Are Gastric Cancer Resection Margin Proteomic Profiles More Similar to Those from Controls or Tumors?

Priscila F. Aquino,¹,‡ Juliana S. G. Fischer,¹,‡ Ana G. C. Neves-Ferreira,§ Jonas Perales,§ Gilberto B. Domont,‖ Gabriel D. T. Araújo,‖ Valmir C. Barbosa,† Jucilana Viana,¶ Sidney R. S. Chalub,‡ Antonia Q. Lima de Souza,§ Maria G. C. Carvalho,§ Afonso D. Leão de Souza,† and Paulo C. Carvalho*‡

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

Supporting Information

ABSTRACT: A strategy for treating cancer is to surgically remove the tumor together with a portion of apparently healthy tissue surrounding it, the so-called “resection margin”, to minimize recurrence. Here, we investigate whether the proteomic profiles from biopsies of gastric cancer resection margins are indeed more similar to those from healthy tissue than from cancer biopsies. To this end, we analyzed biopsies using an offline MudPIT shotgun proteomic approach and performed label-free quantitation through a distributed normalized spectral abundance factor approach adapted for extracted ion chromatograms (XICs). A multidimensional scaling analysis revealed that each of those tissue-types is very distinct from each other. The resection margin presented several proteins previously correlated with cancer, but also other overexpressed proteins that may be related to tumor nourishment and metastasis, such as collagen alpha-1, ceruloplasmin, calpastatin, and E-cadherin. We argue that the resection margin plays a key role in Paget’s “soil to seed” hypothesis, that is, that cancer cells require a special microenvironment to nourish and that understanding it could ultimately lead to more effective treatments.

KEYWORDS: gastric cancer, shotgun proteomics, microenvironment, resection margin

INTRODUCTION

Gastric cancer is responsible for a high mortality rate and affects people of all ages.¹ 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.² 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.³

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.⁴ 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.⁴,⁵

Richard Caprioli’s group introduced a shift in paradigm on how these resection margins are studied by employing Matrix
Assisted Laser Desorption Ionization (MALDI) imaging mass spectrometry. Briefly, MALDI imaging constitutes a strategy for analyzing the spatial distribution of ion signals related to biomolecules such as peptides, proteins, and small molecules, usually from tissue on a microscope slide. Patterns of mass spectral peaks can determine, for example, a drug distribution or boundaries between tissues. With MALDI imaging, Caprioli’s group pointed to various molecular changes, undetected by immunohistochemistry and morphology assessments, and showed that what was previously diagnosed as a histologically “normal” resection margin contained many molecular characteristics similar to the tumor. They finally concluded that “cells near a tumor aren’t so normal” and that, as seen from a molecular perspective, the resection margin looked more like the tumor than the normal cells even though their morphology did not show it yet. Another example of MALDI mass spectrometry application has been on determining if the resection margin contained many resection margins really are. Recent results now pose the resection margin as a treasure trove for understanding tumorigenesis, tumor growth, and the mechanisms behind metastasis: the tissue surrounding the tumor provides means to nourish it. Here, we further tackle the problem of studying the resection margin by employing Multidimensional Protein Identification Technology (MudPIT) to compare biopsies from gastric cancers, their resection margins, and from corresponding regions of control subjects. Briefly, MudPIT constitutes a large-scale strategy for identifying and quantifying proteins by digesting them and employing peptide chromatographic separation online with tandem mass spectrometry. 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 extracted ion chromatograms (XICs). 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.

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 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 - no. 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 on four patients, of which three were females. Briefly, resection margins were macroscopically defined during the operation as a 10 cm rim of healthy-looking tissue surrounding the tumor. Four control biopsies were obtained during upper endoscopy according to Bormann’s classification for control subjects; three of the subjects were females. Our criterion for classifying a subject as control was by not detecting traces of cancer according to endoscopic evaluation. All biopsies were obtained from the stomach, specifically from the gastric antrum. Each biopsy was then subtyped and the clinical stage of the disease was determined according to the Tumor, Node, and Metastasis (TNM) classification of the American Joint Committee on Cancer (AJCC); from the four tumors, three were classified as T4 and one as T3. Only histological type adenocarcinoma was considered in this work.

