

## RESEARCH ARTICLE

## Open Access



# Increased circulating resistin levels in early-onset breast cancer patients of normal body mass index correlate with lymph node negative involvement and longer disease free survival: a multi-center POSH cohort serum proteomics study

Bashar Zeidan<sup>1</sup>, Antigoni Manousopoulou<sup>2</sup>, Diana J. Garay-Baquero<sup>2,3</sup>, Cory H. White<sup>3,7</sup>, Samantha E. T. Larkin<sup>1</sup>, Kathleen N. Potter<sup>1</sup>, Theodoros I. Roumeliotis<sup>2,5</sup>, Evangelia K. Papachristou<sup>2,6</sup>, Ellen Copson<sup>1</sup>, Ramsey I. Cutress<sup>1</sup>, Stephen A. Beers<sup>1</sup>, Diana Eccles<sup>1</sup>, Paul A. Townsend<sup>4,8\*</sup> and Spiros D. Garbis<sup>1,2\*</sup>

## Abstract

**Background:** Early-onset breast cancer (EOBC) affects about one in 300 women aged 40 years or younger and is associated with worse outcomes than later onset breast cancer. This study explored novel serum proteins as surrogate markers of 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 5-year disease-free survival (DFS) (good outcome group,  $n = 203$ ) or DFS of less than 2 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 Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) cohort ( $n = 181$ ) was used to validate expression levels of the 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 and human epidermal growth factor receptor 2 status on patients' outcome.

(Continued on next page)

\* Correspondence: [paul.townsend@manchester.ac.uk](mailto:paul.townsend@manchester.ac.uk); [s.d.garbis@soton.ac.uk](mailto:s.d.garbis@soton.ac.uk)

<sup>4</sup>Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science, Manchester Biomedical Research Centre, Health Innovation Manchester, Manchester, UK

<sup>1</sup>Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, UK

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

**Results:** A total of 5346 unique proteins were analysed (peptide FDR  $p \leq 0.05$ ). Of these, 812 were differentially expressed in the good vs poor outcome groups 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 ( $p < 0.0001$ ,  $z\text{-score} > 2$ ). Survival analysis showed that resistin overexpression was associated with improved DFS.

**Conclusions:** Higher circulating resistin correlated with node-negative patients and longer DFS independent of BMI and ER status in women with EOBC. Overexpression of serum resistin in EOBC may be a surrogate indicator of improved prognosis.

**Keywords:** Quantitative serum proteomics, Resistin, Insulin resistance, Glycolysis/gluconeogenesis, lymph-node involvement, Early-onset breast cancer

## Background

Approximately one in 300 women aged 40 years or younger is 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 utilized quantitative LC-MS proteomic methods that do not depend on prior affinity enrichment or depletion of plasma/serum which 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].

## 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-centre prospective observational study of EOBC patients, aged 40 years or younger and treated with standard therapies according to local protocols (Additional file 1: Section 1) [1, 2, 10]. Patients included in this study received anthracycline-based chemotherapy. For the discovery phase, patients were selected based on the 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 full patient clinico-pathological characteristics are presented in Table 1. The study design is summarized in Fig. 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 (Additional file 1: Sections 1A and 1B) [1, 2]. For the good outcome group, using the randomization function of

**Table 1** Clinical characteristics of the discovery cohort

Clinical characteristics	Good outcome group	Poor outcome group	<i>p</i> value
<i>n</i>	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 <sup>2</sup> )			
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	< 0.0001
Negative	43	88	
Unknown	22	0	
PR status			
Positive	87	75	0.43
Negative	79	86	
Unknown	42	35	
HER2 receptor status			
Positive	53	82	0.77
Negative	59	92	
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	

