

1 Are Gastric Cancer Resection Margin Proteomic Profiles More Similar 2 to Those from Controls or Tumors?

3 Priscila F. Aquino,^{†,#} Juliana S. G. Fischer,^{‡,#} Ana G. C. Neves-Ferreira,[§] Jonas Perales,[§]
4 Gilberto B. Domont,^{||} Gabriel D. T. Araujo,^{||} Valmir C. Barbosa,[⊥] Jucilana Viana,^{||} Sidney R. S. Chalub,[○]
5 Antonia Q. Lima de Souza,^{||} Maria G. C. Carvalho,[⊗] Afonso D. Leao de Souza,[†] and Paulo C. Carvalho^{*:‡}

6 [†]Departamento de Química, Universidade Federal do Amazonas, Amazonas, Brazil

7 [‡]Instituto Carlos Chagas, Fiocruz, Parana, Brazil

8 [§]Laboratorio de Toxinologia, Instituto Oswaldo Cruz, Rio de Janeiro, Brazil

9 ^{||}Proteomics Unit, Rio de Janeiro Proteomics Network, Departamento de Bioquímica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

11 [⊥]Programa de Engenharia de Sistemas e Computacao, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

12 ^{||}Escola Superior de Ciencias da Saude, Universidade do Estado do Amazonas, Amazonas, Brazil

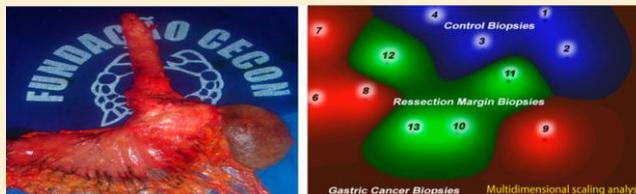
13 [○]Departamento de Cirurgia Digestiva, Universidade Federal do Amazonas, Amazonas, Brazil

14 [⊗]Departamento de Patologia, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

15 ^{*:‡} Supporting Information

16 **ABSTRACT:** A strategy for treating cancer is to surgically
17 remove the tumor together with a portion of apparently
18 healthy tissue surrounding it, the so-called “resection margin”,
19 to minimize recurrence. Here, we investigate whether the
20 proteomic profiles from biopsies of gastric cancer resection
21 margins are indeed more similar to those from healthy tissue
22 than from cancer biopsies. To this end, we analyzed biopsies
23 using an offline MudPIT shotgun proteomic approach and
24 performed label-free quantitation through a distributed normalized spectral abundance factor approach adapted for extracted ion
25 chromatograms (XICs). A multidimensional scaling analysis revealed that each of those tissue-types is very distinct from each
26 other. The resection margin presented several proteins previously correlated with cancer, but also other overexpressed proteins
27 that may be related to tumor nourishment and metastasis, such as collagen alpha-1, ceruloplasmin, calpastatin, and E-cadherin.
28 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
29 microenvironment to nourish and that understanding it could ultimately lead to more effective treatments.

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



31 INTRODUCTION

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

46 A common problem when dealing with cancer is recurrence:
47 a patient may suffer from the same cancer or metastasis even

48 after curative surgery. To lower the chances of recurrence, the
49 surgeon removes a rim of “healthy tissue” around the tumor,
50 namely, the resection margin. This margin varies widely
51 depending on the site and extent of the disease, so it is very
52 difficult to define or establish standards.⁴ After removal, it is
53 further examined by a pathologist to search for cancer cells and
54 ultimately define how to treat the patient and establish other
55 medical procedures. A “negative microscopic margin” (i.e.,
56 cancer cells that were not detected by the pathologist) is
57 correlated with a good follow-up and survival rate; a “positive
58 resection margin”, especially in the case of pancreatic cancer, is
59 correlated with a poor survival rate.^{4,5}

Richard Caprioli’s group introduced a shift in paradigm on
60 how these resection margins are studied by employing Matrix 61

Received: July 7, 2012

62 Assisted Laser Desorption Ionization (MALDI) imaging mass
63 spectrometry.⁶ Briefly, MALDI imaging constitutes a strategy
64 for analyzing the spatial distribution of ion signals related to
65 biomolecules such as peptides, proteins, and small molecules,
66 usually from tissue on a microscope slide. Patterns of mass
67 spectral peaks can determine, for example, a drug distribution
68 or boundaries between tissues.⁷ With MALDI imaging,
69 Caprioli's group pointed to various molecular changes,
70 undetected by immunohistochemistry and morphology assess-
71 ments, and showed that what was previously diagnosed as a
72 histologically "normal" resection margin contained many
73 molecular characteristics similar to the tumor.^{8,9} They finally
74 concluded that "cells near a tumor aren't so normal" and that,
75 as seen from a molecular perspective, the resection margin
76 looked more like the tumor than the normal cells even though
77 their morphology did not show it yet. Another example of
78 MALDI mass spectrometry application has been on defining
79 sets of mass spectral peaks that may aid in the diagnosis and,
80 possibly, in detecting gastric cancer in a very early stage.¹⁰
81 Further experimentation is required to identify the proteins
82 from which these spectral peaks could have originated.