#### Protein Solubilization with RapiGest and Trypsin Digestion

All biopsies were pulverized with liquid nitrogen. Each protein pellet was resuspended independently with RapiGest SF according to the manufacturer’s instructions to a final concentration of 0.1% of RapiGest. The samples were quantified using the BCA protein assay Kit (Sigma-Aldrich) as per the manufacturer’s instructions. One hundred micrograms of each sample was reduced with 20 mM of dithiothreitol (DTT) at 60 °C for 30 min. The samples were cooled to room temperature and incubated, in the dark, with 66 mM of iodoacetamide (IAA) for 20 min. Afterward, all samples were digested overnight with trypsin (Promega) at the ratio of 1/50 (w/w) (E/S) at 37 °C. Following digestion, all reactions were acidified with 10% formic acid (1% final concentration) to stop the proteolysis. The samples were centrifuged for 15 min at 60 000 RCF to remove insoluble material.

#### Evaluation of Protein Profile by 1D Polyacrylamide Gel Electrophoresis

Fifteen micrograms of each sample (control, tumor, and resection margin) was 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 and 10% acetic acid in water. Subsequently, the gel was stained with Coomassie blue R-250 for 2 h and destained with 40% ethanol and 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 was 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 PolySulfethyl A strong cation-exchange column (50 × 1 mm; PolyLC, Inc., Columbia, MD) 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, VA). The flow rate at the tip of the reverse column was 100 nL/min when the mobile phase composition was 95% H2O, 5% acetonitrile, and 0.1% formic acid. The Orbitrap mass spectrometer was set to the data-dependent acquisition mode with a dynamic exclusion of 90 s.

---

**Journal of Proteome Research**

dx.doi.org/10.1021/pr300612x | J. Proteome Res. 2012, 11, 5836-5842
One MS survey scan was followed by nine MS/MS scans using collision activated dissociation with a normalized collision energy of 35. Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (Thermo, San Jose, CA).

**Shotgun Proteomic Data Analysis**

**Protein Sequence Database.** MS1 and MS2 spectra were extracted from raw files using RawXtractor. Sequences from *Homo sapiens* were downloaded from the UniProt consortium on January 1, 2012; we used these sequences to prepare search database according to the semilabeled decoy guidelines. This database included all *H. sapiens* sequences, *H. pylori, Epstein–Barr* virus, plus those from 127 common contaminants (e.g., keratins, trypsin). Each sequence was used to generate two additional decoy sequences, one tagged as MiddleReversed (labeled decoy) and the other as PairReversed (unlabeled decoy); this was accomplished using PatternLab's Search Database Generator. Our final database contained 599,998 sequences. We recall that the semilabeled decoy approach aims to enable a postevaluation of result quality.

**Peptide Sequence Matching.** The ProLuCID search engine was used to compare experimental MS2 against those theoretically generated from our sequence database and select the most likely peptide sequence candidates. Briefly, the search was limited to fully and semitryptic peptide candidates; we imposed carbamidomethylation as a fixed modification and oxidation of Methionine as a variable modification. The search engine accepted peptide candidates within a 70-ppm tolerance from the measured precursor m/z and used the XCorr and Z-Score as the primary and secondary search engine scores, respectively.

**Assessment of Peptide Sequence Matches (PSMs).** The validity of the PSMs was assessed using the Search Engine Score as the primary and secondary search engine scores, to enable a postevaluation of result quality. This approach aims to postaccept PSMs with a 10-ppm tolerance and proteins supported by two or more PSMs. Results were postprocessed to only accept PSMs with a 70-ppm tolerance from the measured precursor m/z and used the XCorr and Z-Score as the primary and secondary search engine scores, respectively.

