊNCIAS BIBLIOGRÁFICAS .....	122
7. ANEXOS .....	138

7.1 Parecer do Comitê de ética em pesquisa - UFAM .....	138
7.2 Parecer do Comitê de ética em pesquisa - FCECON .....	139
7.3 Termo de Consentimento Livre e Esclarecido (TCLE) .....	140
7.4 Questionário.....	142
7.5 Artigo aceito para a publicação na Genetics and Molecular Research .	143
7.6 Artigo submetido ao Journal of Proteome Research.....	152

**LISTA DE SIMBOLOS E ABREVIACOES**

<b>1 D</b>	Unidimensional
<b>Abs.</b>	Absorbncia
<b>AJCC</b>	<i>American Joint Committee on Cancer</i>
<b>AMP</b>	Adenosina monofosfato.
<b>ATP</b>	Adenosina trifosfato
<b>BCA</b>	cido bicinchonnico
<b>bp</b>	Par de bases (do ingls <i>base pair</i> )
<b>BSA</b>	Albumina de soro bovino (do ingls <i>Bovine Serum Albumin</i> ).
<b>CA</b>	Antgeno carboidrato
<b>CAD</b>	Dissociao por coliso induzida (do ingls <i>Collision Activated Dissociation</i> )
<b>CE</b>	Eletroforese capilar
<b>CEA</b>	Antgeno carcinoembrionrio
<b>CECON</b>	Centro de controle em oncologia
<b>CEP</b>	Comit de tica em Pesquisa
<b>COL11A1</b>	Colgeno $\alpha$ -1
<b>DNA</b>	cido desoxirribonucleico
<b>dNIAF</b>	Distribuio normalizada do fator de abundncia do io
<b>dNSAF</b>	Distribuio normalizada de fatores de abundncia espectrais
<b>DNTPs</b>	Desoxirribonucleotdeos Fosfatados
<b>DTT</b>	Ditiotreitol
<b>EBV</b>	<i>Epstein-Barr vrus</i>
<b>EDTA</b>	<i>Ethylenediamine tetraacetic acid.</i>
<b>ESI</b>	<i>Eletrospray Ionization</i>
<b>ETD</b>	Dissociao por transferncia de eltrons (do ingls <i>Electron Transfer Dissociation</i> )
<b>FBLN1</b>	Fibulina-1

<b>FDR</b>	<i>False Discovery Rate</i>
<b>GC-MS</b>	Cromatografia gasosa acoplada à espectrometria de massa
<b>GC-MS/MS</b>	Cromatografia gasosa acoplada à espectrometria de massa em tandem
<b>HCD</b>	Dissociação por alta energia (do inglês <i>High Energy Dissociation</i> )
<b>HE</b>	Hematoxilina e eosina
<b>HPLC</b>	<i>High Performance Liquid Chromatography.</i>
<b>IAA</b>	Iodoacetamida
<b>IARC</b>	Agência internacional para pesquisa sobre câncer
<b>INCA</b>	Instituto Nacional do Câncer
<b>LC-MS</b>	Cromatografia líquida acoplada à espectrometria de massa
<b>LC-MS/MS</b>	Cromatografia líquida acoplada à espectrometria de massa em tandem
<b>LMW</b>	Padrão de peso molecular (do inglês <i>Low Molecular Weight</i> )
<b>LTQ</b>	<i>Linear Trap Quadrupolo</i>
<b>M</b>	Molar
<b><i>m/z</i></b>	Razão massa/carga
<b>MALDI</b>	<i>Matrix Assisted Laser Desorption Ionization</i>
<b>MDS</b>	Escalonamento multidimensional (do inglês <i>Multidimensional scaling analysis</i> )
<b>MM</b>	Massa molecular
<b>MS</b>	Espectrometria de massas (do inglês <i>Mass Spectrometry</i> ).
<b>MS/MS</b>	Espectrometria de massas em Tandem (do inglês <i>Tandem Mass Spectrometry</i> )
<b>MS1</b>	Espectro do perfil de massas
<b>MS2</b>	Espectro do perfil de massas de íons dissociados
<b>MudPIT</b>	<i>Multidimensional Protein Identification Technology</i>
<b>N</b>	Normalidade
<b>NAD+</b>	Forma oxidada de nicotinamida adenina dinucleotídeo (do inglês <i>Nicotinamide Adenine Dinucleotide - NAD</i> )

<b>NADH</b>	Forma reduzida de nicotinamida adenina dinucleotídeo (do inglês <i>Nicotinamide Adenine Dinucleotide - NAD</i> )
<b>NIST</b>	<i>National Institute of Standards and Technology</i>
<b>p.p.m.</b>	Parte por milhão
<b>PBS</b>	Tampão fosfato-salino (do inglês <i>Phosphate Buffered Saline</i> ).
<b>PCNA</b>	Antígeno nuclear de proliferação celular
<b>PCR</b>	Reação em cadeia pela polimerase (do inglês <i>Polymerase Chain Reaction</i> )
<b>PGAs</b>	Pepsinogênios
<b>PSA</b>	Antígeno prostático específico
<b>RBF</b>	<i>Radial-Basis Function</i>
<b>RCF</b>	<i>Relative Centrifugal Force</i>
<b>R<sub>f</sub></b>	Fator de retenção
<b>RF</b>	Radiofrequência
<b>RMN</b>	Ressonância magnética nuclear
<b>RNA</b>	Ácido ribonucleico
<b>RP</b>	Fase Reversa (do inglês <i>Reverse Phase</i> )
<b>RT- PCR</b>	<i>Real-Time Polymerase Chain Reaction</i>
<b>SCX</b>	Troca catiônica forte (do inglês <i>Strong Cation Exchange</i> )
<b>SDS</b>	Dodecil sulfato de sódio
<b>SDS-PAGE</b>	Eletroforese em gel de poliacrilamida (do inglês <i>Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis</i> )
<b>SELDI</b>	<i>Surface-enhanced laser desorption/ionization</i>
<b>Taq</b>	Enzima Taq DNA polimerase
<b>TBE</b>	Tris/Borato/EDTA
<b>TEMED</b>	Tetrametiletilenodiamina
<b>TFA</b>	Ácido trifluoracético

<b>TGF-<math>\beta</math></b>	Fator de transformação do crescimento beta ( <i>do inglês Transforming growth factor beta</i> ).
<b>TNM</b>	Tumor, Nódulo e classificação da Metástase
<b>TOF</b>	Tempo de voo ( <i>do inglês Time Of Flight</i> )
<b>Tris-HCl</b>	<i>Tris (hydroxymethyl) aminomethane - Hydrochloric acid</i>
<b>w/v</b>	Razão peso/volume ( <i>do inglês weight/volume</i> )
<b>x g</b>	<i>Times gravity</i>
<b>XIC</b>	Cromatogramas de íons extraídos

## LISTA DE FIGURAS

Figura 1- Processo de cancerização da célula, onde um agente cancerígeno entra em contato com as células normais desencadeando a malignização dessas .....	25
Figura 2 - O polimorfismo no promotor ILB1 relacionado ao risco de câncer gástrico indica a troca da base nitrogenada citosina (C) por timina (T), gerando uma sequência TATAAA, o que aumenta a força desse promotor. ....	32
Figura 3 - Fluxo de trabalho de um típico experimento de proteômica desde a extração até a identificação e caracterização das proteínas usando dados de espectrometria de massas em tandem.....	38
Figura 4 - Esquema da eletroforese em gel de poliacrilamida, onde ao aplicar um campo elétrico, as proteínas do padrão (P) e das amostras migram através da malha de acrilamida do pólo negativo (ânodo) em direção ao pólo positivo (cátodo) .....	40
Figura 5 - Tecnologia de identificação de proteínas multidimensional (MudPIT), a qual utiliza uma cromatografia de nanofluxo com coluna de troca iônica seguido de coluna de fase reversa acoplada a espectrometria de massa em tandem .....	43
Figura 6 - Esquema de um espectrômetro de massas. ....	45
Figura 7 - Micro gotas são expelidas do cone de Taylor, à medida que se aplica uma diferença de potencial entre a extremidade do capilar e a entrada no espectrômetro de massas, sofrendo ainda a dessolvatação, resultando em gotículas altamente carregadas para o analisador. ....	47
Figura 8 – (a) O quadrupolo “filtra” e “guia” os íons até a armadilha de íons. O íon representado pelo caminho azul demonstra uma trajetória instável, colidindo em uma das varetas de onde o campo elétrico se origina. O íon representado pelo caminho vermelho possui uma trajetória estável, seguindo para a armadilha de íons. b) A armadilha de íons ( <i>ion trap</i> ) seleciona os íons de acordo com sua razão $m/z$ à medida que a intensidade dos campos eletromagnéticos cresce fazendo com que os íons saiam do “aprisionamento” e sejam expelidos e detectados, obtendo-se o MS <sup>1</sup> . ....	48

Figura 9 – Esquema do espectrômetro de massas Orbitrap Velos (Thermo, San José, Califórnia) com a fonte ionizante eletrospray (ESI) e os analisadores do tipo quadrupolo/armadilha de íons (linear trap) e Orbitrap.....	49
Figura 10 - A sequência de aminoácidos do peptídeo PVNFKFLSH pode ser elucidada com o MS2 (espectrometria de massa em tandem) acima pela diferença de massa entre “picos” .....	50
Figura 11 - Cromatograma exemplificando as diferenças nos padrões de expressão entre os pacientes com e sem câncer.....	53
Figura 12 - Expressão diferencial no <i>Surface-enhanced laser desorption/ionization - Time of flight</i> (SELDI-TOF) em picos na razão m/z 4665 em amostras de pacientes com estágio I/II versus estágio III/IV de câncer gástrico. C73a = Estágio I, C7-A = Estágio II, C70-b = Estágio III e C5 = Estágio IV. ....	54
Figura 13 - Amostras a serem pulverizadas (margem de ressecção e câncer).....	67
Figura 14 - Eletroforese em gel de poliacrilamida do produto da reação em cadeia pela polimease (PCR) para detecção de sequências de DNA de EBV.....	83
Figura 15 - Eletroforese em gel de poliacrilamida do produto amplificado da reação em cadeia pela polimerase (PCR) para detecção de sequências de DNA- EBV das amostras 11, 12, 13, 14,15 e 16.....	83
Figura 16 - Estômago de um paciente exemplificando a área com câncer e a margem de ressecção. ....	85
Figura 17 – Curva padrão e equação da reta utilizada.....	88
Figura 18 - Eletroforese em gel de poliacrilamida das amostras selecionadas para o estudo proteômico.....	90
Figura 19 - Diagrama de Venn das proteínas identificadas na análise das bandas diferencialmente expressadas no gel unidimensional. ....	91
Figura 20 - Multifuncionalidades que a filamina A pode apresentar, entre as quais se destacam: regulação da adesão celular, da integrina e motilidade celular.....	92

Figura 21 - Cromatograma da amostra seis na troca iônica referente a um paciente com câncer. ....	95
Figura 22 – Cromatograma da amostra doze na troca iônica referente à margem de ressecção de um paciente. ....	96
Figura 23 – Cromatograma da amostra dois na troca iônica referente a um paciente controle. ....	96
Figura 24 - Diagrama de Venn de proteínas diferencialmente expressadas identificadas na análise da tecnologia de identificação de proteínas multidimensional (MudPIT). ....	97
Figura 25 - Análise do TFold comparando perfis proteômicos de proteínas identificadas em duas ou mais replicatas biológicas de amostras de indivíduos controle frente às da margem de ressecção . ....	102
Figura 26 - A análise de <i>Multidimensional Scaling</i> (MDS) dos dados obtidos a partir da tecnologia de identificação de proteínas multidimensional (MudPIT) de amostras de tecidos de indivíduos controle, com câncer e a respectiva margem de ressecção. ....	103
Figura 27 – Comparação dos perfis metabolômicos das amostras dos indivíduos controle, pacientes com câncer e a correspondente margem de ressecção. ....	108
Figura 28 - Comparação dos perfis medianos (normalizados) do câncer e do controle assim como a diferença mediana entre eles. ....	110
Figura 29 - Comparação dos perfis medianos (normalizados) do câncer e da margem de ressecção, e a diferença mediana entre eles. ....	111
Figura 30 - Comparação dos perfis medianos (normalizados) do controle e da margem de ressecção, e a diferença mediana entre eles. ....	112
Figura 31 – Dendrograma das distâncias euclidianas das réplicas das amostras normalizadas usadas nesse estudo. ....	115
Figura 32 - Dendrograma das distâncias euclidianas da média das amostras normalizadas usadas nesse estudo. ....	116

**LISTA DE GRÁFICOS**

Gráfico 1 – Incidência do câncer gástrico por faixa etária, em mulheres e homens, registrada pela Fundação Centro de controle em oncologia (Manaus) no ano de 2000 a 2004. ....	26
Gráfico 2 - Estimativa para o ano 2012 das taxas relativas de incidência por 100 mil novos casos de câncer gástrico nas regiões do Brasil. Cálculo realizado para as capitais. ....	29

**LISTA DE TABELAS**

Tabela 1 – Causas de câncer gástrico em humanos.....	29
Tabela 2 – Metabólitos diferenciáveis encontrados em cromatogramas de cromatografia gasosa/espectrometria de massas (CG/MS) ao comparar pacientes com câncer gástrico e normais.....	56
Tabela 3 - Características clínicas e experimentais para a presença do vírus Epstein-Barr (EBV) detectada pela reação em cadeia pela polimerase em indivíduos com e sem câncer.....	81
Tabela 4 – Características dos indivíduos envolvidos no estudo proteômico.....	86
Tabela 5 – Valores e médias das absorvâncias das concentrações do padrão. ....	87
Tabela 6 – Valores e médias das absorvâncias de cada amostra analisada. ....	87
Tabela 7 – Quantidade de microgramas totais e a concentração de cada amostra durante a dosagem.....	88
Tabela 8 – Informações analíticas referentes à concentração de proteínas da dosagem e do sobrenadante.....	89
Tabela 9 – Características dos indivíduos envolvidos no estudo metabolômico. ....	105
Tabela 10 – Agrupamento de k-Means com a mediana dos valores normalizados para os três tipos de tecido utilizados no estudo .....	118

## **1. INTRODUÇÃO**

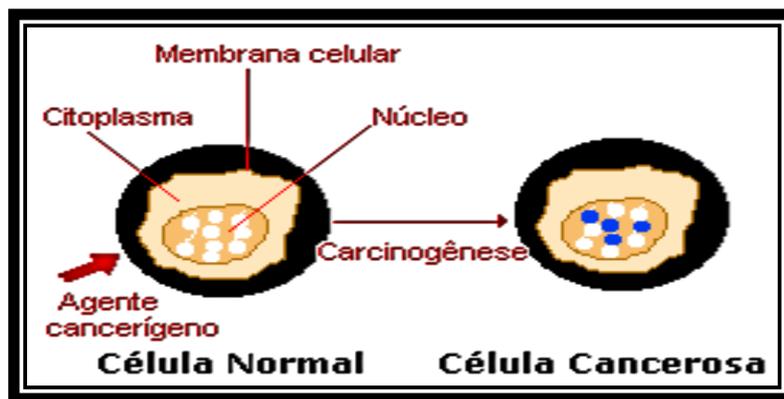
### **1.1 Câncer**

O câncer não é uma doença única, mas um acúmulo anormal de células causado por alterações genéticas e epigenéticas, ou seja, uma doença com expressão gênica alterada considerada multifatorial pelo envolvimento poligênico e com relação ambiental (Gartner et al., 1999; Nussbaum et al., 2008). De maneira simplificada, o organismo humano possui um ciclo celular controlado, no qual existe um arranjo de genes que regulam o crescimento celular, fazendo com que as células proliferem à medida que sofram mitose, e a morte celular, removendo as células de um tecido que sofreram desgaste. O desequilíbrio entre esses processos aliados a possíveis mutações nas células somáticas, visto que podem ocorrer vários erros de replicação por base de DNA durante o tempo de vida de um adulto, podem ocasionar o desenvolvimento do câncer (Nussbaum et al., 2008).

Uma vez que as células cujo material genético está alterado passam a receber instruções erradas para as suas atividades, por exemplo: a produção de certas proteínas vitais ao metabolismo, as quais ao perderem a funcionalidade ou deixarem de agir corretamente afetam o equilíbrio do metabolismo como um todo proporcionando um processo em cascata, no qual o conjunto de mutações vai se intensificando e a doença se prolifera rapidamente. Este pode ser hereditário como uma herança através da linhagem germinativa e, portanto, já inerente em cada célula do corpo. Ou esporádico (o mais comum) como resultado de uma desordem no processo citado acima (Nussbaum et al., 2008).

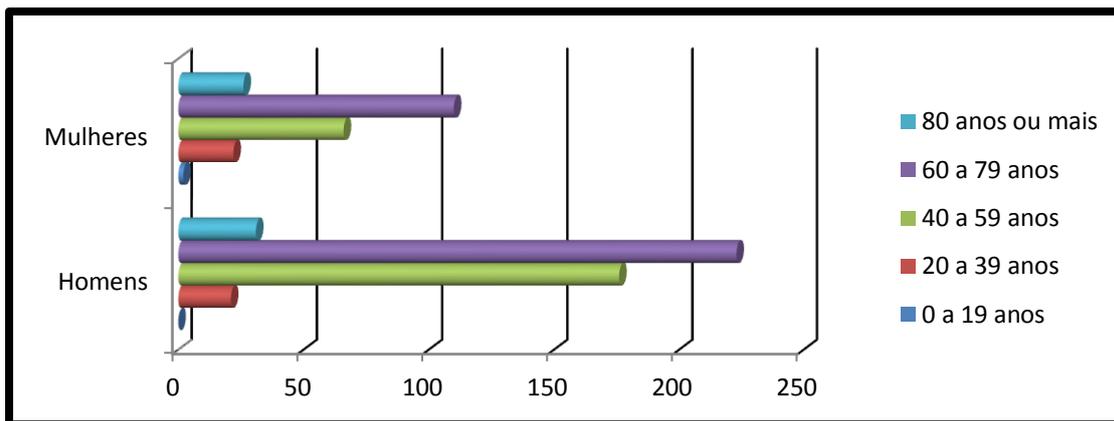
Os genes nos quais as mutações que causam esta doença ocorrem, são classificados em dois tipos: os oncogenes e os genes supressores tumorais. Ambos estão diretamente relacionados ao crescimento celular seja pela codificação de proteínas de sinalização ou de regulação ou outra característica envolvida no processo. Em especial, os oncogenes são genes mutantes dos proto-oncogenes, uma classe de genes normais que participam do ciclo celular. Eles têm um efeito dominante e uma vez ativados ou hiperexpressos, é iniciada a mudança do fenótipo de uma célula, resultando na malignização (cancerização) das células normais (**Figura 1**). Essas células diferentes são denominadas cancerosas (Nussbaum et al., 2008; [http://www.inca.gov.br/conteudo\\_view.asp?id=322](http://www.inca.gov.br/conteudo_view.asp?id=322)).

A incidência do câncer varia entre as populações de diferentes países e regiões e responde mundialmente por mais de 20% de todas as mortes. Em países desenvolvidos, essa doença é responsável por mais de 10% da renda destinada a cuidados médicos (Nussbaum et al., 2008). Entre os fatores envolvidos no mecanismo de carcinogênese estão: alterações múltiplas acumuladas, incluindo inativação de genes supressores tumorais, ativação de oncogenes e anormalidades nos reguladores do ciclo celular e dos fatores de crescimento (Dantas et al., 2009).



**Figura 1-** Processo de cancerização da célula, onde um agente cancerígeno entra em contato com as células normais desencadeando a malignização dessas. Fonte: [http://www1.inca.gov.br/conteudo\\_view.asp?id=322](http://www1.inca.gov.br/conteudo_view.asp?id=322)

Para a maioria dos cânceres, as taxas de sobrevivência dependem da detecção precoce da doença; tipicamente, quanto mais precocemente o câncer é detectado e diagnosticado, maior a taxa de sobrevivência. Na região amazônica, segundo uma análise no banco de dados da FCECON (Fundação Centro de Controle de Oncologia do Estado do Amazonas, uma instituição referência em detecção, prognóstico e tratamento do câncer), contida no INCA (**Gráfico 1**), a maior incidência ocorre na população com faixa etária entre 60 a 79 anos, havendo em geral mais vítimas do sexo masculino. Além disso, os marcadores existentes como, por exemplo: CA 125 – câncer de ovário, CA 15.3 – câncer de mama, CA 19.9 – câncer gastrointestinal, CEA – câncer de cólon, mama, pulmão e pancreático (Diamandist, 2004; Almeida et al., 2007), deixam a desejar em termos de especificidade e sensibilidade, tornando-se necessária a identificação de novas proteínas séricas e metabólitos associados a tumores que possam ser usados clinicamente como novos biomarcadores e permitam detectar o câncer precocemente e monitorar a terapia ou a recorrência da doença (Barril et al., 1999).



**Gráfico 1** – Incidência do câncer gástrico por faixa etária, em mulheres e homens, registrada pela Fundação Centro de controle em oncologia (Manaus) no ano de 2000 a 2004. Fonte: <http://www.inca.gov.br/cancernobrasil/2010/docs>.

## 1.2 Câncer gástrico

O câncer gástrico é uma neoplasia maligna e como tal se diferencia dos tumores benignos, que são autolimitados em seu crescimento e não invadem tecidos adjacentes (embora alguns tumores benignos sejam capazes de se tornarem malignos). Os tumores malignos apresentam-se predominantemente sob a forma de três tipos histológicos: o adenocarcinoma, responsável por 95% dos tumores gástricos, o linfoma, diagnosticado em cerca de 3% dos casos, e o estromal gastrointestinal – GIST (<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/estomago>). A maioria surge no antro e no corpo gástrico, havendo ainda dois subtipos histológicos comuns, o intestinal e o difuso (Schulz, 2005).

Dentre todos os tipos de câncer, o gástrico é um dos mais comuns e o segundo em número de mortes no mundo. A maioria dos pacientes é diagnosticada com a doença em estado avançado, o que diminui drasticamente as opções de tratamento e resulta em uma média de sobrevivência em torno de 5 anos apenas para 25% dos casos (Ebert et al., 2006).

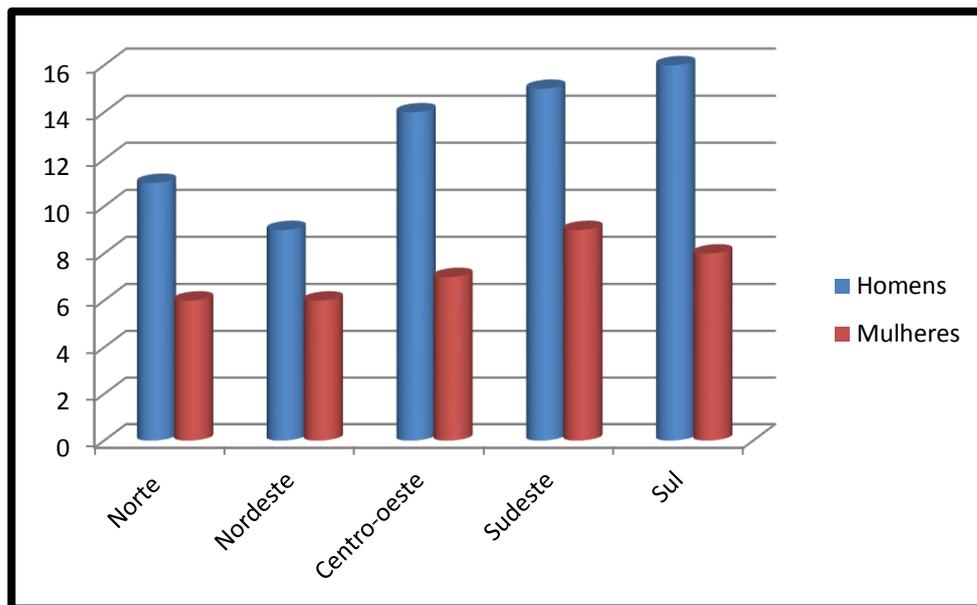
Conforme citado anteriormente, há prevalência do sexo masculino, como maior detentor de câncer gástrico, sobretudo acima dos 40 anos, porém essa diferença de incidência por sexo não tem sido relacionada a questões hormonais particularmente e sim ao próprio tipo histológico ou a localização do tumor (Mauad et al, 2000; Theuer et al., 1996). Embora na faixa etária mais jovem (abaixo de 40 anos) a incidência seja de apenas 5% do total dos casos, a sobrevida é bem inferior, pois esses pacientes são diagnosticados tardiamente, uma vez que suas lesões podem ser confundidas com patologias benignas (Mauad et al, 2000; Theuer et al., 1996). Quanto à ocorrência nesse grupo mais jovem, não há dados substanciais que apontem uma diferença na distribuição por sexo (1:1). Todavia, houve um aumento nos estudos relacionados a esse assunto, sendo iniciado por McNeer (McNeer, 1941) e se estendendo até os dias de hoje. Em Manaus, segundo estudos realizados na FCECON sob a responsabilidade de um dos colaboradores do projeto, Dr. Sidney Chalub, foi observada uma porcentagem maior, cerca de 10%, de pacientes com câncer gástrico nas pessoas com menos de 40 anos de idade. Em geral parece não haver barreiras quanto à faixa etária, pois foi relatado diagnóstico positivo até numa criança de 10 anos de idade (Mauad et al, 2000).

No contexto mundial, o Japão é o que apresenta a mais alta taxa de incidência deste tipo de câncer na faixa etária acima de 50 anos (59,9 a 91,6/100.000 em homens e 23,8 a 38,9/100.000 em mulheres), sendo esta associada principalmente à alimentação como, por exemplo: peixes conservados em sal. A incidência de câncer de estômago em imigrantes japoneses residentes em São Paulo excede àquela encontrada nos brasileiros

não japoneses em cerca de 30%, sendo pouco menor que a incidência observada nos japoneses residentes no Japão. (Parkin et al., 1997).

No Brasil, o câncer de estômago em homens é o segundo mais frequente nas regiões Norte (11/100.000) e Nordeste (9/100.000). Nas regiões Centro-Oeste (14/100.000), Sul (15/100.000) e Sudeste (14/100.000), é o quarto. Para as mulheres é o quarto mais frequente na Região Norte (6/100.000); é o quinto na Região Centro-Oeste (7/100.000) e nas demais regiões, Sul (8/100.000), Sudeste (9/100.000) e Nordeste (6/100.000), é o sexto. Novamente, ele apresenta maior incidência em homens, porém nas mulheres, corresponde a cerca de 5% de todos os tumores femininos. Estima-se para o Brasil que no ano de 2012, o número de novos casos desse tipo de câncer será de 12.670 nos homens e de 7.420 nas mulheres. Estes valores correspondem a um risco estimado de 13 casos novos a cada 100 mil homens e 7 para cada 100 mil mulheres (Guerra et al., 2005; <http://www.inca.gov.br/estimativa/2012/>).

Tendo em vista a enorme quantidade de cânceres conhecidos, o câncer gástrico possui certa prevalência na região Amazônica. Como pode ser observado no **Gráfico 2**, os índices de taxa relativas para novos casos de câncer gástrico calculado com base em cada 100.000 habitantes nas capitais da região norte se assemelham aos de outras regiões do país. Estudos anteriores realizados por Guerra et al.(2005) relacionam o desenvolvimento do câncer gástrico ao baixo poder econômico e hábitos alimentares, e em especial na região Norte pode estar associado ao maior consumo de farinha de mandioca, alimento que pode liberar cianeto e modificar o pH da mucosa gástrica.



**Gráfico 2** - Estimativa para o ano 2012 das taxas relativas de incidência por 100 mil novos casos de câncer gástrico nas regiões do Brasil. Cálculo realizado para as capitais. Fonte: <http://www.inca.gov.br/estimativa/2012/>.

### 1.3 Origem do câncer gástrico

Diversos fatores podem contribuir individualmente ou em conjunto para o surgimento dessa neoplasia, desde fatores genéticos até hábitos alimentares, alguns destes descritos na **tabela 1** (Schulz, 2005).

**Tabela 1** – Causas de câncer gástrico em humanos.

Fatores envolvidos no câncer gástrico
<i>Helicobacter pylori</i>
Nitrosaminas
Baixa consumo de frutas frescas e vegetais
Álcool
Fumo
Pré-disposição genética pelo alto risco de mutações nos genes ( <i>CDH1</i> )
Pré-disposição genética pelo polimorfismo ( <i>IL1B</i> )

O *Helicobacter pylori* é considerado o principal agente etiológico em mais de 95% das gastrites crônicas e responsável pelo aumento do risco de câncer em até nove vezes (César et al, 2002), sendo considerado um carcinógeno humano. Em torno de 50% da população mundial possui cepas dessa bactéria, porém menos de 10% desenvolve doença inflamatória e úlcera, e um número menor ainda vem a apresentar câncer. A interação entre as células epiteliais da mucosa do estômago e a bactéria ocorre mediante a atuação de várias proteínas bacterianas, por exemplo: a proteína VacA tem efeitos diretos sobre a regulação do ciclo celular e a apoptose das células epiteliais gástricas, ou seja, ela induz a vacuolização e morte das células do hospedeiro. Isso leva à liberação de nutrientes e ainda facilita a introdução da *H. pylori* através da barreira mucosa. Além disso, essa proteína também é imunossupressora, diminuindo as respostas por macrófagos e células-T (Schulz, 2005).

A ação deste agente carcinogênico é indireta, provocando gastrite, metaplasia e displasia. A infecção é adquirida na infância e a sua prevalência entre a população está relacionada com o *status* sócio-econômico, onde geralmente pessoas mais pobres apresentam a infecção por um tempo mais prolongado (César et al, 2002).

O risco do câncer gástrico também pode estar relacionado a presença de co-cancerígenos na dieta, especificamente sal e álcool excessivo, os quais podem atuar como irritantes, agravando a destruição dos tecidos e ocasionando uma inflamação da mucosa gástrica, o que pode corroborar para uma infecção por *H. pylori* (César et al, 2002). As nitrosaminas, ( $R_2N-N=O$ ), são também potentes agentes cancerígenos e estão relacionadas a diversos

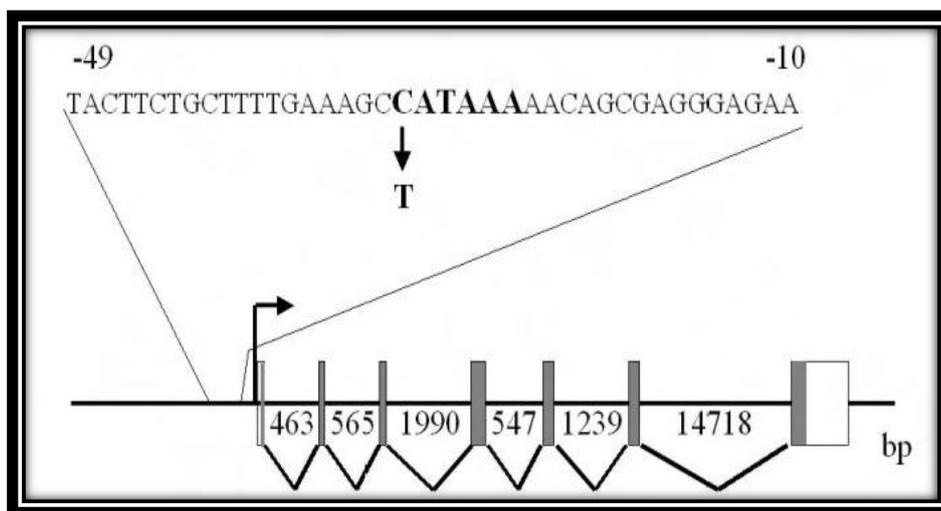
gêneros alimentícios, pois os nitritos e nitratos usados para conservar alguns tipos de alimentos, como carnes, peixes, salsichas e outros tipos de enlatados, são transformados nelas no estômago. Elas são agentes alquilantes que ao reagirem com bases de DNA podem causar mutações. Algumas dessas mutações foram encontradas em cânceres gástricos nos genes: TP53, CDH1 (E-Caderina), e CTNNB1 ( $\beta$ -Catenina) (Schulz, 2005; Macdonald et al., 2004).

Entre os fatores genéticos têm-se as mutações no gene CDH1, responsável por codificar a E-caderina. Esta é uma glicoproteína de membrana, que possui a função de mediar a adesão homolítica entre as células epiteliais, ou seja, ela conecta o citoesqueleto pela forma de cateninas. Algumas dessas mutações são pequenas deleções na fita ou aquelas que afetam o encaixe do sítio de domínio de adesão, ocasionando uma proteína mais curta que o normal, que resulta na perda da habilidade dessas interações. Essa disfunção da E-caderina acarreta um padrão de crescimento difuso das células tumorais, e pode contribuir também para a desregulação da via WNT, o que geralmente ocorre durante a progressão do tumor. Raros casos familiares deste tipo de câncer são causados por mutações germinativas (hereditárias) da E-caderina, e o gene CDH1 se comporta como um clássico gene supressor tumoral para esse tipo específico de doença (Schulz, 2005; Alberts et al., 2008).

Outros fatores genéticos são os possíveis polimorfismos principalmente no gene IL1B, o qual codifica as proteínas citocinas (um grupo de moléculas envolvidas na emissão de sinais entre as células durante o desencadeamento das respostas imunes) em humanos. Uma alteração no promotor pode influenciar o nível de expressão da citocina. Na **figura 2**, é mostrado um

exemplo de polimorfismo desse gene, onde a troca entre as bases nitrogenadas citonina por timina altera substancialmente a intensidade do promotor desse gene, pois os alelos-T de transcrição se ligam mais fortemente do que em alelos-C, logo a expressão de IL1B é mais fortemente induzida. A presença desses alelos-T está associada a aproximadamente dez vezes o aumento do risco de úlceras gástricas e do câncer (Schulz, 2005; Macdonald et al., 2004; Alberts et al., 2008).

O polimorfismo nesses genes, incluindo os polimorfismos IL1B-31C/T está na grande maioria das vezes em ligação de desequilíbrio com outro grupo de genes. Significando que uma forma particular (alelo) do gene IL1B é encontrada junto com outra forma particular do gene IL1R. Este efeito age de maneira sinérgica com o genótipo da *Helicobacter pylori*, pois se, por exemplo: o risco de câncer gástrico para determinada população é de cerca de 2 a 3 vezes mais elevado com a presença dessa bactéria; ele pode aumentar cerca de até 100 vezes mais nos portadores do gene IL1B-31T (Schulz, 2005).



**Figura 2** - O polimorfismo no promotor ILB1 relacionado ao risco de câncer gástrico indica a troca da base nitrogenada citosina (C) por timina (T), gerando uma sequência TATAAA, o que aumenta a força desse promotor. Fonte: Schulz, 2005.

Diante dos fatores acima citados, não é surpresa que o câncer gástrico seja considerado uma doença multifatorial. Contudo, vale enfatizar que essa neoplasia começa a demonstrar indícios de uma diminuição em certos países, inclusive no Brasil. A explicação vem de uma melhor conservação dos alimentos aliados à modificação no hábito alimentar, ou seja, o aumento da ingestão de frutas, legumes e verduras frescas. Essa alimentação se torna um fator protetor visto que esses alimentos têm vitaminas com propriedades antioxidantes. Outra característica pertinente é a diminuição nas taxas referentes a permanência da infecção pela bactéria *Helicobacter pylori*, devido ao surgimento de medicamentos mais eficazes juntamente com uma melhoria no saneamento básico e mudanças no estilo de vida da população.

#### **1.4 Biomarcadores**

O advento da engenharia genética, o surgimento e o desenvolvimento de novas metodologias e a conclusão do sequenciamento do genoma humano em 2001, possibilitaram à área da biologia molecular um grande avanço, tornando possível o estudo de maneira mais abrangente e complexa dos sistemas biológicos, e o desafio da busca por biomarcadores.

No caso específico do câncer, um biomarcador pode ser entendido como qualquer molécula ou característica biológica que indique alterações do estado fisiológico normal para o patológico. É definido como uma característica biológica que fornece informações tanto para diagnóstico, prognóstico, valor preditivo ou informações terapêuticas sobre uma determinada doença

(Carvalho et al., 2006a). Alguns aspectos fundamentais devem ser considerados quando se procura um biomarcador:

1. A seleção do material biológico a ser utilizado, pois vários tipos de amostras estão disponíveis para a pesquisa de câncer, por exemplo: o tecido tumoral e os fluidos do corpo tais como soro ou plasma;
2. A definição do alvo de um biomarcador: DNA, RNA ou proteínas. Pois estes podem ser utilizados como biomarcadores, e a escolha do tipo de alvo tem implicações relevantes;
3. A incorporação das características de diagnóstico precoce da neoplasia e de sua origem, estabelecimento da extensão da doença, de maneira a permitir o acompanhamento temporário das mudanças do tumor (Silveira, 2005).

Da mesma forma, se um marcador é usado num cenário de previsão uma correlação do biomarcador com a quantificação do estágio tumoral também é desejável. Finalmente, reprodutibilidade e facilidade de interpretação são valiosos recursos aplicáveis (Jimeno et al., 2006).

Portanto, há um enorme esforço para identificar, caracterizar e validar biomarcadores significativos, isso porque o sucesso de desenvolvimento destes representa um passo em frente à individualização do diagnóstico precoce, na terapia e acompanhamento a partir de uma perspectiva orientada clinicamente, com ênfase nos biomarcadores que são candidatos a serem incorporados nas etapas sucessivas de diagnóstico, seleção terapêutica e monitoramento da eficácia terapêutica (Jimeno et al., 2006).

Finalmente, vale ressaltar que a grande maioria dos trabalhos objetivando a busca de biomarcadores tem como resultado um painel de candidatos. Isto porque para ser considerado como um biomarcador, ele deve

passar por cinco fases experimentais, as quais são: 1) a exploração pré-clínica; 2) testes clínicos e validação; 3) retrospectiva longitudinal; 4) triagem prospectiva e 5) controle da doença. Esse processo é bastante complicado, uma vez que o número de pacientes necessário para cada etapa é elevado, passando de centenas de indivíduos e também pela necessidade da presença de vários órgãos estarem envolvidos para a efetiva utilização e aprovação, como por exemplo: a NIST (*National Institute of Standards and Technology*). Por isso que o presente trabalho busca candidatos a biomarcadores para ajudar a compreender melhor a doença em um nível molecular, porém não a aplicabilidade imediata no diagnóstico clínico (Pepe et al., 2001; Barker, 2003).

### **1.5 Biologia molecular – Reação em cadeia pela polimerase (PCR-*Polymerase chain reaction*)**

A biologia molecular tem o estudo dirigido especialmente para a estrutura e função do material genético, os produtos de expressão correspondentes e as proteínas; de maneira a investigar a interação entre os diversos sistemas celulares, principalmente no que diz respeito à relação entre o DNA, RNA e a síntese protéica (Alberts et al., 2008). Entre as diversas técnicas existentes que auxiliam nesse estudo tem-se o PCR, reação em cadeia pela polimerase. Essa técnica avançou bastante desde a primeira utilização por Saiki e colaboradores (Saiki et al., 1985).