83 In all, the literature leaves us with no choice but to redefine
84 what these resection margins really are. Recent results now
85 pose the resection margin as a treasure trove for understanding
86 tumorigenesis, tumor growth, and the mechanisms behind
87 metastasis: the tissue surrounding the tumor provides means to
88 nourish it. Here, we further tackle the problem of studying the
89 resection margin by employing Multidimensional Protein
90 Identification Technology (MudPIT) to compare biopsies
91 from gastric cancers, their resection margins, and from
92 corresponding regions of control subjects. Briefly, MudPIT
93 constitutes a large-scale strategy for identifying and quantifying
94 proteins by digesting them and employing peptide chromato-
95 graphic separation online with tandem mass spectrometry.¹¹
96 Relative protein quantitation is obtained by acquiring and
97 normalizing their peptide extracted ion chromatograms
98 according to the distributed Normalized Ion Abundance Factor
99 (dNIAB) approach. In summary, the latter is accomplished by
100 porting the spectral counting normalization procedure
101 described by Zhang et al. to extracted ion chromatograms
102 (XICs).¹² We argue that our approach is complementary to
103 existing MALDI imaging approaches, which are advantageous
104 in providing ion peak data related to a precise tissue location.
105 On the other hand, MudPIT is capable of performing protein
106 identification in large scale. Moreover, MALDI and ESI
107 ionizations have been described to be complementary.¹³
108 All biopsies were obtained from patients or control subjects
109 from the city of Manaus in the state of Amazonas, Brazil, and
110 were negatively diagnosed for the presence of *Helicobacter*
111 *pylori* (the main etiologic agent). Our main goal has been to
112 investigate whether the resection margin is indeed predom-
113 inantly similar to control tissue by using MudPIT.

114 MATERIAL AND METHODS

115 Subjects

116 This study was approved by the Ethics Committee of the
117 Federal University of Amazonas (CEP/UFAM: MEMO - no.
118 0057.0.115.000-11-CAAE). The samples were collected at the
119 Oncology Control Foundation Center of the Amazonas State
120 (FCECON), a very prestigious Brazilian institution. After
121 signing informed consent, biopsies from tumor and resection
122 margins were obtained by operating on four patients, of which

three were females. Briefly, resection margins were macro- 123
scopically defined during the operation as a 10 cm rim of 124
healthily-looking tissue surrounding the tumor. Four control 125
biopsies were obtained during upper endoscopy according to 126
Borrmann's classification for control subjects; three of the 127
subjects were females. Our criterion for classifying a subject as 128
control was by not detecting traces of cancer according to 129
endoscopic evaluation. All biopsies were obtained from the 130
stomach, specifically from the gastric antrum. Each biopsy was 131
then subtyped and the clinical stage of the disease was 132
determined according to the Tumor, Node, and Metastasis 133
(TNM) classification of the American Joint Committee on 134
Cancer (AJCC); from the four tumors, three were classified as 135
T4 and one as T3. Only histological type adenocarcinoma was 136
considered in this work. 137

Protein Solubilization with RapiGest and Trypsin Digestion 138

All biopsies were pulverized with liquid nitrogen. Each protein 139
pellet was resuspended independently with RapiGest SF 140
according to the manufacturer's instructions to a final 141
concentration of 0.1% of RapiGest. The samples were 142
quantified using the BCA protein assay Kit (Sigma-Aldrich) 143
as per the manufacturer's instructions. One hundred micro- 144
grams of each sample was reduced with 20 mM of dithiothreitol 145
(DTT) at 60 °C for 30 min. The samples were cooled to room 146
temperature and incubated, in the dark, with 66 mM of 147
iodoacetamide (IAA) for 20 min. Afterward, all samples were 148
digested overnight with trypsin (Promega) at the ratio of 1/50 149
(w/w) (E/S) at 37 °C. Following digestion, all reactions were 150
acidified with 10% formic acid (1% final concentration) to stop 151
the proteolysis. The samples were centrifuged for 15 min at 60 152
000 RCF to remove insoluble material. 153

Evaluation of Protein Profile by 1D Polyacrylamide Gel 154 Electrophoresis 155

Fifteen micrograms of each sample (control, tumor, and 156
resection margin) was added to Lammeli buffer and heated for 157
5 min at 100 °C, and subsequently subjected to 1D 158
electrophoresis on 12% polyacrylamide gel. After running the 159
gel, it was fixed for 30 min with 40% ethanol and 10% acetic 160
acid in water. Subsequently, the gel was stained with 161
Coommassie blue R-250 for 2 h and destained with 40% 162
ethanol and 10% acetic acid in water. After scanning, we visually 163
select bands of interest to be excised, digested with trypsin, and 164
have their protein profiles analyzed by liquid chromatography/ 165
tandem mass spectrometry LC/MS/MS. 166

LC/LC/MS/MS Data Acquisition 167

Fifty micrograms of the digested peptide mixture was desalted 168
using reverse phase column manually packed in a tip using the 169
Poros R2 resin (Applied Biosystems). The desalted peptides 170
were resuspended in a solution composed of 0.1% TFA and 171
30% acetonitrile and then introduced into PolySulfethyl A 172
strong cation-exchange column (50 × 1 mm; PolyLC, Inc., 173
Columbia, MD) using Ettan HPLC system GE Healthcare). A 174
linear salt gradient was applied from 0 to 800 mM NaCl and 175
the absorbance was monitored at 215 and 280 nm; six salt steps 176
fractions were obtained, desalted once again and analyzed on a 177
reversed phase column coupled to an Orbitrap Velos mass 178
spectrometer (Thermo, San Jose, VA). The flow rate at the tip 179
of the reverse column was 100 nL/min when the mobile phase 180
composition was 95% H₂O, 5% acetonitrile, and 0.1% formic 181
acid. The Orbitrap mass spectrometer was set to the data- 182
dependent acquisition mode with a dynamic exclusion of 90 s. 183

184 One MS survey scan was followed by nine MS/MS scans using
185 collision activated dissociation with a normalized
collision

186 energy of 35. Mass spectrometer scan functions and HPLC
187 solvent gradients were controlled by the Xcalibur data system
188 (Thermo, San Jose, CA).