**Protein Quantitation.** The MS1 files were deisotoped and decharged using YADA. SEPro’s quantitation module (SEProQ) was then used to obtain the XICs from the deconvoluted MS1 files and link them with the corresponding PSMs. The XICs were normalized according to the dNIAF approach, which employs the same procedure as the distributed Normalized Spectral Abundance Factors (dNSAF) approach, but instead of relying on quantitation by spectral counts it uses the XICs extracted from the deconvoluted MS1. We recall that dNSAF normalization capitalizes on unique peptide signals to distribute the signal from peptides that are shared between proteins.

**Differential Expression.** We used PatternLab’s Approximately Area Proportional Venn Diagram module to pinpoint proteins uniquely identified in a tissue-type; the analysis only considered proteins found in two or more biological replicates from that tissue-type (i.e., control, margin, or cancer). As for proteins common to two or more biological replicates, we used PatternLab’s TFold module using a q-value of 0.05 to pinpoint those that are differentially expressed. We recall that the TFold module uses a theoretical FDR estimator to maximize identifications satisfying both a fold-change cutoff that varies with the test parameter as a power law and a stringency criterion that aims to fish out lowly abundant proteins that are likely to have had their quantitations compromised.

**Multidimensional Scaling (MDS) Analysis.** An MDS analysis was employed to aid in interpreting similarities within our data set. For this, we implemented an algorithm, termed Buzios, of which we integrated into the PatternLab for Proteomics environment. Buzios maps each vector from an N-dimensional space, where N corresponds to the number of proteins identified in all analyses, onto a two-dimensional space. The mapping is such that each input dimension corresponds to the quantitation obtained for a given protein. The final outcome is a representation of each vector as a dot in a two-dimensional space. This is done by attempting to respect their similarities in the high-dimensional space as measured by a normalized dot product. As abiding to this similarity criterion in a lower-dimensional space is usually not possible, an approximation is obtained by solving the problem of finding two-dimensional representations \(x_1, \ldots, x_I\) that minimize the function

\[
\sum_{i<j} (||x_i - x_j||^2 - \delta_{ij})^2 w_{ij}
\]

where each \(\delta\) is one of the aforementioned similarities and each \(w\) is a weight to penalize outliers. The weights are attributed as follows. First, for each class, its centroid is calculated in the high-dimensional space. Second, vectors are ordered in a nondecreasing order according to their Euclidian distances to the centroid. Finally, each vector’s weight is set to \(1/\text{rank}\).

**Available Data**

The raw mass spectra files, the PatternLab intermediary files, search database, the SEPro identification files and Excel spreadsheets listing the protein identification data are available for download at http://max.ioc.fioruz.br/pcarvalho/2012aquino. The PatternLab modules used in this work are available for download at http://pcarvalho.com/patterlab.

**RESULTS AND DISCUSSION**

**Proteins Uniquely Identified to a State**

The Venn diagram comparing the proteins found in the control, cancer, and resection margin biopsies is described in
Figure 1. Even though the Venn diagram shows some proteins to be unique to a tissue-type, we point out that such is not necessarily true; they might be present in lower abundance and thus below our experiment a detection capability for the given sample complexity. A list of the proteins corresponding to each of the diagram’s areas is available in Supporting Information (zip file). Next we discuss some of these proteins.

**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 preneoplastic 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.23,24 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.25

**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 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.26,27 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.28,29 One potential role of COL11A1 upregulation has been described as distinguishing between premalignant and malignant lesions in stomach cancer.26

**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,30 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.31

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

**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.33,34

**Annexin 1.** Annexin 1 has been linked with tumorigenesis in glioblastomas35 and urothelial carcinomas.36

**Proteins Uniquely Identified in the Cancer Biopsies**

**Tumor Protein D52.** This protein has been previously associated with other types of cancer such as ovarian,37 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.38 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 nononcogenic mediators supporting cancer growth.39 The prognostic significance of PCNA expression has been previously described for gastric carcinomas.40

**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.41,2 Histopathological studies strongly suggest that its elevated presence is topographically associated with the invasion front of gastric adenocarcinomas and clinically correlated with an increased risk of local invasion and metastasis.42

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

**TFold Differential Expression analysis**

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

When comparing the resection margin with cancer, we detected a downregulation of gelsolin in the tumor. This result is well aligned with previous reports. However, we hypothesize that the corresponding up-accumulation in the resection margin...
could be linked to metastasis, as gelsolin increases permeability and has been linked with tumor mobility. A complete list of proteins pinpointed by the TFold analyses is found in Supporting Information (zip file).

