Breast Cancer Research

Circulating resistin in early-onset breast cancer patients with normal body mass index correlates with disease-free survival and lymph node involvement: An agnostic quantitative proteomics study from the multi-center POSH¶ cohort

--Manuscript Draft--

Manuscript Number:	BRCR-D-17-00362R1
Full Title:	Circulating resistin in early-onset breast cancer patients with normal body mass index correlates with disease-free survival and lymph node involvement: An agnostic quantitative proteomics study from the multi-center POSH¶ cohort
Article Type:	Research article
Funding Information:	
Abstract:	Background: Early-onset breast cancer (EOBC) affects about one in 300 women aged 40 years or less and is associated with worse outcomes than later onset breast cancer. This study explored serum protein markers of adverse prognosis in patients with EOBC. Methods: Serum samples from EOBC patients (stages 1-3) were analysed using agnostic high-precision quantitative proteomics. Patients received anthracycline-based chemotherapy. The discovery cohort (n=399) either had more than five-year disease-free survival (DFS) (good outcome group, n=203) or DFS of less than two years (poor outcome group, n=196). Expressed proteins were assessed for differential expression between the two groups. Bioinformatics pathway and network analysis in combination with literature research were used to determine clinically relevant proteins. ELISA analysis against an independent sample set from the POSH cohort (n=181) was used to validate expression levels of selected target. Linear and generalized linear modelling was applied to determine the effect of target markers, body mass index (BMI), lymph node involvement (LN), oestrogen receptor (ER), progesterone receptor (PR) and HER2 status on patients' outcome. Results: A total of 5,346 unique proteins were analyzed (peptide FDR $p \leq 0.05$). Of these, 812 were differentially expressed in the good vs. poor outcome group and showed significant enrichment for the insulin signalling ($p=0.01$) and the glycolysis/gluconeogenesis ($p=0.01$) pathways. These proteins further correlated with interaction networks involving glucose and fatty acid metabolism. A consistent nodal protein to these metabolic networks was resistin (upregulated in the good outcome group, $p=0.009$). ELISA validation demonstrated resistin to be upregulated in the good outcome group ($p=0.04$), irrespective of BMI and ER status. LN involvement was the only covariate with a significant association with resistin measurements ($p=0.004$). An ancillary in silico observation was the induction of the inflammatory response, leucocyte infiltration, lymphocyte migration and recruitment of phagocytes ($p<0.0001$, $z > 2$). Survival analysis showed that resistin overexpression was associated with improved DFS. Conclusions: Lower circulating resistin correlated with worse DFS independent of BMI and ER status in women with EOBC. Node positive patients had lower levels of resistin. Low resistin levels in EOBC may be a surrogate indicator of worse breast cancer specific prognosis.
Corresponding Author:	Spiros D Garbis University of Southampton Faculty of Medicine UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Southampton Faculty of Medicine
Corresponding Author's Secondary Institution:	
First Author:	Bashar Zeidan
First Author Secondary Information:	

Order of Authors:	Bashar Zeidan
	Antigoni Manousopoulou
	Diana J Garay-Baquero
	Samantha ET Larkin
	Cory H White
	Kathleen N Potter
	Theodoros I Roumeliotis
	Evangelia K Papachristou
	Ellen Copson
	Ramsey I Cutress
	Stephen A Beers
	Diana Eccles
	Paul A Townsend
	Spiros D Garbis
Order of Authors Secondary Information:	
Response to Reviewers:	17 December 2017 Dear Professor Chodosh, Thank you for the opportunity to submit a revised version of our manuscript entitled: "Circulating resistin in early-onset breast cancer patients with normal body mass index correlates with disease-free survival and lymph node involvement: An agnostic quantitative proteomics study from the multi-center POSH¶ cohort" by Bashar Zeidan, Antigoni Manousopoulou, Diana J. Garay-Baquero, Cory H. White, Samantha E.T. Larkin, Kathleen N. Potter, Theodoros I. Roumeliotis, Evangelia K. Papachristou, Ellen Copson, Ramsey I. Cutress, Stephen A. Beers, Diana Eccles, Paul A. Townsend and Spiros D. Garbis (ID: BRCR-D-17-00362) for consideration for publication in Breast Cancer Research. We also thank the reviewers for their very helpful and insightful comments. In order to address these, we have materially enhanced the quality of our manuscript using a substantially expanded quantitative proteome, comprehensive bioinformatics interrogation, and targeted ELISA validation experiments against an independent multi-center cohort at a statistically significant number of samples. We also employed a more sophisticated biostatistical analysis approach using linear and generalized linear modelling. Below follows a point-by-point reply to the concerns raised by the reviewers. We would like to verify that all authors have made a substantial contribution to the information or material submitted for publication, and have read and approved the final manuscript. All authors have no direct or indirect commercial financial incentive associated with publishing the article. The results presented in this paper have not been published previously in whole or in part, except in abstract form. The corresponding author acknowledges full responsibility for dealing with all editorial matters having to do with the procession of the paper until its publication. We look forward to hearing from you. Lastly, all mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008443. The reviewer account details are as follows: Username: reviewer67391@ebi.ac.uk and Password: nQwGmmGi. Respectfully yours, Spiros D. Garbis, BSc, PhD Corresponding author Tel: +44 7554 944 362

Email: S.D.Garbis@soton.ac.uk
Reviewer reports:

Reviewer #1: This work has looked into potential role of circulating adipokine expression for the prognosis of post-treatment response, overall survival and potential risk of long-term insulin resistance in non-obese patients with early-onset breast cancer (EOBC). They mainly used high-precision quantitative mass spectrometry proteomics on a POSH cohort. They found among 117 differentiated adipokines, resistin was found to be up-regulated in the good outcome group [proteomics, p=0.003; ELISA, p=0.03) irrespective of BMI, ER51 and LN status. This could be a candidate marker of longer OS in non-obese patients with EOBC. The work is well designed and has some interesting results. I have the following comments.

1. The conclusion is only based on one cohort as authors acknowledged. Are there any other public similar cohorts for doubling confirming the results?

Author response
We thank the reviewer for highlighting this issue. To the best of our knowledge there are no publicly available data on the adipocytokine levels in normal-weight women with early onset breast cancer. To further verify the validity of our findings in the revised manuscript, we have measured the levels of resistin, a key adipocytokine, using ELISA against an independent set of patients (n=181) from the multi-centre and UK nationwide POSH cohort. No other such cohort was available nationally or internationally. The revised manuscript has been substantially revised to account for the additional validation experiments with extensive biostatistical processing using linear and generalized linear modelling, which further confirm and potential clinical utility of the findings and conclusions made in the original manuscript version.

2. The authors should look into resistin in other breast cancer cohorts in addition EOBC and see if this protein has any significance.

Author response
We thank Reviewer #1 for this interesting observation. Examining the levels of resistin in other types of breast cancer (e.g. post-menopausal breast cancer cases in normal weight or overweight/obese women) has been reported by other groups and was beyond the scope of the present study targeting early onset cancer patients (25 to 40 years old) with normal weight. However, further assessment of resistin expression using the current study pipeline constitutes a future perspective for our group. The revised manuscript now includes this future perspective in the discussion section.

3. Are there other proteins besides resistin showing such better outcome? What about down-regulated proteins?

Author response
We think this comment is highly relevant and we thank Reviewer #1 for bringing up this issue. Indeed, except for resistin 811 additional up- and down-regulated adipokines were differentially expressed between the good vs. poor outcome groups out of a total of 5346 protein profiled. This expanded proteome coverage was made possible by including the analysis results from the additional, higher molecular weight proteomes observed against the same discovery sample set from the original manuscript version, which only included the low molecular weight proteome. However, validating other up- or down-regulated proteins was beyond the scope of the present study. We focused on serum resistin given its strong and interdependent association with the insulin signaling, glycolysis/gluconeogenesis, sugar and fatty acid metabolism and immunological pathways as a candidate marker of EOBC prognosis. Such a substantive, more pleiotropic association of resistin with these biological pathways was made possible because of the expanded proteome coverage and its biological interpretation with a combination of commercial and licensed bioinformatics software tools. Lastly, this more comprehensive assessment of resistin, within the context of the EOBC cohort examined, lead to the generation of a novel hypothesis that we intend to examine as future perspective. These elements are fully described in the revised manuscript text with additional figures, tables and references.

4. I understand this is a proteomic biomarker study. I am curious the gene expression levels for these 117 adipokines and if they are differentially expressed.

Author response
We thank Reviewer #1 for this interesting observation. It would be of great interest to examine the gene expression levels for these adipokines in the adipose tissue of EOBC patients, however such samples were not available from women in the POSH cohort.

5. Are there any literature reported proteomic biomarkers? If yes, the authors should compare them with resistin and discuss about it.

Author response
We thank Reviewer #1 for pointing this out and we apologise for the oversight. An expanded literature review of biomarker discovery efforts using quantitative proteomics approaches and how it relates to our study approach is now included in the Introduction section. However, our study constitutes the first-ever observation focusing on resistin expression, a key adipokine protein, at the serological level of EOBC patients using a unique depletion-free quantitative proteomics approach. Furthermore, the most comprehensive serum proteome coverage observed to date, thanks to the technical merits of our unique methodological approach, further solidified the potential clinical utility of resistin as a novel prognosis marker of EOBC patients.

Reviewer #2: In the manuscript by B Zeidan et al. the authors describe studies aimed at the identification and validation of the protein resistin in non obese early-onset breast cancer patients and shown increased presence in patients with a good overall survival as defined by >=5 years overall survival. Differential protein analysis comparing good outcome vs poor outcome identified many proteins in the serum. Utilizing pathway analysis, resistin was identified for further study. It would be of interest if there was discussion on any other pathways that might be pursued in future studies.

Author response
We thank Reviewer #2 for pointing this out and we apologise for the oversight. Our revised manuscript has expanded on the full repertoire of quantitative proteomics measurements performed to the sera of EOBC patients. Specifically, this expanded proteome coverage was made possible by including the analysis results from the additional, higher molecular weight proteomes observed against the same discovery sample set from the original manuscript version, which only included the low molecular weight proteome. This expanded differentially expressed proteome allowed for a more comprehensive biochemical and molecular biology inference to be made. In the revised manuscript we have included new figures (Figure 2 and Figure 5) along with the respective text in the Methods, Results and Discussion sections describing with more detail the proteomic results. In particular, the insulin-signalling pathway, glycolysis/gluconeogenesis, glucose/fatty acid metabolism and immune response were significantly enriched in the differentially expressed proteins between good and poor outcome groups. The revised manuscript expands on how these pathways are manifested in EOBC patients and their prognosis. Consequent to this approach a novel hypothesis was derived that now claims that individuals with early breast cancer who have relatively higher resistin levels may provide an environment from which tumours are less likely to metastasise.

The levels of resistin were validated in individual samples, however, the difference between the two groups was small albeit had statistical significance. Ideally, if more of the POSH sample could be accessed for individual testing to provide a larger sample size, that would be ideal.
Author response
We thank Reviewer #2 for raising this important issue. As part of the revised manuscript, we performed individual ELISA measurements of resistin in an independent validation cohort (n=181). As stated in the method section of the revised manuscript: “The size of the validation cohort was based on the logistic models requiring a minimum of 10 events per predictor variable (see references below, included in the revised manuscript), which in our case included ER, PR, HER-2, LN, and BMI status. For the validation cohort, the same inclusion and exclusion criteria as

described above were applied but, additionally, samples used in the discovery phase were excluded".

References

Concato J, Peduzzi P, Holford TR, Feinstein AR: Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol 1995, 48(12):1495-1501.

Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995, 48(12):1503-1510.

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996, 49(12):1373-1379.

The lack of correlation between Resistin and BMI, ER status, and LN status could be due to the small sample size and would warrant further analysis in an independent data set or sampling more from the POSH cohort.

Author response

We thank Reviewer #2 for this interesting and highly relevant observation. Our revised manuscript thoroughly addresses this concern. Namely, based on the individual serum analysis performed to the independent validation cohort (n=181), as described in our reply to the previous comment, linear and generalized linear modelling was applied to determine the effect of BMI, lymph node (LN), ER, PR and HER2 status on resistin expression. The up-regulation of resistin in the good relative to the poor outcome group in the validation cohort (p=0.04), was not dependent of the BMI. LN involvement was the only covariate with a significant effect on resistin measurements (p=0.004). Furthermore, increased circulating resistin positively correlates with disease-free survival. LN negative compared to LN positive patients had higher levels of resistin. Additional figures and tables with corresponding text in the result, discussion and supplementary sections (including the above descriptions) have been added to account for this multi-parametric assessment.

