

1                   **Applying GAN-based data augmentation to improve transcriptome-based**  
2   **prognostication in breast cancer**

3  
4                   Working title: Data augmentation for improved breast cancer prognostication

5  
6   Cristiano Guttà<sup>1, \*</sup>, Christoph Morhard<sup>2</sup>, Markus Rehm<sup>1, 3, \*</sup>

7  
8                   1 Institute of Cell Biology and Immunology, University of Stuttgart, Allmandring 31, 70569  
9                   Stuttgart, Germany

10                   2 ProKanDo GmbH, Elfriede-Breitenbach-Straße 38a, 71640 Ludwigsburg, Germany

11                   3 Stuttgart Research Center Systems Biology, University of Stuttgart, Nobelstrasse 15, 70569  
12                   Stuttgart, Germany

13  
14                   **\*To whom correspondence should be addressed:**

Prof Dr Markus Rehm	Cristiano Guttà
Institute of Cell Biology and Immunology	Institute of Cell Biology and Immunology
University of Stuttgart	University of Stuttgart
Allmandring 31, 70569 Stuttgart, Germany	Allmandring 31, 70569 Stuttgart, Germany
E-mail: <a href="mailto:markus.morrison@izi.uni-stuttgart.de">markus.morrison@izi.uni-stuttgart.de</a>	E-mail: <a href="mailto:cristiano.gutta@izi.uni-stuttgart.de">cristiano.gutta@izi.uni-stuttgart.de</a>
ORCID ID 0000-0001-6149-9261	

15

16                   **Abstract**

17                   Established prognostic tests based on limited numbers of transcripts can identify high-risk  
18                   breast cancer patients yet are approved only for individuals presenting with specific clinical  
19                   features or disease characteristics. Deep learning algorithms could hold potential for  
20                   stratifying patient cohorts based on full transcriptome data, yet the development of robust  
21                   classifiers is hampered by the number of variables in omics datasets typically far exceeding  
22                   the number of patients. To overcome this hurdle, we propose a classifier based on a data

23 augmentation pipeline consisting of a Wasserstein generative adversarial network (GAN) with  
24 gradient penalty and an embedded auxiliary classifier to obtain a trained GAN discriminator  
25 (T-GAN-D). Applied to 1244 patients of the METABRIC breast cancer cohort, this classifier  
26 outperformed established breast cancer biomarkers in separating low- from high-risk patients  
27 (disease specific death, progression or relapse within 10 years from initial diagnosis).  
28 Importantly, the T-GAN-D also performed across independent, merged transcriptome datasets  
29 (METABRIC and TCGA-BRCA cohorts), and merging data improved overall patient  
30 stratification. In conclusion, GAN-based data augmentation therefore allowed generating a  
31 robust classifier capable of stratifying low- vs high-risk patients based on full transcriptome  
32 data and across independent and heterogeneous breast cancer cohorts.  
33

## 34 **Introduction**

35 Breast cancer is the tumor with the highest incidence in women, accounting for 2.3 million  
36 new diagnoses and 685,000 deaths worldwide in 2020. According to the World Health  
37 Organization, nearly eight million patients were diagnosed with breast cancer in the five years  
38 before 2020, making it the most prevalent tumor disease worldwide <sup>1</sup>. In current clinical  
39 practice, the expression of estrogen receptor (ER), progesterone receptor (PR), and human  
40 epidermal growth factor receptor 2 (HER2) is determined by immunohistochemistry (IHC),  
41 with the expression patterns defining to which molecular subtype (luminal A, luminal B,  
42 HER2 positive or enriched and triple-negative breast cancer) individual tumors belong.  
43 Prognosis differs between these subtypes, and subtyping informs treatment plans in patients in  
44 which surgical resection of the tumor alone is insufficient <sup>2</sup>. However, substantial response  
45 heterogeneities to the current standard of care treatments can be observed in populations of  
46 breast cancer patients <sup>3</sup>, highlighting the need for additional prognostic markers that could  
47 serve to identify high risk patients that could instead benefit from alternative treatments or for  
48 which the burden from inefficient standard of care treatments could be avoided <sup>4</sup>.

49 Various multi-gene activity tests based on transcript abundance have been developed to assist  
50 in the clinical management of breast cancer (e.g. Oncotype DX <sup>5</sup>, MammaPrint <sup>6,7</sup>, Prosigna  
51 <sup>8,9</sup>, OncoMasTR<sup>10</sup>) and received regulatory approval as prognostic tests <sup>11</sup>. Despite the  
52 prognostic value of these assays, their use is restricted to only subsets of patients with specific  
53 clinical characteristics (e.g. cancer stage, receptor or lymph node status, tumor size,  
54 menopause state, age group) <sup>12-14</sup>. It would therefore be desirable if more generally applicable  
55 prognostic tests based on transcriptome data could be developed.

