Complement C3f serum levels may predict breast cancer risk in women with gross cystic disease of the breast

Aldo Profumo\(^a\), Rosa Mangerinib, Alessandra Rubagottib, Paolo Romanoa, Gianluca Damonted, Pamela Guglielminic, Angelo Facchiano\(^f\), Fabio Ferrif, Francesco Ricci\(^c\), Mattia Rocca\(^a\), Francesco Boccardob, \(^b,c,⁎\)

\(^a\)Biopolymers and Proteomics Unit, IRCCS AOU San Martino-IST (San Martino University Hospital and National Cancer Research Institute), Genoa, Italy
\(^b\)Academic Unit of Medical Oncology (Medical Oncology B), IRCCS AOU San Martino—IST (San Martino University Hospital and National Cancer Research Institute), Genoa, Italy
\(^c\)Department of Internal Medicine, School of Medicine, University of Genoa, Genoa, Italy
\(^d\)Department of Experimental Medicine and Center of Excellence for Biomedical Research (CEBR), School of Medicine, University of Genoa, Genoa, Italy
\(^e\)Bioinformatics, Institute of Food Sciences, National Research Council, Avellino, Italy
\(^f\)Department of Science and High Technology, Insubria University, Como, Italy

**ARTICLE INFO**

Article history:
Received 30 January 2013
Accepted 13 April 2013

Keywords:
Complement C3f
Serum
Peptidome profile
MALDI-TOF
Breast cancer risk

**ABSTRACT**

Gross cystic disease (GCDB) is a breast benign condition predisposing to breast cancer. Cryopreserved sera from GCDB patients, some of whom later developed a cancer (cases), were studied to identify potential risk markers. A MALDI-TOF mass spectrometry analysis found several complement C3f fragments having a significant increased abundance in cases compared to controls. After multivariate analysis, the full-length form of C3f maintained a predictive value of breast cancer risk. Higher levels of C3f in the serum of women affected by a benign condition like GCDB thus appears to be correlated to the development of breast cancer even 20 years later.

Biological significance
Increased complement system activation has been found in the sera of women affected by GCDB who developed a breast cancer, even twenty or more years later. C3f may predict an increased breast cancer risk in the healthy population and in women affected by predisposing conditions.

© 2013 Elsevier B.V. All rights reserved.

**1. Introduction**

One of the most relevant problems in breast cancer prevention is how to identify the women who might be at higher risk of developing this disease and, therefore, who might gain the greatest benefit from periodic surveillance with new imaging technologies or from chemoprevention measures. Gross cystic disease of the breast (GCDB) is a common benign disease of the mammary gland, affecting some 7% of women in Western countries [1]. It is associated with a 2–4 fold increased risk of developing breast cancer, probably as a consequence of the evolution of the proliferative epithelial changes which are commonly associated with this benign condition [2]. Though breast cancer risk appears to be associated with cyst type and the intracystic level of steroids or growth factors [3–5], so far no specific serum risk marker has been reported yet in the women affected by GCDB. To this end, proteomic studies could be quite informative, especially when nanotechnologies are applied.
In particular, mass spectrometry (MS) is well suited for the detection of low molecular weight proteins or peptides that might not be detectable with traditional techniques, and that might well represent a dynamic reflection of tissue function. A few studies have addressed the possibility of applying a proteomic approach to the analysis of cyst fluid or nipple aspirates in the diagnosis of breast cancer [6–8]; more encouraging results have been achieved by applying these techniques to the study of the serum peptidome of women diagnosed with breast cancer or to follow disease outcome of patients after breast surgery [9–11]. In particular, Villanueva et al. [12] identified specific degradation patterns associated with breast cancer and able to generate proteomic signatures potentially useful to diagnose this disease. However, so far just a few studies used biological fluids, collected before breast cancer diagnosis, in proteomic profiling approaches to this disease. Namely, Opstal-van Winden et al. [13] performed a serum surface-enhanced laser desorption/ionization (SELDI)-time of flight (TOF) analysis at the peptidome level and identified two signals, of m/z 3323 and m/z 8938 respectively, able to identify the women subsequently developing a breast cancer up to three years before cancer was diagnosed. These findings, though preliminary, prompted us to investigate the possible association of serum peptidome profiles with breast cancer risk in a group of women diagnosed with GCDB but followed for a much longer period of time (up to 20 or more years since blood collection). On the basis of preliminary work on the matrix assisted laser desorption/ionization (MALDI)-TOF MS analysis of cryopreserved sera of a group of healthy donors [14], we have begun by analyzing the sera of a group of women diagnosed with GCDB even twenty or more years before developing breast cancer and those of an appropriately matched group of comparable women who were still breast cancer-free by the same date limit. The results of this analysis form the object of the present article.