É um método que faz uso de oligonucleotídeos (*primers*) como demarcadores de regiões específicas da molécula de DNA para a ação da DNA polimerase, possibilitando a amplificação de um número ilimitado de sequências-alvo. É bastante sensível e o número de ciclos de amplificação

varia de acordo com o objetivo e as condições utilizadas. Para quantificar os produtos desse método da PCR, é geralmente usado o RT-PCR (*Real-Time Polymerase chain reaction*), onde é possível monitorar a quantidade de produto amplificado em tempo real (Alberts et al., 2008; Bartlett e Stirling, 2003).

No câncer gástrico, esta técnica de PCR tem sido empregada para um estudo mais direcionado para a ocorrência e função de certos genes envolvidos nesta doença, entre eles têm-se os genes virais como o do vírus de *Epstein-Barr* (EBV). O EBV é classificado como um cancerígeno do grupo 1, de acordo com a agência internacional para pesquisa sobre câncer (IARC). Pertence à família gama-herpes, que tem uma molécula de DNA linear responsável pela codificação de aproximadamente 100 proteínas (Lima e Rabenhorst, 2006; Silva e Zucoloto, 2003). A incidência desse vírus nesse tipo de câncer tem sido encontrada num percentual de 2 a 18% da população mundial (Burke et al., 1990; Hayashi et al., 1998; Koriyama et al., 2001; Lopes et al., 2004). Um artigo recente no Brasil descreve essa incidência como de aproximadamente 8% entre os casos diagnosticados no Ceará (Lima et al., 2011).

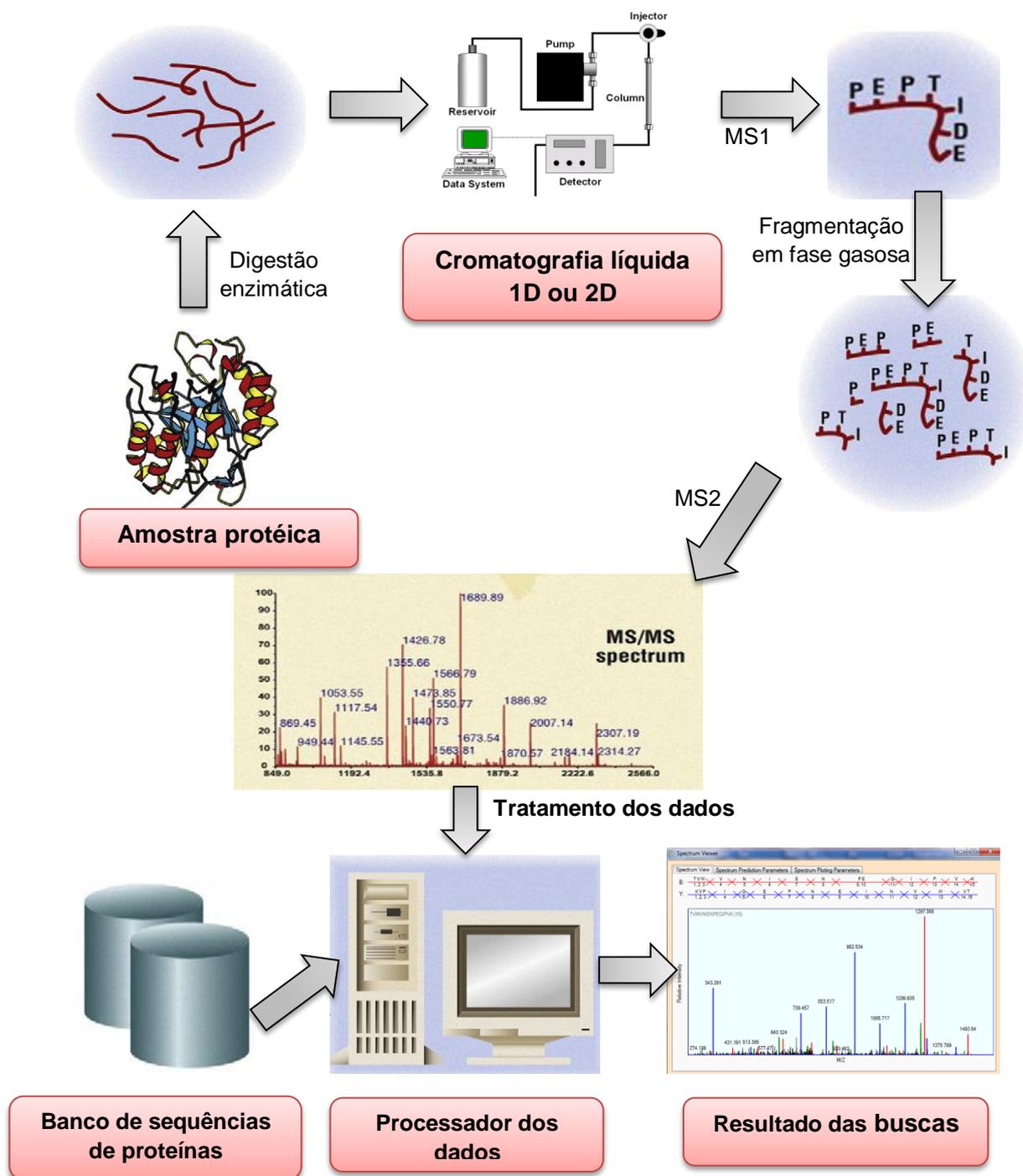
### **1.6 Proteômica**

As proteínas são a ação do organismo, ou seja, são as moléculas executoras das células, as quais realizam praticamente tudo com relação ao metabolismo. Exemplos são enzimas que catalisam processos bioquímicos; outras são reguladoras que contribuem para a expressão do genoma; várias são estruturais; entre outras diversas funções que elas podem ter. O proteoma pode ser definido como o completo conteúdo protéico expresso por um genoma ou por uma célula ou um tipo de tecido. De maneira mais abrangente neste

contexto estão inclusas a quantificação, localização, determinação de modificações pós-traducionais, mutações e polimorfismos das proteínas. A proteômica é o conjunto de técnicas que estuda esse conteúdo protéico presente em diferentes partes do organismo e como ele reage às mudanças com o tempo e as várias condições submetidas (Eidhammer et al., 2007). Através da proteômica pode-se fazer um estudo sistemático de diversas propriedades das proteínas numa forma paralela, podendo fornecer assim uma detalhada descrição da estrutura, função e controle dos sistemas biológicos. Tal fato permite, portanto, estudar perfis de expressão protéica, onde comparação de estados fisiológicos diferentes é possível (Huber, 2003; Patterson et al., 2003).

O proteoma se modifica ao longo do desenvolvimento de um organismo, diferente do genoma (conjunto de genes), o qual permanece praticamente o mesmo durante toda a vida. Logo, não é surpresa que as informações contidas no genoma não sejam suficientes para conhecer completamente seu equivalente protéico, pois o DNA não fornece toda informação sobre a estrutura, interação e funcionalidades das proteínas. Além de que geralmente as proteínas são modificadas pós-traducionalmente (Zhang et al., 2010)

O estudo da proteômica requer metodologias que possuam sensibilidade, reprodutibilidade, rapidez, facilidade e automação, fatores esses que resultam em uma análise de custo relativamente alto. Com o avanço nas técnicas proteômicas torna-se possível uma “visualização” do proteoma, o que pode auxiliar no reconhecimento de padrões podendo trazer impacto direto na clínica médica entre outras aplicabilidades (Carvalho et al., 2006a). Na **Figura 3**, é mostrado um fluxo de trabalho que geralmente é empregado nessa área.



**Figura 3** - Fluxo de trabalho de um típico experimento de proteômica desde a extração até a identificação e caracterização das proteínas usando dados de espectrometria de massas em tandem. Fonte: Modificado de Cottrell, 2011.

### 1.7 Separação de Proteínas

As técnicas de separação de proteínas são capazes de simplificar um complexo de amostras biológicas, possibilitando um aumento no número de

identificações confiáveis quando acopladas a espectrometria de massa (Veenstra et al., 2006). Logo, a escolha do método é geralmente o primeiro passo de uma aplicação proteômica. Alguns exemplos de técnicas que podem ser empregadas são: eletroforese unidimensional e/ou bidimensional em gel (SDS-PAGE), e cromatografia líquida (por exemplo: fase reversa e troca iônica, por afinidade, entre outras ou combinação delas).

### **1.7.1 Descrição das técnicas utilizadas nesse trabalho:**

#### ➤ Eletroforese em gel desnaturante de poliacrilamida (SDS-PAGE)

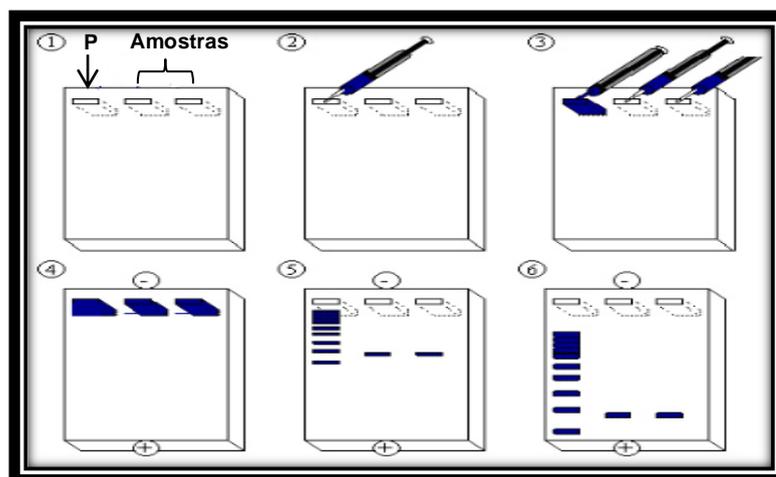
É um método utilizado para a análise das massas moleculares de proteínas. Este utiliza o dodecilssulfato de sódio (SDS) para desnaturar as proteínas no gel de poliacrilamida, ou seja, convertê-las numa estrutura linear, conferindo uma densidade de carga uniforme, o que minimiza assim possíveis interferências. O uso do ditioneitol (DTT) ou 2-mercaptoetanol, como agentes redutores em altas temperaturas desfazendo as pontes dissulfetos, também auxiliam na eliminação da estrutura tridimensional dos polipeptídeos. Tais fatores são usados com o propósito de que a separação dependa apenas do volume molar (que na grande maioria das vezes está fortemente correlacionado à massa molecular) de cada amostra a ser analisada.

O gel é uma matriz constituída de polímeros de acrilamida com ligações cruzadas de *N-N*-metil-bisacrilamida, cuja porosidade é escolhida com base na composição do alvo a ser escolhido. Quanto maior a concentração de acrilamida, menores serão os poros da malha formada. Já o SDS é um

detergente anfótero com uma carga negativa dentro de uma ampla faixa de pH, cuja calda hidrofóbica interage com as cadeias polipeptídicas.

Ao aplicar um campo elétrico, as moléculas migram através da malha de acrilamida do pólo negativo (ânodo) em direção ao pólo positivo (cátodo), **Figura 4.** Dependendo do seu tamanho, cada proteína se moverá diferentemente: as proteínas menores migrarão mais rapidamente, enquanto que as maiores terão mais dificuldade em atravessar a malha do gel e, assim, se moverão mais lentamente.

Quando proteínas marcadoras (pesos moleculares conhecidos de diversas cadeias polipeptídicas) são usadas na análise, pode-se inferir a mobilidade eletroforética da proteína de interesse. Devido uma reta que pode ser utilizada como padrão para o cálculo aproximado do peso molecular. Esta é obtida pelo logaritmo da massa molecular de proteínas conhecidas no eixo y e pela distância percorrida de cada banda ( $R_f$ ) pertencente ao padrão até a frente do gel no eixo x (Fischer, 2010).



**Figura 4** - Esquema da eletroforese em gel de poliacrilamida, onde ao aplicar um campo elétrico, as proteínas do padrão (P) e das amostras migram através da malha de acrilamida do pólo negativo (ânodo) em direção ao pólo positivo (cátodo). Fonte: Modificado de <http://biomedicinabrasil.blogspot.com>.

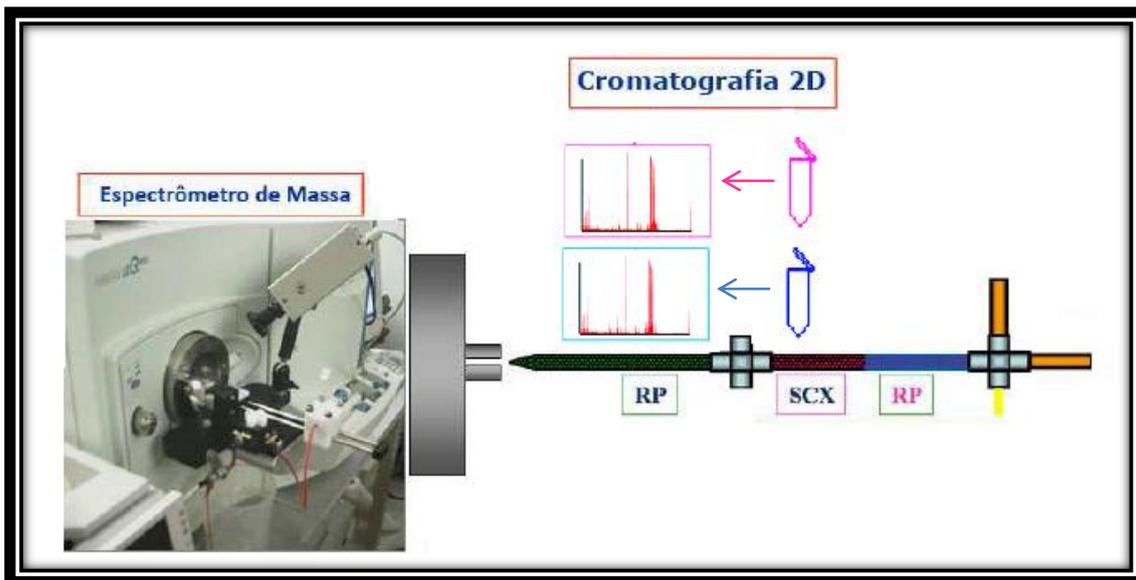
➤ Tecnologia de identificação de proteínas multidimensional (MudPIT)

A motivação de aumentar o número de proteínas identificadas em uma análise de misturas complexas levou ao desenvolvimento da tecnologia de identificação de proteínas multidimensional – MudPIT (Lin et al., 1999; Washburn et al., 2001). MudPIT combina cromatografia de troca iônica (SCX) intercalada com cromatografia de fase reversa (RP) diretamente acoplada a espectrometria de massas em tandem (**Figura 5**). Ela pode ser *on-line* (com uma coluna de cromatografia bi-fásica) ou *off-line* (a cromatografia de troca iônica é feita independentemente para então submeter as frações dessa primeira dimensão a cromatografia de fase reversa). Ambas têm vantagens e desvantagens: o modo *off-line* oferece maior flexibilidade, podendo ser otimizada a separação; o modo *on-line* por sua vez é um processo automatizado, o que evita a perda e contaminação de amostra (Motoyama et al., 2008).

A cromatografia de troca iônica (SXC) separa as proteínas de acordo com a carga através das interações eletrostáticas das cadeias laterais dos aminoácidos carregados com o tipo de resina usada. A separação se inicia com o aumento gradual da concentração iônica permitindo uma eluição diferencial do analito, visto que os de menor afinidade com a resina irão eluir primeiro. As matrizes disponíveis comercialmente para esse tipo de análise podem variar quanto ao tipo e grau de interação, tamanho da partícula e natureza química da resina, pois essas podem ser aniônicas ou catiônicas. Contudo, todas permitem o fracionamento de proteínas pela carga superficial, portanto a escolha do tipo de matriz depende do tipo de amostra.

A cromatografia de fase reversa (RP) é um método que utiliza uma fase estacionária não polar, onde a sílica (polar) foi derivatizada com hidrocarbonetos que variam em comprimento de C<sub>4</sub> a C<sub>18</sub>. A fase C<sub>18</sub> é a mais usada, por ser a mais indicada para peptídeos, além de oferecer retenção e seletividade para uma ampla gama de compostos contendo diferentes grupos polares e não polares na sua superfície. A fase móvel consiste em combinações de gradientes de água e solventes polares orgânicos, geralmente acetonitrila ou metanol. Nesse método, as moléculas mais polares irão eluir primeiro e conforme o aumento da concentração do solvente orgânico menos polar, aquelas moléculas que ficaram retidas na coluna irão eluir também. Outro quesito importante nesse processo é o tempo de retenção de cada molécula, que será específico para cada um, dependente também do grau de hidrofobicidade (Fischer, 2010; Neverova et al., 2005).

Na abordagem MudPIT, o complexo de peptídeos é aplicado a uma coluna de troca iônica, onde uma fração de peptídeos absorvida é deslocada para uma coluna de fase reversa quando submetida a um “pulso” de concentração da solução salina crescente. Os peptídeos retidos na coluna de fase reversa são então submetidos a um gradiente hidrofóbico com o aumento na concentração de solvente orgânico para serem eluídos para o espectrômetro de massas e identificados (Fischer, 2010; Zhang et al., 2010).



**Figura 5** - Tecnologia de identificação de proteínas multidimensional (MudPIT), a qual utiliza uma cromatografia de nanofluxo com coluna de troca iônica seguido de coluna de fase reversa acoplada a espectrometria de massas em tandem. Fonte: Modificado de Fischer, 2010.

### 1.8 Identificação de Proteínas

A identificação das proteínas pode ser realizada mediante o uso da espectrometria de massas em conjunto com algoritmos especializados. A massa molecular de um peptídeo, juntamente com um perfil de massas resultante da fragmentação do mesmo pode ser adquirida experimentalmente por espectrometria de massas, fornecendo o  $MS^1$  (*survey scan*) e o  $MS^2$  (também conhecido como o *peptide fragmentation fingerprint*). Tais informações podem ser submetidas a uma ferramenta de busca, por exemplo: SEQUEST (Eng et al., 1994), X-Tandem (Craig et al., 2004), Mascot (Perkins et al., 1999), ou ProLuCID (Xu et al., 2006). Estas ferramentas comparam o espectro experimental com teóricos gerados a partir de um banco de sequências de peptídeos e fornece como resultado a sequência para o qual o espectro teórico mais se assemelha ao experimental. Os resultados devem ser

probabilisticamente validados usando uma ferramenta como *DTASelect* ou *Search Engine Processor* (Cociorva et al., 2006; Carvalho et al., 2012a).

### 1.8.1 Espectrometria de massas

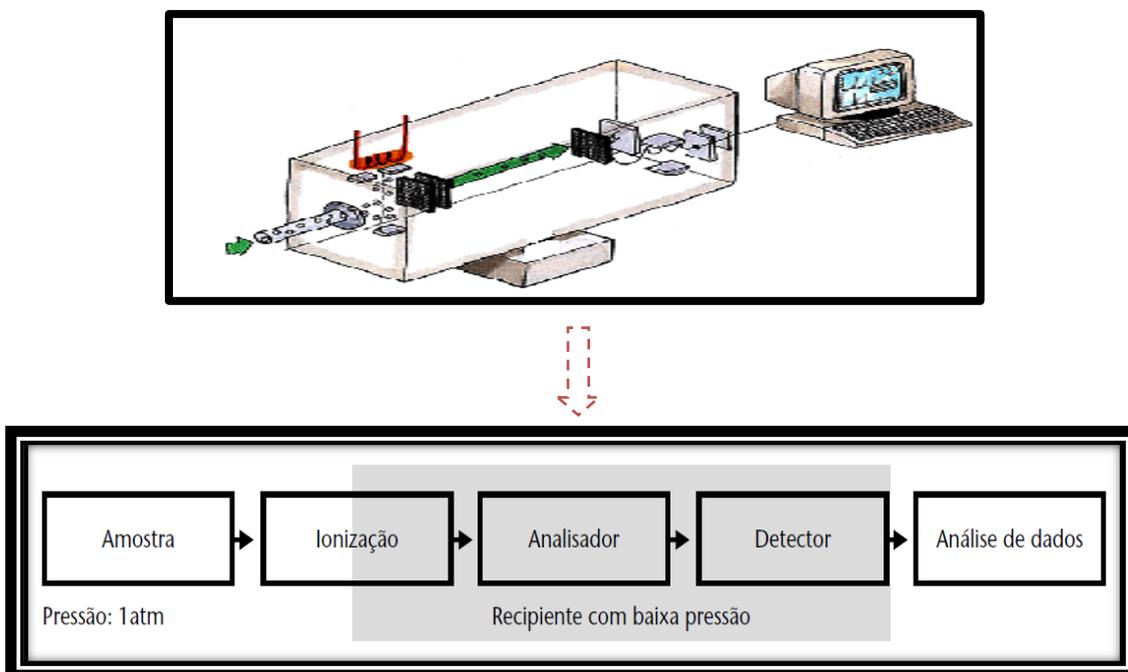
A espectrometria de massas (MS) é uma das ferramentas analíticas usadas em estudos relacionados a áreas da biologia, da medicina, alimentícia e tecnológica. Através dela podem-se comparar padrões entre duas condições específicas, estudar modificações protéicas pós-traducionais, realizar fragmentações visando a elucidação e confirmação de determinada estrutura, além de poder ser usada para quantificar biomoléculas, tais como proteínas, peptídeos, carboidratos, entre outros. A MS é considerada o estado da arte na análise de sequência de peptídeos (Liebler, 2002; Ferreira et al., 2009).

#### ➤ Funcionamento do espectrômetro de massas

O espectrômetro de massa é um instrumento analítico capaz de converter moléculas neutras em íons na forma gasosa e separá-los de acordo com a sua razão massa/carga ( $m/z$ ), utilizando para isso campos elétricos ou eletromagnéticos. Uma análise no espectrômetro de massas inicia com a introdução da amostra no equipamento por injeção direta ou através de um equipamento de separação acoplado: cromatógrafo gasoso (GC), o cromatógrafo líquido (LC) ou aparelho de eletroforese capilar (CE). Caso esse processo seja automatizado uma maior quantidade de amostras será analisada num curto período de tempo. Em seguida, os íons são gerados e separados por meio de sua relação massa-carga ( $m/z$ ) em um analisador de massas e detectados qualitativamente e/ou quantitativamente por meio de um detector, o

qual registra a corrente iônica oriunda do analisador na forma de íons. A magnitude do sinal elétrico em função da razão  $m/z$  é convertida por um processador de dados, o qual gera o espectro de massas correspondente (Gross, 2004 apud Ferreira et al., 2009; Carvalho, 2010). Em resumo, um espectrômetro de massas é composto basicamente, por uma fonte ionizante, analisador (es) e detector (es), conforme esquematizado na **Figura 6** (Carvalho, 2006b; Carvalho, 2010).

Contudo, essa tecnologia cujos dados permitem interpretar as informações contidas no genoma (Mann et al., 2003), só passou a ser usada no estudo de biomoléculas após a criação de novas técnicas de ionização que não degradassem as macromoléculas, ou seja, técnicas mais brandas, nas quais se destacam: *Matrix Assisted Laser Desorption Ionization* - MALDI (Karas et al., 1987) e *ElectroSpray Ionization* – ESI (Fenn et al., 1989).

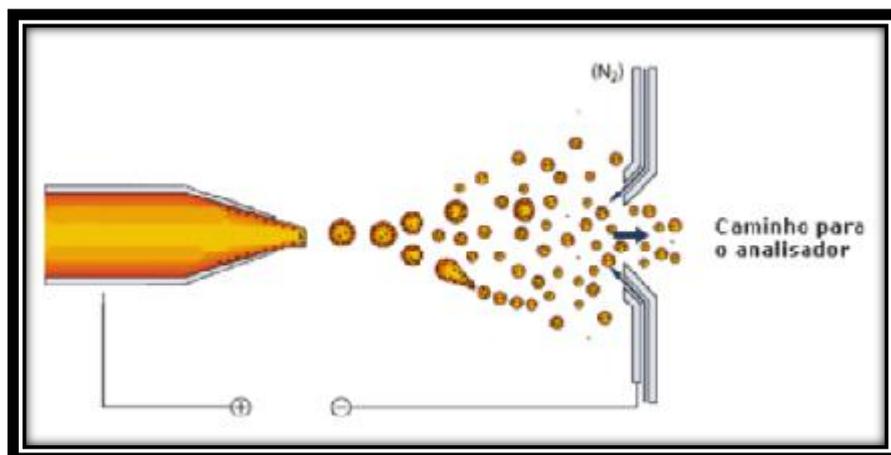


**Figura 6** - Esquema de um espectrômetro de massas. Fonte: Adaptado de Carvalho et al., 2006b.

O MALDI é um método no qual a amostra é co-cristalizada numa matriz composta de ácido de peso molecular conhecido. Essa matriz absorve a energia do laser e a transfere para o analito acidificado. O aquecimento rápido pelo laser causa a dessorção da matriz e os íons do analito ( $[M+H]^+$ ) entram na fase gasosa. Várias centenas de disparos de laser são necessários para atingir uma relação sinal/ruído aceitável para a detecção do íon. O analisador geralmente empregado nesse método é o tempo de voo (TOF), no qual os íons cruzam uma região livre de campo elétrico numa velocidade inversamente proporcional a raiz quadrada da razão  $m/z$ . Logo, os íons com menor razão  $m/z$  chegam mais rapidamente ao detector. Algumas desvantagens desse método são a baixa reprodutibilidade e a dependência da preparação de amostra (Yates et al., 2009; Eidhammer et al., 2007; Fischer, 2010).

A ionização por spray de elétrons, ESI, ao contrário do MALDI, produz íons a partir de uma solução. Pela aplicação de uma diferença de potencial entre a extremidade do capilar na câmara de ionização e o orifício de entrada no espectrômetro de massas concentram-se os íons na extremidade da gota da solução, onde se têm duas forças atuantes: a força eletrostática e a tensão superficial. À medida que há o aumento da densidade das cargas positivas, a forma esférica das micro-gotas se modifica para de um cone, denominado “cone de Taylor”. Ao intensificar esse processo, há uma desestabilização do menisco que causa a ruptura do cone, transformando-o em um “spray eletrolítico” de micro-gotas altamente carregadas (**Figura 7**), as quais sofrem dessolvatação pelo aumento da temperatura, diminuindo a concentração do solvente, aumentando as forças eletrostáticas repulsivas até o limite de Rayleigh, quando ocorre a “explosão coulombiana”. Então, essas gotículas

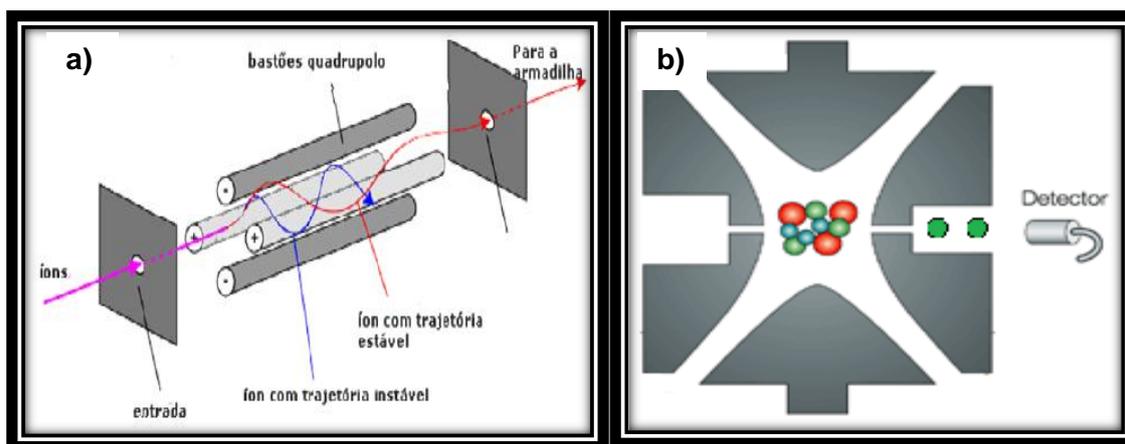
altamente ionizadas desfeitas transferem carga às moléculas do analito, os quais seguem até o analisador como íons discretos. A ESI é bastante utilizada devido à compatibilidade com vários instrumentos, e o desenvolvimento do micro e do nano-ESI possibilitou uma melhoria na sensibilidade do método (Carvalho, 2006b; Carvalho, 2010; Yates et al., 2009).



**Figura 7** - Micro gotas são expelidas do cone de Taylor, à medida que se aplica uma diferença de potencial entre a extremidade do capilar e a entrada no espectrômetro de massas, sofrendo ainda a dessolvatação, resultando em gotículas altamente carregadas para o analisador. Fonte: Fischer, 2010.

Pode-se usar dois analisadores em conjunto com o ESI, por exemplo: o conjunto armadilha de íons / Orbitrap. Neste, os íons oriundos da fonte de ionização são selecionados e “guiados” até a armadilha de íons (*ion trap*) por um campo elétrico quadrupolar originado de quatro varetas paralelas, na qual cada par de varetas opostas são eletricamente conectadas para estabelecer um campo eletromagnético. De maneira que a variação na radiofrequência (RF) com a voltagem e na frequência angular do potencial de RF do ângulo inicial faz com que os íons possuam uma variação senoidal, porém com uma trajetória estável (sem colisão contra os polos) ao longo do filtro (**Figura 8a**) (Carvalho, 2010).

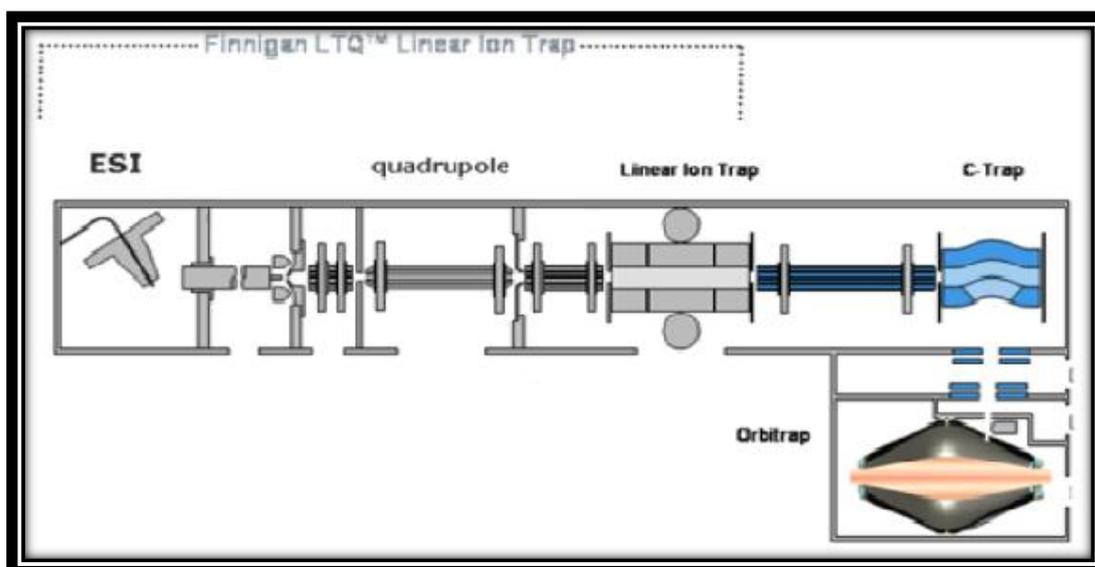
Na armadilha de íons, os íons são separados de acordo com sua razão  $m/z$  ao passo que se aplicam campos eletromagnéticos de frequência constante, porém cujas intensidades crescem. Quando a frequência natural dos íons se iguala a frequência imposta na armadilha, os íons são expelidos sendo obtido o MS1, o qual é o espectro de massas que representa o perfil da razão  $m/z$  dos peptídeos ionizados (**Figura 8b**).



**Figura 8– (a)** O quadrupolo “filtra” e “guia” os íons até a armadilha de íons. O íon representado pelo caminho azul demonstra uma trajetória instável, colidindo em uma das varetas de onde o campo elétrico se origina. O íon representado pelo caminho vermelho possui uma trajetória estável, seguindo para a armadilha de íons. Fonte: (a) Carvalho, 2010. **Figura 8- (b)** A armadilha de íons (*ion trap*) seleciona os íons de acordo com sua razão  $m/z$  à medida que a intensidade dos campos eletromagnéticos cresce fazendo com que os íons saiam do “aprisionamento” e sejam expelidos e detectados, obtendo-se o MS<sup>1</sup>. (b) Modificado de Glish et al., 2003.

Atualmente existem inúmeros modelos de espectrômetro de massas, cada um otimizado para uma particularidade. O Orbitrap Velos (Thermo, San José, Califórnia) (**Figura 9**), que é considerado estado-da-arte, foi o utilizado nesse trabalho. Ele contém uma fonte ionizante eletrospray (ESI), analisadores do tipo “armadilha de íons” (Linear-trap ou LTQ) e Orbitrap. Este último analisador foi inventado por Makarov em 1999 (Makarok, 2000) e possui uma

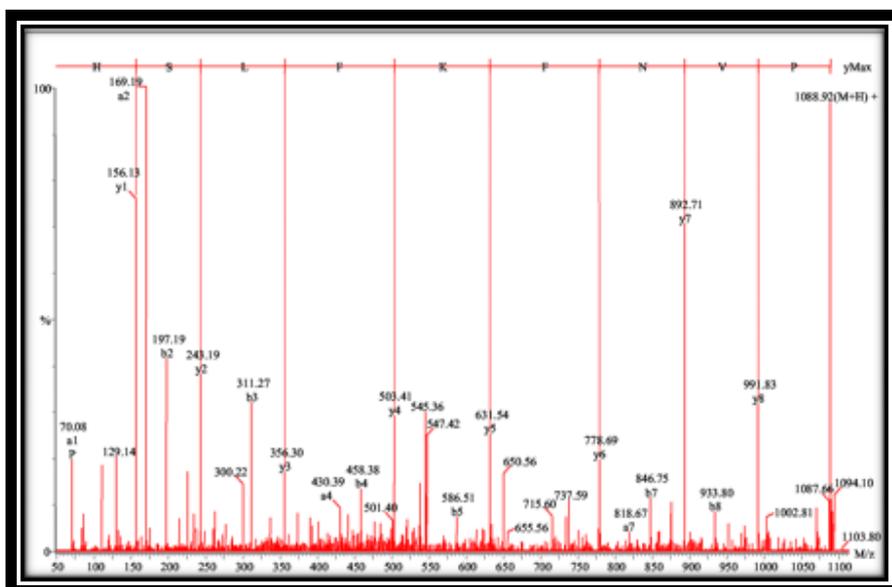
órbita que aprisiona os íons em torno do eletrodo central e oscilam harmonicamente ao longo do eixo com frequências relativas aos valores de  $m/z$  até que atinjam uma frequência característica. Tal aprisionamento deve-se ao campo eletrostático aplicado, no qual a atração eletrostática do eletrodo interno se iguala à força centrípeta. A frequência desta órbita pode ser facilmente demonstrada a ser inversamente proporcional à raiz quadrada da razão  $m/z$  dos íons. A determinação da frequência dos íons presentes é feita aplicando a transformada de Fourier ao sinal que provém do detector. Além disso, o analisador do tipo Orbitrap se caracteriza pela alta resolução (até 200 000) e acurácia de aproximadamente 1 a 5 ppm (Makarok, 2000).



**Figura 9** – Esquema do espectrômetro de massas Orbitrap Velos (Thermo, San José, Califórnia) com a fonte ionizante eletrospray (ESI) e os analisadores do tipo quadrupolo/armadilha de íons (linear trap) e Orbitrap. Fonte: Carvalho, 2010.

Este espectrômetro é capaz de realizar experimentos de espectrometria de massa em tandem ( $MS^n$ ), sendo fundamental para a proteômica *shotgun* na identificação de peptídeos e proteínas (**Figura 10**). Ele pode dissociar as moléculas por três técnicas: dissociação por colisão induzida (CAD),

dissociação por transferência de elétrons (ETD) e dissociação por alta energia (HCD). Porém a mais usada neste tipo de abordagem *shotgun* é a dissociação por colisão induzida, sendo eficaz, nesta configuração com íons com carga +2 e +3. Com relação aos analisadores presentes, ambos podem trabalhar em paralelo, onde a configuração mais utilizada é a obtenção do MS<sup>1</sup> no Orbitrap, e as análises subsequentes de MS<sup>2</sup> no *linear trap* - LTQ, fazendo assim proveito da alta resolução do Orbitrap e velocidade do LTQ (Carvalho, 2010; Han et al., 2008; Motoyama et al., 2008). A sensibilidade e precisão deste aparelho o fizeram o escolhido para as análises deste trabalho.



**Figura 10** - A sequência de aminoácidos do peptídeo PVNFKFLSH pode ser elucidada com o MS2 (espectrometria de massa em tandem) acima pela diferença de massa entre “picos”.

Fonte: Rioli et al., 2003.

### ➤ Bioinformática

Uma vez obtidos os dados, ou seja, os resultados inerentes à massa molecular (MM) dos peptídeos mediante o MS<sup>1</sup>, assim como a informação relativa à sequência de aminoácidos dos peptídeos, contidas nos espectros de fragmentação (MS<sup>2</sup>) e também tendo o conhecimento da enzima que realizou a

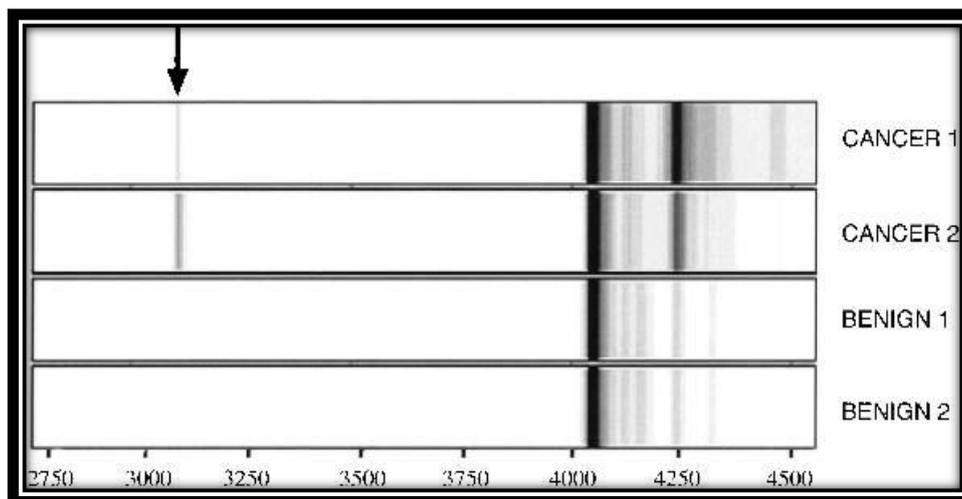
digestão, neste caso, tripsina, podem-se usar *softwares* de busca. Para cada espectro de massas experimental obtido, estas ferramentas são capazes de apontar a sequência peptídica existente em um banco de dados, que indica qual espectro teórico mais se assemelha ao experimental; é fornecido em conjunto um *score* de medida de semelhança. Os softwares mais conhecidos e comumente empregados são: o SEQUEST, ProLuCID e o Mascot. No SEQUEST e ProLuCID a métrica de semelhança é conhecida como correlação cruzada. Os dados devem ser pós-processados por uma ferramenta como *DTASelect* (Cociorva et al., 2006) ou *Search Engine Processor* (Carvalho et al., 2012a). Estes se baseiam em uma distribuição de “*scores*” obtidos da identificação de sequências artificiais, inexistentes no organismo, para poder estabelecer “*scores*” de *cut-off* de forma a convergir para uma resposta confiável. Em seguida, estas ferramentas fornecem a lista de proteínas existentes a partir dos peptídeos identificados.

