SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
GASTRIC CANCER
THERMODYNAMICS

Received
1 October 2014

Accepted
29 December 2014

Published
23 January 2015

Correspondence and requests for materials should be addressed to A.V.-C. (adrianvc@unizar.es) or O.A. (oabifra@unizar.es)

Deconvolution Analysis for Classifying Gastric Adenocarcinoma Patients Based on Differential Scanning Calorimetry Serum Thermograms

Sonia Vega[1], María Asuncion Garcia-Gonzalez[2,3,4], Angel Lanas[3,4,5,6], Adrian Velazquez-Campoy[1,7,8] & Olga Abian[1,2,3,4]

[1]Institute of Biocomputation and Physics of Complex Systems (BIFI), Joint-Unit IQFR-CSIC-BIFI, University of Zaragoza, Zaragoza, Spain, [2]Instituto Aragonés de Ciencias de la Salud (IACS), [3]IIS Aragón, Zaragoza, Spain, [4]Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd), Spain, [5]Servicio de Aparato Digestivo. Hospital Clínico Universitario LB, Spain, [6]Department of Medicine. University of Zaragoza, Spain, [7]Department of Biochemistry and Molecular and Cell Biology, University of Zaragoza, Zaragoza, Spain, [8]Fundacion ARAID, Government of Aragon, Zaragoza 50018, Spain.

Recently, differential scanning calorimetry (DSC) has been acknowledged as a novel tool for diagnosing and monitoring several diseases. This highly sensitive technique has been traditionally used to study thermally induced protein folding/unfolding transitions. In previous research papers, DSC profiles from blood samples of patients were analyzed and they exhibited marked differences in the thermal denaturation profile. Thus, we investigated the use of this novel technology in blood serum samples from 25 healthy subjects and 30 patients with gastric adenocarcinoma (GAC) at different stages of tumor development with a new multiparametric approach. The analysis of the calorimetric profiles of blood serum from GAC patients allowed us to discriminate three stages of cancer development (I to III) from those of healthy individuals. After a multiparametric analysis, a classification of blood serum DSC parameters from patients with GAC is proposed. Certain parameters exhibited significant differences ($P < 0.05$) and allowed the discrimination of healthy subjects/patients from patients at different tumor stages. The results of this work validate DSC as a novel technique for GAC patient classification and staging, and offer new graphical tools and value ranges for the acquired parameters in order to discriminate healthy from diseased subjects with increased disease burden.

The precise determination of the thermally-induced conformational transitions of biomolecules can be performed using DSC (Differential Scanning Calorimetry) due to its high sensitivity. In particular, the thermodynamic parameters of protein thermal denaturation (unfolding) can be determined directly by this technique[1]. In 2007, Chaires and co-workers proposed DSC technique as a potential tool for disease diagnosis and monitoring through the analysis of blood plasma from patients[2–5]. The DSC thermogram of blood plasma from healthy subjects contrasted with those from patients with different diseases (from inflammatory to oncology pathologies)[2–10]. Taneva and co-workers confirmed these preliminary studies and reported new data revealing marked multiple myeloma-induced modifications in blood serum thermograms[11].They also reported colorectal cancer-specific alterations in the thermal response of blood plasma proteome[12]. All these works have contributed to the validation of DSC as a potential non-invasive tool for diagnosing and discriminating several malignancies.

The underlying hypotheses in applying DSC in clinical diagnosis are: 1) the thermogram acquired from the thermal denaturation reflects the complex protein and metabolite composition of the plasma sample (metabolites may not undergo conformational transitions, but they can interact with proteins modulating their thermal stability)[2–5]; and 2) pathologies and disorders trigger alterations in protein and metabolite composition in plasma (up- or down-regulation of specific proteins and the presence/absence of metabolites specifically related to the disease), which will be mirrored in distorted thermally-induced conformational transitions and, therefore, distorted thermograms when compared to those from healthy subjects. One of the main advantages of using DSC with plasma samples is that a minimally invasive assay such as a routine blood test analysis could help to: 1)

diagnose the disease at an early stage; 2) monitor the remission of the disease or relapse in treated patients; and 3) anticipate the decision making process during medical treatment by predicting the evolution of the disease.

DSC blood plasma analysis has been applied to different cancer patients[6–12] and different profile patterns have been determined. With a final goal of implementing and including DSC tests within routine clinical analyses of patients with different diseases, there are some important specific requirements that need to be fulfilled and considered before the test could be used in clinical practice (even before sensitivity, specificity, precision and accuracy of the test could be assessed). Patients in the study should be precisely characterized and classified in order to minimize errors in defining the profile parameters for a certain disease. In addition, it is necessary to optimize experimental protocols and data analysis methodologies in order to avoid or minimize potential errors (e.g. determination of protein concentration). Furthermore, it is necessary to develop and implement a quantitative methodology for a multiparametric data analysis able to capture the different features between healthy and unhealthy individuals, as well as provide numerical ranges to discriminate the key parameters.

Gastric adenocarcinoma (GAC) ranks as the fourth most common cancer and the second most frequent cause of cancer deaths worldwide[13]. GAC is well known to be a heterogeneous and complex disease, and it is noteworthy that distinct clinical, epidemiological, and molecular features have been reported among tumors arising from the cardia or non-cardia region within the stomach, and among intestinal and diffuse histological subtypes[14,15]. These phenotypic differences are determined by the combined effects of multiple environmental and host risk factors. Hence, the main goal of this work is to validate the data analysis method developed in our group and its application in the classification of a group of patients with different GAC stages.

Methods

Subjects. Consecutive Spanish Caucasian patients with primary GAC identified by endoscopic and pathological diagnosis at the Hospital Clínico Universitario Lozano Blesa in Zaragoza, Spain, from 2010 to 2011 were invited to take part in the study. A total of 30 GAC patients were initially selected as cases. Gastric tumors were grouped according to their anatomical location as cardia GAC (located at the gastroesophageal junction) and non-cardia or distal GAC. Moreover, non-cardia GACs were classified according to the histological type as intestinal, diffuse, or undetermined[16]. Patients with local recurrence of GAC, non-adenocarcinoma histology, previous history of other malignancies, absence of blood samples, or refusal to participate in the study were considered non-eligible. At the time of inclusion, detailed information was recorded concerning age, gender, smoking habits, tumor-node-metastasis stage (TNM stage) according to the Union for International Cancer Control/American Joint Committee on Cancer (UICC/AJCC) classification, presence of metastases, tumour location, and histological subtype.

