Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

1 **A novel patient-specific protein expression-based algorithmic model for immune**
2 **checkpoint inhibition therapy response prediction in a pan-cancer cohort**

3 Oluwaseun Adebayo Bamodu, MD., PhD[1, 2, 3#], Wei-Hong Cheng, MD[1#], Oluleke Bamodu,
4 MEng[4], Wei-Hwa Lee, MD., PhD[2, 5], Kang-Yun Lee, MD., PhD[3,6], Liang-Shun Wang, MD.,
5 PhD[1, 2, 3], Tsu-Yi Chao, MD., PhD[1,2, 3, 7*], Chi-Tai Yeh, PhD [1, 2, 3*]
6
7 [1]Department of Hematology and Oncology, Cancer Center, Taipei Medical University - Shuang
8 Ho Hospital, New Taipei City 235, Taiwan. 16625@s.tmu.edu.tw; [2]Department of Medical
9 Research & Education, Taipei Medical University - Shuang Ho Hospital, New Taipei City 235,
10 Taiwan. ctyeh@s.tmu.edu.tw; [3]Graduate Institute of Clinical Medicine, School of Medicine,
11 Taipei Medical University, Taipei City 110, Taiwan; [4]Center for Sustainable Energy
12 Technologies, University of Nottingham, Ningbo, Republic of China.
13 oluleke.bamodu@nottingham.edu.cn; [5]Department of Pathology, Taipei Medical University -
14 Shuang Ho Hospital, New Taipei City 235, Taiwan. whlpath97616@s.tmu.edu.tw; [6]Division of
15 Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University - Shuang Ho
16 Hospital, New Taipei City 235, Taiwan. leekangyun@tmu.edu.tw; [7]Taipei Cancer Center, Taipei
17 Medical University, Taipei City, Taiwan. 10575@s.tmu.edu.tw
18
19 [#] Equal Contribution
20 * Corresponding author(s):
21 Chi-Tai Yeh, PhD
22 Department of Medical Research and Education, Taipei Medical University - Shuang Ho
23 Hospital, New Taipei City 23561, Taiwan; Tel: +886-2-2490088 ext. 8881,
24 Fax: +886-2-2248-0900, E-mail: ctyeh@s.tmu.edu.tw
25
26 Tsu-Yi Chao, MD., PhD
27 Department of Hematology and Oncology, Cancer Center, Taipei Medical University - Shuang
28 Ho Hospital, New Taipei City 23561, Taiwan; Tel: +886-2-2490088 ext. 8885, Fax: +886-2-
29 2248-0900, E-mail: 10575@s.tmu.edu.tw
30
31
32
33

© 2019 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

34 **Abstract**

35

36 **Background:** Accurate prediction of patients' response to therapy is clinically indispensable,
37 howbeit challenging. With increased understanding of the human genome and malignancies,
38 there is the renaissance of *in silico* pharmacogenomics with renewed interest in drug response
39 predictability based on gene-drug interaction. **Objective:** Evidence-based transcript-proteome
40 profiling is essential for synthesizing clinically applicable algorithms for predicting response to
41 anticancer therapy, including immune checkpoint blockade (ICBT); thus, saving physicians' time,
42 reducing polypharmacy, and curtailing unnecessary treatment expense. In this study, we tested
43 and validated the hypothesis that a selected proteomic signature in ICBT-naïve patients is
44 sufficient for the prediction of response to ICBT. **Methods:** Using a multimodal approach
45 consisting of computational pharmacogenomics, transcript-proteome analytics, mathematical
46 modeling, and machine learning systems; we delineated therapy-sensitivity and stratified patients
47 into graduated response groups based on their proteomic profile. Protein expression levels in our
48 cohort tissue specimens were evaluated based on T cell- and non-T cell- inflamed phenotypes by
49 immunohistochemistry. **Results:** We established β-catenin, PDL1, CD3 and CD8 expression-
50 based ICBT response model. Statistical regression models validated the predictive association
51 between our predefined algorithms and therapeutic outcome. Interestingly, our 4-gene prediction
52 classifier was constitutively independent of tumor tissue origin, correctly stratified patients into
53 high-, low-, and non- responders pre-treatment, with high prediction accuracy, and exhibited
54 good association with patients' performance status and prognosis ($p < 0.01$). **Conclusion:** Our
55 findings demonstrate the possibility of accurate proteomics based ICBT response prediction and
56 provide a putative basis for drug response prediction based on selective proteome profile in
57 untreated cancer patients.

58

59 ***Keywords:*** Pharmacogenomics, immune checkpoint blockade, immunotherapy, drug response
60 prediction, algorithm, mathematical model, precision medicine, personalized medicine

61

62 *Working title:* 4-gene immune checkpoint blockade response prediction algorithm

63

64

65

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

66 **Introduction**

67 The ability to accurately predict the predisposition to respond or actual response of
68 individual cancer patients to treatment, including immune checkpoint blockade, *ab initio*, based
69 on the clinical and molecular profile of such patients is very essential for the effectual
70 individualization of anticancer therapy, or what is now termed 'personalized or precision
71 medicine' (1- 5). The relevance of pharmacogenomics in physicians' therapeutic decisions, in
72 cognizance of the dynamic individual molecular and medical phenotypes, as well as the
73 differential efficacy and toxicity of anticancer agents in different cancer patients (6, 7), cannot be
74 overstated. The last decade has witnessed significant advances in our ability to predict drug
75 sensitivity and identify their toxicities based on variations in the human genomic or proteomic
76 profile, birthing two distinct but related disciplines, pharmacogenomics and pharmacoproteomics
77 (8 - 10). The integral distinction of the two disciplines is highlighted by the fact that genomic
78 profiling provides bio-information at the transcript level, which often lacks direct correlation
79 with the human proteome pattern that ultimately modulate physiopathological processes,
80 including carcinogenesis and disease progression (11, 12). Nevertheless, cumulative evidence
81 suggests that the alternating dynamism of the human molecular profile or phenotype, is not only
82 indicative, but also predictive of therapeutic response and tolerance (1, 6 - 10), thus, rendering
83 them actionable. The clinical relevance of this actionable molecular phenotypic dynamism is
84 evident in the the presence or absence of gene signatures which are indicative of activated $CD8^+$
85 T cells infiltration within the tumor niche or invasive margin, commonly referred to as the T cell-
86 inflamed and non-T cell-inflamed tumor microenvironment (TME) phenotype, respectively (13).

87 The T cell-inflamed phenotype is characterized by pre-existing intra-tumoral or marginal
88 presence of $CD8^+$ T cells which has been shown to be positively correlated with response to
89 programmed death-1 (PD-1) receptor inhibition and tumor regression (14). This may be
90 connected with the constant immune-editing of these types of tumors, enhancing their

91 immunogenic antigen expression and presentation, while concurrently limiting their immune-
92 inhibitory mechanisms to elicit self-destruction, consequently making these tumors responsive to
93 immune-modulating therapies, such as the ICBT. In contrast, the non-T cell-inflamed phenotype
94 is characterized by reduced $CD8^{+}$ T cell population, aberrant immune-inhibitory signaling
95 including programmed cell death protein 1 (PD-1), programmed cell death 1 ligand 1 (PD-L1),
96 and cytotoxic T-lymphocyte associated protein 4 (CTLA-4) signals, with reduced therapeutic
97 efficacy (14, 15). In addition, the aberrant activation of the Wnt/β-catenin signaling pathway has
98 been implicated in the non-T cell-inflamed phenotype, and inversely correlated with PD-L1 and
99 activated CD8 expression (13, 16). Despite increased knowledge of 'drug-genome' or 'drug-
100 proteome' interactions, renaissance of patient-tailored medicine, and advances in computational
101 pharmacology; the prediction of anticancer drug response *in silico* remains a clinical challenge,
102 partly because of the multifactorial polyetiologism and intrinsic heterogeneity of tumors (17).
103 Hence, an *in-silico* model that accurately predicts drug efficacy based on the molecular
104 constitution of patients' tumor would be a step in the right direction towards precision medicine.