189 Shotgun Proteomic Data Analysis

190 Protein Sequence Database. MS1 and MS2 spectra were
191 extracted from raw files using RawXtractor.¹⁴ Sequences from
192 Homo sapiens were downloaded from the UniProt consortium
193 on January 1, 2012; we used these sequences to prepare search
194 database according to the semilabeled decoy guidelines.¹⁵ This
195 database included all H. sapiens sequences, H. pylori, Epstein-
196 Barr virus, plus those from 127 common contaminants (e.g.,
197 keratins, trypsin). Each sequence was used to generate two
198 additional decoy sequences, one tagged as MiddleReversed
199 (labeled decoy) and the other as PairReversed (unlabeled
200 decoy); this was accomplished using PatternLab's Search
201 Database Generator.¹⁶ Our final database contained 599 998
202 sequences. We recall that the semilabeled decoy approach aims
203 to enable a postevaluation of result quality.^{15,17}

204 Peptide Sequence Matching. The ProLuCID search
205 engine was used to compare experimental MS2 against those
206 theoretically generated from our sequence database and select
207 the most likely peptide sequence candidates.¹⁸ Briefly, the
208 search was limited to fully and semitryptic peptide candidates;
209 we imposed carbamidomethylation as a fixed modification and
210 oxidation of Methionine as a variable modification. The search
211 engine accepted peptide candidates within a 70-ppm tolerance
212 from the measured precursor m/z and used the XCorr and Z-
213 Score as the primary and secondary search engine scores,
214 respectively.

215 Assessment of Peptide Sequence Matches (PSMs).
216 The validity of the PSMs was assessed using the Search Engine
217 Processor (SEPro).¹⁶ Identifications were grouped by charge
218 state (+2 and \geq +3) and then by tryptic status (fully tryptic,
219 semitryptic), resulting in four distinct subgroups. For each
220 result, the ProLuCID XCorr, DeltaCN and ZScore values were
221 used to generate a Bayesian discriminator. The identifications
222 were sorted in a nondecreasing order according to the
223 discriminator score. A cutoff score was established to accept a
224 false-discovery rate (FDR) of 1% based on the number of
225 labeled decoys. This procedure was independently performed
226 on each data subset, resulting in a false-positive rate that was
227 independent of tryptic status or charge state. Additionally, a
228 minimum sequence length of six amino acid residues was
229 required. Results were postprocessed to only accept PSMs with
230 less than 10 ppm and proteins supported by two or more
231 independent evidence (e.g., identification of a peptide with
232 different charge states, a modified and a nonmodified version of
233 the same peptide, or two different peptides). This last filter led
234 to a 0% FDR in all search results at the labeled and unlabeled
235 decoy levels for all our sample analyses.

236 Protein Quantitation. The MS1 files were deisotoped and
237 decharged using YADA.¹⁹ SEPro's quantitation module
238 (SEProQ) was then used to obtain the XICs from the
239 deconvoluted MS1 files and link them with the corresponding
240 PSMs. The XICs were normalized according to the dNIAF
241 approach, which employs the same procedure as the distributed
242 Normalized Spectral Abundance Factors (dNSAF) approach,¹²
243 but instead of relying on quantitation by spectral counts it uses
244 the XICs extracted from the deconvoluted MS1. We recall that
245 dNSAF normalization capitalizes on unique peptide signals to

distribute the signal from peptides that are shared between
proteins.

Differential Expression. We used PatternLab's Approx-
imately Area Proportional Venn Diagram module to pinpoint
proteins uniquely identified in a tissue-type;²⁰ the analysis only
considered proteins found in two or more biological replicates
from that tissue-types (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 t test p-value 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.²¹

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 of the PatternLab for
Proteomics environment.²² We recall that MDS is used to
represent data from a high-dimensional space in a lower-
dimensional one, typically of two or three dimensions, to allow
for visual access to patterns. Buzios takes as input the sparse
matrix generated by SEPro, which summarizes the quantitations
of all proteins from all experiments, with contents as described
previously.²² Briefly, each of the I rows (viz., a vector
corresponding to one of the subjects involved) includes the
results from a MudPIT analysis. 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, \dots, x_I that minimize the
function

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

where each δ 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.fiocruz.br/pcarvalho/>
2012aquino. The PatternLab modules used in this work are
available for download at <http://pcarvalho.com/patternlab>.

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

303 Figure 1. Even though the Venn diagram shows some proteins
304 to be unique to a tissue-type, we point out that such is not

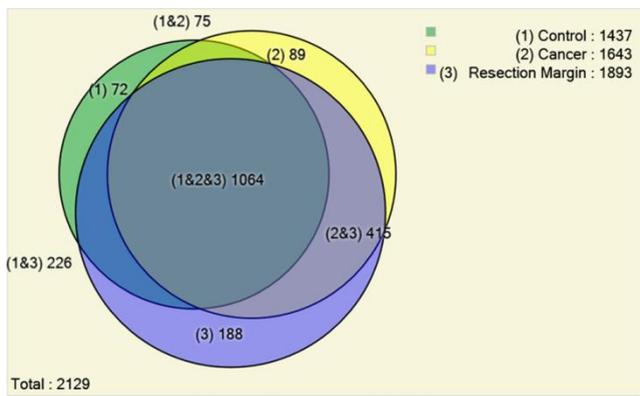


Figure 1. Venn diagram comparing proteins identified from biopsies of control subjects, cancer patients, and the corresponding resection margins. Only proteins found in two or more biological replicates were considered.