56 The rapid advances in high-throughput sequencing technologies make tumor transcriptome  
57 data from larger patient cohorts increasingly available. The accessibility of -omics databases  
58 and companion clinical information now also encourages the application of deep learning  
59 (DL) methods to the oncology field, with the aim of learning and extracting features within

60 large scale data that are not readily accessible by classical statistical and pattern recognition  
61 approaches. It is hoped that from DL-based methods tools can be developed that can aid in  
62 further advancing cancer diagnosis, prognosis or predicting treatment efficacy in the future <sup>15</sup>.  
63 DL algorithms such as convolutional neural networks (CNN) were originally applied for  
64 image analysis but could be successfully repurposed to take non-image objects as input, such  
65 as RNA-seq data <sup>16</sup>. One of the major pitfalls when applying DL models to transcriptome  
66 datasets is the typical imbalance between the number of quantified mRNAs (high) and the  
67 number of patients (low), which can lead to overfitting when solving classification tasks <sup>17</sup>. In  
68 addition, low numbers of samples or patients that represent one category (e.g. good prognosis)  
69 come at the risk of capturing patterns that are not robust when applied to larger populations <sup>18</sup>.  
70 Feature selection strategies <sup>19</sup>, under- and over-sampling <sup>20</sup> are three strategies that may help  
71 mitigating effects arising from imbalanced source data. An alternative strategy lies in novel  
72 data augmentation approaches, such as generative adversarial networks (GANs), by which  
73 source datasets can be enriched with artificially generated additional data. GANs are typically  
74 applied to imaging data and are composed of two subnetworks, the generator and the  
75 discriminator. While the former produces synthetic images, the latter is challenged to  
76 discriminate fake vs. real images. Reiterating this process, the generator learns to produce  
77 images with features that can no longer be separated from the real images by the  
78 discriminator, with these generated images then enriching the source dataset <sup>21</sup>. In comparison  
79 to other generative models, GANs are currently preferred due to their computational speed  
80 and the quality of the generated images <sup>22</sup>. In addition, they exhibit a lower risk of overfitting  
81 classifiers and are less susceptible to the impact of non-pertinent image features (such as  
82 brightness) when enriching training data with synthetic images <sup>23</sup>. For example, GANs have  
83 been applied in the medical field to generate synthetic magnetic resonance, computed  
84 tomography or positron emission tomography images <sup>24</sup>. Aside from image-data, different

85 GAN implementations were also successfully applied to transcriptome data for cancer  
86 diagnosis<sup>25,26</sup>, staging<sup>27</sup> and subtyping<sup>28</sup>.  
87 The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, hereafter  
88 MB)<sup>29</sup> and The Cancer Genome Atlas – Breast Invasive Carcinoma (TCGA-BRCA, hereafter  
89 TCGA)<sup>30</sup> cohorts represent two of the largest and most exhaustively annotated breast cancer  
90 datasets, including, in addition to mRNA expression data, features such as patient  
91 demographics, cancer staging, receptor statuses, and follow-up information such as survival  
92 times. Despite not being directly interoperable due to different sequencing technologies, these  
93 datasets can serve as use cases to test new DL-based prognostication approaches.  
94 In this study, we therefore set out to develop a prognostication framework that used the  
95 trained discriminator of a GAN architecture as a standalone classifier and compared its  
96 performance to classical breast cancer biomarkers and a classical CNN.

97

## 98 **Materials and methods**

### 99 **Data integration**

100 The METABRIC (MB) dataset was used to develop the prototype network implementation.  
101 Transcriptome data (median Z-scores), overall survival (OS), disease specific survival (DSS)  
102 and associated clinical records were downloaded from cbioportal.org<sup>31,32</sup>. The dataset was  
103 integrated with locoregional and distant recurrence information retrieved from Rueda et al.<sup>33</sup>  
104 and *Risk of Recurrence – Proliferation* (ROR-P) scores reported by Xia et al.<sup>9</sup>. Clinical  
105 records, OS, DSS and progression free interval (PFI) of the validation TCGA-BRCA cohort  
106 (TCGA) were integrated from cbioportal.org<sup>31,32</sup> and Liu et al. 2018<sup>34</sup>, respectively. To  
107 merge the mRNA expression data of the two cohorts, normalized transcriptome datasets were  
108 downloaded using the R package MetaGxBreast<sup>35</sup>. The transcript amounts were rescaled as  
109 described by Gendoo et al.<sup>35</sup> so that the 2.5 percentile corresponds to -1 and the 97.5  
110 percentile corresponds to +1. Subsequently, transcripts overlapping between the two cohorts

111 and with quantitative information missing in not more than five patients were retained,  
112 resulting in transcripts for  $m = 14042$  genes. The R script used to download and rescale the  
113 datasets is available in the Zenodo repository <sup>36</sup>.

114

### 115 **Inclusion criteria and category definition**

116 Both cohorts were filtered to exclude normal-like subtype samples <sup>9,37,38</sup> and patients for  
117 which less than 10 years of follow-up time from diagnosis were available. Low and high risk  
118 categories were defined according to published clinical records <sup>8,9</sup> as follows:

119 - high risk patients:

120       ○ MB cohort: disease specific death, locoregional or distant recurrence event  
121       recorded before 10 years from initial diagnosis;

122       ○ TCGA cohort: disease specific death, progression, local recurrence or distant  
123       metastases before 10 years from initial diagnosis.

124 - low risk patients: none of the above-mentioned events recorded before 10 years from  
125       initial diagnosis.

126 In total, 1248 patients of the MB cohort ( $n = 567$  high risk,  $n = 681$  low risk) and 165 patients  
127 of the TCGA cohort ( $n = 132$  high risk,  $n = 33$  low risk) satisfied the inclusion criteria. Four  
128 patients from each cohort were excluded after merging due to insufficient expression data.

129

### 130 **Survival analysis and accuracy**

131 Log-rank testing was used to compare predicted low vs high risk patients over a follow-up  
132 time of 10 years. Kaplan-Meier (KM) survival curves were computed using GraphPad Prism  
133 8 (GraphPad Software, San Diego, CA). The area between the curves (ABC) displayed on the  
134 KM graphs for the pooled predictions was calculated as follows:

135 - Low risk AUC minus Predicted low risk AUC;

136 - Predicted low risk AUC minus Predicted high risk AUC;

137 - Predicted high risk AUC minus High risk AUC.

