Article:
Al Amri, WS, Allinson, LM, Baxter, DE orcid.org/0000-0001-8830-6802 et al. (7 more authors) (2020) Genomic and expression analyses define MUC17 and PCNX1 as predictors of chemotherapy response in breast cancer. Molecular Cancer Therapeutics, 19 (3). pp. 945-955. ISSN 1535-7163

https://doi.org/10.1158/1535-7163.MCT-19-0940

Genomic and expression analyses define MUC17 and PCNX1 as predictors of chemotherapy response in breast cancer

Waleed S Al Amri[1,2], Lisa M Allinson[1], Diana E Baxter[1], Sandra M Bell[1], Andrew M Hanby[1,3], Stacey J Jones[4], Abeer M Shaaban[5], Lucy F Stead[1], Eldo T Verghese[3], Thomas A Hughes[1*]

[1]School of Medicine, University of Leeds, Leeds, UK
[2]Department of Histopathology and Cytopathology, The Royal Hospital, Muscat, Oman
[3]Department of Histopathology, St. James's University Hospital, Leeds, UK
[4]Department of Breast Surgery, Leeds Teaching Hospitals NHS Trust, Leeds, UK
[5]Histopathology and Cancer Sciences, Queen Elizabeth Hospital Birmingham and University of Birmingham, Birmingham, UK

***Corresponding author**: Thomas A Hughes, Wellcome Trust Brenner Building, St James's University Hospital, Leeds, LS9 7TF, UK. tel +44 113 3431984, email t.hughes@leeds.ac.uk

Running title: MUC17 and PCNX1 predict chemoresponse

CoI statement: The authors declare no potential conflicts of interest

ABSTRACT

Poor prognosis breast cancers are treated with cytotoxic chemotherapy, but often without any guidance from therapy predictive markers since universally-accepted markers are not currently available. Treatment failure, in the form of recurrences, is relatively common. We aimed to identify chemotherapy predictive markers and resistance pathways in breast cancer. Our hypothesis was that tumour cells remaining after neoadjuvant chemotherapy (NAC) contain somatic variants causing therapy resistance, while variants present pre-NAC but lost post-NAC cause sensitivity.

Whole exome sequencing was performed on matched pre- and post-NAC cancer cells, which were isolated by laser microdissection, from 6 cancer cases, and somatic variants selected for or against by NAC were identified. Somatic variant diversity was significantly reduced after therapy ($p<0.05$). MUC17 variants were identified in 3 tumours and were selected against by NAC in each case, while PCNX1 variants were identified in 2 tumours and were selected for in both cases, implicating the function of these genes in defining chemoresponse. *In vitro* knock-down of MUC17 or PCNX1 was associated with significantly increased or decreased chemotherapy sensitivity respectively ($p<0.05$), further supporting their roles in chemotherapy response. Expression was tested for predictive value in two independent cohorts of chemotherapy-treated breast cancers (n=53, n=303). Kaplan-Meier analyses revealed that low MUC17 expression was significantly associated with longer survival after chemotherapy, while low PCNX1 was significantly associated with reduced survival.

We concluded that therapy-driven selection of somatic variants allows identification of chemotherapy response genes. With respect to MUC17 and PCNX1, therapy-driven selection acting on somatic variants, in vitro knock-down data concerning drug sensitivity, and survival analysis of expression levels in patient cohorts all define the genes as mediators of and predictive markers for chemotherapy response in breast cancer.

Introduction

Breast cancer is the most common malignancy in women and the second most common overall, and is estimated to cause more than 600,000 deaths annually worldwide (1). Breast cancer has been at the forefront of new cancer treatments, exemplified by the range of molecularly targeted therapies, for example therapies targeting estrogen receptor (ER) or her2, and more recently CDK4/6 or PARP (2). Despite these advances, traditional cytotoxic chemotherapy is still used in treatment of more than a third of primary breast cancer patients (3), and the vast majority of metastatic patients. Included in this are poorer prognosis classifications of primary breast cancer, such larger and more advanced ER-positive cancers, triple negative cancers (negative for ER, her2 and progesterone receptor) (4), which lack suitable alternative systemic therapies, and her2-positive subtypes, which are treated with her2 targeted agents combined with chemotherapy (5). Cytotoxic chemotherapy gives improved outcomes overall (6), but treatment failure is relatively common, as evidenced by recurrences after primary disease, or progression and eventual death in the metastatic setting. Unfortunately, understanding of chemoresistance pathways is incomplete, and most patients are not currently stratified using markers to cytotoxic chemotherapy agents to which they are most likely to respond, or spared chemotherapy if good responses to any are unlikely. An exception may be use of the multigene assay Oncotype DX, which has gained some traction for prediction of chemotherapy benefit in some patient groups in some countries, although evidence supporting this use is not universally-accepted (7). Therefore, there is an urgent need to identify resistance pathways and use these insights to develop and validate predictive markers, allowing improved response rates through stratification, and to identify resistance-related therapeutic targets, which could allow inhibition of resistance thereby eliciting responses in otherwise resistant cancers (8).

Chemotherapy for primary breast cancer is increasingly given in the neoadjuvant setting for at least two reasons. Firstly, tumours can be down-staged, facilitating less radical surgical resections (9). Secondly, it provides an opportunity to assess tumour responses to specific agents using longitudinal imaging, allowing switching to alternative agents if initial responses are inadequate (10). Neoadjuvant chemotherapy (NAC) also provides a powerful research opportunity since matched primary tumour samples from before (diagnostic biopsies) and after therapy (resections) may be available. With respect to chemotherapy response, a relevant hypothesis is that tumour cells that remain post-NAC have characteristics associated with therapy resistance, while characteristics lost from tumour cells in the matched pre-NAC sample are associated with sensitivity; therefore, comparative analysis of

these matched samples can give insights into both resistance and sensitivity (11). Based on this hypothesis, investigators have examined expression pre-NAC and post-NAC and identified individual molecules such as bFGF (12), families of molecules such as xenobiotic drug pumps (13), or pathway-level gene expression changes (14) that were associated with relative resistance. We have taken a genomics approach using these matched tissues. We have determined and compared sub-clonal somatic mutational profiles throughout the exome of purified cancer cells in a cohort of breast tumours pre- and post-NAC, to our knowledge the first such study. Using these data, we aimed to identify genes that host somatic variants selected for or against by NAC; these genes would represent candidate mediators of chemoresponse whether through selection of mutations, as in our screen, or through other pathways acting on gene expression more generally, as testing in our validation cohorts. In this way, we expected to gain understanding of molecular pathways that define chemoresponses, and to provide potential predictive markers and chemosensitising therapeutic targets.

Materials and methods

Ethics and clinical samples

Ethical approval was obtained from Leeds (East) REC (#06/Q1206/180). This study used formalin-fixed paraffin-embedded (FFPE) tissue held by Leeds Teaching Hospitals NHS Trust (Leeds, UK). Samples were from 3 cohorts of female patients diagnosed with primary breast cancer within the trust. 1) 6 patients treated with NAC, after diagnoses post-2013. Cases were selected based on: partial responses to epirubicin/cyclophosphamide (defined by longitudinal imaging); ER-positive tumours (defined by clinical ER assessment); presence of sufficient cancer cells in diagnostic biopsies and resections (assessed using hematoxylin and eosin stained sections). Samples were available for matched pre-NAC biopsies, post-NAC resections, and normal tissue from resections >1cm outside tumour margins. 2) 53 NAC-treated patients, diagnosed 1/1/2005 to 30/4/2013; a subset of a published cohort (15). Exclusion criteria were patients who did not undergo surgery, had metastatic disease, lacked imaging follow-up, or had complete pathological responses. Duplicate 0.6mm cores representing post-NAC tissue were available in TissueMicroarrays (TMAs). 3) 303 patients treated with adjuvant chemotherapy, diagnosed 1/2006 to 12/2010. Exclusion criteria were patients who received any neoadjuvant therapy, had metastatic disease, or had second breast cancers subsequent to initial diagnosis/treatment prior to 2006. TMAs containing treatment-naïve resection tissue (three independent 0.6mm cores per case) were assembled

using manual tissue microarrayers (Beecher Instruments, WI, USA), after representative tissue areas (central and peripheral when possible) were identified microscopically on whole tissue blocks/sections by consultant histopathologists (ETV, AMH).