The control group consisted of 25 sex- and age- (±5 years) matched Spanish Caucasian community volunteers apparently cancer-free, with no previous history of gastric disease, recruited from the out-patient clinical services at the hospital. Individuals with evidence for past or present gastric ulcer, immunosuppressive disorders, and major systemic diseases were excluded.

Approximately 10 mL of peripheral blood from each patient and control subject were collected into serum separator tubes for subsequent DSC analysis. Once processed, 200 μL serum samples were aliquoted and stored at −80°C until analysis. All participants gave written informed consent to the study protocol, which was previously approved and conducted in accordance with the Ethical Review Board for Clinical Research of the Regional Government (CEIC Aragon). All experimental protocols were approved by CEIC Aragon. All experiments were carried out in accordance with the approved guidelines.

Protein concentration determination. Serum protein concentration was measured by Bradford protein assay (Bio-Rad) using purified bovine serum albumin (BSA) 100× (10 mg/mL, New England BioLabs) in phosphate buffered saline (PBS) as standard. Absorbance at 595 nm of two dilutions from each serum sample (1 : 17000 and 1 : 11333) was measured in triplicate in a Synergy HT multimode microplate reader (BioTek Instruments).

Thermograms were not normalized according to the total protein concentration. The multiparametric analysis reported in this work employs a final set of parameters independent of sample protein concentration. Serum is a complex mixture of proteins, and, therefore, total protein concentration determined by colorimetric methods is considerably affected by inherent uncertainties.

Differential Scanning Calorimetry (DSC). The heat capacity of serum samples was measured as a function of temperature, $C_P(T)$, using a high-sensitivity differential scanning VP-DSC microcalorimeter (MicroCal, Northampton, MA). Serum samples and reference solutions were properly degassed and carefully loaded into the cells to avoid bubble formation. The baseline of the instrument was routinely recorded before experiments. Experiments were performed in diluted serum samples (1 : 25 in PBS) at a scanning rate of 1°C/min. No precipitation/aggregation occurred during thermal denaturation. Thermograms were baseline-corrected and analyzed using software developed in our laboratory implemented in Origin 7 (OriginLab).

Data Analysis. We have developed a phenomenological model in which the complex thermogram was deconvoluted in several individual transitions (peaks), modelling each individual transition by the logistic peak or Hubbert function:

$$C_P(T) = C_{P,0} + \frac{4A\ exp\left(-\frac{T-T_c}{w}\right)}{\left(1+exp\left(-\frac{T-T_c}{w}\right)\right)^2} \quad (1)$$

where A is the height of the peak (equivalent to the maximal unfolding heat capacity $C_{P,max}$), T_c is the center of the peak (equivalent to the mid-transition temperature T_m), and w is the width of the peak ($C_P(T_c \pm w) \approx 0.8\ A$ and $C_P(T_c \pm 2\ w) \approx 0.4\ A$). The offset parameter $C_{P,0}$ (found to be always very close to 0 in the experimental data analysis), was included as an adjustable parameter to counterbalance errors from baseline correction. This function is able to reproduce accurately a two-state protein unfolding curve even with a protein concentration normalization error, or when a stabilizing interacting ligand is present. Moreover, it is a simple function and it reproduces protein unfolding data much better than other similar curves (see Results).

From our experience, a minimum set of six individual curves was necessary for reproducing the serum thermograms:

$$C_P(T) = C_{P,0} + \sum_{i=1}^{6} \frac{4A_i\ exp\left(-\frac{T-T_{c,i}}{w_i}\right)}{\left(1+exp\left(-\frac{T-T_{c,i}}{w_i}\right)\right)^2} \quad (2)$$

Adding more terms in Eq. 2 does not improve the analysis; on the contrary, the function gets over-parameterized and degeneracy and correlation among the parameters arises during non-linear fitting analysis.

Therefore, for any given serum thermogram eighteen parameters (A_i, $T_{c,i}$ and w_i, for each of the six individual transitions), were obtained after data analysis. This set of parameters constitutes the basis for the multiparametric comparative quantitative methodology aimed at establishing classification criteria among healthy subjects and GAC patients. Polar or polygonal plots, constructed with the three sets of parameters (A_i, $T_{c,i}$ and w_i, with i = 1 to 6), can be used as a graphical tool for classification (see Results).

Other useful statistical parameters can be defined for each thermogram, as explained in the Results section. In particular, the area under the curve (AUC), and the average temperature or the first moment of the thermogram, T_{av}, defined according to the following expressions:

$$AUC = \sum_j C_P(T_j)$$
$$T_{av} = \frac{\sum_j C_P(T_j) T_j}{\sum_j C_P(T_j)} \quad (3)$$

where j runs over the entire range of experimental points in the thermogram. The deconvolution of $C_P(T)$ into six individual logistic peak curves (Eq. 2) allows the analytical calculation of AUC:

$$AUC = \sum_{i=1}^{6} 4A_i w_i \quad (4)$$

The height of the second peak, A_2, was selected as a normalizing factor. The reasons for applying this normalization were based on the fact that A_2 is one of the main differential features observed among thermograms (see Results), and that normalizing with this peak makes protein concentration normalization unnecessary along the data analysis. Thus, the normalized area under the curve, AUC_n, and the normalized heights, $A_{i,n}$, were defined as:

$$AUC_n = \sum_j \frac{C_P(T_j)}{A_2}$$
$$A_{i,n} = \frac{A_i}{A_2} \quad (5)$$

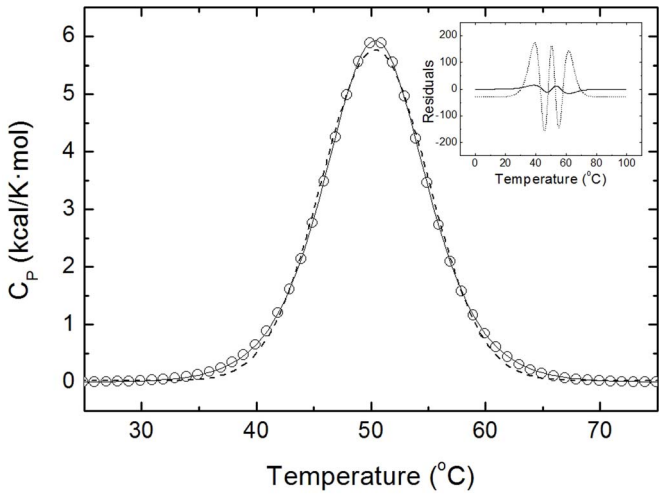

Figure 1 | A two-state protein unfolding curve was simulated corresponding to a mid-transition temperature of 50°C, an unfolding enthalpy of 70 kcal/mol, and an unfolding heat capacity change of 1.5 kcal/K·mol (open circles). Non-linear fitting analysis using the Hubbert function (continuous line) provides an excellent fit of the simulated data ($R^2 = 0.99998$, $\chi^2 = 189$); the Gaussian function (discontinuous line) provides an acceptable fit only ($R^2 = 0.99792$, $\chi^2 = 18980$). Inset: Residual plot for each fitting analysis.

It is important to indicate that only the parameter A_i would be dependent on protein concentration (i.e. it would be affected by protein concentration normalization uncertainties), while $T_{c,i}$ and w_i are independent of normalizing factors (i.e. concentration- or scale-independent). Additional protein concentration-independent parameters for classifying GAC serum thermograms (area of the normalized heights polygon, AP_n; skewness of the thermogram, g_1) will be defined in the Results section.

Results

Deconvolution of thermograms into individual transitions. First of all, we assessed the appropriateness of the logistic peak function (Eq. 1) to reproduce protein thermal unfolding curves. Several types of curves were simulated using conventional protein unfolding models (two-state unfolding, three-state unfolding, and ligand dissociation coupled unfolding models). An excellent fit was obtained in all cases. Using other peak-shaped functions (e.g. Gaussian curve) lead to significantly less satisfactory results. Figure 1 shows a representative result from this test.

Next, the thermograms from serum samples were analyzed using Eq. 2. As anticipated, the combination of six logistic peak curves was able to successfully reproduce the experimental curves. Figure 2 shows typical thermograms and their deconvolution into individual thermal transitions applying the phenomenological model. It can be clearly seen that the global thermal behavior of serum samples, reflecting its protein and metabolite composition, is captured by the deconvolution procedure. Furthermore, individual traits, corresponding to individual transitions that in principle could be identified with major proteins in serum plasma[4], can be also observed. For example, peaks 2, 3 and 4 constitute the main transitions in healthy subjects; however, peak 2 is largely attenuated in GAC patients (Figure 2). Thus, peak 2 is, by far, one of the main discriminating features between subjects, and it was selected as a normalization factor.

The height of a given peak is related to the concentration and unfolding enthalpy of the protein components associated with that peak. If a certain metabolite interacts with any of the protein components responsible for that peak, it will also affect the height and the center of that peak. Thus, the changes observed between thermograms from healthy and diseased subjects reflect the interplay between disease-specific protein up-/down-regulation (affecting the concentration and, therefore, the heights, A_i, of the different peaks) and the presence of disease-specific metabolites (potentially interacting with serum proteins and affecting the centers, $T_{c,i}$, of the peaks; stabilizing ligands induce increments in $T_{c,i}$, whereas destabilizing ligands induce decrements in $T_{c,i}$). Therefore, in principle it is not straightforward to assign and explain the specific changes observed between two given thermograms in terms of serum composition and components interactions, and a phenomenological approach is more convenient. In general, it was found that the heights of the individual peaks showed larger modifications between thermograms than the corresponding widths and centers. Yet, as mentioned above, this fact does not mean that the differences were due only or mainly to changes in the concentration of specific proteins.

Graphical multiparametric thermogram comparison between subjects. Next, we constructed three parameter-specific polar or polygonal plots for each thermogram. Using the six heights A_i, six centers $T_{c,i}$ and six widths w_i associated with the individual

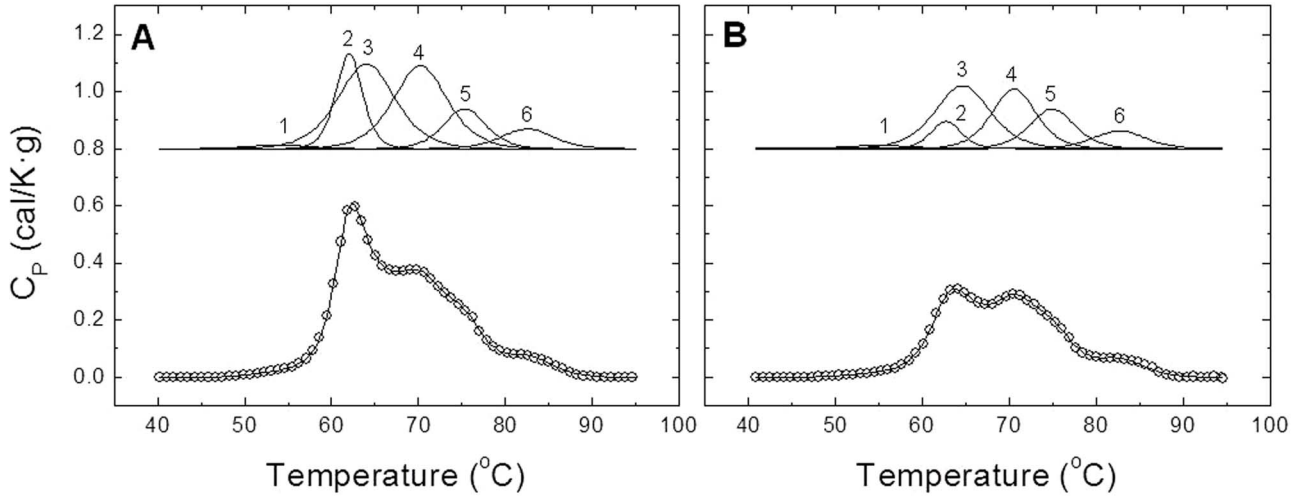

Figure 2 | Experimental serum thermograms from a healthy subject (A) and a patient with stage I gastric adenocarcinoma (B). (Bottom plot) Global thermograms showing the experimental points (open circles; one of every three experimental points is shown for clarity) and the fitting curve (continuous line); (Top plot) Deconvolution of the global thermogram showing the individual numbered transitions.

transitions in a given thermogram, three irregular hexagons were plotted in a way that the vertices were situated from the geometric center at a distance equal to a given parameter value. Figure 3 shows the graphical comparison between some of the healthy subjects and some of the gastric cancer patients at different tumour stages (stages I to III). Because errors in protein concentration normalization would only alter the heights of the peaks, normalized heights, according to Eq. 5, have been plotted.