305 necessarily true; they might be present in lower abundance and
306 thus below our experiment a detection capability for the given
307 sample complexity. A list of the proteins corresponding to each
308 of the diagram's areas is available in Supporting Information
309 (zip file). Next we discuss some of these proteins.

310 Proteins Uniquely Identified in the Resection Margins

311 Pepsinogen (PGA). PGAs 4 and 5, group I are inactive
312 precursors to pepsin A synthesized in the cells of the stomach
313 mucous membrane. Some studies report the association of
314 pepsinogen expression with preneoplastic and neoplastic
315 changes of the stomach mucosa, as well as its significance in
316 cases of gastric cancer, especially to screening as a predictor,
317 irrespective of *H. pylori* infection.^{23,24} Another study suggests
318 that the pepsinogen group I is useful for the early detection of
319 recurrent gastric cancer, as it was observed that the values of
320 pepsinogen become elevated with the recurrence and increase
321 with time. On the other hand, in patients with no recurrence,
322 the levels of this protein does not demonstrate a substantial
323 difference.²⁵

324 Collagen Alpha-1 (COL1A1). Collagen is a protein that
325 acts in cell adhesion and is found in the extracellular matrix.
326 Zhao et al. described COL1A1 as a marker for premalignant
327 lesions in cancer. As we only identified COL1A1 in resection
328 margins our findings support previous reports linking this
329 protein with cell migration, angiogenesis, and tissue morpho-
330 genesis.^{26,27} The literature also points out that COL1A1 was
331 found overexpressed in gastric cancers as compared to controls
332 and linked this protein with invasion and metastasis.^{28,29} One
333 potential role of COL1A1 upregulation has been described as
334 distinguishing between premalignant and malignant lesions in
335 stomach cancer.²⁶

336 Ceruloplasmin. Ceruloplasmin is a glycoprotein synthe-
337 sized in the liver and transports copper in the serum. Previous
338 works suggest this protein to be involved in angiogenesis and
339 neovascularization,³⁰ being therefore aligned with the soil (i.e.,
340 resection margin) to seed (i.e., tumor) model. In another study,
341 Scanni et al. correlated the levels of ceruloplasmin with the
342 prognosis for gastrointestinal cancer by showing that higher
343 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 ulti-
mately promote tumor growth.^{33,34}

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

355 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

376 Proteins Identified in the Cancer Biopsies and Resection 377 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
microfibrils, 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}

396 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

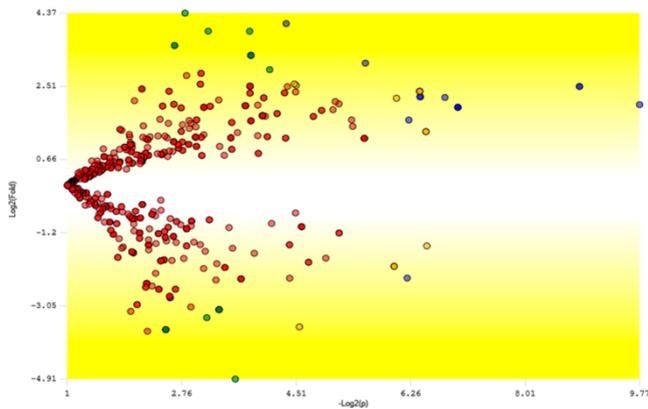


Figure 2. TFold analysis comparing the proteomic profile of proteins identified in two or more biological replicates from control subjects versus resection margins. Each protein is mapped as a dot on the plot according to its $-\text{Log}_2(\text{p-value})$ (x-axis) and $\text{Log}_2(\text{Fold change})$ (y-axis). Red dots are proteins that satisfy neither the variable fold-change cutoff nor the FDR cutoff $\alpha = 0.05$. Green dots are those that satisfy the fold-change cutoff but not α . Orange dots are those that satisfy both the fold-change cutoff and α but are lowly abundant proteins and therefore most likely have their quantitations compromised. Finally, blue dots are those that satisfy all statistical filters. Dots in the upper part of the plot correspond to proteins overexpressed in the resection margin.

could be linked to metastasis, as gelsolin increases permeability and has been linked with tumor mobility.⁴⁷ 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,

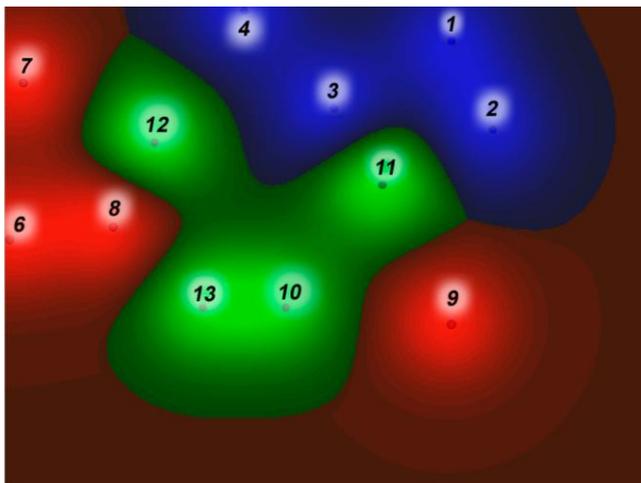


Figure 3. Multidimensional scaling analysis (MDS) of MudPIT data obtained from control, resection margin, and gastric cancer biopsies. The number besides each dot represents the corresponding patient's ID. The blue, green, and red regions delimit the regions for control, resection margin, and cancer,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

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.