2. Materials and methods

2.1. Study design and ethical aspects

Serum samples were obtained from 600 patients affected by GCDB at the time of first cyst aspiration. Samples were collected between 1985 and 1993, and stored at −80 °C. A proportion of these women were followed up at our institution and a few of them have been found to have developed a breast cancer. Many more were lost to follow-up and were diagnosed with breast cancer at other hospitals. Information about the destiny of these women in relation to the development of a breast cancer was obtained by consulting the Genoa Tumor registry.

At the date limit of December 31th 2010, fifty out of 600 women were found to have developed a breast cancer but for only thirty women a number of serum aliquots adequate to be processed for the purposes of the present research were available. In fact a large subgroup of the original cohort had been enrolled in previous sera-epidemiological studies and no residual serum aliquots were available any longer for them. Data concerning women reproductive history and family history of breast cancer had been recorded at the time of clinical examination. The thirty women developing a breast cancer following first cyst aspiration and for whom an adequate number of serum aliquots was actually available served as cases. These women were individually matched (according to a 1:2 ratio) with 60 women who, by the same date limit, were still breast cancer free and who served as controls. Matching was done on the basis of age at first cyst aspiration (cases: median age 43.5 years, range 34–54; controls: median age 44 years, range 35–54) and family history of breast cancer (positive in 24% of both cases and controls). Moreover, as the duration of cryopreservation could potentially affect the protein profile, as it was previously evidenced by us [14], cases and controls were also rigorously matched for the length of cryopreservation period (cases: median length 22.5 years, range 19–24; controls: median length 22.5 years, range 20–24). All the women had provided their consent to blood drawing and cryopreservation; moreover the present study project was approved by the Ethics Committee of the IRCCS AOI San Martino Hospital and National Cancer Research Institute of Genoa, Italy. The permission from the Italian Data Protection Authority has been also obtained to get the information relative to the women health condition for the women lost to follow up.

2.2. Analytical procedures

Acetonitrile, methanol and water were LiChroSolv from Merck (Darmstadt, Germany), trifluoroacetic acid (TFA) and α-cyano-4-hydroxycinnamic acid (CHCA) were from Fluka (St. Louis, MO).

Serum samples preparation, MALDI-TOF/MS and MS/MS analysis procedures were described in a previous paper [14]. In order to avoid procedural bias, both case and control samples were randomly distributed during processing and analysis. All spectra were acquired in quadruplicate in a mass range from 800 to 3000 m/z.

2.2.1. Solid phase extraction

As functionalized superparamagnetic beads appear to be best suited in preparation of low molecular weight (LMW) fraction from complex matrices, samples were processed using Dynabeads® RPC 18 (Invitrogen Dynal, Irvine, CA) with C18 alkyl-modified surface, which are intended for peptide concentration in samples.