It can be observed from these polygonal plots that the normalized heights A_i/A_2 show larger inter-group variability than the centers $T_{c,i}$ or the widths w_i, and that the normalized heights increase with the TNM stage of the GAC. In addition, there was some intra-group variability regarding the heights A_i/A_2 and the widths w_i.

Global and local geometric thermogram parameters for classifying subjects. Two central tendency parameters were calculated from each thermogram: the area under the curve (AUC) and the first moment or average temperature of the thermogram (T_{av}). Because AUC might be affected by protein concentration normalization errors, AUC was normalized according to Eq. 5. T_{av} was slightly different from the median temperature (temperature value dividing the thermogram into two equal halves), but it was always very close and usually both differed in much less than 1°C.

From Figure 4A it is apparent that T_{av} increases with AUC_n and, importantly, the different patient groups cluster in well-defined regions (Figure 4A). Therefore, AUC_n is one of the main explicative parameters for classifying patients.

Observing the normalized height polygonal plot, it is possible to summarize the observed increment in normalized heights according to the severity of the disease in a single parameter: the area of the normalized height polygon, AP_n, which can be expressed analytically as:

$$AP_n = \sum_{i=1}^{6} \frac{\sqrt{3}}{4} \frac{A_i A_{i+1}}{A_2^2} \quad with\ A_7 = A_1 \tag{6}$$

It can be clearly seen that AP_n increases with AUC_n and, again, the different patient groups cluster in well-defined regions (Figure 4B). It is important to point out that such trend is not present if the area of the non-normalized height polygon is employed, even if normalized by protein concentration (not shown).

Other global and local geometric parameters, such as those obtained by normalizing by A_3, A_4 and A_5, were evaluated and plotted in order to explore other possibilities; however, no clear trend and patient clustering were observed. In particular, normalizing AUC according to A_3, another important thermogram transition, did not improve the results (not shown). Since thermograms corresponding to GAC patients showed a marked attenuation in the second individual transition compared to the healthy subjects, skewness g_1 and kurtosis g_2 were calculated for the thermograms:

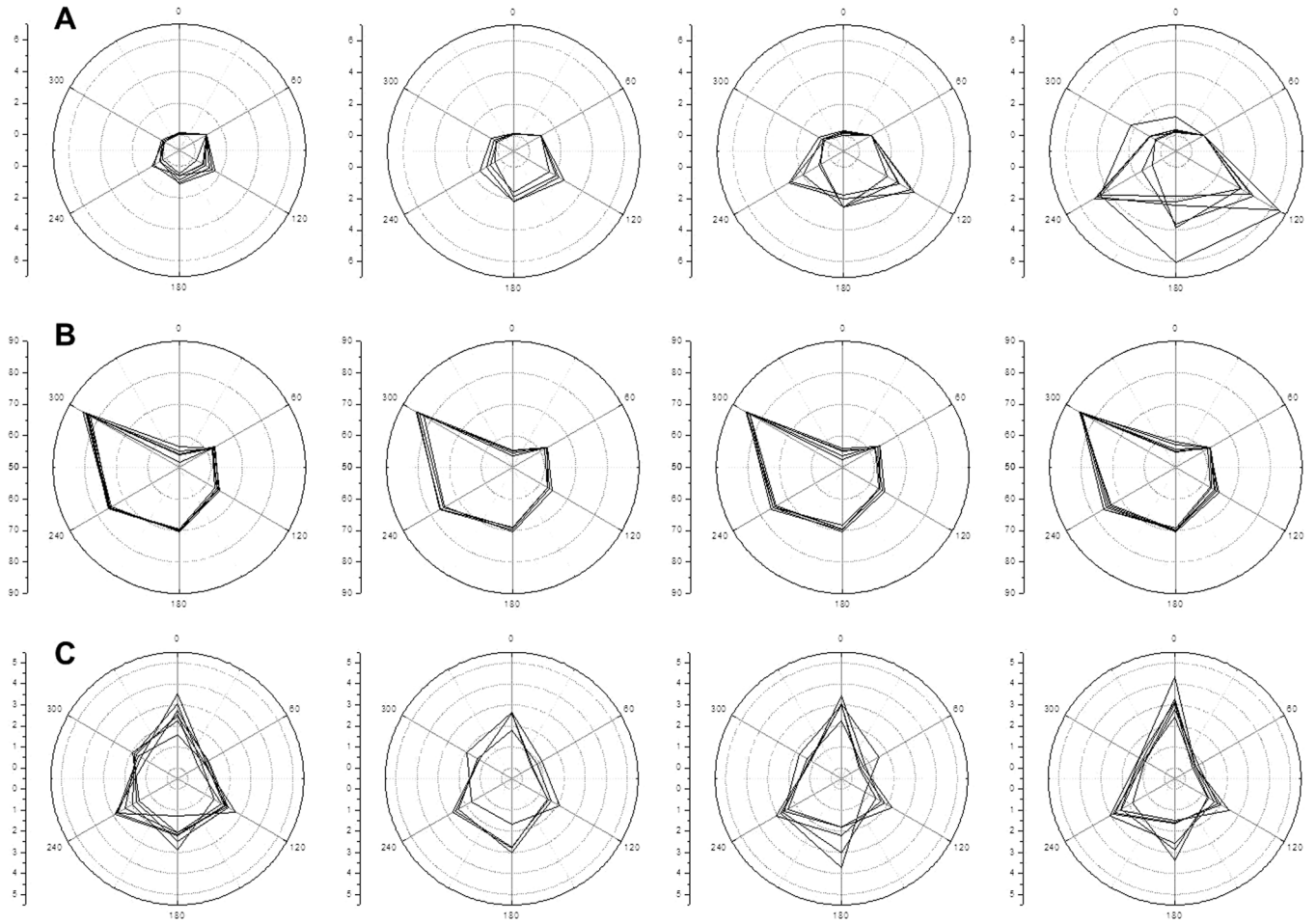

Figure 3 | Polygonal plots for normalized heights A_i/A_2 (top), centers $T_{c,i}$ (middle) and widths w_i (bottom) for selected patients corresponding to the four subject groups: healthy subjects (first column), stage I-GAC patients (second column), stage II-GAC patients (third column), and stage-III GAC patients (fourth column). Each vertex corresponds to a given thermal transition, starting from 0° (first transition) to 360° (sixth transition). The distance from the center to any vertex is equal to the corresponding value of the parameter.