**Table 1** Clinical characteristics of the discovery cohort (Continued)

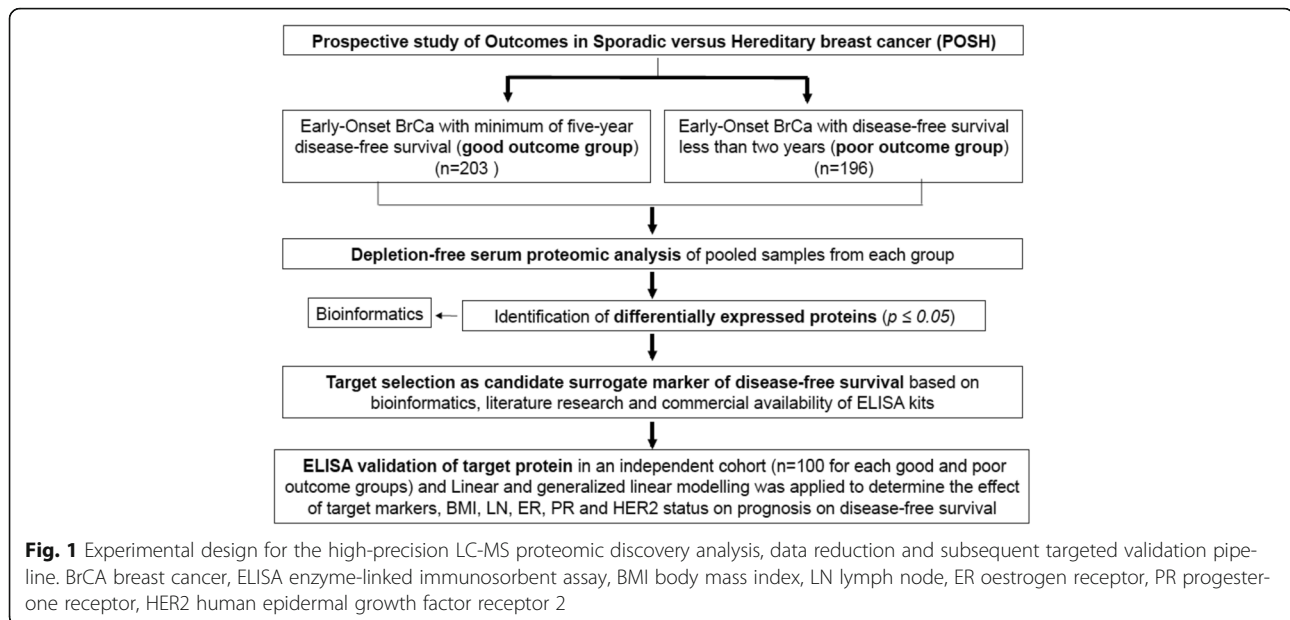
Clinical characteristics	Good outcome group	Poor outcome group	<i>p</i> value
ECMF	28	31	
FEC + docetaxel	22	14	
AC	16	16	
EC + paclitaxel	15	12	
EC + paclitaxel + gemcitabine	8	8	
EC	8	6	
Null	22	8	
Other	15	30	

*A* adriamycin, *B* body mass index, *C* cyclophosphamide, *E* epirubicin, *ER* oestrogen receptor, *F* 5-fluorouracil, *HER2* human epidermal growth factor receptor 2, *M* methotrexate, *PR* progesterone receptor, *SD* standard deviation  
\*Unpaired T-test between groups

Microsoft Excel (2011), individual 20- $\mu$ l aliquots from 102 and 101 specimens respectively were 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  $\mu$ l from each sample pool was mixed with 400  $\mu$ l 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  $\mu$ g protein content derived from the respective SEC segments was prepared. Briefly, the segmented protein fractions were subjected to dialysis purification and lyophilized to dryness. The purified proteins were re-solubilized in 200  $\mu$ l dissolution buffer (0.5 M triethylammonium bicarbonate, 0.05% SDS), quantified and subjected to proteolysis with trypsin using a standardized protocol. The tryptic peptide mixtures generated from each of the four segments were then isobaric stable isotope labelled with the iTRAQ reagents (per manufacturer specifications) for each of the good and poor outcome groups (and their biological replicates), and were pooled. The resulting iTRAQ peptides were initially fractionated with alkaline C<sub>8</sub> reverse phase (RP) liquid chromatography [13, 15]. Each peptide fraction was further separated with on-line nano-capillary C<sub>18</sub> reverse phase liquid chromatography under acidic conditions, subjected to nano-spray 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  $\log_2$  transformed. Proteins identified with a minimum of two unique peptides and a one-sample  $t$ -test  $p \leq 0.05$  were considered as differentially expressed between the 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 Additional file 2: Section 2: Supplementary Methods - Serum Proteomics.