105 Based on increasingly compelling evidence that chemosensitivity is strongly dependent on
106 alterations in the cancer proteogenomic landscape, this present study investigated the feasibility
107 of predicting drug response and now present an histology-independent model for predicting
108 response to pembrolizumab (anti-PD-1), ipilimumab (anti-CTLA-4), or nivolumab (anti-PD-1)
109 single-agent or combinatorial ICBT in patients with cancer, based on selected non-T cell-
110 inflamed and T cell-inflamed markers, namely, β-catenin (CTNNB1), PDL1 (CD274), CD3
111 (CD3G), and CD8 (CD8A). Thus, our model focuses on predicting therapeutic response based on
112 altered genomic and proteomic patterns in the tumors of ICBT- naive patients. Our study is
113 consistent with the evidence-based understanding that the pre-treatment identification of patients
114 with predisposition of responsiveness or sensitivity to a given therapy is dependent, at least in
115 part, on the driving signaling pathways activated in cancer patients and the correct delineation of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

116 the active molecular components of that pathway (6 – 9); the knowledge of which generally
117 stems from established patient-specific expression portraits consisting of several genomic or
118 proteomic endpoints (9).

119 This study presents an approach that integrates several statistical and
120 computational algorithms to predict *in vivo* drug response, using models trained on
121 immunohistochemistry staining data. For model development, the approach was tested on
122 23 ICBT-naive clinical cases from the Taipei Medical University - Shuang Ho Hospital
123 (TMU-SHH), from which β-catenin, PDL1, CD3 and CD8 protein expression data and
124 sensitivity to anti-CTLA-4 and/or anti-PD1 drugs were comparatively evaluated. Our results
125 demonstrate the feasibility of creating a machine learning-based prediction model that captures
126 the variability in patients' drug response from basic human proteome data.

127 Transcending *in silico* prediction of drug response, our findings further beg the case for a
128 multidisciplinary approach in oncology clinics and provides a computational framework for
129 cancer patient stratification based on individualized molecular profiling, as our prediction model
130 is specifically directed towards eliciting therapeutic response, improving performance status, and
131 minimizing drug-related toxicities in cancer patients. This we believe is potentially vital for
132 personalized medicine by coupling the molecular phenotype of patients to drug sensitivity.

133

134

135

136

137

138

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

139 **Materials and Methods**

140 ***Study population***

141 All patients donated their tissue to TMU-SHH for further study in accordance to the TMU-SHH
142 research ethics standard. The study was approved by the Joint Institutional Review Board of
143 Taipei Medical University (TMU-JIRB No.: N201802036). In this retrospective study, we
144 collected 23 pairs of matched primary or metastatic solid cancer tissue. The median patient age
145 was 61± 8.83 years. Patients' tissue samples were obtained prior to initiation of ICBT, and
146 evaluation of CTNNB1, CD274, CD3G and CD8A protein expression was carried out by
147 immunohistochemistry. The patients either received pembrolizumab (anti-PD-1), nivolumab
148 (anti-PD-1) and/or ipilimumab (anti CTLA-4), as single agent or combinatorial therapy.
149 Correlative analyses of patients' proteome profile with therapeutic response were performed.
150 Tumor response to the ICBT was evaluated bi- weekly for the first 12 weeks and thereafter every
151 month, using standard imaging techniques such as X-ray, computed tomography (CT) or
152 magnetic resonance imaging (MRI). Initial clinical assessment was performed based on the
153 modified Response Evaluation Criteria in Solid Tumors (mRECIST) (18), before a final review
154 by an independent imaging review committee consisting of three radiologists and two
155 oncologists who used mRECIST and the immune-related Response Criteria (irRC) (19). The
156 independent imaging review committee was blinded to the therapeutic modalities and evaluated
157 disease progression by comparative analyses of the X-rays, CT or MRI images obtained before
158 enrollment and during the study. Drug-related toxicities were assessed according to the National
159 Cancer Institute Common Terminology Criteria for Adverse Events, version 4.03 (CTCAE
160 v.4.03).

161 ***Inclusion and exclusion criteria for consideration in study***

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

162 We considered eligible male and female adults aged ≥18 years with cytohistological- and/or
163 imaging-confirmed primary, metastatic, or recurrent solid malignancies of WHO grades I to IV,
164 based on relevant guidelines for diagnosis of such cancer, who had progressed or were intolerant
165 to one or more standard chemotherapy regimen, and from whom we had obtained informed
166 consent to procure tissue samples by biopsy or acquire existing formalin-fixed paraffin-
167 embedded tissues. Patients with history of, or active autoimmune disorder, positive hepatitis B/C
168 or HIV, uncontrolled or unstable cardiovascular disease, previous exposure to ICBT, or
169 anticancer therapy within 4 weeks before initiation of ICBT, were excluded.

170 ***Bioinformatics evaluation and validation of panel of selected biomarkers***

171 A systematic analyses of public clinical data was performed to evaluate the expression profile
172 and impact of CTNNB1, CD274, CD3G and CD8A proteins on survival endpoints, particularly
173 the relapse-free survival (RFS) and overall survival (OS) in TCGA and TARGET pan-cancer
174 (PANCAN) cohort (n = 18, 802). The PANCAN cohort consisted of normal solid tissues,
175 primary tumor, metastatic, recurrent, additional new, and additional metastatic solid tumors of
176 >12 histological origins, including breast carcinoma (BRCA), head and neck squamous
177 carcinoma (HNSC), glioblastoma (GBM), bladder carcinoma (BLCA), lung adenocarcinoma
178 (LUAD), lung squamous cell carcinoma (LUSC), colon adenocarcinoma (COAD), ovarian
179 carcinoma (OV), rectal adenocarcinoma (READ), kidney renal clear cell carcinoma (KIRC),
180 uterine cervical and endometrial carcinoma (UCEC), and lymphoblastic acute myeloid leukemia
181 (LAML).

182

183 ***Defining the differential expression of our genes-of-interest***

184 The TCGA and TARGET PANCAN cohort (n = 18,802) data was processed and re-analyzed to
185 determine the differential expression of CTNNB1, CD274, CD3G and CD8A in tumor samples

186 of different tissue origins. Specifically, for each sample, RNA-Seq by Expectation-Maximization
187 (RSEM) -quantified mRNA expression was normalized across all samples to generate the RSEM
188 normalized counts (RSEM norm_counts). Samples with gene expression $\log_{10}$RSEM
189 norm_counts at least 0.5 SD above the mean were defined as over-expressed. Those with
190 $\log_{10}$RSEM norm_counts at least 0.5 SD below the mean were defined as under-expressed. The
191 expression with $\log_{10}$RSEM norm_counts within 0.5 SD were defined as intermediate.

192

193 ***Immunohistochemical Staining***

194 ICBT-naïve tumor samples obtained from cancer patients enrolled for the study were fixed in
195 10% formalin and embedded in paraffin. This was followed by deparaffinization, graduated
196 rehydration and staining with hematoxylin and eosin (H&E). For immunohistochemical staining,
197 antigen retrieval and probing of the deparaffinized tissue blocks with primary anti-β-catenin
198 (1 : 100, #D10A8, Cell Signaling Technology, Inc., Danvers, MA, USA), anti-PDL1/CD274
199 (clone SP142) (1:50, #M4420, Spring Bioscience, Pleasanton, CA, USA), anti-CD3 (1 : 150,
200 #M4620, Spring Bioscience), anti-CD8 (1 : 20, #VP-C324, Vector Labs, Burlingame, CA,USA)
201 antibodies, or isotype IgG control, overnight at 4°C. This was followed by washing and
202 incubation of tissue blocks with biotinylated secondary antibody, then by horseradish peroxidase
203 (HRP)-conjugated streptavidin (LSAB® 1 system HRP kit, Agilent Dako, Santa Clara, CA, USA).
204 The slides were rinsed and the chromogen 3,3-diaminobenzidine (DAB) hydrochloride added for
205 color development. Finally, sections were rinsed in double-distilled water (ddH_2O),
206 counterstained with Mayer's hematoxylin, and mounted with DPX mounting medium (#06522;
207 Sigma-Aldrich Corporation, St. Louis, MO, USA) for evaluation. Photo-images were captured
208 with a Photometrics Scientific CoolSnap cf CCD camera (Nikon, Lewisville, TX, USA).