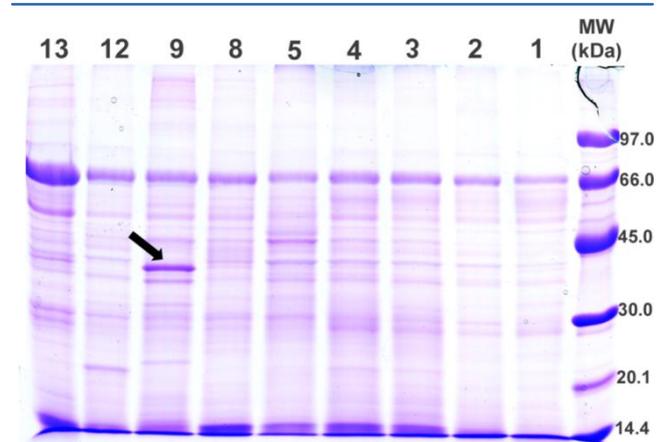


Figure 4. 1D gel analysis of protein profiles obtained from cancer (lanes 8, 9), resection margin (lanes 12, 13), and control (lanes 1, 2, 3, 4) biopsies. The arrow marks an overexpressed band in sample #9.

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 analysis²² (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";⁴⁸ moreover, TGF- β was found to be overexpressed in our tumor tissues proteomic profiles as listed in our Venn Diagram results.

455 **FINAL CONSIDERATIONS**

456 Here, we compared protein profiles of cancer, resection margin,
 457 and control biopsies to investigate whether the resection
 458 margin profiles are more similar to those from cancer or control
 459 biopsies. During this comparison, we pinpointed several key
 460 proteins that have been previously correlated with the disease.
 461 For example, we highlighted several proteins that could be
 462 linked with tumor growth and were found upregulated in the
 463 margin, thus lending support to the soil to seed hypothesis.
 464 While our goal has not been to investigate biomarkers, as this
 465 requires a much larger cohort, our results do nevertheless make
 466 it clear that the resection margin has very specific features that
 467 deserve a better understanding and could aid in the
 468 development of future treatments. Our MDS analysis revealed
 469 limitations in our differential proteomic strategy (which,
 470 incidentally, is adopted in various works for analyzing
 471 differential expression in data). A comparison of shotgun
 472 proteomic profiles without considering independent sample
 473 analyses, by strategies such as our 1D gel analysis or our MDS
 474 algorithm, could lead to missing important information. In our
 475 case, the TFold analysis missed two striking features related to
 476 overexpressed proteins in sample #9. Although these proteins
 477 were also found in other patients, a considerable standard
 478 deviation exists and in turn blinds most common statistical
 479 strategies, including the Venn Diagram complemented by the
 480 TFold analysis that we employed. Nevertheless, our data-
 481 analysis approach enabled us to better investigate the unique
 482 features that explained that subject's status as an outlier. The
 483 take-home lesson is on the importance that should be given to
 484 each individual sample, as each patient may present unique
 485 features. Finally, here we described a complementary approach
 486 to MALDI imaging. Although spatial resolution has been lost as
 487 a consequence, in-depth information for protein identification
 488 was obtained. We conclude that a combination of these
 489 strategies must be further explored to better investigate the role
 490 of the resection margin as well as of any other problematic
 491 tissue that includes spatial data.

492 **ASSOCIATED CONTENT**493 **Supporting Information**

494 Supplementary Figure 1, 1D gel analysis of protein profiles
 495 obtained from cancer (lanes 6, 7, 8, 9) and resection margin
 496 (lanes 10, 11, 12, 13) biopsies; list of proteins pinpointed by
 497 the TFold analyses (zip file); Bend and MudPIT identifications
 498 (Excel files). This material is available free of charge via the
 499 Internet at <http://pubs.acs.org>.

500 **AUTHOR INFORMATION**501 **Corresponding Author**

502 *Address: Rua Prof. Algacyr Munhoz Mader, 3775 CIC 81350-
 503 010 Curitiba/PR, Brazil. Tel.: +55(41)3316-3230. Fax: +55(41)
 504 3316-3267. E-mail: paulo@pcarvalho.com.

505 **Author Contributions**

506 #These authors contributed equally.

507 **Notes**

508 The authors declare no competing financial interest.

509 **ACKNOWLEDGMENTS**

510 The authors thank FAPERJ, CAPES, CNPq, Fiocruz 30/2006-
 511 CAPES, PDTIS, Nova Analitica, and Fundacao do Cancer for
 512 financial support.