### 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, 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 randomization 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 (Table 2). 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 (USCN Life Sciences Inc., Wuhan, P.R. China) according to the manufacturer's protocols. Absorbance was measured with the GloMax<sup>®</sup> Discover, Promega plate reader (Thermo Fisher Scientific). Data were analysed in Prism (version 7.0a). Statistical analyses of the ELISA measurements were based on Welch's two-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 a balance of samples in groups and each group is well above the

**Table 2** Clinical characteristics of the validation cohort

Clinical characteristics	Good outcome group	Poor outcome group	<i>p</i> value
<i>n</i>	90	91	
Age (years)			
Median	37	35	0.35
Range	26–40	18–40	
Relapse (years)			
Median	9.2	1.0	< 0.0001
Range	5.0–11.2	0.3–2.0	
BMI (kg/m <sup>2</sup> )*			
Mean	23.3	23.2	0.84
SD	2.1	2.3	
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	45	26	0.001
Positive	45	65	
Undetermined	0	0	
ER status			
Positive	59	41	0.003
Negative	31	50	
Unknown	0	0	
PR status			
Positive	42	24	0.001
Negative	32	52	
Unknown	16	15	
HER2 receptor status			
Positive	24	35	0.47
Negative	49	49	
Unknown	17	7	
Triple-negative tumours	17	22	
Resection margin			
R0 resection	67	67	
R1 resection	7	12	
Unknown	16	12	
Chemotherapy			
FEC	27	28	

**Table 2** Clinical characteristics of the validation cohort (Continued)

Clinical characteristics	Good outcome group	Poor outcome group	<i>p</i> value
ECMF	22	18	
FEC + docetaxel	5	14	
AC	5	5	
EC + paclitaxel	5	4	
EC + paclitaxel + gemcitabine	2	4	
EC	5	1	
Null	10	2	
Other	9	15	

A adriamycin, B/ body mass index, C cyclophosphamide, E epirubicin, ER oestrogen receptor, F 5-fluorouracil, HER2 human epidermal growth factor receptor 2, M methotrexate, PR progesterone receptor, SD standard deviation \**p* = 0.13 between groups (unpaired *t* test)

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–N3 = 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 *coefstest* 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 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 positive (good outcome)

and below which a prediction was determined to be negative (poor outcome). The area under the curve (AUC) measure was calculated 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>). Kaplan–Meier Plotter assesses the effects of 54,675 genes on patient DFS using 5143 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 5346 unique proteins (peptide FDR-corrected  $p \leq 0.05$ ) from all four HP-SEC-derived segments (Additional file 3: Section 3). Of these, 812 proteins were differentially expressed between the good and poor outcome groups ( $p \leq 0.05$ ,  $\geq 2$  unique peptides) (Additional file 4: Section 4) and were subjected to further bioinformatics analysis. The mass spectrometry proteomics data have been deposited with 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 the good and poor outcome groups ( $p = 0.015$ , KEGG Pathway analysis using DAVID) (Fig. 2a). MetaCore pathway analysis identified glycolysis/gluconeogenesis as a significantly enriched process in the differentially expressed proteins between the good and poor outcome groups ( $p < 0.011$ , FDR corrected) (Fig. 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 the good and poor outcome groups. Resistin was a key molecular participant in this network (Fig. 2c), and was chosen for targeted validation based on its previously reported role in breast cancer biology and insulin resistance risk [26–36].

### Resistin ELISA validation measurements

Resistin was measured to be upregulated in the good outcome group from the proteomic discovery stage using pooled serum samples ( $p = 0.009$ ) (Figure 3a). The upregulation 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 (SD) = 114.2 (114.5) ng/ml; poor outcome group,  $n = 91$ , mean (SD) = 86.8 (57.7) ng/ml;  $p = 0.04$ ) (Fig. 3b) (Additional file 5: 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 (Fig. 4a) along with a cost function with equivalent penalties for false negatives and false positives (Fig. 4b, c). 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 (Fig. 4d).