138 The ABCs values are shown on the graphs in the abovementioned order top to bottom. The

139 AUC was computed using GraphPad Prism 8 (GraphPad Software, San Diego, CA).

140 Univariate and multivariate hazard ratios were calculated using the function *coxph* from the

141 R's library *survival* (v. 3.4.0, <https://www.r-project.org/>).

142

### 143 **GAN architecture**

144 The architecture was based on a Wasserstein<sup>39</sup> GAN<sup>21</sup> with gradient penalty<sup>40</sup> and an

145 auxiliary classifier<sup>41</sup> as a variant of a conditional GAN implementation<sup>42</sup>, yielding a AC-

146 WGAN-GP architecture. The Wasserstein loss was implemented to reduce vanishing

147 gradients and mode collapse<sup>43</sup> in the early phases of the training when the discriminator

148 outperformed the generator. Stability was improved by exchanging the weights clipping

149 approach described in Arjovsky et al.<sup>39</sup>, with the gradient penalty described in Gulrajani et al.

150 <sup>40</sup>. To create a conditional GAN, an auxiliary classifier network was implemented<sup>41</sup>, resulting

151 in a more stable training process and reduced mode collapse compared to the standard

152 conditional GAN approach, supplying labels to both discriminator and generator<sup>43</sup>. A z-

153 vector of size 250 was fed as input for the generator. Following good training practice<sup>44</sup>,

154 strided convolutions with step size 2, batch normalization and LeakyRELU as activation

155 function were used. Since using batch normalization in the discriminator and/or the ADAM

156 optimizer led to an unstable training process, batch normalization<sup>45</sup> was only used in the

157 generator, and RMSprop was selected as the activation function. A shallow network

158 consisting of two layers in both the discriminator and the generator led to the most stable

159 training process, due to the smaller number of trainable parameters compared to deeper

160 networks. Hyperparameters were tuned empirically, selecting 1000 epochs for the training

161 process. Three “discriminator-only” training runs were performed before each full network

162 training run, and the generated pictures were subsequently smoothed with a final convolution

163 layer with one filter and stride size of 1. The GAN architecture generated expression data of  
164 size 144x144 when using the entire transcriptome dataset of the MB cohort alone ( $m = 18543$   
165 genes) and 120x120 when merging the MB and TCGA cohorts ( $m = 14042$  genes). In the  
166 latter setting, expression profiles with less than 14,440 transcripts were filled with random  
167 values, leading to better convergence. The resulting trained GAN Discriminator (T-GAN-D)  
168 was then used as an independent classifier to discriminate low and high risk patients. The  
169 Python code and the input files used to generate the predictions are available in the Zenodo  
170 repository<sup>36</sup>.

171

## 172 **CNN architecture**

173 As the performance of the CNN implemented as the GAN's discriminator showed satisfactory  
174 performance, a similar architecture was used as a benchmark classifier. Batch normalization  
175 was employed to ensure shorter training periods and RELU was used as the activation  
176 function. A fixed training length of 1250 epochs was set due to the limited sample size and to  
177 generate comparable iterations.

178 The accuracy of both classifiers was calculated dividing the number of correct classifications  
179 by the total number of classifications performed.

180

## 181 **Results**

182

### 183 **The METABRIC and BRCA-TCGA cohorts lend themselves as use cases for data** 184 **augmentation and development of prognostication classifiers**

185 One of the major challenges of machine learning applied to -omics data and companion  
186 medical records is the imbalance between the high amounts of variables compared to the  
187 limited number of patients available. Even in the case of breast cancer, one of the most  
188 frequent and widely studied malignant neoplasms, this limitation is apparent in the two major



189 public transcriptome datasets, namely the MB cohort (n = 1904 patients, m = 18543  
190 transcripts) and the TCGA cohort (n = 1101 patients, m = 20532 transcripts). This imbalance  
191 is exacerbated for prognostic analyses that require long-term (10 years) follow-up information  
192 and the application of further exclusion criteria (see methods), reducing cohort sizes to n =  
193 1248 and n = 165, respectively (**Fig. 1A, B**). Both cohorts behaved notably different, with  
194 patients in the MB cohort on average having an overall substantially better prognosis in  
195 overall survival and relapse-free, progression-free or disease specific survival (**Fig. 1C, D**).  
196 This is likely attributable to the MB dataset largely consisting of stage I and stage II patients  
197 (89.5% of patients with reported disease stage at diagnosis), whereas stage III and IV patients  
198 are more prominent in the TCGA dataset (40.4% of individuals with available disease stage at  
199 diagnosis). Despite these differences, the high risk subgroups of both cohorts showed  
200 comparable median survival times (MB = 31.9 months [**Fig. 1E**], TCGA = 26.3 months [**Fig.**  
201 **1F**]). Due to the limited sizes of these cohorts, they lend themselves as suitably challenging  
202 use-cases for applying and testing data augmentation for improving prognostication. In  
203 particular, we set out to implement a classifier based on a data augmentation network for  
204 improved patient stratification in the MB cohort, to subsequently validate robustness and  
205 transferability by integrating the independent TCGA cohort.