209

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

210 ***Unsupervised machine learning techniques used in this study.***

211 For ICBT effect prediction, we used a data cluster of 5 variables, namely 4 proteome expression
212 variables and 1 drug response variable labeled high, low, or null. The 4 proteome expression
213 profile variables served as predictors, while the drug response was our predicted variable. The R-
214 project software Random forest (RF) module (20) was used for preliminary classification and
215 regression. The RF module generalizes classification tree algorithm by ensembling several
216 decision trees instead of growing a single decision tree. Each decision tree was trained on
217 observation subsets, and features of the learning set were derived from a bootstrap sample based
218 on the original learning set, while data was bucketed into smaller sets called tree nodes by using
219 a random subset of the whole variables set. For the prediction of specific outcomes from our
220 vector of input data (X_0, X_1, ..., X_n), the input data were applied to each single tree in the forest
221 to generate one output value Y_{Ti} per tree. The final prediction Y was derived by combining all
222 Y_{Ti}, calculating the mean, the weighted average or the median of all $Y_{Ti.}$ in RF, since
223 approximately 30% of the bootstrap sample cases were excluded ("out-of-bag") from the tree-
224 growing process, the new instance classification decision was by majority voting over all trees.
225 These ''out-of-bag'' cases were used for unbiased evaluation of the algorithm performance and
226 test prediction accuracy (21). The Instance-based learning (IB1) or collaborative filtering
227 technique commonly referred to as the nearest neighbor (NNge) technique performed with the
228 Waikato Environment for Knowledge Analysis (WEKA v.3.9.1) software package (22, 23) was
229 used to create optimal classifiers for the ICBT responses. The IB1 is a basic instance-based
230 machine learning module which finds training sets/instances that are nearest to a given test
231 instance by using the normalized euclidean distance, and consequently predict same class as the
232 training set. The IB1 is similar to the k-nearest neighbors (KNN, IBK), however, for IB1, k=1.
233 The IB1 has a characteristically fast training stage and uses the entire training set in the
234 prediction stage. For classification of new instances, training sets which are nearest in distance to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

235 the new instance are generated, and the new instance is classified based on majority voting
236 among the nearest neighbors (23). NNge is a non-nested generalized examplars classifier based
237 on hyperrectangles splitting procedure, pruning of non-generalized examplars and presentation
238 order of training instances. Unlike the IB1, NNge is rule-based (24). We performed a 10-fold
239 cross-validation for all classifiers and regression analyses to evaluate the prediction performance
240 of the machine learning models, with all the instances being randomly assigned to one of 10 sets.
241 For each cross-validation, 1 set was used for testing and the other 9 used for machine training. At
242 the end of validation, the mean error across the 10 tests was estimated (23). Unsupervised
243 hierarchical clustering was performed using online free statistics and forecasting softwares.
244 Distances were computed based on correlation, the distance type was Euclidean, the clustering
245 method was group average unweighted pair-group, and scale type was standard deviation.

246

247 ***Statistical Analyses***

248 All statistical analyses were performed using the Statistical Package for the Social Sciences
249 (SPSS; IBM Corp., Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk,
250 NY: IBM Corp.). Paired categorical variables were assessed using the Pearson's $\chi 2$ test. Survival
251 analyses were performed by log-rank tests, using the Kaplan-Meier plots. Survival variables
252 were analyzed using the univariate and multivariate Cox proportional hazard model. All tests
253 were 2-sided and p-value < 0.05 was considered statistically significant.

254

255

256

257

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

258 **Results**

259

260 ***CTNNB1, CD274, CD3G and CD8A are functionally associated and are cancer progression-***
261 ***relevant***

262 Since the interaction between functionally associated or related genes is often reflected in
263 biological networks, such that each interacting component gene called a module, constitute part
264 of the network motifs, including positive/ negative feedback, auto-regulatory, directed bow-tie,
265 bi-fan, or feed-forward loops, using the Cytoscape KEGG pathway analysis (http://cytoscape.org)
266 and STRING version 10.5 platform (https://string-db.org), we established probable interactions
267 between CTNNB1, CD274, CD3G and CD8A, as depicted by the mostly bi-fan motif-motif
268 interaction between the selected modules in the T cell receptor (TCR) and primary
269 immunodeficiency signaling pathways, principally reflecting T cell inflamed and non-T cell-
270 inflamed TME (25) (**Figure 1A and B**). This network consisting of CTNNB1, CD274, CD3G
271 and CD8A suggest that inter-molecular activities between these genes play a critical role in the
272 modulation of vital biological events or responses, such as T cell activity, immune status and
273 immunodeficiency. To establish the clinical relevance of this gene-set to tumor growth and or
274 progression, combining data from The Cancer Genome Atlas (TCGA) and Therapeutically
275 Applicable Research To Generate Effective Treatments (TARGET) databases, we analyzed the
276 pan-cancer (PANCAN) data (n = 18802), on the UCSC Cancer Browser portal (26). Analyses of
277 TCGA+TARGET PANCAN data revealed significantly enhanced expression of CTNNB1 in
278 cancer samples compared to CD274, CD3G and CD8A (**Figure 1C and D**). This data while
279 corroborating, at least in part, interaction between the four selected genes, does indicate a
280 converse functional relationship between CTNNB1 and the others. In furtherance to their
281 suggested role in immune status modulation, we sought to understand how variations in this
282 panel of genes affect systemic response to disease progression. To do this, we accessed and re-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

283 analyzed the GSE19234, GPL570 HG-U133_Plus_2 Affymetrix Human Genome U133 Plus 2.0
284 Array data from Gene Expression Omnibus (GEO) originally on gene expression profiling,
285 mitotic index (MI), and quantification of tumor infiltrating leukocytes (TILs) and CD3+ cells in
286 metastatic lesions as a molecular basis to develop improved methods for predicting patient
287 survival (27). Expression heat map generated from our statistical analyses confirmed an inverse
288 correlation between the expression of CTNNB1 and CD274, CD3G or CD8A. While sex had no
289 apparent influence on the expression profile of the 4 selected genes, CTNNB1 increased with
290 disease progression from IIIA to IV, with concomitant decrease in CD274, CD3G and CD8A
291 expression levels (**Figure 1E**). These data do indicate that CTNBB1, CD274, CD3G and CD8A
292 are functionally associated and their expression and/or activities are associated with cancer
293 progression.

294

295 ***CTNNB1, CD274, CD3G and CD8A mediate human immune cells activity and anticancer***
296 ***therapeutic response***

297 Having established that CTNNB1, CD274, CD3G and CD8A are functionally associated
298 and are cancer progression-relevant, we investigated if and how this 4-gene panel affects or
299 influences sensitivity to anticancer therapy. First, we analyzed the tumor expression data from
300 the neoadjuvant trial of cisplatin monotherapy in triple negative breast cancer patients [A-AFFY-
301 44, AFFY_HG_U133_PLUS_2, E-GEOD-18864] consisting of 84 samples and 54675 genes.
302 Generated heatmap indicate strong association between up-regulated CD3G, CD8A or CD274
303 expression, and higher Miller-Payne grade (MPG ≥3) for histological response to chemotherapy,
304 while conversely, high CTNNB1 expression was mostly associated with MPG ≤ 2 (**Figure 2A**).
305 In addition, using the AFFY_HG_U133_PLUS_2, E-GEOD-28702 dataset for 83 CRC patients
306 subjected to FOLFOX therapy, we observed a positive correlation between enhanced CD3G,
307 CD8A or CD274 expression levels and response to therapy, while high CTNNB1 expression