Laser capture microdissection (LCM) and DNA extraction

LCM to purify cancer cells was performed broadly as described previously (16); further details are included in Supplementary Methods. For normal tissue blocks, cellular tissue was identified macroscopically and was manually macro-dissected. QIAamp MinElute Columns DNA FFPE Tissue or AllPrep DNA/RNA FFPE Kits (Qiagen; Dusseldorf, Germany) were used to extract DNA, which was quantified using Quant-iT PicoGreen dsDNA kits (ThermoFisher; MA, USA) and quality assessed using the Agilent 2200 TapeStation (Agilent Technologies; CA, USA) (all following the manufactures' protocols).

Library preparation, sequencing and data analysis

Whole exome sequencing libraries were prepared from 0.2-1.2µg DNA using NEBNext Ultra DNA Library Prep Kits (New England Biolabs; MA, USA) and SureSelectXT reagents (Agilent Technologies; CA, USA) (see Supplementary Methods for protocols). Indexed libraries were pooled and sequenced on the HiSeq 3000 (Illumina; CA, USA), with pair end reads (2x150bp). Sequence data have been deposited at the European Genome-phenome Archive, under accession number EGAS00001003626 (https://ega-archive.org). Exome data were analysed to identify somatic single nucleotide/small indel variants by Edinburgh Genomics Laboratory (Edinburgh, UK), as described in Supplementary Methods.

Cell culture, transfection, and chemoresponse assays

MCF7 cells were purchased (ATCC; Manassas, USA) and cultured in DMEM, 10% FCS (ThermoFisher; MA, USA), 95% air/5% CO_2 at 37°C. Cell line identity was confirmed (STR profiles, Leeds Genomics Service, UK) and cultures were consistently mycoplasma negative (MycoAlert, Lonza; Basal, Switzerland). ON-TARGETplus siRNA-SMARTpools (Dharmacon; CO, USA) were used to perform targeted/control knock-downs (sequences in Supplementary Methods). Cells were seeded at 10,000 (96-well plate) or 500,000 cells/well (6-well plate) and then incubated overnight. 25nM or 50nM final siRNAs concentrations were prepared in serum-free medium (OptiMEM, ThermoFisher; MA, USA) and transfection complexes prepared using DharmaFECT formula 1 (Dharmacon; CO, USA) following the manufacturer's protocols. Serum-free complexes were added to cells with 4 volumes of complete fresh medium and 24h later medium was removed and replaced with further fresh medium. Epirubicin hydrochloride (Sigma Aldrich, MO, USA) was prepared as a 10mM stock solution in water, and was diluted to working concentrations in medium. MTT (3-(4,5-

Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) assays were performed as described previously (17). For colony forming assays, cells were transfected and epirubicin-treated (24h), before cells were harvested in trypsin-EDTA (ThermoFisher; MA, USA) and replated in fresh (epirubicin-free) medium at 500 or 1000 cells (depending on expected survival to ensure colony numbers were countable) per 10cm culture dish (Corning; MA, USA). Cells were cultured undisturbed for 14 days. Medium was removed and cells were fixed in 3ml acetic acid/methanol 1:7 (vol/vol) for 5min. Colonies were stained with 0.5% crystal violet (Sigma Aldrich, MO, USA) for 1min. Colonies were counted macroscopically, regarding >50 cells as a colony. To validate colony counting reproducibility, 10 plates with different colony numbers were scored independently by authors WSAA and LMA; concordance between scorers was very high (Spearman's rank $r=0.948$ $p<0.0001$). Intracellular epirubicin uptake assays were performed using flow cytometry on the Attune Acoustic focusing cytometer (ThermoFisher; MA, USA) with fluorescent detection in BL-3 as described previously (18).

Quantitative (q)PCR

RNA was extracted using ReliaPrep RNA cell Minipreps (Promega; WI, USA) and quantified using the NanoDrop 2000/2000c Spectrophotometer (ThermoFisher; MA, USA), according to manufacturers' protocols. High-capacity cDNA RT kits (ThermoFisher; MA, USA) were used for reverse transcription following the manufacturer's protocols. qPCR was performed using Taqman protocols/assays: MUC17 (#Hs00959753s1), PCNX1 (#Hs00900449m1), ABCB1 (#Hs00184500m1), ABCC1 (#Hs01561483m1), ABCG2 (#Hs01053790m1), RPL19 (#Hs02338565), ACTB (#Hs99999903m1) (ThermoFisher; MA, USA). Reactions were prepared in technical duplicates/triplicates and assayed in a QuantStudio 5 instrument (Applied Biosystems, ThermoFisher; MA, USA) using standard mode and cycling conditions. QuantStudio Design and Analysis Software (ThermoFisher; MA, USA) was used to calculate CTs and average of replicates was taken for each sample. The $2^{-\Delta\Delta CT}$ method was performed to calculate fold differences in gene expression, using RPL19 and ACTB as normalisers (19).

ImmunoHistoChemistry (IHC)

All steps were at room temperature unless otherwise stated. TMA blocks were sectioned at 3 or 5µm onto Superfrost Plus slides (ThermoFisher; MA, USA). Slides were dewaxed through xylene, brought through absolute ethanol, and then into running water. Antigen retrieval was performed using 10mM citric acid buffer, pH 6.0 and microwave heating (10min, high power). Endogenous peroxidase was blocked with 10min in 0.3% hydrogen peroxide. Slides were washed in running water and rinsed with Tris-Buffered Saline (TBS). Slides were transferred to humidified chambers and 100µl of antibody diluent reagent

solution (ThermoFisher; MA, USA) was added for 5min. After that, 1:250 anti-MUC17 rabbit polyclonal antibody (#ab122184, Abcam; Cambridge, UK) or 1:500 anti-PCNX1 rabbit polyclonal antibody (#ab220503, Abcam; Cambridge, UK) was added and incubated overnight (4°C) or 2h respectively. Negative controls were antibody diluent solution alone. Slides were washed with TBS-T (0.1% Tween20, Sigma Aldrich; MO, USA) twice and TBS twice. Slides were treated with 100µl SignalStain Boost IHC detection reagent (HRP, Rabbit) (Cell Signaling Technology; MA, USA) for 30min, and were washed with TBS-T and TBS as previously. 100µl SignalStain DAB substrate working solution (Cell Signaling Technology; MA, USA) was added for 5min and slides were washed in running water. Slides were counterstained with Mayer's Haematoxylin, washed in running water, incubated in Scott's water (1min), and washed again in running water. Finally, slides were dehydrated through absolute ethanol and xylene and mounted under coverslips in DPX (Sigma Aldrich; MO, USA). Slides were digitally scanned (Aperio; CA, USA), and were scored on-line. Scoring protocols were devised in consultation with AMH (consultant breast histopathologist) to record the variation seen across the cohorts. Cytoplasmic MUC17 was scored for intensity only (it was expressed similarly in all tumour cells of each case, therefore positive cell proportion was not informative): 0, no staining; 1, weak staining; 2, moderate staining; 3, strong staining. PCNX1 was scored based on intensity of nuclear staining (scores between 0-3 as above) and proportion of tumour cells showing positive nuclear staining (0, 0%; 1, 1-5%; 2, 6-25; 3, 26-75%, 4, >75%), with final scores being the sum (0-7). All cores were scored by author WSAA, and to validate scoring reproducibility, 10% of cores were scored additionally and independently by AMH; concordance between scorers was assessed as near perfect or substantial (Kappa scores of 0.8 for MUC17 and 0.7 for PCNX) (20).

Statistical analyses

Statistical analyses were performed as described in figure legends, using GraphPad Prism (v7) for Mann-Whitney tests, paired Student's T tests, or 2-way ANOVAs. ROC curves, Spearman's rank correlation analyses, and Kaplan-Meier survival analyses were performed using SSPS (v25).

Results

NAC is associated with an overall reduction in genetic diversity in breast tumours

Our first aim was to perform whole exome sequencing on tumour cells pre- and post-NAC from a small group of primary breast cancers. We focused on ER-positive cancers – a strategy we found to be essential since typically this cancer type is relatively resistant to

NAC as compared to other breast cancer subtypes (21) and therefore numbers of tumour cells after treatment would be increased and post-treatment genomic analysis would be possible. We also minimised variation within our cohort in clinico-pathological features, therapeutic regimen, and treatment response in an effort to increase the likelihood of identifying resistance or sensitivity pathways in common between individuals. Consequently, 6 patients were selected with tumours that had partial therapy responses as defined by clinical MRI monitoring during therapy to an epirubicin/cyclophosphamide regimen, and were ER-positive, her2-negative, and mainly ductal carcinomas of no special type (Table S1). Tissue was available pre- and post-NAC and, importantly, cancer cells were isolated by LCM (Fig S1) to allow extraction of tumour DNA with minimal contamination by normal stromal DNA, thereby maximising our ability to detect somatic variants in relatively rare clones. Matched normal DNA was also extracted from normal tissues for each individual. DNA was subjected to whole exome sequencing.