206

## 207 **A trained GAN discriminator robustly identifies low and high risk breast cancer** 208 **patients**

209 To tackle the problem of data scarcity, we implemented a GAN architecture to augment  
210 transcriptomic data of the MB cohort and tested the performance of a trained discriminator in  
211 stratifying breast cancer patients. First, individual patient transcriptome profiles were rescaled  
212 and converted into arrays of pixels (**Fig. 2A i**) in order to use these images as an input for the  
213 GAN. Independent of these true patient data, the generator created images representing the

214 transcript profiles of synthetic hypothetical patients together with their category (low or high  
215 risk) (**Fig. 2A ii**). After being exposed to a fraction of the real transcriptome images and  
216 associated categories, its adversary, the discriminator network then tried to distinguish fake  
217 from real transcriptome images for high or low risk patients (**Fig. 2A iii**). Reiterating this  
218 training process over 1000 epochs, the generator learned to create realistic synthetic  
219 transcriptome images for high and low risk categories, which then could be used to augment  
220 the original MB cohort data. Associated characteristics of this process (discriminator loss,  
221 discriminator class loss, generator loss) are shown in **Supplementary Fig. 1**. Using this  
222 approach, the discriminator learned to identify features relevant for the risk category  
223 definition, aided by the synthetic profiles that enriched the real training data at each epoch.  
224 The trained GAN discriminator (T-GAN-D) resulting from this process then was used as a  
225 standalone classifier to categorize images from the test fraction of the cohort into the high or  
226 low risk categories (**Fig. 2A iv**), thus prognosticating patient outcome.

227 We first implemented and tested the T-GAN-D for its prognostic capability using follow-up  
228 and mRNA expression data of the prototyping MB cohort, consisting of  $n = 1248$  individuals  
229 and  $m = 18543$  genes. Within this cohort, we independently cross-validated (CV) five-fold  
230 with randomly composed training data. Kaplan-Meier curves and log rank testing for each run  
231 yielded significant class separations in 4 out of 5 iterations (**Fig. 2B, Supplementary Fig.**  
232 **2A**). Pooling the results so that each patient of the MB dataset was present once in the  
233 survival analysis, the T-GAN-D separated high and low risk patients with high statistical  
234 significance ( $p\text{-value} = 2.71\text{E-}12$ ) (**Fig. 2C**). To obtain a reference performance baseline, a  
235 classical CNN was challenged with the same task, using the same training and test sets for  
236 each iteration. The CNN yielded class separations with a  $p < 0.05$  in only two out of five  
237 iterations (**Fig. 2D, Supplementary Fig. 2B**). In the pooled comparison, the CNN performed  
238 well yet failed to outperform the T-GAN-D in separating low vs. high risk patients (**Fig. 2E,**  
239 **Supplementary Table 1**). These results therefore demonstrate that the reiterative learning

240 process of a GAN to train its discriminator and use it as an independent classifier provides a  
241 more robust and slightly improved patient stratification than a classical DL approach.

242

### 243 **Introducing an independent cohort improves MB patient classification**

244 A common limitation of predictors and classifiers is their limited robustness and  
245 transferability to independent datasets. This might arise from overfitting or overtraining  
246 within the initial cohort but also from heterogeneity and batch effects between source  
247 datasets. For validating our approach further, we therefore merged the mRNA expression data  
248 of the MB and TCGA cohorts, which originally were quantified with bead-based microarray  
249 technology (Illumina Human V3) or RNA-Seq (Illumina HiSeq) platforms respectively<sup>46</sup>, by  
250 rescaling the expression of transcripts overlapping between the two cohorts ( $m = 14042$ ). We  
251 then retrained the discriminator using the entire TCGA data plus a fraction of the MB data  
252 from the merged dataset and generated predictions on an independent subset of MB patients  
253 (**Fig. 3A**), using five-fold cross-validation. The T-GAN-D again separated patients into low  
254 and high-risk categories with high statistical significance (**Fig. 3B, Supplementary Fig. 3A**).  
255 The CNN trained and tested with the same data performed similarly well (**Fig. 3C,**  
256 **Supplementary Fig. 3B**). The T-GAN-D trained on the merged and reduced dataset also  
257 showed improved accuracy when compared to all settings where both a CNN or the GAN  
258 were trained on the full or reduced MB dataset alone (**Supplementary Table 1, 2**). Therefore,  
259 in our setting, rescaling and converting transcriptome profiles into images was sufficient to  
260 successfully merge the two cohorts without the need for further preprocessing steps and  
261 allowed to stratify patients into high and low risk classes.

262

263 **The T-GAN-D outperforms classical outcome predictors and accurately stratifies early**  
264 **stage patients into risk categories**

265 We next compared the performance of CNN and GAN based classifications to other  
266 established clinical markers in breast cancer. These included a scoring system based on a  
267 multi-transcript signature (Risk-of-recurrence - proliferation, [ROR-P]), estrogen receptor  
268 status (ER), human epidermal growth factor receptor 2 status (HER2), and progesterone  
269 receptor status (PR). Likewise, tumor staging was included, yet was available for only 911 out  
270 of 1248 patients of the MB cohort. The hazard ratios (HR) obtained from a univariate analysis  
271 were comparable for ROR-P, HER2 or tumor staging as classifiers, and similar HRs were also  
272 obtained for the CNN and T-GAN-D classifiers developed from only the MB transcriptome  
273 dataset (**Fig. 4A**). Interestingly, the T-GAN-D classifier resulting from the merged cohort data  
274 returned a mean  $HR > 2.0$  ( $\pm 0.4$ ), thereby surpassing all other markers. This feature was even  
275 more pronounced in a multivariate analysis including ER, HER2 and PR biomarkers (**Fig.**  
276 **4B**). When reducing the MB cohort to those patients for which staging information was  
277 available, HRs based on staging and T-GAN-D were comparable (**Fig. 4C**). To test whether  
278 both classifiers might be redundant, we performed a T-GAN-D based survival analysis within  
279 the tumor stage I and stage II subcohorts, which dominate the MB dataset. T-GAN-D based  
280 classification allowed separating high and low risk patients within both tumor stages (**Fig. 4D,**  
281 **E**), indicating non-redundancy of the T-GAN-D classification to tumor staging information.  
282 Taken together, these results show that training through data augmentation can enhance the  
283 prognostic performance of DL classifiers, and in this case surpasses individual classical  
284 biomarkers. In addition, the T-GAN-D performed well in prognostication of early stage breast  
285 cancer cases.