Overall, the authors acknowledge and addressed the shorting comings of the study adequately and sought to verify the differential protein using an orthogonal technology (microarrays) and an independent data set. The statistical analysis used throughout the study is sound.

Reviewer #3: The manuscript by Zeidan et al attempts to uncover protein biomarkers associated with prognosis for chemotherapy response and OS in early onset breast cancer patients. The topic is of high importance to the field, but despite this enthusiasm is diminished by several weaknesses that appear to limit impact.

1. Despite the title claims that serum resistin levels appear to correlate with chemotherapy response/outcome, as the authors themselves state- the discovery set is marred by a hard-wired bias insofar as the poor outcome group has higher LN frequency and triple-negative tumors. They do not appear to deploy the correct statistical methods to correct for the interactions and it is not apparent to me that these biases have been corrected for. Simply showing that resistin levels do not seem to correlate with LN and ER status alone is not the same thing as showing that resistin levels predict outcome independent of ER and LN. It seems to me that the authors have not provided or described the evidence to the latter.

Author response

We thank Reviewer #3 for these interesting and insightful comments. We apologise for not including information on the triple negative tumours of the discovery cohort in the initial manuscript. This information for patients in both the discovery and validation cohorts has been included in the revised Tables 1 and 2. The triple negative tumours between the two groups of the discovery sample sets were comparable (n= 32 and 35 for the good and poor outcome groups respectively).
As part of the revised manuscript, we performed individual ELISA measurements of

resistin in an independent validation cohort (n=181). Such a sample size for the validation cohort was based on the logistic models requiring a minimum of 10 events per predictor variable (see references below, included in the revised manuscript) for each of the good and poor outcome groups, which in our case included ER, PR, HER2, LN, and BMI status. For the validation cohort, analogous inclusion and exclusion criteria were applied. Additionally, linear and generalized linear modelling was applied to determine the effect of BMI, lymph node (LN), ER, PR and HER2 status on resistin expression. The up-regulation of resistin in the good relative to the poor outcome group in the validation cohort (p=0.04) was not dependent of the BMI and ER status. Survival analysis showed that resistin had a moderate effect upon disease-free survival. LN involvement was the only covariate with a significant effect on resistin measurements (p=0.004). Furthermore, increased circulating resistin positively correlates with disease-free survival independent of BMI and ER status in women with EOBC. LN negative compared to LN positive patients had higher levels of resistin. Additional Figures and Tables with corresponding text (including the above) in the Results and Discussion sections have been added to account for this multi-parametric statistical assessment.

References

Concato J, Peduzzi P, Holford TR, Feinstein AR: Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol 1995, 48(12):1495-1501.

Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995, 48(12):1503-1510.

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996, 49(12):1373-1379.

2. To judge more fully the impact of the findings, the authors must validate their findings from the discovery set in a blinded independent study set of samples, applying an appropriate cut-point and determining ROC output. This is especially so since the authors did not use independent serum samples in discovery - used only pooled samples.

Author response

We thank Reviewer #3 for raising this essential issue. In the revised manuscript we performed individual ELISA measurements of resistin in an independent set of 181 EOBC patients from the UK nationwide and multi-centre POSH cohort. Specifically, of the randomly selected patients, n=90 from the good outcome group and n=91 samples from the poor outcome group were subjected to ELISA analysis. An ROC output was determined based on the resistin measurements from the independent validation cohort and an appropriate cut-off point was applied. Additional Figures with corresponding text descriptions in the result and discussion sections have been added to account for the ROC assessment.

3. The authors should comment on the use of post-treatment samples for their discovery set. There is no information about this- how long after treatment was the serum obtained? Are there differences in the timing of the serum draw between patients? How do the authors know that the differences in the poor v good outcome aren't due to preanalytical confounders in this aspect?

Author response

We thank Reviewer #3 for raising this issue and we apologise for this oversight. The revised manuscript now includes Standard Operation Procedures used for the POSH study as the Supplementary Section 1. Suitable description with references are included in the revised manuscript. We followed vigilant measures including strict SOP adherence for sample collection, preparation and storage, standardised and automated MS analysis and further validation of all samples in the same expirement to eliminate potential “batch effect”.

	4. The authors should comment on why it would be that nearly 25% of the entire blood proteome appears to be different between the two subgroups studied. Why would you expect such a systemic difference and would this lead one to believe that the resistin findings are non-specific? Is this a methodological weakness when using pooled samples for discovery? Author response We thank Reviewer #3 for this interesting and highly useful comment. In the present study, we identified a total of 5,346 unique proteins were analyzed (peptide FDR $p \leq 0.05$). Of these, 812 proteins were differentially expressed in the good vs. poor outcome group and showed significant enrichment for the insulin signalling pathway ($p=0.015$). This translates to about a 15.2 % of the total proteome profile, which constitutes a reasonable degree of differential expression for a given quantitative proteomics study. This differentiated proteome may indeed reflect a distinct tumour biology background in EOBC patients with good vs. poor outcome but could also be partly attributed to the pooling of samples used for the discovery phase. However, the extensive pooling that was used between the biological replicates for the discovery set of experiments was used to normalize out the inherent heterogeneity of clinical presentation between patients while at the same time preserving the more consistent, and thus potentially more constitutively important, differentially expressed proteins between the good and poor outcome groups. To address the potential non-specific and/or false positive biomarker discovery, we have conducted a further independent validation set of ELISA analysis against individual samples. To address accurate protein inference, ELISA was used as the measurement approach for the validation cohort as it allowed the analysis of the intact form of resistin whereas discovery proteomics allows the assessment of its expression at the derived peptide level that resulted from the trypsin proteolysis step. The differential serum resistin expression made with ELISA was concordant with the quantitative proteomic findings. This indicates a real biological trend and the specific pathway(s) involved in such observation will be explored in future work. Furthermore, the higher levels of resistin have been verified at the tissue level using a publicly available microarray database to correlate with good prognosis in breast cancer, suggesting its tissue-specificity.
Additional Information:	
Question	**Response**

Circulating resistin in early-onset breast cancer patients with normal body mass index correlates with disease-free survival and lymph node involvement: An agnostic quantitative proteomics study from the multi-center POSH¶ cohort

Bashar Zeidan[1], Antigoni Manousopoulou[2], Diana J. Garay-Baquero[2,3], Cory H. White[3%], Samantha E.T. Larkin[1], Kathleen N. Potter[1], Theodoros I. Roumeliotis[2&], Evangelia K. Papachristou[2@], Ellen Copson[1], Ramsey I. Cutress[1], Stephen A. Beers[1], Diana Eccles[1], Paul A. Townsend[4,*] and Spiros D. Garbis[1,2*]

¶ The **P**rospective study of **O**utcomes in **S**poradic versus **H**ereditary breast cancer

[1]Cancer Sciences Unit, Faculty of Medicine, University of Southampton, UK, [2]Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, UK; [3]Clinical and Experimental Sciences Unit, Faculty of Medicine, University of Southampton, UK; [4]Division of Cancer Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Sciences Centre, University of Manchester, UK. [&]Current address: The Institute for Cancer Research, London, UK; [@]Current address: Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK; [%] Current address: Merck Cambridge Exploratory Science Center, Boston, USA

PAT and SDG jointly led the study and are senior co-corresponding authors

To whom correspondence should be addressed:
Spiros D. Garbis, PhD, Faculty of Medicine, University of Southampton
E-mail address: S.D.Garbis@soton.ac.uk
Paul A. Townsend, PhD, Faculty of Biology, Medicine and Health, University of Manchester
E-mail: paul.townsend@manchester.ac.uk

Running title: Circulating resistin in early-onset breast cancer

Keywords: Quantitative serum proteomics, iTRAQ, resistin, insulin resistance, glycolysis/gluconeogenesis, early-onset breast cancer

Abstract

Background: Early-onset breast cancer (EOBC) affects about one in 300 women aged 40 years or less and is associated with worse outcomes than later onset breast cancer. This study explored serum protein markers of adverse prognosis in patients with EOBC.

Methods: Serum samples from EOBC patients (stages 1-3) were analysed using agnostic high-precision quantitative proteomics. Patients received anthracycline-based chemotherapy. The discovery cohort (n=399) either had more than five-year disease-free survival (DFS) (good outcome group, n=203) or DFS of less than two years (poor outcome group, n=196). Expressed proteins were assessed for differential expression between the two groups. Bioinformatics pathway and network analysis in combination with literature research were used to determine clinically relevant proteins. ELISA analysis against an independent sample set from the POSH cohort (n=181) was used to validate expression levels of selected target. Linear and generalized linear modelling was applied to determine the effect of target markers, body mass index (BMI), lymph node involvement (LN), oestrogen receptor (ER), progesterone receptor (PR) and HER2 status on patients' outcome.

Results: A total of 5,346 unique proteins were analyzed (peptide FDR $p \leq 0.05$). Of these, 812 were differentially expressed in the good vs. poor outcome group and showed significant enrichment for the insulin signalling ($p=0.01$) and the glycolysis/gluconeogenesis ($p=0.01$) pathways. These proteins further correlated with interaction networks involving glucose and fatty acid metabolism. A consistent nodal protein to these metabolic networks was resistin (upregulated in the good outcome group, $p=0.009$). ELISA validation demonstrated resistin to be upregulated in the good outcome group ($p=0.04$), irrespective of BMI and ER status. LN involvement was the only covariate with a significant association with resistin measurements ($p=0.004$). An ancillary *in silico* observation was the induction of the inflammatory response, leucocyte infiltration, lymphocyte migration and recruitment of phagocytes ($p<0.0001$, $z > 2$). Survival analysis showed that resistin overexpression was associated with improved DFS.

Conclusions: Lower circulating resistin correlated with worse DFS independent of BMI and ER status in women with EOBC. Node positive patients had lower levels of resistin. Low resistin levels in EOBC may be a surrogate indicator of worse breast cancer specific prognosis.

Abstract word count: 350

Introduction

Approximately one in 300 women aged 40 years or are diagnosed with breast cancer in the UK and young age at diagnosis is associated with worse clinical outcomes and greater likelihood of genetic susceptibility (http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer) [1, 2]. Current prognostic biomarkers are based on tumour characteristics, tumour grade and stage and receptor status. Host factors that may influence prognosis are not currently included in commonly used models [3]. Identifying novel host markers associated with EOBC prognosis may improve our understanding and management of this subgroup of patients.

As a quantitative proteomics approach, the use of chemical labelling with isobaric stable isotope reagents, such as isobaric Tags for Relative and Absolute Quantitation (iTRAQ) and Tandem Mass Tags (TMT), has been applied in combination with liquid chromatography – mass spectrometry (LC-MS) techniques for the discovery of candidate cancer biomarkers in serum or plasma [4, 5]. Such methodological approaches provide the distinct advantage of simultaneously measuring protein expression under the same instrumental analysis conditions thereby reducing experimental bias and improving relative quantitative accuracy and precision [6]. An iTRAQ LC-MS approach that also used a peptide-based affinity enrichment pre-treatment step was applied to plasma samples derived from stage I-III breast cancer patients relative to healthy volunteers [7]. Another iTRAQ LC-MS study that used affinity-depletion of the high-abundant proteins was applied to serum samples derived from post-menopausal breast cancer patients relative to healthy controls [8]. In this study, however, we utilised quantitative LC-MS proteomic methods that do not depend on prior affinity enrichment or depletion of plasma/serum that may compromise their analysis for clinically relevant protein markers [5, 9]. In this capacity, the entire serum protein content was subjected to quantitative proteomic analysis. Using serum from a cohort study of early onset breast cancer cases, we explored the potential for

quantitative discovery proteomics to reveal novel markers of poor outcome in young women with EOBC [2].

Materials and Methods

Patient inclusion criteria

The present study included patients with early-stage (T1-T3) invasive breast carcinoma, diagnosed between January 2000 and December 2007 from the Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) cohort, a UK-wide multi-center prospective observational study of EOBC patients, aged 40 years or younger and treated with standard therapies according to local protocols (**Supplementary Section 1**) [1, 2, 10]. Patients included in this study received anthracycline-based chemotherapy. For the discovery phase, patients were selected based on period of disease-free follow up to provide a discovery cohort enriched for poor and for good outcomes. The good outcome group comprised 203 randomly selected patients with disease-free survival (DFS) of at least 5 years following treatment. The poor outcome group included 196 patients who experienced local recurrence, new primary contralateral and/or distant metastasis and/or death within 2 years of initial diagnosis. The patient full clinico-pathological characteristics are detailed in **Table 1**. The study design is summarized in **Figure 1.**

Serum procurement and processing

Peripheral blood samples were drawn from patients in the POSH cohort at their local cancer unit and processed and stored in accordance with the POSH SOPs – (Supplementary Methods) [1, 2]. For the good outcome group, using the randomization function of Microsoft Excel (2011), individual 20 µL aliquots from 102 and 101 specimens were respectively pooled together to create two biological replicate pools (good outcome groups 1 and 2). Identical procedures were undertaken for the poor outcome group, with 98

samples being pooled in each biological replicate (poor outcome groups 1 and 2). An aliquot of 100 µL from each sample pool was mixed with 400 µL 6 M Guanidine in 9:1 Water: methanol and subjected to High Performance Size-Exclusion Chromatography (HP SEC) and dialysis exchange for the serum protein pre-fractionation and purification steps [9, 11-14].