Some modifications were made to the Dynabeads® RPC 18 manufacturer’s protocol. Briefly, 40 μl of beads suspension were washed one time with water and three times with 100 μl of 200 mM NaCl, 0.1 % TFA. The beads were re-suspended in 20 μl of water, mixed with 50 μl of serum and left at room temperature for 5 min. After incubation, the tube was placed in the manual magnetic particle concentrator (Dynal MPC®, Invitrogen Dynal) and the supernatant was discarded. The peptides-enriched beads were washed three times with 300 μl of 0.1 % TFA in water and the bound fraction was eluted by incubation, at room temperature for 2 min, with 12 μl of a 1:1 acetonitrile:water solution, to which 3.5 pmoles/μl of an internal standard peptide (MW 1419.76) were added. The standard peptide, used to normalize the data, was directly spiked in the eluting solution to avoid any interference during the binding reaction between analytes and the magnetic beads. In Supplementary Fig. S1 we report the analysis by Tricine–SDS–urea–PAGE of the fraction obtained by solid phase extraction using Dynabeads® RPC 18 with C18 alkyl-modified surface. As can be seen, the extracted serum sample (lane1) appears to
contain some residual LMW proteins and a not well defined fraction at MW lower than 3.5 kDa (box). This fraction, which is very difficult to visualize by SDS–PAGE, represents the subject of our MALDI/TOF analysis. Unfortunately, the mass range covered by our mass spectrometer did not allow us to extend the analysis to species with a higher molecular weight.

2.2.2. API–MALDI/TOF MS analysis

For API–MALDI/TOF analysis, several parameters (matrix composition, sample/matrix ratio, spotted volume, laser power) have been tested in the previous work [14] to find the best experimental conditions. Based on the results obtained, a standard procedure was established: one volume of the eluted sample was manually mixed with two volumes of premade matrix solution (6.2 mg of CHCA in 1 ml of 36% methanol, 56% acetonitrile and 8% water), 0.5 µl of this mixture was spotted onto the MALDI target plate, and allowed to dry at room temperature. API–MALDI/TOF analysis was performed in reflectron positive mode on a 6210 Time of Flight LC/MS (Agilent, Santa Clara, CA) coupled with an atmospheric pressure PDF-MALDI Ion Source (Agilent) equipped with a 337 nm nitrogen laser. The following voltages were applied: fragmentor, 300 V; skimmer, 60 V; OCT RF, 300 V. The acquisition laser power was set at 35% of maximum (Peak Power 75 kW, Pulse Energy 300 mJ). Data acquisition was automated through the software Mass Hunter (Agilent), and programmed to accumulate 600 shots per spectrum. The irradiation program was automated using the spiral motion control of the PDF-MALDI Ion Source. The instrument was externally calibrated using Tuning Mix (Agilent). The nominal resolution of the instrument was 20,000 (17,000, observed) and the nominal mass accuracy (with internal calibration) was <2 ppm.

2.2.3. API–MALDI/ion trap MS/MS analysis

The MS/MS analysis of selected peptides was performed coupling the API–MALDI ion source to an 1100 MSD ion trap instrument (Agilent). The value set for the focalization lenses was: skimat1: 85 V; octapole: 3.5 V; trap drive: 65. The spectra were collected in the ICC mode and the maximum accumtime was 300 ms. The acquisition laser power was set at 35% of maximum (Peak Power 75 kW, Pulse Energy 300 mJ). Other API–MALDI parameters were the same as the MS analysis.

2.3. Data processing

For each spectrum, raw data generated by the mass spectrometer analysis software were exported as an Excel spreadsheet, filtered by using the LabView-based MS-BASELINER tool (see below), and finally processed by using Geena [15] (http://bioinformatics.istge.it/geena/), a recently developed software tool that aims at automating some of the fundamental elaboration steps involved in the analysis of m/z and abundance data from MALDI/TOF MS experiments. Geena processing methods are based on original algorithms and include the following steps: a) preprocessing of spectra replicates, which consists in a series of optional elaborations including isotopic peaks joining, normalization of the abundance of each m/z signal against that of a peptide used as an internal standard, and peak selection; b) computing average spectra for replicated analysis of samples; and c) alignment of average spectra. While the peak selection method currently implemented in Geena is sufficient for many applications, some noisy datasets may require a special treatment. For this reason, we have adopted an alternative method, which is currently implemented into a stand-alone LabView program, MS-BASELINER, that can be freely downloaded from the Geena website. Geena results were exported as a matrix of m/z values and relative abundances, and after a visual inspection to resolve a few spectral ambiguities, 96 candidate signals were obtained. Sporadic signals, presumably due to specific patient related variables such as diet, menopausal status, and drug assumption, have not been considered.