### Prediction of biological function directionality (induction or inhibition)

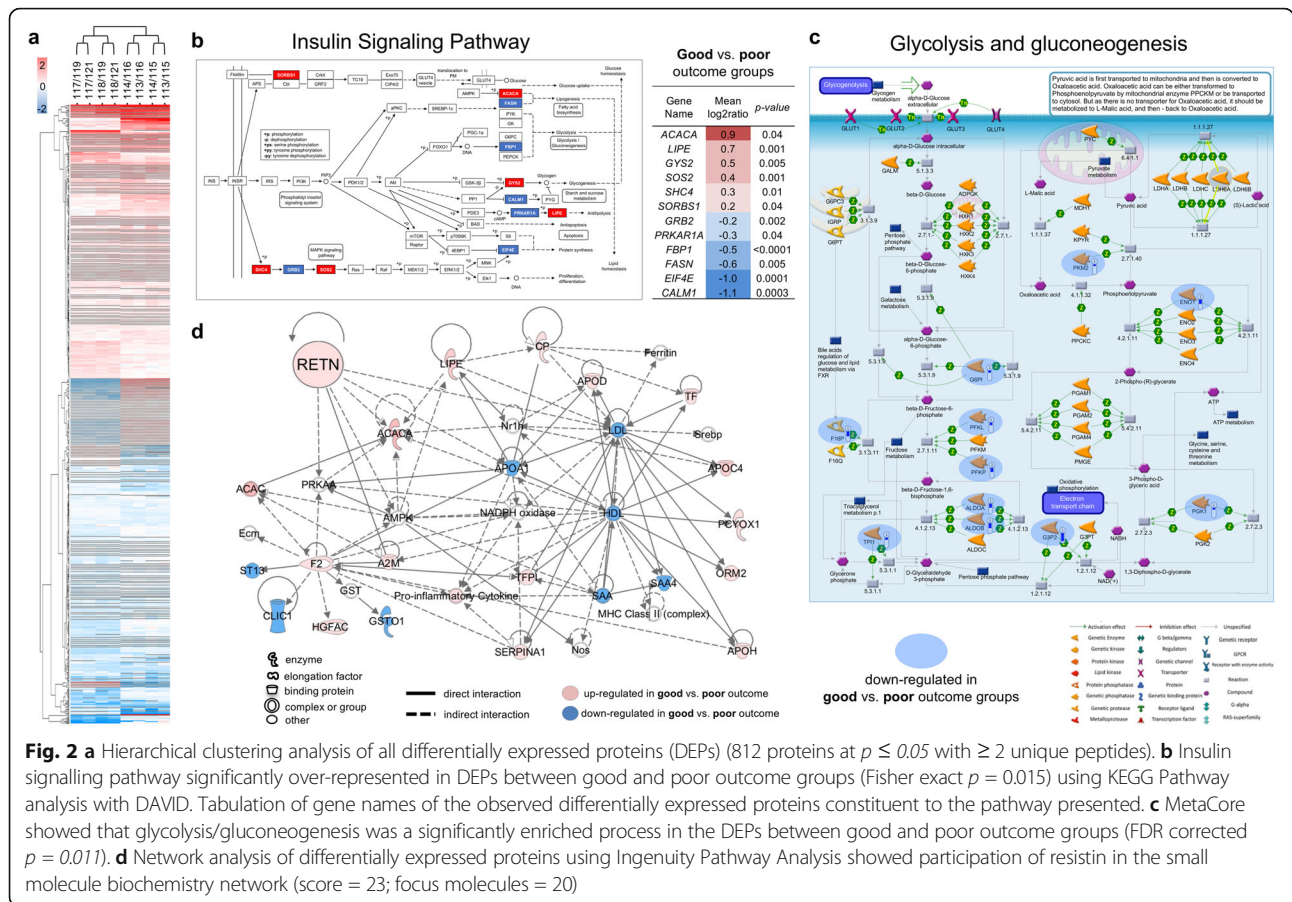
The Diseases & Functions module of Ingenuity Pathway Analysis demonstrated that inflammatory response, leukocyte infiltration, lymphocyte migration and recruitment of phagocytes were significantly induced biological processes based on the downstream differentially expressed proteins of the good vs poor outcome groups. Resistin was specifically found to participate in the activation of leukocyte infiltration (Fig. 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 (Additional file 6: Section 6). LN involvement was found to correlate with worse patient outcome ( $p = 0.004$ ) 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 (SD) = 124.8 (107.5) ng/ml; LN-positive group,  $n = 110$ , mean (SD) = 84.7 (75.6) ng/ml;  $p = 0.0037$ , Welch's two-sample *t* test) (Fig. 3c, Additional file 6: Section 6).