Quantitative LC-MS Proteomics

For each sample pool, 100 µg protein content derived from the respective SEC segments were prepared. Briefly, the segmented protein fractions were subjected to dialysis purification and lyophilized to dryness. The purified proteins were re-solubilized in 200 µL dissolution buffer (0.5M triethylammonium bicarbonate, 0.05% SDS), quantified, and subjected to proteolysis with trypsin using a standardized protocol. The tryptic peptide mixtures per each of the four segments (covering a wide molecular weight range between 1 X 10^6 – 3K Da) were then isobaric stable isotope labelled with the iTRAQ reagents for each of the good and poor outcome groups and their biological replicates) in accordance to manufacturer specifications, and pooled. The resulting iTRAQ peptides were initially fractionated with alkaline C_8 Reverse Phase (RP) liquid chromatography [13, 15]. Each peptide fraction was further separated with on-line nano-capillary C_{18} reverse phase liquid chromatography under acidic conditions, subjected to nanospary ionization and measured with ultra-high resolution mass spectrometry using the hybrid ion-trap / FT-Orbitrap Elite platform [12-14, 16]. Reporter ion ratios derived from unique peptides were used for the relative quantitation of each respective protein. Raw reporter ion intensity values were median-normalized and log_2transformed. Proteins identified with a minimum of two unique peptides and a one-sample T-test of $p \leq 0.05$ were considered as differentially expressed between good and poor outcome groups and were further subjected to bioinformatics analysis [12, 15, 17, 18]. A detailed description of the quantitative proteomics approach used can be found in the **Supplementary Section 2**.

Bioinformatics analysis

Hierarchical clustering of the differentiated proteins was performed using Cluster 3.0 (C Clustering Library 1.52) and Java Treeview (version 1.1.6r4) such that distances were calculated using the Euclidean based metric and then clustered using the complete linkage method. MetaCore (Clarivate Analytics, Boston, MA, USA), Ingenuity Pathway Analysis, including its Diseases & Functions module (Qiagen, Silicon Valley, CA, USA) and DAVID Bioinformatics Resources 6.8 [National Institute of Allergy and Infectious Diseases (NIAID), NIH] (https://david.ncifcrf.gov/), were applied to differentially expressed proteins analysed with at least two unique peptides to identify significantly over- represented networks and gene ontology (GO) terms. Fisher exact and FDR-corrected $p \leq 0.05$ was considered significant.

Single-blinded ELISA measurements in the validation cohort

To replicate the accuracy of relative quantitation of a target protein, ELISA was performed against individual sera derived from an independent validation sample set within the POSH cohort and sharing analogous inclusion criteria with the discovery sample set. As high-BMI levels may constitute a confounding factor for resistin expression, a normal BMI status was used as an additional inclusion criterion. For the ELISA validation a single-blinded design was used, wherein assignment of patient IDs to a good or poor outcome group was unavailable to the analyst performing the measurements and uncovered by an independent clinician after the measurements were completed. In particular, the validation cohort was comprised of 200 samples (n=100 good outcome patients and n=100 poor outcome patients), randomly selected from the POSH cohort using the randomisation function of Microsoft Excel (2011). Of the randomly selected patients, sufficient serum volume was only available for 90 and 91 samples from the good and poor outcome groups respectively. The size of the validation cohort was based on the logistic models requiring a minimum of 10 events per predictor variable [19-21], which in our study included ER, PR,

HER2, LN, and BMI status. The ELISA measurements were performed using a resistin sandwich ELISA kit according to the manufacturer's protocols (USCN Life Sciences Inc, Wuhan, P. R. China). Absorbance was measured with the GloMax® Discover, Promega plate reader (Thermo Fisher Scientific). Data was analysed in Prism (version 7.0a). Statistical analyses of the ELISA measurements were based on the Welch's 2-sample t-test for unequal variances to assess significant differences between groups at $p \leq 0.05$. This test was deemed appropriate as there is balance of samples in groups and each group is well above the suggested level of 15 per group which allows control of the type I error rate even in non-normal distributions [22-24].

Linear and generalized linear modelling

Modelling patient outcome in the validation cohort as a function of resistin and other variables was performed using generalized linear modelling and the function *glm* within the R statistical computing environment (https://www.R-project.org/) and using the logit link function appropriate for the binomial family. For linear modelling of resistin as a function of BMI, lymph node (LN) involvement (N0=negative; N1-3=positive), ER (Allred Score: 0-2= negative; 3-8=Positive) PR (0-2=negative; 3-8=positive) and HER2 status (0, 1+=negative; 2+=equivocal; 3+=positive), the linear modelling function *lm* was utilized (https://www.R-project.org/). The reference for each categorical variable was as follows: LN=negative; ER=negative, PR=negative, HER2=negative. All coefficients were tested with the function *coeftest* available within R (https://www.R-project.org/).

ROC and AUC analysis

A prediction vector was generated with the predict function in R and then merged with a vector of true outcome results. To determine a threshold by which a prediction would be considered a positive (good outcome result) a receiver operating characteristic (ROC)

curve was generated by selecting 101 potential threshold values between 0 and 1 with a 0.01 step-size and calculating the true positive and false positive rates for each threshold value. The cost function for these threshold values was the sum of the false positives and false negatives given the threshold setting. These results indicated that a threshold of 0.5 was reasonable above which, a prediction was determined to be a positive (good outcome) and below which a prediction was determined to be a negative (poor outcome). The AUC (area under the curve) measure was calculated by using the auc function in the pROC package available within R.

In silico survival analysis in breast cancer tissue samples

A meta-analysis based biomarker assessment of resistin in breast cancer tissue samples was performed using the online software tool Kaplan Meier Plotter (http://kmplot.com). The Kaplan Meier Plotter assesses the effects of 54,675 genes on patient DFS using 5,143 breast cancer samples with a mean follow-up of 200 months [25].

Results

Quantitative proteomic analysis and in silico bioinformatics interpretation

Quantitative proteomics yielded a total of 5,346 unique proteins (peptide FDR corrected $p \leq 0.05$) from all 4 HP-SEC derived segments (**Supplementary Section 3**). Of these, 812 proteins were differentially expressed between the good and poor outcome group (*$p \leq 0.05$, ≥ 2 unique peptides*) (**Supplementary Section 4**) and were subjected to further bioinformatics analysis. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008443.

Pathway and Network Analysis

Significant enrichment was observed for the insulin pathway in the differentially expressed proteins between good and poor outcome group (*p=0.015,* KEGG Pathway analysis using DAVID) (**Figure 2A**). MetaCore pathway analysis identified glycolysis/gluconeogenesis as a significantly enriched process in the differentially expressed proteins between good and poor outcome groups ($p < 0.011$, FDR corrected) (**Figure 2B**). Ingenuity Pathway Analysis identified small molecule biochemistry, in particular glucose and fatty acid metabolism, as a significantly over-represented network (score=23, focus molecules=20) in the differentially expressed proteins between good and poor outcome groups. Resistin was a key molecular participant in this network (**Figure 2C**), and based on its previously reported role in breast cancer biology and insulin resistance risk [26-36], was chosen for targeted validation.

Resistin ELISA validation measurements

Resistin was measured to be up-regulated in the good outcome group from the proteomic discovery stage using pooled serum samples [*p=0.009*]. (**Figure 3A**). The up-regulation of serum resistin in the good outcome group relative to the poor outcome group was confirmed with ELISA against the validation cohort [good outcome group; n=90, Mean value (SD) = 114.2 (114.5) ng/mL] [poor outcome group; n=91, Mean value (SD) = 86.8 (57.7) ng/mL] ($p = 0.04$) (**Figure 3B**) (**Supplementary Section 5**).

ROC/AUC and KM Survival analysis

To determine the predictive power of resistin for outcome, a receiver-operating characteristic curve (ROC) was generated (**Figure 4A**) along with a cost function with equivalent penalties for false negatives and false positives (**Figure 4B** and **4C)**. The AUC measure of the ROC curve indicated a moderate level of success for utilizing resistin measures to predict outcome. Using the measure of true positives, true negatives, false

positives, and false negatives, serum resistin provided an accuracy of 0.652, a sensitivity of 0.667, and a specificity of 0.637.

Finally using publically available data, *in silico* Kaplan Meier survival analysis showed a longer disease-free survival in patients with higher resistin levels at the tissue level for up to 200 months (**Figures 4D).**

Prediction of biological function directionality (induction or Inhibition)

The Diseases & Functions module of Ingenuity Pathway Analysis demonstrated that inflammatory response, leucocyte infiltration, lymphocyte migration and recruitment of phagocytes were significantly induced biological processes based on the downstream differentially expressed proteins of the good vs. poor outcome groups. Resistin was specifically found to participate in the activation of leucocyte infiltration (**Figure 5**).

Linear and Generalized Linear Modelling

Both linear and generalized linear modelling techniques were utilized to determine which covariates would relate to DFS and resistin expression (**Supplementary Section 6**). LN involvement was found to correlate with worse patient outcome (*p-value = 0.004*) and demonstrated a significant difference in mean value of resistin between LN groups. More specifically, LN negative patients had significantly higher resistin levels compared to those with LN involvement [LN negative group: n=71, Mean value (SD) = 124.8 (107.5) ng/mL; LN positive group: n=110, Mean value (SD) = 84.7 (75.6) ng/mL; *p = 0.0037,* Welch's two-sample t-test]. (**Figure 3C, Supplementary Section 6**).

Discussion

Improvements made in breast cancer survival have been associated with the wider use of neo/adjuvant chemotherapy such as anthracycline/taxane-based treatment [37].

Routine immunohistochemical analysis is used for both prognosis and predictive markers of response to hormonal therapy and trastuzumab (ER / PR and HER2 respectively). Young age [38, 39] and obesity [2] at breast cancer diagnosis have been reported to be independent prognostic markers of adverse disease outcome. The aim of this study was to find serum proteomic markers of additional prognostic relevance to EOBC outcomes.

This study implemented a high-precision quantitative serum proteomics discovery analysis followed by targeted serum ELISA-based validation in an independent sample set of non obese EOBC patient samples (**Figure 1**). The applied proteomics method achieved the highest degree of proteome coverage in breast cancer serum to date (5,346 unique proteins with peptide FDR $p \leq 0.05$). The methodological features that led to this comprehensive proteome result were its ability to analyze non-depleted serum that also contains exosome-derived proteins in addition to directly secreted proteins, as reported [9, 12, 14]. Such an in-depth analysis was deemed essential for the unbiased interrogation of expected systemic effects and their affiliated biological pathways and networks induced by treatment.

Hierarchical clustering analysis of all 812 differentially expressed proteins (DEPs) is presented in heatmap format in **Figure 2A**. The DEPs were then subjected to canonical pathway analysis, which achieved significant enrichment for the insulin signaling (*p=0.015*) (**Figure 2B**) and glycolysis/gluconeogenesis pathways (*p=0.011)* (**Figure 2C**). Interestingly, the majority of observed proteins that encoded for both these pathways were of exosomal origin, as listed in the manually curated ExoCarta Web-based compendium (http://www.exocarta.org) [40-42]. Of relevance, all enzymes mapping to the glycolysis/gluconeogenesis pathway, were upregulated in the poor outcome group, suggesting that poor prognosis patients catabolize glucose more actively compared to patients with longer survival (**Figure 2C**). One noteworthy enzyme found to be upregulated in the poor outcome group was the Pyruvate Kinase M2 isoform (PKM2) known to play an important role in tumorigenesis. As observed in different types of cancers, including breast

cancer, pyruvate kinase expression shifts to the PKM2 isoform in order to utilize glucose more efficiently to generate biomass under anaerobic conditions [43]. The functional involvement of the insulin signaling and the glycolysis/gluconeogenesis pathways were further verified with Ingenuity Pathway Analysis that showed significant enrichment for glucose and fatty acid metabolism (**Figure 2D**) and included resistin, a secreted protein, as one of its key nodal components. We focused on serum resistin given its association with the insulin signaling and glycolysis/gluconeogenesis pathway as a candidate marker of EOBC prognosis.