$$m_k = \frac{\sum_j C_P(T_j)\,(T_j - T_{av})^k}{\sum_j C_P(T_j)}$$

$$g_1 = \frac{m_3}{m_2^{3/2}} \qquad (7)$$

$$g_2 = \frac{m_4}{m_2^2}$$

and only g_1 showed a clear trend and subject clustering according to severity of the disease (Figure 4C).

The results shown previously can be summarized providing the value ranges for the main global and local discriminating parameters (Table 1, Figure 5). According to a parametric (two-sample t-Student test) and a non-parametric hypothesis test (Mann-Whitney test), some differences are statistically significant ($P < 0.05$), as indicated in Figure 5.

The main analysis in this study was carried out comparing thermograms obtained from healthy individuals and thermograms obtained from GAC patients stratified according to the stage of the tumour (TNM stages I to III). However, analysis according other categories such as localization of the tumour (cardia/non-cardia), histopathological type (intestinal/diffuse) or status of infection with *Helicobacter pylori*, did not reveal any clear trend or defined patient clustering (data not shown).

Discussion

Differential scanning calorimetry (DSC) has been recently proven to be a suitable technique for detecting differences among blood plasma proteome/interactome in healthy individuals and patients from a variety of pathologies[6–12]. Hence, this technique offers a rapid, inexpensive, non-invasive, easy procedure to be included as a complementary tool in standard clinical tests with a direct application in patient staging and classification, as well as in screening and monitoring risk-related subpopulations or groups of patients under surveillance.

The simplicity of the practical procedure for obtaining the thermogram from blood serum/plasma is counterbalanced by difficulties in the interpretation of the data and in the extraction of useful information. Blood serum/plasma is a complex system and a realistic data analysis of the thermogram based on biophysical models on protein unfolding and its coupling with ligand interactions is unfeasible. Up to now only qualitative or semiquantitative data analysis procedures based on evaluating metric-associated thermogram similarity[4,11,12] or comparing apparent geometric features in the thermogram[17,18] have been developed and applied with certain, but limited, success. Therefore, there is a need for new data analysis methodologies providing graphical or numerical comparison tools, as well as practical numerical ranges for thermogram classification. Of course, other experimental and/or clinical techniques can be employed in parallel with DSC in order to obtain additional useful information from blood serum/plasma; however the full capability of DSC as an analytical technique to extract information from blood serum/plasma has not been explored in depth yet.

In this report we studied serum samples from healthy controls and patients with GAC at different TNM stages. The main goal was to develop a robust data analysis methodology based on a phenomenological model that allows a detailed description of the serum/plasma thermogram and able to reproduce the experimental data, providing graphical comparison tools, as well as parameter value ranges for patient classification.

The thermograms from healthy subjects were clearly different from those corresponding to GAC patients. Because the serum is a complex mixture of many proteins and metabolites (many of them interacting with proteins), the development of a model for data analysis based on well-known realistic thermally-induced conformational transition model in proteins is unfeasible: 1) although the behavior of plasma can be accounted for by about a dozen of proteins[2–5], using realistic models for protein unfolding (e.g. two-state and three-state unfolding mechanisms) relies on the precise concentration determination for individual proteins, whereas in our case only total concentration of protein can be determined; 2) accounting for the modulating effect of protein-interacting metabolites on protein thermal stability is not viable, because information about metabolite identities and concentrations, as well as binding affinities and enthalpies would be required. The identification of proteins and