513 **REFERENCES**

- (1) Parkin, D. M.; Bray, F.; Ferlay, J.; Pisani, P. Global cancer 514
 statistics, 2002. *CA Cancer J. Clin.* 2005, 2 (55), 74–108. 515
- (2) Ebert, M. P.; Niemeyer, D.; Deininger, S. O.; Wex, T.; Knippig, 516
 C.; Hoffmann, J.; Sauer, J.; Albrecht, W.; Malferteiner, P.; Rocken, C. 517
 Identification and confirmation of increased fibrinopeptide a serum 518
 protein levels in gastric cancer sera by magnet bead assisted MALDI- 519
 TOF mass spectrometry. *J. Proteome Res.* 2006, 9 (5), 2152–2158. 520
- (3) Theuer, C. P.; de, V. C.; Keese, G.; French, S.; Arnell, T.; 521
 Tolmos, J.; Klein, S.; Powers, W.; Oh, T.; Stabile, B. E. Gastric 522
 adenocarcinoma in patients 40 years of age or younger. *Am. J. Surg.* 523
 1996, 5 (172), 473–476. 524
- (4) Shen, J. G.; Cheong, J. H.; Hyung, W. J.; Kim, J.; Choi, S. H.; 525
 Noh, S. H. Influence of a microscopic positive proximal margin in the 526
 treatment of gastric adenocarcinoma of the cardia. *World J.* 527
Gastroenterol. 2006, 24 (12), 3883–3886. 528
- (5) Neoptolemos, J. P.; Stocken, D. D.; Dunn, J. A.; Almond, J.; 529
 Beger, H. G.; Pederzoli, P.; Bassi, C.; Dervenis, C.; Fernandez-Cruz, 530
 L.; Lacaine, F.; Buckels, J.; Deakin, M.; Adab, F. A.; Sutton, R.; Imrie, 531
 C.; Ihse, I.; Tihanyi, T.; Olah, A.; Pedrazzoli, S.; Spooner, D.; Kerr, D. 532
 J.; Friess, H.; Buchler, M. W. Influence of resection margins on survival 533
 for patients with pancreatic cancer treated by adjuvant chemoradiation 534
 and/or chemotherapy in the ESPAC-1 randomized controlled trial. 535
Ann. Surg. 2001, 6 (234), 758–768. 536
- (6) Chaurand, P.; Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Caprioli, 537
 R. M. New developments in profiling and imaging of proteins from 538
 tissue sections by MALDI mass spectrometry. *J. Proteome Res.* 2006, 11 539
 (5), 2889–2900. 540
- (7) Castellino, S.; Groseclose, M. R.; Wagner, D. MALDI imaging 541
 mass spectrometry: bridging biology and chemistry in drug develop- 542
 ment. *Bioanalysis* 2011, 21 (3), 2427–2441. 543
- (8) Oppenheimer, S. R.; Mi, D.; Sanders, M. E.; Caprioli, R. M. 544
 Molecular analysis of tumor margins by MALDI mass spectrometry in 545
 renal carcinoma. *J. Proteome Res.* 2010, 5 (9), 2182–2190. 546
- (9) Cottingham, K. MALDI MS reveals that cells near a tumor aren't 547
 so normal. *J. Proteome Res.* 2010, 5 (9), 2049. 548
- (10) Kim, H. K.; Reyzer, M. L.; Choi, I. J.; Kim, C. G.; Kim, H. S.; 549
 Oshima, A.; Chertov, O.; Colantonio, S.; Fisher, R. J.; Allen, J. L.; 550
 Caprioli, R. M.; Green, J. E. Gastric cancer-specific protein profile 551
 identified using endoscopic biopsy samples via MALDI mass 552
 spectrometry. *J. Proteome Res.* 2010, 8 (9), 4123–4130. 553
- (11) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale 554
 analysis of the yeast proteome by multidimensional protein 555
 identification technology. *Nat. Biotechnol.* 2001, 3 (19), 242–247. 556
- (12) Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Refinements 557
 to label free proteome quantitation: how to deal with peptides shared 558
 by multiple proteins. *Anal. Chem.* 2010, 6 (82), 2272–2281. 559
- (13) Siuzdak, G. The emergence of mass spectrometry in biochemical 560
 research. *Proc. Natl. Acad. Sci. U.S.A.* 1994, 24 (91), 11290–11297. 561
- (14) McDonald, W. H.; Tabb, D. L.; Sadygov, R. G.; MacCoss, M. J.; 562
 Venable, J.; Graumann, J.; Johnson, J. R.; Cociorva, D.; Yates, J. R., III 563
 MS1, MS2, and SQT-three unified, compact, and easily parsed file 564
 formats for the storage of shotgun proteomic spectra and 565
 identifications. *Rapid Commun. Mass Spectrom.* 2004, 18 (18), 566
 2162–2168. 567
- (15) Barboza, R.; Cociorva, D.; Xu, T.; Barbosa, V. C.; Perales, J.; 568
 Valente, R. H.; Franca, F. M.; Yates, J. R., III; Carvalho, P. C. Can the 569
 false-discovery rate be misleading? *Proteomics* 2011, 20 (11), 4105– 570
 4108. 571
- (16) Carvalho, P. C.; Fischer, J. S.; Xu, T.; Cociorva, D.; Balbuena, T. 572
 S.; Valente, R. H.; Perales, J.; Yates, J. R., III; Barbosa, V. C. Search 573
 Engine Processor: filtering and organizing peptide spectrum matches. 574
Proteomics 2012, 12 (7), 944–949. 575
- (17) Cociorva, D.; Carvalho, P. C. Toward objective evaluation of 577
 proteomic algorithms. *Nat. Methods* 2012, 5 (9), 455–456. 578
- (18) Xu, T.; Venable, J. D.; Park, S. K.; Cociorva, D.; Lu, B.; Liao, L.; 579
 Wohlschlegel, J.; Hewel, J.; Yates, J. R., III. ProLuCID, a fast and 580