2.4. Statistical analysis

The data spread sheet elaborated by Geena was imported into the TMeV program and analyzed using various statistical algorithms such as hierarchical clustering and significance analysis of microarray (SAM). Then, we analyzed the odds associated with each individual peptide. All continuous values are presented as the mean ± standard deviation. The statistical analysis was conducted using the χ² test for categorical variables and a t test for continuous variables when appropriate. Univariate and multivariate logistic regression were used to compute odds ratio (OR), P value, and 95% confidence interval (95% CI) comparing cases and controls. All ORs were adjusted for the potential confounders, namely: age, family history and cryopreservation time. The regression models were constructed using the backward elimination procedure. The model’s goodness of fit was evaluated using the R² index and the level of statistical significance was set at P ≤ 0.05. All P values were two-tailed. For data analysis we used the IBM software Statistical Package for Social Sciences (SPSS) version 19.0 for Windows (SPSS Inc., Chicago, Illinois, USA).

2.5. Data validation

The validation of the SAM analysis was carried out by implementing a bootstrap resampling of original data and by evaluating the frequency of occurrence of peptide signals. The bootstrap method works by constructing a large number of new random data sets by resampling from the original data set [16,17]. Bootstrap may actually be implemented with or without replacement. In the former case, the random data sets may include the same sample twice of even more times (hypothetically, the same sample could randomly be resampled all times and the whole resulting data set could be constituted only by replications of the same sample). In the latter case, this replication is avoided. Due to the relatively small number of samples involved in our analysis, we preferred to avoid the risk of bias that would have been introduced by replicated samples in the validation data sets and we therefore applied bootstrap without replacement.

3. Results

The good quality of our MS spectra can be appreciated in Fig. 1. After data processing and statistical analysis, 9 out of 96 signals showed an increased intensity in cases as compared to
controls. The peak list with the abundance means is shown in Table 1. Fig. 2A shows the distribution of normalized abundances of the 9 peptides in cases (n = 30) and controls (n = 60). Fig. 2B shows a heat map of the intensities of the discriminating peaks.

The identities of the selected signals were assigned by comparing their experimental masses with literature values. Most of them have been found to be C3f complement fragments differing from each other for the removal of a single amino acid. One of the two remaining peaks is a fragment of high molecular weight (HMW) kininogen (m/z 2209.08) while the signal having m/z 964.43 has not been identified yet. The analysis carried out using the bioinformatics tool FindPept (http://expasy.org) allowed verifying the consistency of our high-resolution experimental masses with the assigned sequences. In addition, the sequence of the C3f peptides was also confirmed by MS/MS experiments performed coupling the API–MALDI ion source to an ion trap analyzer (see Supplementary Fig. S2).

To confirm the reliability of our MALDI TOF data we performed an additional experimental validation. Briefly, we selected two aliquots of a single serum sample showing high levels of C3f fragments and performed two independent solid phase extractions. Then, the eluted peptides were analyzed, in quadruplicate, by MALDI/TOF mass spectrometry, and results were normalized against the internal standard (see Ref. [14]). The normalized abundances of C3f fragments, reported as mean and standard deviation in Fig. 3A, show the good reproducibility of the analytical procedures.

Our results were further validated by implementing a bootstrap resampling of original data without replacement, then evaluating the frequency of occurrence of significant peptide signals. We therefore choose to build 100 random data sets constituted each by 75 out of the 90 samples of the original data set, keeping the same 2:1 ratio between controls and cases. In each random data set, the 75 samples included 50 controls and 25 cases. A SAM analysis was then performed for each of the 100 random data sets. Signals that significantly differ between cases and controls were then selected and a summary analysis was performed. Results of this analysis are shown in Fig. 3B where the number of signals that were found to be significantly different in the two groups at least in one of the SAM runs is reported against the number of times they were found to be significant. The figure confirms in a very effective way the results of the SAM analysis that was carried out on the original data set. All the signals (mainly C3f fragments) that were identified by this analysis appear to be largely the most selected by the SAM analysis performed on the 100 random data sets.

To evaluate the breast cancer risk related with the panel of identified signals, we estimated the odds ratio(s) (OR), adjusted for age at first cyst aspiration, family history of breast cancer, and cryopreservation duration time. Table 2 puts in evidence a higher presence of the signals in cases than in controls, resulting in a significant increase in breast cancer odds for each of them.