308 strongly correlated with lack of response (**Figure 2B**). Since the pathological complete response
309 (pCR) is a indicator of shrinking tumors and a measure of response to neoadjuvant chemotherapy
310 (28), we further reanalyzed the GSE4779, GPL1352 HG-U133_X3P Affymetrix Human X3P
311 Array data (n = 102) which was originally intended to elucidate the relationship between tumor-
312 host interactions and the efficacy of chemotherapy (28). Our heatmap showed that most patients
313 with concomitantly high expression levels of CD3G, CD8A, CD274 and low CTNNB1 were
314 associated with pCR, while none with enhanced expression of CTNNB with low CD3G, CD8A,
315 and CD274 achieved pCR; additionally, tumor size ≤ T2 and/or node involvement N0 positively
316 correlated with the overexpression of CD3G, CD8A and CD274 with low CTNNB1, while
317 CTNNB1 expression was associated with tumor size >T2, and node involvement of N1-N2
318 (**Figure 2C**). Furthermore, using the A-AFFY-33, AFFY_HG_U133A, E-GEOD-25055 dataset
319 for discovery of genomic predictor of response and survival following neoadjuvant taxane-
320 anthracycline chemotherapy in breast cancer (n = 310), we evaluated the probable correlation
321 between pathological response in the form of residual cancer burden (RCB), distant relapse-free
322 survival (DRFS), chemosensitivity, Diagonal Linear Discriminant Analysis (DLDA30) 30-gene
323 pCR predictor, and the differential expression of CTNNB1, CD3G and CD8A. Our resultant
324 heatmap demonstrates a strong positive relationship between high CD3G, high CD8A, pCR,
325 presence of DRFS, chemosensitivity, positive DLDA30 and RCB index ≤ 1; Conversely,
326 CTNNB1 expression positively correlated with chemo-insensitivity, residual disease (RD), RCB
327 > 1, negative DLDA30 and DRFS = 0 (**Figure 2D**). These data, regardless of tumor histological
328 origin or tissue type, highlight the potential role of CD3G, CD8A, CD274, and CTNNB1 as
329 indicators of activated immune status and modulators of therapeutic response in cancer patients,
330 and do suggest a predictive role for them in anticancer immune response.

331

332 ***Computational evaluation and validation of CTNNB1, CD3G, CD8A, and CD274 gene-set as***
333 ***prediction model for immune checkpoint blockage therapy response***

334 Consistent with our objective of using patients' baseline protein expression profile to
335 predict their response to ICBT, we established a procedural/approach protocol, an overview of
336 which is shown in **Figure 3A**. For preliminary tests, we assessed internally generated
337 immunohistochemistry data (**Figure 3B**) by using several known machines learning algorithms,
338 including the random forests (RF), correlation matrix (Cor Matrix) and principal component
339 analyses (PCA). All proteomic classifier components of our algorithms performed consistently
340 well with high computational efficiency, which is necessary for unbiased cross-validation
341 analyses. Our correlation matrix data show the correlation coefficients between the proteomic
342 variables, thus further corroborating the strong inverse relationship between CTNNB1
343 expression and that of CD3G ($r = -0.446$, $p = 0.033$), CD8A ($r = -0.514$, $p = 0.012$) and
344 CD3G/CD8A ratio ($r = -0.096$, $p = 0.662$), as well as a mild positive association with CD274 ($r =$
345 0.028, $p = 0.898$) (**Figure 3C and Supplementary Table S1**). Further, we used the PCA, a data
346 compression and dimension-reduction procedure, to transform our original set of proteomic
347 variables into small subgroups of linear combinations representative of most of the variances in
348 the original set, thus revealing the inherent ability of individual patient protein profile to provide
349 relevant personalized cancer biology information. As illustrated in **Figure 3D**, the expression
350 profile of the selected panel of proteins recapitulates the intrinsic propensity to respond or not to
351 ICBT, regardless of tumor tissue of origin, histological subtype and stage, when the digitally-
352 quantified protein expression levels were plotted on the first two principal components (PCs) of a
353 proteome profile matrix, where components 1and 2 are CTNNB1 and CD274, respectively. This
354 data is suggestive of the feasibility of CTNNB1, CD274, CD3G and CD8A protein expression
355 profile to serve as surrogate for unmeasured proteomic and non-proteomic phenotypes, thus,
356 providing additional support for this approach.

The expression levels of CTNNB1, CD274, CD3G and CD8A proteins before initiation of therapy accurately predict immune checkpoint blockade therapy response in a pan-cancer cohort

Having selected a gene-set based on computational analyses, we sought to correlate their expression profile with ICBT response in our exploratory pan-cancer cohort and identify the best gene-set-based classifiers with highest intrinsic therapeutic response predictive accuracy. Thus, we investigated the feasibility of predicting ICBT response of each patient based on the patient's protein expression data exclusively. Firstly, we generated an intensity-based graphical visualization of patients' CTNNB1, CD274, CD3G and CD8A protein expression values, and the relationship between the protein variables, before initiation of ICBT, in an expression matrix. Our data was corroborative of the inverse correlation between the tumor enhancer CTNNB1, and the immune-based anticancer network proteins CD274, CD3G and CD8A (**Figure 4A**). To understand how each protein component of the selected gene-set contribute to patient phenotype and therapeutic response, we constructed a loadings plots showing how the CTNNB1, CD274, CD3G and CD8A protein variables contribute to creating each principal component (PC) in the prediction model, where PC1, PC2, PC3, PC4 and PC5 were CTNNB1, CD274, CD3G, CD8A and CD3G/CD8A, respectively (**Figure 4B and Supplementary Figure S1**). This data suggest that the behavior of each protein variable is dependent on the dominant protein signal in the model. Thus, despite being based on the presumption that the prediction model was a one-component model, the loadings plot revealed that all protein components contributed in varying degrees to creating the model, and thus, are all relevant in understanding the total phenotypic and therapeutic variability in the cohort. Having established that all protein variables were relevant in the prediction model, we then generated cluster structures from patients' proteomic data which recapitulate constitutive propensity to respond to ICBT by projecting the multivariate proteomic data into a two dimension (2-D) projection plot using unsupervised K-means clustering

382 algorithm for visual appreciation of how the proteomic data stratifies the cohort (n = 23) into
383 binary (**Figure 4 C**) or tripartite (**Figure 4 D**) ICBT response clusters. The performance of this
384 classification algorithm was tested by applying the classifier to a 10-fold cross-validation. Next
385 we performed a neural network predictive modeling using the random hold-back validation
386 method to estimate the coefficient of multiple determinations, R^2, which is indicative of how the
387 models explain the response variability of the cohort. As shown in **Table 2**, the generalized R^2
388 was 0.86 and 0.98 for the training and validation sets, respectively. Per proteomic variable,
389 against CTNNB1, CD8A had the best R^2 of 0.63 and 0.90 for the training and validation sets,
390 respectively; as well as the lowest random mean square of errors (RMSE) of 2.88 and 1 for the
391 training and validation sets, respectively. This data is suggestive of the constitutive potential of
392 variations in CD274, CD3G, and most significantly, CD8A data to explain the variability in the
393 CTNNB1 response data. Therefore, to better understand and characterize the clusters of patients
394 formed in Figures 4C and D, we used the Youden plot which was originally intended for
395 assessing inter-group biases and intra-group variation problems in the context of inter-group
396 experimentation. We adapted the Youden plot to graphically stratify patient's response to ICBT
397 based on CTNNB1 and CD8A protein expression. The patients in the upper left quadrant, lower
398 left quadrant, and lower right quadrant, were classified as '*high responders*', '*low responders*',
399 and '*null responders*'. The upper right quadrant contained ambivalent responders, with those
400 above the 45° degree diagonal line classed as '*high responders*', while those below were classed
401 '*low responders*' (**Figure 4E**). To err on the side of caution, analyses of the quality of this
402 prediction using the Matthews correlation coefficient (MCC), our binary classification
403 (*responder*, *non-responder*) model achieved a MCC of 1, using the formula:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

$$\mathrm{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}\text{, where:}$$

- TP = true positives
- FP = false positives
- TN = true negatives
- FN = false negatives

404

405 and a Kruskal-Shepard non-metric multidimensional scaling (nMDS) analysis of the tripartite 406 (*high*-, *low*-, and *null-response*) model showed that almost all the points in our 407 obtained/observed rank and the predicted/target rank fall along a straight line, with R^2 on axis 1, 408 2 and 3 being 0.614, 0.539 and 0, respectively, while the stress was 0.028, suggesting an almost 409 perfect analysis (**Figure 4F**).