### **1.9 Marcadores Protéicos Aplicáveis na Detecção de Processos Cancerígenos**

O desenvolvimento de métodos analíticos mais sofisticados proporciona novas metas para a caracterização da doença, diagnóstico precoce e desenvolvimento de medicamentos. Além disso, as análises de mutações de genes têm proporcionado uma visão sobre a patogênese e progressão a partir de lesões pré-invasivas de câncer invasivo. Dessa maneira, a busca por marcadores séricos nesse caso fundamenta-se no conceito de que algumas substâncias específicas sejam liberadas pelas células tumorais ou pelo organismo hospedeiro em resposta à presença do tumor. Logo, um marcador

tumoral pode ser definido como qualquer substância que esteja correlacionada com a apresentação clínica do tumor, seu crescimento e mudança de tamanho após o tratamento (Carvalho et al, 2005). Nesse contexto, o perfil dos genes e da expressão de proteínas tem avançado e contribuído com a compreensão na identificação de genes que são altamente expressas em cânceres, fornecendo mais detalhes sobre os recursos clínicos patológicos e revelando novos aspectos relacionados para o processo de invasão dos tecidos por estes tumores.

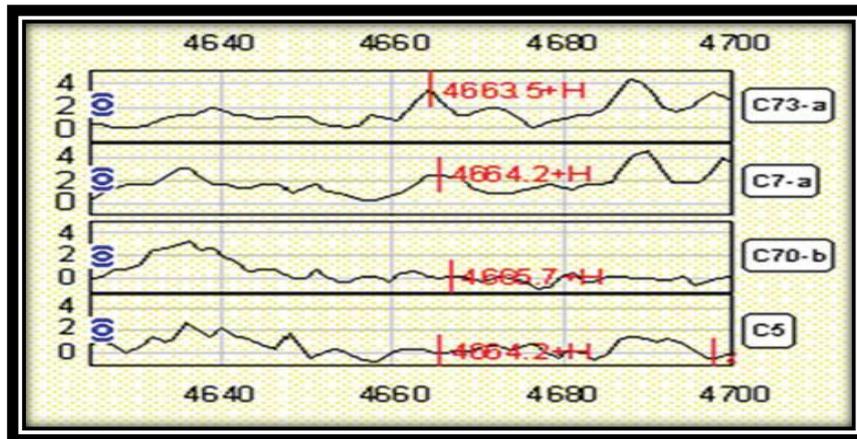
A expansão do conhecimento referente aos perfis de pacientes com câncer está produzindo novas metas, mas também novos marcadores. Como exemplo pode-se citar a descoberta de fatores genéticos cuja presença predispõe o câncer pancreático, como a associação de genes BRCA2 com defeitos da anemia Fanconi e a sensibilidade à mitomicina C. Outros exemplos utilizando a análise proteômica têm apresentando resultados aparentemente satisfatórios, como a análise do perfil protéico realizado no SELDI-TOF que apresentou no espectro de massas diferenças nos padrões de expressão entre os pacientes com e sem câncer (**Figura 11**). Além de que, nesse trabalho foi possível ver que apenas o diagnóstico pela biópsia possui falhas quanto à detecção precoce da mesma, pois pacientes considerados saudáveis pela biópsia e pelo antígeno PSA, foram classificados pela espectrometria de massas como detentores de câncer, o que ficou comprovado cinco anos mais tarde pela biópsia, ou seja, o perfil protéico já estava modificado cinco anos mais cedo, mesmo que este não fosse perceptível ao atual diagnóstico clínico existente (Petricoin et al., 2002a).



**Figura 11-** Cromatograma exemplificando as diferenças nos padrões de expressão entre os pacientes com e sem câncer. Fonte: Petricoin et al., 2002a.

Em outro estudo com SELDI-TOF (Petricoin et al, 2002b), baseando-se apenas nos picos diferencialmente intensificados, foi possível diagnosticar os casos de câncer de ovário em 100% dos casos e de mulheres não afetadas pela doença em 95%. Já no estudo realizado com pacientes de câncer gástrico foi possível distinguir em particular pacientes nos estágios I e II, III e IV (**Figura 12**) desse tipo de câncer (Zhao et al., 2010). Trabalhos semelhantes a esses impulsionaram a aplicação dessa técnica a outros tipos de tumores e tecidos (Li et al., 2002; Yasui et al., 2003).

Portanto, a existência de padrões específicos reforça a hipótese de que biomoléculas podem originar-se de microambientes do tumor hospedeiro (Zhao et al., 2010) permitindo sua análise, e que a poderosa combinação da proteômica com os novos avanços de técnicas instrumentais como a espectrometria de massas pode permitir grandes modificações na clínica médica atual. Tais conhecimentos eventualmente levam a abordagem de um tratamento mais individualizado da doença em estudo. (Jimeno et al., 2006).



**Figura 12** - Expressão diferencial no *Surface-enhanced laser desorption/ionization - Time of flight* (SELDI-TOF) em picos na razão  $m/z$  4665 em amostras de pacientes com estágio I/II versus estágio III/IV de câncer gástrico. C73a = Estágio I, C7-A = Estágio II, C70-b = Estágio III e C5 = Estágio IV. Fonte: (Zhao et al., 2010).

### 1.10 Metabolômica

A metabolômica é outra abordagem das ciências “ômicas” que é usada na descoberta de biomarcadores para certas doenças, desenvolvimento de drogas, em áreas do meio ambiente e de nutrição, entre outros (Chen et al., 2011). Ela é o conjunto de técnicas utilizadas juntamente com abordagens de reconhecimento de padrões e bioinformática, de maneira semelhante à proteômica, que permite uma avaliação quantitativa dos metabólitos endógenos, ou seja, ela possui um importante papel no estudo dos componentes e produtos do metabolismo do organismo em resposta a uma estimulação fisiopatológica ou modificação genética. Considere-se que um número elevado de metabólitos humanos é desconhecido, os quais podem ser de milhares a dezenas de milhares ao comparar com os dados conhecidos da genômica e da proteômica (Spratlin et al., 2009; Kotlowska et al., 2011). Contudo, a abordagem metabolômica pode ainda ser realizada para muitos fins, focando-se nos perfis dos espectros de massas e selecionando apenas

picos de interesse, por exemplo, os diferentes em presença/ausência ou em intensidades relativas. Isto é, na maioria dos casos não é necessário identificar todos os metabólitos presentes numa amostra biológica.

A origem da metabolômica data de várias décadas, com aplicações iniciais no campo de erros inatos do metabolismo, toxicologia e nutrigenômica funcional. Contudo, essa ciência só foi impulsionada no final dos anos 90, a partir da identificação de câncer em amostras de sangue pela ressonância magnética nuclear. Individualmente ou agrupado como perfil metabolômico, a detecção de metabólitos pode ser realizada em células, tecidos e fluidos biológicos tanto por ressonância magnética nuclear (RMN) como por espectrometria de massas (MS), os dois métodos mais aceitos para a mensuração de metabólitos (Spratlin et al., 2009). As técnicas baseadas na espectrometria de massas incluem cromatografia gasosa/espectrometria de massas (GC/MS), GC-MS/MS, cromatografia líquida/espectrometria de massas (LC/MS) e LC-MS/MS (Hu et al., 2011).

Comparada com a genômica ou proteômica, a metabolômica reflete mudanças no fenótipo e, portanto na função. Em contrapartida, ela é a visão terminal do sistema biológico, não permitindo a representação do aumento ou decréscimo das proteínas e dos genes. É necessária a compreensão de todas as ciências “ômicas” para um maior entendimento do sistema biológico humano, uma vez que elas são complementares entre si (Spratlin et al., 2009).

Recentemente, tem havido bons avanços nas tecnologias metabolômicas, possibilitando aos cientistas detectar milhares de compostos numa simples análise, seus biomarcadores potenciais, e finalmente identificar os metabólitos relevantes (Chen et al., 2009). Em especial, para o câncer

gástrico, tem havido indícios de possíveis metabólitos encontrados numa análise de GC/MS, os quais estão descritos na **Tabela 2**, tendo sua expressão elevada ou ausente quando comparado entre os pacientes normais e detentores desse tipo de câncer. É importante frisar que alguns deles estão diretamente envolvidos em vias metabólicas, como por exemplo: da serina e da fosfoserina (Wu et al., 2010).

**Tabela 2** – Metabólitos diferenciáveis encontrados em cromatogramas de cromatografia gasosa/espectrometria de massas (CG/MS) ao comparar pacientes com câncer gástrico e normais. Fonte: Wu et al., 2010.

<b>Metabólitos</b>	<b>Classe química</b>
<b>Ácido heptanodióico</b>	Ácido orgânico
<b>Ácido Propanóico</b>	Ácido orgânico
<b>Fenantrenol</b>	Álcool
<b>Butanotriol</b>	Álcool
<b>L-valina</b>	Aminoácido
<b>Acetamida</b>	Outros
<b>L-isoileucina</b>	Aminoácido
<b>Serina</b>	Aminoácido
<b>Ácido butanoico</b>	Ácido orgânico
<b>Oxazoletiona</b>	Outros
<b>Naftaleno</b>	Outros
<b>L- glutamina</b>	Aminoácido
<b>Fosfoserina</b>	Aminoácido
<b>L-altrose</b>	Carboidratos
<b>L- manofuranose</b>	Carboidratos
<b>Galactofuranosídeo</b>	Carboidratos
<b>Mio-inositol</b>	Poliol
<b>D- ribofuranose</b>	Carboidratos

Resumidamente, a importância de trabalhos como o presente pode ser demonstrada por diversos fatores:

- ✓ Na região norte do Brasil, em especial no Amazonas, o investimento em pesquisas relacionadas aos mais diversos tipos de câncer, em especial com relação ao câncer gástrico, é inferior ao da média mundial; a maioria dessas pesquisas está relacionada a tentativa de compreensão de questões genômicas envolvidas nesse processo;
- ✓ A necessidade da descoberta de novos biomarcadores para a detecção precoce de várias doenças, em particular o câncer. Atualmente, são poucos os marcadores, nem sempre específicos para cada tipo de câncer. O conhecimento desses candidatos a marcadores específicos pode resultar no futuro num diagnóstico mais personalizado e menos tardio aumentando assim a chance de sobrevivência da pessoa;
- ✓ Na Região Amazônica, em especial em Manaus, há uma grande incidência de casos de cânceres gástricos que conforme relatos de especialistas locais da área apresentam certas particularidades frente aos cânceres gástricos existentes em outros lugares. Por exemplo, a maior taxa de incidência regional deste tipo de câncer ocorre em pessoas de 35 a 60 anos. Contudo, a ocorrência entre jovens de 15 a 30 anos é superior à média mundial para essa faixa etária. Ou seja, há uma intrigante ocorrência dessa doença na população jovem na Região, necessitando maiores investigações para seu melhor entendimento;
- ✓ Há vários resultados na literatura de estudos de proteomas e metabolomas por espectrometria de massas, com resultados bastante relevantes, indicando

que metodologias semelhantes podem ser promissoras para o diagnóstico clínico no futuro;

✓ A pesquisa e o ensino multidisciplinar, particularmente com enfoque de diagnóstico clínico com o uso de novas técnicas de análise precisam ser alavancados na Região, pois há carências de especialistas para atuar nesse tipo de abordagem.

## 2. OBJETIVOS

### **2.1 Objetivo geral**

- Realizar um estudo comparativo de tecidos de pacientes com câncer gástrico e de indivíduos controles (sem a doença) através de análises proteômicas e metabolômicas utilizando espectrometria de massas.

### **2.2 Objetivos específicos**

- Analisar preliminarmente as amostras de câncer gástrico, da margem de ressecção e do controle mediante a técnica de PCR (*Polymerase Chain Reaction*);
- Realizar análises por espectrometria de massas de perfis proteômicos e metabolômicos de câncer gástrico do tipo adenocarcinoma (95% dos casos), para melhor caracterizar o perfil molecular dos mesmos;
- Selecionar candidatos a marcadores proteômicos e metabolômicos específicos para o tipo de câncer em estudo;
- Avaliar estatisticamente a consistência dos resultados dos perfis proteômicos e metabolômicos com os dados da avaliação clínica dos pacientes envolvidos no estudo.

### 3. MATERIAIS E MÉTODOS

O projeto foi submetido e aprovado pelo Comitê de Ética em Pesquisa (CEP) da Universidade Federal do Amazonas sob Nº 0057.0.115.000-11-CAAE. Sendo este juntamente com a coleta do material biológico também aprovado pelo Comitê Científico da Fundação CECON.

#### ***3.1 Seleção de pacientes com câncer e indivíduos controle***

Para o trabalho prospectivo comparativo de pacientes com cânceres gástricos avaliados por espectrometria de massas, foram selecionados pacientes com câncer, a respectiva margem de ressecção e controles de acordo com os seguintes critérios:

❖ Critérios de inclusão

- Pacientes com cânceres gástricos diagnosticados através de biópsia e endoscopia, tendo como diagnóstico histológico adenocarcinoma; E que tivessem indicação cirúrgica pelo corpo clínico da fundação CECON.
- Pacientes maiores de idade que concordaram em participar da pesquisa e independente do sexo;
- Os pacientes controles (sem a doença) foram aqueles portadores de dispepsia gástrica comprovados com endoscopia para excluir qualquer possibilidade de câncer.

❖ Critérios de exclusão

- Os voluntários não deveriam apresentar histórico ou qualquer evidência de HIV ou vírus da hepatite B ou C;

- A ocorrência prévia ou atual das doenças citadas acima e o histórico de outros tipos de carcinoma foi anotada e no caso de possível interferência nas análises, a amostra correspondente foi descartada;

### **3.2 Coleta e preservação das amostras**

Em todos os casos o consentimento escrito dos voluntários (Termo de Consentimento Livre e Esclarecido – Anexo 1) foi obtido mantendo a devida privacidade. Além disso, foi respondido um questionário pelos pacientes permitindo ao fim da pesquisa correlacionar os dados laboratoriais com a situação particular de cada voluntário. Vale ressaltar que foram seguidas todas as normas do conselho de ética institucional.

As amostras foram coletadas na Fundação CECON sob a orientação de cirurgião do aparelho digestivo, especialista em Oncologia. Os indivíduos do grupo controle foram escolhidos durante o processo de endoscopia segundo a classificação de Bormann. Sendo obtidos vinte espécimes não tumorais coletados durante o exame endoscópico referente ao antro e corpo gástrico. Já os indivíduos do grupo com câncer gástrico e a respectiva margem de ressecção foram escolhidos mediante a compatibilidade das informações do prontuário de cada paciente com os critérios de inclusão desse trabalho e o estágio cirúrgico foi determinado de acordo com o Tumor, Nódulo e classificação da Metástase (TNM) da *American Joint Committee on Cancer* (AJCC). Sendo obtidos durante o ato operatório vinte espécimes tumorais coletados de peça cirúrgica e vinte espécimes respectivos à margem de ressecção dessa mesma peça pertencentes ao antro gástrico. Resumidamente, as margens de ressecção foram definidas macroscopicamente após a retirada do tumor, onde foram delimitados

aproximadamente 10 cm a partir da localização tumoral, totalizando ao final sessenta amostras.

O espécime tumoral, a margem de ressecção e o tecido sadio foram armazenados em microtubos tipo eppendorf de 2,0 mL com tampa e estéreis, os quais foram transportados em gelo comum. Todo material coletado foi armazenado a -80 °C.

### **3.3 Caracterização dos tumores**

Para a caracterização e a classificação do câncer gástrico foi realizada a endoscopia digestiva e o material obtido (biópsias) foi analisado em HE (hematoxilina e eosina) para o diagnóstico de adenocarcinoma.

A seguir, se encontra a descrição de como foi feita a análise dessas amostras coletadas, sendo que este trabalho foi dividido em três partes principais:

#### **PARTE A – Estudo de biologia molecular das amostras de câncer gástrico**

#### **3.4 Estudo da presença da infecção pelo vírus Epstein-Barr (EBV) em tecidos de cânceres gástricos através da biologia molecular**

Neste estudo o objetivo foi analisar a ocorrência de infecção viral nos pacientes, onde o material (DNA) foi coletado a partir da pulverização do tecido com nitrogênio líquido ou na forma de escovado diretamente dos tecidos de pacientes controles (n=6) sendo cinco do sexo feminino e um do sexo masculino, e de pacientes com câncer (sendo dez amostras da região tumoral e dez amostras das respectivas margens de ressecção, totalizando 20

amostras de 10 pacientes), composto de quatro pacientes do sexo feminino e seis do masculino.

#### **3.4.1 A Extração de DNA**

Os tecidos foram imersos em nitrogênio líquido (-196 °C) e posteriormente macerados para fragmentação do tecido ou foram diretamente escovados. As escovas contendo o material biológico foram mergulhadas num tubo cônico de centrifugação de 1,5 mL contendo 500 µL de tampão de lise celular (Tris-HCl 10 mM, pH 7,5; EDTA 10mM; NaCl 10 mM; SDS 2%) e 15 µL de proteínase K a 10 mg/mL. O material obtido das escovas ou fragmentos dos tecidos foram então incubados por duas horas a 56 °C no tampão de lise celular. Após a incubação acrescentou-se 500 µL de fenol/clorofórmio/álcool isoamílico (25:24:1) a cada tubo, e agitou-se por cerca de um minuto. Depois da agitação os tubos foram centrifugados a 4480 x *g* (RCF), durante 5 minutos. Em seguida, a fase aquosa (fase superior) foi transferida para outro tubo onde foram acrescentados 15 µL de NaCl 5M e 1.000 µL de etanol a 95%. O material foi novamente incubado a -20 °C por 12 horas e posteriormente centrifugado durante 20 minutos a 8.000 r.p.m. Após esta centrifugação, o sobrenadante de cada tubo foi aspirado e os tubos foram mantidos abertos para secagem à temperatura ambiente. Por fim, o *pellet* de DNA foi ressuspensão em 50 µL de água milli-Q e estocado a -20 °C para uso posterior.

#### **3.4.2 A Análise da presença viral por Reação em cadeia pela polimerase (PCR – “*Polimerase Chain Reaction*”)**

O DNA extraído foi submetido à reação em cadeia pela polimerase (PCR) para confirmação da sua funcionalidade, usando iniciadores para o éxon

5 do gene codificante da proteína p53, que amplificam um fragmento de 274 bp (Pestaner et al., 1994). A análise da infecção foi realizada para o vírus Epstein-Barr (EBV), no qual para a detecção do EBV foram utilizados os iniciadores TC67 e TC69 (Saito et al., 1989). Visto que estes iniciadores amplificam um fragmento de 288 bp correspondente a região Bam M que é repetida dez vezes no genoma do EBV.

As reações de PCR foram realizadas utilizando 1 µL de DNA e 12 µL da mistura de reação contendo: 5 µL de PCR 10 x solução tampão, 1,5 µL de 50 mM MgCl<sub>2</sub> (3 µL para análise de EBV), 1 µL de 10 mM dNTPs, 0,5 µL de Taq 5 U/µl, 1 µL de 10 µM de iniciador direto, 1 µL de 10 µM de iniciador reverso e H<sub>2</sub>O para volume final de 50 µL. As condições de amplificação foram: 5 minutos de desnaturação a 95 °C, seguido de 35 ciclos de 95 °C durante um minuto, 60 °C durante um minuto e 72 °C durante um minuto. A extensão final foi realizada por 10 minutos a 72 °C e o produto amplificado foi mantido a 4 °C. Em todas as reações foram utilizados controles positivo e negativo. Como controle positivo da reação foi utilizado à linhagem celular Raji. Já no controle negativo o DNA foi substituído por água destilada (Lattario et al., 2008). As sequências dos iniciadores utilizados (oligonucleotídeos) neste trabalho foram:

- Região Bam M do EBV.

TC67 5` CAG GCT TCC CTG CAA TTT TAC AAG CGG 3`.

TC69 5` CCC AGA AGT ATA CGT GGT GAC GTA GA 3`.

- Éxon 5 do gene codificante da proteína p53.

Direto 5' GCA ACC AGC CCT GTC GTG TCT CCA 3'.

Reverso 5' GGA ATT CTG TTC ACT TGT GCC CTG ACT TTC AAC 3'.

### **3.4.3 A visualização dos produtos amplificados por eletroforese em gel de poliacrilamida.**

Os produtos da PCR foram analisados por eletroforese em gel de poliacrilamida contendo: 4,44 mL de H<sub>2</sub>O; 2,0 mL de TBE 5x; 3,32 mL de (30:0,8) Acrilamida-Bisacrilamida; 240 µL de 10% AMP e 3 µL de TEMED. Sendo que a 10 µL de produto amplificado de DNA foram adicionados 3 µL do tampão de amostra e essa mistura foi aplicada em um gel a 10%. A eletroforese foi corrida a 80 V.

Após a corrida, o gel foi corado por nitrato de prata. Inicialmente o gel foi imerso na solução fixadora (5 mL de etanol, 0,4 mL de ácido acético e 45 mL de água destilada) por 10 minutos. Em seguida, foi feita a imersão na solução de prata (0,1 g de AgNO<sub>3</sub> em 50 mL de água destilada) por 10 minutos (coloração pela prata) e por fim, após lavagem com 50 mL de água destilada, o gel foi revelado na seguinte solução: 1,5 g de NaOH, 0,4 mL de formaldeído 37% em 50 mL de água destilada. A revelação foi interrompida pela solução fixadora (Rosenbauer e Riesner, 1987).

O tamanho aproximado dos fragmentos amplificados foi estimado utilizando-se o marcador de peso molecular *LMW (Low Molecular Weight – BioRad)*.

## **PARTE B – Estudo proteômico das amostras de câncer gástrico**

### ***3.5 Estudo exploratório comparativo de candidatos a biomarcadores protéicos dos cânceres gástricos mediante a técnica de MudPIT off-line***

Nesta parte o objetivo foi analisar qualitativa e quantitativamente o maior número de proteínas possível através de tratamento estatístico e por comparação dos dados dos pacientes com câncer gástrico, da margem de ressecção e dos indivíduos controle, identificar aquelas relacionadas a patologia.

#### **3.5.1 B Pulverização do tecido**

Inicialmente, as amostras (pequenos pedaços de tecidos - **Figura 13**) foram colocadas em envelopes de papel de alumínio, os quais foram pulverizadas em nitrogênio líquido. A retirada desses fragmentos pulverizados ocorreu mediante a transferência para um tubo eppendorf de 2,0 mL e adição de um volume específico de RapiGest® (Waters) a uma concentração de 0,2% (w/v), o qual foi determinado pela observação da quantidade de tecido pulverizado em cada tubo, e em seguida colocados no gelo. Para deixar o RapiGest® nessa concentração foi adicionado 500 µL de bicarbonato de amônio (NH<sub>4</sub>HCO<sub>3</sub>) a 50 mM. Vale ressaltar que o RapiGest® é um detergente compatível com espectrometria de massa ajudando assim a solubilização das proteínas e se desfazendo ao acidificar a solução no momento de parar a digestão enzimática.



**Figura 13** - Amostras a serem pulverizadas (margem de ressecção e câncer).

Feito isso, as amostras foram submetidas a um vórtex por vinte segundos, depois centrifugadas a 5 °C por 31 min a 7280 x *g* (RCF). Posteriormente, o sobrenadante foi retirado, e o pellet resultante de cada amostra foi guardado para o método de amplificação, PCR – *Polymerase Chain Reaction*. Todas as amostras foram guardadas sob-refrigeração a -20 °C.

### **3.5.2 B Dosagem de Proteínas**

A dosagem de proteínas foi realizada pelo método do ácido bicinchonínico (BCA - Sigma), no qual se utiliza dois reagentes, o reagente A (ácido bicinchonínico, carbonato de sódio, tartarato de sódio e bicarbonato de sódio em 0,1 N de NaOH, pH=11,25) e reagente B (sulfato de cobre-pentaidratado 4% (w/v) (Smith et al., 1985), sendo o padrão albumina sérica bovina (BSA) a 1,0 mg/mL (0,15 M de NaCl, 0,05% de azida sódica). Resumidamente, cada amostra foi diluída cinco vezes em tampão fosfato salino (PBS), de onde retirou-se 100 µL e foi adicionado 2 mL da mistura composta

por uma parte do reagente B e cinquenta partes do reagente A, deixando incubar numa estufa a 37 °C por 30 minutos. O mesmo foi realizado para os tubos contendo o padrão de BSA nas concentrações de 10, 20, 30, 40 e 50 mg/mL. O experimento foi realizado em duplicata, e a leitura foi feita num espectrofotômetro para placa de Elisa (VERSAmax) sendo a absorbâncias das amostras medidas a 562 nm.

### **3.5.3 B Eletroforese unidimensional**

Para a eletroforese unidimensional, utilizou-se 15 µg de proteína. Retirou-se o volume correspondente a 15 µg do sobrenadante e adicionou-se 12 µL do tampão Laemmli (duas vezes concentrado com SDS). As amostras passaram por um vórtex e foram centrifugadas por alguns minutos. Sendo aquecidas num banho maria a 100 °C por 5 minutos. O gel de poliacrilamida preparado foi a 12% (Gel de corrida: 3,6 mL de água milli-Q; 2,5 mL de Tris-HCl 1,5 M pH 8,8; 4,0 mL de acrilamida; 50,0 µL de persulfato de amônio; 5,0 µL de TEMED. Gel concentrador a 4%: 6,2 mL de água Milli-Q; 2,5 mL de Tris-HCl 0,5 M pH 6,8; 1,3 mL de acrilamida; 50,0 µL de persulfato de amônio; 10,0 µL de TEMED), conforme instruções do fabricante (Mini Protean II). As condições elétricas utilizadas para a corrida foram: 120 V; 400 mP durante 90 minutos.

### **3.5.4 B Coloração com Coomassie R-250**

Os géis foram retirados das placas e imersos em 200 mL de solução fixadora (40% etanol; 10% ácido acético; 50% de água Milli-Q) por cerca de sessenta minutos. Passado esse tempo, os géis foram corados com

Coomassie R-250 0,2% (40% etanol; 10% ácido acético; 50% de água Milli-Q) durante a noite. No outro dia, descartou-se o corante e adicionou-se novamente a solução descorante (citada acima) por sessenta minutos no mínimo. Feito esse processo, os géis foram escaneados.

### **3.5.5 B Digestão trípica em gel de poliacrilamida**

Após visualizar algumas proteínas diferencialmente expressadas, as bandas foram cortadas em tamanho aproximado de 1 mm<sup>2</sup>, com auxílio de um bisturi, e maceradas dentro de um microtubo.

As bandas recortadas do gel foram transferidas separadamente para microtubos novos. Uma banda conhecida foi utilizada como controle positivo (uma banda do padrão de peso molecular). Para descorar as bandas de coomassie foram feitas três lavagens de quinze minutos com 200 µL de solução contendo 50% de acetonitrila e 25 mM de bicarbonato de amônio (pH 8,8). Posteriormente, o pedaço de gel foi desidratado por cinco minutos com 200 µL de acetonitrila (100%) e seco por centrifugação a vácuo (*SpeedVac - Savant*).

Após esse processo, as proteínas contidas nos pedaços de géis foram reduzidas com 100 µL de DTT (10 mg de DTT/mL de bicarbonato de amônio 100 mM pH 8,0) a 65 mM, por trinta minutos, à 56 °C. Removeu-se essa solução de DTT e alquilou-se com 100 µL de iodoacetamida (37 mg de iodoacetamida/mL de bicarbonato de amônio 100 mM pH 8,0) a 200 mM, por trinta minutos, à temperatura ambiente, no escuro. Removeu-se a solução de iodoacetamida, lavando os pedaços de gel com 200 µL de bicarbonato de amônio 100 mM, por dez minutos. Depois de retirar essa solução, eles foram

desidratados com 200  $\mu\text{L}$  de acetonitrila (100%) por cinco minutos. Para hidratar o gel, a acetonitrila foi removida e foi adicionado 200  $\mu\text{L}$  de bicarbonato de amônio 100 mM, por dez minutos. Removeu-se o bicarbonato e os pedaços de gel foram novamente desidratados com 200  $\mu\text{L}$  de acetonitrila, sendo essa removida posteriormente. Esse último processo foi repetido mais uma vez. Então, os géis foram secos numa centrífuga a vácuo (*SpeedVac - Savant*).

Foram adicionados 30  $\mu\text{L}$  da solução de reidratação e digestão contendo 40 mM de bicarbonato de amônio e tripsina 20 ng/ $\mu\text{L}$  (Promega), deixando durante a noite a 37 °C.

No dia seguinte, os peptídeos foram extraídos do gel pela adição de 30 $\mu\text{L}$  da solução contendo 5% de ácido fórmico e 50% de acetonitrila por alguns segundos no vórtex, quinze minutos em repouso e dois minutos no ultrassom, sendo transferidos para outro tubo previamente marcado. Esse procedimento foi repetido duas vezes e, por fim, as amostras foram concentradas até o volume aproximado de 5  $\mu\text{L}$  por centrifugação à vácuo (*SpeedVac - Savant*). As amostras foram armazenadas a -20 °C para posterior análise por espectrometria de massa.

### **3.5.6 B Digestão trípica de proteínas em solução**

Para a digestão de proteínas em solução, utilizou-se 100  $\mu\text{g}$  de proteínas. Após retirar o volume correspondente a 100  $\mu\text{g}$  do sobrenadante e colocá-los num outro microtubo, foi acrescentado bicarbonato de amônio ( $\text{NH}_4\text{HCO}_3$ ) 50 mM para resultar em uma concentração de 0.1% (w/v) devido ser essa a concentração indicada pelo fabricante Rapigest® para esse tipo de digestão.

Sabendo que para cada 20  $\mu\text{L}$  do sobrenadante, acrescenta-se 5  $\mu\text{L}$  de ditionotreitól (DTT) a 100 mM, calculou-se a partir dos volumes existentes de cada amostra o volume de DTT necessário para uma concentração final de 20 mM. Posteriormente a essa adição, incubou-se as amostras num banho-maria a 60 °C por trinta minutos. E foram deixadas para resfriar a temperatura ambiente. Na sequência, foi adicionado iodoacetamida (IAA), uma vez que a cada 20  $\mu\text{L}$  de sobrenadante, tem-se a adição de 5  $\mu\text{L}$  de IAA a 400 mM, calculou-se novamente os volumes necessários de IAA para uma concentração final de 66,7 mM em cada amostra. Incubou-se as amostras por quinze minutos à temperatura ambiente no escuro.

Terminado esse tempo, adicionou-se tripsina (Promega – 20  $\mu\text{g}$ ) na proporção de 1/50 (m/m) de enzima/substrato. Incubou-se a 37 °C por 20 horas. Parou-se a reação com o volume correspondente em cada amostra de ácido fórmico a 10% (v/v) em água para uma concentração final de 1% (v/v).

### **3.5.7 B Preparação de mini-colunas (Poros R2)**

A dessalinização e retirada de outras impurezas das amostras ocorreu mediante ao uso de mini-colunas (Zip-Tip modificado) com a resina Poros® R2 (*Applied Biosystems*). A capacidade da resina é 5  $\mu\text{g}$  proteína/1 $\mu\text{L}$  de resina. Portanto, como a quantidade necessária de amostra para a troca iônica forte é 50  $\mu\text{g}$ , seria necessária 10  $\mu\text{L}$  de resina. Porém, para garantir a não saturação da coluna, ela foi preparada duas vezes a quantidade de resina para 50  $\mu\text{g}$ , ou seja, foi preparado colunas de 20  $\mu\text{L}$  (1,4 cm) usando tip de 200  $\mu\text{L}$  para as amostras antes da troca iônica e colunas de 8  $\mu\text{L}$  (1,0 cm) usando tip de 20  $\mu\text{L}$  para as frações obtidas depois da troca iônica.

A eluição e o equilíbrio das colunas 20  $\mu$ L foram da seguinte forma:

- Lavou-se 2 x 40  $\mu$ L com acetonitrila 100%;
- Equilibrou-se 4 x 40  $\mu$ L com TFA 1%;
- Adicionou-se a amostra em TFA 1%, preparado na hora;
- Lavou-se 6 x 40  $\mu$ L TFA 0,1%;
- Elui-se 2 x 40  $\mu$ L com TFA 0,1% + Acetonitrila 70%.

A eluição e o equilíbrio das colunas 8  $\mu$ L foram da seguinte forma:

- Lavou-se 2 x 20  $\mu$ L com acetonitrila 100%;
- Equilibrou-se 4 x 20  $\mu$ L com TFA 1%;
- Adicionou-se a amostra em TFA 1%, preparado na hora;
- Lavou-se 6 x 20  $\mu$ L TFA 0,1%;
- Elui-se 2 x 20  $\mu$ L com TFA 0,1% + Acetonitrila 70%.

Essas amostras dessalinizadas foram secas no *Speed Vac* (Savant). Ressuspensas em 55  $\mu$ L de TFA 0,1% + 30% de acetonitrila para aquelas a serem submetidas à troca iônica e em 40  $\mu$ L de TFA 0,5% para aquelas frações provenientes da troca iônica.

### **3.5.8 B Cromatografia de troca iônica forte (SCX)**

Cinquenta microgramas da mistura de peptídeos dessalinizadas e ressuspensas em 55  $\mu$ L de TFA 0,1% + 30% de acetonitrila foram introduzidas em uma coluna de troca catiônica forte *PolySulfoethyl A* (50 x 1 mm; PolyLC Inc., Columbia, MD, EUA) usando um sistema de HPLC (*high performance liquid chromatograph*) do *Ettan GE Healthcare*. Neste foi aplicado um gradiente linear salino de 0 a 700 mM de NaCl, durante vinte e cinco minutos, seguido de

dois minutos de 700 a 800 mM de NaCl, sendo monitorado as absorvâncias a 215, 280 e 320 nm. Conforme os perfis descritos em cada cromatograma, foi feito um *pool* de modo a se obter apenas 6 frações salinas (6 passos salinos) de cada amostra, sendo que cada uma foi dessalinizada com resina Poros R2 (*Applied Biosystems*), conforme descrito acima no item 3.5.7 B. Posteriormente, foram eluídas em tubo próprio, o qual foi carregado, pelo sistema *auto-sampler*, em coluna de fase reversa acoplada diretamente ao espectrômetro de massa Orbitrap-Velos (Thermo, San Jose).

### **3.5.9 B Aquisição de dados no Orbitrap Velos (Thermo, San Jose)**

O espectrômetro de massa Orbitrap-Velos (Thermo, San Jose) foi programado para adquirir os dados no modo *data-dependent* usando uma lista de exclusão dinâmica de sessenta segundos, visando aumentar o número de peptídeos identificados, pois cada vez que um peptídeo mais intenso é submetido à análise por MS<sup>2</sup>, a sua razão *m/z* é temporariamente incluída nessa lista para que outro peptídeo de sinal menos abundante seja submetido também à análise por MS<sup>2</sup> evitando assim a análise do mesmo peptídeo. Foram obtidos os espectros de MS<sup>1</sup> seguido por nove espectros de MS<sup>2</sup> usando dissociação por colisão induzida com uma energia normalizada de colisão de trinta e cinco. A velocidade de fluxo na ponta da coluna de fase reversa foi de cerca de 100 nL/min sendo a fase móvel composta de 95% de H<sub>2</sub>O, 5% de acetonitrila e 0,1% de ácido fórmico. A aquisição dos *scans* de MS e os gradientes dos solventes foram controlados pelo programa *Xcalibur* (Thermo, San Jose, Califórnia).

### 3.5.10 B Análise de dados pela proteômica *Shotgun*

#### ✓ Identificação das proteínas

Os espectros de MS<sup>1</sup> e MS<sup>2</sup> foram convertidos para o formato *raw* usando o programa *RawExtractor* (McDonald et al., 2004). O banco de dados foi preparado de acordo com as diretrizes da semi-marcação “falsa” (*semi-labeled decoy*) (Barboza et al., 2011), a partir de sequências de *Homo sapiens* obtidas do UniProt em 1 de janeiro de 2012, incluindo também as sequências de 127 contaminantes mais comuns na espectrometria de massas (Exemplos: queratina, tripsina). Vale ressaltar que essa abordagem de semi-marcação “falsa” permite uma pós-avaliação da qualidade dos resultados (Barboza et al., 2011; Yates et al., 2012). Visto que para calcular o nível de confiança e a taxa de falso-positivo, um banco de dados no formato *semi-labeled decoy* foi usado para que cada sequência utilizada gerasse duas sequências “falsas” adicionais, a primeira marcada como *MiddleReversed (labeled decoy)* e a outra marcada como *PairReversed (unlabeled decoy)*.

As identificações foram indicadas pelo ProLuCID *search engine*, o qual foi usado para comparar o MS<sup>2</sup> obtido experimentalmente com o espectro teórico gerado a partir do banco de dados e selecionar os mais prováveis candidatos (Xu et al., 2006). Neste, a busca foi limitada a candidatos trípticos e semi-trípticos, a carbamidometilação foi imposta como uma modificação fixa, e a oxidação da metionina como a modificação variável. Foram aceitas sequências com tolerância de 70 ppm a partir da medida *m/z* do precursor, sendo que o *XCorr* e o *Z-Score* foram utilizados como métrica de semelhança primária e secundária. A validação foi realizada pelo *Search Engine Processor*

(Carvalho et al., 2012a). Resumidamente, as identificações foram agrupadas pelo estado de carga (+2 e  $\geq$  +3) e, em seguida, pelo *status* tríptico (totalmente tríptica, semi-tríptica), resultando em quatro subgrupos distintos. Para cada resultado, os valores do ProLuCID XCorr, DeltaCN e ZScore foram usados para gerar um discriminador Bayesiano. As identificações foram classificadas em uma ordem não decrescente de acordo com a pontuação do discriminador.

As pontuações mínimas (*cut-off*) foram estabelecidas para que as identificações tivessem 1% de *false discovery rate* (FDR) baseado no número de labeled decoys. Este procedimento foi realizado de forma independente em cada subconjunto de dados, resultando em uma taxa de falso-positivo que era independente do estado tríptico ou do estado de carga. Adicionalmente, uma restrição a sequências com seis ou mais resíduos de aminoácidos foi imposta. Os resultados foram pós-processados para aceitar apenas proteínas que tivessem duas ou mais evidências independentes na identificação (por exemplo, a identificação de um peptídeo com diferentes estados de carga, uma versão modificada e não modificada do mesmo peptídeo, ou dois peptídeos diferentes). Este último filtro foi feito para que todos os resultados de busca obtivessem um FDR de 0% no nível *labeled and unlabeled decoy*.