### Table 1 – The mean abundance of the selected peaks.

<table>
<thead>
<tr>
<th>Signals (m/z)</th>
<th>Controls (n = 60) mean (SE)</th>
<th>Cases (n = 30) mean (SE)</th>
<th>P &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>964.43</td>
<td>2.83 (1.27)</td>
<td>10.18 (2.54)</td>
<td>0.013</td>
</tr>
<tr>
<td>1211.63</td>
<td>3.81 (1.23)</td>
<td>19.12 (4.56)</td>
<td>0.003</td>
</tr>
<tr>
<td>1348.67</td>
<td>1.28 (0.52)</td>
<td>6.58 (1.65)</td>
<td>0.004</td>
</tr>
<tr>
<td>1449.72</td>
<td>5.46 (1.99)</td>
<td>23.34 (5.08)</td>
<td>0.002</td>
</tr>
<tr>
<td>1690.88</td>
<td>0.50 (0.24)</td>
<td>2.71 (0.79)</td>
<td>0.01</td>
</tr>
<tr>
<td>1777.89</td>
<td>0.69 (0.30)</td>
<td>4.06 (1.26)</td>
<td>0.014</td>
</tr>
<tr>
<td>1864.96</td>
<td>2.93 (0.78)</td>
<td>10.40 (2.12)</td>
<td>0.002</td>
</tr>
<tr>
<td>2021.04</td>
<td>0.44 (0.26)</td>
<td>4.79 (1.71)</td>
<td>0.018</td>
</tr>
<tr>
<td>2209.08</td>
<td>1.10 (0.39)</td>
<td>5.02 (1.37)</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Fig. 1 – A typical spectrum of a GCDB case serum sample generated by API–MALDI/TOF analysis.
Logistic regression analysis was then performed to assess which of the investigated signals would maintain a significant predictive value of the risk of developing a breast cancer. As it is shown in Table 3, only the full-length form of C3f (m/z 2021.04) maintained a statistically significant predictive value (OR = 7.1; P = 0.04) after multivariate analysis. Noteworthy, in

Fig. 2 – Panel A: Distribution of normalized abundances of cases and controls. Black horizontal lines represent the median for each group. Panel B: Supervised hierarchical clustering of serum peptide data, using average-linkage and Euclidean Distance as a distance metric, expressed as normalized ion intensities, derived from the two groups of samples (cases in cyan and controls in blue). Columns represent samples (per group), and rows represent m/z peaks. The heat map scale, ranging from bright green (lowest) to bright red (highest), of normalized peptide intensities is shown below it.
this analysis the HMW kininogen (m/z 2209.08) and the unknown peak (m/z 964.43) lost their predictive value.

Finally, we have also investigated the qualitative changes that may occur in the C3f degradation pattern between cases and controls. For this purpose, we evaluated the ratio between the remaining most abundant degraded forms of C3f. Briefly, for each sample, in which the signals having m/z 1211.63, m/z 1348.67, m/z 1449.72, and m/z 1864.96 were simultaneously represented, we calculated the percentage of the abundance of any individual peptide in respect to their total amount. Then, an average value, with its standard deviation, was calculated for the whole sample set. As is shown in Fig. 4, no significant differences were found.

4. Discussion

Most proteomic profiling studies on breast cancer so far available have looked for novel diagnostic markers, while the search for new predictive biomarkers is limited to few studies investigating the possibility to predict for tumor outcome following surgery and treatment monitoring. The major aim of our study was to identify a serum peptidome signal predictive of the risk of developing a breast cancer in women affected by GCDB constituting a subgroup of otherwise healthy women, though they bear a higher breast cancer risk. Several hypotheses have been proposed to explain the origin of the serum
LWM peptidome. Some authors suggest that a combination of endopeptidase and exopeptidase activities [12,18] could be responsible for peptide fragments generation from specific precursor proteins. Recently, Yi et al. [19] reported experimental evidence that stable isotopically labeled peptides spiked into serum samples are subjected to degradation by intrinsic aminopeptidase and carboxypeptidase activities as a function of time. Other authors suggested thrombin as responsible for a broad spectrum proteolysis in human serum samples [20]. Furthermore, plasma kallikrein is able to cleave HMW kininogen, and subsequently cleavage products are rapidly degraded by plasma proteases [21,22]. Villanueva et al. [12] proposed that the serum peptidome profile could be not only cancer-specific but also cancer type-specific, although in some cases an overlapping of peptide signatures was observed.