- 581 sensitive tandem mass spectra-based protein identification program
582 Mol. Cell. Proteomics 2006, S174 (5)
- 583 (19) Carvalho, P. C.; Xu, T.; Han, X.; Cociorva, D.; Barbosa, V. C.;
584 Yates, J. R., III. YADA: A tool for taking the most out of high-
585 resolution spectra. *Bioinformatics* 2009, No. 25, 2734–2736.
- 586 (20) Carvalho, P. C.; Fischer, J. S.; Perales, J.; Yates, J. R.; Barbosa, V.
587 C.; Barenboim, E. Analyzing marginal cases in differential shotgun
588 proteomics. *Bioinformatics* 2011, 2 (27), 275–276.
- 589 (21) Carvalho, P. C.; Yates, J. R., III; Barbosa, V. C. Improving the
590 TFold test for differential shotgun proteomics. *Bioinformatics* 2012, 28
591 (12), 1652–1654.
- 592 (22) Carvalho, P. C.; Fischer, J. S.; Chen, E. I.; Yates, J. R., III;
593 Barbosa, V. C. PatternLab for proteomics: a tool for differential
594 shotgun proteomics. *BMC Bioinf.* 2008, No. 9, 316.
- 595 (23) Oishi, Y.; Kiyohara, Y.; Kubo, M.; Tanaka, K.; Tanizaki, Y.;
596 Ninomiya, T.; Doi, Y.; Shikata, K.; Yonemoto, K.; Shiota, T.;
597 Matsumoto, T.; Iida, M. The serum pepsinogen test as a predictor of
598 gastric cancer: the Hisayama study. *Am. J. Epidemiol.* 2006, 7 (163),
599 629–637.
- 600 (24) Kitahara, F.; Kobayashi, K.; Sato, T.; Kojima, Y.; Araki, T.;
601 Fujino, M. A. Accuracy of screening for gastric cancer using serum
602 pepsinogen concentrations. *Gut* 1999, 5 (44), 693–697.
- 603 (25) Kodama, M.; Koyama, K.; Tsuburaya, Y.; Ishikawa, K.; Koyama,
604 H.; Narisawa, T.; Uesaka, Y. Group I pepsinogen for early detection of
605 gastric cancer recurrence after total gastrectomy. *World J. Surg.* 1990, 1
606 (14), 94–99.
- 607 (26) Zhao, Y.; Zhou, T.; Li, A.; Yao, H.; He, F.; Wang, L.; Si, J. A
608 potential role of collagens expression in distinguishing between
609 premalignant and malignant lesions in stomach. *Anat. Rec.* 2009, 5
610 (292), 692–700.
- 611 (27) Skubitz, A. P. Adhesion molecules. *Cancer Treat. Res.* 2002,
612 No. 107, 305–329.
- 613 (28) Yasui, W.; Oue, N.; Ito, R.; Kuraoka, K.; Nakayama, H. Search
614 for new biomarkers of gastric cancer through serial analysis of gene
615 expression and its clinical implications. *Cancer Sci.* 2004, 5 (95), 385–
616 392.
- 617 (29) Oue, N.; Hamai, Y.; Mitani, Y.; Matsumura, S.; Oshimo, Y.;
618 Aung, P. P.; Kuraoka, K.; Nakayama, H.; Yasui, W. Gene
619 expression
620 profile of gastric carcinoma: identification of genes and tags potentially
621 involved in invasion, metastasis, and carcinogenesis by serial analysis of
622 gene expression. *Cancer Res.* 2004, 7 (64), 2397–2405.
- 623 (30) Kunapuli, S. P.; Singh, H.; Singh, P.; Kumar, A. Ceruloplasmin
624 gene expression in human cancer cells. *Life Sci.* 1987, 23 (40), 2225–
625 2228.
- 626 (31) Scanni, A.; Tomirotti, M.; Licciardello, L.; Annibaldi, E.; Biraghi,
627 M.; Trovato, M.; Fittipaldi, M.; Adamoli, P.; Curtarelli, G. Variations
628 in serum copper and ceruloplasmin levels in advanced gastrointestinal
629 cancer treated with polychemotherapy. *Tumori* 1979, 3 (65), 331–
630 338.
- 631 (32) Storr, S. J.; Mohammed, R. A.; Woolston, C. M.; Green, A. R.;
632 Parr, T.; Spiteri, I.; Caldas, C.; Ball, G. R.; Ellis, I. O.; Martin, S. G.
633 Calpastatin is associated with lymphovascular invasion in breast
634 cancer. *Breast* 2011, 5 (20), 413–418.
- 635 (33) Chan, A. O. E-cadherin in gastric cancer. *World J. Gastroenterol.*
636 2006, 2 (12), 199–203.
- 637 (34) Debruyne, P.; Vermeulen, S.; Mareel, M. The role of the E-
638 cadherin/catenin complex in gastrointestinal cancer. *Acta Gastro-
639 enterol. Belg.* 1999, 4 (62), 393–402.
- 640 (35) Yang, Y.; Liu, Y.; Yao, X.; Ping, Y.; Jiang, T.; Liu, Q.; Xu, S.;
641 Huang, J.; Mou, H.; Gong, W.; Chen, K.; Bian, X.; Wang, J. M.
642 Annexin 1 released by necrotic human glioblastoma cells stimulates
643 tumor cell growth through the formyl peptide receptor 1. *Am. J. Pathol.*