In agreement with the discovery cohort (**Figure 3A**), resistin was found to be upregulated in the good outcome group in the normal weight validation cohort (**Figure 3B**). To address accurate protein inference, ELISA was used as the measurement approach for the validation cohort as it allowed the analysis of the intact form of resistin whereas discovery proteomics allows the assessment of its expression at the derived peptide level that resulted from the trypsin proteolysis step.

In this work, both linear and generalized linear regression analysis confirmed ER, PR, and HER2 exhibited a significant degree of interdependence ($p < 0.05$) (**Supplementary Section 6**). A receiver operating characteristic (ROC) curve (**Figure 4A)** and associated cost curve (**Figure 4B**) were used to assess the value of resistin in outcome prediction between the two groups in this study, The AUC measure of the ROC curve indicated a moderate level of success for utilizing resistin measures to predict outcome. Using the measure of true positives, true negatives, false positives, and false negatives (**Figure 4C**), serum resistin provided an accuracy of 0.652, a sensitivity of 0.667, and a specificity of 0.637. We explored resistin expression at the tissue level using an *in silico* meta-analysis micro-array database, the Kaplan Meier plotter software tool (http://kmplot.com/analysis/), Consistent with the serum observations in our current study, this analysis showed that high tissue levels of resistin were associated with longer disease-free survival ($p<0.001$) (**Figure 4D**).

Resistin is a pro-inflammatory molecular that has been implicated in obesity-mediated type 2 diabetes. Obesity is a host factor that adversely influences breast cancer prognosis [2] [42]. There is evidence that insulin resistance may develop after breast cancer adjuvant therapy [41] and a recent prospective study, reported that increased resistin levels coincided with the concurrent increase in serum insulin and insulin resistance following treatment (surgery followed by chemotherapy and radiotherapy) among stage II-III breast cancer patients in an adiposity independent way [35]. It is therefore possible that derangement of glucose metabolism through insulin resistance may be a result of late toxic effects of chemotherapy possibly due to impaired pancreatic beta-cell function. However, in our present study all patients received chemotherapy and so any differential effect cannot be due to the chemotherapy alone. Recent reports strongly suggest that resistin production in humans is largely from macrophages rather than adipose tissue [30, 33, 44]. Insulin pathophysiology has been associated with inflammatory markers independent of BMI in subjects at risk of type-2-diabetes [45]. Additionally, in transgenic mice, production of human resistin from macrophages was associated with increased inflammation and contributed to the acquisition of insulin resistance [33]. Our current proteomic findings add to the evidence suggesting resistin is a potential surrogate marker of disturbed insulin pathophysiology and inflammation that could provide an explanation for the observed association between higher resistin level and improved DFS.

As an ancillary finding, resistin levels were significantly higher in LN positive vs. LN negative patients, irrespective of outcome group ($p = 0.0037$) (**Figure 3C**). A regression model further examined this trend where LN status demonstrated a significant association with resistin measurements. Resistin overexpression was found to correlate with node negative status (p-value = 0.0428). This trend in combination with the results from the association testing, provide further evidence that resistin and nodal status could be linked (**Supplementary Information 6**). During inflammation, macrophages can be both a major source of resistin and themselves able to respond to resistin in an autocrine loop, leading to

an increase in pro-inflammatory 'M1-like' macrophages and a reduction in anti-inflammatory 'M2-like' macrophages [33, 46]. Given that the lymph node status existed at presentation and all patients received chemotherapy, we considered whether the over-expression of resistin *per se*, may have influenced the tumour micro-environment to exert a suppressive effect on tumour cell motility or extravasation. The association of anti-inflammatory 'M2-like' monocytes and macrophages with metastases in preclinical models [47] provides a possible mechanism whereby increased resistin levels could lead to a lower potential for metastatic development possibly through promoting or reflecting a pre-existing pro-inflammatory tumour microenvironment. To further explore this hypothesis, the *post priori* examination of the downstream differentially expressed proteins between the good vs. poor outcome groups using the Diseases & Functions module of Ingenuity Pathway analysis identified the inflammatory response, leucocyte infiltration (also implicating resistin), lymphocyte migration and recruitment of phagocytes to be significantly induced biological processes ($p < 0.0001$, *z-score* > 2) (**Figure 5**). Overall, improved prognosis associated with increased resistin levels may indicate an immunomodulatory role of this protein during early breast tumour development limiting the ability of the tumour primary cells to spread to distant sites. Further examining the mechanistic link between circulating resistin levels and patient LN status was beyond the scope of the present study, future studies will be required to explore this hypothesis. This is a relatively small study, and a larger follow-up study is warranted, ideally with pre-treatment serum samples to determine whether the observed specific correlation with metastasis to axillary lymph nodes holds true in all ages. A potential technical limitation was the sample pooling strategy used at the discovery phase, which did not permit the assessment of anticipated inter-individual heterogeneity in protein expression levels. However, extensive sample pooling is more likely to find larger, more consistent, protein differences that are therefore more likely to replicate. In addition the accuracy of relative protein quantitation for resistin was validated with ELISA

measurements against individual serum specimens from a separate validation cohort, and from an *in silico* analysis of an independent cohort at a tissue level.

Conclusions

A high-precision serum proteomics based pipeline identified increased serum resistin to positively correlate with disease-free survival independent of BMI in women with EOBC. High resistin levels were associated with better survival and correlated with less axillary lymph node involvement at presentation. We hypothesize that individuals with early breast cancer who have relatively higher resistin levels may provide an environment from which tumours are less likely to metastasise. Further prospective studies are needed to confirm these findings and elucidate the mechanistic role of resistin in EOBC patients.

List of abbreviations

POSH: Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; HPLC: High performance liquid chromatography; iTRAQ: isobaric tags for relative and absolute quantitation; LC-MS: liquid chromatography-mass spectrometry; OS: overall survival; EOBC: early-onset breast cancer, ROC: receiver operating characteristic, AUC: area under the curve, FN: false negative, FP: false positive, TN: true negative, TP: true positive, HR: hazard ratio.

Declarations

- Ethics approval and consent to participate: The study received ethical approval from the South and West Multi-centre Research Ethics Committee (MREC 00/6/69). POSH is a multicenter prospective observational cohort study of 3000 young women diagnosed with breast cancer in the UK between 2000 and 2008 (http://www.southampton.ac.uk/medicine/research/posh.page). All participating patients signed an informed consent form. The cohort was previously described and a detailed study protocol was published in 2007 [1, 2].
- Consent for publication: Not applicable
- Availability of data and material: All mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PXD008443.
- Competing interests: The authors declare that they have no competing interests
- Funding: Wessex Cancer Trust, Wessex Medical Research (Grants N11 and N12), Hope for Guernsey, the University of Manchester,
- Authors' contributions: B.Z. designed study, performed experiments, interpreted results and wrote manuscript; A.M. interpreted results and wrote manuscript; S.E.T.L., T.I.R., and E.K.P. performed experiments; C.H.W. performed biostatistical

analysis; K.N.P. sample procurement; E.C., R.I.C., D.E. interpreted results and wrote manuscript, P.A.T. designed study and edited manuscript; S.D.G. designed study, interpreted results and wrote manuscript.

- Acknowledgements: We are indebted to Mr. Roger Allsopp and Mr. Derek Coates for establishing a mass spectrometry laboratory in the University of Southampton. We are grateful to the Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank within the University of Southampton and in particular Anna Scibior and Jenna Watt for providing the serum samples. We thank the Visiting Professor Program of King Saud University, Riyadh, Saudi Arabia for their financial support.

References:

1. Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG, group Ps: **Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol**. *BMC Cancer* 2007, **7**:160.
2. Copson E, Eccles B, Maishman T, Gerty S, Stanton L, Cutress RI, Altman DG, Durcan L, Simmonds P, Lawrence G *et al*: **Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study**. *J Natl Cancer Inst* 2013, **105**(13):978-988.
3. Candido Dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, van den Broek AJ, Ellis IO, Green A, Rakha E *et al*: **An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation**. *Breast Cancer Res* 2017, **19**(1):58.
4. Zeidan BA, Townsend PA, Garbis SD, Copson E, Cutress RI: **Clinical proteomics and breast cancer**. *Surgeon* 2015, **13**(5):271-278.
5. Moulder R, Bhosale SD, Goodlett DR, Lahesmaa R: **Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling**. *Mass Spectrom Rev* 2017.
6. Rauniyar N, Yates JR, 3rd: **Isobaric labeling-based relative quantification in shotgun proteomics**. *J Proteome Res* 2014, **13**(12):5293-5309.
7. Meng R, Gormley M, Bhat VB, Rosenberg A, Quong AA: **Low abundance protein enrichment for discovery of candidate plasma protein biomarkers for early detection of breast cancer**. *J Proteomics* 2011, **75**(2):366-374.
8. Opstal-van Winden AW, Krop EJ, Karedal MH, Gast MC, Lindh CH, Jeppsson MC, Jonsson BA, Grobbee DE, Peeters PH, Beijnen JH *et al*: **Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case-control study**. *BMC Cancer* 2011, **11**:381.

9. Garbis SD, Roumeliotis TI, Tyritzis SI, Zorpas KM, Pavlakis K, Constantinides CA: **A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance RP chromatography/nESI-MS2 for the in-depth analysis of the serum proteome and phosphoproteome: application to clinical sera derived from humans with benign prostate hyperplasia**. *Anal Chem* 2011, **83**(3):708-718.

10. Copson ER, Cutress RI, Maishman T, Eccles BK, Gerty S, Stanton L, Altman DG, Durcan L, Wong C, Simmonds PD *et al*: **Obesity and the outcome of young breast cancer patients in the UK: the POSH study**. *Ann Oncol* 2015, **26**(1):101-112.

11. Johnston HE, Carter MJ, Cox KL, Dunscombe M, Manousopoulou A, Townsend PA, Garbis SD, Cragg MS: **Integrated Cellular and Plasma Proteomics of Contrasting B-cell Cancers Reveals Common, Unique and Systemic Signatures**. *Mol Cell Proteomics* 2017, **16**(3):386-406.

12. Al-Daghri NM, Al-Attas OS, Johnston HE, Singhania A, Alokail MS, Alkharfy KM, Abd-Alrahman SH, Sabico SL, Roumeliotis TI, Manousopoulou-Garbis A *et al*: **Whole serum 3D LC-nESI-FTMS quantitative proteomics reveals sexual dimorphism in the milieu interieur of overweight and obese adults**. *J Proteome Res* 2014, **13**(11):5094-5105.

13. Al-Daghri NM, Alokail MS, Manousopoulou A, Heinson A, Al-Attas O, Al-Saleh Y, Sabico S, Yakout S, Woelk CH, Chrousos GP *et al*: **Sex-specific vitamin D effects on blood coagulation among overweight adults**. *Eur J Clin Invest* 2016, **46**(12):1031-1040.

14. Larkin SE, Johnston HE, Jackson TR, Jamieson DG, Roumeliotis TI, Mockridge CI, Michael A, Manousopoulou A, Papachristou EK, Brown MD *et al*: **Detection of candidate biomarkers of prostate cancer progression in serum: a depletion-**

free 3D LC/MS quantitative proteomics pilot study. *Br J Cancer* 2016, **115**(9):1078-1086.

15. Manousopoulou A, Koutmani Y, Karaliota S, Woelk CH, Manolakos ES, Karalis K, Garbis SD: **Hypothalamus proteomics from mouse models with obesity and anorexia reveals therapeutic targets of appetite regulation**. *Nutr Diabetes* 2016, **6**:e204.

16. Papachristou EK, Roumeliotis TI, Chrysagi A, Trigoni C, Charvalos E, Townsend PA, Pavlakis K, Garbis SD: **The shotgun proteomic study of the human ThinPrep cervical smear using iTRAQ mass-tagging and 2D LC-FT-Orbitrap-MS: the detection of the human papillomavirus at the protein level**. *J Proteome Res* 2013, **12**(5):2078-2089.

17. Giannogonas P, Apostolou A, Manousopoulou A, Theocharis S, Macari SA, Psarras S, Garbis SD, Pothoulakis C, Karalis KP: **Identification of a novel interaction between corticotropin releasing hormone (Crh) and macroautophagy**. *Sci Rep* 2016, **6**:23342.