#### ✓ **Quantificação de proteínas**

Os arquivos de MS<sup>1</sup> foram deconvoluídos usando YADA (Carvalho et al., 2009). O módulo de quantificação do SEPro (SEProQ) foi então utilizado para a obtenção dos cromatogramas de íons extraídos (XIC) a partir dos arquivos de MS<sup>1</sup> deconvoluídos e correlacionados com as sequências de peptídeos correspondentes. Os XICS foram normalizados de acordo com a abordagem de distribuição normalizada do fator de abundância do íon (dNIAF).

O dNIAF emprega um processo semelhante a distribuição normalizada de fatores de abundância espectrais (dNSAF) (Zhang et al., 2010) porém, em vez de depender da quantificação por contagem espectral, ele usa os XICS extraídos dos MS1 deconvoluídos. Lembrando que a normalização do dNSAF baseia-se na quantificação de peptídeos exclusivos para distribuir a quantificação do sinal de peptídeos que são compartilhados entre proteínas.

O módulo do programa *PatternLab* chamado Diagrama de Venn de área proporcional foi usado para identificar as proteínas exclusivamente identificadas em cada tipo de tecido (Carvalho et al., 2011). Esta análise considerou apenas proteínas encontradas em duas ou mais replicatas biológicas a partir das condições do experimento (ou seja, controle, margem de ressecção ou câncer). Para as proteínas encontradas em duas ou mais condições e em duas ou mais replicatas biológicas dessas condições, foi usado o módulo TFold do programa *PatternLab* utilizando um  $q$ -valor de 0.05. Vale frisar que esse módulo introduz uma abordagem de seleção de proteínas expressadas considerando uma grandeza de variação de expressão (*fold-change*) e um  $q$ -valor para garantir que a expressão diferencial obtida não foi ao acaso. O algoritmo aplica um *fold-change* variável para maximizar o número de proteínas diferencialmente expressadas satisfazendo um FDR (Carvalho et al. 2012b; Fischer, 2010).

Visando auxiliar na interpretação de semelhanças no conjunto de dados obtidos foi empregada a análise de escalonamento multidimensional (*Multidimensional scaling analysis*). Esse escalonamento é usado para representar dados de um espaço em alta dimensão em um menor, tipicamente de duas e três dimensões, para permitir o acesso visualmente de padrões.

Dessa maneira, foi implementado um algoritmo, chamado Búzios (Aquino et al., 2012).

Búzios recebe como entrada uma matriz esparsa gerada a partir do SEPro. Neste, cada linha (vetor) inclui os resultados da análise MudPIT, onde o Búzios mapeia cada vetor de um espaço N-dimensional, no qual N corresponde ao número de proteínas identificadas em todas as análises num espaço bidimensional. O mapeamento é de maneira que cada dimensão corresponde à quantificação obtida de uma dada proteína. Logo, o resultado final é uma representação de cada vetor como um ponto num espaço bidimensional. Isto é feito pela tentativa de respeitar as semelhanças em um espaço dimensional elevado medindo com um ponto do produto normalizado. Como geralmente a estas “distâncias de similaridade” em espaço dimensional baixo não é possível, este pode ser formalizado por uma otimização do seguinte problema:

$$\sum_{i < j} (\|x_i - x_j\| - \delta_{i,j})^2 w_i w_j,$$

Onde:  $x$  é um vetor num espaço bidimensional,  $\delta$  é uma medida de semelhança e  $w$  é um peso para penalizar os discrepantes. Os pesos são atribuídos da seguinte forma: primeiro, para cada classe, o centróide é calculado num espaço dimensional elevado. Em segundo lugar, os vetores são ordenados em ordem decrescente de acordo com a sua distância euclidiana do centróide.

## **PARTE C: Estudo metabolômico das amostras de câncer gástrico.**

### ***3.6 Estudo exploratório comparativo de perfis metabolômicos de cânceres gástricos.***

Neste estudo o objetivo foi verificar a possibilidade de diferenciar biópsias tumorais, de suas margens de ressecção e de indivíduos controle através da análise de seus perfis metabolômicos por espectrometria de massas seguida do tratamento estatístico dos dados.

#### **3.6.1 C Preparação da amostra**

As amostras de cada tipo de tecido (n=16) foram pulverizadas sob nitrogênio líquido de maneira semelhante ao item 3.5.1 B. Sendo que neste caso, após a retirada dos fragmentos pulverizados foram adicionados 1000 µL de acetato de etila/metanol (1:1) para extrair os metabólitos de média e alta polaridade. Em seguida, essas amostras foram submetidas a um vórtex por alguns segundos, a uma centrifuga por 2 min a 7504 x g (RCF), a um ultrassom por dez minutos e repouso por 24 horas. Passado esse tempo, o líquido sobrenadante foi transferido para outro tubo *ependorf* de 2,0 mL, sendo evaporado em um dissecador. Após a obtenção da amostra seca, foi realizada uma normalização para que as intensidades das respectivas bandas na corrente de íons totais fossem semelhantes, como segue. Para as amostras dos indivíduos “controle” foram adicionados 2000 µL de metanol grau HPLC. Nas amostras de câncer e da respectiva margem de ressecção foi realizada uma diluição, a qual se constituiu inicialmente na adição de 1000 µL de metanol grau HPLC a amostra seca, sendo essas as soluções estoque. Destas

soluções foram retirados para outros tubos 200 µL das amostras de câncer e 300 µL das de margem de ressecção. Cada solução foi completada para 2000 µL com metanol grau HPLC.

### **3.6.2 C Aquisição de dados no LCQ *Fleet* (Thermo-Scientific)**

As amostras foram injetadas em triplicatas mediante inserção direta em um espectrômetro de massas *ion trap*, modelo LCQ *Fleet* (Thermo Scientific), utilizando como fonte de ionização o *electrospray*, operando no modo positivo. A concentração das amostras foi normalizada como descrito no item 3.6.1 C e analisada na faixa de *m/z* de 600-1000 Da. As condições utilizadas do aparelho foram: *sheath gas flow rate* - 9 psi; *auxiliary gas flow rate* – 3 psi; *sweep gas flow rate* – 0 psi; *spray voltage* – 5 KV; *capillary temp* – 260 °C; *capillary voltage* – 26 V; *tube lens* - 120. Foi utilizado o *loop* de 5 µL e o fluxo de infusão de 25 µL/min de metanol grau HPLC. Os espectros obtidos foram processados através do programa Xcalibur.

### **3.6.3 C Agrupamento não supervisionado dos perfis de metabólitos**

Uma análise estatística com intuito exploratório foi realizada, objetivando verificar se as observações resultantes para os perfis por espectrometria de massas podem vir a ser empregadas para o desenvolvimento de métodos estatísticos para a classificação de indivíduos com câncer, margem de ressecção e o controle. Considerando o estágio inicial da pesquisa, a intenção não foi aplicar nem propor métodos de Análise Discriminante (Johnson e Wichern, 2007), mas investigar se os dados obtidos dos perfis de massas teriam informação necessária para agrupar, ou segmentar, os tecidos de

câncer, de margem de ressecção e de controle (sem a doença). Para este fim, foram empregadas técnicas de Análise de Agrupamentos (AA), como métodos hierárquicos e não hierárquicos, visto que estas técnicas têm a intenção de formar grupos com observações apresentadas e, como a origem de cada uma das observações (os perfis) é conhecida, foi possível analisar os grupos formados por estes métodos com relação às observações oriundas dos três tipos de tecidos (Johnson e Wichern, 2007). Apesar de a análise ter sido realizada em triplicata, a primeira replicata foi retirada do estudo devido à qualidade dos resultados obtidos nesta. De maneira que, as análises foram realizadas apenas com os dados de duas replicatas e também considerando como entrada do perfil a mediana das replicatas em cada ponto de  $m/z$  do espectro. Toda a implementação computacional dos métodos empregados foi obtida com o emprego de funções do *software* estatístico R, Versão 2.13.0 (*R Development Core Team*, 2009).

#### 4. RESULTADOS E DISCUSSÃO

##### **4 A: Estudo da presença da infecção pelo vírus Epstein-Barr (EBV) em tecidos de cânceres gástricos através da biologia molecular**

Neste estudo, a idade variou de 26 a 71 anos, com uma média de 47 anos para os seis controles e de 53 anos para os dez pacientes com câncer e as respectivas margens de ressecção (RM). A infecção viral por EBV assim como a presença de *H. pylori* nos pacientes analisados estão apresentadas na **Tabela 3** abaixo.

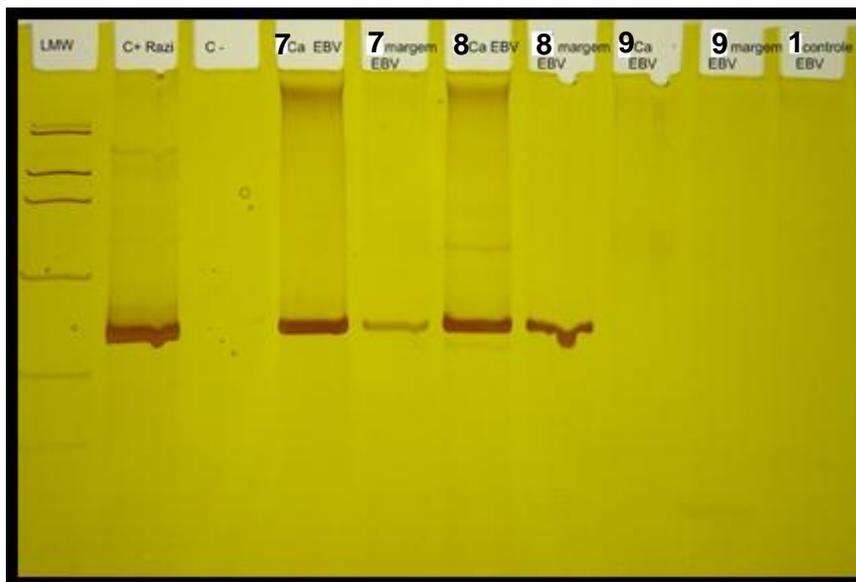
**Tabela 3** - Características clínicas e experimentais para a presença do vírus Epstein-Barr (EBV) detectada pela reação em cadeia pela polimerase em indivíduos com e sem câncer.

N <sup>o</sup>	Tipo	Sexo	Idade	Diagnóstico	<i>H. pylori</i>	EBV	EBV RM
1	Cont	F	46	Gastrite	-	-	NA
2	Cont	F	26	Gastrite	-	-	NA
3	Cont	F	52	Gastrite	-	-	NA
4	Cont	M	43	Gastrite	-	-	NA
5	Cont	F	55	Gastrite	-	-	NA
6	Cont	F	62	Gastrite	-	-	NA
7	Can	F	49	T4N3M0	-	+	+
8	Can	F	71	T3NXM0	-	+	+
9	Can	M	48	T4NXM0	-	-	-
10	Can	F	46	T4N0M0	-	-	-
11	Can	F	46	T4NXN0	+	-	+
12	Can	M	47	T3N3M0	+	+	-

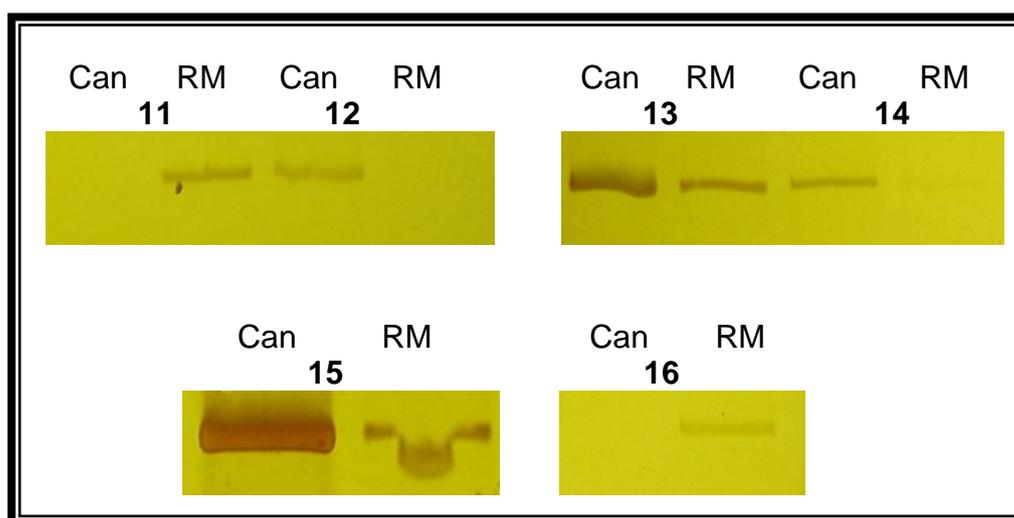
13	Can	M	68	T3NXM0	-	+	+
14	Can	M	50	T4NXM1	-	+	+
15	Can	M	40	T1NXM0	-	-	+
16	Can	M	62	T4N2M1	-	+	+

**Legenda:** Cont, RM, e Can significam amostra de controle, margem de ressecção, e câncer. M e F representam masculino e feminino, respectivamente. + e - indicam a presença ou ausência de *H. pylori* e EBV; NA é não aplicável.

Pela técnica de PCR foi possível verificar que dos dez pacientes com câncer gástrico cinco apresentaram infecção por EBV (50%) tanto no tumor quanto na respectiva margem de ressecção, um apenas no tumor e dois somente na margem de ressecção, de maneira que no total o EBV foi encontrado em 80% (8/10) dos pacientes com câncer. Em contraste, em nenhum dos seis indivíduos “controle”, o DNA do EBV foi detectado (**Figura 14 e Figura 15**). Vale ressaltar que a presença da bactéria *H. pylori* mostrada na tabela 3 acima foi detectada na análise clínica, ou seja, pelo resultado da biópsia descrita no prontuário dos pacientes.



**Figura 14** - Eletroforese em gel de poliácridamida do produto da reação em cadeia pela polimease (PCR) para detecção de sequências de DNA de EBV. Caneleta 1: LMW – Padrão de peso molecular (100; 200; 400; 800; 1600 e 2200 bp). Caneleta 2: Controle positivo, células Raji. Caneleta 3: Controle negativo, com ausência de DNA na mistura de reação. Os números correspondem aos DNAs obtidos dos pacientes com câncer gástrico (7,8 e 9) seguidos de suas margens de ressecção cirúrgica correspondentes. A última canaleta corresponde ao DNA de estômago de um indivíduo controle.



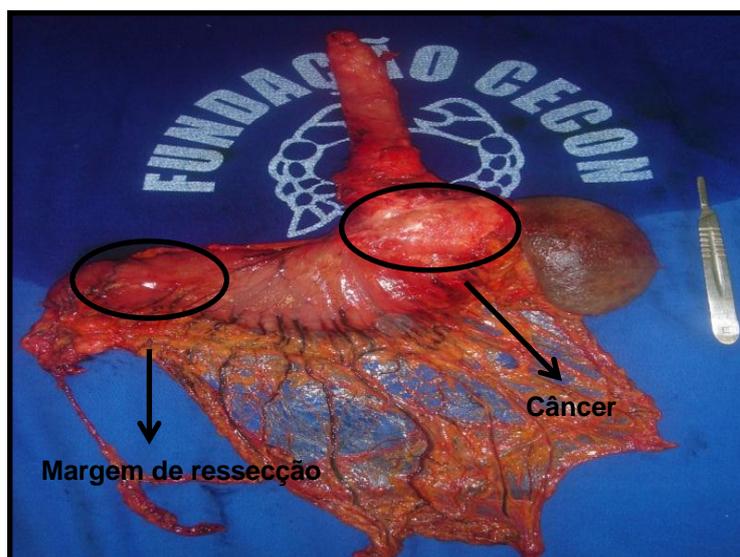
**Figura 15** - Eletroforese em gel de poliácridamida do produto amplificado da reação em cadeia pela polimerase (PCR) para detecção de sequências de DNA- EBV das amostras 11, 12, 13, 14,15 e 16, onde Can: câncer e RM: margem de ressecção.

O EBV é um vírus pertencente a família Herpesviridae, o qual é caracterizado como um vírus oncogênico ligado a alguns tipos de câncer, especialmente desde a descoberta do linfoma de Burkitt na biópsia, que foi o primeiro vírus a ser diretamente associado ao câncer humano (Pattle e Farrell, 2006). O número elevado de infecção por EBV latente pode estar relacionado com carcinogênese (Lima e Rabenhorst, 2006). Embora a evidência de infecção por EBV tenha sido relatada em alguns estudos anteriores de câncer gástrico (Lima et al., 2011; Fukayama, 2010) assim como a detecção em fases iniciais e avançadas dessa doença (Shibata et al., 1991; Sample e Sample, 2008), o mecanismo do efeito cancerígeno de EBV é ainda desconhecido. Além disso, segundo a literatura há uma baixa correlação entre esse vírus e o câncer gástrico (Lima e Rabenhorst, 2006; Lima et al., 2011), sendo que a bactéria *H. pylori* é considerada o principal agente causador deste tipo de câncer (Asaka et al., 1994; Komoto et al., 1998).

Em contraste com os registros da literatura, os resultados obtidos no presente estudo, ainda que preliminares, sugerem que o EBV pode ser considerado como um fator de risco para esta doença, especialmente no estado do Amazonas, visto que a infecção pela *H. pylori* foi detectada em apenas dois casos. Assim, levanta-se uma hipótese importante de uma alta correlação (80%) entre EBV e câncer gástrico na Amazônia, merecendo uma investigação mais aprofundada por parte da comunidade científica onde uma quantidade maior de amostras será necessária para uma maior representatividade estatística desses fatos.

#### **4 B: Estudo exploratório comparativo de candidatos a biomarcadores protéicos de cânceres gástricos mediante à técnica de MudPIT off-line**

Por se tratar de uma abordagem criteriosa que demanda tempo e altos custos, foi analisada a composição proteômica de quatro amostras de tecido tumoral de pacientes com câncer, quatro amostras das respectivas margens de ressecção (**Figura 16**) desses pacientes e cinco amostras de tecido gástrico não tumoral dos indivíduos controles. A escolha dessas amostras foi realizada de maneira aleatória. Algumas características das mesmas se encontram descritas na **tabela 4** a seguir, onde é possível observar que os controles possuem 42 anos em média, e os pacientes com câncer possuem 54 anos em média.



**Figura 16** - Estômago de um paciente exemplificando a área com câncer e a margem de ressecção.

**Tabela 4** – Características dos indivíduos envolvidos no estudo proteômico.

<b>Nº de amostra</b>	<b>Nº do indivíduo</b>	<b>Tipo de amostra</b>	<b>Idade</b>	<b>Sexo</b>	<b>Diagnóstico clínico ou cirúrgico</b>
1	12	Controle	46	Feminino	Gastrite
2	29	Controle	26	Feminino	Gastrite
3	16	Controle	52	Feminino	Gastrite
4	15	Controle	43	Masculino	Gastrite
5	4	Controle	55	Feminino	Gastrite
6	25	Câncer	49	Feminino	T4N3M0
7	11	Câncer	71	Feminino	T3NXM0
8	30	Câncer	48	Masculino	T4NXM0
9	19	Câncer	46	Feminino	T4N0M0
10	25	Margem de ressecção	49	Feminino	T4N3M0
11	11	Margem de ressecção	71	Feminino	T3NXM0
12	30	Margem de ressecção	48	Masculino	T4NXM0
13	19	Margem de ressecção	46	Feminino	T4N0M0

Na dosagem de proteínas pelo método do ácido bicinchonínico foram obtidos os valores de absorbâncias para as concentrações do padrão com as respectivas médias (**Tabela 5**) e de cada amostra (**Tabela 6**).

**Tabela 5** – Valores e médias das absorvâncias das concentrações do padrão.

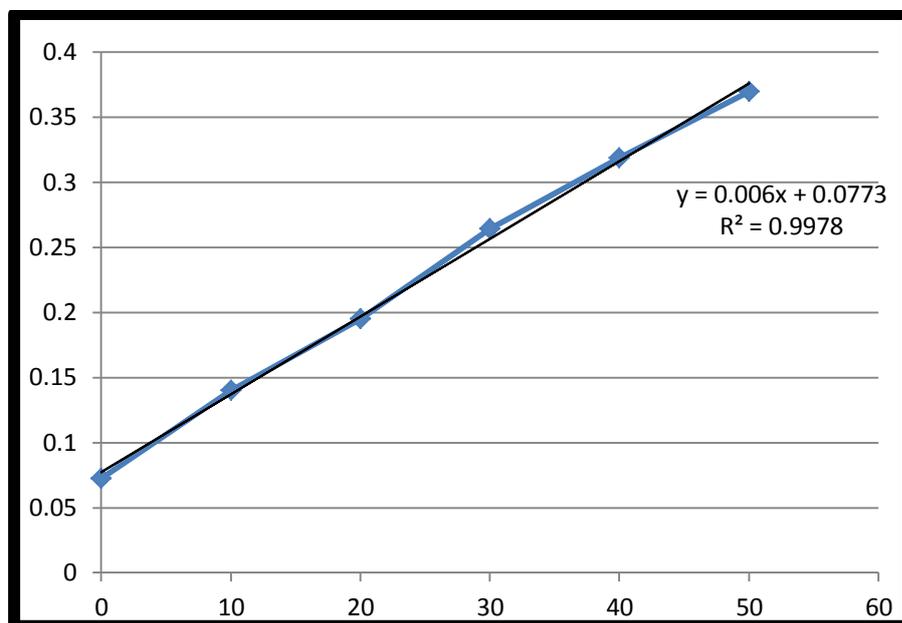
<b>Padrão</b>	<b>Abs. 1</b>	<b>Abs. 2</b>	<b>Média</b>
<b>P0</b>	0,071	0,074	0,073
<b>P10</b>	0,138	0,142	0,140
<b>P20</b>	0,203	0,187	0,195
<b>P30</b>	0,249	0,279	0,264
<b>P40</b>	0,314	0,323	0,319
<b>P50</b>	0,387	0,352	0,370

**Tabela 6** – Valores e médias das absorvâncias de cada amostra analisada.

<b>Nº de amostra</b>	<b>Abs. 1</b>	<b>Abs. 2</b>	<b>Média</b>
<b>1</b>	0,201	0,189	0,195
<b>2</b>	0,173	0,177	0,175
<b>3</b>	0,192	0,198	0,195
<b>4</b>	0,242	0,252	0,247
<b>5</b>	0,194	0,203	0,199
<b>6</b>	0,346	0,294	0,320
<b>7</b>	0,316	0,322	0,319
<b>8</b>	0,376	0,364	0,370
<b>9</b>	0,285	0,277	0,281
<b>10</b>	0,337	0,325	0,331
<b>11</b>	0,243	0,243	0,243
<b>12</b>	0,310	0,312	0,311
<b>13</b>	0,405	0,407	0,406

Com os valores das médias das absorvâncias das concentrações do padrão, construiu-se a curva padrão (**Figura 17**), através da qual se obteve a equação da reta com um  $R^2$  de 0,9978. Posteriormente, calculou-se a

quantidade de microgramas totais ( $\mu\text{g}$ ), assim como a concentração obtida de cada amostra durante a dosagem. (**Tabela 7**).



**Figura 17** – Curva padrão e equação da reta utilizada.

**Tabela 7** – Quantidade de microgramas totais e a concentração de cada amostra durante a dosagem.

Nº de amostra	$\mu\text{g}$ total	$\mu\text{g}/\mu\text{L}$
1	19,620	0,981
2	16,283	0,814
3	19,617	0,981
4	28,283	1,414
5	20,200	1,010
6	40,450	2,022
7	40,283	2,014
8	48,783	2,439
9	33,950	1,697
10	42,283	2,114

11	27,617	1,381
12	38,950	1,947
13	54,783	2,739

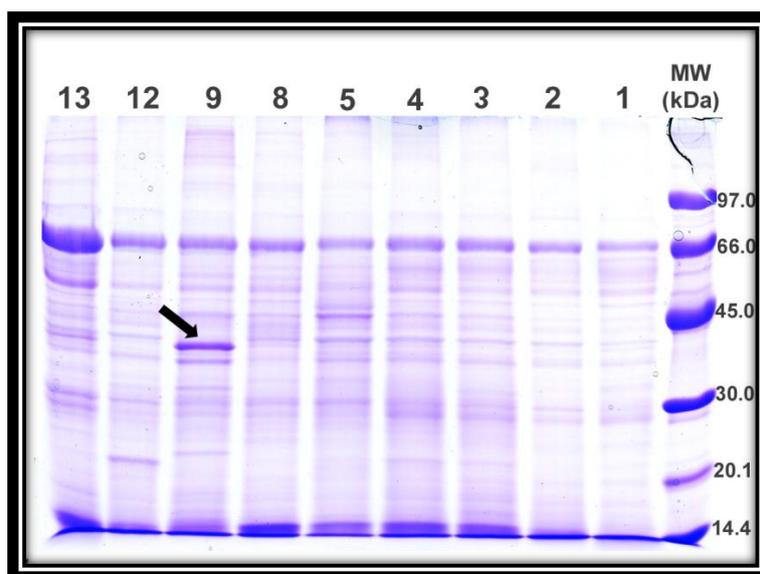
Sabendo o fator de diluição utilizado na dosagem, calculou-se a concentração de proteínas totais de cada amostra no frasco de origem. E com a informação do volume obtido do sobrenadante, calculou-se também a quantidade total de cada amostra proveniente da pulverização (**Tabela 8**).

**Tabela 8** – Informações analíticas referentes à concentração de proteínas da dosagem e do sobrenadante.

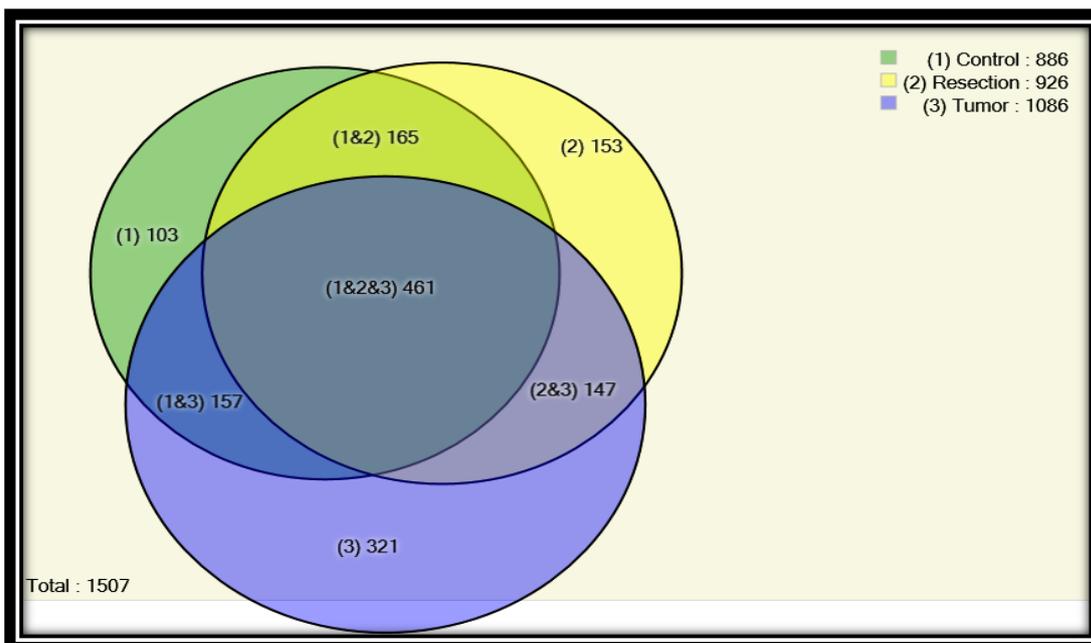
Nº de amostra	Diluição	Concentração $\mu\text{g}/\mu\text{L}$ (dosagem)	Total da Amostra ( $\mu\text{L}$ )	Quantidade total ( $\mu\text{g}$ ) (Sobrenadante)
1	5	4,904	65	318,771
2	5	4,071	65	264,604
3	5	4,904	65	318,771
4	5	7,071	65	459,604
5	5	5,050	65	328,250
6	5	10,112	75	758,437
7	5	10,071	90	906,375
8	5	12,196	90	1097,625
9	5	8,487	90	763,875
10	5	10,571	75	792,812
11	5	6,904	90	621,375
12	5	9,737	90	876,375
13	5	13,696	95	1301,104

#### 4.1 B Análise do gel unidimensional – 1 D

As amostras dos pacientes analisados nesse estudo proteômico foram submetidas à eletroforese em gel de poliacrilamida, o qual após ser corado com coomassie R-250, foi possível observar *bandas* diferencialmente expressadas, em especial a banda da amostra 9 indicado na **Figura 18**. As proteínas contidas nas bandas cortadas foram identificados pelo *Search Engine Processor* (Carvalho et al., 2012a). Através deste foi possível identificar proteínas únicas ao tumor, à margem de ressecção, ao controle assim como aquelas encontradas em um ou mais desses estados biológicos estudados descritos pelo diagrama de *Venn*, totalizando 1507 proteínas identificadas (**Figura 19**).



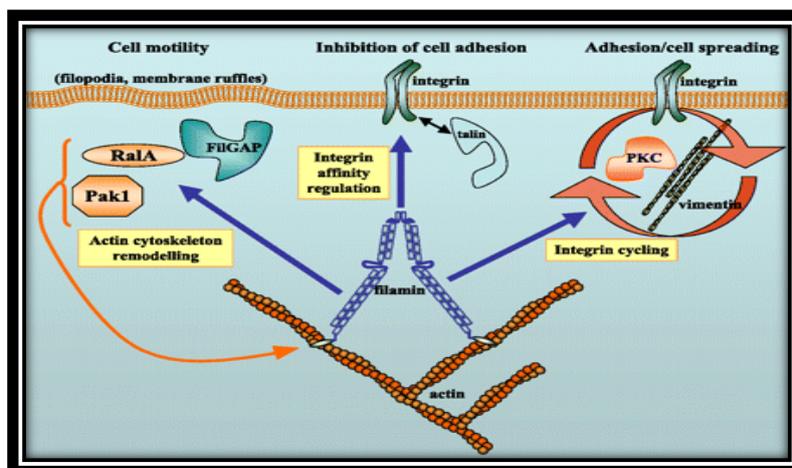
**Figura 18** - Eletroforese em gel de poliacrilamida das amostras selecionadas para o estudo proteômico. Caneletas 1, 2, 3, 4: amostras de indivíduos controle. Caneletas 8, 9: amostras da região tumoral dos pacientes com câncer; Caneletas 12 e 13: as respectivas margens de ressecção dos pacientes com câncer.



**Figura 19** - Diagrama de Venn das proteínas identificadas na análise das bandas diferencialmente expressadas no gel unidimensional.

A quantificação relativa dessas proteínas foi realizada conforme a contagem espectral (*spectrum count*). Essa pode ser entendida como uma medida relativa de quantificação onde esta atribui um número de espectros identificados em uma molécula (Carvalho et al., 2008). Ao analisar separadamente a *banda* da amostra nove (**Figura 18**) é possível verificar que as proteínas com as maiores mudanças na quantificação foram tropomiosina 2- $\beta$  e filamina-A por serem as que apresentaram as maiores contagens espectrais em todas as três análises da banda em questão dessa amostra (*download disponível* em: <http://max.ioc.fiocruz.br/pcarvalho/2012aquino>). As filaminas são proteínas citoplasmáticas que se ligam à actina, sua função tem sido relacionada à remodelação dinâmica da rede de actina do citoesqueleto durante a migração de células e a sinalização do receptor de integrina. Conseqüentemente, elas têm sido relatadas por interagir com um grande

número de proteínas celulares de grande diversidade funcional (**Figura 20**) (Ravid et al., 2008).



**Figura 20** - Multifuncionalidades que a filamina A pode apresentar, entre as quais se destacam: regulação da adesão celular, da integrina e motilidade celular. Fonte: MacPherson e Fagerholm, 2010.

As mutações na filamina A têm sido relatadas em diversas doenças, em que a perda de função celular tem sido considerada como causa da migração celular neural deficiente em resposta a estímulos microambientais, além de provocar defeitos no sistema vascular (Ravid et al., 2008; Xu et al., 2010).

A tropomiosina 2- $\beta$  é uma importante proteína que também se liga à actina, sendo responsável pela regulação/estabilização mecânica da actina, como por exemplo: a contração muscular mediante a regulação da ligação de miosina (Bharadwaj e Prasad, 2002). Alguns estudos têm mostrado que isoformas diferentes da tropomiosina podem executar funções fisiológicas distintas, permitindo a essas isoformas uma regulação específica em resposta à transformação celular (Braverman et al., 1996; He et al., 2004).

Esta família de proteínas é geralmente suprimida em células transformadas e sua diminuição tinha sido relatada em alguns carcinomas humanos (Alaiya et al., 2000; Chen et al., 2004). Com relação ao câncer gástrico, isoformas da

tropomiosina 2- $\beta$  foram encontradas superexpressadas em tumores não metastáticos (He et al., 2004). Os resultados dessa análise unidimensional estão de acordo com a literatura, uma vez que a tropomiosina está superexpressada na amostra nove, a qual é referente a um paciente com câncer que pelo diagnóstico clínico/cirúrgico não possui metástase à distância, o que explica o seu aumento e não a diminuição observada em tecidos tumorais metastáticos (Chen et al., 2004). Sugerindo assim que a tropomiosina 2- $\beta$  pode estar envolvida no câncer gástrico e também contribuir para a respectiva metástase, pois a possível diminuição dessa proteína verificada na literatura durante essa invasão dos tecidos adjacentes pode ser decorrente da ação de células cancerígenas, as quais prejudicam o processo de diferenciação celular das células à medida que elas vão se proliferando nos tecidos.

Contudo, ao analisar as proteínas identificadas em todas as análises de cada tipo de tecido estudado demonstrado pelo diagrama de Venn (**Figura 19**) verifica-se nas proteínas únicas ao tecido tumoral que a tropomiosina 2- $\beta$  possui uma contagem espectral muito baixa com relação às demais. A super expressão da tropomiosina 2- $\beta$  mostrada na **figura 18** foi bem particular e característica apenas da amostra nove desse paciente, necessitando de maiores estudos e da aplicação de outras técnicas complementares para auxiliar na melhor compreensão de proteínas tão diferentemente expressadas com relação à mesma e diferentes tipos de tecidos.

Outras proteínas unicamente identificadas no tumor foram: a ornitina carbamoiltransferase, pertencente ao ciclo da úreia e responsável por catalisar a reação entre o carbamoil fosfato e a ornitina para formar a citrulina e o fosfato;

a sinemina também chamada de desmulina, a qual é um filamento intermediário que tem a função de integrar o estress mecânico e manter a integridade estrutural das células eucarióticas; Serpina B5 também chamada de Maspina é uma proteína da família Serpina, ou seja, inibidor de serina protease. Ela pode também funcionar como um supressor de células epiteliais e sua expressão pode suprimir a capacidade de invasão das células cancerosas a outros tecidos (Sager et al., 1997). Por outro lado, a sua expressão tem sido associada com o risco de câncer de próstata, mama assim como em tumores no pâncreas (Kashima et al., 2008; Shao et al., 2008).

Entre as proteínas exclusivas do controle têm-se: glutamato desidrogenase a qual é uma enzima que converte glutamato em  $\alpha$ -cetoglutarato, e vice-versa. O aumento da produção de  $\alpha$ -cetoglutarato pode fornecer energia a ser usada no ciclo do ácido cítrico para então produzir ATP; o isocitrato desidrogenase, que também participa no ciclo do ácido cítrico sendo responsável pela descarboxilação oxidativa do isocitrato produzindo  $\alpha$ -cetoglutarato e  $\text{CO}_2$ , enquanto converte  $\text{NAD}^+$  para NADH; a pirofosfatase inorgânica, que cataliza a conversão da pirofosfatase apresentando um importante papel no metabolismo lipídico, na absorção de cálcio e na formação óssea e na síntese do DNA, assim como outras transformações químicas (Orimo et al., 1971; Usui et al., 2010).

Na margem de ressecção entre as proteínas exclusivas foram identificadas algumas relacionadas a vias metabólicas como, por exemplo: a 6-fosfogluconato desidrogenase da via das pentoses e a frutose 1,6 bifosfatase da glicólise. A proteína serina/ treonina fosfatase e a desmina foram outras

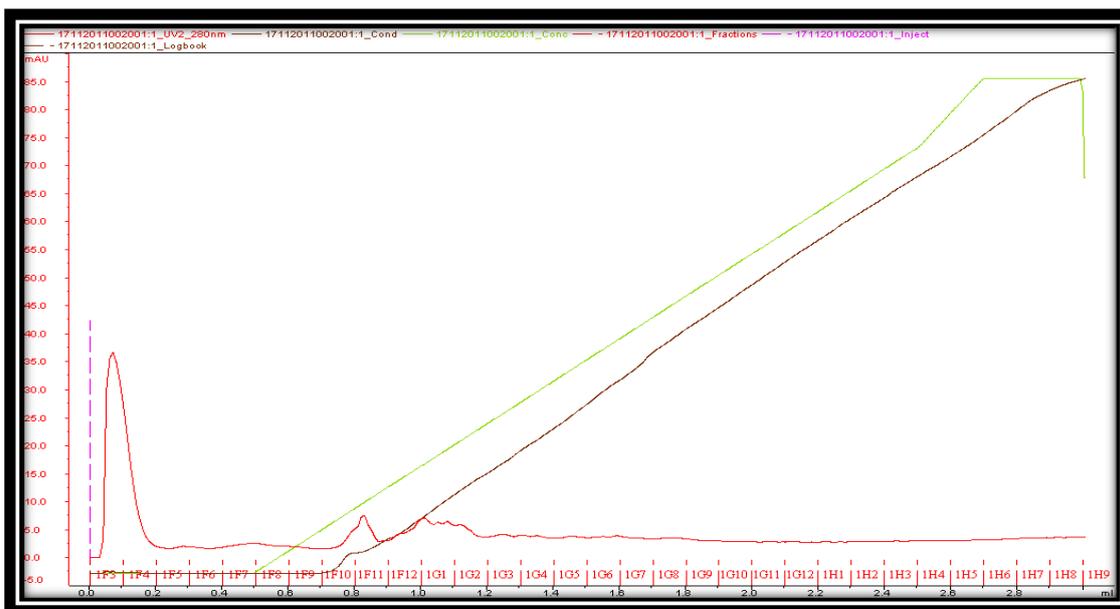
proteínas encontradas também com uma elevada contagem espectral nesse grupo.