In the present study, statistical analysis of data obtained by API–MALDI/TOF MS of cryopreserved serum samples has singled out several signals differing between cases and controls; these signals are related to C3f complement fragments, a fragment of HMW kininogen, and an unidentified signal having m/z 964.43 respectively. All these signals were individually associated with an increased breast cancer risk; however only the m/z 2021.04 C3f fragment maintained a statistically significant predictive value after multivariate analysis, thus proposing itself as a "cancer marker" able to identify the women with the highest risk of developing a breast cancer.

The full length form of C3f is a 17 amino acids peptide having a mass of 2001.1 Da (monoisotopic), and derives from the cleavage of C3b into iC3b by factors I and H [23]. Since the C3f peptide is an internal sequence of the complement C3 protein, its presence in serum can be related exclusively to a specific enzymatic action. Subsequently, it may be degraded by intrinsic peptidase activities, to form the typical degradation pattern observed in the MS spectrum of cryopreserved sera.

The level of the C3f-derived peptides has been previously reported to increase in the serum of patients diagnosed with several neoplasms, including breast, bladder, thyroid and prostate cancer [12,24]. Our findings put in evidence, for the first time, that a complement cascade, and the increased serum concentration of C3f in patients with GCDB, who developed a breast cancer even 20 or more years later.

The complement system is involved in the innate immunity and plays an important role in the surveillance against tumors [25]. Indeed, since the early nineties, immunohistochemistry studies, performed on the tissue samples of breast carcinomas, have demonstrated the deposition of C5b-9 complexes on tumor cell surface, indicating persistent complement activation [26]. Cancer cells are able to activate the complement cascade, and the increased serum concentration of C3f in patients with GCDB, who developed a breast cancer during follow-up, may give rise to two opposing interpretations. On one hand, it could suggest an activation of the complement system by "dormant" breast cancer micro-foci hypothetically associated with breast cysts [5]. It is known that, even at early stages of tumorigenesis, transformed cells may express cancer-specific markers able to initiate the so called "Cancer Immunoediting Process", composed of three phases: elimination, equilibrium, and escape [27]. In the first phase, the developing tumor may be eradicated by innate and adaptive immunity and represents the classical concept of cancer immunosurveillance which protects the host from

<table>
<thead>
<tr>
<th>Signals (m/z)</th>
<th>Controls (n = 60) +/−</th>
<th>Cases (n = 30) +/−</th>
<th>OR a (95% CI)</th>
<th>P &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>964.43</td>
<td>8/52</td>
<td>12/18</td>
<td>4.52 (1.56–13.10)</td>
<td>0.005</td>
</tr>
<tr>
<td>1211.63</td>
<td>11/49</td>
<td>18/12</td>
<td>7.59 (2.71–21.20)</td>
<td>0.000</td>
</tr>
<tr>
<td>1348.67</td>
<td>8/52</td>
<td>13/17</td>
<td>5.45 (1.87–15.89)</td>
<td>0.002</td>
</tr>
<tr>
<td>1449.72</td>
<td>13/47</td>
<td>18/12</td>
<td>6.98 (2.45–19.94)</td>
<td>0.000</td>
</tr>
<tr>
<td>1690.88</td>
<td>5/55</td>
<td>10/20</td>
<td>5.50 (1.67–18.7)</td>
<td>0.005</td>
</tr>
<tr>
<td>1777.89</td>
<td>6/54</td>
<td>10/20</td>
<td>5.74 (1.68–19.56)</td>
<td>0.005</td>
</tr>
<tr>
<td>1846.96</td>
<td>14/46</td>
<td>18/12</td>
<td>6.18 (2.21–17.33)</td>
<td>0.001</td>
</tr>
<tr>
<td>2021.04</td>
<td>3/57</td>
<td>11/19</td>
<td>13.36 (3.02–59.03)</td>
<td>0.001</td>
</tr>
<tr>
<td>2209.08</td>
<td>9/51</td>
<td>15/15</td>
<td>6.41 (2.22–18.50)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

+ : Presence of signal; − : absence of signal.