MDS Analysis

To investigate the closeness of the control, resection margin, and tumor clusters of subjects, we performed multidimensional scaling as described in Materials and Methods. The clustering result is displayed in Figure 3.

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 that the resection margin has very specific features and should not be seen as healthy tissue. As for the outlier subject, it motivated us to further investigate our samples and rethink the computational approach employed for this analysis. In this regard, we proceeded with a 1D gel analysis to verify whether any obvious pattern showed up to discriminate sample #9. The result of this 1D gel is found in Figure 4. Supporting Information Figure 1 shows a complementary 1D gel analysis including profiles from additional samples.

By visually inspecting the 1D gel, we clearly noticed a bold band, which unarguably is overexpressed only in sample #9. It is important to note that the MDS analysis we performed provides no direct evidence that the reason for isolating #9 is specifically due to the alteration observed in the 1D gel; nevertheless it is suggestive. This band was then excised from the gel, as were the bands in the equivalent regions from the other lanes. Proteins were extracted from these bands and their contents trypsinized and analyzed by LC/MS/MS on our Orbitrap XL. By performing an ACFold analysis (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, these two proteins play a key role in “one mechanism by which the switch to a TGF-β tumorigenic response occurs”; moreover, TGF-β was found to be overexpressed in our tumor tissues proteomic profiles as listed in our Venn Diagram results.
**FINAL CONSIDERATIONS**

Here, we compared protein profiles of cancer, resection margin, and control biopsies to investigate whether the resection margin profiles are more similar to those from cancer or control biopsies. During this comparison, we pinpointed several key proteins that have been previously correlated with the disease. For example, we highlighted several proteins that could be linked with tumor growth and were found upregulated in the margin, thus lending support to the soil to seed hypothesis. While our goal has not been to investigate biomarkers, as this requires much larger cohort, our results do nevertheless make it clear that the resection margin has very specific features that deserve a better understanding and could aid in the development of future treatments. Our MDS analysis revealed limitations in our differential proteomic strategy (which, incidentally, is adopted in various works for analyzing differential expression in data). A comparison of shotgun proteomic profiles without considering independent sample analyses, by strategies such as our 1D gel analysis or our MDS algorithm, could lead to missing important information. In our case, the TFold analysis missed two striking features related to overexpressed proteins in sample #9. Although these proteins were also found in other patients, a considerable deviation exists and in turn blinks most common statistical procedures. During this comparison, we pinpointed several key features that explained that subject status as an outlier. The take-home lesson is on the importance that should be given to features that the analysis approach enabled us to better investigate the unique TG Seg analysis that we employed. Nevertheless, our data deviation exists and in turn blinds most common statistical procedures. Our MDS analysis revealed limitations in our differential proteomic strategy (which, incidentally, is adopted in various works for analyzing differential expression in data). A comparison of shotgun proteomic profiles without considering independent sample analyses, by strategies such as our 1D gel analysis or our MDS algorithm, could lead to missing important information. In our case, the TFold analysis missed two striking features related to overexpressed proteins in sample #9. Although these proteins were also found in other patients, a considerable deviation exists and in turn blinks most common statistical procedures. During this comparison, we pinpointed several key features that explained that subject status as an outlier. The take-home lesson is on the importance that should be given to features that explained that subject status as an outlier. The take-home lesson is on the importance that should be given to features that explained that subject status as an outlier.

**REFERENCES**

(9) Cottingham, K. MALDI MS reveals that cells near a tumor aren’t so normal. J. Proteome Res. 2010, 9 (9), 2049.