#### 4.2 B Análise dos dados obtidos com a tecnologia multidimensional de identificação de proteínas (MudPIT) *off-line*

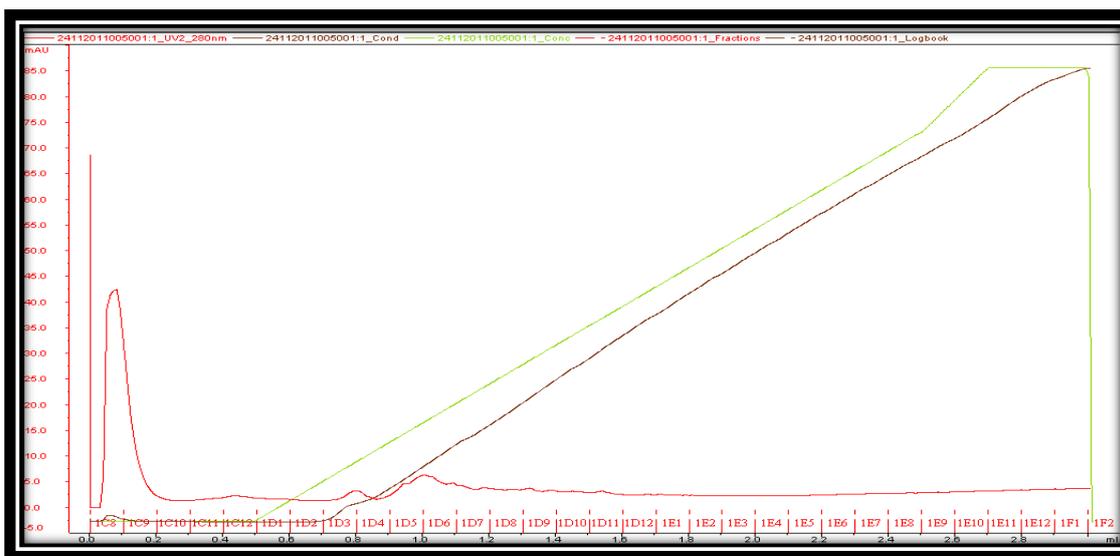
As amostras submetidas à troca iônica apresentaram cromatogramas bem semelhantes ao mostrado na **Figura 21** para os pacientes com câncer, na **Figura 22** para as respectivas margens de ressecção desses pacientes e na **Figura 23** para os controles. Sendo que de maneira geral o comportamento de todas as amostras apresentaram características de eluição bem próximas, porém com diferenças na concentração de proteínas.



**Figura 21** - Cromatograma da amostra seis na troca iônica referente a um paciente com câncer.



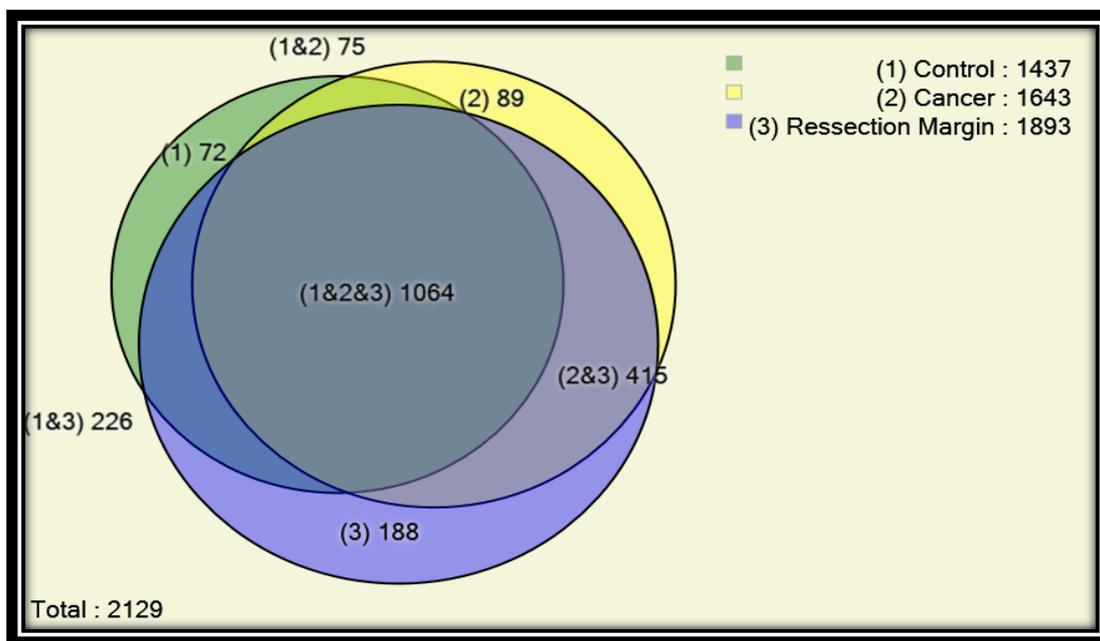
**Figura 22** – Cromatograma da amostra doze na troca iônica referente à margem de ressecção de um paciente.



**Figura 23** – Cromatograma da amostra dois na troca iônica referente a um paciente controle.

Nessa análise do MudPIT, foi possível identificar 2129 proteínas, sendo 1064 proteínas comuns às três condições biológicas estudadas e 415 proteínas correlacionadas entre a região tumoral e a margem de ressecção (**Figura 24**). Ao comparar ambas as análises, observa-se que a análise por MudPIT proporcionou uma maior identificação de proteínas em relação aquelas

identificadas no gel unidimensional. A amostra cinco referente a um controle não está presente na análise de dados por MudPIT devido ter sido perdida durante o procedimento.



**Figura 24** - Diagrama de Venn de proteínas diferencialmente expressadas identificadas na análise da tecnologia de identificação de proteínas multidimensional (MudPIT).

Entre as diversas proteínas identificadas através de cada secção do diagrama de Venn, levando em consideração também os valores da contagem espectral, unicamente expressadas nas amostras tumorais, têm-se:

- ✚ Proteína tumoral D52: Esta proteína tem sido associada a alguns tipos de câncer como, por exemplo: o de ovário (Byrne et al., 2010). No entanto, não há relatos da superexpressão dessa proteína no caso de câncer gástrico.

- ✚ Zípper de próstata leucina: Esta é um membro da família da proteína tumoral D52, correlacionada ao câncer de próstata (Wang et al., 2004). Vale ressaltar que no estudo apenas um paciente era masculino, de maneira que ao analisar os dados verificou-se tal proteína estava presente nesse paciente. Contudo, inesperadamente ela também foi identificada com seis peptídeos, o

que reafirma a sua presença na amostra de um paciente de 71 anos, do sexo feminino deste grupo.

✚ Antígeno nuclear de proliferação celular (PCNA): Este é essencial para a replicação do DNA e reparação de danos, formação da cromatina, e a progressão do ciclo celular. Dado essas diversas funções, o PCNA é descrito como um dos mediadores essenciais não-oncogênicos, o qual auxilia no crescimento do câncer (Punchihewa et al., 2012). Para o câncer gástrico, a expressão do PCNA tem sido descrito para fins de prognóstico (Konno et al., 2001).

Entre as proteínas identificadas unicamente nas margens de ressecção, têm-se:

✚ Pepsinogênio: Os pepsinogênios (PGA<sub>s</sub>) cinco e quatro, grupo I, são precursores inativos da pepsina A, os quais são formados em células da membrana mucosa do estômago. Alguns estudos relatam a associação da expressão do pepsinogênio com alterações pré-neoplásicas e neoplásicas da mucosa do estômago, e a significância como um possível indicador em casos de câncer gástrico, independentemente da infecção por *H. pylori* (Oishi et al., 2006; Kitahara et al., 1999). Assim como, sugere-se que o grupo do pepsinogênio I seja útil na detecção precoce da recorrência de câncer gástrico uma vez que os valores desta proteína tornam-se elevados nessa condição e com a passagem do tempo; enquanto em pacientes sem recorrência, os níveis desta proteína não demonstraram uma diferença substancial (Kodama et al., 1990).

✚ Colágeno  $\alpha$  -1 (COL11A1): O colágeno é uma proteína que atua na adesão celular, se encontra na matriz extracelular e tem sido correlacionado a

processos ligados a carcinogênese. Por exemplo: Zhao et al. descreveu o COL11A1 como um marcador para lesões pré-malignas no câncer. Como ela foi identificada apenas nas margens de ressecção, a sua presença pode indicar uma luz sobre a migração celular, angiogênese e morfogênese do tecido (Zhao et al., 2009; Skubitz, 2002). Esta proteína foi citada na literatura como superexpressa em câncer gástrico, em comparação com controle e também tem sido associada a processos de metástase (Yasui et al., 2004; Oue et al., 2004), assim como há relatos de sua super expressão em lesões pré-malignas e malignas no tipo de câncer em estudo (Zhao et al., 2009).

✚ Ceruloplasmina: É uma glicoproteína que é sintetizada no fígado. Estudos prévios sugerem que esta proteína pode estar envolvida na angiogênese e neovascularização, estando, portanto alinhada com a hipótese do solo e semente (margem de ressecção e tumores) (Mathot e Stenninger, 2012). Em um estudo realizado por Scanni et al. os níveis de ceruloplasmina tem sido correlacionados com o prognóstico para pacientes com câncer gastrointestinal, mostrando que níveis mais altos estavam relacionados com a evolução clínica.

✚ Calpastina: A regulação desta proteína tem sido associada com a invasão linfovascular do câncer de mama, podendo assim desempenhar um papel na disseminação metastática inicial (Storr et al., 2011).

✚ E-caderina: A caderina é uma molécula de adesão, sendo o protótipo da classe caderinas que liga as cateninas para formar o citoesqueleto. Evidências recentes têm demonstrado que a E-caderina desempenha um papel importante na fase inicial da tumorigênese pela modulação da sinalização intracelular para então promover o crescimento do tumor (Chan, 2006; Debruyne et al, 1999).

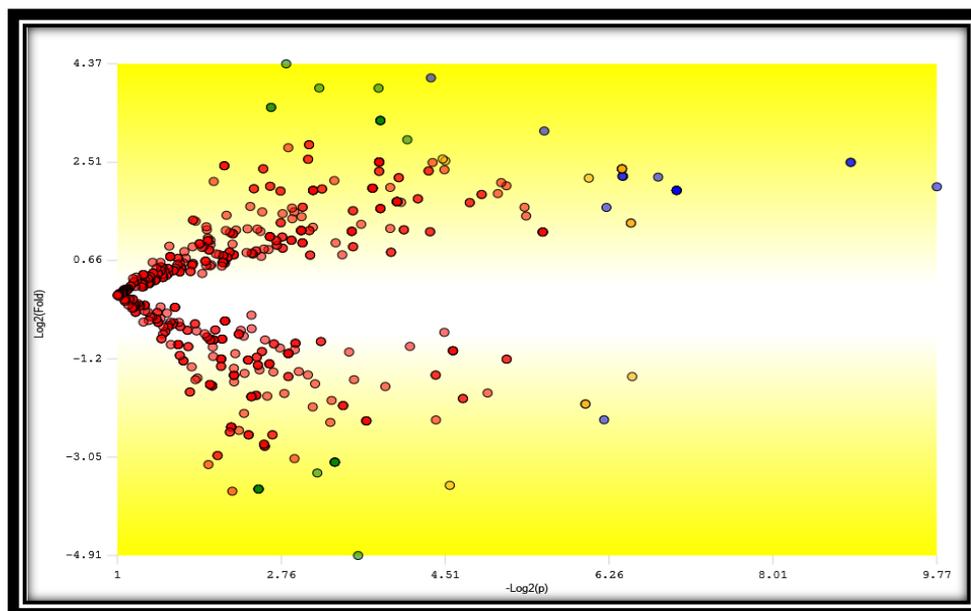
Já entre as proteínas identificadas em amostras tumorais e nas margens de ressecção correspondentes, mas não nos controles, têm-se:

✚ Fibronectina: É uma glicoproteína de matriz, a qual desempenha um papel importante nas ligações celulares, no crescimento, diferenciação, migração e espalhamento das células. A sua expressão é geralmente elevada em numerosas patologias, incluindo na estimulação do crescimento de células cancerígenas e inibição da apoptose (Fischer et al., 2008; Ebert et al., 2006). Estudos histopatológicos sugerem também que a presença elevada é topograficamente associada à invasão de adenocarcinomas gástricos e está clinicamente correlacionada com um aumento do risco de invasão local e metástase (Nejjari et al., 2004).

✚ Fibulina-1 (FBLN1): Esta proteína é uma glicoproteína de ligação ao cálcio que se encontra em associação com estruturas da matriz extracelular, como por exemplo: microfibrilas e fibras elásticas (Cheng et al., 2008). Vários estudos sugerem a interação entre a fibulina-1 e a fibronectina (Argraves et al., 1989; Twal et al., 2001). Essa proteína tem sido associada com a progressão e a desregulação tumoral ocorrida em uma gama de cânceres humanos como, por exemplo: câncer de próstata e de mama (Greene et al., 2003; Wlzlinski et al., 2007). No câncer gástrico, a hipermetilação do promotor da fibulina 1 tem sido detectada em linhagens de células e tecidos de carcinoma primário. O mecanismo molecular de como a FBLN1 age na tumorigênese ainda não está completamente compreendido. Contudo, sabe-se que a sua expressão pode suprimir a motilidade celular, inibir a fosforilação da cinase regulada pelo sinal extracelular e pode reduzir o nível de cálcio intracelular, o que é importante

para a ativação de múltiplas cascatas de sinalização tais como a via de Ras/MAPK (Cheng et al., 2008; Twal et al., 2001).

De maneira complementar à análise de expressão diferencial realizada através do diagram de Venn, o TFold foi empregado utilizando um  $q$ -valor de 0.05. Neste foi comparada a margem de ressecção frente à amostra tumoral e os controles. A **figura 25** mostra a comparação da margem de ressecção com o controle, onde quinze proteínas (pontos azuis) foram consideradas diferencialmente expressadas satisfazendo todos os filtros estatísticos; vinte proteínas (pontos verdes) passaram na grandeza de variação (*fold change*), porém não passaram no estimador teórico da taxa de falso positivo (FDR), indicando que a diferença foi ao acaso; quatorze proteínas (pontos laranja) que satisfazem tanto o *fold change* como o estimador teórico, porém são proteínas pouco abundantes que provavelmente tiveram sua quantificação comprometida. Seiscentos e trinta proteínas (pontos vermelhos) não satisfizeram nenhum dos critérios estatísticos usados. Sendo que as proteínas na parte superior dessa figura correspondem às proteínas superexpressadas na margem de ressecção.

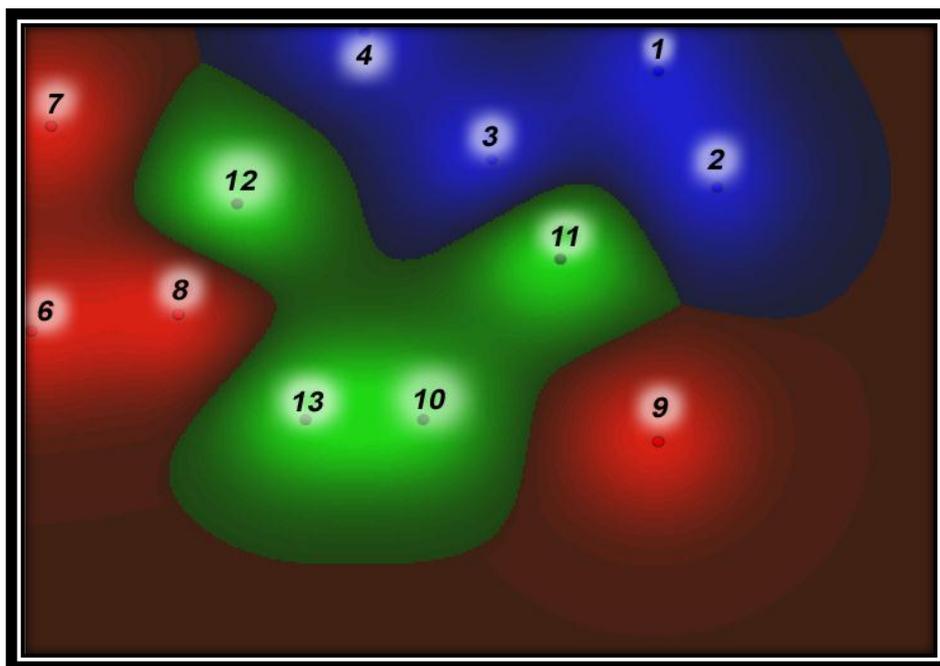


**Figura 25** - Análise do TFold comparando perfis proteômicos de proteínas identificadas em duas ou mais replicatas biológicas de amostras de indivíduos controle frente às da margem de ressecção. Cada proteína foi mapeada como um ponto no gráfico de acordo com  $-\text{Log}_2(p\text{-valor})$  (eixo x) e  $\text{Log}_2(\text{Fold change})$  (eixo y). Os pontos vermelhos não passaram no *Fold-change* nem no critério do *p*-valor; os pontos laranjas apresentaram baixa quantificação, portanto foram descartadas); os pontos verdes passaram no *Fold-change* mas não no *p*-valor; e os pontos azuis passaram em ambos os testes.

Ao comparar pelo TFold a margem de ressecção frente à amostra tumoral, foram encontrados os seguintes resultados: oito proteínas que se encaixam nos critérios dos pontos azuis descritos acima, dezenove proteínas correspondem a pontos verdes, treze proteínas a pontos laranja e setecentos e vinte e duas proteínas a pontos vermelhos. Além disso, ao fazer essa comparação da margem *versus* câncer, foi detectada uma baixa regulação da gelsolina nas amostras tumorais, o que também foi encontrado em outros estudos (Zhu et al., 2012). Segundo esses autores a baixa regulação dessa proteína encontrada na margem de ressecção correspondente pode estar

ligada à metástase, visto que esta proteína aumenta a permeabilidade e tem sido associada com a mobilidade do tumor (Zhu et al., 2012).

A análise *Multidimensional Scaling* (MDS) foi empregada para investigar a proximidade de agrupamento entre os indivíduos controle, com câncer e a respectiva margem de ressecção conforme descrito anteriormente no item 3.5.10 B. Os resultados desse agrupamento são mostrados na **Figura 26**.



**Figura 26** - A análise de *Multidimensional Scaling* (MDS) dos dados obtidos a partir da tecnologia de identificação de proteínas multidimensional (MudPIT) de amostras de tecidos de indivíduos controle, com câncer e respectiva margem de ressecção. O número em cada ponto representa a identificação da amostra correspondente. As regiões azuis, verdes e vermelhas delimitam as regiões do controle, margem de ressecção e câncer, respectivamente. Sendo que os limites foram realizados usando uma função discriminante *radial-basis function* (RBF).

Ao examinar os dados provenientes da análise MDS, pode-se inferir algumas observações. Curiosamente, os indivíduos do grupo controle e das margens de ressecção parecem ter sido agrupados mais efetivamente e mais próximo quando comparado àqueles das amostras tumorais, no qual foi obtido

um ponto separadamente dos demais, ou seja, a região verde que representa a margem de ressecção está um pouco mais separada da região azul (controle) do que da região vermelha (câncer). Contudo, parece um erro agrupar os membros da margem de ressecção com aqueles do controle, já que apesar de próximos apresentam delimitações de classificação bem distintas. O que reforça a visão de que a margem possui características muito específicas, se diferenciando, portanto do tecido saudável. Quanto a amostra nove classificada separadamente do grupo tumoral, conforme descrito na análise da eletroforese unidimensional, ela apresenta um padrão discriminatório comparada aos demais, fazendo com que a abordagem computacional utilizada nesta análise fosse repensada, conforme limitação já descrita anteriormente no item 4.1 B. Resumidamente, através da análise ACFold foi possível identificar a tropomiosina 2- $\beta$  e a filamina-A em elevada concentração na amostra nove, justificando sua separação das demais amostras de câncer. Segundo a literatura, foi levantada a hipótese de que, em conjunto, estas duas proteínas desempenham um papel chave em “um mecanismo pelo qual ocorre a troca de uma resposta tumorigênica TGF- $\beta$ ”, ou seja, essas duas proteínas apresentam um mecanismo pelo qual ocorre a troca de um fator de crescimento  $\beta$  que auxilia na propagação das células cancerígenas (Assinder e Cole, 2011).

Todos os arquivos de espectros de massas *raw*, os arquivos intermediários do *PatternLab*, e a completa lista de identificações das proteínas do gel unidimensional e do MudPIT para todos os tipos de tecido estão disponíveis para *download* em: <http://max.ioc.fiocruz.br/pcarvalho/2012aquino>.

#### **4 C: Estudo exploratório comparativo de perfis metabolômicos de cânceres gástricos**

A análise metabolômica foi realizada com biópsias gástricas de dezesseis indivíduos controle, dezesseis pacientes com câncer e as correspondentes dezesseis margens de ressecção, totalizando quarenta e oito amostras analisadas. Algumas características desses indivíduos encontram-se descritas na **tabela 9** a seguir, onde é possível observar que os indivíduos controle tinham em média 41 anos, enquanto que os com câncer tinham uma média de 55 anos.

**Tabela 9** – Características dos indivíduos envolvidos no estudo metabolômico.

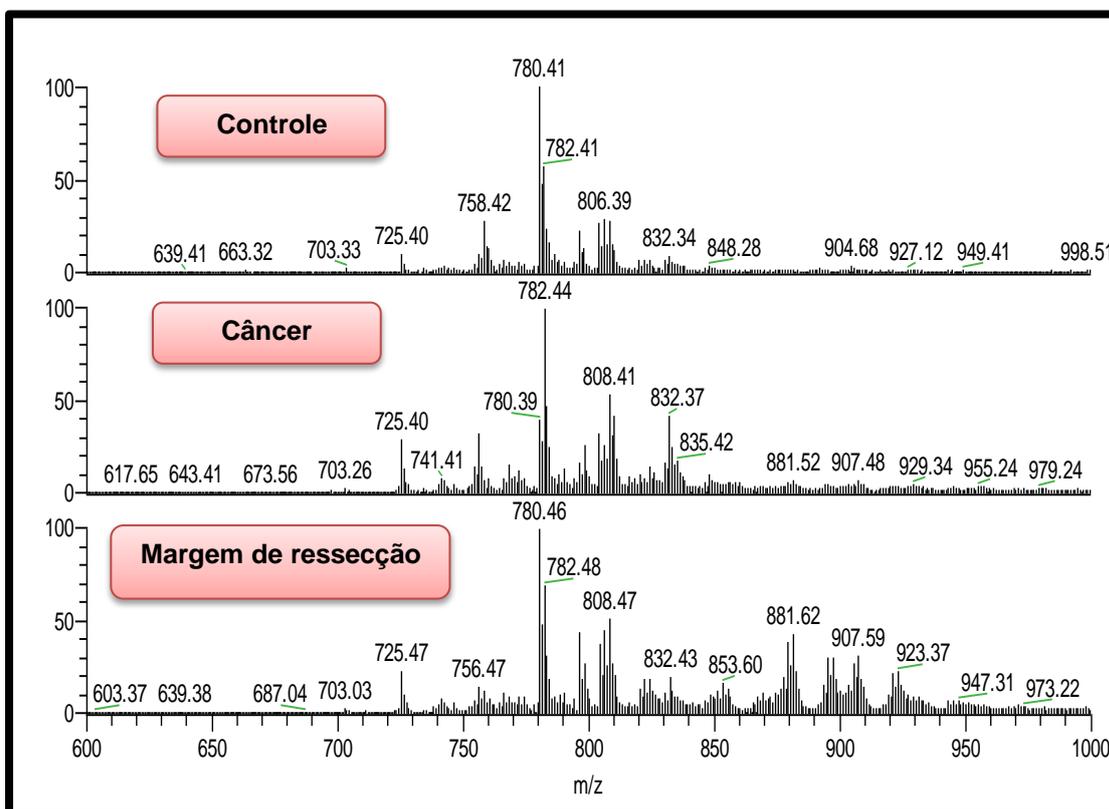
<b>Nº da amostra</b>	<b>Nº do indivíduo</b>	<b>Tipo de amostra</b>	<b>Idade</b>	<b>Sexo</b>	<b>Diagnóstico clínico ou cirúrgico</b>	<b>H. Pylori</b>
1	22	Controle	28	Feminino	Gastrite	Positivo
2	37	Câncer	62	Feminino	T2NXM0	Negativo
3	37	Margem de ressecção	62	Feminino	T2NXM0	Negativo
4	27	Controle	33	Feminino	Gastrite	Negativo
5	2	Controle	44	Feminino	Gastrite	Positivo
6	5	Controle	34	Feminino	Gastrite	Negativo
7	34	Câncer	66	Masculino	T4NXM1	Negativo
8	34	Margem de ressecção	66	Masculino	T4NXM1	Negativo
9	31	Câncer	59	Masculino	T2NXM0	Negativo
10	31	Margem de ressecção	59	Masculino	T2NXM0	Negativo
11*	12	Controle	46	Feminino	Gastrite	Negativo
12	20	Controle	34	Feminino	Gastrite	Positivo

<b>13*</b>	30	Câncer	48	Masculino	T4NXM0	Negativo
<b>14*</b>	30	Margem de ressecção	48	Masculino	T4NXM0	Negativo
<b>15*</b>	29	Controle	26	Feminino	Gastrite	Positivo
<b>16</b>	23	Controle	19	Feminino	Gastrite	Negativo
<b>17*</b>	15	Controle	43	Masculino	Gastrite e presença de agregados de linfócitos	Positivo
<b>18*</b>	19	Câncer	46	Feminino	T4NXM0	Negativo
<b>19*</b>	19	Margem de ressecção	46	Feminino	T4NXM0	Negativo
<b>20</b>	18	Controle	50	Feminino	Gastrite	Positivo
<b>21</b>	14	Controle	60	Feminino	Gastrite	Positivo
<b>22</b>	13	Margem de ressecção	46	Feminino	T4NXM0	Positivo
<b>23</b>	13	Câncer	46	Feminino	T4NXM0	Positivo
<b>24*</b>	25	Margem de ressecção	49	Feminino	T4N3M0	Negativo
<b>25*</b>	25	Câncer	49	Feminino	T4N3M0	Negativo
<b>26</b>	9	Câncer	47	Masculino	T3NXM0	Positivo
<b>27</b>	9	Margem de ressecção	47	Masculino	T3NXM0	Positivo
<b>28</b>	24	Margem de ressecção	50	Masculino	T4NXM1	Negativo
<b>29</b>	24	Câncer	50	Masculino	T4NXM1	Negativo
<b>30</b>	8	Controle	31	Feminino	Gastrite	Negativo
<b>31*</b>	11	Câncer	71	Feminino	T3NXM0	Negativo

<b>32*</b>	11	Margem de ressecção	71	Feminino	T3NXM0	Negativo
<b>33</b>	4	Controle	55	Feminino	Gastrite	Negativo
<b>34</b>	6	Controle	72	Feminino	Gastrite	Negativo
<b>35</b>	38	Câncer	52	Feminino	T4NXM1	Negativo
<b>36</b>	38	Margem de ressecção	52	Feminino	T4NXM1	Negativo
<b>37</b>	39	Câncer	61	Masculino	T2NXM0	Positivo
<b>38</b>	39	Margem de ressecção	61	Masculino	T2NXM0	Positivo
<b>39</b>	33	Câncer	62	Masculino	T4N2M1	Negativo
<b>40</b>	33	Margem de ressecção	62	Masculino	T4N2M1	Negativo
<b>41</b>	10	Margem de ressecção	68	Masculino	T3NXM0	Negativo
<b>42</b>	10	Câncer	68	Masculino	T3NXM0	Negativo
<b>43</b>	32	Câncer	40	Masculino	T1NXM0	Negativo
<b>44</b>	32	Margem de ressecção	40	Masculino	T1NXM0	Negativo
<b>45</b>	26	Controle	44	Masculino	Gastrite e pólipo hiperplásico de antro	Negativo
<b>46</b>	17	Controle	32	Feminino	Gastrite	Negativo
<b>47</b>	40	Margem de ressecção	60	Masculino	T3NXM0	Positivo
<b>48</b>	40	Câncer	60	Masculino	T3NXM0	Positivo

\*Amostras utilizadas também na análise proteômica.

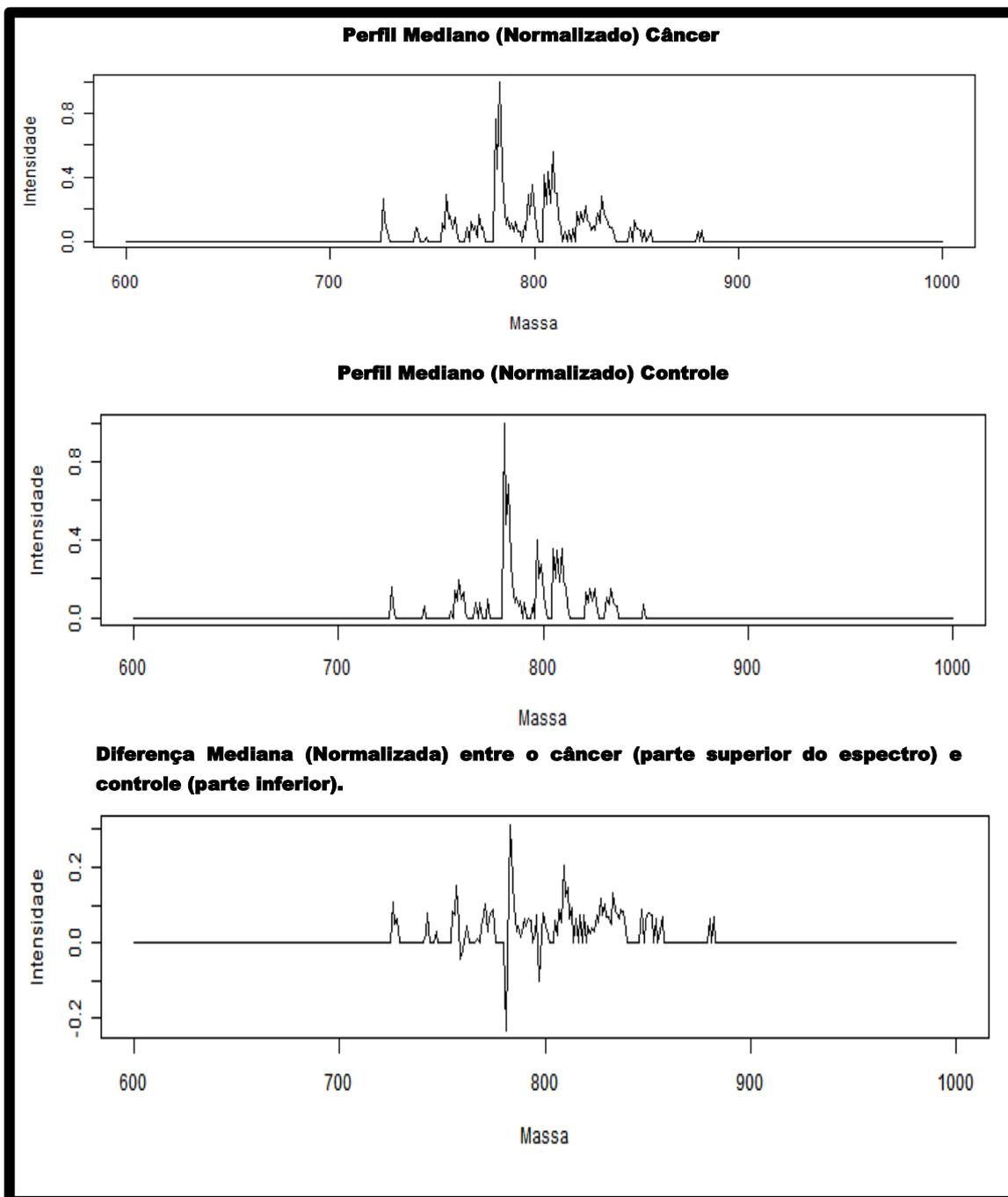
A **figura 27** exemplifica os perfis metabolômicos dos tipos de tecidos analisados, onde se pode verificar que os perfis das amostras dos indivíduos controle são semelhantes aos do câncer, apresentando apenas diferenças relativas à intensidade ou ausência de alguns picos de razão  $m/z$ .



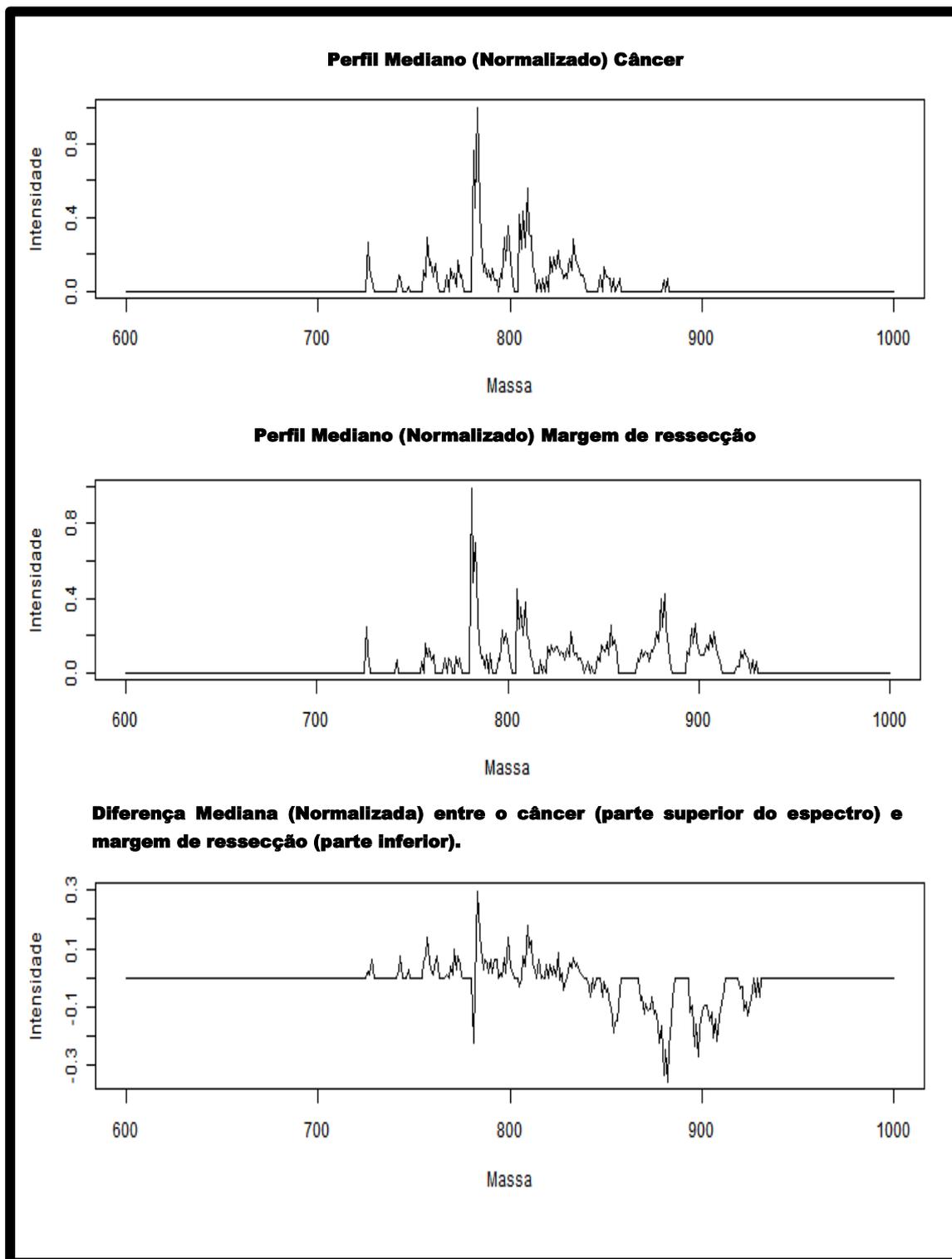
**Figura 27** – Comparação dos perfis metabolômicos das amostras dos indivíduos controle, pacientes com câncer e a correspondente margem de ressecção.

Tal semelhança visualizada nos perfis entre os controles e os cânceres foi confirmada pelo perfil mediano normalizado das amostras (**Figura 28**) onde se verifica também e de forma mais clara a diferença mediana entre eles, ou seja, a variação na presença ou ausência de determinado pico assim como na sua intensidade em cada tipo de tecido. Fica assim demonstrado que os perfis de tais tipos de tecidos não são tão semelhantes conforme aparentam pelos perfis medianos. Essas diferenças de presença ou intensidades relativas de picos indicam prováveis substâncias produzidas, suprimidas ou cujas

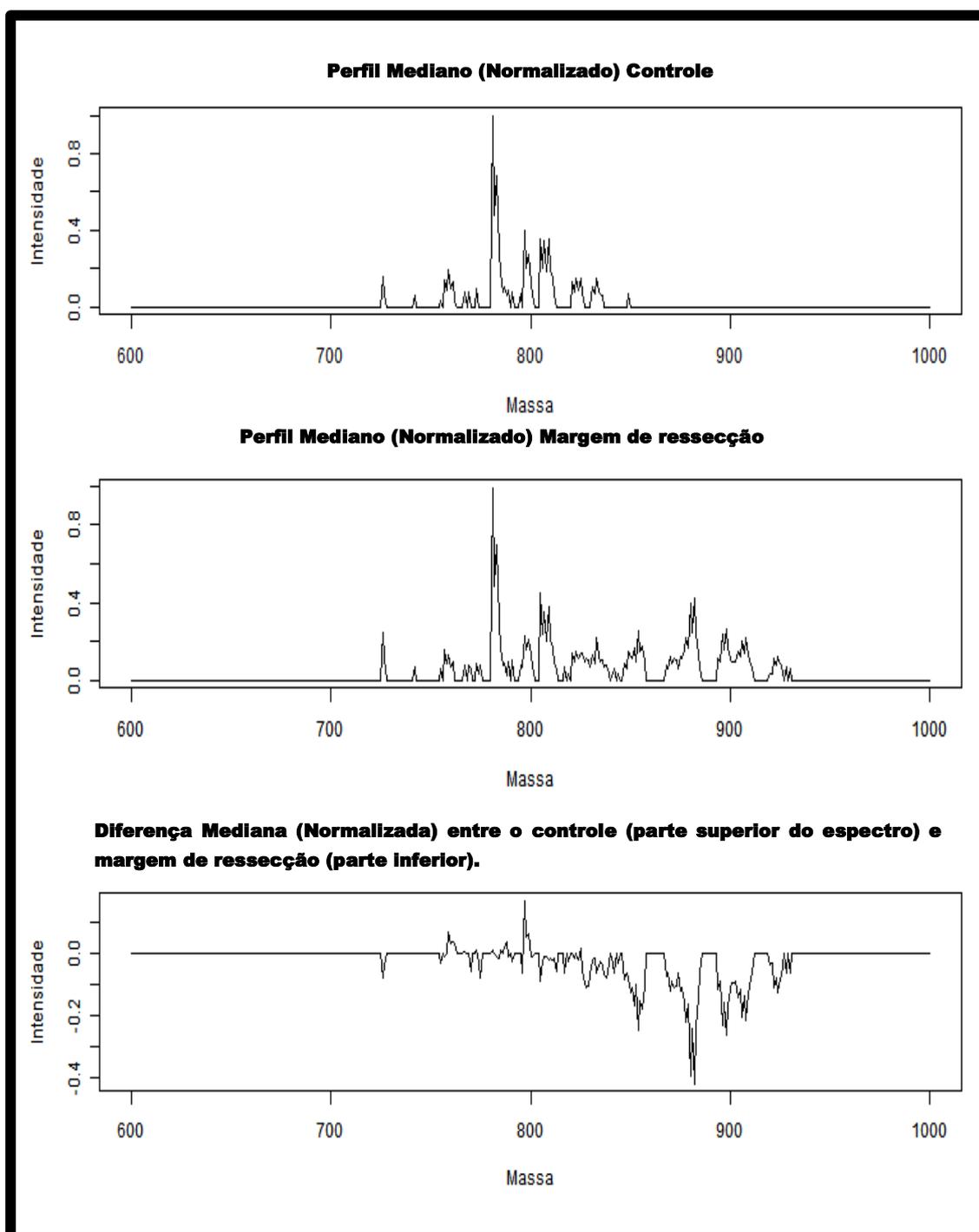
produções apenas alteraram em amplitudes como consequência do câncer gástrico. Surpreendentemente na presente análise metabolômica os efeitos dessa doença foram mais evidenciados no perfil mediano da margem de ressecção, em comparação com os perfis medianos do câncer e do controle (**Figura 29 e 30**). Essa observação está de acordo com algumas evidências do resultado da proteômica, em que foram observadas proteínas na margem de ressecção, que segundo a literatura algumas delas já foram relacionadas à progressão tumoral assim como foi verificado que a margem de ressecção possui uma característica distinta quando comparada aos demais tipos de tecido analisado. Fato esse que também foi visualizado na metabolômica, ao verificar o perfil bem diferente da margem de ressecção na região de  $m/z$  de 850-930. Logo, pode se sugerir que o perfil distinto da margem de ressecção seja a manifestação de uma batalha bioquímica entre as células do organismo e as cancerígenas, as primeiras em estado de defesa e as últimas em estado de ataque invasivo, inclusive pela busca de nutrição para a própria sobrevivência e propagação nos tecidos próximos e distantes (metástase).