Table 2 – Odds ratio (OR) associated to the presence/absence of each signal.

Table 3 – Logistic Regression Model including the seven signals of C3f, the HMW kininogen (m/z 2209.08) and the unknown peak (m/z 964.43).

<table>
<thead>
<tr>
<th>m/z</th>
<th>OR a (95% CI)</th>
<th>P &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>964.43</td>
<td>2.09 (0.50–8.70)</td>
<td>0.3</td>
</tr>
<tr>
<td>1211.63</td>
<td>5.11 (0.50–52.33)</td>
<td>0.5</td>
</tr>
<tr>
<td>1349.72</td>
<td>1.81 (0.21–15.80)</td>
<td>0.6</td>
</tr>
<tr>
<td>1846.96</td>
<td>0.59 (0.05–7.08)</td>
<td>0.7</td>
</tr>
<tr>
<td>1348.67</td>
<td>0.55 (0.05–6.55)</td>
<td>0.6</td>
</tr>
<tr>
<td>1690.88</td>
<td>1.99 (0.25–15.71)</td>
<td>0.5</td>
</tr>
<tr>
<td>1777.89</td>
<td>0.41 (0.04–4.69)</td>
<td>0.5</td>
</tr>
<tr>
<td>2021.04</td>
<td>7.10 (1.04–48.38)</td>
<td>0.04</td>
</tr>
<tr>
<td>2209.08</td>
<td>1.11 (0.19–6.44)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

* Adjusted for age, family history of breast cancer and time of cryopreservation.
tumor formation. The equilibrium phase is the period of immuno-mediated latency after incomplete cancer cells removal in the elimination phase, where tumor cells may be either maintained chronically or modified to generate new tumor variants. During the third phase, these variants may eventually escape the immunological control and tumor growth may proceed unrestrained by immune pressure. These concepts may explain the sometimes very late onset of overt tumors in the GCDB patients with high serum levels of C3f. On the other hand, the complement system might be activated in all cases of GCDB, as a consequence of an altered micro-environment generated by the presence of macrocysts, in particular the type I cysts, that are associated to a constant inflammatory condition of the mammary gland. According to this assumption, only the women with massive C3b inactivation, mediated by factor I and H, would present increased level of C3f [28]. C3b inactivation would result in a lower surveillance against transformed cells, which would take advantage of an immune evasion mechanism mediated by factors I and H. This interpretation is in accord with the data reported by other authors [29] who showed a lower expression of C3c complement fragment in tissues from infiltrating ductal breast carcinomas. In fact C3c is generated from iC3b by the action of factor I, during the complement inactivation process that leads to the removal of cell-bound complement components from tumor cell membrane. A down regulation of complement C3c component in tumor tissues may, therefore, suggest the putative involvement in breast tumors of complement inhibitory mechanisms.

In our study, solely the presence of the full-length C3f peptide appears to be associated with a significantly increased breast cancer risk after multivariate analysis. We have also investigated the putative qualitative changes in the C3f degradation pattern between cases and controls, and, apart from the 2021.04 peptide, we failed to find any significant difference in the ratio between the remaining most abundant degraded forms of C3f (Fig. 4). This suggests that, in the specific subgroup of patients with GCDB enrolled in our study, the higher concentration of C3f derived peptides in the sera of women who subsequently developed a breast cancer, is likely to be the result of a different concentration of the C3b precursor protein rather than of a different peptidase activity. This assumption could also explain the higher concentration of full length C3f in the cases group. Indeed, under hypothetical equivalent condition of intrinsic aminopeptidase and carboxypeptidase activities, the persistence of undigested C3f could result from higher baseline levels of full length substrate. In the cases group we correlated the variability of C3f serum levels with age at first cyst aspiration, time to breast cancer diagnosis and family history of breast cancer...