18. Manousopoulou A, Gatherer M, Smith C, Nicoll JAR, Woelk CH, Johnson M, Kalaria R, Attems J, Garbis SD, Carare RO: **Systems proteomic analysis reveals that clusterin and tissue inhibitor of metalloproteinases 3 increase in leptomeningeal arteries affected by cerebral amyloid angiopathy**. *Neuropathol Appl Neurobiol* 2017, **43**(6):492-504.

19. Concato J, Peduzzi P, Holford TR, Feinstein AR: **Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy**. *J Clin Epidemiol* 1995, **48**(12):1495-1501.

20. Peduzzi P, Concato J, Feinstein AR, Holford TR: **Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates**. *J Clin Epidemiol* 1995, **48**(12):1503-1510.

21. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: **A simulation study of the number of events per variable in logistic regression analysis**. *J Clin Epidemiol* 1996, **49**(12):1373-1379.

22. Kuhnast C, Neuhauser M: **A note on the use of the non-parametric Wilcoxon-Mann-Whitney test in the analysis of medical studies**. *Ger Med Sci* 2008, **6**:Doc02.

23. Lumley T, Diehr P, Emerson S, Chen L: **The importance of the normality assumption in large public health data sets**. *Annu Rev Public Health* 2002, **23**:151-169.

24. Sawilowsky SS, Hillman SB: **Power of the independent samples t test under a prevalent psychometric measure distribution**. *J Consult Clin Psychol* 1992, **60**(2):240-243.

25. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z: **An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients**. *Breast Cancer Res Treat* 2010, **123**(3):725-731.

26. Zimmerlin L, Donnenberg AD, Rubin JP, Basse P, Landreneau RJ, Donnenberg VS: **Regenerative therapy and cancer: in vitro and in vivo studies of the interaction between adipose-derived stem cells and breast cancer cells from clinical isolates**. *Tissue Eng Part A* 2011, **17**(1-2):93-106.

27. Koerner A, Kratzsch J, Kiess W: **Adipocytokines: leptin--the classical, resistin--the controversical, adiponectin--the promising, and more to come**. *Best Pract Res Clin Endocrinol Metab* 2005, **19**(4):525-546.

28. Georgiou GP, Provatopoulou X, Kalogera E, Siasos G, Menenakos E, Zografos GC, Gounaris A: **Serum resistin is inversely related to breast cancer risk in premenopausal women**. *Breast* 2016, **29**:163-169.

29. Lee JO, Kim N, Lee HJ, Lee YW, Kim SJ, Park SH, Kim HS: **Resistin, a fat-derived secretory factor, promotes metastasis of MDA-MB-231 human breast cancer cells through ERM activation**. *Sci Rep* 2016, **6**:18923.

30. Sun CA, Wu MH, Chu CH, Chou YC, Hsu GC, Yang T, Chou WY, Yu CP, Yu JC: **Adipocytokine resistin and breast cancer risk**. *Breast Cancer Res Treat* 2010, **123**(3):869-876.

31. Adeghate E: **An update on the biology and physiology of resistin**. *Cell Mol Life Sci* 2004, **61**(19-20):2485-2496.

32. Lee YC, Chen YJ, Wu CC, Lo S, Hou MF, Yuan SS: **Resistin expression in breast cancer tissue as a marker of prognosis and hormone therapy stratification**. *Gynecol Oncol* 2012, **125**(3):742-750.

33. Qatanani M, Szwergold NR, Greaves DR, Ahima RS, Lazar MA: **Macrophage-derived human resistin exacerbates adipose tissue inflammation and insulin resistance in mice**. *J Clin Invest* 2009, **119**(3):531-539.

34. Lu LJ, Gan L, Hu JB, Ran L, Cheng QF, Wang RJ, Jin LB, Ren GS, Li HY, Wu KN *et al*: **On the status of beta-cell dysfunction and insulin resistance of breast cancer patient without history of diabetes after systemic treatment**. *Med Oncol* 2014, **31**(5):956.

35. Coskun T, Kosova F, Ari Z, Sakarya A, Kaya Y: **Effect of oncological treatment on serum adipocytokine levels in patients with stage II-III breast cancer**. *Mol Clin Oncol* 2016, **4**(5):893-897.

36. Liu Z, Shi A, Song D, Han B, Zhang Z, Ma L, Liu D, Fan Z: **Resistin confers resistance to doxorubicin-induced apoptosis in human breast cancer cells through autophagy induction**. *Am J Cancer Res* 2017, **7**(3):574-583.

37. Schott AF, Hayes DF: **Defining the benefits of neoadjuvant chemotherapy for breast cancer**. *J Clin Oncol* 2012, **30**(15):1747-1749.

38. Narod SA: **Breast cancer in young women**. *Nat Rev Clin Oncol* 2012, **9**(8):460-470.

39. Mandusic V, Dimitrijevic B, Nikolic-Vukosavljevic D, Neskovic-Konstantinovic Z, Kanjer K, Hamann U: **Different associations of estrogen receptor beta isoforms, ERbeta1 and ERbeta2, expression levels with tumor size and survival in early- and late-onset breast cancer**. *Cancer Lett* 2012, **321**(1):73-79.

40. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Jois M, Chilamkurti N *et al*: **ExoCarta: A Web-Based Compendium of Exosomal Cargo**. *J Mol Biol* 2016, **428**(4):688-692.

41. Simpson RJ, Kalra H, Mathivanan S: **ExoCarta as a resource for exosomal research**. *J Extracell Vesicles* 2012, **1**.

42. Mathivanan S, Fahner CJ, Reid GE, Simpson RJ: **ExoCarta 2012: database of exosomal proteins, RNA and lipids**. *Nucleic Acids Res* 2012, **40**(Database issue):D1241-1244.

43. Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC: **The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth**. *Nature* 2008, **452**(7184):230-233.

44. Schwartz DR, Briggs ER, Qatanani M, Sawaya H, Sebag IA, Picard MH, Scherrer-Crosbie M, Lazar MA: **Human resistin in chemotherapy-induced heart failure in humanized male mice and in women treated for breast cancer**. *Endocrinology* 2013, **154**(11):4206-4214.

45. Shetty GK, Economides PA, Horton ES, Mantzoros CS, Veves A: **Circulating adiponectin and resistin levels in relation to metabolic factors, inflammatory markers, and vascular reactivity in diabetic patients and subjects at risk for diabetes**. *Diabetes Care* 2004, **27**(10):2450-2457.

46. Zuniga MC, Raghuraman G, Hitchner E, Weyand C, Robinson W, Zhou W: **PKC-epsilon and TLR4 synergistically regulate resistin-mediated inflammation in human macrophages**. *Atherosclerosis* 2017, **259**:51-59.

47. Noy R, Pollard JW: **Tumor-associated macrophages: from mechanisms to therapy**. *Immunity* 2014, **41**(1):49-61.

Table and Figure Legends

Table 1. Clinical characteristics of discovery cohort

Table 2. Clinical characteristics of the validation cohort

Figure 1. Experimental design for the high-precision LC-MS proteomic discovery analysis, data reduction and subsequent targeted validation pipeline

Figure 2. (A) Hierarchical clustering analysis of all differentially expressed proteins (DEPs) (812 proteins at $p \leq 0.05$ with ≥ 2 unique peptides) **(B)** The insulin signaling pathway was significantly over-represented in the DEPs between good and poor outcome groups (Fisher exact p=0.015) using KEGG Pathway analysis with DAVID. Tabulation of the Gene names of the observed differentially expressed proteins constituent to the pathway is provided. **(C)** MetaCore showed that glycolysis/gluconeogenesis was a significantly enriched process in the DEPs between good and poor outcome groups (FDR corrected $p=0.011$). **(D)** Network analysis of differentially expressed proteins using Ingenuity Pathway Analysis showed participation of resistin in the small molecule biochemistry network (score=23; Focus molecules=20).

Figure 3 (A) Serum proteomic analysis of resistin showed higher circulating levels in good compared to poor outcome group. Each points represents the $\log_2$ratio of the reporter ion intensity of each clinical group (good or poor outcome respectively) over the mean of all four reporter ion intensities from both clinical groups produced from a given unique peptide [Good vs. poor outcome iTRAQ mean log2ratio=0.2, SD=0.13 between biological replicates, p=0.009]. (**B**) Resistin ELISA measurements across individual samples from the validation cohort in the good outcome group [n=90, Mean value (SD) = 114.2 (114.5)

ng/mL] compared to the poor outcome group [n=91, Mean value (SD) = 86.8 (57.7) ng/mL] (*p = 0.04*) **(C)** Resistin expression was higher in LN negative vs. LN positive patients, irrespective of outcome group [LN negative group: n=71, Mean value (SD) = 124.8 (107.5) ng/mL; LN positive group: n=110, Mean value (SD) = 84.7 (75.6) ng/mL; *p = 0.0037,* Welch's 2-sample t-test].

Figure 4. (**A**) Receiver operating characteristic (ROC) curve with area under the curve (AUC) = 0.6352 (**B**) Cost function with equivalent penalties for false negatives and false positives (**C**) Distribution plot of 30 false negatives (FN), 33 false positives (FP), 58 true negatives (TN) and 60 true positives (TP). (**D**) *In silico* Kaplan Meier survival analysis of resistin expression at the tissue level.

Figure 5. Significant induction (*p<0.0001*) of the inflammatory response, leucocyte infiltration, lymphocyte migration and recruitment of phagocytes in the good vs. poor outcome group based on downstream differentiated proteins. A z-score > 2 signifies a positive induction effect.

Supplementary Sections Legends

Supplementary Sections 1A and B. POSH Serum Procurement SOPs

Supplementary Section 2. Serum Proteomics Method

Supplementary Section 3. Total Serum Proteome

Supplementary Section 4. Differentially expressed proteins in good vs. poor outcome groups

Supplementary Section 5. ELISA measurements for resistin

Supplementary Section 6. Linear and Generalized Linear Modelling of Resistin, ER, PR, LN and HER-2 clinical parameters

Table 1. Clinical characteristics of discovery cohort

Clinical characteristics	Good outcome group	Poor outcome group	p-value
n	**203**	**196**	
Age (years)			
Median	37	36	0.89
Range	25-40	18-41	
Relapse (years)			
Median	9.3	1.3	<0.0001
Range	5.0-10.2	0.4-2.0	
BMI (kg/m^2)			
Mean	25.2	26.3	0.13
SD	5.1	5.4	
Histology			
Invasive Ductal Carcinoma	203	190	
Invasive Lobular Carcinoma	0	0	
Unknown	0	6	
Grade 1	10	6	
Grade 2	75	47	
Grade 3	114	137	
Unknown	4	6	
Lymph node status			
Negative	104	61	<0.0001
Positive	95	127	
Undetermined	4	8	
ER status			
Positive	138	108	
Negative	43	88	<0.0001
Unknown	22	0	
PR status			
Positive	87	75	
Negative	79	86	0.43
Unknown	42	35	
HER2 receptor status			
Positive	53	82	
Negative	59	92	0.77
Unknown	91	22	
Triple negative tumours	32	35	
Resection margin			
R0 resection	142	141	
R1 resection	24	22	
Unknown	37	33	
Chemotherapy			
FEC	69	71	
ECMF	28	31	
FEC + Decotaxel	22	14	
AC	16	16	
EC + Paclitaxel	15	12	
EC + Paclitaxel + Gemcitabine	8	8	
EC	8	6	
Nul	22	8	
Other	15	30	

Note: A: Adriamycin; C: Cyclophosphamide; E: Epirubicin; F: 5 FU; M: Methotrexate; $^+$ p-value=0.13 between groups (unpaired T-Test

Table 2. Clinical characteristics of the validation cohort

Clinical characteristics	Good outcome group	Poor outcome group	p-value
n	**90**	**91**	
Age (years)			
Median Range	37 26-40	35 18-40	0.35
Relapse (years)			
Median Range	9.2 5.0-11.2	1.0 0.3-2.0	<0.0001
BMI (kg/m^2)$^+$			
Mean SD	23.3 2.1	23.2 2.3	0.84
Histology			
Invasive Ductal Carcinoma	83	83	
Invasive Lobular Carcinoma	6	7	
Unknown	1	1	
Grade 1	2	1	
Grade 2	30	16	
Grade 3	57	73	
Unknown	1	1	
Lymph node status			
Negative Positive Undetermined	45 45 0	26 65 0	0.001
ER status			
Positive Negative Unknown	59 31 0	41 50 0	0.003
PR status			
Positive Negative Unknown	42 32 16	24 52 15	0.001
HER2 receptor status			
Positive Negative Unknown	24 49 17	35 49 7	0.47
Triple negative tumours	17	22	
Resection margin			
R0 resection	67	67	
R1 resection	7	12	