## Discussion

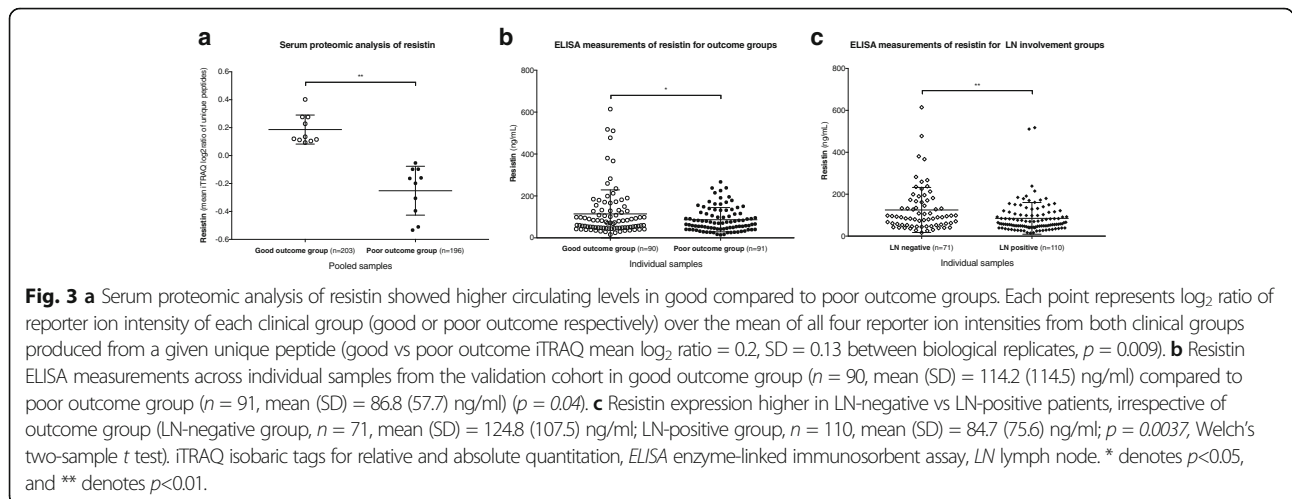
Improvements made in breast cancer survival have been associated with the wider use of neo/adjuvant chemotherapy such as anthracycline/taxane-based treatment [37]. Routine immunohistochemical analysis is used for both prognosis and predictive markers of response to hormonal therapy and trastuzumab (ER/PR and HER2 respectively). Young age [38, 39] and obesity [2] at breast cancer diagnosis have been reported to be independent prognostic

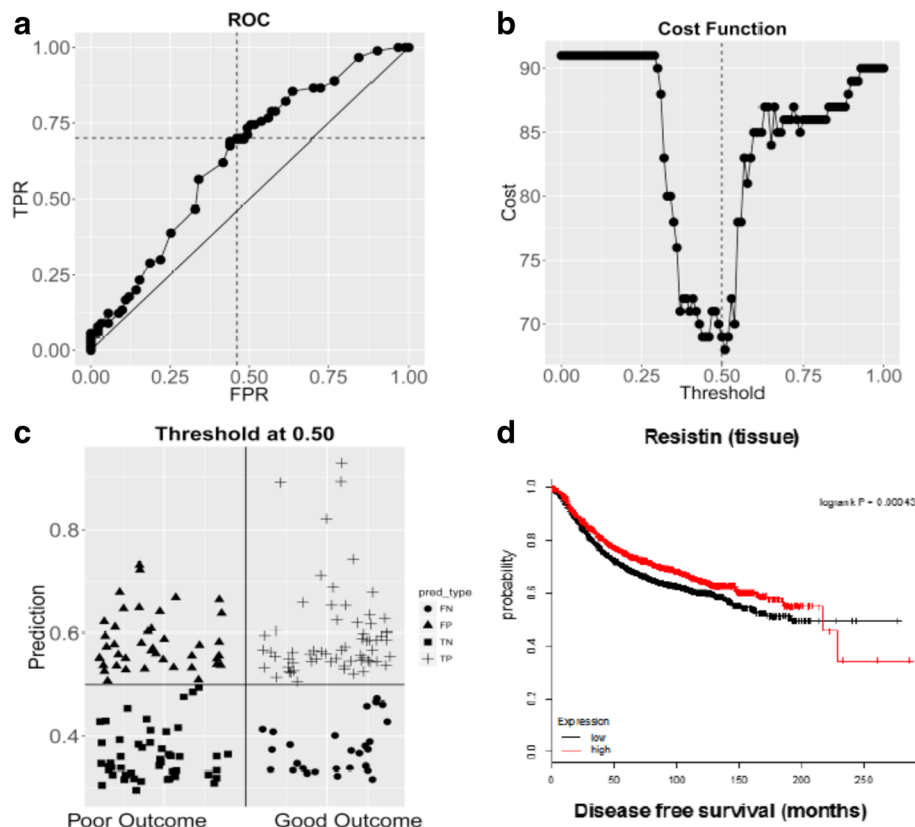


markers of adverse disease outcome. The aim of this study was to find serum proteomic markers of additional prognostic relevance to EOBC outcomes.

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 (Fig. 1).

The applied proteomics method achieved the highest degree of proteome coverage in breast cancer serum to date (5346 unique proteins with peptide FDR  $p \leq 0.05$ ). The methodological feature that led to this comprehensive proteome result was its ability to analyse non-depleted serum that also contains exosome-enriched and other extracellular vesicle-derived proteins in addition





**Fig. 4** **a** Receiver operating characteristic (ROC) curve of the true positive rate (TPR) versus the false positive rate (FPR) with 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

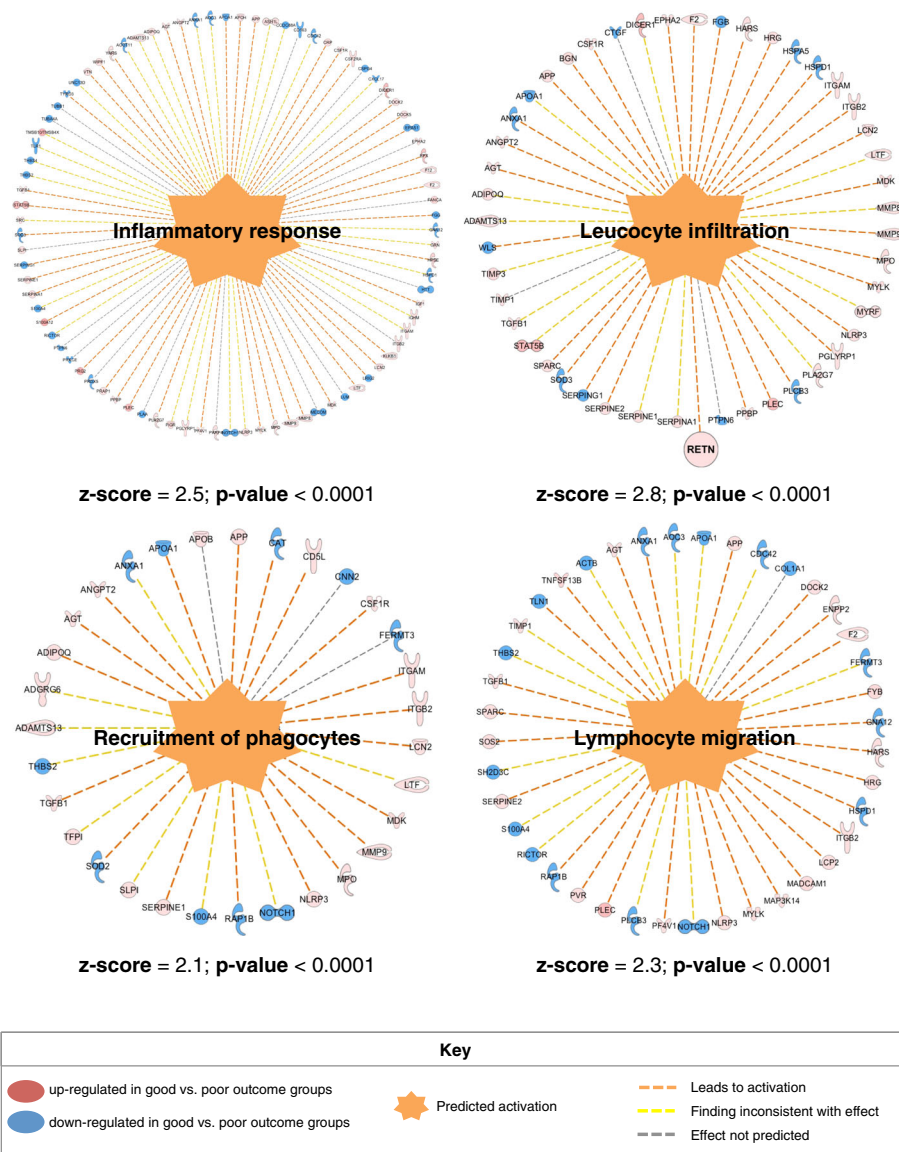
to directly secreted proteins, as reported previously [9, 12, 14]. Such an in-depth analysis was deemed essential for the unbiased interrogation of expected systemic effects and their affiliated biological pathways and networks induced by treatment.

Hierarchical clustering analysis of all 812 differentially expressed proteins (DEPs) is presented in heatmap format in Fig. 2a. The DEPs were then subjected to canonical pathway analysis, which achieved significant enrichment for the insulin signalling ( $p = 0.015$ ) (Fig. 2b) and glycolysis/gluconeogenesis ( $p = 0.011$ ) pathways (Fig. 2c). Interestingly, the majority of observed proteins that encoded for both of 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 (Fig. 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 signalling and the glycolysis/gluconeogenesis pathways were further verified with Ingenuity Pathway Analysis that showed significant enrichment for glucose and fatty acid metabolism (Fig. 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 signalling and glycolysis/gluconeogenesis pathways as a candidate marker of EOBC prognosis.

In agreement with the discovery cohort (Fig. 3a), resistin was found to be upregulated in the good outcome group in the normal weight validation cohort (Fig. 3b). To address accurate protein inference, ELISA was used as the measurement approach for the validation cohort because it allowed the analysis of the intact form of resistin, whereas bottom-up proteomics, as used in this study, allows the assessment of protein expression at the derived peptide level resulting from the trypsin proteolysis step.





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

In this work, both linear and generalized linear regression analysis confirmed that ER, PR and HER2 exhibited a significant degree of interdependence ( $p < 0.05$ ) (Additional file 6: Section 6). A receiver operating characteristic (ROC) curve (Fig. 4a) and associated cost curve (Fig. 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 (Fig. 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 ( $p < 0.001$ ) (Fig. 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 alone (also known to contain macrophages) [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 that 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 ( $p = 0.0037$ ) (Fig. 3c). A regression model further examined this trend where the LN status demonstrated a significant association with resistin measurements. Resistin overexpression was found to correlate with node-negative status ( $p = 0.0428$ ). This trend, in combination with the results from the association testing, provide further evidence that resistin and nodal status could be linked (Additional file 6: Section 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 overexpression 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 spread by promoting 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 ( $p < 0.0001$ ,  $z\text{-score} > 2$ ) (Fig. 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 at all ages. A potential technical limitation was the sample pooling strategy used in 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 the *in-silico* analysis of an independent cohort at the 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 less axillary lymph node involvement at presentation and better survival. 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 metastasize. Further prospective studies are needed to confirm these findings and elucidate the mechanistic role of resistin in EOBC patients.

## Additional files

**Additional file 1:** Sections 1A and 1B presenting POSH Serum Procurement SOPs. (ZIP 244 kb)

**Additional file 2:** Section 2 presenting the serum proteomics method. (PDF 82 kb)

**Additional file 3:** Section 3 presenting the total serum proteome. (PDF 1522 kb)

**Additional file 4:** Section 4 presenting differentially expressed proteins in good vs poor outcome groups. (PDF 489 kb)

**Additional file 5:** Section 5 presenting ELISA measurements for resistin. (PDF 356 kb)

**Additional file 6:** Section 6 presenting linear and generalized linear modelling of resistin, ER, PR, LN and HER-2 clinical parameters. (DOCX 108 kb)

## Abbreviations

AUC: Area under the curve; EOBC: Early-onset breast cancer; HPLC: High-performance liquid chromatography; iTRAQ: Isobaric tags for relative and absolute quantitation; LC-MS: Liquid chromatography–mass spectrometry; POSH: Prospective study of Outcomes in Sporadic versus Hereditary breast cancer; ROC: Receiver operating characteristic

## Acknowledgements

The authors are indebted to Mr Roger Allsopp and Mr Derek Coates for establishing a mass spectrometry laboratory at the University of Southampton. They 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. The authors thank the Visiting Professor Program of King Saud University, Riyadh, Saudi Arabia for their financial support. PAT would like to thank University of Manchester for financial support. This work is dedicated to the memory of Diana Dilkevic.