286

287 **The T-GAN-D stratifies TCGA patients despite these being scarcely represented**

288 After observing that introducing TCGA patients into the training set of the T-GAN-D did not  
289 degrade, but improved the stratification of MB patients, we tested the performance of the  
290 classifier on the smaller TCGA dataset. To do this, we trained the discriminator using the  
291 entire MB data plus a fraction of the TCGA data from the merged dataset and generated  
292 predictions on an independent subset of TCGA patients (**Fig. 5A**), using five-fold cross-  
293 validation. The T-GAN-D correctly predicted 78% of the cases (**Fig. 5B, Supplementary**  
294 **Fig. 4, Supplementary Table 3**). In contrast, when trained on the MB dataset alone, the T-  
295 GAN-D was not able to separate high and low risk patients (**Fig. 5C, Supplementary Fig. 4**),  
296 achieving an overall accuracy of only 43% (**Supplementary Table 3**). Therefore, the addition  
297 to the training set of a comparably small number of TCGA patients ( $n = 129$ ) to the larger MB  
298 cohort ( $n = 1244$ ) was sufficient to drastically improve the performance of the T-GAN-D  
299 predicting TCGA patient outcome. This demonstrates that even if the training set is largely  
300 dominated by patients belonging to one cohort, the introduction of a limited number of  
301 samples of a second, differently balanced dataset appears sufficient to possibly capture  
302 relevant patterns that contribute to achieving improved prognostic performance.

303

## 304 **Discussion**

305 The increasing availability and routine acquisition of large scale genomic data encourage the  
306 repurposing and application of AI to the field of oncology in order to identify novel means for  
307 improved and personalized prediction of prognosis<sup>47</sup>. In this study, we developed a DL-based  
308 tool to stratify high vs. low risk breast cancer patients according to full transcriptome profiles.  
309 Using the MB and TCGA cohorts as use cases, we converted expression data into images and  
310 used the trained discriminator of our GAN architecture as a standalone prognostic classifier.  
311 Our results show that the T-GAN-D performed better than classical outcome predictors and  
312 maintained robust performance when merging the two cohorts.

313 AI has already been applied to breast cancer based on different classes of data, to inform  
314 diagnosis, treatment planning and prognosis<sup>48,49</sup>. For example, pattern recognition and data  
315 augmentation proved to be promising approaches to assist in generating accurate diagnoses  
316 from mammography images<sup>50,51</sup>. Transcriptome data were also employed to develop ML-  
317 based analysis pipelines for breast cancer subtyping, diagnosis, patient stratification and  
318 identification of altered pathways<sup>52</sup>, and these techniques may improve the accuracy of  
319 cancer prognosis in the future. However, shortcomings must be taken into account, as  
320 applicable also to currently available breast cancer datasets. When dealing with low sample  
321 size - high dimension datasets such as the MB and TCGA cohorts, common DL classification  
322 algorithms such as neural networks may be prone to overfitting<sup>53</sup>. Multi-gene signatures  
323 based on the expression of a lower number of transcripts may circumvent this problem, but  
324 are applicable only to subsets of patients with specific clinical characteristics<sup>11-14</sup>. To tackle  
325 these problems, we aimed at developing a more universally applicable algorithm that takes  
326 advantage of GAN's data augmentation and generalizing capability. In our training strategy,  
327 the T-GAN-D was exposed not only to a subset of original data, but also to the synthetic  
328 patients generated by the generator in each epoch. This approach for the augmentation of  
329 training data was demonstrated before to aid a discriminator in learning hidden features and  
330 correlations<sup>54,55</sup>. When compared to a classic CNN, the T-GAN-D showed comparable, yet  
331 slightly improved performance. Other GAN implementations have been applied to the MB or  
332 TCGA cohorts in the past, addressing different aims such as the generation of missing data<sup>56</sup>,  
333 the identification of multi -omics signatures<sup>57</sup> and prognostication<sup>58</sup>. While showing  
334 encouraging results, these prior works limited the follow up time to 5 years and focused on  
335 death events only. Besides considering longer follow up times, the inclusion of progression or  
336 recurrence events in the class definition can be considered a more exhaustive assessment of a  
337 patient's risk category, since OS or DSS alone may be insufficient especially in early stage  
338 screenings<sup>59</sup>. In addition, short follow up times were shown to affect the prognostication

339 performance of ML algorithms leading to low sensitivity, mostly due to the insufficient  
340 occurrence of recurrence or death events <sup>60</sup>.

341 We demonstrated that the conversion of transcriptome profiles into images allowed the  
342 integration of independent transcriptome datasets. To date, the majority of gene expression  
343 databases cannot be directly integrated due to different sequencing technologies, protocols or  
344 batch effects, with the consequence of producing merely qualitative results in a meta-analysis  
345 fashion or unveiling evidences that remain cohort-specific <sup>61</sup>. To test if our conversion  
346 strategy could allow a straightforward integration of heterogenous datasets, we challenged the  
347 T-GAN-D in assessing the risk category of MB patients, training the network with a subset of  
348 MB patients plus the entire TCGA cohort. Introducing patients belonging to a different cohort  
349 improved the performance of the classifier, which in our case outperformed established  
350 clinical biomarkers and a published ROR-P signature <sup>9</sup> in uni- and multi-variate analyses.