Unknown	16	12
Chemotherapy		
FEC	27	28
ECMF	22	18
FEC + Docetaxel	5	14
AC	5	5
EC + Paclitaxel	5	4
EC + Paclitaxel + Gemcitabine	2	4
EC	5	1
Nul	10	2
Other	9	15

Figure 1

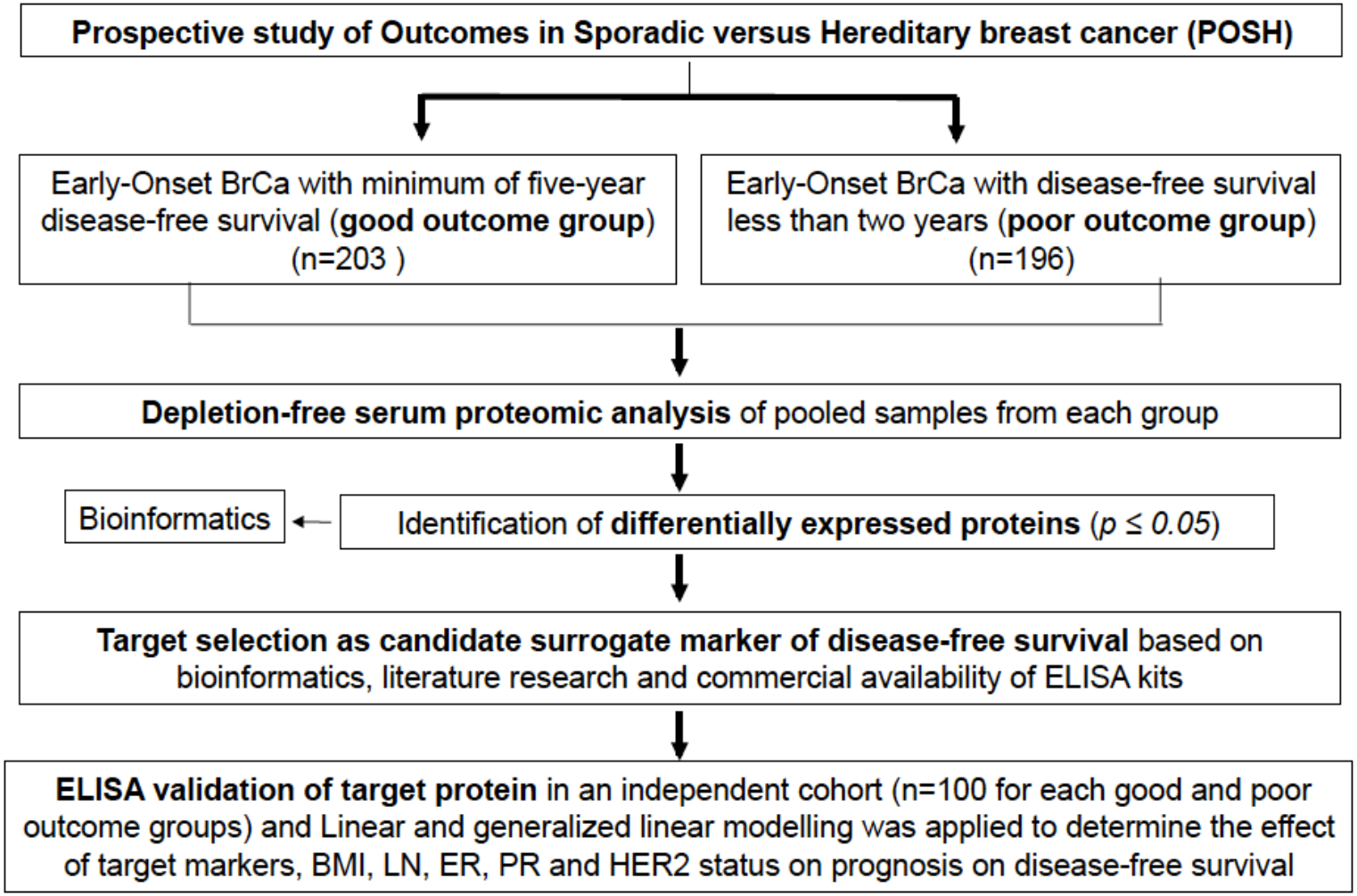
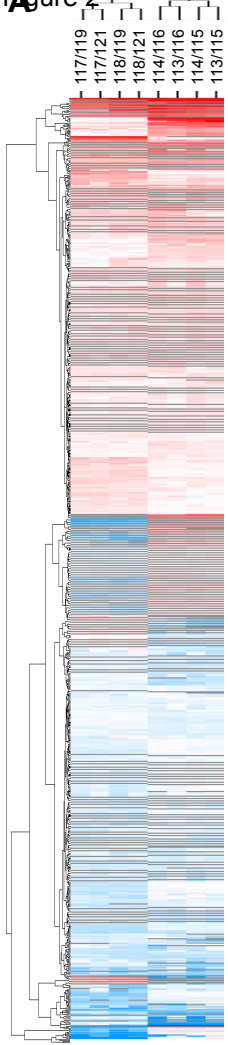
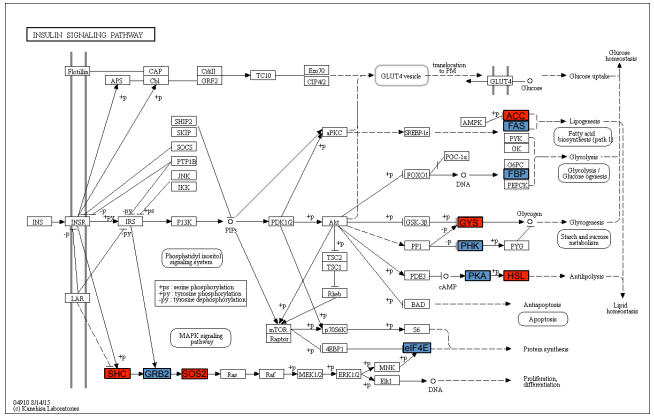


Figure 2

A



B

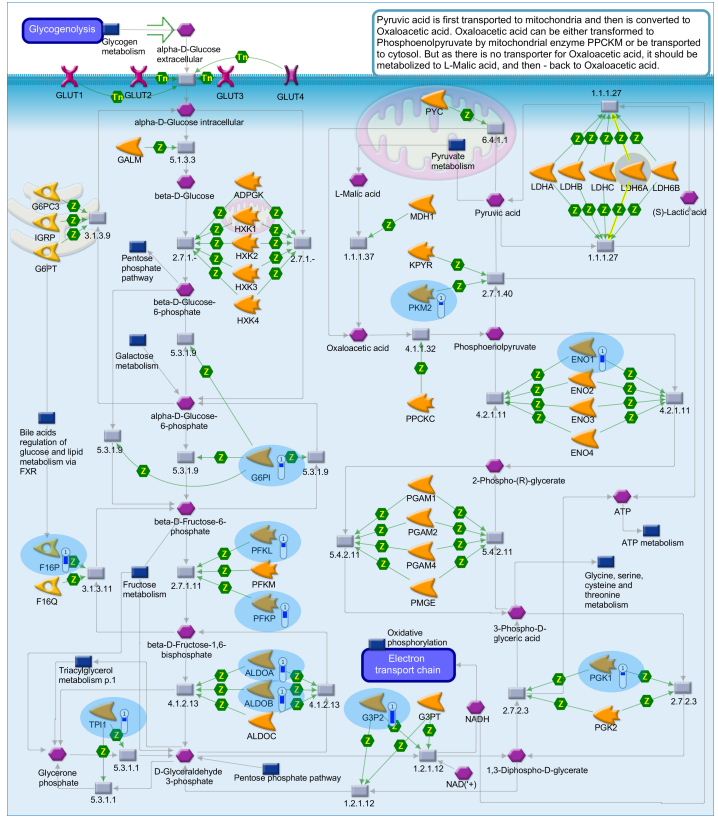


Good vs. poor outcome groups

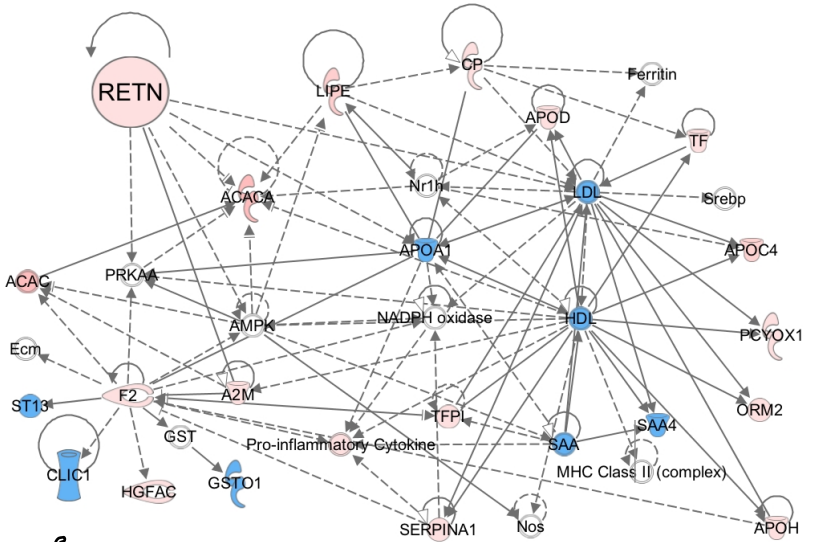
Gene Name	Mean log2ratio	p-value
ACACA	0.9	0.04
LIPE	0.7	0.001
GYS2	0.5	0.005
SOS2	0.4	0.001
SHC4	0.3	0.01
SORBS1	0.2	0.04
GRB2	-0.2	0.002
PRKAR1A	-0.3	0.04
FBP1	-0.5	<0.0001
FASN	-0.6	0.005
EIF4E	-1.0	0.0001
CALM1	-1.1	0.0003