Quality control analyses of exome data are shown in Table S2, demonstrating some variation in quality as expected particularly in the context that some samples were prepared by LCM from small, FFPE biopsies therefore input DNA was limiting. Somatic variants in cancer samples were identified and their numbers, along with overlaps between matched pre- and post-NAC samples, are shown in Venn diagrams (Fig 1), while the full list of variants is available in Table S3. Our analysis provided sensitivity to detect mutant variants represented in very rare clones, for example the rarest variant detected had a mutant allele frequency of only 0.012. Mutations in PIKC3 and TP53, the genes most commonly mutated in breast cancer, were detected in 1 and 2 cases respectively, which was in line with expectations within a small cohort (reportedly present in 42% and 15% of ER-positive cases overall (22)). Numbers of variants showed substantial variation between samples (45-1434) as previously reported in breast cancer, which is regarded as genetically heterogeneous (23); nevertheless, some consistent observations can be made. There were large and significant differences in mutational load between pre- and post-NAC, with pre-NAC samples containing a mean of 434 variants, while post-NAC this was reduced to 148 (paired T test, $p=0.03$). Also, overlaps between somatic variants in matched pre- and post-NAC samples was relatively small, with a mean number of variants shared between the matched samples of 34 (range 3-141), representing only 13.1% of the totals from pre-NAC samples and 19.1% from post-NAC. Both these observations are in line with those in related previous studies (24,25) We concluded that NAC causes substantial changes in the clonal composition of breast cancers, with an overall reduction in somatic variant diversity.

MUC17 or PCNX1 variants are changed in representation post-NAC in multiple patients

We selected variants in genes MUC17 and PCNX1 for further analysis, on the basis of identification in multiple patients, consistency in representation in either pre- or post-NAC samples (ie consistency in potential selection by NAC), and predictions of functional consequences of variants (focusing on coding changes with at least one example of a strong prediction of a damaging phenotype). MUC17 was found to have 3 different somatic missense single nucleotide variants in 3 separate patients (50% of the cohort). All these variants were identified in pre-NAC samples, while MUC17 somatic variants were not present in the matched (or indeed any) post-NAC samples. This distribution was compatible with the hypothesis that these MUC17 variants are associated with relative chemotherapy sensitivity, since post-NAC they had apparently been eliminated. PCNX1 was found to have 2 different somatic deletions leading to frame-shifts in 2 separate patients (33% of the cohort); both deletions were identified in post-NAC samples but were not present in any pre-NAC samples, a distribution compatible with the hypothesis that the deletions are associated with relative chemotherapy resistance. Further details of the variants, the sequencing evidence for them, and their predicted consequences are shown in Table 1. MUC17 variants cause missense changes to the encoded protein within its extra-cellular domain that were defined as being of 'moderate' predicted functional impact. Variant Thr3809Met is of particular note, since this is immediately adjacent to the transmembrane EGF-like domain of the protein. This EGF-like domain is a region of likely functional importance, since it is believed to direct EGFR mediated oncogenic signalling (26), supporting the potential negative functional impact of the variant. PCNX1 variants cause frame-shifts at residues 564 or 623 (of 2341 residues), which are predicted to cause premature termination, and may induce nonsense-mediated decay of transcripts (27). In both cases, the likely predicted effect is a loss of gene function.

***MUC17 and PCNX1 mediate response to chemotherapy in MCF7 cells* in vitro**

Next, we aimed to use a cell line model to screen MUC17 and PCNX1 for roles in defining chemoresponse, in order to support further potential analyses in clinical cohorts; it is worth emphasising that the intention here was to provide a further *in vitro* screening step only, whereas subsequent separate analyses would provide the essential, robust, clinical validation of any impact of candidate proteins on chemotherapy response to meet our key aim of defining therapy predictive markers. Our strategy was to mimic the loss of function mutations using siRNA-mediated knock-down, and to assess impact on chemosensitivity *in vitro*; therefore, we needed a cell line that was representative of the same breast cancer subtype as the cancers used for sequencing (ER-positive/her2-negative), and we needed to

ensure that MUC17 and PCNX1 were wild-type and expressed (therefore potentially functional). MCF7 cells were identified as suitable from their receptor expression status (28), and MUC17 and PCNX1 expression/sequence data available from the Cancer Cell Line Encyclopaedia (https://portals.broadinstitute.org/ccle).

First, we assessed whether siRNA could successfully manipulate MUC17 or PCNX1 expression. MCF7 cells were transfected with siRNAs targeted against MUC17 or PCNX1, or with non-targeted control siRNAs, and qPCR was used to quantify relative expression post-transfection (Fig 2A). Both gene products were successfully and significantly targeted for up to 96h; knock-down efficiency was up to 99% (PCNX1 48h). Knock-down was also confirmed at the protein level (Fig S2). Then, we assessed sensitivities of cells after these transfections to the chemotherapeutic epirubicin, which was the key component of the regimen used in the clinical cases that were sequenced above. MCF7 cells were transfected with MUC17 or PCNX1 targeted siRNA, or with non-targeted control as before, and then after 24h, cells were treated with a range of epirubicin doses. MTT assays were performed after 24, 48 and 72h of drug treatment (Fig 2B); note that these times represent 48, 72 and 96h post-transfection, time-points at which suitable knock-down is maintained (Fig 2A). As expected, cells showed reduced survival after epirubicin treatment that was dose- and time-dependent. No significant differences were noted when comparing sensitivities of control and targeted siRNA cells at 24 or 48h for either gene, or at 72h for PCNX1. However, at 72h the MUC17 siRNA treated cells demonstrated significantly increased sensitivity to epirubicin (two-way ANOVA, $p=0.0018$). In accordance with this, a trend for increased sensitivity after MUC17 knock-down was also seen at 24h and 48h. Next, we assessed epirubicin-sensitivity after these transfections using clonogenic survival assays, in which cells were treated with drug and then cultured in drug-free medium for 2 weeks to allow assessment of proportions of cells that retained long-term proliferative potential. This assay is sensitive to terminal cellular damage that does not cause *immediate* cell death, and is consequently more reflective of some aspects of clinical cancer treatments. Cells were transfected and treated with epirubicin, and their potential to grow into viable colonies was assessed (Fig 2C). Targeting MUC17 with siRNA significantly reduced survival after epirubicin treatment, whereas, targeting PCNX1 significantly increased survival (Mann Whitney tests for individual doses, $p<0.05$, and two-way ANOVA to assess overall chemo-response, $p<0.0001$). We concluded that reduced MUC17 expression caused relative epirubicin sensitisation, a conclusion supported the elimination of cells with presumed loss of function MUC17 variants by epirubicin-based therapy in the initial patient group. By contrast, we concluded that reduced PCNX1 expression caused relative epirubicin resistance, a conclusion supported by the survival of cells with loss of function PCNX1 variants in patients after therapy.

We also investigated potential mechanisms by which MUC17 or PCNX1 impact on chemosensitivity. We noted that cells transfected with siRNA targeted against MUC17 expressed significantly decreased levels of xenobiotic pumps ABCB1 (encoding P-glycoprotein) and ABCC1 (encoding MRP-1), while those transfected with siRNA targeted against PCNX1 expressed significantly increased levels of ABCG2 (encoding BCRP) (Fig S3A). These pumps have been implicated as mediators of chemoresistance by actively exporting many chemotherapeutics (29). Accordingly, transfection with siRNA against MUC17 was associated with increased cellular loading of epirubicin while transfection with siRNA against PCNX1 was associated with reduced loading (Fig S3B).

MUC17 and PCNX1 expression predicts survival after chemotherapy

Having determined that levels of MUC17 and PCNX1 potentially impact on epirubicin efficacy *in vitro* and in patients, we were interested to assess whether expression levels of these proteins could predict chemotherapy response in larger cohorts. Our hypothesis was not that somatic mutations impact on overall expression frequently, since our own data demonstrated that the mutations when present were in minority clones of the tumours (Table 1); rather, we wished to test whether the genes we had implicated in chemoresistance through selection acting on somatic mutations had wider chemoresistance roles through differential expression. Therefore, we assembled two further independent sets of breast cancer samples from patients who had been treated with chemotherapy in either the neoadjuvant (n=53) or the adjuvant setting (n=303). Clinico-pathological data concerning patients and their tumours are shown in Table 2. It is worth noting that both cohorts included a variety of breast cancer subtypes that were treated with various chemotherapy regimens, unlike the work above that focused only on ER-positive disease and epirubicin; this is because we were keen to examine whether any effects we identified were limited to specific subgroups or were generally applicable to all breast cancers. Breast cancer samples were collected into Tissue MicroArrays (TMAs), comprising up to 3 separate cores from each tumour. MUC17 or PCNX1 expression was assessed using immunohistochemistry and was scored semi-quantitatively on scales of 0-3 for MUC17 or 0-7 for PCNX1 (Fig 3A). Mean scores from multiple cores were calculated to give scores for individual cases (Fig 3B). Scores did not show significant correlations with the standard breast cancer prognostic markers, estrogen receptor status (positive/negative), tumour grade (1, 2, 3), lymph nodes metastasis (positive/negative) or molecular subtype (triple negative vs others) (Table S4).

Kaplan-Meier survival analyses were performed to determine whether expression of MUC17 or PCNX1 was significantly related to either disease free survival (DFS) or disease specific

survival (DSS). Cut-offs were applied to dichotomise patients into two groups based on low or high expression of each marker. These cut-offs were defined objectively using receiver operator curve analyses (30) to give the best balance between sensitivity and specificity for prediction of clinical outcome (see Table S5). For the neoadjuvant cohort (Fig 4A), MUC17 showed a significant relationship with DFS, in which patients with low MUC17 expression survived considerably longer than those with high levels (by a mean of 823 days; log rank $p<0.02$). This relationship was not significant for DSS, and PCNX1 did not show significant relationships with survival.

In the adjuvant cohort (Fig 4B), which is substantially larger therefore statistically more powerful, further significant findings were made. MUC17 again showed a significant relationship with DFS, with low MUC17 expression defining improved survival (by 208 days; $p<0.03$), and this was also reflected in significantly improved disease specific survival (by 159 days; $p<0.04$). PCNX1 expression was again not significantly related to survival, however the plots showed separation between low and high expression groups that was in accordance with expectations from our data concerning selection of somatic variants by NAC and *in vitro* sensitivity, specifically that low expression could be associated with relative resistance and therefore poor survival. These trends encouraged us to investigate the markers *in combination* in this adjuvant cohort; therefore, we examined survival in groups with high MUC17 *and* low PCNX1 (both potentially indicative of worse survival), or low MUC17 *and* high PCNX1 (both potentially indicative of improved survival). Significant differences in both DFS and DSS were revealed in the directions expected when either of these groups was compared to the remainder of the cohort (Fig 4C; $p=0.008$ to $p=0.022$), showing greater significance and larger differences in mean survival between the two groups than with either marker alone. Patients with high MUC17 *and* low PCNX1 had shorter DFS and DSS by 353 and 365 days respectively, while those with low MUC17 *and* high PCNX1 had longer DFS and DSS by 297 and 544 days. We also performed subgroup analyses, testing associations in groups with different receptor expressions, focusing on the ER-positive, ER-negative, her2-positive or triple negative groups, or chemotherapy regimens, focusing on treatment with anthracyclines with or without taxanes. This was particularly relevant since our initial work was solely in the context of ER-positive/her2-negative disease with anthracyclines and no taxanes. Low MUC17 was a significant predictor of improved survival in both ER-positive (DFS $p=0.03$; DSS $p<0.02$) and ER-negative disease (DFS $p<0.03$), and also in her2-positive (DFS $p<0.03$) and triple negative (DFS $p<0.04$) groups (Fig S4A), suggesting that the marker has value across subtypes. PCNX1 also demonstrated significant relationships with survival in both ER-positive (DFS $p<0.05$) and ER-negative disease (DFS $p<0.05$), and in the triple negative group (DFS $p<0.05$), but not in

the her2-positives, hinting at some subgroup specificity (Fig S4B). Concerning different therapies, neither marker was a significant predictor of outcome in patients treated with anthracyclines and taxanes, while both markers significantly predicted outcome after anthracycline-based therapy that lacked taxanes (MUC17: DFS $p<0.01$, DSS $p<0.04$; PCNX: DFS $p<0.04$) (Fig S4C). Overall, we concluded that expression levels of MUC17 or PCNX1 significantly predict survival after anthracycline-based chemotherapy in a range of breast cancer subtypes.

Discussion

NAC provides opportunities to characterise cancer cells that are relatively therapy resistant (cells that remain after therapy) or relatively sensitive to therapy (cells present before but not after therapy). However, assessment of these cells presents logistical and methodological challenges (11). One key issue is obtaining the matched cancer cells pre- and post-NAC that are necessary to allow comparison. Diagnostic biopsies are taken from women with breast symptoms, at which time it is not even possible to identify those with cancer, let alone the small minority who go on to receive NAC; therefore, it is impractical to collect research material at this point. Moreover, the NAC treatment pathway does not normally include further pre-NAC tissue sampling opportunities. The result is that diagnostic biopsies are often the only practical pre-NAC tissue available, but unfortunately these are very small and FFPE-treated, which can limit down-stream analyses. By contrast, difficultly post-NAC is caused by relative lack of cancer cells within tissues rather than sizes of available tissue *per se*. Successful NAC reduces tumour size and cancer cellularity (31), therefore cancer cells can be comparatively scarce post-NAC, within an abnormal fibrotic stroma (32). This last issue explains why we felt it was necessary to perform LCM of cancer cells, and thereby avoid wasteful sequencing of normal DNA derived from stromal cells and potentially-flawed comparisons of variant prevalence between cancer cell-rich (pre-NAC) and cancer cell-poor (post-NAC) samples. Concerns regarding the representative nature of these samples remain even after LCM. Diagnostic biopsies sample only a small proportion of the total tumour volume, and it is unlikely they contain all the tumour genetic diversity, while even use of resection samples, which *could* contain all the genetic diversity, is limited by the proportion of tumour cells analysed (33). Despite all these issues, studies have been published comparing characteristics of breast tumours pre- and post-NAC at the levels of protein (13), transcript (14,25), and genome (25,34). To our knowledge, our study is the first to examine whole exome data pre- and post-NAC, while employing LCM in this context to give improvements in the accuracy of assessment of changes in mutant allele frequency.

Nevertheless, there are two highly-related previous studies. Balko et al used targeted sequencing of selected exons of 196 cancer-related genes to investigate changes in variants representation in 20 triple negative breast cancers, treated with a range of different NAC regimens (24). The authors found a minority of variants to be present in both matched samples, which mirrors our findings (Fig 1), while overall they found variants in cell-cycle regulators and PI3K/mTOR pathway genes to be enriched post-NAC. Recently, Kim et al published the first study to investigate matched pre- and post-NAC exomes in breast cancer, in which they also examined 20 triple negative breast cancers (25). The difficulty of low tumour cellularity was exemplified in this work, as some post-treatment samples were assessed by histopathology as containing as little as 10%, or even 0% tumour cells. Like our study, the key findings show that therapy drives expansion of resistant clones, or contraction of sensitive clones, in the background of an overall decrease in variants. Interestingly, in both studies, the authors tried to adjust their mutant allele frequencies for differences in tumour cellularity to allow comparison between matched samples, which represents an alternative approach to LCM, although at the cost of using 90% or more of read depth on normal genome.

In accordance with the published literature describing profound heterogeneity in somatic variants between breast cancers (22), we found no variants in common between any of our 6 tumours, despite our attempt to minimise variation in histopathological features of the group (Table S1). However, there was commonality between patients in genes hosting variants at relatively low mutant allele frequencies and in their distribution between pre- or post-NAC samples, allowing us to identify MUC17 and PCNX1 as candidate regulators of chemoresponse (Table 1). It is interesting to note that somatic variants in these genes have not been identified previously as candidate driver mutations in the wealth of cancer genome studies available in breast cancer (35); this may relate to the paired pre- and post-treatment design of our study, which provides a strategy to prioritise rare sub-clonal variants for further analysis based on a change in representation after treatment, whereas important rare sub-clonal variants can be lost in the noise in a single time-point study. Furthermore, MUC17 and PCNX1 demonstrated roles in defining chemoresponse *in vitro* (Fig 2) and in two separate cohorts of patient samples (Fig 4) that were compatible with the selection acting on the somatic variants identified in our initial cohort, allowing us to propose these genes as functional mediators of, and biomarkers for chemotherapy response. For both genes, this represents a novel finding.

MUC17 belongs to the mucin family of O-glycosylated, high molecular weight, glycoproteins that contribute to mucus at mucosal surfaces (36). Little is known about MUC17 function,

however it is relevant to note that up-regulation of several mucins, most compellingly MUC1 and MUC4, has been shown to induce chemoresistance (37), potentially by forming a physical barrier reducing tumoural drug concentrations, or by inhibiting apoptosis. In pancreatic and cervical cancer cells, MUC1 has also been linked to up-regulation of the xenobiotic transporter ABCB1 (38,39), which is associated with chemoresistance (29). The work presented here is the first to identify and test a role for MUC17 in chemoresistance, and to investigate a potential mechanism: regulation of ABCB1 and ABCC1, similarly to MUC1. Interestingly, MUC17 somatic mutations in cancer have very recently been investigated in two seemingly conflicting studies: MUC17 was defined as a recurrently mutated gene in multiple cancer types (although specifically not in breast) (40) while, by contrast, MUC17 was identified as having significantly fewer somatic mutations than expected (41), suggesting wild type function was required in 'successful' cancer cells. Our interpretation of this conflict is that the depth of tumour cell sequencing, as highlighted above, may be critical for detecting MUC17 mutations, which were in only ~18% of the tumour cells at most in our data (Table 1), and this was in the context that we had purified the cancer genomes from the normal genomes of stromal cells that can comprise a large majority in many tumour samples. Even less is known about human PCNX1 (pecanex homolog 1) and its functions. The Drosophila protein, pecanex, is a positive regulator of Notch-signalling in neurogenic tissues (42). However, expression of the mammalian homolog, PCNX1, could not be detected in neuronal tissues, and a potential role was inferred in spermatogenesis because of expression in testis (43). The link with Notch is intriguing in the context of our study, since Notch-signalling has also been implicated as a positive regulator of chemoresistance (18). However, this relationship can not explain how reduced/low PCNX1 could associate with chemoresistance, since reduced PCNX1 function would potentially inhibit Notch-signalling and thereby act against Notch-induced chemoresistance. The only manuscript describing a cancer-related function for PCNX1 shows it to be a positive regulator of the oncoprotein Skp2 in lung cancer (44). Interestingly, this function is independent of PCNX1 protein, since the transcripts act as a competitive endogenous RNAs that derepresses Skp2 expression by sequestering inhibitory microRNAs. Skp2 has itself been reported as a chemoresistance mediator (45), but again this relationship is the reverse of that which would explain our findings for PCNX1, therefore chemoresistance mechanisms directly downstream of PCNX1 remain unknown. In the case of PCNX1, where variants were identified only after chemotherapy, it is interesting to speculate about whether these variants were present pre-therapy below the detection level, or whether they were actually caused by chemotherapy itself and then selected through relative resistance. Our assessment is that it is most likely they were previously present. This is based on the abilities of the chemotherapy agents (epirubicin/cyclophosphamide) to

induce different classes of mutations. Both PCNX1 variants detected were small deletions, while cyclophosphamide and doxorubicin (an anthracycline closely related to epirubicin) have been shown not to induce somatic deletions at levels above background (46).

In conclusion, we present genomics and expression data from 3 separate cohorts to show that MUC17 and PCNX1 are potential markers for chemotherapy stratification. In addition, since mucins are well-established therapeutic targets in cancer (47), MUC17 presents an attractive case for therapeutic inhibition in breast cancer chemo-sensitisation strategies.

Acknowledgments

Funding: WSAA and DEB were funded by studentships from the Ministry of Higher Education, Sultanate of Oman and the University of Leeds respectively. LFS was supported by a University Academic Fellowship from the University of Leeds. AMS is funded by Birmingham CRUK Cancer Centre. The funders had no role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions: WSAA, LMA, DEB, AMH, SJJ, AMS and ETV all carried out experiments, secured data or carried out analyses for the manuscript. SMB, AMS, and ETV provided essential resources for the work. WSAA, AMH, LFS and TAH carried out or oversaw data analyses. AMH, ETV and TAH conceived the project. WSAA and TAH wrote the draft manuscript. All authors contributed to refining the manuscript and approved the final draft.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin **2018**;68(6):394-424 doi 10.3322/caac.21492.
2. Nur Husna SM, Tan HT, Mohamud R, Dyhl-Polk A, Wong KK. Inhibitors targeting CDK4/6, PARP and PI3K in breast cancer: a review. Ther Adv Med Oncol **2018**;10:1758835918808509 doi 10.1177/1758835918808509.
3. NCRAS, CRUK. Chemotherapy, Radiotherapy and Tumour Resections in England: 2013-2014. London: NCRAS; 2017.
4. Lebert JM, Lester R, Powell E, Seal M, McCarthy J. Advances in the systemic treatment of triple-negative breast cancer. Curr Oncol **2018**;25(Suppl 1):S142-S50 doi 10.3747/co.25.3954.
5. Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L. Treatment of HER2-positive breast cancer: current status and future perspectives. Nat Rev Clin Oncol **2011**;9(1):16-32 doi 10.1038/nrclinonc.2011.177.
6. Albain K, Anderson S, Arriagada R, Barlow W, Bergh J, Bliss J*, et al.* Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. Lancet **2012**;379(9814):432-44 doi 10.1016/S0140-6736(11)61625-5.
7. Harnan S, Tappenden P, Cooper K, Stevens J, Bessey A, Rafia R*, et al.* Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer: a systematic review and economic analysis. Health Technol Assess **2019**;23(30):1-328 doi 10.3310/hta23300.
8. Tan W, Yang M, Yang H, Zhou F, Shen W. Predicting the response to neoadjuvant therapy for early-stage breast cancer: tumor-, blood-, and imaging-related biomarkers. Cancer Manag Res **2018**;10:4333-47 doi 10.2147/CMAR.S174435.
9. Karakatsanis A, Tasoulis MK, Warnberg F, Nilsson G, MacNeill F. Meta-analysis of neoadjuvant therapy and its impact in facilitating breast conservation in operable breast cancer. Br J Surg **2018**;105(5):469-81 doi 10.1002/bjs.10807.
10. Cain H, Macpherson IR, Beresford M, Pinder SE, Pong J, Dixon JM. Neoadjuvant Therapy in Early Breast Cancer: Treatment Considerations and Common Debates in Practice. Clin Oncol (R Coll Radiol) **2017**;29(10):642-52 doi 10.1016/j.clon.2017.06.003.
11. Selli C, Sims AH. Neoadjuvant Therapy for Breast Cancer as a Model for Translational Research. Breast Cancer (Auckl) **2019**;13:1178223419829072 doi 10.1177/1178223419829072.
12. Li S, Payne S, Wang F, Claus P, Su Z, Groth J*, et al.* Nuclear basic fibroblast growth factor regulates triple-negative breast cancer chemo-resistance. Breast Cancer Res **2015**;17:91 doi 10.1186/s13058-015-0590-3.
13. Kim B, Fatayer H, Hanby AM, Horgan K, Perry SL, Valleley EM*, et al.* Neoadjuvant chemotherapy induces expression levels of breast cancer resistance protein that predict disease-free survival in breast cancer. PLoS One **2013**;8(5):e62766 doi 10.1371/journal.pone.0062766.
14. Gonzalez-Angulo AM, Iwamoto T, Liu SY, Chen HQ, Do KA, Hortobagyi GN*, et al.* Gene Expression, Molecular Class Changes, and Pathway Analysis after Neoadjuvant

Systemic Therapy for Breast Cancer. Clinical Cancer Research **2012**;18(4):1109-19 doi 10.1158/1078-0432.Ccr-11-2762.
15. Gahlaut R, Bennett A, Fatayer H, Dall BJ, Sharma N, Velikova G*, et al.* Effect of neoadjuvant chemotherapy on breast cancer phenotype, ER/PR and HER2 expression - Implications for the practising oncologist. Eur J Cancer **2016**;60:40-8 doi 10.1016/j.ejca.2016.03.006.
16. Verghese ET, Drury R, Green CA, Holliday DL, Lu X, Nash C*, et al.* MiR-26b is down-regulated in carcinoma-associated fibroblasts from ER-positive breast cancers leading to enhanced cell migration and invasion. J Pathol **2013**;231(3):388-99 doi 10.1002/path.4248.
17. Langlands FE, Dodwell D, Hanby AM, Horgan K, Millican-Slater RA, Speirs V*, et al.* PSMD9 expression predicts radiotherapy response in breast cancer. Molecular cancer **2014**;13(1):73 doi 10.1186/1476-4598-13-73.
18. Kim B, Stephen SL, Hanby AM, Horgan K, Perry SL, Richardson J*, et al.* Chemotherapy induces Notch1-dependent MRP1 up-regulation, inhibition of which sensitizes breast cancer cells to chemotherapy. Bmc Cancer **2015**;15 doi ARTN 634 10.1186/s12885-015-1625-y.
19. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **2001**;25(4):402-8 doi 10.1006/meth.2001.1262.
20. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics **1977**;33(2):363-74.
21. Gentile LF, Plitas G, Zabor EC, Stempel M, Morrow M, Barrio AV. Tumor Biology Predicts Pathologic Complete Response to Neoadjuvant Chemotherapy in Patients Presenting with Locally Advanced Breast Cancer. Annals of surgical oncology **2017**;24(13):3896-902 doi 10.1245/s10434-017-6085-y.
22. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490(7418):61-70 doi 10.1038/nature11412.
23. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW*, et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature **2012**;486(7403):353-60 doi 10.1038/nature11143.
24. Balko JM, Giltnane JM, Wang K, Schwarz LJ, Young CD, Cook RS*, et al.* Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. Cancer discovery **2014**;4(2):232-45 doi 10.1158/2159-8290.CD-13-0286.
25. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T*, et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell **2018**;173(4):879-93 e13 doi 10.1016/j.cell.2018.03.041.
26. Joshi S, Kumar S, Choudhury A, Ponnusamy MP, Batra SK. Altered Mucins (MUC) trafficking in benign and malignant conditions. Oncotarget **2014**;5(17):7272-84 doi 10.18632/oncotarget.2370.
27. Raimondeau E, Bufton JC, Schaffitzel C. New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. Biochem Soc T **2018**;46:503-12 doi 10.1042/Bst20170427.
28. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. Breast Cancer Res **2011**;13(4):215 doi 10.1186/bcr2889.

29. Thorne JL, Hanby AM, Hughes TA. The Molecular Pathology of Chemoresistance During the Therapeutic Response in Breast Cancer. In: Khan A, Ellis I, Hanby A, Cosar E, Rakha E, Kandil D, editors. Precision Molecular Pathology of Breast Cancer. Volume 10. New York, USA: Springer; 2015. p 291-307.
30. Zlobec I, Steele R, Terracciano L, Jass JR, Lugli A. Selecting immunohistochemical cut-off scores for novel biomarkers of progression and survival in colorectal cancer. J Clin Pathol **2007**;60(10):1112-6 doi 10.1136/jcp.2006.044537.
31. Rajan R, Poniecka A, Smith TL, Yang Y, Frye D, Pusztai L*, et al.* Change in tumor cellularity of breast carcinoma after neoadjuvant chemotherapy as a variable in the pathologic assessment of response. Cancer **2004**;100(7):1365-73 doi 10.1002/cncr.20134.
32. Naidoo K, Pinder SE. The importance of histological assessment after neoadjuvant therapy and the need for standardisation. Clin Radiol **2018**;73(8):693-9 doi 10.1016/j.crad.2018.01.005.
33. Clark BZ, Onisko A, Assylbekova B, Li X, Bhargava R, Dabbs DJ. Breast cancer global tumor biomarkers: a quality assurance study of intratumoral heterogeneity. Mod Pathol **2019**;32(3):354-66 doi 10.1038/s41379-018-0153-0.
34. Chung YR, Kim HJ, Kim M, Ahn S, Park SY. Clinical implications of changes in the diversity of c-MYC copy number variation after neoadjuvant chemotherapy in breast cancer. Scientific reports **2018**;8(1):16668 doi 10.1038/s41598-018-35072-5.
35. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X*, et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature **2016**;534(7605):47-54 doi 10.1038/nature17676.
36. Corfield AP. Mucins: a biologically relevant glycan barrier in mucosal protection. Biochim Biophys Acta **2015**;1850(1):236-52 doi 10.1016/j.bbagen.2014.05.003.
37. Jonckheere N, Skrypek N, Van Seuningen I. Mucins and tumor resistance to chemotherapeutic drugs. Biochim Biophys Acta **2014**;1846(1):142-51 doi 10.1016/j.bbcan.2014.04.008.
38. Jin W, Liao X, Lv Y, Pang Z, Wang Y, Li Q*, et al.* MUC1 induces acquired chemoresistance by upregulating ABCB1 in EGFR-dependent manner. Cell death & disease **2017**;8(8):e2980 doi 10.1038/cddis.2017.378.
39. Nath S, Daneshvar K, Roy LD, Grover P, Kidiyoor A, Mosley L*, et al.* MUC1 induces drug resistance in pancreatic cancer cells via upregulation of multidrug resistance genes. Oncogenesis **2013**;2:e51 doi 10.1038/oncsis.2013.16.
40. Liu B, Hu FF, Zhang Q, Hu H, Ye Z, Tang Q*, et al.* Genomic landscape and mutational impacts of recurrently mutated genes in cancers. Mol Genet Genomic Med **2018**;6(6):910-23 doi 10.1002/mgg3.458.
41. Gorlov IP, Gorlova OY, Amos CI. Untouchable genes in the human genome: Identifying ideal targets for cancer treatment. Cancer Genet **2019**;231-232:67-79 doi 10.1016/j.cancergen.2019.01.005.
42. Yamakawa T, Yamada K, Sasamura T, Nakazawa N, Kanai M, Suzuki E*, et al.* Deficient Notch signaling associated with neurogenic pecanex is compensated for by the unfolded protein response in Drosophila. Development **2012**;139(3):558-67 doi 10.1242/dev.073858.
43. Geisinger A, Alsheimer M, Baier A, Benavente R, Wettstein R. The mammalian gene pecanex 1 is differentially expressed during spermatogenesis. Bba-Gene Struct Expr **2005**;1728(1-2):34-43 doi 10.1016/j.bbaexp.2005.01.010.

44. Li JQ, Tian HH, Pan JC, Jiang N, Yang J, Zhou CW*, et al.* Pecanex functions as a competitive endogenous RNA of S-phase kinase associated protein 2 in lung cancer. Cancer Letters **2017**;406:36-46 doi 10.1016/j.canlet.2017.07.030.
45. Gong J, Zhou Y, Liu D, Huo J. F-box proteins involved in cancer-associated drug resistance. Oncol Lett **2018**;15(6):8891-900 doi 10.3892/ol.2018.8500.
46. Szikriszt B, Poti A, Pipek O, Krzystanek M, Kanu N, Molnar J*, et al.* A comprehensive survey of the mutagenic impact of common cancer cytotoxics. Genome biology **2016**;17:99 doi 10.1186/s13059-016-0963-7.
47. Dhanisha SS, Guruvayoorappan C, Drishya S, Abeesh P. Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. Crit Rev Oncol Hemat **2018**;122:98-122 doi 10.1016/j.critrevonc.2017.12.006.

Gene	Patient number	Position, Reference, Variant, Description	Variant frequency, Depth	Effect	Impact SnpEff Polyphen SIFT
MUC17	2	Chr7, 101042842 C T SNV	0.184 165	Missense p.Thr3809Met	Moderate Damaging 0.092
	4	Chr7, 101033319 T C SNV	0.052 187	Missense p.Ser635Pro	Moderate Benign 0.802
	6	Chr7, 101038440 C A SNV	0.034 165	Missense p.Arg2342Ser	Moderate Benign 1
PCNX1	4	Chr14, 70978026 CACAGG C Deletion	0.118 30	Frameshift/ premature termination p.Thr564FS	High n/a n/a
	6	Chr14, 70978204 GAT G Deletion	0.136 45	Frameshift/ premature termination p.Asp623FS	High n/a n/a

Table 1. Somatic variants in MUC17 or PCNX1 that changed representation after NAC were identified in 3 out of 6 patients. Details included are: chromosomal position in the reference genome, mutant allele frequency (relative to 1), mutant read depth, and effects of variants on encoded proteins in terms of primary protein sequence and predictions of functional impact (from SnpEff, Polyphen, and SIFT). Note: SIFT (Sorting Intolerant from Tolerant) values represent predicted functional impact on scale of 0-1, with 0 being the more damaging end of the scale. SNV, single nucleotide variant.

	Neoadjuvant n=53 (%)	Adjuvant n=303 (%)
Histopathological diagnosis:		
Ductal NST	39 (74)	222 (73)
Lobular	2 (4)	21 (7)
Metaplastic	1 (2)	7 (2)
Mixed	6 (11)	47 (16)
Others	5 (9)	6 (2)
Grade:		
1	4 (8)	17 (6)
2	25 (47)	122 (40)
3	24 (45)	164 (54)
Lymph node positive	33 (62)	112 (37)
ER positive	16 (30)	206 (68)
her2 positive	38 (72)	66 (22)
Chemotherapy: anthracycline-based		
- without taxanes	17 (32.1)	149 (49)
- with taxanes	27 (50.9)	116 (38)
- with others	9 (17)	38 (13)

Table 2. Clinico-pathological details of the neoadjuvant and adjuvant cohorts.

Note: ER and her2 status was unknown in 5 neoadjuvant cases and her2 status was unknown in 2 adjuvant cases. NST, no special type; pos, positive; neg, negative

Figure legends

Figure 1. Numbers and distribution of somatic variants identified in breast cancers pre- and post-NAC. Venn diagrams illustrate the numbers of somatic variants identified in the cancer cells of 6 numbered breast tumours. Variants were identified pre-NAC (left) and post-NAC (right), with those shared between these time-points within the intersect.

Figure 2. MUC17 and PCNX1 regulate chemoresponse *in vitro*. MCF7 cells were transfected with siRNA targeted against MUC17 (left) or PCNX1 (right), or with non-targeting siRNA control. A) MUC17 (left) or PCNX1 (right) expression was assessed by qPCR 48 to 96h post-transfection. Error bars represent SEM of technical replicates. B) Cells were treated with doses of epirubicin 24h after transfection, and relative cell survival was determined a further 24, 48, or 72h later using MTT assays. Error bars represent SEM of 3 fully-independent experiments. Overall significance between targeted siRNA vs control was analysed using 2-way ANOVA tests. C) Cells were treated with doses of epirubicin 24h after transfection, before relative cell survival was determined using clonogenic survival assays. Error bars represent SEM of 3 fully-independent experiments. *indicates significant differences at specific doses ($p<0.05$ Mann Whitney test). Overall significance between targeted siRNA vs control was analysed using 2-way ANOVA tests.

Figure 3. MUC17 and PCNX1 are variably expressed across different breast cancers. Expression levels of MUC17 or PCNX1 were determined by immunohistochemistry in the cancer cells of breast cancer resections from cohorts of patients treated with neoadjuvant (n=53) or adjuvant (n=303) chemotherapy. A) Representative images of staining are shown for MUC17 (left) and PCNX1 (right). Individual tumour cores were scored to quantify expression on a scale of 0-3 (MUC17) or 0-7 (PCNX1). For MUC17, examples of moderate (scored as 2) and strong (scored as 3) staining are shown. For PCNX1, examples of moderate staining (scored as 2) in 25-75% of tumour cells (scored as 3; total 5), and of strong staining (3) in >75% of tumour cells (4; total 7) are shown. B) Histograms showing score distributions across the neoadjuvant and adjuvant cohorts. Case scores were determined from means of all cores scored for that case, and were rounded to the closest whole integer in these histograms.

Figure 4. Expression of MUC17 and PCNX1 predict survival after chemotherapy. Expression levels of MUC17 or PCNX1 were determined by immunohistochemistry in the cancer cells of breast cancer resections from cohorts of patients treated with neoadjuvant (A; n=53) or adjuvant (B and C; n=303) chemotherapy. Expression levels were dichotomised

objectively into low or high groups using ROC analyses. Kaplan-Meier analyses were performed comparing disease free survival (DFS) or disease specific survival (DSS) in two groups. A and B) Analyses compare groups with low or high expression of MUC17 (left) or of PCNX1 (right). C) Analyses compare the group with high MUC17 expression *and* low PCNX1 expression (high M/low P) to the remainder of the cohort ("rest") (left), or the group with low MUC17 expression *and* high PCNX1 expression (low M/high P) to the rest (right). Significance values by log rank tests are shown. ns, not significant.

Figures and tables

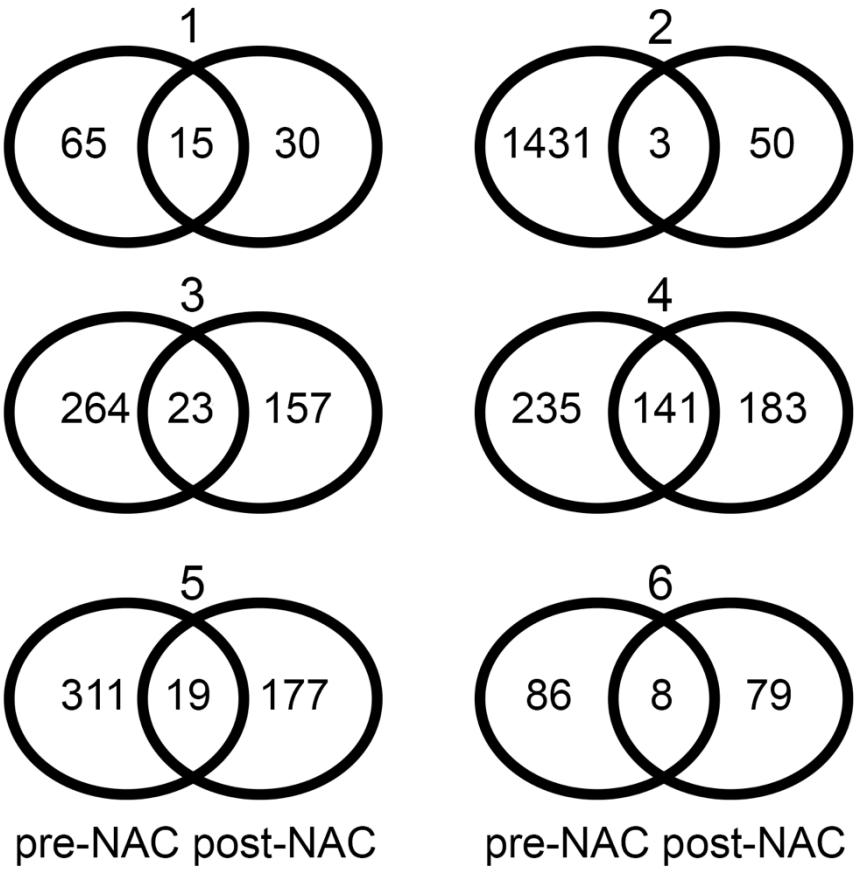

Figure 1.

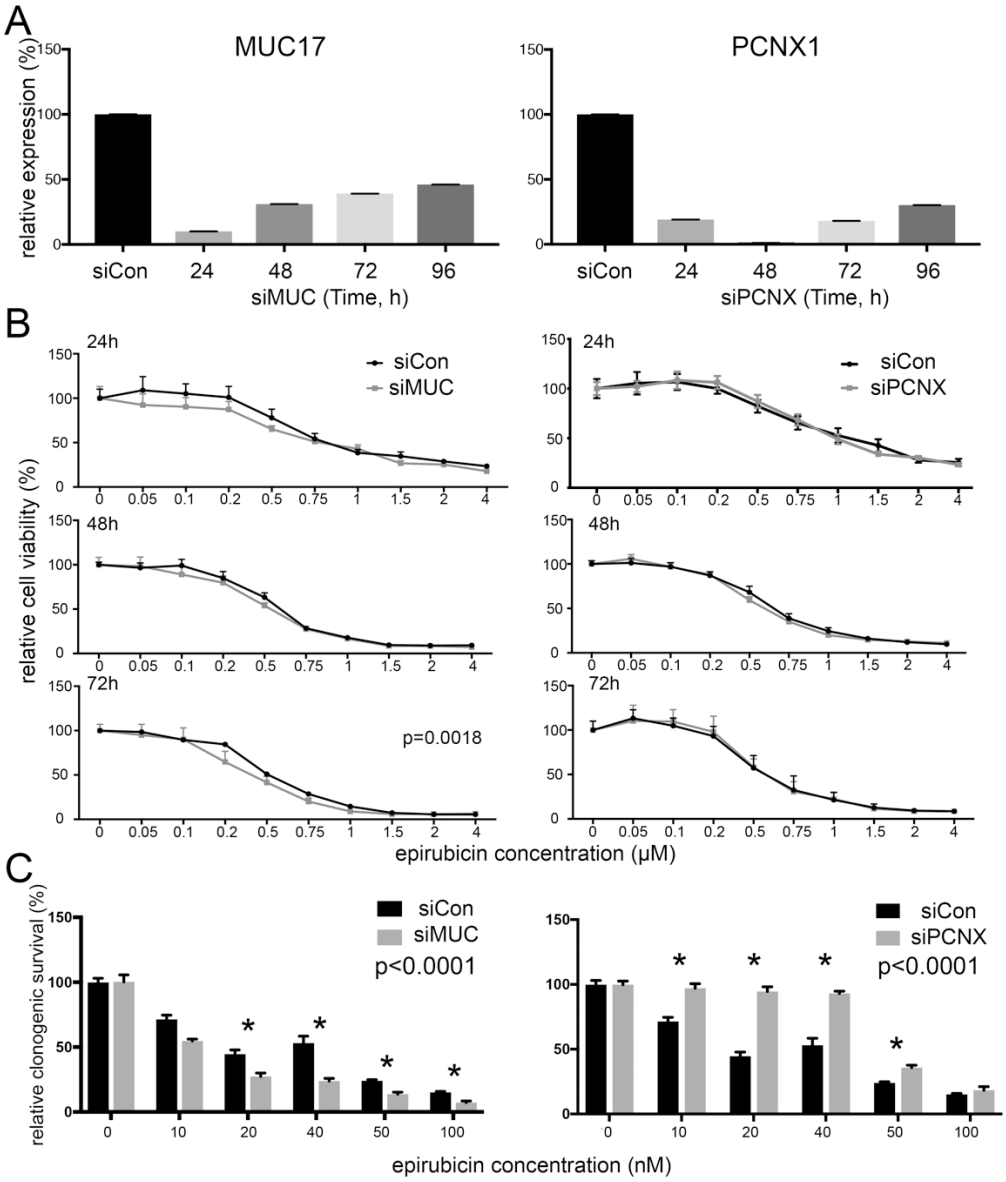
A
MUC17
PCNX1
relative expression (%)
siCon
24
48
72
96
siMUC (Time, h)
siPCNX (Time, h)
B
24h
48h
72h
siCon
siMUC
siPCNX
relative cell viability (%)
p=0.0018
epirubicin concentration (μM)
C
siCon
siMUC
siPCNX
p<0.0001
relative clonogenic survival (%)
epirubicin concentration (nM)

Figure 2.

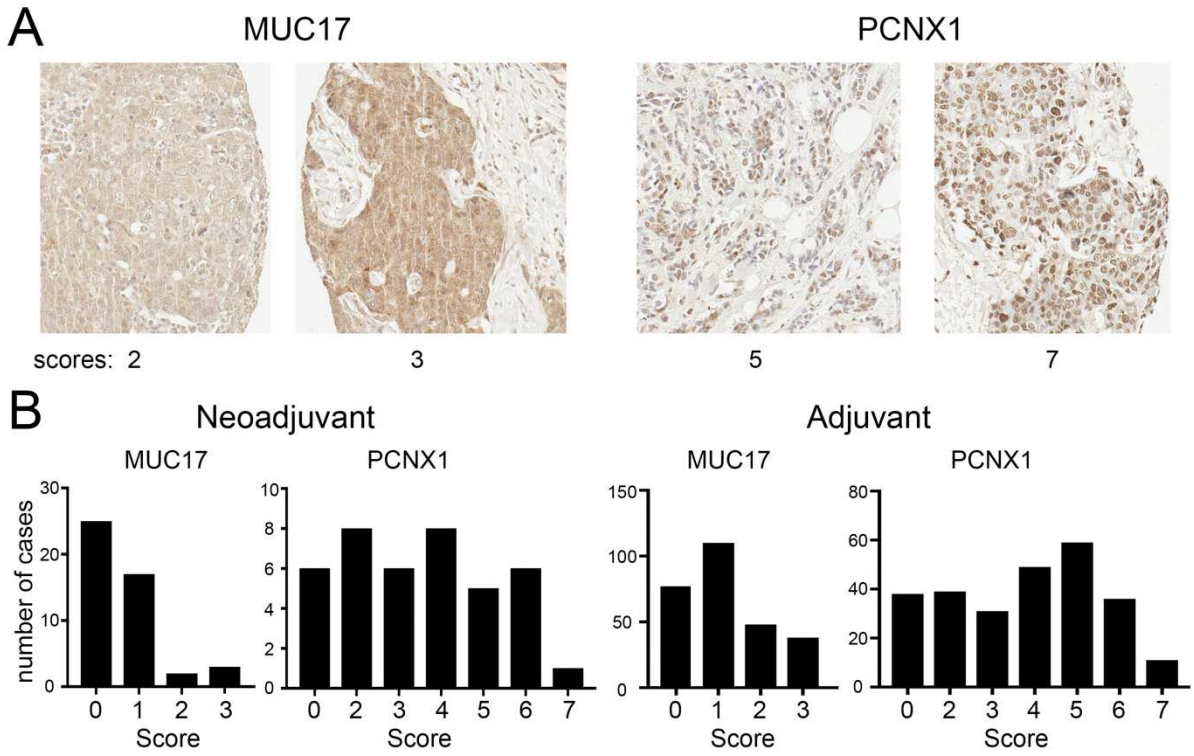
A
MUC17
PCNX1
scores: 2
3
5
7
B
Neoadjuvant
Adjuvant
MUC17
PCNX1
MUC17
PCNX1
number of cases
Score
Score
Score
Score

Figure 3.

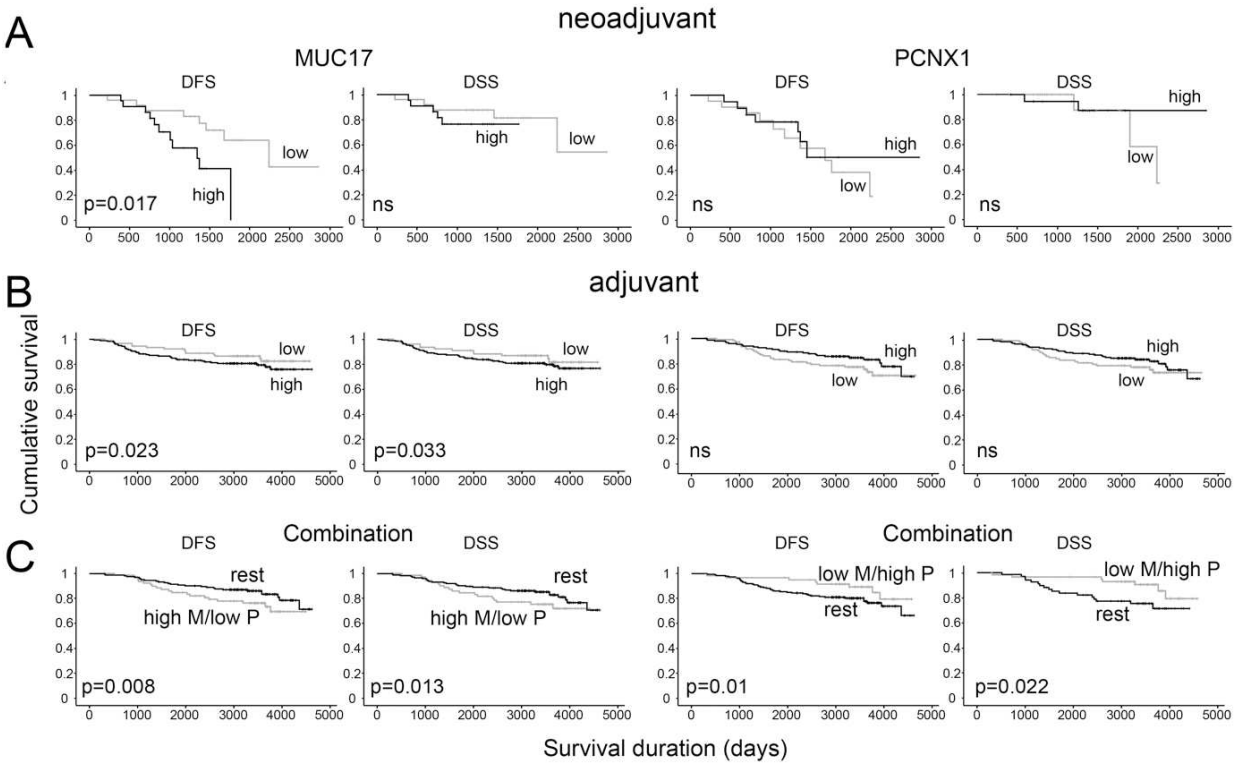
A
neoadjuvant
MUC17
PCNX1
DFS
DSS
low
high
p=0.017
ns
B
adjuvant
Cumulative survival
p=0.023
p=0.033
C
Combination
rest
high M/low P
low M/high P
p=0.008
p=0.013
p=0.01
p=0.022
Survival duration (days)

Figure 4.