## Funding

This study was supported by Wessex Cancer Trust, Wessex Medical Research (Grants N11 and N12), Hope for Guernsey, University of Manchester, EU-FP7/HELLENIC NSRF (BIOEXPLORE), EU-FP7 Marie Curie (CANOMICS), Annual Adventures in Research—University of Southampton, EU-Excellence II—Systems Biology Framework FRA-SYS (Grant 4072). The MRC Manchester Academic Health Science Centre Technology Hub: Clinical Proteomics Centre for Stratified Medicine (Grant MR/M008959/1).

## Availability of data and materials

All mass spectrometry proteomics data have been deposited with the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PXD008443.

## Authors' contributions

BZ designed the study, performed experiments, interpreted results and wrote the manuscript. AM interpreted results and wrote the manuscript. DJG-B, SETL, TIR and EKP performed experiments. CHW performed biostatistical modelling and analysis. KNP performed sample procurement. EC, RIC, SAB and DE interpreted results and wrote the manuscript. PAT designed the study, raised funding and edited the manuscript. SDG raised funding, designed the study, interpreted results and wrote the manuscript. All authors read and approved the final manuscript.

## 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 multicentre 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 described previously and a detailed study protocol was published in 2007 [1, 2].

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests. PAT would like to note he is cofounder of Karus Therapeutics Limited, sits on a number of company advisory boards and consults to help various companies but none are involved with or influenced this study.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, UK. <sup>2</sup>Institute for Life Sciences, University of Southampton, Southampton, UK. <sup>3</sup>Clinical and Experimental Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, UK. <sup>4</sup>Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science, Manchester Biomedical Research Centre, Health Innovation Manchester, Manchester, UK. <sup>5</sup>Present address: The Institute for Cancer Research, London, UK. <sup>6</sup>Present address: Cancer Research

Cambridge Institute, University of Cambridge, Cambridge, UK. <sup>7</sup>Present address: Merck Exploratory Science Center, Cambridge, MA, USA. <sup>8</sup>Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

Received: 17 July 2017 Accepted: 25 January 2018

Published online: 22 March 2018

## References

- Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG, POSH Steering Group. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. *BMC Cancer*. 2007;7:160.
- 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):978–88.
- 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.
- Zeidan BA, Townsend PA, Garbis SD, Copson E, Cutress RI. Clinical proteomics and breast cancer. *Surgeon*. 2015;13(5):271–8.
- Moulder R, Bhosale SD, Goodlett DR, Lahesmaa R. Analysis of the plasma proteome using iTRAQ and TMT-based isobaric labeling. *Mass Spectrom Rev*. 2017. <https://doi.org/10.1002/mas.21550>. [Epub ahead of print].
- Rauniyar N, Yates JR 3rd. Isobaric labeling-based relative quantification in shotgun proteomics. *J Proteome Res*. 2014;13(12):5293–309.
- 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):366–74.
- 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.
- Garbis SD, Roumeliotis TI, Tyrizis 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):708–18.
- 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):101–12.
- 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.
- 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):5094–105.
- 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):1031–40.
- 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):1078–86.
- Manousopoulou A, Koutmani Y, Karaliota S, Woelk CH, Manolakas 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.
- 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):2078–89.

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):1495–501.
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):1503–10.
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):1373–9.
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; 23:151–69.
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):240–3.
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):725–31.
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 controversial, adiponectin—the promising, and more to come. *Best Pract Res Clin Endocrinol Metab*. 2005;19(4):525–46.
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;29:163–9.
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):869–76.
31. Adeghate E. An update on the biology and physiology of resistin. *Cell Mol Life Sci*. 2004;61(19–20):2485–96.
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):742–50.
33. Qatanani M, Szwegold 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):531–9.
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):893–7.
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):574–83.
37. Schott AF, Hayes DF. Defining the benefits of neoadjuvant chemotherapy for breast cancer. *J Clin Oncol*. 2012;30(15):1747–9.
38. Narod SA. Breast cancer in young women. *Nat Rev Clin Oncol*. 2012;9(8): 460–70.
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):73–9.
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):688–92.
41. Simpson RJ, Kalra H, Mathivanan S. ExoCarta as a resource for exosomal research. *J Extracell Vesicles*. 2012;1:18374–80.
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):D1241–4.
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):230–3.
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):4206–14.
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):2450–7.
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;259:51–9.
47. Noy R, Pollard JW. Tumor-associated macrophages: from mechanisms to therapy. *Immunity*. 2014;41(1):49–61.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