C

Glycolysis and gluconeogenesis



D



down-regulated in good vs. poor outcome groups

Figure 3

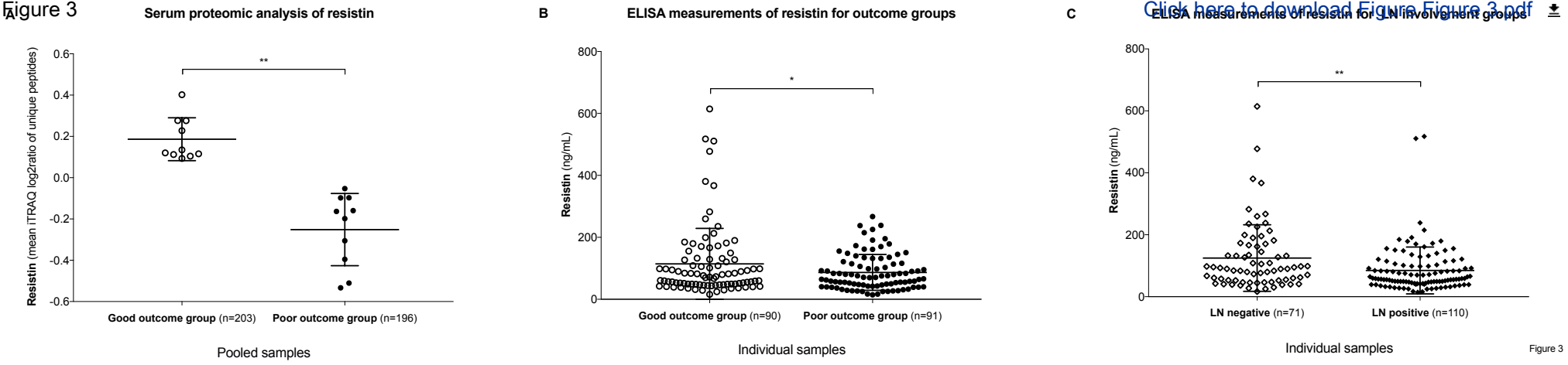
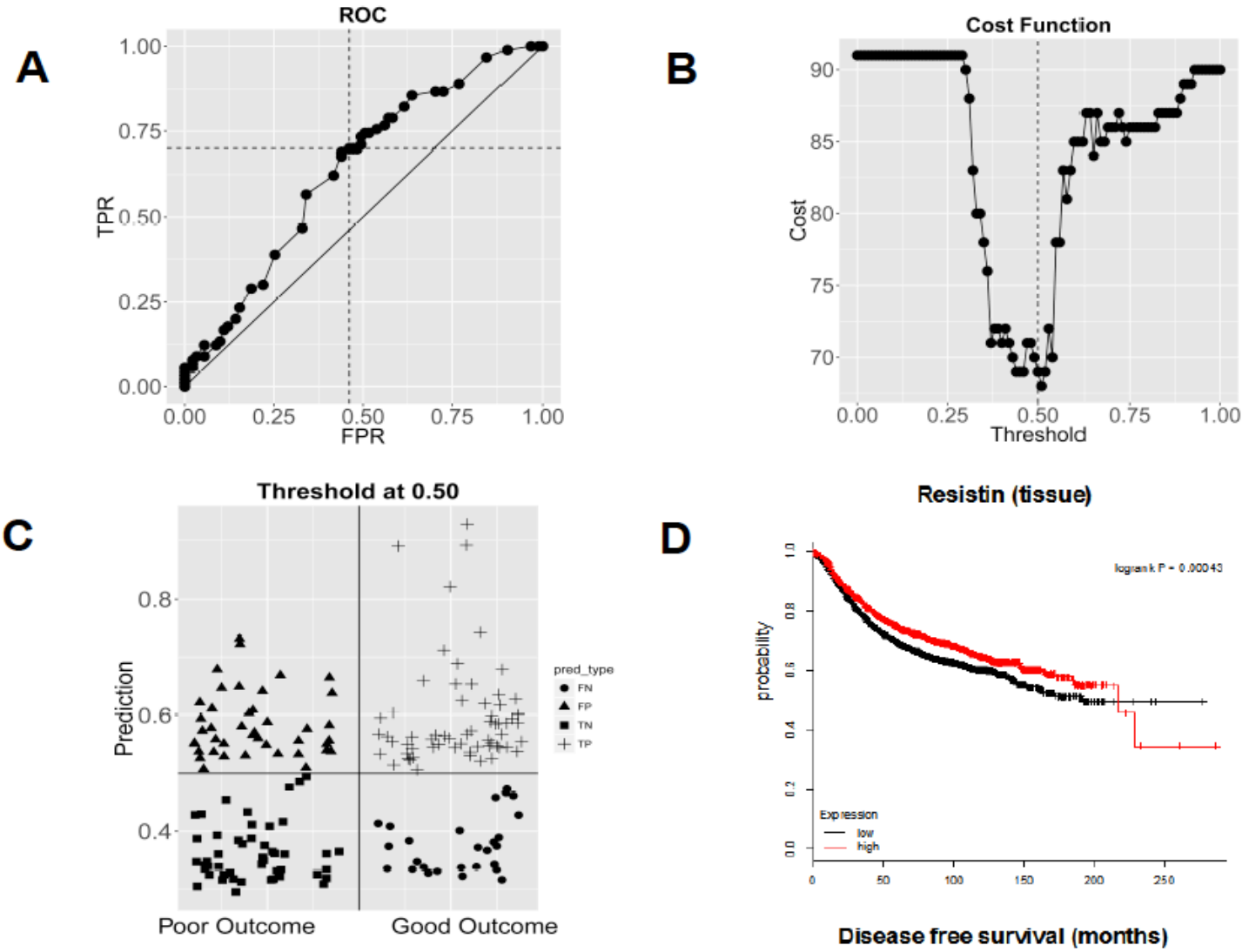
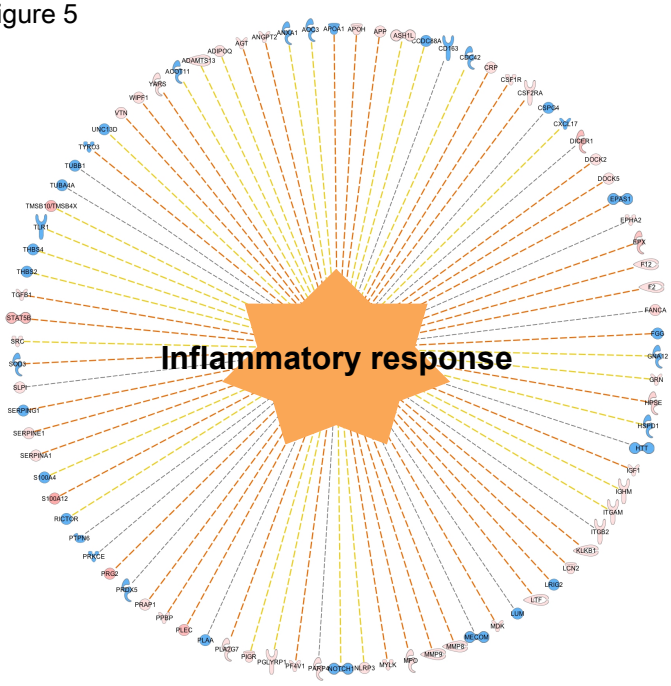


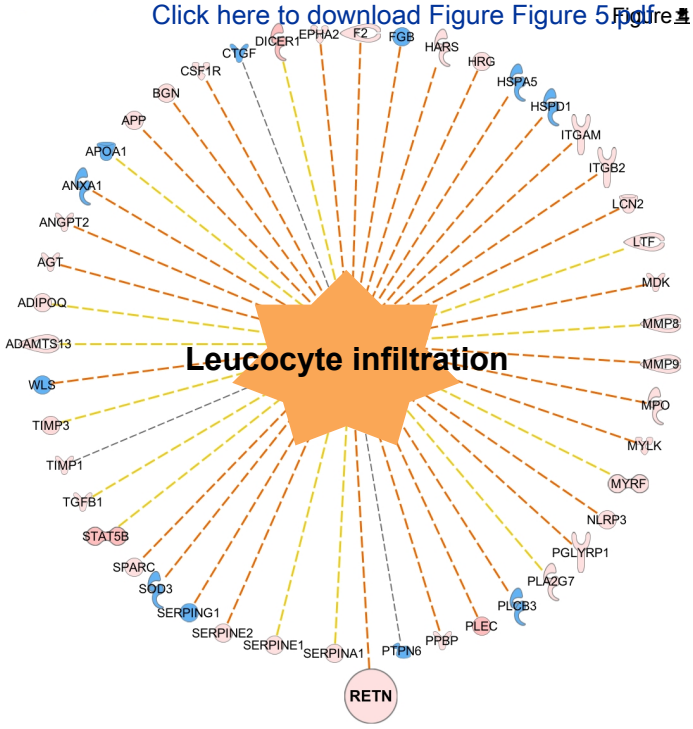
Figure 3

Figure 4

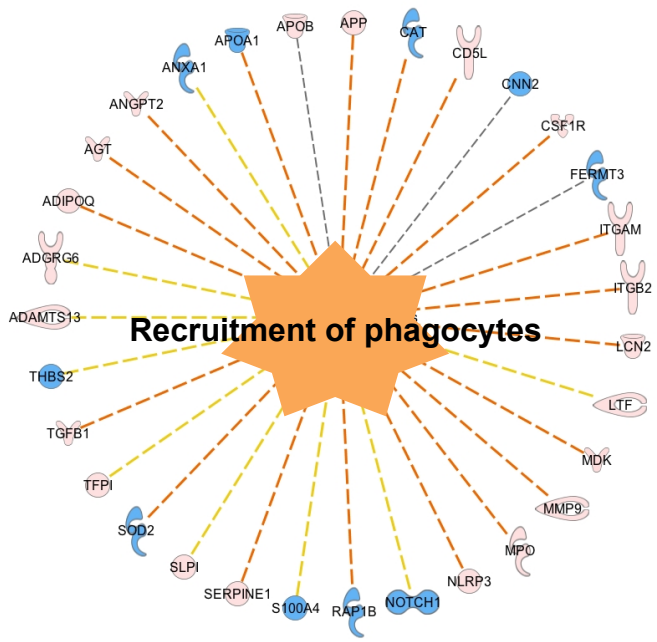




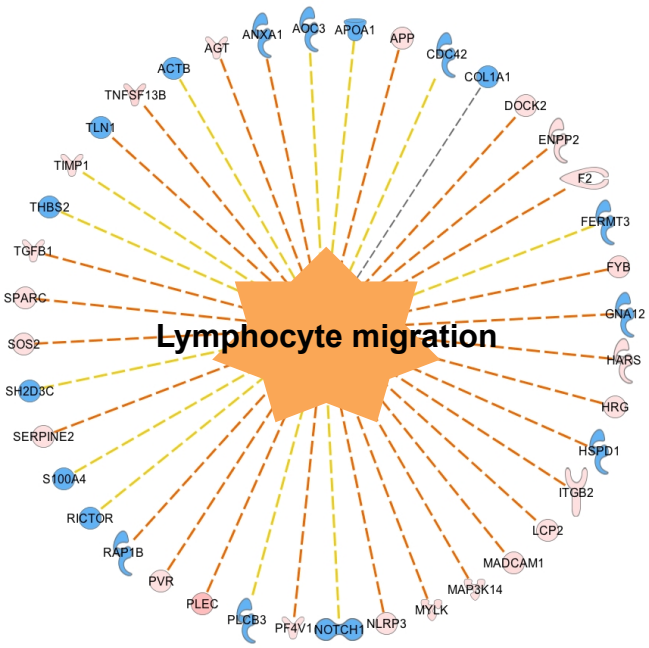
z-score = 2.5; **p-value** < 0.0001



z-score = 2.8; **p-value** < 0.0001



z-score = 2.1; **p-value** < 0.0001



z-score = 2.3; **p-value** < 0.0001

Key

up-regulated in good vs. poor outcome groups

down-regulated in good vs. poor outcome groups

Predicted activation

Leads to activation

Finding inconsistent with effect

Effect not predicted

Medicine

UNIVERSITY OF Southampton

17 December 2017

Dear Professor Chodosh,

Thank you for the opportunity to submit a revised version of our manuscript entitled: **"Circulating resistin in early-onset breast cancer patients with normal body mass index correlates with disease-free survival and lymph node involvement: An agnostic quantitative proteomics study from the multi-center POSH¶ cohort"** by Bashar Zeidan, Antigoni Manousopoulou, Diana J. Garay-Baquero, Cory H. White, Samantha E.T. Larkin, Kathleen N. Potter, Theodoros I. Roumeliotis, Evangelia K. Papachristou, Ellen Copson, Ramsey I. Cutress, Stephen A. Beers, Diana Eccles, Paul A. Townsend and Spiros D. Garbis (**ID: BRCR-D-17-00362)** for consideration for publication in *Breast Cancer Research.*

We also thank the reviewers for their very helpful and insightful comments. In order to address these, we have materially enhanced the quality of our manuscript using a substantially expanded quantitative proteome, comprehensive bioinformatics interrogation, and targeted ELISA validation experiments against an independent multi-center cohort at a statistically significant number of samples. We also employed a more sophisticated biostatistical analysis approach using linear and generalized linear modelling. Below follows a point-by-point reply to the concerns raised by the reviewers.

We would like to verify that all authors have made a substantial contribution to the information or material submitted for publication, and have read and approved the final manuscript. All authors have no direct or indirect commercial financial incentive associated with publishing the article. The results presented in this paper have not been published previously in whole or in part, except in abstract form. The corresponding author acknowledges full responsibility for dealing with all editorial matters having to do with the procession of the paper until its publication.
We look forward to hearing from you.

Lastly, all mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008443. The reviewer account details are as follows: **Username:** reviewer67391@ebi.ac.uk and **Password:** nQwGmmGi.

Respectfully yours,

Spiros D. Garbis, BSc, PhD

Corresponding author

Tel: +44 7554 944 362

Email: S.D.Garbis@soton.ac.uk

Reviewer reports:

Reviewer #1: This work has looked into potential role of circulating adipokine expression for the prognosis of post-treatment response, overall survival and potential risk of long-term insulin resistance in non-obese patients with early-onset breast cancer (EOBC). They mainly used high-precision quantitative mass spectrometry proteomics on a POSH cohort. They found among 117 differentiated adipokines, resistin was found to be up-regulated in the good outcome group [proteomics, p=0.003; ELISA, p=0.03) irrespective of BMI, ER51 and LN status. This could be a candidate marker of longer OS in non-obese patients with EOBC. The work is well designed and has some interesting results. I have the following comments.

1. The conclusion is only based on one cohort as authors acknowledged. Are there any other public similar cohorts for doubling confirming the results?

Author response

We thank the reviewer for highlighting this issue. To the best of our knowledge there are no publicly available data on the adipocytokine levels in normal-weight women with early onset breast cancer. To further verify the validity of our findings in the revised manuscript, we have measured the levels of resistin, a key adipocytokine, using ELISA against an independent set of patients (n=181) from the multi-centre and UK nationwide POSH cohort. No other such cohort was available nationally or internationally. The revised manuscript has been substantially revised to account for the additional validation experiments with extensive biostatistical processing using linear and generalized linear modelling, which further confirm and potential clinical utility of the findings and conclusions made in the original manuscript version.