644 2011, 3 (179), 1504–1512.
- 645 (36) Kang, W. Y.; Chen, W. T.; Huang, Y. C.; Su, Y. C.; Chai, C. Y.
646 Overexpression of annexin 1 in the development and differentiation of
647 urothelial carcinoma. *Kaohsiung J. Med. Sci.* 2012, 3 (28), 145–150.
- 648 (37) Byrne, J. A.; Maleki, S.; Hardy, J. R.; Gloss, B. S.; Murali, R.;
649 Scurry, J. P.; Fanayan, S.; Emmanuel, C.; Hacker, N. F.; Sutherland, R.
650 L.; Defazio, A.; O'Brien, P. M. MAL2 and tumor protein D52
(TPD52) are frequently overexpressed in ovarian carcinoma, but
651 differentially associated with histological subtype and patient outcome. *BMC
652 Cancer* 2010, No. 10, 497.
- 653 (38) Wang, R.; Xu, J.; Saramaki, O.; Visakorpi, T.; Sutherland, W. M.;
654 Zhou, J.; Sen, B.; Lim, S. D.; Mabeesh, N.; Amin, M.; Dong, J. T.;
655 Petros, J. A.; Nelson, P. S.; Marshall, F. F.; Zhou, H. E.; Chung, L. W.
656 PrLZ, a novel prostate-specific and androgen-responsive gene of the
657 TPD52 family, amplified in chromosome 8q21.1 and overexpressed in
658 human prostate cancer. *Cancer Res.* 2004, 5 (64), 1589–1594.
- 659 (39) Punchedewa, C.; Inoue, A.; Hishiki, A.; Fujikawa, Y.; Connelly,
660 M.; Evison, B.; Shao, Y.; Heath, R.; Kuraoka, I.; Rodrigues, P.;
661 Hashimoto, H.; Kawanishi, M.; Sato, M.; Yagi, T.; Fujii, N. 661
662 Identification of small molecule proliferating cell nuclear antigen
663 (PCNA) inhibitor that disrupts interactions with PIP-box proteins and
664 inhibits DNA replication. *J. Biol. Chem.* 2012, 17 (287), 14289–14300.
- 665 (40) Konno, S.; Takebayashi, Y.; Aiba, M.; Akiyama, S.; Ogawa, K.
666 Clinicopathological and prognostic significance of thymidine phos-
667 phorylase and proliferating cell nuclear antigen in gastric carcinoma.
668 *Cancer Lett.* 2001, 1 (166), 103–111.
- 669 (41) Fischer, J. S.; Carvalho, P. C.; Neves-Ferreira, A. G.; da Fonseca,
670 C. O.; Perales, J.; Carvalho, M. G.; Domont, G. B. Anti-thrombin as a
671 prognostic biomarker candidate for patients with recurrent glioblas-
672 toma multiform under treatment with perillyl alcohol. *J. Exp. Ther.*
673 *Oncol.* 2008, 4 (7), 285–290.
- 674 (42) Nejjar, M.; Anderson, W.; Pourreya, C.; Jacquier, M. F.;
675 Scoazec, J. Y.; Remy, L. The role of fibroblasts in the modulation of
676 integrin-dependent interactions between the gastric cell line HGT-1
677 and fibronectin. *Int. J. Cancer* 2004, 4 (112), 560–569.
- 678 (43) Twal, W. O.; Czirok, A.; Hegedus, B.; Knaak, C.; Chintalapudi,
679 M. R.; Okagawa, H.; Sugi, Y.; Argraves, W. S. Fibulin-1 suppression of
680 fibronectin-regulated cell adhesion and motility. *J. Cell Sci.* 2001, Pt 24
681 (114), 4587–4598.
- 682 (44) Argraves, W. S.; Tran, H.; Burgess, W. H.; Dickerson, K. Fibulin
683 is an extracellular matrix and plasma glycoprotein with repeated
684 domain structure. *J. Cell Biol.* 1990, 6 Pt 2 (111), 3155–3164.
- 685 (45) Greene, L. M.; Twal, W. O.; Duffy, M. J.; McDermott, E. W.;
686 Hill, A. D.; O'Higgins, N. J.; McCann, A. H.; Dervan, P. A.; Argraves,
687 W. S.; Gallagher, W. M. Elevated expression and altered processing of
688 fibulin-1 protein in human breast cancer. *Br. J. Cancer* 2003, 6 (88), 689
690 871–878.
- 691 (46) Wlazlinski, A.; Engers, R.; Hoffmann, M. J.; Hader, C.; Jung, V.;
692 Muller, M.; Schulz, W. A. Downregulation of several fibulin genes in
693 prostate cancer. *Prostate* 2007, 16 (67), 1770–1780.
- 694 (47) Zhu, W. Y.; Hunag, Y. Y.; Liu, X. G.; He, J. Y.; Chen, D. D.;
695 Zeng, F.; Zhou, J. H.; Zhang, Y. K. Prognostic evaluation of CapG,
696 gelsolin, P-gp, GSTP1, and Topo-II proteins in non-small cell lung
697 cancer. *Anat. Rec.* 2012, 2 (295), 208–214.
- 698 (48) Assinder, S.; Cole, N. Does TGF-beta induced formation of
699 actin stress fibres reinforce Smad dependent TGF-beta signalling in
700 the prostate? *Med. Hypotheses* 2011, 6 (76), 802–804.