**Figura 28** - Comparação dos perfis medianos (normalizados) do câncer e do controle assim como a diferença mediana entre eles.



**Figura 29** - Comparação dos perfis medianos (normalizados) do câncer e da margem de ressecção, e a diferença mediana entre eles.



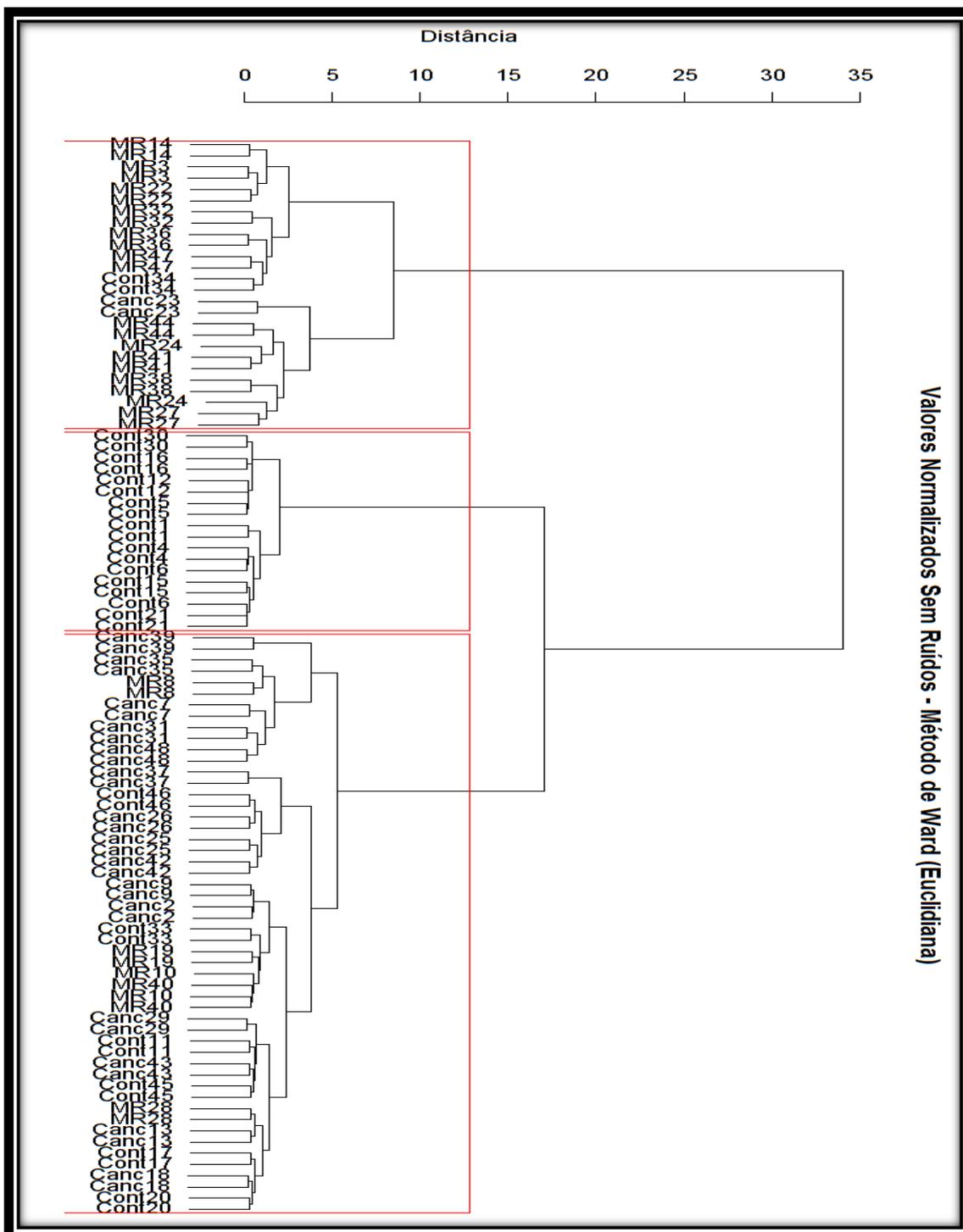
**Figura 30** - Comparação dos perfis medianos (normalizados) do controle e da margem de ressecção, e a diferença mediana entre eles.

Como a extração de informações dos resultados de um experimento químico geralmente envolve a análise de um grande número de variáveis é necessária à utilização de ferramentas estatísticas. Para a análise dos dados obtidos por inserção direta no espectrômetro de massas foi decidido que as

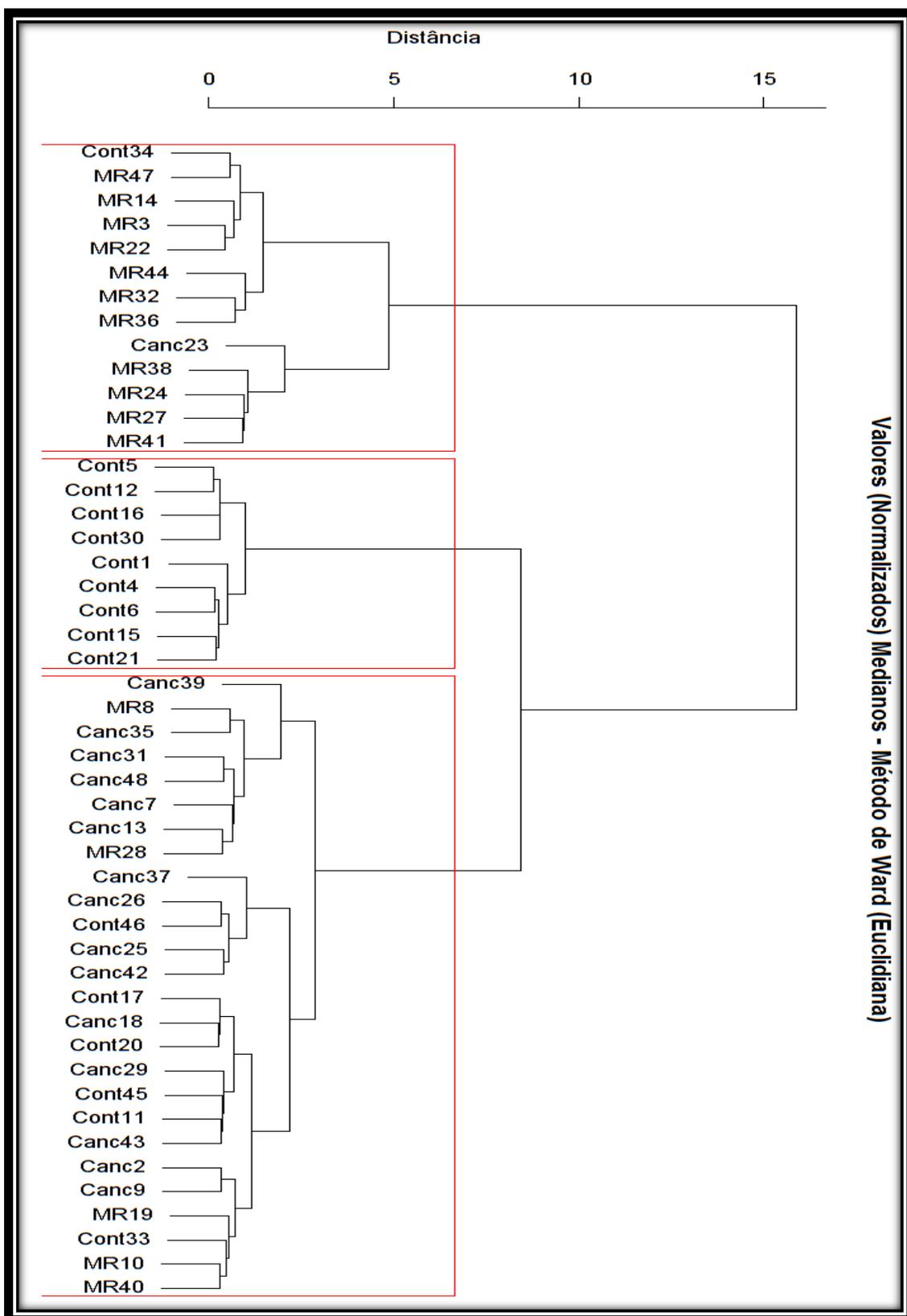
variáveis a serem analisadas seriam todos os íons de  $m/z$  600 a 1000, faixa previamente escolhida por ser a mais rica em informações nos espectros de todas as amostras. Para a normalização dos dados os valores menores ou iguais a 6% da intensidade do íon mais abundante foram considerado como ruído e lhes foi atribuído o valor zero. Esta redução de variáveis permitiu a construção de gráficos bidimensionais contendo maior informação estatística, o que foi possível através de agrupamentos hierárquicos entre as amostras de acordo com suas similaridades, utilizando as variáveis escolhidas e representando-as de maneira bidimensional através de dendrogramas (Johnson e Wichern, 2007; Filho, 2010). Brevemente, a técnica de agrupamento hierárquico interliga as amostras por suas associações, produzindo um dendrograma onde as amostras semelhantes, segundo as variáveis escolhidas, são agrupadas entre si. A suposição básica de sua interpretação é esta: quanto menor a distância entre os pontos, maior a semelhança entre as amostras. Existem muitas maneiras de procurar agrupamentos no espaço  $n$ -dimensional. A mais simples consiste em agrupar pares de pontos que são mais próximos, usando a distância euclidiana, e substituí-los por um novo ponto localizado na metade da distância entre eles (Johnson e Wichern, 2007; Filho, 2010; Santos et al., 2009).

Portanto, devido à natureza exploratória dos métodos hierárquicos na análise de agrupamentos (AA), foram empregados três métodos: o da média das ligações (*Average Linkage*), o centróide (*Centroid*) e o de Ward (*Ward*) (Johnson e Wichern, 2007). Os métodos da média das ligações e o centróide não enfatizaram claramente a presença de grupo nos dados, ou seja, não demonstraram uma nítida classificação discriminatória dos dados obtidos em

grupos diferentes. Por outro lado, o método de Ward, um dos mais eficientes para identificar agrupamento em conjunto de dados e que usa como métrica de semelhança a distância euclidiana de vetores gerados a partir das intensidades de cada  $m/z$ , apresentou um dendrograma indicando de maneira bem definida a presença de três grupos nas observações dos perfis, tanto para as observações com as replicatas, como também, com as medianas das replicatas, como pode ser observado nas **figuras 31 e 32**.



**Figura 31** – Dendrograma das distâncias euclidianas das réplicas das amostras normalizadas usadas nesse estudo.



**Figura 32** - Dendrograma das distâncias euclidianas da média das amostras normalizadas usadas nesse estudo.

A análise dos dendrogramas das distâncias euclidianas revelou que o método foi eficaz em agrupar os três tipos de tecidos empregados no estudo, e corroborou com o perfil mediano, ao revelar indivíduos controle bem próximos de pacientes com câncer. Este fato intrigante necessita de maiores investigações, especialmente porque entre os indivíduos agrupados com o grupo do câncer existe pelo menos um (controle 45) que apresentou no diagnóstico clínico um pólipó hiperplásico de antro, o qual é um crescimento anormal benigno em torno da mucosa intestinal (tumores não cancerígenos) que necessita de acompanhamento regularmente para a investigação de possível malignidade.

Em contrapartida, a metodologia empregada neste estudo apresentou relativa robustez, visto que ao analisar tanto o dendrograma das distâncias euclidianas das réplicas como o das medianas das amostras normalizadas verifica-se que ambas foram bem semelhantes, ou seja, que os dados obtidos durante as duas réplicas foram bem coerentes entre si. Observa-se que praticamente todas as réplicas estão intimamente agrupadas (**Figura 31**). Outro fato observado nessa análise é que existem indivíduos da margem de ressecção que agruparam com os de câncer, corroborando com os resultados obtidos da proteômica de que a margem de ressecção apresenta características mais semelhantes ao grupo do câncer, inclusive a amostra nove (câncer) que demonstrou ser tão distinta das demais na análise proteômica teve sua margem de ressecção (MR 19) agrupada entre os indivíduos com câncer pelo método de Ward.

Além disso, para complementar a análise exploratória dos dados, foi empregado o método não hierárquico *k-Means* (Johnson e Wichern, 2007).

Neste caso, o número de grupos de interesse é fixado e o método forma os grupos de maneira que a variância das observações dentro de cada grupo seja minimizada, sendo considerado na literatura muito eficiente em formar grupos cujas observações componentes sejam similares com relação às características que as descrevem (Hastie et al., 2009). Os resultados obtidos para três grupos ( $k=3$ ), com dez inicializações aleatórias do método, são apresentados na **Tabela 10**.

**Tabela 10** – Agrupamento de k-Means com a mediana dos valores normalizados para os três tipos de tecido utilizados no estudo. Sendo grupo 1 = Margem de ressecção; 2= Controle e 3= Câncer.

<b>Tecido /Grupo</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Formado</b>			
<b>Câncer</b>	1	0	15
<b>Controle</b>	1	9	6
<b>Margem</b>	11	0	5

Ao analisar a **tabela 10**, verifica-se que a análise exploratória dos dados foi eficaz ao classificar os tipos de tecido do câncer e da margem, especialmente os primeiros, com quinze dos dezesseis indivíduos (94%) classificados de maneira correta com os dados do diagnóstico clínico. A margem de ressecção teve onze dos dezesseis indivíduos (69%) agrupados também de maneira coerente com a avaliação clínica. As cinco amostras de margens de ressecção agrupadas com as de câncer, quatro são de indivíduos em estágio avançado, T4 (MR 8, 19, 28 e 40) e uma de paciente em estágio intermediário, T2 (MR 10), de acordo com diagnóstico clínico/cirúrgico. Entre

estes indivíduos em estágio T4, três apresentaram metástase a distância observável (M1), possível causa de terem sido incluídas no grupo de câncer. Por fim, o grupo controle teve a classificação menos efetiva, com apenas nove indivíduos (56%) agrupados de forma coerente com a análise clínica. A alta taxa de amostras controle agrupadas com as de câncer é coerente com a relativa semelhança observada entre os espectros medianos de ambos os tipos (Figura 28), e pode indicar uma limitação da metodologia aplicada (por exemplo, devido a faixa de valores  $m/z$  escolhida) ou mesmo ser um indicativo de que tais indivíduos estejam em estágio inicial da doença, o que confirmaria o potencial da abordagem metabolômica para o diagnóstico precoce. A continuação das investigações, com novas análises dessas amostras e novas amostragens de voluntários, bem como o acompanhamento clínico dos indivíduos envolvidos por um período de alguns anos, é necessária para inferir se o diagnóstico precoce é possível pela abordagem metabolômica adotada nesta dissertação.

## 5. CONSIDERAÇÕES FINAIS

A alta incidência de 80% (8/10) do vírus Epstein-Barr (EBV) nas amostras tumorais e nas margens cirúrgicas correspondentes, e em nenhum dos seis indivíduos controle bem como a presença de *H. Pylori* em apenas dois dos casos estudados sugerem que EBV pode ser considerado como um fator de risco para adenocarcinoma gástrico em Manaus, o que deve ser determinado por estudos adicionais em larga escala.

Ao comparar perfis protéicos das amostras de tecidos de câncer, margem de ressecção e do controle, observou-se que os perfis da margem de ressecção são mais semelhantes aos do câncer do que aos do controle. Durante esta comparação, várias proteínas encontradas já haviam sido previamente correlacionadas com a doença, como, por exemplo: algumas proteínas, as quais segundo a literatura possuem ligação com o crescimento tumoral e que foram encontradas na margem de ressecção, o que reforça a importância do estudo dessa região para o entendimento do processo de tumorigênese.

Embora tenham sido encontradas proteínas diferencialmente expressadas e candidatas a biomarcadores, um estudo muito mais amplo e específico é necessário para tal confirmação e validação. Porém, tais resultados demonstram claramente que a margem de ressecção possui características bem específicas que merecem uma melhor compreensão e pode ajudar no desenvolvimento de futuros tratamentos.

Além disso, a análise por MDS revelou limitações na estratégia da proteômica diferencial empregada no presente estudo. Pois a comparação de perfis proteômicos *shotgun* sem considerar a análise de amostras

independentes, através de ferramentas como a análise do gel 1D ou um algoritmo como MDS, pode ocasionar a perda de informações importantes. Neste trabalho, a análise por TFold perdeu características marcantes relacionadas as proteínas superexpressadas na amostra nove. Embora, tais proteínas tenham sido também encontradas em outros pacientes, um considerável desvio padrão existe assim como lacunas nas estratégias estatísticas mais comuns. Logo, ao empregar uma abordagem complementar simples, é possível investigar as características únicas que explicam a particularidade desse indivíduo. Logo, deve ser dada importância também para a individualidade da amostra, já que cada paciente apresenta características únicas.

A metodologia empregada de análise metabolômica de amostras do estômago de indivíduos com e sem câncer por espectrometria de massas apresentou robustez de acordo com a comparação estatística dos dados. Pelo estudo exploratório usando o método de Ward foi possível agrupar corretamente 72,9% dos três tipos de tecidos, o mesmo percentual obtido pelo método *k-Means*, enquanto os perfis medianos e as respectivas diferenças demonstraram características espectrais típicas de cada grupo. Embora não haja dados suficientes para afirmar se os desvios de classificação observados são devidos a limitações da metodologia empregada ou indicativos de diagnósticos não detectáveis por outros métodos, por exemplo, a indicação precoce da doença, os resultados alcançados estimulam a continuidade das investigações.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ALAIYA, A.; ROBLICK, U.; EGEVAD, L. ; CARLSSON, A.; FRANZÉN, B.; VOLZ, D. et al. Polypeptide expression in prostate hyperplasia and prostate adenocarcinoma. *Anal Cell Pathol.*, 21(1):1-9, 2000.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Molecular Biology of the Cell*. United States of America: Garland Science, 5<sup>o</sup> edição, 2008.

ALMEIDA, J. R. C.; PEDROSA, N. L.; LEITE, J. B.; FLEMING, T. R. P.; CARVALHO, V. H.; CARDOSO, A. A. A. Marcadores tumorais: revisão de literatura. *Revista Brasileira de Cancerologia*, 53(3): 305-316, 2007.

AQUINO, P. F.; FISCHER, J. S. G.; NEVES-FERREIRA, A. G. C.; PERALES, J.; DOMONT, G. B.; ARAUJO, G. D. T. et al. Are gastric cancer resection margin profiles more similar to those from controls or tumors? *Journal of Proteome Research*, submetido em julho de 2012.

ARGRAVES, W. S.; DICKERSON, K.; BURGESS, W. H.; RUOSLAHTI, E. Fibulin, a novel protein that interacts with the fibronectin receptor beta subunit cytoplasmic domain. *Cell*, 58(4): 623-629, 1989.

ASAKA, M.; KIMURA, T.; KATO, M.; KUDO, M.; MIKI, K.; OGOSHI, K. et al. Possible role of Helicobacter pylori infection in early gastric cancer development. *Cancer*. 73: 2691-2694, 1994.

ASSINDER,S.; COLE, N. Does TGF-beta induced formation of actin stress fibres reinforce Smad dependent TGF-beta signalling in the prostate? *Med Hypotheses* 76: 802-804, 2011.

BARBOZA,R.; COCIORVA,D.; XU,T.; BARBOSA,V. C.; PERALES,J. et al. Can the false-discovery rate be misleading? *Proteomics*. 11:4105-4108, 2011.

BARKER, P. E. Cancer biomarker validation: standards and process: roles for the National Institute of Standards and Technology (NIST). *Ann. NY Acad. Sci.* 983: 142-150, 2003.

BARRIL, N. TAJARA, E L. H. Aspectos Moleculares do Câncer Tiroideano. *Arq Bras Endocrinol Metab.* 43(5): 313-324, 1999.

BARTLETT, J. M. S.; STIRLING, D. PCR protocols – Methods in molecular biology. New Jersey: Humana press, 2<sup>o</sup> edição, vol. 226, 2003.

BHARADWAJ, S.; PRASAD, G. L. Tropomyosin-1, a novel suppressor of cellular transformation is downregulated by promoter methylation in cancer cells. *Cancer Letters*, 183 (2): 205-213, 2002.

BRAVERMAN, R. H.; COOPER, H. L.; LEE, H. S.; PRASAD, G.L. Anti-oncogenic effects of tropomyosin: isoform specificity and importance of protein coding sequences. *Oncogene*, 13 (3):537 – 545, 1996.

BURKE, A. P.; YEN, T. S.; SHEKITKA, K. M.; SOBIN, L. H. Lymphoepithelial carcinoma of the stomach with Epstein-Barr virus demonstrated by polymerase chain reaction. *Modern Pathology.* 3: 377-380, 1990.

BYRNE, J. A.; MALEKI, S.; HARDY, J. R.; GLOSS, B. S.; MURALI, R.; SCURRY, J. P. et al. MAL2 and tumor protein D52 (TPD52) are frequently overexpressed in ovarian carcinoma, but differentially associated with histological subtype and patient outcome. *BMC. Cancer* 10: 497, 2010.

CARVALHO, P. C. et al. Detecção de potenciais marcadores moleculares séricos da doença de Hodgkin. *J. Bras Patol Med Lab.*, 41(3):165-168, 2005.

CARVALHO, P. C.; FISCHER J. S. G.; DEGRAVE, W. M.; CARVALHO, M. G. C. Marcadores séricos e espectrometria de massas no diagnóstico do câncer. *J. Bras Patol Med Lab.* 42(6): 431-436, 2006a.

CARVALHO, P. C. Reconhecimento de padrões proteômicos e genômicos por aprendizagem de máquinas para o diagnóstico médico. *Dissertação (Mestrado)* – Departamento de Biologia Celular e Molecular do Instituto Oswaldo Cruz, Rio de Janeiro, 2006b.

CARVALHO, P. C.; HEWEL, J.; BARBOSA, V. C.; YATES, J. R., III Identifying differences in protein expression levels by spectral counting and feature selection. *Genetics and Molecular Research*, 7 (2): 342-356, 2008.

CARVALHO, P. C.; XU, T.; HAN, X.; COCIORVA, D.; BARBOSA, V. C.; YATES, J. R., III YADA: A tool for taking the most out of high-resolution spectra. *Bioinformatics*. 25:2734-2736, 2009.

CARVALHO, P. C. Um ambiente computacional para proteômica. *Tese (Doutorado)* – Departamento de Engenharia de Sistemas e Computação (COPPE), Rio de Janeiro, 2010.

CARVALHO, P. C.; FISCHER, J. S.; PERALES, J.; YATES, J. R., III; BARBOSA, V. C., BAREINBOIM, E. Analyzing marginal cases in differential shotgun proteomics. *Bioinformatics*, 27:275-276, 2011.

CARVALHO, P. C.; FISCHER, J. S.; XU, T.; COCIORVA, D.; BALBUENA, T. S.; VALENTE, R. H.; PERALES, J.; YATES, J. R., III; BARBOSA, V. C. Search Engine Processor: filtering and organizing PSMs. *Proteomics*, 2012a.

CARVALHO, P. C.; YATES, J. R., III; BARBOSA, V. C. Improving the TFC test for differential shotgun proteomics. *Bioinformatics*, 2012b.

CÉSAR, A. C. G., et al. Fatores genéticos e ambientais envolvidos na carcinogênese gástrica. 39(4): 253-259, 2002.

Chan, A. O. E-cadherin in gastric cancer. *World J. Gastroenterol.* 12: 199-203, 2006.

CHEN, J., et al. Serum 27-nor-5 $\alpha$ -cholestane-3,7,12,24,25 pentol glucuronide discovered by metabolomics as potential diagnostic biomarker for epithelium ovarian cancer. *J. Proteome Res.* 10(5): 2625-2632, 2011.

CHEN, J.; KÄHNE, T.; RÖCKEN, C.; GÖTZE, T.; YU, J. ; SUNG, J. J. et al. Proteome analysis of gastric cancer metastasis by two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-mass spectrometry for identification of metastasis-related proteins. *J Proteome Res.*, 3 (5):1009 – 1016, 2004.

CHEN, Y. et al. RRLC-MS/MS-based metabolomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. *Analyst*, 134: 2003–2011, 2009.

CHENG, Y. Y.; JIN, H.; LIU, X.; SIU, J. M.; WONG, Y. P.; NG, E. K.; YU, J. et al. Fibulin 1 is downregulated through promoter hypermethylation in gastric cancer. *Br J Cancer.* 99(12): 2083-2087, 2008.

COCIORVA,D.; TABB,L.; YATES,J. R. Validation of tandem mass spectrometry database search results using DTASelect. *Curr. Protoc. Bioinformatics.* Chapter 13:13.4.1-13.4.14, 2006.

COTTRELL, J. S. Protein identification using MS/MS data. *Journal of Proteomics XX*, 2011.

CRAIG, R.; BEAVIS, R.C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9): 1466-1467, 2004.

DANTAS, E. L. R. et al. Genética do câncer hereditário. *Revista Brasileira de Cancerologia*, 55(3): 263-269, 2009.

DEBRUYNE, P.; VERMEULEN, S.; MAREEL, M. The role of the E-cadherin/catenin complex in gastrointestinal cancer. *Acta Gastroenterol. Belg.* 62:393-402, 1999.

DIAMANDIST, E. P. Mass Spectrometry as a Diagnostic and a Cancer Biomarker Discovery; *Tool.Molecular & Cellular Proteomic*, 3:367–378, 2004.

EBERT, M. P. A. et al. Identification and confirmation of increased fibrinopeptide A serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. *Journal of proteome research*, 5:2152-2158, 2006.

EIDHAMMER, I.; FLIKKA, K.; MARTENS, L.; MIKALSEN, S. *Computational Methods for Mass Spectrometry Proteomics*. England: John Wiley & Sons Ltda, 2007.

ENG, J. K.; MCCORMACK, A. L.; YATES, J. R. III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom*, 5: 976-989, 1994,

FENN, J. B.; MANN, M.; MENG, C. K.; WONG, S. F.; WHITEHOUSE, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246:64 – 71, 1989.

FERREIRA, C. R. et al. Princípios e aplicações da espectrometria de massas em produção animal. *Anais do II Simpósio de Biologia Molecular Aplicada à Produção Animal*. São Paulo: 109-136, 2009.

FILHO, D. B. Aplicação de métodos quimiométricos para a classificação de águas minerais comercializadas em Goiânia – GO. Dissertação (Mestrado) – Departamento de química, Goiânia, 2010.

FISCHER, J. S.; CARVALHO, P. C.; NEVES-FERREIRA, A. G.; DA FONSECA, C. O.; PERALES, J.; CARVALHO, M. G.; DOMONT, G. B. Anti-thrombin as a prognostic biomarker candidate for patients with recurrent glioblastoma multiform under treatment with perillyl alcohol. *J. Exp. Ther. Oncol.* 7: 285-290, 2008.

FISCHER, J. S. G. Proteoma diferencial do plasma de pacientes com glioblastoma recidivo e de células A172 em resposta ao álcool perílico. *Tese (Doutorado)* – Departamento de bioquímica, Rio de Janeiro, 2010.

FUKAYAMA, M. Epstein-Barr virus and gastric carcinoma. *Pathology International* 60: 337–350, 2010.

GARTNER, L. P; HIATT, J. L., *Tratado de Histologia*. Rio de Janeiro: Guanabara Koogan S.A., 1999.

GLISH, G. L.; VACHET, R. W. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery*, 2:140-150, 2003.

GREENE, L. M.; TWAL, W. O.; DUFFY, M. J.; MCDERMOTT, E. W.; HILL, A. D.; O'HIGGINS, N. J.; MCCANN, A. H. et al. Elevated expression and altered processing of fibulin-1 protein in human breast cancer. *Br J Cancer*. 88(6): 871-878, 2003.

GROSS, J. H. *Mass spectrometry: a textbook*. Heidelberg, Germany: Springer Verlag, 518, 2004.

GUERRA, M. R., GALLO, C. V. DE M., AZEVEDO, G.; MENDONÇA, S., Risco de câncer no Brasil: tendências e estudos epidemiológicos mais recentes, *Rev. Bras. de Cancerologia*, 51(3): 227-234, 2005.

HAN, X.; ASLANIAN, A.; YATES III, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* 12(5):483-490, 2008.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2<sup>o</sup> edição. Springer, 2009.

HAYASHI, K.; CHEN, W. G.; CHEN, Y. Y. et al. Deletion of Epstein-Barr virus latent membrane protein 1 gene in Japanese and Brazilian gastric carcinomas, metastatic lesions, and reactive lymphocytes. *American Journal of Pathology*. 152: 191-198, 1998.

HE, Q. Y.; CHEUNG, Y. H.; LEUNG, S. Y.; YUEN, S. T.; CHU, K. M.; CHIU, J. F. Diverse proteomic alterations in gastric adenocarcinoma. *Proteomics*, 4(10):3276-3287, 2004.

<http://biomedicinabrasil.blogspot.com>. Acesso em: 24 de janeiro de 2012.

<http://www.inca.gov.br/cancernobrasil/2010/docs>. Acesso em: 18 de janeiro de 2012.

<http://www.inca.gov.br/estimativa/2012/>. Acesso em: 20 de janeiro de 2012.

[http://www1.inca.gov.br/conteudo\\_view.asp?id=322](http://www1.inca.gov.br/conteudo_view.asp?id=322). Acesso em: 19 de novembro de 2010.

<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/estomago>.

Acesso em: 19 de novembro de 2010.

HU, J. D. et al. Prediction of gastric cancer metastasis through urinary metabolomic investigation using GC/MS. *World J Gastroenterol*, 17(6): 727-734, 2011.

HUBER, L.A. Is proteomics heading in the wrong direction, *Nature Reviews Molecular Cell Biology*, 4:74-80, 2003.

JIMENO A.; HIDALGO M.; Molecular biomarkers: their increasing role in the diagnosis, characterization, and therapy guidance in pancreatic cancer, *Mol Cancer Ther*, 5(4), 2006.

JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 6<sup>o</sup> edição. New Jersey: Prentice Hall, 2007.

KARAS, M.; BACHMANN, D.; BAHR, U.; HILLENKAMP, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Internacional Journal of Mass Spectrometry and Ion Processes*, 78:53-68,1987.

KASHIMA, K.; OHIKE, N.; MUKAI, S.; SATO, M.; TAKAHASHI, M.; MOROHOSHI, T. Expression of the tumor suppressor gene maspin and its significance in intraductal papillary mucinous neoplasms of the pâncreas. *Hepatobiliary Pancreat Dis Int.*, 7(1): 86-90, 2008.

KITAHARA, F.; KOBAYASHI, K.; SATO, T.; KOJIMA, Y.; ARAKI, T.; FUJINO, M. A. Accuracy of screening for gastric cancer using serum pepsinogen concentrations. *Gut*. 44(5): 693-697, 1999.

KOMOTO, K.; HARUMA, K.; KAMADA, T.; TANAKA, S.; YOSHIHARA, M.; SUMII, K. et al. Helicobacter pylori infection and gastric neoplasia: correlation with histological gastritis and tumor histology. *American Journal of Gastroenterology*. 93: 1271-1276, 1998.

KONNO, S.; TAKEBAYASHI, Y.; AIBA, M.; AKIYAMA, S.; OGAWA, K. Clinicopathological and prognostic significance of thymidine phosphorylase and proliferating cell nuclear antigen in gastric carcinoma. *Cancer Letters*. 166: 103-111, 2001.

KORIYAMA, C.; AKIBA, S.; IRIYA, K. et al. Epstein-Barr virus-associated gastric carcinoma in Japanese Brazilians and non-Japanese Brazilians in São Paulo. *Japanese Journal of Cancer Research*. 92: 911-917, 2001.

KOTLOWSKA, A. et al. Urine metabolomics analysis for adrenal incidentaloma activity detection and biomarker Discovery. *Journal of Chromatography B*, 879: 359-363, 2011.

LATTARIO, F.; FURTADO, Y. L.; GONÇALVES, R.; SILVEIRA, F. A. et al. Analysis of human papillomavirus and Epstein-Barr virus infection and aberrant death-associated protein kinase methylation in high grade squamous intraepithelial lesions. *International Journal of Gynecological Cancer*. 18: 785-789, 2008.

LI, J. et al. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem*, 48(8): 1296-304, 2002.

LIEBLER, D. C. *Introduction to Proteomics: Tools for the new biology*. New Jersey: Humana Press, 2002.

LIMA, M. A. P.; FERREIRA, M. V. P.; BARROS, M. A. P.; PARDINI, M. I. M. C.; FERRASI, A. C.; RABENHORST, S. H. B. Epstein-Barr virus (EBV) detection in gastric carcinomas from Ceará and São Paulo states, in Brazil. *J. Bras. Patol. Med. Lab.* 47(2): 171-179, 2011.

LIMA, M. A. P.; RABENHORST, S. H. B. Association of Epstein-Barr virus (EBV) with solid tumors. *Revista Brasileira de Cancerologia*. 52(1): 87-96, 2006.

LOPES, L. F.; BACCHI, M. M.; OLIVEIRA, D. E.; ZANATI, S. G.; ALVARENGA, M.; BACCHI, C. E. Epstein-Barr virus infection and gastric carcinoma in São Paulo State, Brazil. *Brazilian Journal of Medical and Biological Research*. 37: 1707-1712, 2004.

MACDONALD, F.; FORD, C. H. J.; CASSON, A. G. *Molecular Biology of Cancer*. London e New York: Garland Science/Bios Scientific Publishers, 2<sup>o</sup> edição, 2004.

MacPherson, M.; Fagerholm, S. C. Filamin and filamin-binding proteins in integrin regulation and adhesion. Focus on: Filamin A is required for vimentin-mediated cell adhesion and spreading. *Am J. Physiol. Cell Physiol.* 98(2): 206-208, 2010.

MAKAROV, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, 72:1156 – 1162, 2000.

MANN, M. et al., Mass spectrometry-based proteomics. *Nature*, 422: 198-207, 2003.

MATHOT, L.; STENNINGER, J. Behavior of seeds and soil in the mechanism of metastasis: a deeper understanding. *Cancer Sci.* 103(4): 626-631, 2012.

MAUAD, E.C. et al., Câncer gástrico em adultos jovens. *Revista Brasileira de Cancerologia*; 46(3), 299-304, 2000.

MCDONALD,W. H.; TABB,D. L.;SADYGOV,R. G.; MACCOSS,M. J.; VENABLE,J.; GRAUMANN,J.; JOHNSON,J. R.;COCIORVA,D.;YATES,J. R., III MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun. Mass Spectrom.* 18:2162-2168, 2004.

MCNEER, G. Cancer of the stomach in the young. *AJR*; 45: 537-550, 1941.

MOTOYAMA, A.; YATES, J. R. III. Multi-dimensional LC separations in shotgun proteomics. *Analytical Chemistry*, 80(19): 7187-7193, 2008.

NEJJARI, M.; ANDERSON, W.; POURREYRON, C.; JACQUIER, M. F.; SCOAZEC, J. Y.; REMY, L. The role of fibroblasts in the modulation of integrin-dependent interactions between the gastric cell line HGT-1 and fibronectin. *Int. J. Cancer* 112: 560-569, 2004.

NEVEROVA, I.; EYK, J. E. V. Role of chromatographic techniques in proteomic analysis. *Journal of Chromatography B*, 815: 51 – 63, 2005.

NUSSBAUM, R. L.; Willard, H. F.; Mcinnes, R. R. Thompson e Thompson Genética Médica. Rio de Janeiro: Elsevier, 7<sup>o</sup> edição, 2008.

OISHI, Y.; KIYOHARA, Y.; KUBO, M.; TANAKA, K.; TANIZAKI, Y.; NINOMIYA, T. et al. The serum pepsinogen test as a predictor of gastric cancer: the Hisayama study. *Am J Epidemiol*. 163(7): 629-637. 2006.

ORIMO, H.; OHATA, M.; FUJITA, T. Role of inorganic pyrophosphatase in the mechanism of action of parathyroid hormone and calcitonin. *Endocrinology*, 89 (3): 852–858, 1971.

OUE, N.; HAMAI, Y.; MITANI, Y.; MATSUMURA, S.; OSHIMO, Y.; AUNG, P. P.; KURAOKA, K.; NAKAYAMA, H.; YASUI, W. Gene expression profile of gastric carcinoma: identification of genes and tags potentially involved in invasion, metastasis, and carcinogenesis by serial analysis of gene expression. *Cancer Res*. 64: 2397-2405, 2004.

PARKIN, D. M. et al. Cancer incidence in five continents, *IARC Scientific Publications*, 143: 45-173, 1997.

PATTERSON, S. D.; AEBERSOLD, R. H. Proteomics: the first decade and beyond. *Nature Genetics*, 33: 311-323, 2003.

PATTLE, S. B.; FARRELL, P. J. The role of Epstein-Barr virus in cancer. *Expert Opin Biol Ther*. 6(11):1193-1205, 2006.