410

411 ***A patient-specific, evidence-based mathematical predictive model of drug response based on*** 412 ***patients' CTNNB1, CD274, CD3G and CD8A protein profile***

413 In parallel experiments, we examined the feasibility of developing a patient-specific, evidence-414 based mathematical prediction algorithm of ICBT response based on our selected gene-set. Thus, 415 based on accrued exploratory cohort data (n = 12), we determined that our control variable was 416 the malignancy itself, while our controllable variables (x) were CTNNB1, PD274, CD3G, CD8A 417 and CD3G/CD8A, and denoting them with x_1, x_2, x_3, x_4, x_5, respectively, such that:

418
$$x = \text{Decision variable} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = (x_1, x_2, x_3, x_4, x_5)^{\mathrm{T}}$$

419 We then sought to develop a mathematical formula f() that satisfies the equation:

420 $y = f(x) = f(x_1, x_2, x_3, x_4, x_5)$, which translates into,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

421 $$(\text{Predict Response}) = f\left[\text{CTNNB1}, \text{CD274}, \text{CD3G}, \text{CD8A}, (\text{CD3G/CD8A})\right],$$

422 such that the predicted response to ICBT is a function or dependent on variation in CTNNB1,
423 CD274, CD3G and CD8A protein profile, as well as the immune score, CD3G/CD8A.

424 To generate the predictive mathematical model, we decided on the use of basic (constant, input
425 variable, addition, subtraction, multiplication, division), trigonometry (sine, cosine), and
426 exponential (exponential, natural logarithm, power) as the formula building blocks; set
427 significant level (α) was 0.05, and a 30% test set size used to detect curve over-fitting.

428 Using the Eur curve-fitting software, we finally derived and selected the mathematical predictive
429 model:

430 $$\text{Predict Response} = a + b*CD8A + -c/CTNNB1 + d*CD274^2 + -e/CD274^2 - f*CD274*CD8A - g*CD8A^4 - h*CTNNB1^2$$

431

432 Where, *a, b, c, d, e, f, g*, and *h* are values known to us.

433 In generating this mathematical model, CD8A, CD274, CTNNB1, CD3G and CD3G/CD8A
434 appeared in 15, 11, 10, 4 and 2 models, respectively, while the 'Predict Response' appeared 19
435 models (**Figure 5A**). In parallel analyses, across all models, there were 38, 35, 10, 5, and 2
436 occurrences of CD8A, CD274, CTNNB1, CD3G and CD3G/CD8A, respectively; while the
437 'Predict Response' occurred 19 times (**Figure 5B**). We propound that this data delineates the
438 decision vectors/variables whose interaction and variation, actually modulate the patients'
439 pathological phenotype and potential therapeutic response, namely, CD8A, CD274, and
440 CTNNB1. This notion corroborates earlier data from the Youden plot in Figure 4E, and results of
441 our multivariate regression analysis of the cohort proteome profile which automatically selected
442 CD8A and CD274 as independent variables, while ignoring CD3G and CD3G/CD8A in the
443 model, when CTNNB1 was designated as our dependent variable (**Supplementary Table S1**).

444 To confirm the relative impact of each decision variable (x), namely, CD8A, CD274, and
445 CTNNB1 on the target variable (y), namely 'Predict Response', we carried out the variable
446 sensitivity analyses and demonstrated that CD8A exhibited a sensitivity of 1.26, 44% positive, a
447 positive magnitude of 0.67, 56% negative and negative magnitude of 1.73; for CD274,
448 sensitivity was 1.15 and 100% positive with a positive magnitude of 1.15, while CTNNB1
449 exhibited a sensitivity of 0.32, 100% negative and a negative magnitude of 0.32 (**Figure 5C**),
450 based on a model equation of the form $y = f(x_1, x_2....)$, and defining the influence metrics of x on
451 y as follows: *Sensitivity* $= \overline{\left|\frac{\partial y}{\partial x}\right|} . \frac{\sigma(x)}{\sigma(y)}$ was evaluated at all input data points. The proportion of
452 data points where $\frac{\partial y}{\partial x} > 0$ or $\frac{\partial y}{\partial x} < 0$ was considered as the *percentage positive* or *percentage*
453 *negative*, respectively. The *positive magnitude* or *negative magnitude* was estimated as
454 $\overline{\left|\frac{\partial y}{\partial x}\right|} . \frac{\sigma(x)}{\sigma(y)}$ at all points where $\frac{\partial y}{\partial x} > 0$ or $\frac{\partial y}{\partial x} < 0$, respectively. $\frac{\partial y}{\partial x}$ is the partial derivative of y with
455 respect to x; $\sigma(x)$ and $\sigma(y)$ are the standard deviation of x and y in the input data; $|x|$ represents
456 the absolute value of x and $\bar{x}$ denotes the mean of x.

457 The complexity of our mathematical ICBT response predictive model pales in the face of its
458 significantly high accuracy, as demonstrated by the Pareto error/complexity analysis (**Figure 5D**),
459 and the $Observed/\Pr edicted \approx 1$ plot which shows that the actual observed input values used in
460 the training and validation data line up in a 1:1 ratio against the values predicted by the selected
461 mathematical prediction model (**Figure 5E**). This predictive accuracy was further corroborated
462 by the results of our statistical analyses of the quality of our novel mathematical patient-specific
463 prediction model which revealed that the area under the receiver operating characteristics (ROC)
464 curve (AUC) for our CTNNB1-CD274-CD3G-CD8A gene-set as predictor of response to ICBT
465 was 1 (95% CI 0.94-0.99, SE = 0.00001, p = 0.0001) (**Figure 5F**). Our algorithm exhibits a R^2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

466 goodness of fit and correlation coefficient of ~1, a maximum error (ME) of 0.011, mean squared
467 error (MSE) of 1.34 x 10^{-5}, and a mean absolute error (MAE) of 0.002 (**Supplementary Table**
468 **S2**).

469 Finally, in a double-blind trial, as we remained blind to patients' clinicopathological
470 characteristics, we tested the functionality and clinical accuracy of the generated prediction
471 algorithm. As a demonstration of the predictive accuracy and clinical applicability of our model,
472 we now present 3 clinical cases validating our prediction model.

473 Patient 4, a 56 years old male cigarette smoker presented with persistent lower back pain (L4 to
474 S2) of about a month; A definitive diagnosis of right upper left lung adenocarcinoma, with
475 multiple lung, bone, and brain metastases, cT2N3M1b, clinical stage IV, EGFR L858R(+) was
476 established, with an Eastern Cooperative Oncology Group performance status (ECOG PS) of 1.
477 He was started on Tarceva® (erlotinib), and the addition of an immune checkpoint inhibitor as
478 an adjuvant was considered. Digital quantification of his CTNNB1, CD274, CD3G and CD8A
479 immunohistochemistry profile were 36.3%, 4.5%, 2.1%, and 2.0%, respectively. Based on the
480 mathematical model, we predicted 'null response' *ab initio;* Computed tomography (CT)
481 imaging revealed that the tumor was non-respondent to the therapy, even after 5 cycles of
482 Keytruda® (pembrolizumab) (**Figure 6A**).

483 Patient 6, a 75 years old male with a history of nasopharygeal carcinoma, presented at our
484 facility with *status post*- combined chemoradiotherapy (CCRT) in relapse with touch bleeding
485 over the growing tumor, and would not consent to further chemotherapy, citing no appreciable
486 effect. ECOG PS = 1. Immune checkpoint blockade was being considered. Using our
487 mathematical model, we predicted patient as a 'high responder'. After medical counseling on
488 therapeutic options, patient gave informed consent and was commenced on Opdivo® (nivolumab)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

489 3mg/kg q3w. After 4 cycles of ICBT, patient's status had improved with good clinical and image
490 response (**Figure 6B**).