2. The authors should look into resistin in other breast cancer cohorts in addition EOBC and see if this protein has any significance.

Author response

We thank Reviewer #1 for this interesting observation. Examining the levels of resistin in other types of breast cancer (e.g. post-menopausal breast cancer cases in normal weight or overweight/obese women) has been reported by other groups and was beyond the scope of the present study targeting early onset cancer patients (25 to 40 years old) with normal weight. However, further assessment of resistin expression using the current study pipeline constitutes a future perspective for our group. The revised manuscript now includes this future perspective in the discussion section.

3. Are there other proteins besides resistin showing such better outcome? What about down-regulated proteins?

Author response

We think this comment is highly relevant and we thank Reviewer #1 for bringing up this issue. Indeed, except for resistin 811 additional up- and down-regulated adipokines were differentially expressed between the good vs. poor outcome groups out of a total of 5346 protein profiled. This expanded proteome coverage was made possible by including the analysis results from the additional, higher molecular weight proteomes observed against the same discovery sample set from the original manuscript version, which only included the low molecular weight proteome. However, validating other up- or down-regulated proteins was beyond the scope of the present study. We focused on serum resistin given its strong and interdependent association with the insulin signaling, glycolysis/gluconeogenesis, sugar and fatty acid metabolism and immunological pathways as a candidate marker of EOBC prognosis. Such a substantive, more pleiotropic association of resistin with these biological pathways was made possible because of the expanded proteome coverage and its biological interpretation with a combination of commercial and licensed bioinformatics software tools. Lastly, this more comprehensive assessment of resistin, within the context of the EOBC cohort examined, lead to the generation of a novel hypothesis that we intend to examine as future perspective. These elements are fully described in the revised manuscript text with additional figures, tables and references.

4. I understand this is a proteomic biomarker study. I am curious the gene expression levels for these 117 adipokines and if they are differentially expressed.

Author response

We thank Reviewer #1 for this interesting observation. It would be of great interest to examine the gene expression levels for these adipokines in the adipose tissue of EOBC patients, however such samples were not available from women in the POSH cohort.

5. Are there any literature reported proteomic biomarkers? If yes, the authors should compare them with resistin and discuss about it.

Author response

We thank Reviewer #1 for pointing this out and we apologise for the oversight. An expanded literature review of biomarker discovery efforts using quantitative proteomics approaches and how it relates to our study approach is now included in the Introduction section. However, our study constitutes the first-ever observation focusing on resistin expression, a key adipokine protein, at the serological level of EOBC patients using a unique depletion-free quantitative proteomics approach. Furthermore, the most comprehensive serum proteome coverage observed to date, thanks to the technical merits of our unique methodological approach, further solidified the potential clinical utility of resistin as a novel prognosis marker of EOBC patients.

Reviewer #2: In the manuscript by B Zeidan et al. the authors describe studies aimed at the identification and validation of the protein resistin in non obese early-onset breast cancer patients

and shown increased presence in patients with a good overall survival as defined by >=5 years overall survival. Differential protein analysis comparing good outcome vs poor outcome identified many proteins in the serum. Utilizing pathway analysis, resistin was identified for further study. It would be of interest if there was discussion on any other pathways that might be pursued in future studies.

Author response

We thank Reviewer #2 for pointing this out and we apologise for the oversight. Our revised manuscript has expanded on the full repertoire of quantitative proteomics measurements performed to the sera of EOBC patients. Specifically, this expanded proteome coverage was made possible by including the analysis results from the additional, higher molecular weight proteomes observed against the same discovery sample set from the original manuscript version, which only included the low molecular weight proteome. This expanded differentially expressed proteome allowed for a more comprehensive biochemical and molecular biology inference to be made. In the revised manuscript we have included new figures (**Figure 2** and **Figure 5**) along with the respective text in the Methods, Results and Discussion sections describing with more detail the proteomic results. In particular, the insulin-signalling pathway, glycolysis/gluconeogenesis, glucose/fatty acid metabolism and immune response were significantly enriched in the differentially expressed proteins between good and poor outcome groups. The revised manuscript expands on how these pathways are manifested in EOBC patients and their prognosis. Consequent to this approach a novel hypothesis was derived that now claims that individuals with early breast cancer who have relatively higher resistin levels may provide an environment from which tumours are less likely to metastasise.

The levels of resistin were validated in individual samples, however, the difference between the two groups was small albeit had statistical significance. Ideally, if more of the POSH sample could be accessed for individual testing to provide a larger sample size, that would be ideal.

Author response

We thank Reviewer #2 for raising this important issue. As part of the revised manuscript, we performed individual ELISA measurements of resistin in an independent validation cohort (n=181). As stated in the method section of the revised manuscript: “The size of the validation cohort was based on the logistic models requiring a minimum of 10 events per predictor variable (see references below, included in the revised manuscript), which in our case included ER, PR, HER-2, LN, and BMI status. For the validation cohort, the same inclusion and exclusion criteria as described above were applied but, additionally, samples used in the discovery phase were excluded”.

References

Concato J, Peduzzi P, Holford TR, Feinstein AR: **Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy**. *J Clin Epidemiol* 1995, **48**(12):1495-1501.

Peduzzi P, Concato J, Feinstein AR, Holford TR: **Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates**. *J Clin Epidemiol* 1995, **48**(12):1503-1510.

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: **A simulation study of the number of events per variable in logistic regression analysis**. *J Clin Epidemiol* 1996, **49**(12):1373-1379.

The lack of correlation between Resistin and BMI, ER status, and LN status could be due to the small sample size and would warrant further analysis in an independent data set or sampling more from the POSH cohort.

Author response

We thank Reviewer #2 for this interesting and highly relevant observation. Our revised manuscript thoroughly addresses this concern. Namely, based on the individual serum analysis performed to the independent validation cohort (n=181), as described in our reply to the previous comment, linear and generalized linear modelling was applied to determine the effect of BMI, lymph node (LN), ER, PR and HER2 status on resistin expression. The up-regulation of resistin in the good relative to the poor outcome group in the validation cohort (*p=0.04*), was not dependent of the BMI. LN involvement was the only covariate with a significant effect on resistin measurements (*p=0.004*). Furthermore, increased circulating resistin positively correlates with disease-free survival. LN negative compared to LN positive patients had higher levels of resistin. Additional figures and tables with corresponding text in the result, discussion and supplementary sections (including the above descriptions) have been added to account for this multi-parametric assessment.

Overall, the authors acknowledge and addressed the shorting comings of the study adequately and sought to verify the differential protein using an orthogonal technology (microarrays) and an independent data set. The statistical analysis used throughout the study is sound.

Reviewer #3: The manuscript by Zeidan et al attempts to uncover protein biomarkers associated with prognosis for chemotherapy response and OS in early onset breast cancer patients. The topic is of high importance to the field, but despite this enthusiasm is diminished by several weaknesses that appear to limit impact.

1. Despite the title claims that serum resistin levels appear to correlate with chemotherapy response/outcome, as the authors themselves state- the discovery set is marred by a hard-wired bias insofar as the poor outcome group has higher LN frequency and triple-negative tumors. They do not appear to deploy the correct statistical methods to correct for the interactions and it is not apparent to me that these biases have been corrected for. Simply showing that resistin levels do not seem to correlate with LN and ER status alone is not the same thing as showing that resistin levels

predict outcome independent of ER and LN. It seems to me that the authors have not provided or described the evidence to the latter.

Author response

We thank Reviewer #3 for these interesting and insightful comments. We apologise for not including information on the triple negative tumours of the discovery cohort in the initial manuscript. This information for patients in both the discovery and validation cohorts has been included in the revised Tables 1 and 2. The triple negative tumours between the two groups of the discovery sample sets were comparable (n= 32 and 35 for the good and poor outcome groups respectively).

As part of the revised manuscript, we performed individual ELISA measurements of resistin in an independent validation cohort (n=181). Such a sample size for the validation cohort was based on the logistic models requiring a minimum of 10 events per predictor variable (see references below, included in the revised manuscript) for each of the good and poor outcome groups, which in our case included ER, PR, HER2, LN, and BMI status. For the validation cohort, analogous inclusion and exclusion criteria were applied. Additionally, linear and generalized linear modelling was applied to determine the effect of BMI, lymph node (LN), ER, PR and HER2 status on resistin expression. The up-regulation of resistin in the good relative to the poor outcome group in the validation cohort (*p=0.04*) was not dependent of the BMI and ER status. Survival analysis showed that resistin had a moderate effect upon disease-free survival. LN involvement was the only covariate with a significant effect on resistin measurements (*p=0.004*). Furthermore, increased circulating resistin positively correlates with disease-free survival independent of BMI and ER status in women with EOBC. LN negative compared to LN positive patients had higher levels of resistin. Additional Figures and Tables with corresponding text (including the above) in the Results and Discussion sections have been added to account for this multi-parametric statistical assessment.

References

Concato J, Peduzzi P, Holford TR, Feinstein AR: **Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy**. *J Clin Epidemiol* 1995, **48**(12):1495-1501.

Peduzzi P, Concato J, Feinstein AR, Holford TR: **Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates**. *J Clin Epidemiol* 1995, **48**(12):1503-1510.

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: **A simulation study of the number of events per variable in logistic regression analysis**. *J Clin Epidemiol* 1996, **49**(12):1373-1379.

2. To judge more fully the impact of the findings, the authors must validate their findings from the discovery set in a blinded independent study set of samples, applying an appropriate cut-point and

determining ROC output. This is especially so since the authors did not use independent serum samples in discovery - used only pooled samples.

Author response

We thank Reviewer #3 for raising this essential issue. In the revised manuscript we performed individual ELISA measurements of resistin in an independent set of 181 EOBC patients from the UK nationwide and multi-centre POSH cohort. Specifically, of the randomly selected patients, n=90 from the good outcome group and n=91 samples from the poor outcome group were subjected to ELISA analysis. An ROC output was determined based on the resistin measurements from the independent validation cohort and an appropriate cut-off point was applied. Additional Figures with corresponding text descriptions in the result and discussion sections have been added to account for the ROC assessment.

3. The authors should comment on the use of post-treatment samples for their discovery set. There is no information about this- how long after treatment was the serum obtained? Are there differences in the timing of the serum draw between patients? How do the authors know that the differences in the poor v good outcome aren't due to preanalytical confounders in this aspect?

Author response

We thank Reviewer #3 for raising this issue and we apologise for this oversight. The revised manuscript now includes Standard Operation Procedures used for the POSH study as the **Supplementary Section 1**. Suitable description with references are included in the revised manuscript. We followed vigilant measures including strict SOP adherence for sample collection, preparation and storage, standardised and automated MS analysis and further validation of all samples in the same expirement to eliminate potential "batch effect".

4. The authors should comment on why it would be that nearly 25% of the entire blood proteome appears to be different between the two subgroups studied. Why would you expect such a systemic difference and would this lead one to believe that the resistin findings are non-specific? Is this a methodological weakness when using pooled samples for discovery?

Author response

We thank Reviewer #3 for this interesting and highly useful comment. In the present study, we identified a total of 5,346 unique proteins were analyzed (peptide FDR $p \leq 0.05$). Of these, 812 proteins were differentially expressed in the good vs. poor outcome group and showed significant enrichment for the insulin signalling pathway ($p=0.015$). This translates to about a 15.2 % of the total proteome profile, which constitutes a reasonable degree of differential expression for a given quantitative proteomics study. This differentiated proteome may indeed reflect a distinct tumour biology background in EOBC patients with good vs. poor outcome but could also be partly attributed to the pooling of samples used for the discovery phase. However, the extensive pooling that was used between the biological replicates for the discovery set of experiments was used to normalize out the inherent heterogeneity of clinical presentation between patients

while at the same time preserving the more consistent, and thus potentially more constitutively important, differentially expressed proteins between the good and poor outcome groups.

To address the potential non-specific and/or false positive biomarker discovery, we have conducted a further independent validation set of ELISA analysis against individual samples. To address accurate protein inference, ELISA was used as the measurement approach for the validation cohort as it allowed the analysis of the intact form of resistin whereas discovery proteomics allows the assessment of its expression at the derived peptide level that resulted from the trypsin proteolysis step. The differential serum resistin expression made with ELISA was concordant with the quantitative proteomic findings. This indicates a real biological trend and the specific pathway(s) involved in such observation will be explored in future work. Furthermore, the higher levels of resistin have been verified at the tissue level using a publicly available microarray database to correlate with good prognosis in breast cancer, suggesting its tissue-specificity.