PEPE, M. S. et al. Phases of biomarker development for early detection of cancer. *J. Nat. Cancer Inst*. 93: 1054-1061, 2001.

PERKINS, D.N.; PAPPIN, D.J.; CREASY, D. M.; COTTRELL, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18): 3551-3567, 1999.

PESTANER, C., BIBBO, M., BOBROSKI, L., SESHAMMA, T. et al. Potential of the in situ Polymerase Chain Reaction in diagnostic cytology. *Acta Cytological*, 38: 676-680, 1994.

PETRICOIN, E. F. et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 359(9306): 572-577, 2002b.

PETRICOIN, E. F. III et al., Serum proteomic patterns for detection of prostate cancer. *Journal of the national cancer institute*, 94(20): 1576-1578, 2002a.

R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2009 (<http://www.R-project.org>).

RAVID, D.; CHUDERLAND, D.; LANDSMAN, L.; LAVIE, Y.; REICH, R.; LISCOVITCH, M. Filamin A is a novel caveolin-1-dependent target in IGF-I-stimulated cancer cell migration. *Experimental cell Research*, 314: 2762 – 2773, 2008.

RIOLI, V. et al. Novel Natural Peptide Substrates for Endopeptidase 24.15, Neurolysin, and Angiotensin-converting Enzyme. *The journal of Biological Chemistry*, 278(10): 8547-8555, 2003.

ROSENBAUER, V., RIESNER, D. Temperature-gradient gel electrophoresis-thermodynamic analysis of nucleic acids and proteins in purified form and extracellular extracts. *Biophys Chem*. 26(2– 3):235–246, 1987.

SAGER, R.; SHENG, S., PEMBERTON, P.; HENDRIX, M. J. Maspin - A tumor suppressing serpin. *Adv. Exp. Med. Biol*. 425: 77–88, 1997.

SAIKI, R., SCHARF, S., FALOONA, F., MULLIS, K., HORN, G., AND ERLICH, H. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-54, 1985.

SAITO, I., SERVENIUS, B., COMPTON, T., FOX, R.I. Detection of Epstein-Barr virus DNA by polymerase chain reaction in blood and tissue biopsies from patients with Sjogren's syndrome. *Journal of Experimental Medicine*. 169: 2191-219, 1989.

SAMPLE, J. T.; SAMPLE, C. E. Epstein-Barr virus: Molecular Biology. *Encyclopedia of Virology*, Oxford: Elsevier Ltd., 3<sup>o</sup> edição: 157-167, 2008.

SANTOS, R. B.; JÚNIOR, V. L.; CASTRO, V. O. B.; GRECO, S. J. Uso de métodos quimiométricos e mecânico-quanticos na análise de terpenóides e fenilpropanóides bioativos contra o *Aedes aegypti*. *Orbital*. 1(4): 273-288, 2009.

SCANNI, A.; TOMIROTTI, M.; LICCIARDELLO, L.; ANNIBALI, E.; BIRAGHI, M.; TROVATO, M.; FITTIPALDI, M.; ADAMOLI, P.; CURTARELLI, G. Variations in serum copper and ceruloplasmin levels in advanced gastrointestinal cancer treated with polychemotherapy. *Tumori* 65: 331-338, 1979.

SCHULZ, W. A. *Molecular Biology of Human Cancers – An advanced student's textbook*. Springer, Dordrecht, The Netherlands, 2005.

SHAO, L. J.; SHI, H. Y.; AYALA, G.; ROWLEY, D.; ZHANG, M. Haploinsufficiency of the maspin tumor suppressor gene leads to hyperplastic lesions in prostate. *Cancer Res*. 68 (13): 5143–51, 2008.

SHIBATA, D.; TOKUNAGA, M.; UEMURA, Y.; SATO, E.; TANAKA, S.; WEISS, L. M. Association of Epstein-Barr virus with undifferentiated gastric carcinoma with intense lymphoid infiltration lymphoepitheliomalike carcinoma. *American Journal of Pathology*. 139: 469-474, 1991.

- SILVA, A. R. e ZUCOLOTO, S. O. The role of Epstein-Barr virus in human tumorigenesis. *Medicina (Ribeirão Preto)*. 36: 16-23, 2003.
- SILVEIRA, A. S. Câncer ginecológico: Diagnóstico e tratamento. In: Gil RA. *Fatores prognósticos, preditivos e marcadores tumorais no câncer ginecológico*. Florianópolis: 135-152, 2005.
- SKUBITZ, A. P. Adhesion molecules. *Cancer Treat. Res.* 107:305-329, 2002.
- SMITH, P. K.; KROHN, R. I.; HERMANSON, G.T. et al. Measurement of protein using bicinchoninic acid. *Anal. Biochem.*, 150: 76-85,1985.
- SPRATLIN, J. L.; SERKOVA, N. J. and ECKHARDT, S. G. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res.*, 15: 431 – 440, 2009.
- STORR,S. J.; MOHAMMED, R. A.; WOOLSTON, C. M.; GREEN, A. R.; PARR, T.; SPITERI, I.; CALDAS, C.; BALL, G. R.; ELLIS, I. O.; MARTIN, S. G. Calpastatin is associated with lymphovascular invasion in breast cancer. *Breast* 20: 413-418, 2011.
- THEUER, C.P. et al. Gastric adenocarcinoma in patients 40 years of age or younger. *Am J Surg*: 172(5), 473-477,1996.
- TWAL, W.O.; CZIROK, A.; HEGEDUS, B.; KNAAK, C.; CHINTALAPUDI, M. R.; OKAGAWA, H.; SUGI, Y.; ARGRAVES, W. S. Fibulin-1 suppression of fibronectin-regulated cell adhesion and motility. *J Cell Sci.* 114(24): 4587-4598, 2001.
- USUI, Y.; UEMATSU, T.; UCHIHASHI, T. et al. Inorganic polyphosphate induces osteoblastic differentiation *J. Dent. Res.*, 89 (5): 504–509, 2010.
- VEENSTRA, T. D.; YATES III, J. R. *Proteomics for Biological Discovery*. New Jersey: John Wiley & Sons, 2006.

WANG, R. ; XU, J.; SARAMÄKI, O.; VISAKORPI, T.; SUTHERLAND, W. M.; ZHOU, J. et al. PrLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in human prostate cancer. *Cancer Res.* 64(5): 1589-94, 2004.

WASHBURN, M. P.; WOLTERS, D.; YATES, J. R. III. Large – scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19: 242-247, 2001.

WLAZLINSKI, A.; ENGERS, R.; HOFFMANN, M. J.; HADER, C.; JUNG, V.; MÜLLER, M.; SCHULZ, W. A. Downregulation of several fibulin genes in prostate cancer. *Prostate.* 67(16): 1770-80, 2007.

WU, H. et al. Metabolomic investigation of gastric cancer tissue using gas chromatography/mass spectrometry. *Anal Bioanal Chem*, 396:1385–1395, 2010.

XU, T. et al. ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. *Mol Cell Proteomics*, 5, 2006.

XU, Y.; BISMAR, T. A.; SU, J.; XU, B.; KRISTIANSEN, G.; VARGA, Z.; TENG, L. et al. Filamin A regulates focal adhesion disassembly and suppresses breast cancer cell migration and invasion. *J Exp Med.* 207(11): 2421-2437, 2010.

YASUI, W.; OUE, N.; ITO, R.; KURAOKA, K.; NAKAYAMA, H. Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications. *Cancer Sci.* 95: 385-392, 2004.

YASUI, Y. et al. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, 4(3): 449-63, 2003.

YATES, J. R., III; PARK, S. K.; DELAHUNTY, C. M.; XU, T.; SAVAS, J. N.; COCIORVA, D.; CARVALHO, P. C. Toward objective evaluation of proteomic algorithms. *Nat. Methods* 9:455-456, 2012.

YATES, J.; RUSE, C. I.; NAKORCHEVSKY, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Reviews Biomed. Eng.*, 11:49-79, 2009.

ZHANG, X.; FANG, A.; RILEY, C. P.; WANG, M.; REGNIER, F. E.; BUCK, C. Multi-dimensional liquid chromatography in proteomics – a review. *Analytica Chimica Acta*, 664:101 – 113, 2010.

ZHANG, Y.; WEN, Z.; WASHBURN, M. P.; FLORENS, L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* 82: 2272-2281, 2010.

ZHAO, C. H. et al. Five serum proteins identified using SELDI-TOF-MS as potential biomarkers of gastric cancer. *Jpn J Clin Oncol*, 40(4); 336–342, 2010.

ZHAO, Y.; ZHOU, T.; LI, A.; YAO, H.; HE, F.; WANG, L.; SI, J. A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. *Anat. Rec. (Hoboken.)* 292:692-700, 2009.

ZHU, W. Y.; HUNAG, Y. Y.; LIU, X. G.; HE, J. Y.; CHEN, D. D.; ZENG, F.; ZHOU, J. H.; ZHANG, Y. K. Prognostic evaluation of CapG, gelsolin, P-gp, GSTP1, and Topo-II proteins in non-small cell lung cancer. *Anat Rec (Hoboken)* 295: 208-214, 2012.

## 7. ANEXOS

### 7.1 Parecer do Comitê de ética em pesquisa - UFAM



UNIVERSIDADE FEDERAL DO AMAZONAS  
COMITÊ DE ÉTICA EM PESQUISA – CEP/UFAM



## PARECER DO COMITÊ DE ÉTICA EM PESQUISA

O Comitê de Ética em Pesquisa da Universidade Federal do Amazonas aprovou, em reunião ordinária realizada nesta data, por unanimidade de votos, o Projeto de Pesquisa protocolado no CEP/UFAM com CAAE nº. 0057.0.115.000-11, intitulado: **“PROSPECÇÃO DE CANDIDATOS A BIOMARCADORES ONCOLÓGICOS DE CÂNCERES GÁSTRICOS POR ESPECTROMETRIA DE MASSAS”**, tendo como Pesquisadora Responsável Priscila Ferreira de Aquino.

Sala de Reunião da Escola de Enfermagem de Manaus – EEM da Universidade Federal do Amazonas, em Manaus/Amazonas, 27 de abril 2011.

Prof. MSc. Plínio José Cavalcante Monteiro  
Coordenador CEP/UFAM

Escola de Enfermagem de Manaus – EEM/UFAM

Rua Teresina, 4950 – Adrianópolis – CEP: 69057-070 – Manaus-AM – Fone: (92) 3305-5130 – E-mail: cep@ufam.edu.br

## 7.2 Parecer do Comitê de ética em pesquisa - FCECON



### AUTORIZAÇÃO PARA REALIZAÇÃO DE PESQUISA

Declaramos para os devidos fins que **autorizamos a realização** do estudo “Prospecção de candidatos a biomarcadores oncológicos de cânceres gástricos por espectrometria de massas”. O estudo tem como responsáveis Priscila Ferreira de Aquino Afonso Duarte Leão de Souza, Paulo Carvalho, Sidney Chalub além de outros professores da Universidade Estadual do Amazonas e Universidade Federal do Amazonas.

O estudo tem como objetivo geral objetivo geral é “Contribuir para o conhecimento do uso da espectrometria de massas para a detecção de candidatos a marcadores proteômicos e metabólicos em amostras de tecido de pacientes com câncer gástrico” e objetivos específicos: 1- Desenvolver metodologias e realizar análises de perfis proteômicos e metabólicos de cânceres gástricos por espectrometria de massas para a busca de candidatos a biomarcadores; 2- Comparar os perfis proteômicos e metabólicos de pacientes controles com os de pacientes com câncer gástrico; 3- Selecionar candidatos a marcadores proteômicos específicos para o tipo de câncer em estudo; 4- Selecionar candidatos a marcadores metabólicos de cânceres gástricos; 5- Avaliar estatisticamente a consistência dos resultados dos perfis proteômicos e metabólicos com os dados da avaliação diagnóstica (biópsia) dos pacientes com câncer gástrico e com a condição de cada candidato envolvido no estudo: saudável ou com câncer.

O referido protocolo de pesquisa foi apreciado pelo Comitê Científico da Fundação CECON e após os ajustes solicitados, **obteve aprovação**. Além disso, apresenta a anuência do responsável pelo serviço de Cirurgia do Aparelho Digestivo da Fundação CECON.

Considerando que o referido protocolo já foi aprovado pelo CEP da UFAM, sugerimos que os ajustes realizados sejam reapresentados ao mesmo com os adendos e alterações realizadas.

O estudo trará grandes contribuições científicas e irá promover o fortalecimento dos grupos de pesquisa da Fundação CECON, tendo em vista que um dos pesquisadores responsáveis faz parte da equipe clínica do serviço de cirurgia gástrica e também dos grupos de pesquisa da Fundação CECON.

Manaus, 15 de Junho de 2011

A handwritten signature in blue ink, appearing to read 'Kátia'.

Kátia Luz Torres Silva  
Diretora de Ensino e Pesquisa  
Fundação CECON.

### 7.3 Termo de Consentimento Livre e Esclarecido (TCLE)

UNIVERSIDADE FEDERAL DO AMAZONAS  
DEPARTAMENTO DE QUÍMICA



#### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

Convidamos o (a) Sr (a). para participar de um Projeto de Pesquisa com o título “**PROSPECÇÃO DE BIOMARCADORES ONCOLÓGICOS DE CÂNCERES GÁSTRICOS POR ESPECTROMETRIA DE MASSAS**”. Este projeto a ser realizado na Universidade Federal do Amazonas (UFAM) em parcerias com outras instituições, visa ajudar no desenvolvimento de metodologias novas para a detecção precoce do câncer, o que aumenta a chance de curar esta doença. Serão feitos estudos com o objetivo de identificar substâncias que permitam diferenciar a população com câncer gástrico dos pacientes sadios. O(s) pesquisador (es) Priscila Ferreira de Aquino, Afonso Duarte Leão de Souza e Paulo Costa Carvalho, responsáveis pelo projeto, pedem autorização para retirar uma amostra do tecido do seu estômago que será utilizada apenas para esta pesquisa. Caso necessário, esta amostra de tecido também poderá ser enviada ao Rio de Janeiro onde parte da pesquisa será feita. Neste caso, o transporte do material biológico seguirá as normas institucionais previamente estabelecidas. As análises em busca das substâncias referidas acima ocorrerão em aparelhos apropriados para este tipo de pesquisa, os quais são chamados de cromatógrafos e espectrômetros de massas de alta e baixa resolução. Os dados obtidos com esses aparelhos serão tratados por técnicas de análises estatísticas e de informática para a identificação dos metabólitos e/ou proteínas relacionadas ao câncer.

Se você autorizar esta coleta, seu tecido será utilizado nas análises. O que não for usado não será guardado e sim descartado com o cuidado adequado para materiais biológicos humanos. A coleta do tecido será durante o ato operatório ( ) ou exame endoscópico ( ), não representando nenhuma dor ou qualquer outro sintoma a você.

Se depois de autorizar a coleta, o Sr(a) não quiser que seu tecido seja usado, tem o direito e a liberdade de retirar seu consentimento em qualquer

fase da pesquisa, seja antes ou depois da coleta do tecido, independente do motivo e sem prejuízo do atendimento que está recebendo. O(a) Sr(a) não terá nenhuma despesa e também não receberá nenhum pagamento. A sua participação voluntária é importante para o melhor conhecimento desta doença (câncer) e de como pode ser mais bem diagnosticada antes que comprometa a recuperação dos pacientes.

Os resultados da pesquisa serão analisados e publicados, mas sua identidade não será divulgada, sendo guardada em segredo. Para qualquer outra informação, o(a) Sr.(a) poderá entrar em contato com os pesquisadores pelo telefones (92) 8803-9310 e 8123-1079 ou pelos e-mails: **priscila\_aquino23@hotmail.com**, **souzadq@ufam.edu.br** e **paulo@pcarvalho.com**.

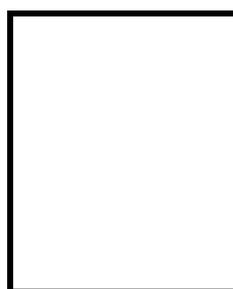
#### Consentimento Pós-Informação

Eu, \_\_\_\_\_,  
fui informado(a) sobre o que os pesquisadores querem fazer e porque precisa da minha colaboração, e entendi a explicação. Por isso, eu concordo em participar do projeto, sabendo que não vou ganhar nada e que posso sair quando quiser. Estou recebendo uma cópia deste documento, assinada, que vou guardar.

\_\_\_\_\_  
Assinatura do participante

ou

\_\_\_\_\_  
Pesquisadora Responsável



Impressão do dedo polegar

Caso não saiba assinar

\_\_\_\_ - \_\_\_\_ - \_\_\_\_  
Data

\_\_\_\_ - \_\_\_\_ - \_\_\_\_  
Data

#### 7.4 Questionário

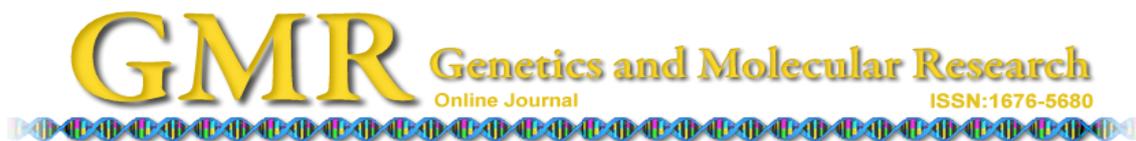
- Nome:
- Idade:
- Sexo:
- Histórico de doença na família:
- Caso sim, qual a doença e relação de parentesco?
  
- Possui alguma outra doença associada:
- Caso sim, qual?
- Está tomando medicamentos?
- Caso sim, qual?
- Passou por algum tratamento ou intervenção cirúrgica recentemente?
- Caso sim, qual?
- Possui dispepsia gástrica?
- Estágio clínico do câncer:
- Estágio cirúrgico do câncer:
- Subtipo histológico:

---

Assinatura - Paciente.

Manaus, de 2011.

## 7.5 Artigo aceito para a publicação na *Genetics and Molecular Research*



### **Epstein-Barr virus DNA associated with gastric adenocarcinoma and adjacent non-cancerous mucosa in patients from Manaus, Brazil**

**Running title: Evaluation of EBV DNA in gastric cancer**

P.F. de Aquino<sup>1</sup>, P.C. Carvalho<sup>2</sup>, J.S. da Gama Fischer<sup>2</sup>, A.Q.L. de Souza<sup>3</sup>, J.S. Viana<sup>3</sup>, S.R.S. Chalub<sup>4</sup>, A.D.L. de Souza<sup>1</sup> and M.G.C. Carvalho<sup>5\*</sup>

1 – Departamento de Química, Universidade Federal do Amazonas, Amazonas, Brasil

2 – Instituto Carlos Chagas, Fiocruz, Paraná, Brazil

3 – Escola Superior de Ciências da Saúde, Universidade do Estado do Amazonas, Amazonas, Brasil

4 – Departamento de Cirurgia Digestiva, Universidade Federal do Amazonas, Amazonas, Brazil

5 – Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

\*Corresponding Author:

Maria da Gloria da Costa Carvalho, MD, PhD  
Laboratório de Controle da Expressão Gênica  
Instituto de Biofísica Carlos Chagas Filho, UFRJ  
Centro de Ciência da Saúde, (CCS)  
Bloco G, Ilha do Fundão  
Rio de Janeiro, RJ, CEP 21949-900, Brazil  
Phone: 55 – 21 2562-6511  
Email: gloria@gcarvalho.org.

### **Abstract**

Gastric cancer is one of most frequent causes of death in Brazil. The city of Manaus has one of the highest incidences of this disease in Brazil. The Epstein-Barr virus (EBV) is a ubiquitous herpesvirus that is classified as a group 1 carcinogen by the International Agency for Research on Cancer. We obtained biopsies from six control subjects and 10 patients with gastric carcinomas living in Manaus. In the patients, the samples were taken from tumors and from

adjacent non-cancerous mucosa. These samples were screened for EBV DNA by PCR to amplify the 288 bp fragments from the Bam M region. The EBV DNA was detected in 8/10 of the tumor cases and in none of the six control subjects. In the positively identified samples, EBV DNA was detected in five corresponding resection margins. Previous research indicated only a weak association between EBV and gastric cancer. We suggest that EBV should be considered as a risk factor for gastric adenocarcinomas in Manaus.

**Keywords:** Epstein-Barr; Gastric cancer; Amazon; Resection margin; PCR.

## INTRODUCTION

Gastric cancer (GC) is one of the most common cancers, and is ranked 2nd among the worldwide cancer mortality rates (Ebert et al., 2006). Several factors may contribute individually or together to the emergence of this neoplasia, including infection with *Helicobacter pylori* (*H. pylori*), dietary habits, genetic factors, and viral infections (Schulz, 2005). GC is mostly found in individuals aged over 50 years; its incidence in individuals under 40 corresponds to only 5% of all cases (Mauad, 2000; Theuer et al., 1996). Most patients are diagnosed with GC in an advanced state, thus drastically reducing the treatment options, survival, and cure (Ebert et al., 2006). In the northern part of Brazil, GC is the 2nd most common cancer in men and the 4th in women, accounting for an estimated 13 and 7 new cases per 100,000 per year for men and women, respectively (Guerra et al., 2005; INCA, 2012). While most GCs are correlated with *H. pylori* infection (Asaka et al., 1994; Komoto et al., 1998), such is not true for cases in the Brazilian state of Amazonas, where most GC

patients are *H. pylori* negative. These facts motivated our group to investigate other risk factors to help explain this high incidence.

The incidence of the Epstein-Barr virus (EBV) with GC has been found to vary from 2 to 18% in the world population. A recent Brazilian report described an incidence of ~ 8% among GCs from Ceará (Lima et al., 2011). Thus, we considered investigating whether EBV could be correlated with GC in the Amazon.

EBV is classified as a group 1 carcinogen according to the International Agency for Research on Cancer (IARC). It belongs to the gamma-herpes family, which has a linear DNA molecule responsible for encoding approximately 100 proteins (Lima and Rabenhorst, 2006; Silva and Zucoloto, 2003). Reports describe more than 90% of our population as being previously exposed to EBV, which can be easily transmitted by salivary contact (Macswen and Crawford, 2003). This virus has 2 distinct life cycles in the human host: a lytic cycle, during which the production of new virions occurs; and a latent form, wherein it remains in a dormant state for the lifetime of the host (Pattle and Farrell, 2006). EBV is mostly well known as the causative agent for mononucleosis (IM), which is usually manifested in adulthood through latent EBV genes (Asaka et al., 1994). In addition to GC, EBV has been associated with malignancies including Burkitt's lymphoma, Hodgkin's disease, peripheral T-cell lymphoma, thymoma, and nasopharyngeal carcinoma (NPC) (Lima and Rabenhorst, 2006). The role of EBV in tumorigenesis is not well understood, but there is evidence correlating its persistence in human epithelial cells with tumorigenesis (Tsao et al., 2012).

## MATERIAL AND METHODS

### Subjects

This study was approved by the ethics committee review board from the Federal University of the Amazon (CEP/UFAM: MEMO - n° 0057.0.115.000-11-CAAE). After signing informed consent, tumor specimens with resection margins were obtained after surgery from 10 GC patients, 4 females. Six biopsies were obtained from control subjects during upper endoscopy; 5 were from females. Subjects were classified as controls if no traces of cancer were detected by endoscopic evaluation. All biopsies were obtained from the stomach. Each biopsy was then subtyped and the clinical stage of the disease was determined according to the Tumor, Node, and Metastasis classification (TNM) from the American Joint Committee on Cancer (AJCC).

### **DNA extraction**

For the PCR analysis, the biopsies were pulverized and genomic DNA was isolated by digestion in 500  $\mu$ L containing 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2% SDS, 10 mM EDTA, pH 8.0, and 15  $\mu$ L proteinase K (10 mg/mL). The material was incubated at 56°C for 2 h; DNA was then extracted by phenol-chloroform-isoamyl alcohol and resuspended in sterile distilled water. The extracted genomic DNA was submitted to polymerase chain reaction (PCR) to confirm DNA integrity using p53 primers to exon 5, generating a 274 bp product (Pestaner et al., 1994).

### **Polymerase Chain Reaction (PCR) for detecting EBV DNA in the biopsies**

The detection for EBV DNA used the primers TC67 and TC69, which amplify a portion of the Bam M region of EBV DNA and produce a 288 bp product (Saito

et al., 1989). Amplification conditions consisted of an initial denaturing step at 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 60°C for 1 min, and 72°C for 1 min. This was followed by a final extension step at 72°C for 5 min. For all reactions we used positive (*viz.*, DNA from Raji cells) and negative (*viz.*, Milli-Q water) controls (Lattario et al., 2008).

### Diagnosis of EBV DNA

Diagnosis was accomplished by visually inspecting for the presence of 288 bp PCR products using a 10% polyacrylamide gel stained with silver nitrate (Rosenbauer and Riesner, 1987). The approximate size of the amplified fragments was calculated using co-electrophoresis with LMW (Low Molecular Weight; BioRad).

## RESULTS

The results from our analysis for the presence of *H. pylori* and EBV DNA are presented in **Table 1**. Briefly, EBV DNA was detected in 80% of the biopsies, among which 50% had positive resection margins. EBV DNA was not found in any control subjects.

**Table 1** - Clinical and experimental features for the presence of EBV from subjects enrolled in the study. For the “Type” column, Cont, RM, and Can stand for Control sample, Resection Margin, and Cancer, respectively. In the “Sex” column, M and F stand for male and female, respectively. The presence or not of *H. pylori* and EBV are indicated as + and -, and NA (i.e., not applicable), respectively.

N <sup>0</sup>	Type	Sex	Age	Diagnosis	H. <i>pylori</i>	EBV DNA Can	EBV DNA RM
1	Cont	F	46	Gastritis	-	-	NA
2	Cont	F	26	Gastritis	-	-	NA
3	Cont	F	52	Gastritis	-	-	NA
4	Cont	M	43	Gastritis	-	-	NA
5	Cont	F	55	Gastritis	-	-	NA
6	Cont	F	62	Gastritis	-	-	NA
7	Can	F	49	T4N3M0	-	+	+
8	Can	F	71	T3NXM0	-	+	+

<b>9</b>	Can	M	48	T4NxM0	-	-	-
<b>10</b>	Can	F	46	T4N0M0	-	-	-
<b>11</b>	Can	F	46	T4NxN0	+	-	+
<b>12</b>	Can	M	47	T3N3M0	+	+	-
<b>13</b>	Can	M	68	T3NxM0	-	+	+
<b>14</b>	Can	M	50	T4NxM1	-	+	+
<b>15</b>	Can	M	40	T1NxM0	-	-	+
<b>16</b>	Can	M	62	T4N2M1	-	+	+

## DISCUSSION

Previous studies have described *H. pylori* as the main causative agent for GC (Asaka et al., 1994; Komoto et al., 1998) and report a low correlation between EBV and GC (Lima and Rabenhorst, 2006; Lima et al., 2011). Here, EBV DNA was detected in 80% of the tumors examined using PCR, and 50% presented EBV DNA in both the gastric cancer specimen and the resection margin. EBV was detected in both early and advanced stages of GC, as aligned with previous reports (Shibata et al., 1991; Sample and Sample, 2008). It is also important to emphasize that *H. pylori* was detected in only 2 cases. As such, although EBV's mechanisms for triggering carcinogenesis are not well understood (Nishikawa et al., 1999), our results suggest that EBV should be considered as a risk factor for this disease, especially in the Amazon state of Brazil.

An increased sample size will be necessary to accomplish a more robust statistics. Regardless, we have raised an important hypothesis of a high correlation between EBV and GC in the Amazon that deserves further investigation by the scientific community in different and larger cohorts.

## ACKNOWLEDGMENTS

The authors thank the medical team of the CECON foundation for assisting with the samples and CAPES, CAPES/Fiocruz 2006/2, CNPq, Fundação do Cancer, and FAPEAM for financial support.

## REFERENCES

Asaka M, Kimura T, Kato M, Kudo M, Miki K, Ogoshi K et al. (1994). Possible role of *Helicobacter pylori* infection in early gastric cancer development. *Cancer*. 73: 2691-2694.

Ebert MP, Niemeyer D, Deininger SO, Wex T, Knipping C, Hoffmann J et al. (2006). Identification and confirmation of increased fibrinopeptide A serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. *Journal of proteome research*. 5: 2152-2158.

Guerra MR, Gallo CVM, Azevedo G and Mendonça S (2005). Risco de câncer no Brasil: tendências e estudos epidemiológicos mais recentes, *Rev. Bras. de Cancerologia*. 51(3): 227-234.

Komoto K, Haruma K, Kamada T, Tanaka S, Yoshihara M, Sumii K et al. (1998). *Helicobacter pylori* infection and gastric neoplasia: correlation with histological gastritis and tumor histology. *American Journal of Gastroenterology*. 93: 1271-1276.

Lattario F, Furtado YL, Gonçalves R, Silveira FA, do Val IC, Almeida G et al. (2008). Analysis of human papillomavirus and Epstein-Barr virus infection and aberrant death-associated protein kinase methylation in high grade squamous intraepithelial lesions. *International Journal of Gynecological Cancer*. 18: 785-789.

Lima, MAP and Rabenhorst, SHB (2006). Association of Epstein-Barr virus (EBV) with solid tumors. *Revista Brasileira de Cancerologia*. 52(1): 87-96.

Lima MAP, Ferreira MVP, Barros MAP, Pardini MIMC, Ferrasi AC and Rabenhorst SHB. (2011). Epstein-Barr virus (EBV) detection in gastric carcinomas from Ceará and São Paulo states, in Brazil. *J. Bras. Patol. Med. Lab.* 47(2): 171-179.

Macswen KF and Crawford DH (2003). Epstein-Barr virus – recent advances. *The Lancet infectious diseases*. 3(3): 131-140.

Mauad, EC, Nogueira JL, Souza JMP, Wohnrath DR, Oliveira ATT, Colli G et al. (2000). Câncer gástrico em adultos jovens. *Revista Brasileira de Cancerologia*. 46(3): 299-304.

Ministério da saúde, Instituto Nacional do Câncer/*Estimativa 2012: incidência de câncer no Brasil*. Available: [<http://www.inca.gov.br/estimativa/2012/estimativa20122111.pdf>]. Accessed: January 20, 2012.

Nishikawa J, Jmai S, Oda T, Kojima T, Okita K and Takada K (1999). Epstein-Barr virus promotes epithelial cell growth in the absence of EBNAQ and LMP1 expression. *Journal of Virology*. 73: 1286-1292.

Pattle SB and Farrell PJ (2006). The role of Epstein-Barr virus in cancer. *Expert Opin Biol Ther*. 6(11): 1193-1205.

Pestaner C, Bibbo M, Bobroski L, Seshamma T and Bagasra O (1994). Potential of the in situ polymerase chain reaction in diagnostic cytology. *Acta Cytological*. 38: 676-680.

Rosenbauer V and Riesner D (1987). Temperature-gradient gel electrophoresis- thermodynamic analysis of nucleic acids and proteins in purified form and extracellular extracts. *Biophys Chem.* 26(2–3): 235-246.

Saito I, Serenius B, Compton T, Fox RI (1989). Detection of Epstein-Barr virus DNA by polymerase chain reaction in blood and tissue biopsies from patients with Sjogren's syndrome. *Journal of Experimental Medicine.* 169: 2191-2198.

Sample, JT and Sample, CE (2008). Epstein-Barr virus: Molecular Biology. *Encyclopedia of Virology*, Third edition, Elsevier Ltd., Oxford, 157-167.

Schulz WA (2005). *Molecular Biology of Human Cancers – An advanced student's textbook.* Springer, Dordrecht, The Netherlands.

Shibata D, Tokunaga M, Uemura Y, Sato E, Tanaka S, Weiss LM (1991). Association of Epstein-Barr virus with undifferentiated gastric carcinoma with intense lymphoid infiltration lymphoepitheliomalike carcinoma. *American Journal of Pathology.* 139: 469-474.

Silva, AR and Zucoloto, SO (2003). The role of Epstein-Barr virus in human tumorigenesis. *Medicina (Ribeirão Preto).* 36: 16-23.

Theuer, CP, de Virgilio C, Keese G, French S, Arnell T, Tolmos J et al., (1996). Gastric adenocarcinoma in patients 40 years of age or younger. *Am J Surg.* 172(5): 473-477.

Tsao SW, Tsang CM, Pang PS, Zhang G, Chen H, Lo KW (2012). The biology of EBV infection in human epithelial cells. *Seminars in cancer Biology.* 22: 137-143.

## 7.6 Artigo submetido ao *Journal of Proteome Research*

Page 1 of 28

Journal of Proteome Research

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Are gastric cancer resection margin proteomic profiles more similar to those from controls or tumors?

Priscila F Aquino<sup>1&</sup>; Juliana S G Fischer<sup>2&</sup>, Ana G C Neves-Ferreira<sup>3</sup>, Jonas Perales<sup>3</sup>, Gilberto B Domont<sup>4</sup>, Gabriel D T Araujo<sup>4</sup>, Valmir C Barbosa<sup>5</sup>, Jucilana Viana<sup>6</sup>, Sidney R S Chalub<sup>7</sup>, Antonia Q Lima de Souza<sup>6</sup>, Maria G C Carvalho<sup>8</sup>, Afonso D Leão de Souza<sup>1</sup>, Paulo C Carvalho<sup>2\*</sup>.

1. Departamento de Química, Universidade Federal do Amazonas, Amazonas, Brazil.
2. Instituto Carlos Chagas, Fiocruz, Paraná, Brazil.
3. Laboratório de Toxinologia, Instituto Oswaldo Cruz, Rio de Janeiro, Brazil.
4. Proteomics Unit, Rio de Janeiro Proteomics Network, Departamento de Bioquímica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
5. Programa de Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
6. Escola Superior de Ciências da Saúde, Universidade do Estado do Amazonas, Amazonas, Brazil.
7. Departamento de Cirurgia Digestiva, Universidade Federal do Amazonas, Amazonas, Brazil.
8. Departamento de Patologia, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

& Equal contributions

\*Corresponding Author

Rua Prof. Algacyr Munhoz Mader, 3775  
CIC 81350-010 Curitiba/PR, Brazil  
Tel.: +55(41)3316-3230  
Fax: +55(41)3316-3267

1  
2  
3 **Abstract:** A strategy for treating cancer is to surgically remove the tumor  
4 together with a portion of apparently healthy tissue surrounding it, the so-called  
5 "resection margin", to minimize recurrence. Here we investigate whether the  
6 proteomic profiles from biopsies of gastric cancer resection margins are indeed  
7 more similar to those from healthy tissue than from cancer biopsies. To this  
8 end, we analyzed biopsies using an offline MudPIT shotgun proteomic  
9 approach and performed label-free quantitation through a distributed normalized  
10 spectral abundance factor approach adapted for extracted ion chromatograms  
11 (XICs). A multidimensional scaling analysis revealed that each of those tissue-  
12 types are very distinct from each other. The resection margin presented several  
13 proteins previously correlated with cancer, but also other overexpressed  
14 proteins that may be related to tumor nourishment and metastasis, such as  
15 collagen alpha-1, ceruloplasmin, calpastatin, and E-cadherin. We argue that the  
16 resection margin plays a key role in Paget's "soil to seed" hypothesis, i.e., that  
17 cancer cells require a special microenvironment to nourish and that  
18 understanding it could ultimately lead to more effective treatments.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 **Keywords:** Gastric cancer, shotgun proteomics, microenvironment, resection  
31 margin.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Introduction

Gastric cancer is responsible for a high mortality rate and affects people of all ages<sup>1</sup>. It is classified according to three histological types: adenocarcinoma, which accounts for 90-95% of the gastric tumors, lymphoma diagnosed in about 3% of the cases and gastrointestinal stromal tumor (GIST). The diagnosis is usually performed only in advanced stages because there are few symptoms during the initial stages; this dramatically decreases the options of treatment and results in a five-year survival rate in only 25% of the cases<sup>2</sup>. It is also reported that the risk of this disease increases with age. Conversely, even though the incidence of gastric cancer is of only around 5% in individuals below 40, these cases are linked with a higher mortality rate as their lesions are usually confused with those from benign pathologies<sup>3</sup>.

A common problem when dealing with cancer is recurrence: a patient may suffer from the same cancer or metastasis even after curative surgery. To lower the chances of recurrence, the surgeon removes a rim of "healthy tissue" around the tumor, namely, the resection margin. This margin varies widely depending on the site and extent of the disease, so it is very difficult to define or establish standards<sup>4</sup>. After removal, it is further examined by a pathologist to search for cancer cells and ultimately define how to treat the patient and establish other medical procedures. A "negative microscopic margin" (i.e., cancer cells that were not detected by the pathologist is correlated with a good follow-up and survival rate; a "positive resection margin", especially in the case of pancreatic cancer, is correlated with a poor survival rate)<sup>4,5</sup>.

Richard Caprioli's group introduced a shift in paradigm on how these resection margins are studied by employing Matrix Assisted Laser Desorption

1  
2  
3 ionization (MALDI) imaging mass spectrometry <sup>6</sup>. Briefly, MALDI imaging  
4  
5 constitutes a strategy for analyzing the spatial distribution of ion signals related  
6  
7 to biomolecules such as peptides, proteins, and small molecules, usually from  
8  
9 tissue on a microscope slide. Patterns of mass spectral peaks can determine,  
10  
11 for example, a drug distribution or boundaries between tissues <sup>7</sup>. With MALDI  
12  
13 imaging, Caprioli's group pointed to various molecular changes, undetected by  
14  
15 immunohistochemistry and morphology assessments, and showed that what  
16  
17 was previously diagnosed as a histologically "normal" resection margin  
18  
19 contained many molecular characteristics similar to the tumor <sup>8,9</sup>. They finally  
20  
21 concluded that "cells near a tumor aren't so normal" and that, as seen from a  
22  
23 molecular perspective, the resection margin looked more like the tumor than the  
24  
25 normal cells even though their morphology did not show it yet. Another example  
26  
27 of MALDI mass spectrometry application has been on defining sets of mass  
28  
29 spectral peaks that may aid in the diagnosis and, possibly, in detecting gastric  
30  
31 cancer in a very early stage <sup>10</sup>. Further experimentation is required to identify  
32  
33 the proteins from which these spectral peaks could have originated.  
34  
35  
36  
37

38  
39 In all, the literature leaves us with no choice but to redefine what these  
40  
41 resection margins really are. Recent results now pose the resection margin as a  
42  
43 treasure trove for understanding tumorigenesis, tumor growth, and the  
44  
45 mechanisms behind metastasis: the tissue surrounding the tumor provides  
46  
47 means to nourish it. Here we further tackle the problem of studying the  
48  
49 resection margin by employing Multi-dimensional Protein Identification  
50  
51 Technology (MudPIT) to compare biopsies from gastric cancers, their resection  
52  
53 margins, and from corresponding regions of control subjects. Briefly, MudPIT  
54  
55 constitutes a large-scale strategy for identifying and quantifying proteins by  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

digesting them and employing peptide chromatographic separation online with tandem mass spectrometry<sup>11</sup>. Relative protein quantitation is obtained by acquiring and normalizing their peptide extracted ion chromatograms according to the distributed Normalized Ion Abundance Factor (dNIAF) approach. In summary, the latter is accomplished by porting the spectral counting normalization procedure described by Zhang et al. to XICs<sup>12</sup>. We argue that our approach is complementary to existing MALDI imaging approaches, which are advantageous in providing ion peak data related to a precise tissue location. On the other hand, MudPIT is capable of performing protein identification in large scale. Moreover, MALDI and ESI ionizations have been described to be complementary<sup>13</sup>.