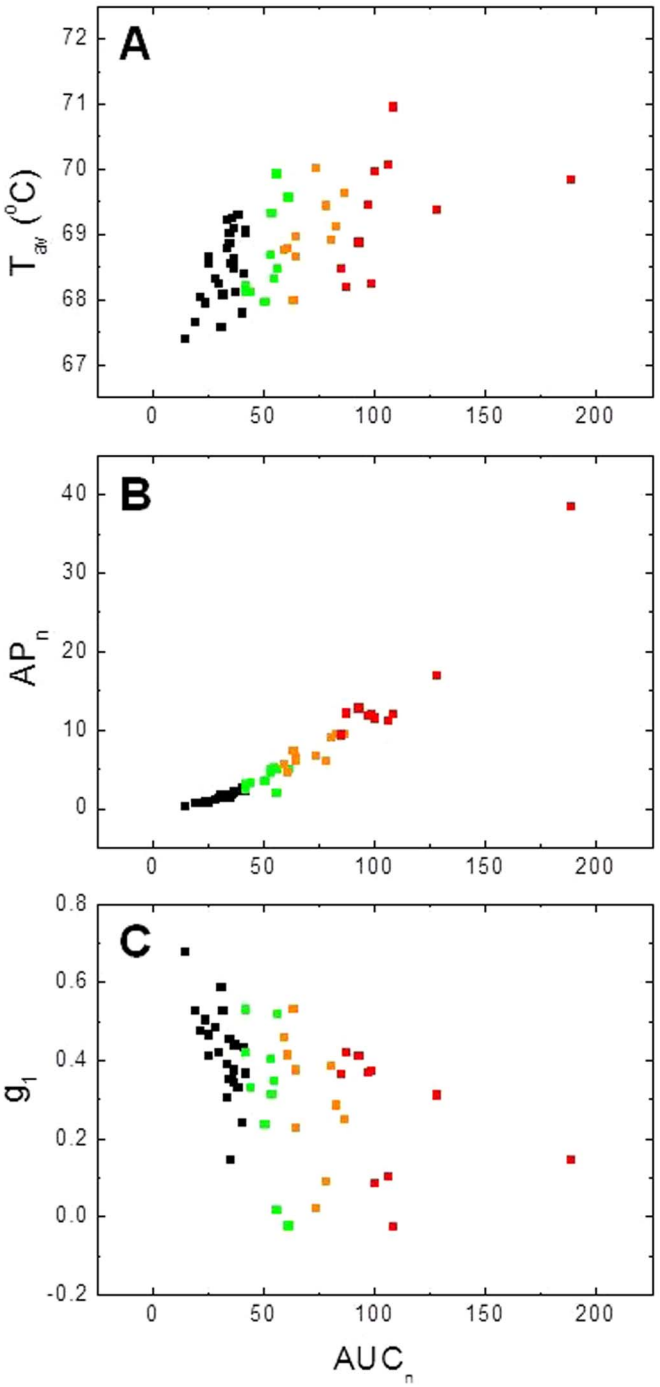

Figure 4 | Global and local parameters from the serum thermograms. (A) T_{av} vs. AUC_n; (B) AP_n vs. AUC_n; (C) g_1 vs. AUC_n. The different subject groups are colour coded: healthy subjects (black), stage I GAC patients (green), stage II GAC patients (orange), stage III GAC patients (red).

Table 1 | Mean values and standard errors for the main discriminating parameters for classifying GAC patients from DSC thermograms

	Healthy	Stage I	Stage II	Stage III
AUC_n	28 ± 5	46 ± 6	66 ± 9	100 ± 20
AP_n	1.3 ± 0.5	4 ± 1	6 ± 1	14 ± 5
T_{av}	68.1 ± 0.2	68.5 ± 0.3	69.2 ± 0.3	69.5 ± 0.4
g_1	0.14 ± 0.04	0.05 ± 0.01	0.03 ± 0.01	0.02 ± 0.01

metabolites responsible for the observed thermogram changes would need the involvement of additional experimental techniques (e.g. gel electrophoresis or mass spectrometry)[19–22]. Nevertheless, our main goal was to explore the possibilities offered by DSC alone and the multiparametric analysis that can be performed by using a phenomenological model.

The phenomenological model employed in this work allows the decomposition of the thermogram into individual thermal transitions, each one characterized by three parameters (height, center and width). Among the different peak-associated curves that were tested, the logistic peak curve or Hubbert curve was the most appropriate for reproducing the data, as confirmed by fitting analysis of protein unfolding curves. Although the individual transitions in the thermogram reflect the unfolding of major protein components in blood plasma and their interaction with metabolites, there is no need for identifying such components.

Global geometric traits (area under the curve AUC, average temperature T_{av}, height- polygon area AP, skewness g_1) together with local geometric traits (parameters from individual transitions: A_i, $T_{c,i}$, w_i) can be employed in order to establish differential features among thermograms and define clustering regions for patients belonging to a given category (healthy, TNM stage I GAC, stage II or stage III). Using global and local geometric traits at the same time helps in overcoming the typical drawbacks of using those separately (either too many details and parameters hiding the essential information, or too few details and parameters loosing essential information).

In particular, the deconvolution of the thermogram into individual components allows a rapid quantitative and qualitative inspection of the main differences between two given thermograms. In addition, polygonal or polar plots provide a graphical tool for a quick review of the main differences among the thermograms and the identification of differential features leading to clustering of patients according to given categories. In the polygonal plots, differences in the individual transitions from two given thermograms are quickly detected as differences in the size and/or the shape (biases or deviations) of the corresponding polygons.