351 The T-GAN-D classifier also stratified early stage breast cancer patients into low and high  
352 risk groups, even though no additional factors such as treatment regimens, age, subtype or  
353 other clinical features were considered when composing the training datasets. Early stage  
354 patients expected to experience recurrence or progression may benefit from more frequent  
355 screenings, yet it remains to be assessed if the transcriptome-based classifier operates  
356 independently of or correlates with other established risk factors.

357 High accuracy in predicting the risk class of the smaller and imbalanced TCGA cohort was  
358 achieved when training the T-GAN-D with a subset of TCGA patients plus the whole MB  
359 dataset. Classical ML algorithms (SVM and random forest, among others) were also shown to  
360 benefit from the combination of TCGA RNA-Seq and MB microarray data, which in a  
361 previous study improved 5 years OS prognostication <sup>62</sup>, but lead to misleadingly high  
362 accuracy due to highly imbalanced classes. Taken together, our results suggest that the T-  
363 GAN-D remains robust when merging cohorts differently balanced between positive and  
364 negative outcomes, and that the network is still able to capture relevant risk patterns when one

365 cohort is heavily underrepresented in the training dataset. Therefore, our classification  
366 framework may allow the integration of new, smaller datasets, lending itself as a suitable  
367 prototype for generating prospective personalized outcome predictions for scarce *de novo*  
368 data.

369 In conclusion, our proof-of-concept study represents an avenue for developing a scalable data  
370 augmentation-based tool that could be a stepping stone towards individualized prognosis in  
371 the future. Molecular high throughput techniques are increasing in quality, resolution and  
372 amount of data produced and are more and more commonly captured in clinical research and  
373 diagnostic environments. It was estimated that within the next decade, between 2 and 40  
374 exabytes of genomic data will be generated every year<sup>63</sup>, with large quantities being related  
375 to human health and disease. GAN-based approaches therefore could become a meaningful  
376 approach to exploit such data for the benefit of patients. In addition, -omics domains other  
377 than transcriptomics likewise have the potential to enter the clinical arena as part of routine  
378 analytical practice, including proteome, metabolome or lipidome data. Such data classes can  
379 readily be integrated with clinical-pathological information<sup>64</sup>, and could be processed with  
380 the assistance of GAN based approaches to improve patient-tailored interventions or  
381 prognostication.

382

### 383 **Acknowledgements**

384 MR and CG receive funding by the Deutsche Forschungsgemeinschaft (DFG, German  
385 Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016 and  
386 acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

387

### 388 **References**

- 389 1. Breast cancer. Available at: [https://www.who.int/news-room/fact-sheets/detail/breast-](https://www.who.int/news-room/fact-sheets/detail/breast-cancer)  
390 cancer. (Accessed: 30th August 2022)



- 391 2. Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and  
392 therapeutic implications. *World J. Clin. Oncol.* **5**, 412 (2014).
- 393 3. Turashvili, G. & Brogi, E. Tumor heterogeneity in breast cancer. *Front. Med.* **4**, 227  
394 (2017).
- 395 4. Cardoso, F. *et al.* 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage  
396 Breast Cancer. *N. Engl. J. Med.* **375**, 717–729 (2016).
- 397 5. Syed, Y. Y. Oncotype DX Breast Recurrence Score®: A Review of its Use in Early-  
398 Stage Breast Cancer. *Mol. Diagnosis Ther.* **24**, 621–632 (2020).
- 399 6. Van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast  
400 cancer. *Nat. 2002 4156871* **415**, 530–536 (2002).
- 401 7. Arc, M. *et al.* A Gene-Expression Signature as a Predictor of Survival in Breast  
402 Cancer. <https://doi.org/10.1056/NEJMoa021967> **347**, 1999–2009 (2002).
- 403 8. Bernard, P. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic  
404 subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- 405 9. Xia, Y., Fan, C., Hoadley, K. A., Parker, J. S. & Perou, C. M. Genetic determinants of  
406 the molecular portraits of epithelial cancers. *Nat. Commun. 2019 101* **10**, 1–13 (2019).
- 407 10. Buus, R. *et al.* Validation of the OncoMASTR risk score in estrogen receptor–  
408 positive/HER2-negative patients: A TransATAC study. *Clin. Cancer Res.* **26**, 623–631  
409 (2020).
- 410 11. Ross, J. S., Hatzis, C., Symmans, W. F., Pusztai, L. & Hortobágyi, G. N.  
411 Commercialized Multigene Predictors of Clinical Outcome for Breast Cancer.  
412 *Oncologist* **13**, 477–493 (2008).
- 413 12. Yao, K., Tong, C. Y. & Cheng, C. A framework to predict the applicability of  
414 Oncotype DX, MammaPrint, and E2F4 gene signatures for improving breast cancer  
415 prognostic prediction. *Sci. Reports 2022 121* **12**, 1–11 (2022).
- 416 13. Kelly, C. M. *et al.* Comparison of the prognostic performance between OncoMasTR