All biopsies were obtained from patients or control subjects from the city of Manaus in the state of Amazonas, Brazil, and were negatively diagnosed for the presence of *Helicobacter pylori* (the main etiologic agent). Our main goal has been to investigate whether or not the resection margin is indeed predominantly similar to control tissue by using MudPIT.

## Material and Methods

### Subjects.

This study was approved by the Ethics Committee of the Federal University of Amazonas (CEP/UFAM: MEMO - n° 0057.0.115.000-11-CAAE). The samples were collected at the Oncology Control Foundation Center of the Amazonas State (FCECON), a very prestigious Brazilian institution. After signing informed consent, biopsies from tumor and resection margins were obtained by operating four patients, of which three were females. Briefly, resection margins were macroscopically defined during the operation as a 10 cm rim of healthy-looking

1  
2  
3 tissue surrounding the tumor. Four control biopsies were obtained during upper  
4  
5 endoscopy according to Bormann's classification for control subjects; three of  
6  
7 the subjects were females. Our criterion for classifying a subject as control was  
8  
9 by not detecting traces of cancer according to endoscopic evaluation. All  
10  
11 biopsies were obtained from the stomach, specifically from the gastric antrum.  
12  
13 Each biopsy was then subtyped and the clinical stage of the disease was  
14  
15 determined according to the Tumor, Node, and Metastasis (TNM) classification  
16  
17 of the American Joint Committee on Cancer (AJCC); from the four tumors, three  
18  
19 were classified as T4 and one as T3. Only histological type adenocarcinoma  
20  
21 was considered in this work.  
22

#### 23 24 25 **Protein solubilization with RapiGest and trypsin digestion.**

26  
27 All biopsies were pulverized with liquid nitrogen. Each protein pellet was re-  
28  
29 suspended independently with RapiGest SF according to the manufacturer's  
30  
31 instructions to a final concentration of 0.1% of RapiGest. The samples were  
32  
33 quantified using the BCA protein assay Kit (Sigma-Aldrich) as per the  
34  
35 manufacturer's instructions. One hundred micrograms of each sample were  
36  
37 reduced with 20 mM of dithiothreitol (DTT) at 60 °C for 30 min. The samples  
38  
39 were cooled to room temperature and incubated, in the dark, with 66 mM of  
40  
41 iodacetamide (IAA) for 20 min. Afterwards, all samples were digested overnight  
42  
43 with trypsin (Promega) at the ratio of 1/50 w/w (E/S) at 37°C. Following  
44  
45 digestion, all reactions were acidified with 10% formic acid (1% final  
46  
47 concentration) to stop the proteolysis. The samples were centrifuged during 15  
48  
49 minutes at 60,000 RCF to remove insoluble material.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60**Evaluation of protein profile by 1D polyacrylamide gel electrophoresis.**

Fifteen micrograms of each sample (control, tumor, and resection margin) were added to Lammeli buffer and heated for 5 min at 100 °C, and subsequently subjected to 1D electrophoresis on 12 % polyacrylamide gel. After running the gel, it was fixed for 30 min with 40% ethanol, 10% acetic acid in water. Subsequently the gel was stained with Coomassie blue R-250 for 2 h and destained with 40% ethanol, 10% acetic acid in water. After scanning we visually select bands of interest to be excised, digested with trypsin, and have their protein profiles analyzed by liquid chromatography / tandem mass spectrometry LC/MS/MS.

**LC/LC/MS/MS data acquisition.**

Fifty micrograms of the digested peptide mixture were desalted using reverse phase column manually packed in a tip using the Poros R2 resin (Applied Biosystems). The desalted peptides were resuspended in a solution composed of 0.1% TFA and 30% acetonitrile and then introduced into PolySulfoethyl A strong cation-exchange column (50 x 1 mm; PolyLC Inc., Columbia, MD, USA) using Ettan HPLC system (GE Healthcare). A linear salt gradient was applied from 0 to 800 mM NaCl and the absorbance was monitored at 215 and 280 nm; six salt steps fractions were obtained, desalted once again and analyzed on a reversed phase column coupled to an Orbitrap Velos mass spectrometer (Thermo, San Jose). The flow rate at the tip of the reverse column was 100nL/min when the mobile phase composition was 95% H<sub>2</sub>O, 5% acetonitrile, and 0.1% formic acid. The Orbitrap mass spectrometer was set to the data-dependent acquisition mode with dynamics exclusion turned on. One MS survey scan was followed by nine MS/MS scans using collision activated

1  
2  
3 dissociation with a normalized collision energy of 35. Mass spectrometer scan  
4  
5 functions and HPLC solvent gradients were controlled by the Xcalibur data  
6  
7 system (Thermo, San Jose, CA).

8  
9  
10 **Shotgun proteomic data analysis.**

11 **Protein sequence database.** MS1 and MS2 spectra were extracted from raw  
12  
13 files using RawXtractor<sup>14</sup>. Sequences from *Homo sapiens* were downloaded  
14  
15 from the UniProt consortium on January 1, 2012; we used these sequences to  
16  
17 prepare search database according to the semi-labeled decoy guidelines<sup>15</sup>.  
18  
19 This database included all *H. sapiens* sequences plus those from 127 common  
20  
21 contaminants (ex.: keratins, trypsin). Each sequence was used to generate two  
22  
23 additional decoy sequences, one tagged as MiddleReversed (labeled decoy)  
24  
25 and the other as PairReversed (unlabeled decoy); this was accomplished using  
26  
27 PatternLab's Search Database Generator<sup>16</sup>. We recall that the semi-labeled  
28  
29 decoy approach aims to enable a post-evaluation of result quality<sup>15,17</sup>.  
30  
31

32  
33 **Peptide sequence matching.** The ProLuCID search engine was used to  
34  
35 compare experimental MS2 against those theoretically generated from our  
36  
37 sequence database and select the most likely peptide sequence candidates<sup>18</sup>.  
38  
39 Briefly, the search was limited to fully and semi-tryptic peptide candidates; we  
40  
41 imposed carbamidomethylation as a fixed modification and oxidation of  
42  
43 Methionine as a variable modification. The search engine accepted peptide  
44  
45 candidates within a 70-ppm tolerance from the measured precursor *m/z* and  
46  
47 used the XCorr and Z-Score as the primary and secondary search engine  
48  
49 scores, respectively.  
50  
51

52  
53 **Assessment of PSMs.** The validity of the Peptide Sequence Matches (PSMs)  
54  
55 was assessed using the Search Engine Processor (SEPro) [16]. Identifications  
56  
57  
58  
59  
60

1  
2  
3 were grouped by charge state (+2 and  $\geq$  +3) and then by tryptic status (fully  
4 tryptic, semi-tryptic), resulting in four distinct subgroups. For each result, the  
5 ProLuCID XCorr, DeltaCN and ZScore values were used to generate a  
6 Bayesian discriminator. The identifications were sorted in a nondecreasing  
7 order according to the discriminator score. A cutoff score was established to  
8 accept a false-discovery rate (FDR) of 1% based on the number of labeled  
9 decoys. This procedure was independently performed on each data subset,  
10 resulting in a false-positive rate that was independent of tryptic status or charge  
11 state. Additionally, a minimum sequence length of six amino acid residues was  
12 required. Results were post-processed to only accept PSMs with less than 10  
13 ppm and proteins supported by two or more independent evidences (e.g.,  
14 identification of a peptide with different charge states, a modified and a non-  
15 modified version of the same peptide, or two different peptides). This last filter  
16 led to a 0% FDR in all search results at the labeled and unlabeled decoy levels  
17 for all our sample analyses.

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36 **Protein quantitation.** The MS1 files were deisotoped and decharged using  
37 YADA <sup>19</sup>. SEPro's quantitation module (SEProQ) was then used to obtain the  
38 XICs from the deconvoluted MS1 files and link them with the corresponding  
39 PSMs. The XICs were normalized according to the dNIAF approach, which  
40 employs the same procedure as the distributed Normalized Spectral Abundance  
41 Factors (dNSAF) approach <sup>12</sup>, but instead of relying on quantitation by spectral  
42 counts it uses the XICs extracted from the deconvoluted MS1. We recall that  
43 dNSAF normalization capitalizes on unique peptide signals to distribute the  
44 signal from peptides that are shared between proteins.

1  
2  
3 **Differential expression.** We used PatternLab's Approximately Area  
4 Proportional Venn Diagram module to pinpoint proteins uniquely identified in a  
5 tissue-type <sup>20</sup>; the analysis only considered proteins found in two or more  
6 biological replicates from that tissue-types (i.e., control, margin, or cancer). As  
7 for proteins common to two or more biological replicates we used PatternLab's  
8 TFold module using a q-value of 0.05 to pinpoint those that are differentially  
9 expressed. We recall that the TFold module uses a theoretical FDR estimator to  
10 maximize identifications satisfying both a fold-change cutoff that varies with the  
11 t-test p-value as a power law and a stringency criterion that aims to fish out  
12 lowly abundant proteins that are likely to have had their quantitations  
13 compromised <sup>21</sup>.

14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27 **Multidimensional Scaling (MDS) analysis.** An MDS analysis was employed to  
28 aid in interpreting similarities within our dataset. For this, we implemented an  
29 algorithm, termed Buzios, of which we integrated into of the PatternLab for  
30 Proteomics environment <sup>22</sup>. We recall that MDS is used to represent data from a  
31 high-dimensional space in a lower-dimensional one, typically of two or three  
32 dimensions, to allow for visual access to patterns. Buzios takes as input the  
33 sparse matrix generated by SEPro, which summarizes the quantitations of all  
34 proteins from all experiments, with contents as described previously <sup>22</sup>. Briefly,  
35 each of the  $I$  rows (viz., a vector corresponding to one of the subjects involved)  
36 includes the results from a MudPIT analysis. Buzios maps each vector from an  
37  $N$ -dimensional space, where  $N$  corresponds to the number of proteins identified  
38 in all analyses, onto a two-dimensional space. The mapping is such that each  
39 input dimension corresponds to the quantitation obtained for a given protein.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
The final outcome is a representation of each vector as a dot in a two-

1  
2  
3 dimensional space. This is done by attempting to respect their similarities in the  
4  
5 high-dimensional space as measured by a normalized dot product. As abiding  
6  
7 to this similarity criterion in a lower-dimensional space is usually not possible,  
8  
9 an approximation is obtained by solving the problem of finding two-dimensional  
10  
11 representations  $x_1, \dots, x_l$  that minimize the function

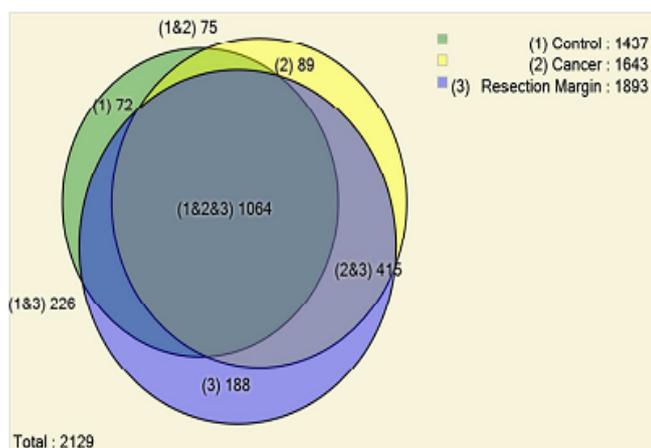
$$\sum_{i < j} (\|x_i - x_j\| - \delta_{i,j})^2 w_i w_j,$$

12  
13  
14  
15  
16  
17 where each  $\delta$  is one of the aforementioned similarities and each  $w$  is a weight  
18  
19 to penalize outliers. The weights are attributed as follows. First, for each class,  
20  
21 its centroid is calculated in the high-dimensional space. Second, vectors are  
22  
23 ordered in a nondecreasing order according to their Euclidian distances to the  
24  
25 centroid. Finally, each vector's weight is set to  $1/\text{rank}$ .  
26  
27  
28  
29

## 30 31 **Results and Discussion**

### 32 33 **Proteins uniquely identified to a state.**

34  
35 The Venn diagram comparing the proteins found in the control, cancer, and  
36  
37 resection margin biopsies is described in Figure 1. Even though the Venn  
38  
39 diagram shows some proteins to be unique to a tissue-type, we point out that  
40  
41 such is not necessarily true; they might be present in lower abundance and thus  
42  
43 below our experiment a detection capability for the given sample complexity. A  
44  
45 list of the proteins corresponding to each of the diagram's areas is available in  
46  
47 Supplementary File 1. Next we discuss some of these proteins.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



**Figure 1.** Venn diagram comparing proteins identified from biopsies of control subjects, cancer patients, and the corresponding resection margins. Only proteins found in two or more biological replicates were considered.

#### Proteins uniquely identified in the resection margins.

**Pepsinogen (PGA).** PGAs 4 and 5, group I are inactive precursors to pepsin A synthesized in the cells of the stomach mucous membrane. Some studies report the association of pepsinogen expression with pre-neoplastic and neoplastic changes of the stomach mucosa, as well as its significance in cases of gastric cancer, especially to screening as a predictor, irrespective of *H. pylori* infection<sup>23,24</sup>. Another study suggests that the pepsinogen group I is useful for the early detection of recurrent gastric cancer, as it was observed that the values of pepsinogen become elevated with the recurrence and increase with time. On the other hand, in patients with no recurrence, the levels of this protein does not demonstrate a substantial difference<sup>25</sup>.

**Collagen alpha-1 (COL11A1).** Collagen is a protein that acts in cell adhesion and is found in the extracellular matrix. Zhao et al. described COL11A1 as a

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

marker for premalignant lesions in cancer. As we only identified COL11A1 in resection margins our findings support previous reports linking this protein with cell migration, angiogenesis, and tissue morphogenesis<sup>26,27</sup>. The literature also points out that COL11A1 was found overexpressed in gastric cancers as compared to controls and linked this protein with invasion and metastasis<sup>28,29</sup>. One potential role of COL11A1 upregulation has been described as distinguishing between pre-malignant and malignant lesions in stomach cancer<sup>26</sup>.

**Ceruloplasmin.** Ceruloplasmin is a glycoprotein synthesized in the liver and transports copper in the serum. Previous works suggest this protein to be involved in angiogenesis and neovascularization<sup>30</sup>, being therefore aligned with the soil (i.e., resection margin) to seed (i.e., tumor) model. In another study, Scanni et al. correlated the levels of ceruloplasmin with the prognosis for gastrointestinal cancer by showing that higher levels were linked with clinical evolution<sup>31</sup>.

**Calpastatin.** This protein's regulation has been associated with lymphovascular invasion in breast cancer, thus playing a role in the initial metastatic dissemination<sup>32</sup>.

**E-cadherin.** Cadherin is an adhesion molecule and E-cadherin is the prototype of class-E cadherin that links to catenins to form the cytoskeleton. Recent evidence shows that E-cadherin plays an important role in the early stage of tumorigenesis by modulating intracellular signaling to ultimately promote tumor growth<sup>33,34</sup>.

**Annexin 1.** Annexin 1 has been linked with tumorigenesis in glioblastomas<sup>35</sup> and urothelial carcinomas<sup>36</sup>.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Proteins uniquely identified in the cancer biopsies.**

**Tumor protein D52.** This protein has been previously associated with other types of cancer such as ovarian<sup>37</sup>, but as far as we know, there are no reports linking its overexpression with stomach cancer.

**Prostate leucine zipper isoform.** This protein is a member of the D52 tumor protein family and has been correlated with prostate cancer<sup>38</sup>. Since the present study has included one single male subject, it would not be inconceivable to hypothesize that this protein is overexpressed precisely on account of this subject. Indeed, by looking in our data, we found this protein to be present in the male's sample. Unexpectedly, we also identified this protein (with six peptides) in the sample from a 71-year old female patient in this group.

**The proliferating cell nuclear antigen (PCNA).** PCNA is essential for DNA replication and damage repair, chromatin formation, and cell cycle progression. Given its diverse functions, PCNA is described as one of the essential non-oncogenic mediators supporting cancer growth<sup>39</sup>. The prognostic significance of PCNA expression has been previously described for gastric carcinomas<sup>40</sup>.

**Proteins identified in the cancer biopsies and resection margins but not in the control biopsies.**

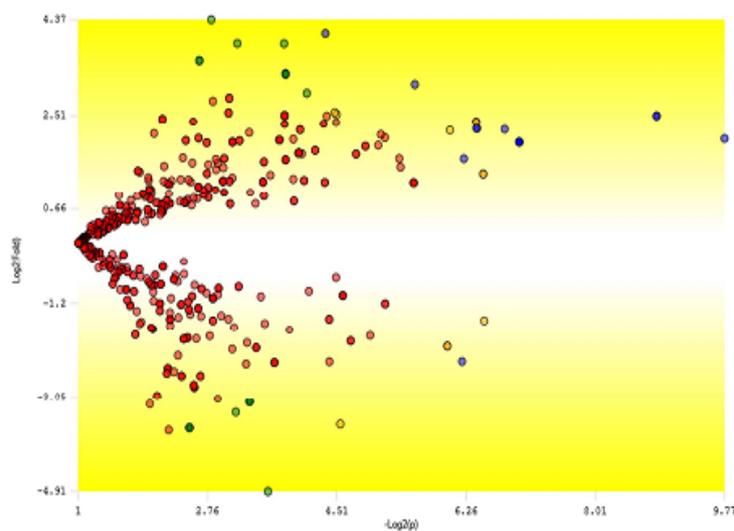
**Fibronectin.** This is a matrix glycoprotein that plays an important role in cellular attachment, growth, and cell spreading, Its expression is increased in numerous, including the stimulation of carcinoma cell growth and the inhibition of apoptosis<sup>2;41</sup>. Histopathological studies strongly suggest that its elevated presence is topographically associated with the invasion front of gastric

1  
2  
3 adenocarcinomas and clinically correlated with an increased risk of local  
4 invasion and metastasis <sup>42</sup>.

5  
6  
7 **Fibulin-1.** This protein is a calcium-binding glycoprotein found in association  
8 with extracellular matrix structures, as microfibrils, basement membranes, and  
9 elastic fibres; it has been shown to modulate cell morphology, growth, adhesion,  
10 and motility <sup>43</sup>. Several studies suggest the interaction between Fibulin-1 and  
11 Fibronectin <sup>44</sup>. Furthermore, it has been associated with tumor progression, its  
12 differential expression occurring in a range of human cancers, such as prostate  
13 cancer and breast cancer <sup>45,46</sup>.

#### 24 **TFold differential expression analysis.**

25  
26 TFold analyses were performed to further compare the resection margin versus  
27 cancer versus controls; an illustrative image of a TFold analysis comparing the  
28 resection margin versus controls is found in Figure 2.



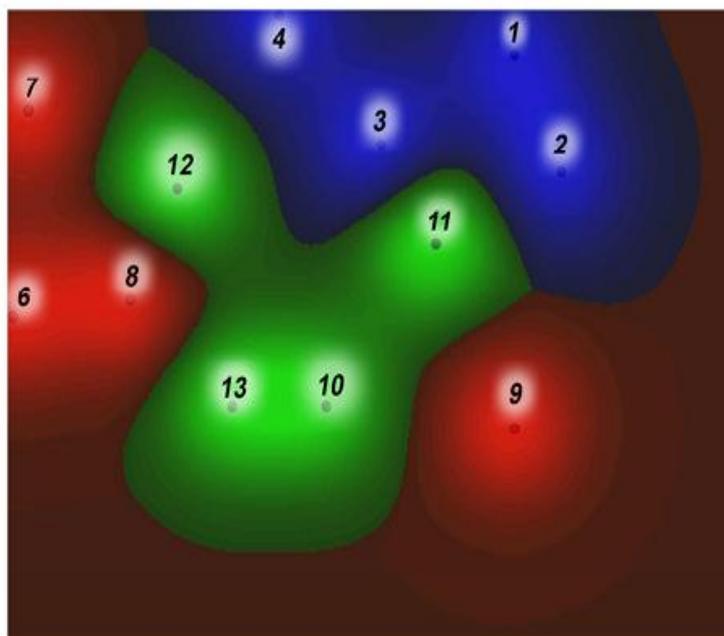
55  
56 **Figure 2.** TFold analysis comparing the proteomic profile of proteins identified  
57 in two or more biological replicates from control subjects versus resection  
58

1  
2  
3 margins. Each protein is mapped as a dot on the plot according to its  $-\text{Log}_2(\text{p-value})$  (x-axis) and  $\text{Log}_2(\text{Fold change})$  (y-axis). Red dots are proteins that  
4 satisfy neither the variable fold-change cutoff nor the FDR cutoff  $\alpha = 0.05$  Green  
5 dots are those that satisfy the fold-change cutoff but not  $\alpha$ . Orange dots are  
6 those that satisfy both the fold-change cutoff and  $\alpha$  but are lowly abundant  
7 proteins and therefore most likely have their quantitations compromised. Finally,  
8 blue dots are those that satisfy all statistical filters. Dots in the upper part of the  
9 plot correspond to proteins overexpressed in the resection margin.  
10  
11  
12

13 When comparing the resection margin with cancer, we detected a  
14 downregulation of gelsolin in the tumor. This result is well aligned with previous  
15 reports. However, we hypothesize that the corresponding up-accumulation in  
16 the resection margin could be linked to metastasis, as gelsolin increases  
17 permeability and has been linked with tumor mobility <sup>47</sup>. A complete list of  
18 proteins pinpointed by the TFold analyses is found in Supplementary File 1.  
19  
20  
21  
22  
23  
24  
25  
26  
27

#### 28 **MDS analysis.**

29 To investigate the closeness of the control, resection margin, and tumor clusters  
30 of subjects, we performed multidimensional scaling as described in Materials  
31 and Methods. The clustering result is displayed in Figure 3.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

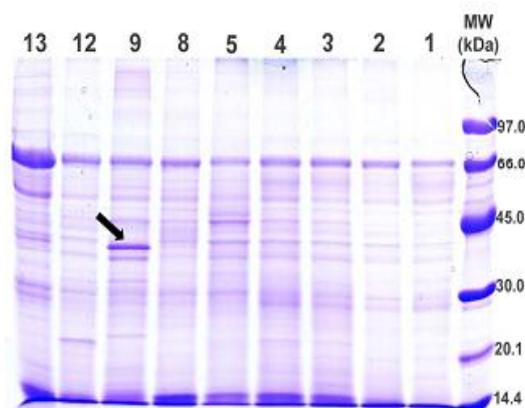


**Figure 3.** Multidimensional scaling analysis (MDS) of MudPIT data obtained from control, resection margin, and gastric cancer biopsies. The number besides each dot represents the corresponding patient's ID. The blue, green, and red regions delimit the regions for control, resection margin, and cancer, respectively. The boundaries were drawn employing a Radial Basis Function (RBF) kernel.

The interpretation of MDS plots is done on an intuitive basis, which naturally opens room for discussion. Be as it may, MDS may help pinpoint outliers in the data and provide insights (though only as from a bird's-eye view). Interestingly, control and resection margin subjects appear to be tightly clustered while, apparently, there is one outlier in the cancer realm. Moreover, although the resection margin subjects are clustered more closely to the control subjects than are the cancer subjects (i.e., the green region that represents the resection margin is somewhat separating the blue (controls) from the red (cancer)), in general it seems hard to mistake members of the resection margin cluster for those of the control cluster. We regard this as strengthening the view

17

1  
2  
3 that the resection margin has very specific features and should not be seen as  
4 healthy tissue. As for the outlier subject, it motivated us to further investigate  
5 our samples and rethink the computational approach employed for this analysis.  
6  
7  
8  
9  
10 In this regard, we proceeded with a 1D gel analysis to verify whether any  
11 obvious pattern showed up to discriminate sample #9. The result of this 1D gel  
12 is found in Figure 4.



17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35 **Figure 4.** 1D gel analysis of protein profiles obtained from cancer (lanes 8, 9),  
36 resection margin (lanes 12, 13), and control (lanes 1, 2, 3, 4) biopsies. The  
37 arrow marks an overexpressed band in sample #9.

38  
39  
40 By visually inspecting the 1D gel, we clearly noticed a bold band, which  
41 unarguably is overexpressed only in sample #9. This band was then excised  
42 from the gel, as were the bands in the equivalent regions from the other lanes.  
43  
44 Proteins were extracted from these bands and their contents trypsinized and  
45 analyzed by LC/MS/MS on our Orbitrap XL. By performing an ACFold analysis  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
22 (data not shown) we were able to establish that the proteins with the greatest  
changes in quantitation were tropomyosin and filamin-A. Indeed, these were the  
ones with the most spectral counts in all three replicate analyses of the band in  
question for sample #9. Interestingly, it has been hypothesized that, together,

1  
2  
3 these two proteins play a key role in "one mechanism by which the switch to a  
4  
5 TGF- $\beta$  tumorigenic response occurs"<sup>48</sup>.  
6  
7

### 8 9 **Final Considerations**

10  
11 Here we compared protein profiles of cancer, resection margin, and  
12 control biopsies to investigate whether the resection margin profiles are more  
13 similar to those from cancer or control biopsies. During this comparison, we  
14 pinpointed several key proteins that have been previously correlated with the  
15 disease. For example, we highlighted several proteins that could be linked with  
16 tumor growth and were found upregulated in the margin, thus lending support to  
17 the soil to seed hypothesis. While our goal has not been to investigate  
18 biomarkers, as this requires a much larger cohort, our results do nevertheless  
19 make it clear that the resection margin has very specific features that deserve a  
20 better understanding and could aid in the development of future treatments. Our  
21 MDS analysis revealed limitations in our differential proteomic strategy (which,  
22 incidentally, is adopted in various works for analyzing differential expression in  
23 data). A comparison of shotgun proteomic profiles without considering  
24 independent sample analyses, by strategies such as our 1D gel analysis or our  
25 MDS algorithm, could lead to missing important information. In our case, the  
26 TFold analysis missed two striking features related to overexpressed proteins in  
27 sample #9. Although these proteins were also found in other patients, a  
28 considerable standard deviation exists and in turn blinds most common  
29 statistical strategies. By employing a simple complementary approach, we were  
30 able to better investigate the unique features that explained that subject's status  
31 as an outlier. The take-home lesson is on the importance that should be given  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 to each individual sample, as each patient may present unique features. Finally,  
4  
5 here we described a complementary approach to MALDI imaging. Although  
6  
7 spatial resolution has been lost as a consequence, in-depth information for  
8  
9 protein identification was obtained. We conclude that a combination of these  
10  
11 strategies must be further explored to better investigate the role of the resection  
12  
13 margin as well as of any other problematic that includes spatial data.  
14  
15  
16  
17

### 18 **Acknowledgements**

19  
20 The authors thank FAPERJ, CAPES, CNPq, Fiocruz 30/2006-CAPES,  
21  
22 PDTIS, Nova Analítica, and Fundação do Câncer for financial support.  
23  
24  
25  
26

### 27 **Supporting Information Available**

28  
29 The raw mass spectra files, the PatternLab intermediary files, and a  
30  
31 complete list of identifications for all conditions are available for download at  
32  
33 <http://max.ioc.fiocruz.br/pcarvalho/2012aguino>. The PatternLab modules  
34  
35 used in this work are available for download at  
36  
37 <http://pcarvalho.com/patternlab>.  
38  
39  
40  
41  
42

### 43 **Reference List**

- 44  
45 [1] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA*  
46 *Cancer J Clin* 2005 Mar;55(2):74-108.  
47  
48 [2] Ebert MP, Niemeyer D, Deininger SO, Wex T, Knippig C, Hoffmann J, et  
49 al. Identification and confirmation of increased fibrinopeptide a serum  
50 protein levels in gastric cancer sera by magnet bead assisted MALDI-  
51 TOF mass spectrometry. *J Proteome Res* 2006 Sep;5(9):2152-8.  
52  
53 [3] Theuer CP, de VC, Keese G, French S, Arnell T, Tolmos J, et al. Gastric  
54 adenocarcinoma in patients 40 years of age or younger. *Am J Surg* 1996  
55 Nov;172(5):473-6.  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [4] Shen JG, Cheong JH, Hyung WJ, Kim J, Choi SH, Noh SH. Influence of a microscopic positive proximal margin in the treatment of gastric adenocarcinoma of the cardia. *World J Gastroenterol* 2006 Jun 28;12(24):3883-6.
- [5] Neoptolemos JP, Stocken DD, Dunn JA, Almond J, Beger HG, Pederzoli P, et al. Influence of resection margins on survival for patients with pancreatic cancer treated by adjuvant chemoradiation and/or chemotherapy in the ESPAC-1 randomized controlled trial. *Ann Surg* 2001 Dec;234(6):758-68.
- [6] Chaurand P, Norris JL, Cornett DS, Mobley JA, Caprioli RM. New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry. *J Proteome Res* 2006 Nov;5(11):2889-900.
- [7] Castellino S, Groseclose MR, Wagner D. MALDI imaging mass spectrometry: bridging biology and chemistry in drug development. *Bioanalysis* 2011 Nov;3(21):2427-41.
- [8] Oppenheimer SR, Mi D, Sanders ME, Caprioli RM. Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma. *J Proteome Res* 2010 May 7;9(5):2182-90.
- [9] Cottingham K. MALDI MS reveals that cells near a tumor aren't so normal. *J Proteome Res* 2010 May 7;9(5):2049.
- [10] Kim HK, Reyzer ML, Choi IJ, Kim CG, Kim HS, Oshima A, et al. Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. *J Proteome Res* 2010 Aug 6;9(8):4123-30.
- [11] Washburn MP, Wolters D, Yates JR, III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001 Mar;19(3):242-7.
- [12] Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem* 2010 Mar 15;82(6):2272-81.
- [13] Siuzdak G. The emergence of mass spectrometry in biochemical research. *Proc Natl Acad Sci U S A* 1994 Nov 22;91(24):11290-7.
- [14] McDonald WH, Tabb DL, Sadygov RG, MacCoss MJ, Venable J, Graumann J, et al. MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom* 2004;18(18):2162-8.
- [15] Barboza R, Cociorva D, Xu T, Barbosa VC, Perales J, Valente RH, et al. Can the false-discovery rate be misleading? *Proteomics* 2011 Oct;11(20):4105-8.

- 1  
2  
3 [16] Carvalho PC, Fischer JS, Xu T, Cociorva D, Balbuena TS, Valente RH,  
4 et al. Search Engine Processor: filtering and organizing PSMs.  
5 Proteomics 2012 Feb 6.  
6
- 7 [17] Yates JR, III, Park SK, Delahunty CM, Xu T, Savas JN, Cociorva D, et al.  
8 Toward objective evaluation of proteomic algorithms. Nat Methods  
9 2012;9(5):455-6.  
10
- 11 [18] Xu T, Venable JD, Park SK, Cociorva D, Lu B, Liao L, et al. ProLuCID, a  
12 fast and sensitive tandem mass spectra-based protein identification  
13 program. Mol Cell Proteomics 2006;5(S174).  
14
- 15 [19] Carvalho PC, Xu T, Han X, Cociorva D, Barbosa VC, Yates JR, III.  
16 YADA: A tool for taking the most out of high-resolution spectra.  
17 Bioinformatics 2009 Aug 14;25:2734-6.  
18
- 19 [20] Carvalho PC, Fischer JS, Perales J, Yates JR, Barbosa VC, Bareinboim  
20 E. Analyzing marginal cases in differential shotgun proteomics.  
21 Bioinformatics 2011 Jan 15;27(2):275-6.  
22
- 23 [21] Carvalho PC, Yates I, Jr., Barbosa VC. Improving the TFCID test for  
24 differential shotgun proteomics. Bioinformatics 2012 Apr 26.  
25
- 26 [22] Carvalho PC, Fischer JS, Chen EI, Yates JR, III, Barbosa VC.  
27 PatternLab for proteomics: a tool for differential shotgun proteomics.  
28 BMC Bioinformatics 2008;9:316.  
29
- 30 [23] Oishi Y, Kiyohara Y, Kubo M, Tanaka K, Tanizaki Y, Ninomiya T, et al.  
31 The serum pepsinogen test as a predictor of gastric cancer: the  
32 Hisayama study. Am J Epidemiol 2006 Apr 1;163(7):629-37.  
33
- 34 [24] Kitahara F, Kobayashi K, Sato T, Kojima Y, Araki T, Fujino MA. Accuracy  
35 of screening for gastric cancer using serum pepsinogen concentrations.  
36 Gut 1999 May;44(5):693-7.  
37
- 38 [25] Kodama M, Koyama K, Tsuburaya Y, Ishikawa K, Koyama H, Narisawa  
39 T, et al. Group I pepsinogen for early detection of gastric cancer  
40 recurrence after total gastrectomy. World J Surg 1990 Jan;14(1):94-9.  
41
- 42 [26] Zhao Y, Zhou T, Li A, Yao H, He F, Wang L, et al. A potential role of  
43 collagens expression in distinguishing between premalignant and  
44 malignant lesions in stomach. Anat Rec (Hoboken ) 2009  
45 May;292(5):692-700.  
46
- 47 [27] Skubitz AP. Adhesion molecules. Cancer Treat Res 2002;107:305-29.  
48
- 49 [28] Yasui W, Oue N, Ito R, Kuraoka K, Nakayama H. Search for new  
50 biomarkers of gastric cancer through serial analysis of gene expression  
51 and its clinical implications. Cancer Sci 2004 May;95(5):385-92.  
52
- 53 [29] Oue N, Hamai Y, Mitani Y, Matsumura S, Oshimo Y, Aung PP, et al.  
54 Gene expression profile of gastric carcinoma: identification of genes and  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- tags potentially involved in invasion, metastasis, and carcinogenesis by serial analysis of gene expression. *Cancer Res* 2004 Apr 1;64(7):2397-405.
- [30] Kunapuli SP, Singh H, Singh P, Kumar A. Ceruloplasmin gene expression in human cancer cells. *Life Sci* 1987 Jun 8;40(23):2225-8.
- [31] Scanni A, Tomirotti M, Licciardello L, Annibali E, Biraghi M, Trovato M, et al. Variations in serum copper and ceruloplasmin levels in advanced gastrointestinal cancer treated with polychemotherapy. *Tumori* 1979 Jun 30;65(3):331-8.
- [32] Storr SJ, Mohammed RA, Woolston CM, Green AR, Parr T, Spiteri I, et al. Calpastatin is associated with lymphovascular invasion in breast cancer. *Breast* 2011 Oct;20(5):413-8.
- [33] Chan AO. E-cadherin in gastric cancer. *World J Gastroenterol* 2006 Jan 14;12(2):199-203.
- [34] Debruyne P, Vermeulen S, Mareel M. The role of the E-cadherin/catenin complex in gastrointestinal cancer. *Acta Gastroenterol Belg* 1999 Oct;62(4):393-402.
- [35] Yang Y, Liu Y, Yao X, Ping Y, Jiang T, Liu Q, et al. Annexin 1 released by necrotic human glioblastoma cells stimulates tumor cell growth through the formyl peptide receptor 1. *Am J Pathol* 2011 Sep;179(3):1504-12.
- [36] Kang WY, Chen WT, Huang YC, Su YC, Chai CY. Overexpression of annexin 1 in the development and differentiation of urothelial carcinoma. *Kaohsiung J Med Sci* 2012 Mar;28(3):145-50.
- [37] Byrne JA, Maleki S, Hardy JR, Gloss BS, Murali R, Scurry JP, et al. MAL2 and tumor protein D52 (TPD52) are frequently overexpressed in ovarian carcinoma, but differentially associated with histological subtype and patient outcome. *BMC Cancer* 2010;10:497.
- [38] Wang R, Xu J, Saramaki O, Visakorpi T, Sutherland WM, Zhou J, et al. PrLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in human prostate cancer. *Cancer Res* 2004 Mar 1;64(5):1589-94.
- [39] Punchihewa C, Inoue A, Hishiki A, Fujikawa Y, Connelly M, Evison B, et al. Identification of Small Molecule Proliferating Cell Nuclear Antigen (PCNA) Inhibitor That Disrupts Interactions with PIP-box Proteins and Inhibits DNA Replication. *J Biol Chem* 2012 Apr 20;287(17):14289-300.
- [40] Konno S, Takebayashi Y, Aiba M, Akiyama S, Ogawa K. Clinicopathological and prognostic significance of thymidine phosphorylase and proliferating cell nuclear antigen in gastric carcinoma. *Cancer Lett* 2001 May 10;166(1):103-11.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [41] Fischer JS, Carvalho PC, Neves-Ferreira AG, da Fonseca CO, Perales J, Carvalho MG, et al. Anti-thrombin as a prognostic biomarker candidate for patients with recurrent glioblastoma multiform under treatment with perillyl alcohol. *J Exp Ther Oncol* 2008;7(4):285-90.
- [42] Nejari M, Anderson W, Pourreyaon C, Jacquier MF, Scoazec JY, Remy L. The role of fibroblasts in the modulation of integrin-dependent interactions between the gastric cell line HGT-1 and fibronectin. *Int J Cancer* 2004 Nov 20;112(4):560-9.
- [43] Twal WO, Czirok A, Hegedus B, Knaak C, Chintalapudi MR, Okagawa H, et al. Fibulin-1 suppression of fibronectin-regulated cell adhesion and motility. *J Cell Sci* 2001 Dec;114(Pt 24):4587-98.
- [44] Argraves WS, Tran H, Burgess WH, Dickerson K. Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *J Cell Biol* 1990 Dec;111(6 Pt 2):3155-64.
- [45] Greene LM, Twal WO, Duffy MJ, McDermott EW, Hill AD, O'Higgins NJ, et al. Elevated expression and altered processing of fibulin-1 protein in human breast cancer. *Br J Cancer* 2003 Mar 24;88(6):871-8.
- [46] Wlazlinski A, Engers R, Hoffmann MJ, Hader C, Jung V, Muller M, et al. Downregulation of several fibulin genes in prostate cancer. *Prostate* 2007 Dec 1;67(16):1770-80.
- [47] Zhu WY, Hunag YY, Liu XG, He JY, Chen DD, Zeng F, et al. Prognostic evaluation of CapG, gelsolin, P-gp, GSTP1, and Topo-II proteins in non-small cell lung cancer. *Anat Rec (Hoboken)* 2012 Feb;295(2):208-14.
- [48] Assinder S, Cole N. Does TGF-beta induced formation of actin stress fibres reinforce Smad dependent TGF-beta signalling in the prostate? *Med Hypotheses* 2011 Jun;76(6):802-4.