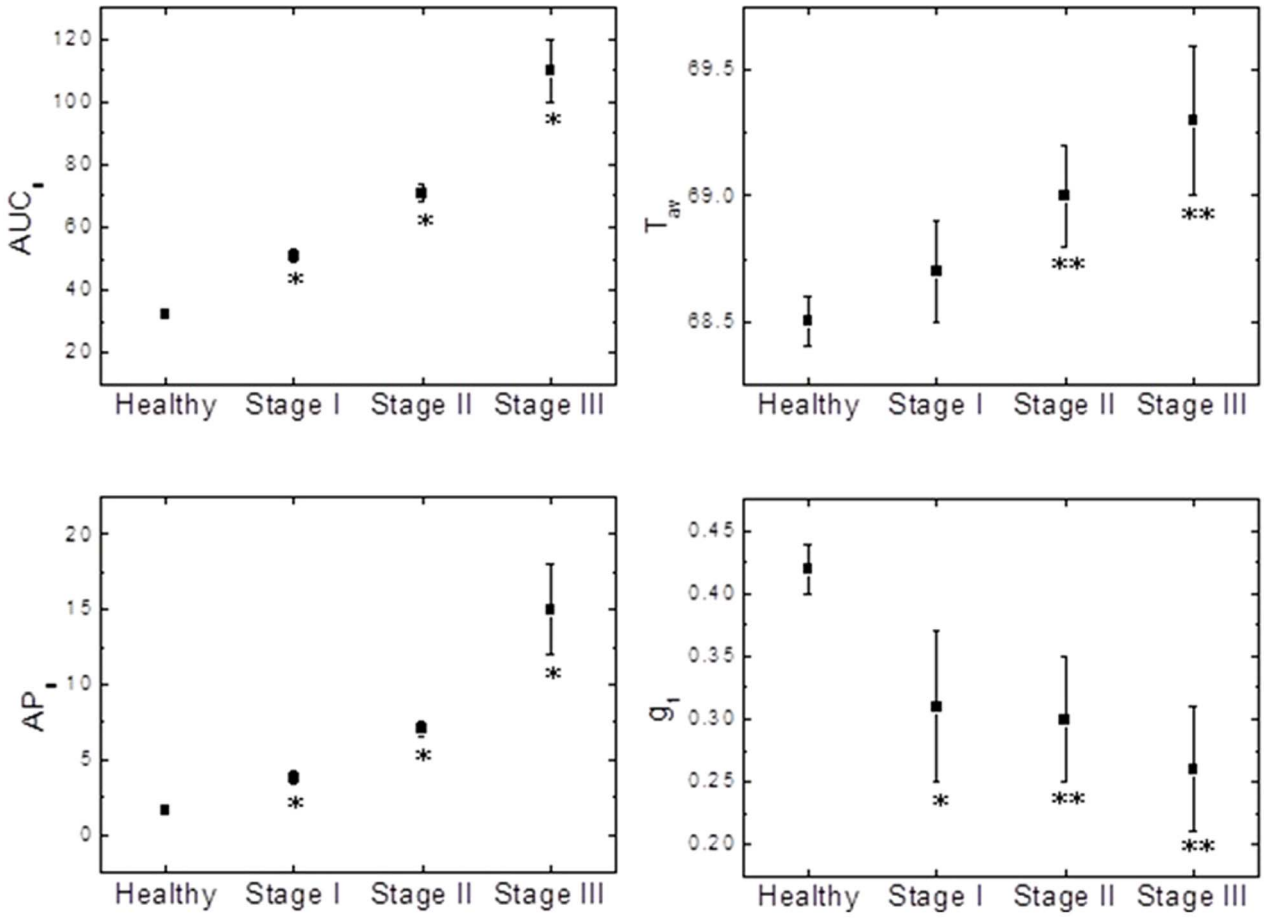

Figure 5 | Mean values (±standard error) for the parameters obtained from the serum thermograms in the different categories. The single asterisk indicates differences statistically significant ($P < 0.05$) regarding the previous subject group. The double asterisk indicates differences statistically significant ($P < 0.05$) regarding the healthy subjects group. Two-sample t-Student (parametric) and Mann-Whitney (non-parametric) tests were employed.

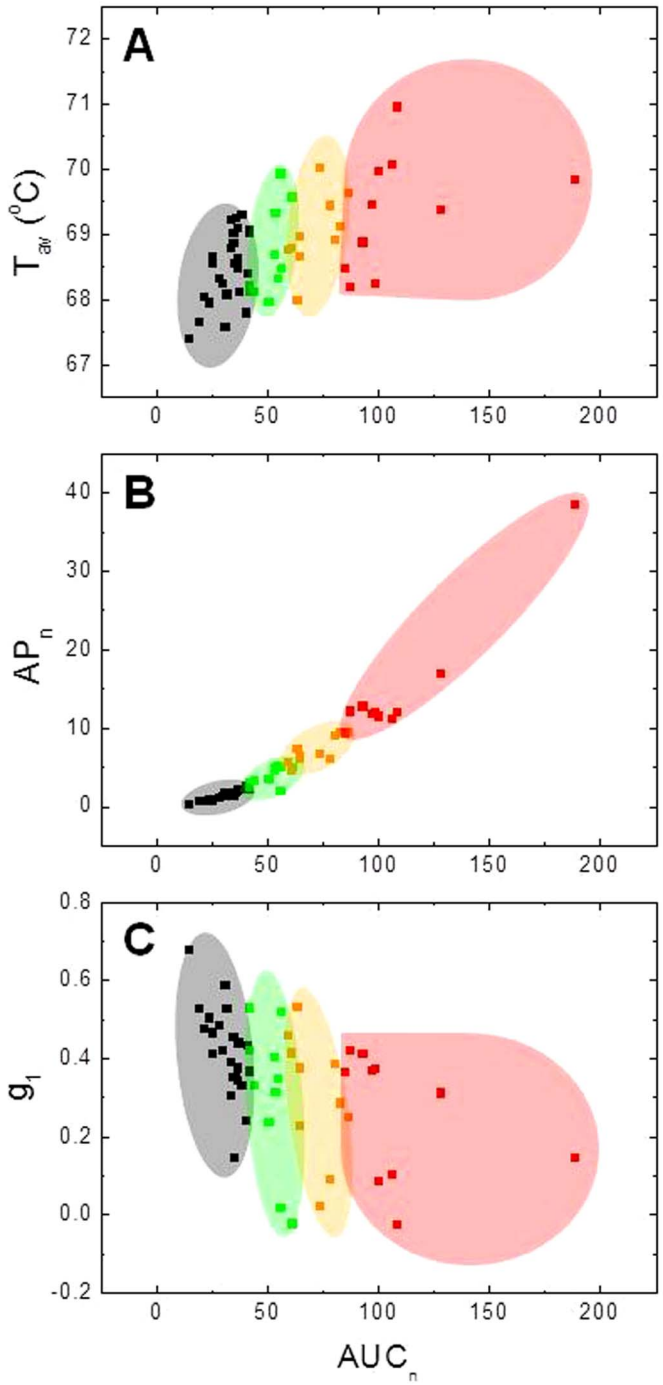

Figure 6 | Clustering regions for subject categories. (A) T_{av} vs. AUC_n; (B) AP_n vs. AUC_n; (C) g_1 vs. AUC_n. The different subject groups are colour coded: healthy subjects (black), stage I GAC patients (green), stage II GAC patients (orange), stage III GAC patients (red). Regions have been sketched considering a larger variability than that corresponding to the uncertainties indicated in Table 1.

The goal of this study was to go deeper into the DSC thermogram data analysis and explore different possibilities for extracting useful information. This is a proof-of-principle preliminary report showing that DSC alone is an invaluable technique for detecting subtle differences in the plasma proteome/interactome of patients that allow discriminating between different progression stages in GAC. Although the number of subjects in this study is not large (25 healthy control subjects, and 30 patients distributed into GAC categories), many of the observed differences are statistically significant (Figure 5). Patient clustering into well-defined regions according to the explicative parameters here defined can be clearly observed (Figure 6). In the cases where the differences that are not statistically significant at this point (e.g. some differences in T_{av} or g_1), an extremely large number of subjects would be required to achieve statistical significance, considering the small differences in the mean values and the rather large variability; in addition, in a particular realization for a given unknown subject that would not have added value because of such large variability.

It is important to emphasize that the most important global and local differential parameters for classifying patients into categories do not have to be the same for every disease. Each disease will require a different set and/or different value ranges of the main discriminating parameters. In the case of GAC patients, the area under the curve, the average temperature, the skewness of the thermogram and the area of the normalized heights polygon as global geometric parameters, and the height of the second individual transition as a local geometric parameter, seem to be the most useful set of parameters for classifying and clustering GAC patients according to the progression stage of the tumor. Thus, an unknown patient will be represented by a set of parameters $\{p_i; i = 1, \ldots, m\}$ obtained from the experimental thermogram, and, depending on the particular values of these parameters, that patient would be ascribed to a certain group associated with a certain disease (with a certain probability).