- 417 and OncotypeDX multigene signatures in hormone receptor-positive, HER2-negative,  
418 lymph node-negative breast cancer.  
419 [https://doi.org/10.1200/JCO.2018.36.15\\_suppl.12074](https://doi.org/10.1200/JCO.2018.36.15_suppl.12074) **36**, 12074–12074 (2018).
- 420 14. Jensen, M. B. *et al.* The Prosigna gene expression assay and responsiveness to adjuvant  
421 cyclophosphamide-based chemotherapy in premenopausal high-risk patients with  
422 breast cancer. *Breast Cancer Res.* **20**, (2018).
- 423 15. Tran, K. A. *et al.* Deep learning in cancer diagnosis, prognosis and treatment selection.  
424 *Genome Med.* 2021 131 **13**, 1–17 (2021).
- 425 16. Sharma, A., Vans, E., Shigemizu, D., Boroevich, K. A. & Tsunoda, T. DeepInsight: A  
426 methodology to transform a non-image data to an image for convolution neural  
427 network architecture. *Sci. Reports* 2019 91 **9**, 1–7 (2019).
- 428 17. Liu, R. & Gillies, D. F. Overfitting in linear feature extraction for classification of  
429 high-dimensional image data. *Pattern Recognit.* **53**, 73–86 (2016).
- 430 18. Barandela, R., Valdovinos, R. M., Salvador Sánchez, J. & Ferri, F. J. The imbalanced  
431 training sample problem: under or over sampling? *Lect. Notes Comput. Sci. (including*  
432 *Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **3138**, 806–814 (2004).
- 433 19. Raghu, V. K., Ge, X., Chrysanthis, P. K. & Benos, P. V. Integrated Theory- and Data-  
434 driven Feature Selection in Gene Expression Data Analysis. *Proceedings. Int. Conf.*  
435 *Data Eng.* **2017**, 1525 (2017).
- 436 20. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic  
437 Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2011).
- 438 21. Goodfellow, I. *et al.* Generative Adversarial Networks. *Commun. ACM* **63**, 139–144  
439 (2014).
- 440 22. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep  
441 Learning. *J. Big Data* **6**, 1–48 (2019).
- 442 23. Bowles, C. *et al.* GAN Augmentation: Augmenting Training Data using Generative

- 443 Adversarial Networks. (2018). doi:10.48550/arxiv.1810.10863
- 444 24. Li, X. *et al.* When medical images meet generative adversarial network: recent  
445 development and research opportunities. *Discov. Artif. Intell.* 2021 11 **1**, 1–20 (2021).
- 446 25. Xiao, Y., Wu, J. & Lin, Z. Cancer diagnosis using generative adversarial networks  
447 based on deep learning from imbalanced data. *Comput. Biol. Med.* **135**, 104540 (2021).
- 448 26. Wei, K., Li, T., Huang, F., Chen, J. & He, Z. Cancer classification with data  
449 augmentation based on generative adversarial networks. *Front. Comput. Sci.* 2022 162  
450 **16**, 1–11 (2021).
- 451 27. Kwon, C. H., Park, S., Ko, S. & Ahn, J. Increasing prediction accuracy of pathogenic  
452 staging by sample augmentation with a GAN. *PLoS One* **16**, e0250458 (2021).
- 453 28. Yang, H., Chen, R., Li, D. & Wang, Z. Subtype-GAN: a deep learning approach for  
454 integrative cancer subtyping of multi-omics data. *Bioinformatics* **37**, 2231–2237  
455 (2021).
- 456 29. Mukherjee, A. *et al.* Associations between genomic stratification of breast cancer and  
457 centrally reviewed tumour pathology in the METABRIC cohort. *npj Breast Cancer*  
458 2018 41 **4**, 1–9 (2018).
- 459 30. The Cancer Genome Atlas Program - NCI. Available at:  
460 <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>.  
461 (Accessed: 30th August 2022)
- 462 31. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles  
463 using the cBioPortal. *Sci. Signal.* **6**, (2013).
- 464 32. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring  
465 multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–4 (2012).
- 466 33. Rueda, O. M. *et al.* Dynamics of breast-cancer relapse reveal late-recurring ER-positive  
467 genomic subgroups. *Nature* **567**, 399–404 (2019).
- 468 34. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-

- 469 Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).
- 470 35. Gendoo, D. M. A. *et al.* MetaGxData: Clinically Annotated Breast, Ovarian and  
471 Pancreatic Cancer Datasets and their Use in Generating a Multi-Cancer Gene  
472 Signature. *Sci. Rep.* **9**, (2019).
- 473 36. Guttà, C., Morhard, C. & Rehm, M. T-GAN-D: a GAN-based classifier for breast  
474 cancer prognostication. (2022). doi:10.5281/ZENODO.7151831
- 475 37. Troester, M. A. *et al.* Racial Differences in PAM50 Subtypes in the Carolina Breast  
476 Cancer Study. *JNCI J. Natl. Cancer Inst.* **110**, 176–182 (2018).
- 477 38. Sweeney, C. *et al.* Intrinsic subtypes from PAM50 gene expression assay in a  
478 population-based breast cancer cohort: Differences by age, race, and tumor  
479 characteristics. *Cancer Epidemiol. Biomarkers Prev.* **23**, 714 (2014).
- 480 39. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein GAN. (2017). Available at:  
481 <https://arxiv.org/abs/1701.07875v3>. (Accessed: 1st March 2022)
- 482 40. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. Improved  
483 Training of Wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **2017-December**, 5768–  
484 5778 (2017).
- 485 41. Odena, A., Olah, C. & Shlens, J. Conditional Image Synthesis With Auxiliary  
486 Classifier GANs. *34th Int. Conf. Mach. Learn. ICML 2017* **6**, 4043–4055 (2016).
- 487 42. Mirza, M. & Osindero, S. Conditional Generative Adversarial Nets. (2014). Available  
488 at: <https://arxiv.org/abs/1411.1784v1>. (Accessed: 1st March 2022)
- 489 43. Kodali, N., Abernethy, J., Hays, J. & Kira, Z. On Convergence and Stability of GANs.  
490 (2017). Available at: <https://arxiv.org/abs/1705.07215v5>. (Accessed: 1st March 2022)
- 491 44. Radford, A., Metz, L. & Chintala, S. Unsupervised Representation Learning with Deep  
492 Convolutional Generative Adversarial Networks. *4th Int. Conf. Learn. Represent. ICLR*  
493 *2016 - Conf. Track Proc.* (2015).
- 494 45. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by