491 Patient 9, a 57-year-old male with documented metastatic urothelial cancer in liver *status post*-
492 cisplatin, ECOG PS 2-3, refused further chemotherapy for lack of perceived effect. Nivolumab
493 was being considered. Using our mathematical model, we predicted 'null response'. Patient gave
494 informed consent to be enlisted in the clinical trial after medical counseling on therapeutic
495 options and was commenced on nivolumab 3mg/kg q2w. After 6 cycles, CT imaging revealed no
496 response to the ICBT, and patient was clinically assessed to be in disease progression (**Figure**
497 **6C**).

498 Together, these clinical cases highlight the feasibility of predicting therapeutic response based on
499 patients' proteome profile, the predictive accuracy of our model, and the clinical applicability of
500 our novel mathematical patient-specific ICBT response prediction algorithm, begging its case for
501 exploration in a larger clinical cohort, thus, adding one more arsenal to the
502 precision/personalized anticancer medicine armory.

503

504

505

506

507

508

509

510

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

511 **Discussion**

512 We tested the hypotheses and demonstrated that baseline genomic alterations in tumors
513 drive disease course, influence tumor susceptibility or insensitivity to therapy, and inherently
514 predict the response of cancer patients to anticancer therapy, particularly immune checkpoint
515 blockade therapy. In the present study, proposing a novel machine learning algorithmic model
516 for biomarkers identification, classifiers construction and patient stratification, we demonstrated
517 that the expression of CTNNB1 is inversely correlated with CD274, CD3G and CD8A
518 expression, and the 4 -gene panel is cancer progression-relevant (**Figure 1**). This is consistent
519 with evidence provided by Stefani Spranger, *et al.*, implicating activated Wnt/β-catenin signaling
520 in the exclusion of a T cell gene expression signature in human metastatic melanoma (16);
521 Analyzing gene expression profile from human melanoma tissue microarray, they suggested the
522 categorization of tumors into "non-T cell-inflamed" or "T cell-inflamed" based on the alterations
523 in genomic constitution of the TME, where the former is largely mediated by active β-catenin
524 signaling. In addition, it was recently demonstrated that epithelial–to-mesenchymal transition
525 (EMT) enriches CD274 in cancer stem cells (CSCs) by a proposed β-catenin/STT3/PD-L1
526 signaling axis, wherein EMT induces N-glycosyltransferase STT3 in a CTNNB-dependent
527 manner, and subsequently stabilizes and up-regulates CD274 (30). Recently, CD274 expression
528 has been correlated with immune cells and immune activity signatures in tumors but was found
529 to be insufficient for response to MAPK inhibition in melanoma cells, as many pretreatments and
530 progressing melanomas showed both CD274 positivity and immune activation signatures (31).
531 However, we demonstrated herein that regardless of tumor histological origin or tissue type,
532 CTNNB1, CD274, CD3G and CD8A mediate human immune cells activity and anticancer
533 therapeutic response, thus highlighting the potential role of CD3G, CD8A, CD274, and
534 CTNNB1 as indicators of activated immune status in cancer patients, and suggesting a predictive
535 role for them in anticancer immune response. In current oncology practice, chemotherapy is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

536 mostly empirical and often not beneficial all cancer patients, illustrating the necessity of a rather
537 more personalized or precise approach in anticancer therapy. It is thus, clinically desirable and
538 therapeutically relevant to be able to predict the likelihood of each individual patient to benefit
539 from a specific therapy. Consistent with this, we find it clinically relevant that a strong positive
540 relationship exist between the enhanced expression of CD3G, CD8A, and CD274, pCR,
541 chemosensitivity, DRFS and DLDA30 positivity and RCB index ≤ 1, while CTNNB1 expression
542 was associated with chemo-insensitivity, residual disease (RD), RCB > 1, negative DLDA30 and
543 DRFS (**Figure 2**). Our data supports the proposition that the CTNNB1-CD274-CD3G-CD8A
544 gene-set is, in part, a surrogate biomarker for pCR which defines patients without residual
545 invasive carcinoma (pTO), absence of nodal metastasis, minimal residual cellularity, with no
546 residual *in situ* carcinoma, and akin to pCR, is adoptable as primary endpoint indicator in pre-
547 operative (neoadjuvant) therapy trials because of its inferential association with long-term
548 survival (**Supplementary Figure S2**). While survival or DRFS can be used as a metric in
549 developing predictors of therapeutic outcomes, surrogate markers such as RCB index which is
550 used to quantify residual disease ranging from pathological complete response to extensive
551 residual disease, after neoadjuvant chemotherapy, based on pathohistological variables including
552 number of involved nodes, size of the largest nodal metastasis and size and percent cellularity of
553 the primary tumor bed (32); and the DLDA30 which is a signature of 30-probe set based on 27
554 genes, selected after comparing many other predictors, to predict pathological response to
555 preoperative chemotherapy (33), also exhibit strong association with our CTNNB1-CD274-
556 CD3G-CD8A gene-set.

557 Furthermore, using machine learning algorithms, we computationally validated the utility
558 of the CTNNB1-CD274-CD3G-CD8A gene-set as prediction model for ICBT response, and as
559 surrogate for unmeasured proteomic and non-proteomic phenotypes (**Figure 3**), thus, providing
560 additional support for our algorithmic approach which determined and constructed proteome-

 doi:10.20944/preprints201908.0131.v1

561 based signature functions that are specific to T cell receptor and primary immunodeficiency
562 pathways in a pan-cancer cohort. Our proteome-based signature functions were fit using training
563 datasets made up of selected proteomic profiles of tumors with respect to nominated immune
564 pathways and existent T cell-inflamed or non-T cell-inflamed nature of patient's tumor. Based on
565 the value of the signature function, tumors with CTNNB1-driven alterations in the cohort were
566 separated from those bearing tumors with CD8A-CD3G-CD274-dominant signaling. In addition,
567 we demonstrated that the expression levels of CTNNB1, CD274, CD3G and CD8A proteins
568 before initiation of therapy accurately predict immune checkpoint blockade therapy response in a
569 pan-cancer cohort (**Figure 4**). Our findings are corroborated by accruing evidence that clinical
570 response to ICBT is elicited exclusively in patients with preexisting $CD3^+$/$CD8^+$ T cell infiltrates
571 in tumor beds with up-regulated CD274, especially as ICBT enhances the intratumoral
572 (re)activation, proliferation and expansion of $CD8^+$ T cells, resulting in tumor regression (15, 25).
573 Furthermore, gene expression analysis revealed that a large percentage of non-T cell-inflamed
574 tumors were characterized by activation of molecular components of WNT/β-catenin pathway,
575 including β-catenin, which are inversely correlated with the expression of CD8α in the tumor,
576 while PD-L1 expression positively correlated with CD8α (15, 16).

577 Having tested our model in pan-cancer cohorts computationally and correctly predicting
578 all 23 clinical cases, it would not be overstatement that our novel 4-gene prediction signature
579 reveals proteomic driver alterations which make it possible to accurately predict clinical and
580 image response to ICBT with AUC of 1, R^2 of ~1, ME of 0.011, MSE of 1.34×10^{-5}, and a MAE
581 of 0.002 (**Figures 5 and 6; Supplementary Table S2**). Our prediction model is easily
582 implemented in clinical practice given it uses a weighted sum of gene expression levels and a
583 simple multi-classification scale.

584 An important feature of the proposed algorithm is that the proteomic classifiers of this 4-
585 gene prediction signature enable the delineation of tumors based on their susceptibility to PD-1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

586 and/or CTLA-4 blockade, as well as demonstrate the clinical feasibility of stratifying cancer
587 patients in ICBT response groups based on the activation status and predominant signaling(s) in
588 the tumor microenvironment. However, this study has a few limitations, including the relatively
589 small cohort size and its single-center nature, thus the need for its validation in a large cohort,
590 multi-institution study.

591 In summary, the present study identified a 4-gene transcriptoproteomic profile for
592 predicting sensitivity to ICBT in cancer patients who were refractory to standard
593 chemotherapeutic agents from a single center using machine learning methodologies which
594 included the construction of optimal classifiers consisting of 4 protein predictors. Our proteome-
595 based predictor of response to ICBT is specifically directed towards eliciting therapeutic
596 response, improving performance status and minimizing toxicity in cancer patients, and had an
597 overall accuracy of 95.7% ($n = 23$, $p < 0.0001$). Finally, this 4-gene prediction algorithm is
598 individual patient-specific, tissue origin-independent, cost- and time- efficient, highly precise
599 and evidence-based.

600

601

602

603

604

605

606

607

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

608 **Acknowledgments**

609 This work was supported by Ministry of Science & Technology, Taiwan: Tsu-Yi Chao (MOST-
610 2019-001), and Wei-Hwa Lee (MOST 2019-002). This study was also supported by grants from
611 Taipei Medical University (TMU-SHH-001) to Tsu-Yi Chao and grants from Taipei Medical
612 University (TMU-SHH-002) to Chi-Tai Yeh.

613 **Authors' contribution**

614 Study conception and design: OAB, CTY, TYC. Data mining, machine learning and
615 mathematical modeling: OAB, OB. Data acquisition, collation and analyses: OAB, WHC, CTY,
616 KYL, TYC. Manuscript writing: OAB, WHC, CTY, TYC. Provision of reagents, experimental
617 infrastructure, and administrative oversight: WHL, KYL, CTY, LSW, TYC. All authors read and
618 approved the final version of the manuscript.

619 **Conflict of interest statement**

620 The authors declare that there are no potential conflicts of interest.

621

622 **Ethical approval**

623 All patients donated their tissue to TMU-SHH for further study in accordance to the TMU-SHH
624 research ethics standard. The study was approved by the Joint Institutional Review Board of
625 Taipei Medical University (TMU-JIRB No.: N201802036).

626

627

628

629

630

631

632

633 **References**

634 1. Niepel M, Hafner M, Pace EA, Chung M, Chai DH, Zhou L, et al. Profiles of Basal and
635 stimulated receptor signaling networks predict drug response in breast cancer lines. Sci
636 Signal. 2013;6(294):ra84. doi: 10.1126/scisignal.2004379.

637 2. Schmidt C. Cancer: reshaping the cancer clinic. Nature 2015;527(7576):S10–1.

638 3. Rubin MA. Health: make precision medicine work for cancer care. Nature
639 2015;520(7547):290–1.

640 **4.** Cully M. Advancing precision medicine in silico. Nat Rev Drug Discov 2015; 14 (5), 311.
641 doi:10.1038/nrd4619

642 **5.** Davidson NE, Armstrong SA, Coussens LM, Cruz-Correa MR, DeBerardinis RJ, Doroshow
643 JH, et al. AACR cancer progress report 2016. Clin Cancer Res 2016;22(Suppl 19):S1–134.

644 **6.** Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al: Personal omics profiling
645 reveals dynamic molecular and medical phenotypes. Cell. 2012, 148: 1293-1307.
646 doi:10.1016/j.cell.2012.02.009.

647 **7.** Schwab M, Schaeffeler E. Pharmacogenomics: a key component of personalized therapy.
648 Genome Med 2012;4:93. doi:10.1186/gm394

649 8. Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics:
650 strategies and challenges. Nat Rev Genet. 2013; 14(1): 23–34.

651 9. Wulfkuhle JD, Edmiston KH, Liotta LA, Petricoin EF 3rd. Technology insight:
652 pharmacoproteomics for cancer--promises of patient-tailored medicine using protein
653 microarrays. Nat Clin Pract Oncol. 2006;3(5):256-68.

654 10. Chambliss AB, Chan DW. Precision medicine: from pharmacogenomics to
655 pharmacoproteomics. Clin Proteomics 2016; 13(25)

656 11. Wei YN, Hu HY, Xie GC, Fu N, Ning ZB, Zeng R, Khaitovich P. Transcript and protein
657 expression decoupling reveals RNA binding proteins and miRNAs as potential modulators
658 of human aging. Genome Biology 2015; 16:41. doi: 10.1186/s13059-015-0608-2

659 12. Shi Z, Barna M. Translating the genome in time and space: specialized ribosomes, RNA
660 regulons, and RNA-binding proteins. Annu Rev Cell Dev Biol. 2015;31:31-54. doi:
661 10.1146/annurev-cellbio-100814-125346.

662 13. Spranger S. Mechanisms of tumor escape in the context of the T-cell-inflamed and the non-
663 T-cell-inflamed tumor microenvironment. Int Immunol 2016; 28 (8): 383-391.

664 14. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade
665 induces responses by inhibiting adaptive immune resistance. Nature. 2014;515(7528):568-71.

666 15. Spranger S, Sivan A, Corrales L, Gajewski TF. Tumor and Host Factors Controlling
667 Antitumor Immunity and Efficacy of Cancer Immunotherapy. Adv Immunol. 2016;130: 75–
668 93. doi:10.1016/bs.ai.2015.12.003.

669 16. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic β-catenin signalling prevents anti-
670 tumour immunity. Nature. 2015;523(7559):231-5. doi: 10.1038/nature14404

671 17. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging
672 Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome.
673 Clin Cancer Res. 2015 Jan 15; 21(2): 249–257.

674 18. Nishino M, Jagannathan JP, Ramaiya NH, Van den Abbeele AD. Revised RECIST guideline
675 version 1.1: What Oncologists want to know and what Radiologists need to know. Am J
676 Roentgenol 2010;195: 281-289. 10.2214/AJR.09.4110

677 19. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the
678 evaluation of immune therapy activity in solid tumors: immune-related response criteria.
679 Clin Cancer Res 2009;15(23):7412–7420.

680 20. http://www.r-project.org Accessed from 04 March – 29 June, 2017

681 21. Hayn D, Walch H, Stieg J, Kreiner K, Ebner H, Schreier G. Plausibility of Individual
682 Decisions from Random Forests in Clinical Predictive Modelling Applications. Stud Health
683 Technol Inform. 2017;236:328-335.

684 22. http://www.cs.waikato.ac.nz/ml/weka/ Accessed from 11 March – 30 June, 2017

685 23. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data
686 mining software: an update. SIGKDD Explor 2009; 11(1): 10-18

687 24. D Zaharie, L Perian, V Negru. A View Inside the Classification with Non-Nested
688 Generalized Exemplars. IADIS European Conference on Data Mining 2011; 24-26

689 25. Gajewski TF, Corrales L, Williams J, Horton B, Sivan A, Spranger S.
690 Cancer Immunotherapy Targets Based on Understanding the T Cell-Inflamed Versus Non-
691 T Cell-Inflamed Tumor Microenvironment. Adv Exp Med Biol. 2017;1036:19-31. doi:
692 10.1007/978-3-319-67577-0_2.

693 26. https://genome-cancer.ucsc.edu/ Accessed from 03 March – 29 June, 2017.

694 27. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19234 Accessed on 15-
695 16/05/2017

696 28. Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, et al. A stroma-
697 related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. Nat
698 Med. 2009; 15(1):68-74.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

699 29. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. Genomic predictor of
700 response and survival following taxane-anthracycline chemotherapy for invasive breast
701 cancer. JAMA 2011; 305(18):1873-81.

702 30. Hsu JM, Xia W, Hsu YH, Chan LC, Yu WH, Cha JH, et al. STT3-dependent PD-
703 L1 accumulation on cancer stem cells promotes immune evasion. Nat Commun.
704 2018;9(1):1908.

705 31. Kakavand H, Rawson RV, Pupo GM, Yang JYH, Menzies AM, Carlino MS, et al. PD-L1
706 Expression and Immune Escape in Melanoma Resistance to MAPK Inhibitors. Clin Cancer
707 Res. 2017;23(20):6054-6061.

708 32. Peintinger F, Sinn B, Hatzis C, Albarracin C, Downs-Kelly E, Morkowski J., et al.
709 Reproducibility of Residual Cancer Burden For Prognostic Assessment of Breast Cancer
710 After Neoadjuvant Chemotherapy. Mod Pathol 2015;28(7):913-920.

711 33. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, et al. Evaluation of a 30-
712 gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response
713 predictor in a multicenter randomized trial in breast cancer. Clin Cancer Res. 2010 Nov
714 1;16(21):5351-61.

715 34.

716

717

718

719

720

721

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

722 **Tables**

723

724 **Table 1.** Comparative correlation indexes for all pairs of the data series based on Pearson r,
725 Spearman rho and Kendall tau correlation, with *p*-values, as well as meta-analysis of the
726 correlation tests.

727

728 **Table 2.** Neural network predictive modeling table showing the data for the training and
729 validation sets based on the protein variables

730

731 **Table 3.** Novel patient-specific mathematical prediction algorithm accurately predicts ICBT
732 response in pan-cancer cohort (n = 23)

733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

756 **Figure Legends**

757

758 **Figure 1. CTNNB1, CD274, CD3G and CD8A are functionally associated and are cancer**
759 **progression relevant.** (A) Schematic illustration of ICBT response-based dysbalance in the
760 tumor microenvironment (B) Functionally associated network motif depicting k-means-clustered
761 molecular components of the KEGG T-cell receptor signaling and primary immunodeficiency
762 pathways, including CTNBB1, CD274, CD3G, and CD8A using the STRING version 10.5
763 platform. Green and red nodes represent the TCR signaling and immunodeficiency target genes,
764 respectively. (C) Visualization of the gene expression profiles and (D) box-and-whiskers plot
765 showing the global distribution of CTNNB1, CD274, CD3G and CD8A gene expression in
766 TCGA and TARGET PANCAN cohort, n=18802. (E) Heatmap based on GEO data, GSE19234,
767 GPL570, show the expression profile of CTNNB1, CD274, CD3G and CD8A in a cohort of
768 melanoma patients (n = 44).

769

770 **Figure 2. CTNNB1, CD274, CD3G and CD8A mediate anticancer therapeutic response.** (A)
771 Heatmap based on GEO data, E-GEOD-18864 showing the correlation between tumor grade,
772 Miller/Payne response, and the differential expression of CTNNB1, CD3G and CD8A from
773 neoadjuvant cisplatin monotherapy in a cohort of 84 triple negative breast cancer patients. (B)
774 Heatmap based on reanalysis of GEO data, E-GEOD-28702, showing the differential expression
775 of the CTNNB1, CD274, CD3G and CD8A in response to MFOLFOX6-based chemotherapy in
776 83 CRC samples. (C) pCR - npCR stratified heatmap based on GEO data, GSE4779, GPL1352,
777 showing the expression profile of the CTNNB1, CD274, CD3G and CD8A gene-set following
778 neoadjuvant anthracycline-based chemotherapy in a cohort of 102 breast cancer patients. (D)
779 Heatmap based on E-GEOD-25055 showing the correlation between pathological response,
780 disease relapse-free survival, chemosensitivity, DLDA30, and the differential expression of
781 CTNNB1, CD3G and CD8A in response to neoadjuvant anthracycline-based chemotherapy in a
782 cohort of 310 BRCA patients.

783

784 **Figure 3. Computational evaluation and validation of CTNNB1, CD3G, CD8A, and CD274**
785 **gene-set as prediction model for immune checkpoint blockage therapy response.** (A) A
786 schematic flow-chart of our prediction model for ICBT response. (B) A photomicrograph of
787 sample tissue slides for patients 3, 4 and 8 stained for CTNNB1, CD274, CD3G and CD8A by
788 immunohistochemistry and protein expression digitally quantified. (C) Multivariate correlation
789 plot for all pairs of data series based on the linear r (pearson's) correlation (*upper panel*), and a

790 scatter plot of intermolecular interaction, with p-value indicated (*lower panel*). (D) Principal
791 component analysis plot based on correlation matrices comparing disassociated protein
792 expression profiles in our clinical samples (n = 23) (*left panel*). Variance vectors show the
793 relative amount of variation each variable, namely, CTNNB1, CD274, CD3G and CD8A
794 contributes to the data set (*right panel*).
795
796
797 **Figure 4. The expression levels of CTNNB1, CD274, CD3G and CD8A proteins before**
798 **therapy initiation accurately predict immune checkpoint blockade therapy response in a**
799 **pan-cancer cohort.** (A) A matrix plot of the pre-ICBT CTNNB1, CD274, CD3G and CD8A
800 protein expression profile of our validation cohort (n = 23). (B) Representative cumulative
801 loadings plot showing how the original CTNNB1, CD274, CD3G and CD8A proteome variables
802 contribute to creating each principal component of our ICBT response prediction model.
803 Projection plot using an unsupervised K-means clustering algorithm, divides the patients' tissue
804 samples (n = 23) into (C) binary, then (D) tripartite ICBT response clusters based on the digitally
805 quantified protein expression attributes. (E) Youden plot adapted to graphically stratify patients'
806 response to ICBT based on CTNNB1 and CD8A protein expression. The inner rectangle
807 represents 2X standard deviation. Values within the rectangle are considered acceptable while
808 those outside are outliers. The 45-degree diagonal reference line through the Manhattan median
809 differentiates responders (above the diagonal) from non-responders (below the diagonal). (F) A
810 Shepard non-metric multidimensional scaling plot of all points on the obtained/observed rank
811 and the target/predicted rank.
812
813 **Figure 5. A patient-specific mathematical predictive model for ICBT response based on**
814 **patients' CTNNB1, CD274, CD3G and CD8A protein variables.** Bar charts of (A) the number
815 of models each variable appears in, and (B) the number of occurrences of each variable across all
816 models. (C) Variable sensitivity analyses chart showing the sensitivity, percentage positive,
817 positive magnitude, percentage negative and negative magnitude of the decision variables,
818 CD8A, CD274 and CTNNB1. (D) Pareto chart showing error/complexity ratio of the selected
819 prediction model. (E) Observed vs. Predicted plot showing how the actual observed values
820 within the training and validation data compare with values predicted by the prediction model.
821 (F) Receiver operating characteristics (ROC) curves for the CTNNB1-CD274-CD3G-CD8A
822 gene-set as predictor of response to ICBT. AUC was 1 (95% CI 0.94-0.99, SE = 0.02, p = 0.0001)
823

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2019 doi:10.20944/preprints201908.0131.v1

Figure 6. Our patient-specific, evidence-based ICBT response mathematical model predicts accurately and is clinically applicable. Computed tomography image showing (A) non-shrinkage of tumor after 5 cycles of pembrolizumab in lung cancer patient predicted to be a 'null response' case, (B) shrinking tumor size after 4 cycles of nivolumab in nasopharyngeal cancer patient predicted as a 'high response' case, and (C) quantitative and qualitative increase in metastatic nodules, after 6 cycles of nivolumab in patient with metastatic liver cancer, predicted as a 'null-response' case.

858 **Supplementary Data**

859

860 **Supplementary Table S1** Multiple regression analysis of cohort proteome profile show protein
861 factors relevant for the prediction model and their statistical significance

862

863 **Supplementary Table S2** Analyses of the quality of our novel patient specific mathematical
864 ICBT response prediction algorithm

865

866 **Supplementary Figure S1.** Representative loadings plot showing how the original CTNNB1,
867 CD274, CD3G and CD8A proteome variables contribute to creating each principal component of
868 our ICBT response prediction model

869

870 **Supplementary Figure S2.** Association of CTNNB1, CD274, CD3G and CD8A with long-term
871 survival of cancer patients. Kaplan-Meier plots showing the correlation between overall survival
872 and (A) CTNNB1, (B) CD274, (C) CD3G and (D) CD8A expression in the TCGA PANCAN12
873 cohort, n=5,158.

874