This work suggests that changes in DSC thermograms from serum samples correlate with the severity of GAC. The physiological cause underlying and triggering the observed differences between DSC thermograms from healthy subjects and GAC patients with different disease burden remains unclear. However, they reveal the potential of the technique and the methodology for establishing biomarkers for disease burden in patients with application in GAC patient staging. The phenomenological model employed in the data analysis provides graphical and numerical tools for quantitating the observed differences among thermograms, as well as for discriminating and clustering subjects. Thus, DSC represents a rapid, inexpensive, non-invasive, easy technique complementary to existing cancer diagnostic and prognostic tools, and appropriate for standard clinical tests with a direct application in patient staging and classification, as well as in screening and monitoring risk-related subpopulations or groups of patients under surveillance.

1. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **12**, 3–18 (1999).
2. Garbett, N. C., Miller, J. J., Jenson, A. B., Miller, D. M. & Chaires, J. B. Interrogation of the plasma proteome with differential scanning calorimetry. *Clin. Chem.* **53**, 2012–2014 (2007).
3. Garbett, N. C., Miller, J. J., Jenson, A. B. & Chaires, J. B. Calorimetric Analysis of the Plasma Proteome. *Semin. Nephrol.* **27**, 621–626 (2007).
4. Garbett, N. C., Miller, J. J., Jenson, A. B. & Chaires, J. B. Calorimetry outside the box: a new window into the plasma proteome. *Biophys. J.* **94**, 1377–1383 (2008).
5. Garbett, N. C., Mekmaysy, C., Helm, C. V., Jenson, A. B. & Chaires, J. B. Differential scanning calorimetry of blood plasma for clinical diagnosis and monitoring. *Exp. Mol. Pathol.* **86**, 186–191 (2009).
6. Michnik, A. *et al.* Differential scanning calorimetry study of blood serum in chronic obstructivepulmonary disease. *J. Therm. Anal. Calorim.* **102**, 57–60 (2010).
7. Zapf, I. *et al.* DSC analysis of human plasma in breast cancer patients. *Thermochim. Acta.* **524**, 88–91 (2011).
8. Fekecs, T., Zapf, I., Ferencz, A. & Lőrinczy, D. Differential scanning calorimetry (DSC) analysis of human plasma in melanoma patients with or without regional lymphnode metastases. *J. Therm. Anal. Calorim.* **108**, 149–152 (2012).
9. Ferencz, A., Fekecs, T. & Lőrinczy, D. Differential Scanning Calorimetry, as a New Method to Monitor Human Plasma in Melanoma Patients with Regional Lymph Node or Distal Metastases. *In Skin Cancer Overview. InTech.* **12**, 141; DOI: 10.5772/25606 (2011).
10. Chagovetz, A. A., Jensen, R. L., Recht, L., Glantz, M. & Chagovetz, A. M. Preliminary use of differential scanning calorimetry of cerebrospinal fluid for the diagnosis of glioblastoma multiforme. *J. Neurooncol.* **105**, 499–506 (2011).
11. Todinova, S., Krumova, S., Gartcheva, L., Robeerst, C. &Taneva, S. G. Microcalorimetry of Blood Serum Proteome: A Modified Interaction Network in the Multiple Myeloma Case. *Anal. Chem.* **83**, 7992–7998 (2011).

12. Todinova, S. *et al.* Calorimetry-based profiling of blood plasma from colorectal cancer patients. *Biochim. Biophys. Acta.* **1820**, 1879–1885 (2012).
13. Parkin, D. M. International variation. *Oncogene.* **23**, 6329–40 (2004).
14. Vieth, M. & Stolte, M. Pathology of early upper GI cancers. *Best Pract. Res. Clin. Gastroenterol.* **19**, 857–69 (2005).
15. Kim, M. A., Lee, H. S., Yang, H. K. & Kim, W. H. Clinicopathologic and protein expression differences between cardia carcinoma and noncardia carcinoma of the stomach. *Cancer.* **103**, 1439–46 (2005).
16. Lauren, P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* **64**, 31–49 (1965).
17. Garbett, N. C., Merchant, M. L., Chaires, J. B. & Klein, J. B. Calorimetric analysis of the plasma proteome: Identification of type 1 diabetes patients with early renal function decline. *Biochim. Biophys. Acta.* **1830**, 4675–4680 (2013).
18. Fish, D. J. *et al*. Statistical analysis of plasma thermograms measured by differential scanning calorimetry. *Biophys. Chem.* **152**, 184–190 (2010).
19. Diamandis, E. P. Serum proteomic profiling by matrix-assisted laser desorption-703 ionization time-of-flight mass spectrometry for cancer diagnosis: next steps. *Cancer Res.* **66**, 5540–5541 (2006).
20. Diamandis, E. P. Peptidomics for cancer diagnosis: present and future. *J. Proteome.* **5**, 2079–2082 (2006).
21. Diamandis, E. P. Validation of breast cancer biomarkers identified by mass 706 spectrometry. *Clin. Chem.* **52**, 771–772 (2006).
22. Villanueva, J. *et al*. Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. *Mol. Cell. Proteomics.* **5**, 1840–1852 (2006).

Acknowledgments

This work was supported by Spanish Ministerio de Ciencia e Innovación (BFU2010-19451 to AVC, PTA2009-2341-I to SV), Ministry of Economy and Competitivity (BFU2013-47064-P to AVC), and Miguel Servet Program from Instituto de Salud Carlos III (CP07/00289 to OA), Fondo de Investigaciones Sanitarias (PI10/00186 to OA, PI11/02578 to AL, PS09/00213 to MAGG), Diputación General de Aragón (Protein Targets Group to AVC, and Digestive Pathology Group B01-2012/2013 to OA), and Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank Dr. Pedro Baptista for helping us with the grammatical revision of this manuscript.

Author contributions

M.A.G.G. and A.L. enrolled patients and collected serum samples; S.V. performed DSC experiments; O.A. and A.V. analyzed data and wrote the main manuscript. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Vega, S., Garcia-Gonzalez, M.A., Lanas, A., Velazquez-Campoy, A. & Abian, O. Deconvolution Analysis for Classifying Gastric Adenocarcinoma Patients Based on Differential Scanning Calorimetry Serum Thermograms. *Sci. Rep.* **5**, 7988; DOI:10.1038/srep07988 (2015).