- 495 Reducing Internal Covariate Shift. *32nd Int. Conf. Mach. Learn. ICML 2015* **1**, 448–  
496 456 (2015).
- 497 46. Craven, K. E., Gökmen-Polar, Y. & Badve, S. S. CIBERSORT analysis of TCGA and  
498 METABRIC identifies subgroups with better outcomes in triple negative breast cancer.  
499 *Sci. Reports 2021 111* **11**, 1–19 (2021).
- 500 47. Wallis, C. How Artificial Intelligence Will Change Medicine. *Nature* **576**, S48 (2019).
- 501 48. Zhang, C. *et al.* Cancer diagnosis with DNA molecular computation. *Nat. Nanotechnol.*  
502 *2020 158* **15**, 709–715 (2020).
- 503 49. Jia, D. *et al.* Breast Cancer Case Identification Based on Deep Learning and  
504 Bioinformatics Analysis. *Front. Genet.* **12**, 767 (2021).
- 505 50. McKinney, S. M. *et al.* International evaluation of an AI system for breast cancer  
506 screening. *Nat. 2020 5777788* **577**, 89–94 (2020).
- 507 51. Desai, S. D., Giraddi, S., Verma, N., Gupta, P. & Ramya, S. Breast Cancer Detection  
508 Using GAN for Limited Labeled Dataset. *Proc. - 2020 12th Int. Conf. Comput. Intell.*  
509 *Commun. Networks, CICN 2020* 34–39 (2020). doi:10.1109/CICN49253.2020.9242551
- 510 52. Liñares-Blanco, J., Pazos, A. & Fernandez-Lozano, C. Machine learning analysis of  
511 TCGA cancer data. *PeerJ Comput. Sci.* **7**, 1–47 (2021).
- 512 53. Liu, B., Wei, Y., Zhang, Y. & Yang, Q. Deep neural networks for high dimension, low  
513 sample size data. in *IJCAI International Joint Conference on Artificial Intelligence*  
514 2287–2293 (2017). doi:10.24963/ijcai.2017/318
- 515 54. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-  
516 level performance on imagenet classification. in *Proceedings of the IEEE International*  
517 *Conference on Computer Vision 2015 Inter*, 1026–1034 (2015).
- 518 55. Shams, S., Platania, R., Zhang, J., Kim, J. & Park, S. J. Deep generative breast cancer  
519 screening and diagnosis. in *Lecture Notes in Computer Science (including subseries*  
520 *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11071**

- 521 LNCS, 859–867 (Springer Verlag, 2018).
- 522 56. Arya, N. & Saha, S. Generative Incomplete Multi-View Prognosis Predictor for Breast  
523 Cancer: GIMPP. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 1–1 (2021).  
524 doi:10.1109/TCBB.2021.3090458
- 525 57. Kim, M., Oh, I. & Ahn, J. An Improved Method for Prediction of Cancer Prognosis by  
526 Network Learning. *Genes (Basel)*. **9**, 1. – 11 (2018).
- 527 58. Hsu, T. C. & Lin, C. Generative Adversarial Networks for Robust Breast Cancer  
528 Prognosis Prediction with Limited Data Size. *Proc. Annu. Int. Conf. IEEE Eng. Med.  
529 Biol. Soc. EMBS 2020-July*, 5669–5672 (2020).
- 530 59. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I.  
531 Machine learning applications in cancer prognosis and prediction. *Comput. Struct.  
532 Biotechnol. J.* **13**, 8–17 (2015).
- 533 60. Boeri, C. *et al.* Machine Learning techniques in breast cancer prognosis prediction: A  
534 primary evaluation. *Cancer Med.* **9**, 3234 (2020).
- 535 61. Carnielli, C. M. *et al.* Combining discovery and targeted proteomics reveals a  
536 prognostic signature in oral cancer. *Nat. Commun.* **9**, 3598 (2018).
- 537 62. Dubourg-Felonneau, G. *et al.* A Framework for Implementing Machine Learning on  
538 Omics Data. (2018). doi:10.48550/arxiv.1811.10455
- 539 63. Stephens, Z. D. *et al.* Big Data: Astronomical or Genomical? *PLoS Biol.* **13**, (2015).
- 540 64. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat. Rev.  
541 Genet.* **19**, 299 (2018).

542  
543

#### 544 **Figure legends**

#### 545 **Fig. 1. MB and TCGA patient demographics and survival**

546 (A) Patients demographics of the MB subcohort. (B) Patients demographics of the TCGA

547 subcohort. (C) Overall and (D) relapse-free, progression-free or disease specific survival of  
548 the MB and TCGA cohorts. (E) Kaplan Meier curves comparing low vs high risk patients of  
549 the MB and (F) the TCGA cohorts.

550

551 **Fig. 2. The T-GAN-D robustly stratifies low and high risk breast cancer patients**

552 (A) Workflow of the data processing, including the schematics of the generator network and  
553 its adversary, the discriminator network. