N-glycolylneuraminic acid serum biomarker levels are elevated in breast cancer patients at all stages of disease.

Lucy K. Shewell[1#], Christopher J. Day[1#], Jamie R. Kutasovic[2], Jodie L. Abrahams[1^], Jing Wang[1], Jessica Poole[1], Colleen Niland[2], Kaltin Ferguson[2], Jodi M. Saunus[2], Sunil R. Lakhani[2, 3], Mark von Itzstein[1], James C. Paton[4], Adrienne W. Paton[4], and Michael P Jennings[1*].

[1]Institute for Glycomics, Griffith University, Gold Coast, Australia

[2]UQ Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, QLD Australia

[3]Pathology Queensland, The Royal Brisbane and Women's Hospital, Herston, QLD Australia

[4]Research Centre for Infectious Diseases, Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia

^Current address: Glycosciences Laboratory, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

[#]These authors contributed equally

*To whom correspondence may be addressed:
Prof Michael P. Jennings, Institute for Glycomics, Griffith University, Gold Coast campus, QLD 4222, Australia, +61 7 5552 7050, Australia, +61 7 5552 7050, m.jennings@griffith.edu.au

Abstract

Background

Normal human tissues do not express glycans terminating with the sialic acid *N*-glycolylneuraminic acid (Neu5Gc), yet Neu5Gc-containing glycans have been consistently found in human tumor tissues, cells and secretions and have been proposed as a cancer biomarker. We engineered a Neu5Gc-specific lectin called SubB2M, and previously reported elevated Neu5Gc biomarkers in serum from ovarian cancer patients using a Surface Plasmon Resonance (SPR)-based assay. Here we report an optimized SubB2M SPR-based assay and use this new assay to analyse sera from breast cancer patients for Neu5Gc levels.

Methods

To enhance specificity of our SPR-based assay, we included a non-sialic acid binding version of SubB, $SubB_{A12}$, to control for any non-specific binding to SubB2M, which improved discrimination of cancer-free controls from early-stage ovarian cancer. We analysed 96 serum samples from breast cancer patients at all stages of disease compared to 22 cancer-free controls using our optimized SubB2M-$_{A12}$-SPR assay. We also analysed a collection of serum samples collected at 6 monthly intervals from breast cancer patients at high risk for disease recurrence or spread.

Results

Analysis of sera from breast cancer cases revealed significantly elevated levels of Neu5Gc biomarkers at all stages of breast cancer. We show that Neu5Gc serum biomarker levels can discriminate breast cancer patients from cancer-free individuals with 98.96% sensitivity and 100% specificity. Analysis

of serum collected prospectively, post-diagnosis, from breast cancer patients at high risk for disease recurrence showed a trend for a decrease in Neu5Gc levels immediately following treatment for those in remission.

Conclusions

Neu5Gc serum biomarkers are a promising new tool for early detection and disease monitoring for breast cancer that may complement current imaging- and biopsy-based approaches.

Keywords: *N*-glycolylneuraminic acid, Neu5Gc, biomarker, breast cancer, diagnostic

73 **Background**

74 Aberrant glycosylation is one of the hallmarks of cancer cells. Normal
75 human tissues do not express glycans terminating with the sialic acid *N*-
76 glycolylneuraminic acid (Neu5Gc) as humans express an inactive cytidine
77 monophosphate *N*-acetylneuraminic acid (Neu5Ac) hydroxylase (CMAH)
78 enzyme (1, 2); the only enzyme known to convert Neu5Ac to Neu5Gc.
79 Nevertheless, Neu5Gc-containing glycans have been consistently found in
80 human tumor tissues, cells and secretions (3-10), and have been proposed as
81 a tumor biomarker (6, 11, 12).

82 Little progress has been made towards the development of a Neu5Gc
83 biomarker-based assay for cancer detection and patient monitoring due to the
84 lack of sufficiently sensitive and specific tools to detect this potential glyco-
85 marker in a clinically relevant biological sample. We have improved upon the
86 current methods for the detection of Neu5Gc (13-15) by developing a lectin
87 with enhanced sensitivity and specificity for this glycan in the context of
88 complex biological samples. This new lectin is derived from the B-subunit of
89 the Shiga toxigenic *Escherichia coli* (STEC) Subtilase cytotoxin (SubAB),
90 which recognizes α2-3 linked Neu5Gc (16). We used structure aided design to
91 engineer this lectin to ablate the recognition of Neu5Ac and to expand the
92 recognition from only α2-3-linked Neu5Gc to include both α2-3 and α2-6
93 Neu5Gc linkages to substituent sugars (17). This improved Neu5Gc-specific
94 lectin was called SubB2M (17, 18). In a SPR-based SubB2M assay, we
95 previously reported that the serum of ovarian cancer patients at all stages of
96 disease has elevated levels of Neu5Gc-containing biomarkers compared to
97 cancer-free females (19). This demonstrated the potential utility of Neu5Gc-

containing biomarkers in the early detection of ovarian cancer, for which there is currently no universally applicable blood-based biomarker. The best currently available biomarker for ovarian cancer is the human cancer antigen 125 (CA125), also known as MUC16, a heavily glycosylated mucin (20). Serum CA125 levels are elevated in approximately 80% of ovarian cancer cases at the time of diagnosis (21). However, CA125 serum levels may also be elevated in non-malignant conditions including pregnancy, endometriosis, ovarian cysts, pelvic inflammatory disease and in the follicular phase of the menstrual cycle (22). As a screening tool for ovarian cancer, longitudinal measurement of CA125 levels has been shown to improve sensitivity and specificity for early detection. However, outcomes from the largest ovarian cancer screening trial to date, the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) show that screening including CA125 did not significantly reduce mortality (23, 24). Hence CA125 is currently the only approved ovarian cancer serum biomarker, limited to monitoring response to therapy and disease recurrence in post-menopausal women (23).

Breast cancer is the most frequently diagnosed cancer among women worldwide and is the leading cause of cancer death in over 100 countries (25). Detection of breast cancer at the early stages is associated with better patient outcomes including lower morbidity and lower mortality rates (26). Mammography is currently the main screening tool for the early detection of breast cancer; however, this method has limitations. For example, the sensitivity of mammography in women with dense breasts is reduced from 85 % to 47-64 % (27), yet these women have an increased risk of developing breast cancer (28). Cancer antigen 15-3 (CA 15-3) is the most widely used

123 serum biomarker for breast cancer, and is approved for monitoring treatment
124 efficacy only, due to the low sensitivity in early detection (29, 30). CA 15-3 is a
125 secreted form of MUC1, a heavily glycosylated mucin (31, 32). High levels of
126 circulating CA 15-3 have been found in breast cancer patients (33); however,
127 like CA125 for ovarian cancer, levels of serum CA 15-3 are also elevated in
128 other physiological conditions, such as pregnancy (34) and coronary heart
129 disease (35). Despite decades of research, there is no single serum
130 biomarker that has proved useful for the early detection or monitoring of
131 recurrence in breast cancer (36, 37).
132 In this study we developed an improved SubB2M-based SPR assay
133 methodology and used this assay to analyze serum samples collected from
134 breast cancer patients to determine whether detection of Neu5Gc biomarkers
135 may be relevant in screening for and monitoring of breast cancer.

136

137 **Methods**

138 *Expression and purification of SubB2M and $SubB_{A12}$*

139 The recombinant SubB2M and $SubB_{A12}$ proteins were expressed and purified
140 as previously described (17, 38). Briefly, SubB2M and $SubB_{A12}$ were
141 expressed in *E. coli* BL21 (DE3) cells transformed with the SubB2M or
142 $SubB_{A12}$ expression constructs, respectively, as His_6-tagged fusion proteins,
143 which were then purified by Ni-NTA affinity chromatography.

144 *Glycan array analysis of SubB2M and $SubB_{A12}$*

145 Neu5Ac/Neu5Gc glycan array slides were purchased from Z-Biotech (Aurora,
146 Colorado, USA). A 16-subarray slide array was used and glycan array
147 analysis of SubB2M and $SubB_{A12}$ was performed as described previously (18)

148 and as described in Supplementary Table 1. The full list of glycans on the
149 array can be found at https://www.zbiotech.com/neu5gc-neu5ac-n-glycan-
150 array.html.
151 *Development and use of the SubB2M-$_{A12}$-SPR assay for Neu5Gc serum*
152 *biomarkers*
153 SPR was conducted using the Biacore S200 system (GE) with immobilization
154 of SubB2M and SubB$_{A12}$ performed essentially as described previously (19).
155 For glycan analysis, SubB2M was immobilized through flow cells 2 and 3 and
156 SubB$_{A12}$ was immobilized through flow cell 4 (capture levels: 5000-6000
157 Response Units (RU) onto a series S sensor chip CM5 (GE) using the NHS
158 capture kit. Glycans purchased from Chemily Glycoscience (Atlanta, GA)
159 were analysed across a five-fold dilution series in PBS at a maximum
160 concentration of 20μM. Analysis was run using single cycle analysis and
161 double reference subtraction on the Biacore S200 evaluation software.
162 For glycoprotein analysis, SubB2M was immobilized through flow cells 2 and
163 3 and SubB$_{A12}$ was immobilized through flow cell 4. Flow cell 1 was run as a
164 blank immobilization. After immobilization, a start up cycle of 0.5 % normal
165 human serum (Sigma-Aldrich, Cat No. H4522) was run over the immobilized
166 SubB proteins for 10 steps of 30 seconds at 30 μL/minute flow rate to
167 condition the chip. A final wash of 10 mM Tris/1 mM EDTA was run for 30
168 seconds at a 30 μL/minute flow rate prior to beginning the data collection.
169 SPR analysis was performed using multi-cycle analysis and double reference
170 (values from flow cell 1 and 0.5 % normal human serum only) subtraction
171 using the Biacore S200 evaluation software. At least two independent SPR
172 runs were performed for each sample set.

173 For analysis of human sera, samples were diluted 1:200 in PBS and analyzed
174 in duplicate in each SPR run as described above. RU values obtained for
175 each serum sample with $SubB_{A12}$ (flow cell 4) were subtracted from the RU
176 values obtained with SubB2M from flow cells 2 and 3 and averaged to obtain
177 the final RUs used for conversion to GPUs. Two independent SPR runs were
178 performed for each sample set.

179 *Development of Glycoprotein Units (GPUs) standard curve for normalization*
180 *of data from the SubB2M-$_{A12}$-SPR assays*

181 SPR was conducted as described above with SubB2M immobilized through
182 flow cells 2 and 3 and $SubB_{A12}$ immobilized through flow cell 4. To generate
183 an internal calibration curve, bAGP and CA125 were combined at starting
184 concentrations of 15 µg/ml and 15 units/ml, respectively, in 0.5 % normal
185 human serum. For further detail, see Supplementary Methods.

186 *Mass spectrometry glycomic analysis of standard glycoproteins*

187 To confirm the presence of Neu5Gc on the glycoprotein standards (bAGP and
188 human CA125) *N* and *O*-glycans were released and analysed by PGC-LC-
189 MS/MS as previously described (39, 40). For further details, see
190 Supplementary Methods.

191 *Human serum samples*

192 <u>Victorian Cancer Biobank samples</u>

193 Serum samples from cancer-free (normal) females and serum samples from
194 patients with Stage I (n=12), Stage II (n=11), Stage IIIC (n=10) and Stage IV
195 ovarian cancer (n=14) were obtained from the Victorian Cancer Biobank and
196 have been described previously (19). 24 serum samples from patients with
197 Stage I breast cancer, 24 with Stage II breast cancer, 24 with Stage III breast

198 cancer and 24 with Stage IV breast cancer were also obtained from the
199 Victorian Cancer Biobank under application 17020. As described in our
200 previous study (19), 'normal' controls are defined as patients with an apparent
201 non-malignancy diagnosis at the time the sample is taken. The serum
202 samples were collected immediately pre-operatively using Serum Separation
203 Tubes (BD) and were processed and stored at -80 °C within two hours of
204 collection. The patient data and serum samples used in this project were
205 provided by the Victorian Cancer Biobank with informed consent from all
206 donors and use of the samples was approved by the Griffith University HREC
207 (GU Ref No: 2017/732) in accordance with the National Statement on Ethical
208 Conduct in Human Research. The majority of the breast cancers in the cohort
209 were the most common form of breast cancer, invasive ductal carcinoma. The
210 remainder included 8 cases of invasive lobular carcinoma and 6 cases of
211 mucinous carcinoma. Information for each of the ovarian cancer serum
212 samples used in this study can be found in our previous publication (19) while
213 information regarding each of the breast cancer serum samples can be found
214 in Supplementary Table 2.
215 Circulating Biomarkers of Relapse in Breast Cancer (Circ.BR) cohort
216 Circ.BR was established in 2013 as part of the Brisbane Breast Bank (41).
217 Patients with breast cancer who are at high risk for disease recurrence or
218 spread (inclusion criteria below) are followed prospectively for 5 years, with
219 serial collection of blood samples taken at 6 monthly intervals and tumor
220 tissue collected at the time of surgery. Human research ethics committees of
221 The University of Queensland (ref. 2005000785) and The Royal Brisbane and
222 Women's Hospital (2005/022) approved the study with written informed

223 consent obtained from each subject. Serial blood samples from 9 patients
224 who experienced a relapse (median 4 samples per patient, range 3-9) and 6
225 patients who were free from recurrence (median 7 samples per patient, range
226 6-8) at the time of the study were analyzed, with a median follow-up of 19.2
227 months for the relapse group, and 43.9 months for the recurrence-free group.
228 Detailed information for each patient in the Circ.BR cohort are shown in
229 Supplementary Table 3.

230 **Circ.BR inclusion criteria**

231 Invasive breast cancer, grade 3 or grade 2 (score >6) invasive cancer and
232 axillary lymph node positive; OR grade 1 and grade 2 (score ≤6) invasive
233 cancer with adverse features such as tumor size >5cm; Family history
234 (NBOCC group 3)/gene carrier; or previous history of breast cancer.

235 *Statistical analysis*

236 All statistical analyses were performed using GraphPad Prism 8.0. The mean
237 GPUs between normal serum samples compared to cancer patient serum
238 samples were analyzed by two-tailed, unpaired *t*-tests, with a *P* value of <0.05
239 considered significant. Optimal cut-off values from Receiver operating
240 characteristics (ROC) analyses were determined by maximizing the sum of
241 specificity and sensitivity.

242

243 **Results**

244 **Development of the optimized SubB2M-$_{A12}$-SPR assay for detection of**
245 **Neu5Gc serum biomarkers.**

246 In a previous study, we analyzed serum samples from ovarian cancer patients
247 with our SubB2M lectin via the highly sensitive method of SPR (19). This

248 assay is based on label-free detection of Neu5Gc biomarkers that bind to the
249 SubB2M lectin and elicit a response in SPR (see Figure 1). In this study we
250 aimed to improve this SPR-based assay by including a parallel analysis of all
251 samples with a non-sialic acid binding version of SubB called $SubB_{A12}$.
252 $SubB_{A12}$ has the most critical amino acid residue of the B subunit required for
253 binding Neu5Gc mutated from a serine to an alanine (Ser12 > Ala12) (16).
254 Mutation of this Ser residue abolishes interactions with the C1 carboxylate
255 group of sialic acid and thus the $SubB_{A12}$ mutant cannot bind any sialylated
256 glycans. Any binding to $SubB_{A12}$ observed with serum samples must be due to
257 non-sialic acid-dependent interactions of serum components with the SubB
258 protein, for example the binding of antibodies that may recognize the SubB
259 portion of the SubAB toxin. The lack of binding to sialylated glycans by the
260 $SubB_{A12}$ mutant has been described previously (16) and was further
261 confirmed herein with an analysis of $SubB_{A12}$ specificity using a Neu5Ac/Gc
262 glycan microarray, where negligible binding was observed to either Neu5Ac or
263 Neu5Gc-containing glycans by this mutant protein compared to SubB2M
264 **(Table S1 and Figure S1)**. The kinetics of the interaction of SubB2M and
265 $SubB_{A12}$ using SPR analysis with a range of paired synthetic oligosaccharides
266 presenting either a Neu5Ac or Neu5Gc are shown in **Figure 1A** and confirms
267 the loss of all sialic acid binding by $SubB_{A12}$. In the optimized SubB2M-$_{A12}$-
268 SPR assay for the analysis of serum samples, the SPR Response Units
269 (RUs) detected with $SubB_{A12}$ for each serum sample are subtracted from the
270 RUs detected with SubB2M to control for any non-Neu5Gc-dependent binding
271 of serum components to SubB2M **(Figure 1B)**.

272 To further refine our SubB2M-$_{A12}$-SPR assay we established a standard curve
273 using a mixture of two commercially available Neu5Gc-containing
274 glycoproteins at known concentrations to be included as an internal control in
275 each SPR analysis to normalize all data from all studies to common units. As
276 the identity and nature of the Neu5Gc glycoconjugates detected in the ovarian
277 cancer patient serum samples from our 2018 study (19) is currently unknown,
278 we selected a combination of glycoproteins representing a high molecular
279 weight glycoprotein with low Neu5Gc glycosylation and a lower molecular
280 weight glycoprotein with high Neu5Gc glycosylation. These glycoproteins
281 were the human tumor antigen CA125 and bovine AGP (bAGP), respectively.
282 We have previously confirmed the presence of Neu5Gc on bAGP (17) and
283 CA125 (18), and this was reconfirmed for both of the control glycoproteins by
284 mass spectrometry analysis **(Figure S2)**. The standard curve generated by
285 this mixture of glycoproteins diluted into 0.5 % normal human serum was used
286 to calibrate the SubB2M-$_{A12}$-SPR assay. The response units (RUs) obtained
287 for each serum sample was converted to Glycoprotein Units (GPUs)
288 (representative standard curve shown in **Figure S3**).

289 In our 2018 ovarian cancer study (19) we analyzed serum samples from
290 subjects in Stage I (n=12), Stage II (n=11), Stage IIIC (n=10) and Stage IV
291 ovarian cancer (n=14) as well as serum samples from 22 cancer-free females.
292 In the current study, we reanalyzed this sample set using our optimized
293 SubB2M-$_{A12}$-SPR assay **(Figure 2)**. Figure 2A shows the data before
294 subtraction of non-specific binding to SubB$_{A12}$ from each serum sample while
295 Figure 2B shows the data after subtraction of SubB$_{A12}$ responses. As we saw
296 with our original analysis of this sample set (19), significantly elevated serum

297 Neu5Gc biomarker levels were detected at all stages of ovarian cancer
298 compared to cancer-free female controls.
299 Receiver-Operating Characteristic (ROC) analyses were performed on the
300 serum Neu5Gc levels detected with the optimized SubB2M-$_{A12}$-SPR assay
301 before and after SubB$_{A12}$ subtraction **(Supplementary Table 4, Figure S4)**.
302 Subtraction of binding due to SubB$_{A12}$ improved the ability of our SPR-based
303 assay to distinguish cancer-free individuals from ovarian cancer patients at all
304 stages to 100 % specificity and 100 % sensitivity.
305

306 **Serum Neu5Gc levels can discriminate breast cancer patients from**
307 **cancer-free individuals with high specificity and sensitivity**

308 We then used the refined SubB2M-$_{A12}$-SPR assay to determine whether
309 elevated levels of Neu5Gc biomarkers could be detected in serum from
310 patients with breast cancer compared to cancer-free controls. We analyzed a
311 collection of breast cancer serum samples across all stages of disease (24
312 Stage I, 24 Stage II, 24 Stage III and 24 Stage IV) with the same set of
313 cancer-free females used for the ovarian cancer analyses. The SubB2M-$_{A12}$-
314 SPR analysis **(Figure 3)** shows that significantly elevated levels of Neu5Gc
315 biomarkers were detected in serum samples from all stages of disease.
316 Detailed clinical information for each of the serum samples can be found in
317 Supplementary Table 2.
318 ROC analyses were performed to assess the ability of serum Neu5Gc levels
319 detected with the optimized SubB2M-$_{A12}$-SPR assay to discriminate cancer-
320 free females from patients from each stage of breast cancer (**Supplementary**
321 **Table 5, Figure S5**). When all stages are considered as one group with an

322 optimal ROC cut-off value (>10.49 GPUs), as would be the case for a
323 diagnostic screen, the SubB2M-$_{A12}$-SPR assay has 98.96% sensitivity and
324 100% specificity to distinguish patients with breast cancer across all stages of
325 disease from cancer-free individuals (**Figure 4**). In summary, this test
326 achieves 100% specificity and 100% sensitivity for patients with Stage II-IV
327 disease, however, due to one individual data point in the Stage I group,
328 below the limit of detection of our assay, the overall sensitivity did not reach
329 100%. These data indicate that the detection of serum Neu5Gc-biomarkers
330 with SubB2M has potential to detect breast cancer at all stages of disease.
331
332 **Analysis of serum Neu5Gc levels using SubB2M has potential utility for**
333 **treatment monitoring in breast cancer**
334 The Circ.BR cohort is a collection of breast cancer patients with serum
335 samples collected at 6 monthly intervals, allowing us the opportunity to
336 analyze Neu5Gc biomarker levels over the course of disease in these patients
337 at high risk for disease recurrence or spread. Detailed clinical information for
338 each patient can be found in Supplementary Table 3. Analysis of the available
339 serum samples from 15 cases (6 cases in remission, 9 cases with relapse)
340 from this cohort showed a trend for a decrease in Neu5Gc levels immediately
341 following the first line of treatment in the cases who did not have a tumor
342 recurrence, but only in some of the recurrence cases. Figure 5 shows a
343 representative plot from one remission case and one relapse case, with the
344 remaining cases shown in Supplementary Figure S6. These data demonstrate
345 the potential utility of the assessment of serum Neu5Gc levels to monitor

346 treatment response during breast cancer and warrants further exploration and
347 validation.
348
349 **Discussion**
350 CA125 is currently the best performing serum biomarker for ovarian
351 cancer, but due to its limitations, it is not currently used as a screening tool
352 (23). Using our new, optimized SubB2M-$_{A12}$-SPR assay, we are able to
353 distinguish cancer-free females from ovarian cancer patients at all stages of
354 disease with 100 % specificity and 100 % specificity.
355 The SubB2M-$_{A12}$-SPR assay is an improved version of our previously
356 described SubB2M-SPR assay (19). Firstly, we included subtraction of non-
357 specific binding of serum components to SubB$_{A12}$. The inclusion of the
358 SubB$_{A12}$ mutant improved the specificity and sensitivity of our assay to
359 distinguish cancer-free females from early-stage ovarian cancer patients.
360 Secondly, we now report Neu5Gc biomarker levels as GPUs. Including a
361 standard curve in each SPR analysis allowed us to convert Neu5Gc
362 biomarker levels in patient serum samples to GPUs to control for assay-to-
363 assay variability and to accurately compare data from independent studies.
364 While screening mammography is used worldwide for early detection of
365 breast cancer, this method has limitations, including accessibility for women in
366 low-income countries (42), and there is currently no single serum biomarker
367 for the early detection of breast cancer. Our analysis of 96 breast cancer
368 patient serum samples from Stages I-IV, showing that serum Neu5Gc levels
369 can distinguish patients with early-stage breast cancer from cancer-free
370 individuals with high specificity and high sensitivity, and has perfect

371 discriminative ability to distinguish later stage breast cancer from cancer-free
372 individuals, indicates that SubB2M detection of serum Neu5Gc levels has the
373 potential to be developed into a tool for the early detection of breast cancer. It
374 is important to note that this cohort, whilst comprising mostly invasive ductal
375 carcinoma, also included the less common invasive lobular carcinoma, and
376 six cases of mucinous carcinoma. Mucinous carcinoma of the breast is a rare
377 form of invasive ductal carcinoma, making up 2-3% of breast cancer patients.
378 That all of these cancers could be differentiated from cancer-free controls by
379 our assay supports the broad applicability of the approach. Our analysis of
380 serially collected serum samples from the Circ.BR cohort showed that
381 SubB2M detection of Neu5Gc levels also has potential to be used to monitor
382 response to treatment, although further investigation of its utility, using a
383 larger cohort, is required.
384 Neu5Gc is well known to be found in human cancer (in cells, tissues and
385 secretions) (3, 5-9, 43), and may be found in trace amounts in some healthy
386 human tissues (44). The source of Neu5Gc in humans is assumed to be
387 dietary incorporation following consumption of animal-derived food. Dietary
388 acquisition of Neu5Gc and metabolic incorporation into human cells forms the
389 basis of the 'xenosialitis' hypothesis, which proposes that autoantibodies
390 against Neu5Gc at the cell surface promote inflammation and pathology (see
391 reference 45 for a recent review) (45). The direct biochemical evidence for
392 dietary uptake and expression in humans involved three human subjects
393 consuming Neu5Gc in the form of porcine submaxillary mucin (44).
394 Alternatively, It has been proposed that human cancer cells can produce
395 endogenous Neu5Gc by providing electron donors that circumvent the CMAH

396 inactivating truncation found in the human enzyme (46). The current study
397 and our previous data (19) report elevated levels of Neu5Gc in serum can
398 differentiate breast and ovarian cancer patients (n=143) from cancer-free
399 controls, and supports the hypothesis of endogenous production of this
400 biomarker by tumor cells in addition to potential dietary uptake.
401 Our development of the purpose-engineered SubB2M lectin implemented
402 in the SubB2M-$_{A12}$-SPR assay provides a sensitive, specific and practical
403 reagent for the measurement of Neu5Gc-containing glycoproteins in serum,
404 which are an under-explored class of circulating tumor biomarkers. The
405 detection of Neu5Gc glycosylation via SubB2M may also be used in
406 conjunction with existing biomarker assays to improve specificity and
407 sensitivity and represents a promising area for future studies.
408
409 **Conclusions**
410 Minimally invasive, rapid and reliable methods for the early detection of
411 ovarian and breast cancer will have a significant impact on patient survival
412 rates. Biomarkers that can accurately and specifically monitor response to
413 treatment and disease progression would also drastically improve patient
414 outcomes. Our SubB2M lectin assay offers the opportunity to exploit
415 measurement of Neu5Gc biomarkers as a path to achieve these outcomes.
416
417
418
419
420

421 **Declarations**

422 **Ethics approval and consent to participate**

423 The patient data and serum samples used in this project were provided by the
424 Victorian Cancer Biobank with informed consent from all donors and use of
425 the samples was approved by the Griffith University HREC (GU Ref No:
426 2017/732). Human research ethics committees of The University of
427 Queensland (ref. 2005000785) and The Royal Brisbane and Women's
428 Hospital (2005/022) approved the Circ.BR study with written informed consent
429 obtained from each subject.

430 **Consent for publication**

431 Not applicable

432 **Availability of data and materials**

433 All data generated or analysed during this study are included in this published
434 article and its supplementary information files.

435 **Competing interests**

436 The following authors (JCP, AWP, CJD and MPJ) declare that they are
437 named inventors on a patent on the SubB2M technology (WO2018085888A1)
438 and a second patent (JCP, AWP, CJD, LKS and MPJ) for improvements to
439 the SubB2M test (2021901444). Both of these patents were licensed to
440 BARD1 Life Sciences, VIC, Australia in 2020.

441 **Funding**

442 This work was supported by the National Health and Medical Research
443 Council (NHMRC) (Program Grant 1071659 to JCP and MPJ, Project Grant
444 1084050 to AWP, Principal Research Fellowship 1138466 to MPJ) and the
445 US Department of Defense (Breakthrough Award Level 1 W81XWH-20-1-

446 0527 to MPJ, CJD, LKS, JLA, JRK, SRL, JCP and AWP). The Brisbane
447 Breast Bank and Circ.BR study are supported by the NHMRC (Program grant
448 APP1113867 to SRL) and the Royal Brisbane and Women's Hospital
449 Foundation grant (to SRL).

450 **Author's contributions**

451 LKS curated all serum samples, analysed and interpreted SPR data,
452 produced figures, contributed to design of the study and was a major
453 contributor to writing of the manuscript, CJD acquired and analysed SPR and
454 glycan array data, produced figures, contributed to conceptualisation and
455 design of the study and contributed to writing of the manuscript, JRK
456 contributed to analysing and interpreting Circ.BR data and writing of the
457 manuscript, JLA acquired and interpreted mass spectrometry data and
458 contributed to review and editing of the manuscript, JW contributed to curation
459 of serum samples, SPR analysis and review and editing of the manuscript, JP
460 acquired glycan array data and contributed to review and editing of the
461 manuscript, CN contributed to the Circ.BR cohort and read and approved the
462 final manuscript, KF contributed to the Circ.BR cohort and read and approved
463 the final manuscript, JMS contributed to the Circ.BR cohort and read and
464 approved the final manuscript SRL contributed to the Circ.BR cohort and read
465 and approved the final manuscript**,** MvI provided resources and read and
466 approved the final manuscript, JCP contributed to conceptualisation and
467 design of the study, provided SubB2M and $SubB_{A12}$ proteins and reviewed
468 and edited the manuscript, AWP contributed to conceptualisation and design
469 of the study, provided SubB2M and $SubB_{A12}$ proteins and reviewed and edited

470 the manuscript, MPJ conceptualised and designed the study, contributed to
471 writing of the manuscript and provided supervision of the study.

472 **Acknowledgements**

473 We thank the Victorian Cancer Biobank for providing all patient data and
474 serum samples. The Victorian Cancer Biobank, through the Cancer Council
475 Victoria as Lead Agency, is supported by the Victorian Government through
476 the Victorian Cancer Agency, a business unit of the Department of Health and
477 Human Services.

478 **References**

479 1 Varki NM, Varki A. Diversity in cell surface sialic acid presentations:
480 implications for biology and disease. *Lab Invest*. 2007;87(9):851-857.
481 2 Chou HH, Hayakawa T, Diaz S, et al. Inactivation of CMP-N-
482 acetylneuraminic acid hydroxylase occurred prior to brain expansion
483 during human evolution. *Proceedings of the National Academy of*
484 *Sciences of the United States of America*. 2002;99(18):11736-11741.
485 3 Inoue S, Sato C, Kitajima K. Extensive enrichment of N-
486 glycolylneuraminic acid in extracellular sialoglycoproteins abundantly
487 synthesized and secreted by human cancer cells. *Glycobiology*.
488 2010;20(6):752-762.
489 4 Samraj AN, Laubli H, Varki N, Varki A. Involvement of a non-human
490 sialic Acid in human cancer. *Front Oncol*. 2014;4:33.
491 5 Higashi H, Hirabayashi Y, Fukui Y, et al. Characterization of N-
492 glycolylneuraminic acid-containing gangliosides as tumor-associated
493 Hanganutziu-Deicher antigen in human colon cancer. *Cancer research*.
494 1985;45(8):3796-3802.

495 6 Malykh YN, Schauer R, Shaw L. N-Glycolylneuraminic acid in human
496 tumours. *Biochimie*. 2001;83(7):623-634.
497 7 Furukawa J, Tsuda M, Okada K, et al. Comprehensive Glycomics of a
498 Multistep Human Brain Tumor Model Reveals Specific Glycosylation
499 Patterns Related to Malignancy. *PloS one*. 2015;10(7):e0128300.
500 8 Tzanakakis GN, Nikitovic D, Katonis P, Kanakis I, Karamanos NK.
501 Expression and distribution of N-acetyl and N-glycolylneuraminic acids
502 in secreted and cell-associated glycoconjugates by two human
503 osteosarcoma cell lines. *Biomedical chromatography : BMC*.
504 2007;21(4):406-409.
505 9 Tzanakakis GN, Syrokou A, Kanakis I, Karamanos NK. Determination
506 and distribution of N-acetyl- and N-glycolylneuraminic acids in culture
507 media and cell-associated glycoconjugates from human malignant
508 mesothelioma and adenocarcinoma cells. *Biomedical chromatography*
509 *: BMC*. 2006;20(5):434-439.
510 10 Labrada M, Dorvignit D, Hevia G, et al. GM3(Neu5Gc) ganglioside: an
511 evolution fixed neoantigen for cancer immunotherapy. *Seminars in*
512 *Oncology*. 2018;45(1):41-51.
513 11 Marquina G, Waki H, Fernandez LE, et al. Gangliosides expressed in
514 human breast cancer. *Cancer research*. 1996;56(22):5165-5171.
515 12 Higashi H, Sasabe T, Fukui Y, Maru M, Kato S. Detection of
516 gangliosides as N-glycolylneuraminic acid-specific tumor-associated
517 Hanganutziu-Deicher antigen in human retinoblastoma cells. *Japanese*
518 *journal of cancer research : Gann*. 1988;79(8):952-956.

519 13 Priego-Capote F, Orozco-Solano MI, Calderon-Santiago M, Luque de
520 Castro MD. Quantitative determination and confirmatory analysis of N-
521 acetylneuraminic and N-glycolylneuraminic acids in serum and urine by
522 solid-phase extraction on-line coupled to liquid chromatography-
523 tandem mass spectrometry. *Journal of chromatography A*.
524 2014;1346:88-96.
525 14 Kitajima K, Varki N, Sato C. Advanced Technologies in Sialic Acid and
526 Sialoglycoconjugate Analysis. *Topics in current chemistry*.
527 2015;367:75-103.
528 15 Diaz SL, Padler-Karavani V, Ghaderi D, et al. Sensitive and specific
529 detection of the non-human sialic Acid N-glycolylneuraminic acid in
530 human tissues and biotherapeutic products. *PloS one*.
531 2009;4(1):e4241.
532 16 Byres E, Paton AW, Paton JC, et al. Incorporation of a non-human
533 glycan mediates human susceptibility to a bacterial toxin. *Nature*.
534 2008;456(7222):648-652.
535 17 Day CJ, Paton AW, Higgins MA, et al. Structure aided design of a
536 Neu5Gc specific lectin. *Scientific reports*. 2017;7(1):1495.
537 18 Wang J, Shewell LK, Paton AW, Paton JC, Day CJ, Jennings MP.
538 Specificity and utility of SubB2M, a new N-glycolylneuraminic acid
539 lectin. *Biochemical and biophysical research communications*.
540 2018;500(3):765-771.
541 19 Shewell LK, Wang JJ, Paton JC, Paton AW, Day CJ, Jennings MP.
542 Detection of N-glycolylneuraminic acid biomarkers in sera from patients
543 with ovarian cancer using an engineered N-glycolylneuraminic acid-

544 specific lectin SubB2M. *Biochemical and biophysical research*
545 *communications*. 2018;507(1-4):173-177.
546 20 Yin BW, Lloyd KO. Molecular cloning of the CA125 ovarian cancer
547 antigen: identification as a new mucin, MUC16. *The Journal of*
548 *biological chemistry*. 2001;276(29):27371-27375.
549 21 Bast RC, Jr., Klug TL, St John E, et al. A radioimmunoassay using a
550 monoclonal antibody to monitor the course of epithelial ovarian cancer.
551 *The New England journal of medicine*. 1983;309(15):883-887.
552 22 Goonewardene T, Hall MR, Rustin GJS. Management of asymptomatic
553 patients on follow-up for
554 ovarian cancer with rising CA-125 concentrations. *Lancet Oncology*.
555 2007;8:813-821.
556 23 Jacobs I, Menon U, Ryan Aea. Ovarian cancer screening and mortality
557 in the UK Collaborative Trial of Ovarian Cancer Screening
558 (UKCTOCS): a randomised controlled trial. . *Lancet*. 2016;387:945-
559 956.
560 24 Menon U, Gentry-Maharaj A, Burnell M, et al. Ovarian cancer
561 population screening and mortality after long-term follow-up in the UK
562 Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a
563 randomised controlled trial. *Lancet*. 2021.
564 25 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A.
565 Global cancer statistics 2018: GLOBOCAN estimates of incidence and
566 mortality worldwide for 36 cancers in 185 countries. *CA: a cancer*
567 *journal for clinicians*. 2018;68(6):394-424.

568 26 Groot MT, Baltussen R, Uyl-de Groot CA, Anderson BO, Hortobagyi
569 GN. Costs and health effects of breast cancer interventions in
570 epidemiologically different regions of Africa, North America, and Asia.
571 *The breast journal*. 2006;12 Suppl 1:S81-90.
572 27 Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of
573 screening mammography, physical examination, and breast US and
574 evaluation of factors that influence them: an analysis of 27,825 patient
575 evaluations. *Radiology*. 2002;225(1):165-175.
576 28 Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk
577 and detection of breast cancer. *The New England journal of medicine*.
578 2007;356(3):227-236.
579 29 Cheung KL, Graves CR, Robertson JF. Tumour marker measurements
580 in the diagnosis and monitoring of breast cancer. *Cancer treatment*
581 *reviews*. 2000;26(2):91-102.
582 30 Harris L, Fritsche H, Mennel R, et al. American Society of Clinical
583 Oncology 2007 update of recommendations for the use of tumor
584 markers in breast cancer. *Journal of clinical oncology : official journal of*
585 *the American Society of Clinical Oncology*. 2007;25(33):5287-5312.
586 31 Gendler SJ, Lancaster CA, Taylor-Papadimitriou J, et al. Molecular
587 cloning and expression of human tumor-associated polymorphic
588 epithelial mucin. *The Journal of biological chemistry*.
589 1990;265(25):15286-15293.
590 32 Nath S, Mukherjee P. MUC1: a multifaceted oncoprotein with a key role
591 in cancer progression. *Trends in molecular medicine*. 2014;20(6):332-
592 342.

593 33 Devine PL, McGuckin MA, Ramm LE, Ward BG, Pee D, Long S.
594 Serum mucin antigens CASA and MSA in tumors of the breast, ovary,
595 lung, pancreas, bladder, colon, and prostate. A blind trial with 420
596 patients. *Cancer*. 1993;72(6):2007-2015.
597 34 Han SN, Lotgerink A, Gziri MM, Van Calsteren K, Hanssens M, Amant
598 F. Physiologic variations of serum tumor markers in gynecological
599 malignancies during pregnancy: a systematic review. *BMC medicine*.
600 2012;10:86.
601 35 Li X, Xu Y, Zhang L. Serum CA153 as biomarker for cancer and
602 noncancer diseases. *Progress in molecular biology and translational*
603 *science*. 2019;162:265-276.
604 36 Tang SS, Gui GP. Biomarkers in the diagnosis of primary and recurrent
605 breast cancer. *Biomarkers in medicine*. 2012;6(5):567-585.
606 37 Loke SY, Lee ASG. The future of blood-based biomarkers for the early
607 detection of breast cancer. *European journal of cancer*. 2018;92:54-68.
608 38 Paton AW, Srimanote P, Talbot UM, Wang H, Paton JC. A new family
609 of potent AB(5) cytotoxins produced by Shiga toxigenic Escherichia
610 coli. *The Journal of experimental medicine*. 2004;200(1):35-46.
611 39 Jensen P, Karlsson N, Kolarich D, Packer N. Structural analysis of N-
612 and O-glycans released from glycoproteins. *Nat Protoc*.
613 2012;7(7):1299-1310.
614 40 Abrahams J, Campbell M, Packer N. Building a PGC-LC-MS N-glycan
615 retention library and elution mapping resource. *Glycoconjugate journal*.
616 2018;35(1):15-29.

617 41 McCart Reed AE, Saunus JM, Ferguson K, Niland C, Simpson PT,
618 Lakhani SR. The Brisbane Breast Bank. *Open Journal of Bioresources*.
619 2018;5:5.
620 42 Rivera-Franco MM, Leon-Rodriguez E. Delays in Breast Cancer
621 Detection and Treatment in Developing Countries. *Breast Cancer:*
622 *Basic and Clinical Research*. 2018;12:1-5.
623 43 Samraj AN, Pearce OM, Laubli H, et al. A red meat-derived glycan
624 promotes inflammation and cancer progression. *Proceedings of the*
625 *National Academy of Sciences of the United States of America*.
626 2015;112(2):542-547.
627 44 Tangvoranuntakul P, Gagneux P, Diaz S, et al. Human uptake and
628 incorporation of an immunogenic nonhuman dietary sialic acid.
629 *Proceedings of the National Academy of Sciences of the United States*
630 *of America*. 2003;100(21):12045-12050.
631 45 Dhar C, Sasmal A, Varki A. From "Serum Sickness" to "Xenosialitis":
632 Past, Present, and Future Significance of the Non-human Sialic Acid
633 Neu5Gc. *Frontiers in immunology*. 2019;10:807.
634 46 Bousquet PA, Sandvik JA, Jeppesen Edin NF, Krengel U. Hypothesis:
635 Hypoxia induces de novo synthesis of NeuGc gangliosides in humans
636 through CMAH domain substitute. *Biochemical and biophysical*
637 *research communications*. 2018;495(1):1562-1566.
638
639
640
641

642 **Figures**

643 **Figure 1. Characterization of the SubB2M-$_{A12}$-SPR assay workflow used**
644 **in this study. A)** SPR analysis of Neu5Ac/Neu5Gc glycan pairs. NCDI: No
645 concentration dependent interaction with glycan up to 20 μM.
646 **B)** Optimized SubB2M-$_{A12}$-SPR assay. Serum from cancer-free (normal)
647 females and ovarian cancer patients were analyzed by SPR with SubB2M
648 immobilized onto the surface of the sensor chip through flow cells 2 and 3 and
649 SubB$_{A12}$ immobilized onto the sensor surface through flow cell 4.
650

651 **Figure 2. Analysis of cancer-free and Stage I-IV ovarian cancer patient**
652 **serum samples with the optimized SubB2M-$_{A12}$-SPR assay.** 22 serum
653 samples from cancer-free (normal) females, 12 patients with Stage I ovarian
654 cancer, 11 with Stage II ovarian cancer, 10 with Stage IIIC ovarian cancer and
655 14 with Stage IV ovarian cancer were analyzed by the optimized SubB2M-
656 SPR assay. The mean GPUs from duplicate analyses for each serum sample
657 determined **A)** before and **B)** after subtraction of binding due to SubB$_{A12}$ are
658 shown. Error bars = ± 1 SD from the mean for each group. Statistical analysis
659 was performed using two-tailed unpaired *t*-tests. **** = *P*-value <0.0001
660 compared to Normal. Duplicate, independent assays were performed with
661 both showing the same trends. One representative assay is shown. **C)**
662 Descriptive statistics of data from ovarian cancer patient serum samples and
663 cancer-free controls.
664

665 **Figure 3. SubB2M-$_{A12}$-SPR analysis of cancer-free and Stage I – IV breast**
666 **cancer serum samples. A)** 22 serum samples from cancer-free females, 24

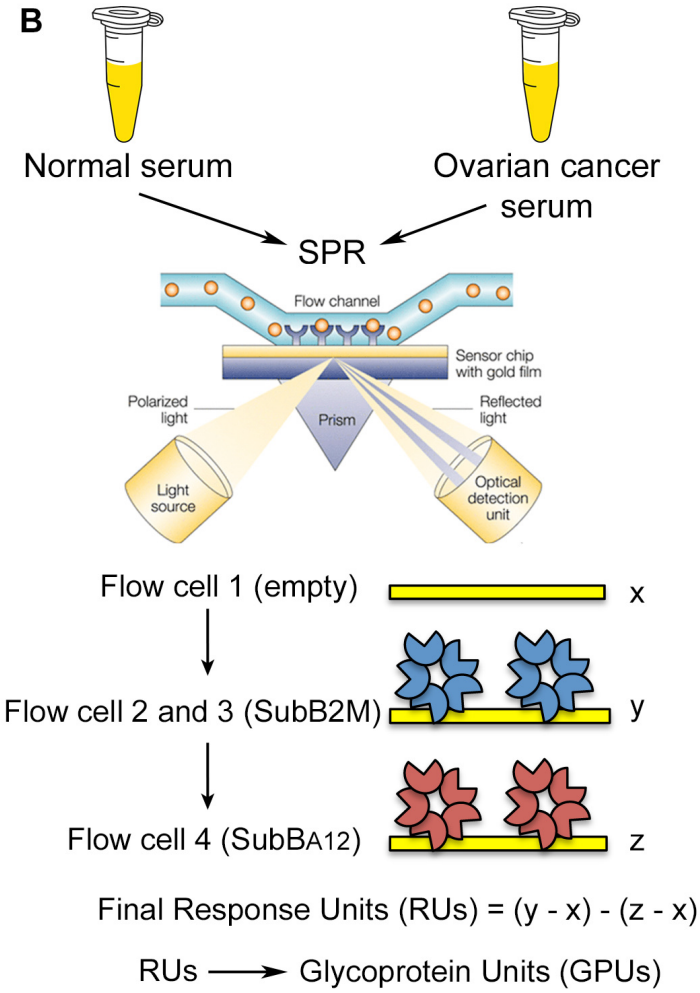
667 Stage I, 24 Stage II, 24 Stage III and 24 Stage IV serum samples were
668 analyzed by SubB2M-$_{A12}$-SPR assay. The mean GPUs from duplicate
669 analyses for each serum sample are shown. Error bars = ± 1 SD from the
670 mean for each group. Statistical analysis was performed using two-tailed
671 unpaired *t*-tests. **** = *P*-value <0.0001 compared to Normal. Duplicate,
672 independent assays were performed with both showing the same trends. One
673 representative assay is shown. Only values above 0 are shown. **B)** Cancer-
674 free controls compared to early stage breast cancer samples only (Stage I
675 and II samples only from Figure 3A). **C)** Descriptive statistics of data from
676 breast cancer patient serum samples and cancer-free controls.
677
678 **Figure 4. ROC curve depicting the ability of serum Neu5Gc levels**
679 **determined by the optimized SubB2M-$_{A12}$-SPR assay to distinguish**
680 **breast cancer patients from cancer-free (normal) individuals.** Sensitivity%
681 (true positive rate; ability to detect disease) is plotted against 100 %-
682 specificity% (false positive rate or 100 %-true negative rate; ability to detect
683 lack of disease). ROC analyses were performed with the data shown in Figure
684 3A using Graphpad Prism 8.0.
685
686 **Figure 5. Representative plots of serum Neu5Gc levels determined by**
687 **SubB2M-$_{A12}$-SPR assay for A) one remission case and B) one relapse**
688 **case from the Circ.BR cohort.** The mean GPUs from duplicate analyses for
689 each serum sample are shown. Error bars = ± 1 SD from the mean for each
690 group. Two independent assays were performed with both showing the same
691 trends with results from one assay presented. Plots for the remaining cases

692 are shown in Supplementary Figure S6. Clinical information for each patient is
693 shown in the top right of each plot with treatment history and metastases
694 overlaid. ALND: Axillary lymph node dissection, ILC: Invasive Lobular
695 Carcinoma, Mast: mastectomy, SNB: sentinel node biopsy, WLE: wide local
696 excision, XRT: radiation therapy.
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716

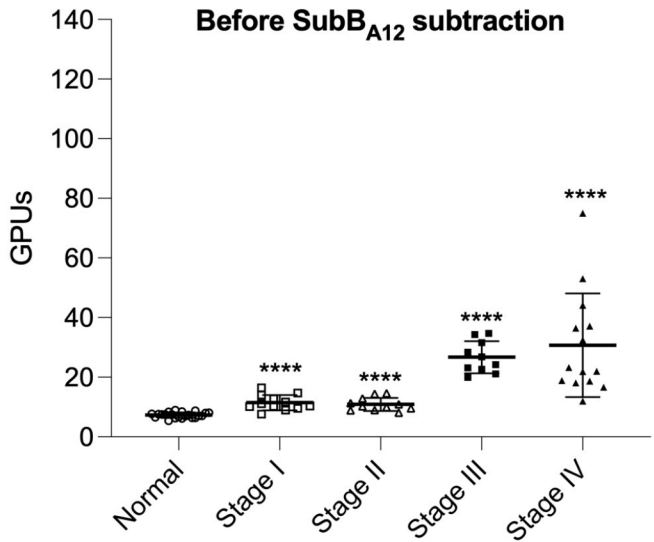
A

Structure	Symbol	SubB2M	$SubB_{A12}$
Neu5Ac	◆	NCDI	NCDI
Neu5Gc	◇	K_D = 18.4nM±2.9	NCDI
lactose	●–β 4–●	NCDI	NCDI
lactosamine	●–β 4–■	NCDI	NCDI
Neu5Ac-α2-3-lac	◆–α 3–●–β 4–■	NCDI	NCDI
Neu5Gc-α2-3-lac	◇–α 3–●–β 4–■	K_D = 16.8nM±2.1	NCDI
Neu5Ac-α2-6-lac	◆–α 6–●–β 4–■	NCDI	NCDI
Neu5Gc-α2-6-lac	◇–α 6–●–β 4–■	K_D = 8.6nM±0.22	NCDI
sLeX-Ac	◆–α 3–●–β 4–■, ■–3 α–▲	NCDI	NCDI
sLeX-Gc	◇–α 3–●–β 4–■, ■–3 α–▲	K_D = 92.4nM±4.2	NCDI

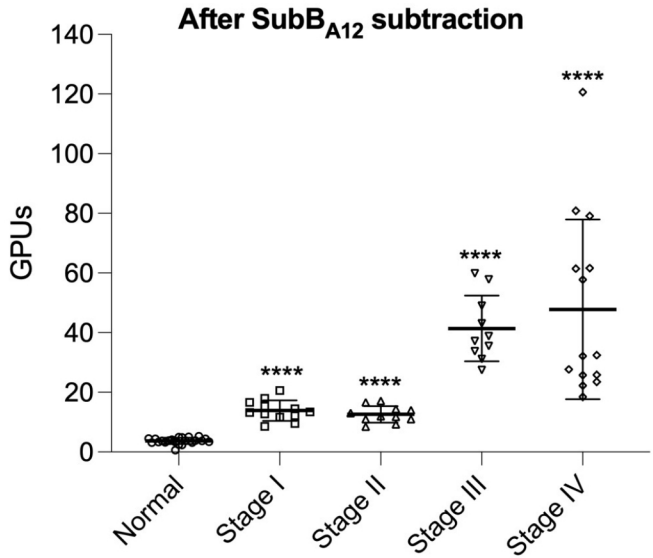
B



A

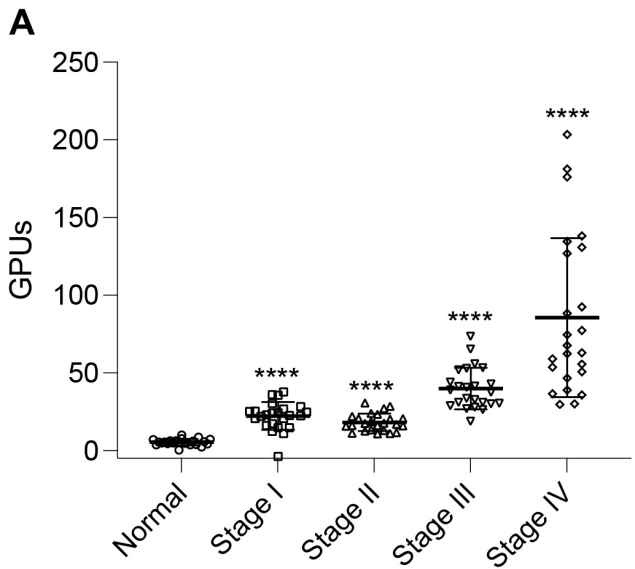


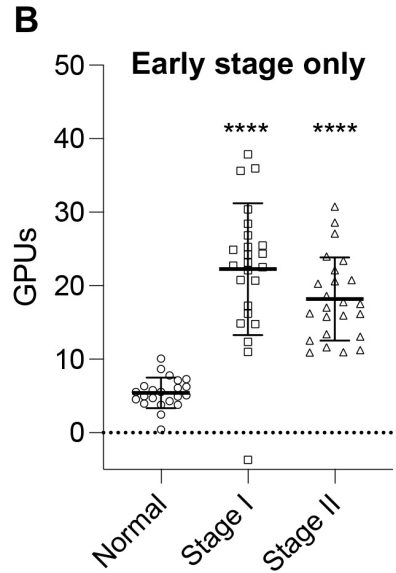
B



C

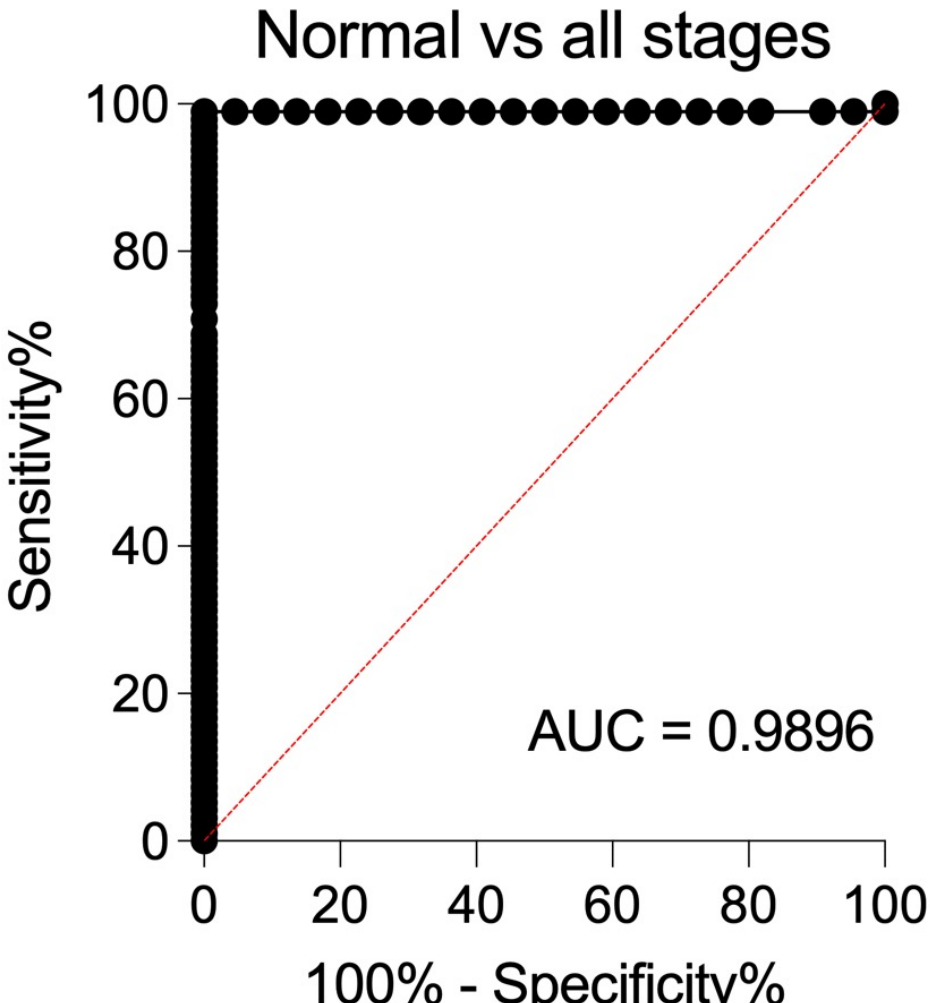
	Normal		Stage I		Stage II		Stage III		Stage IV	
Number of samples	22		12		11		10		14	
SubB$_{A12}$ subtraction	Before	After	Before	After	Before	After	Before	After	Before	After
Minimum	5.38	0.668	7.67	8.553	8.10	8.509	19.99	27.55	11.95	18.49
Maximum	9.02	5.25	16.50	20.62	14.61	17.11	34.75	59.89	75.02	120.6
Mean	7.31	3.72	11.48	13.88	10.89	12.63	26.71	41.42	30.7	47.83
SD	0.933	1.04	2.58	3.434	2.20	2.758	5.4	11.00	17.37	30.14

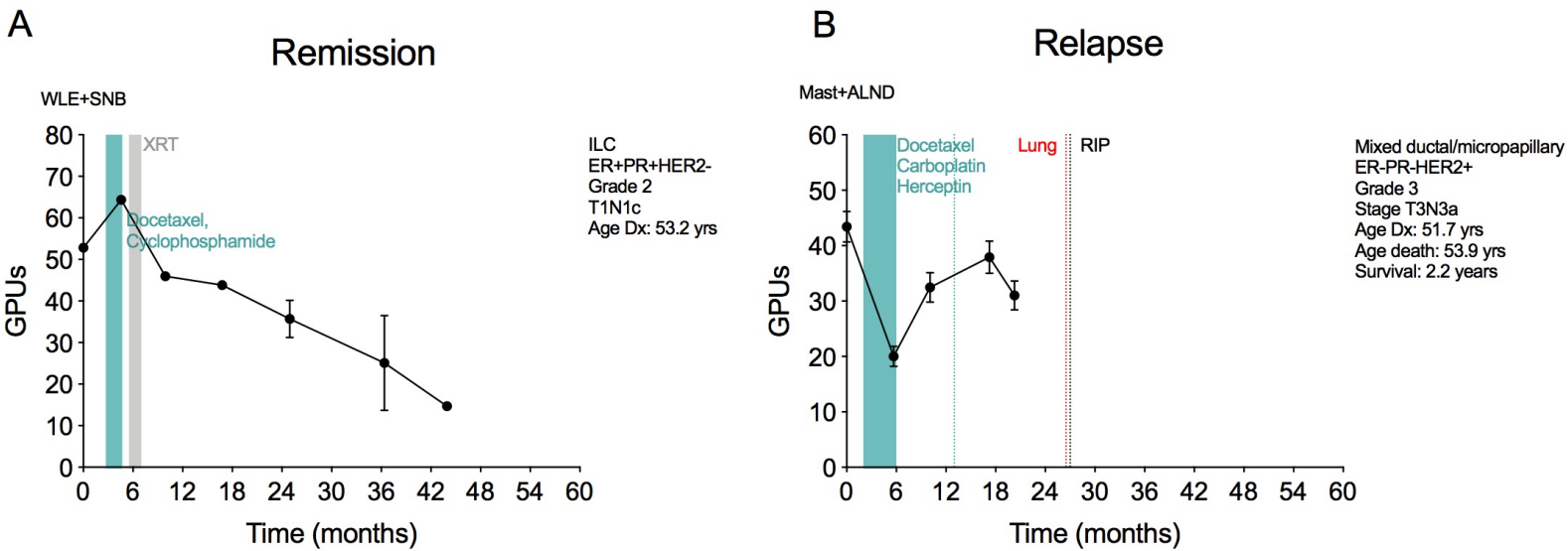




C

	Normal	Stage I	Stage II	Stage III	Stage IV
Number of samples	22	24	24	24	24
Minimum	0.4233	-3.674	10.91	19.00	29.66
Maximum	10.08	37.86	30.74	73.66	203.4
Mean	5.430	22.25	18.18	39.96	85.60
SD	2.076	8.953	5.659	13.28	51.21


Normal vs all stages
Sensitivity%
100
80
60
40
20
0
AUC = 0.9896
0
20
40
60
80
100
100% - Specificity%


A
Remission
WLE+SNB
XRT
Docetaxel,
Cyclophosphamide
ILC
ER+PR+HER2-
Grade 2
T1N1c
Age Dx: 53.2 yrs
GPUs
Time (months)
B
Relapse
Mast+ALND
Docetaxel
Carboplatin
Herceptin
Lung
RIP
Mixed ductal/micropapillary
ER-PR-HER2+
Grade 3
Stage T3N3a
Age Dx: 51.7 yrs
Age death: 53.9 yrs
Survival: 2.2 years
GPUs
Time (months)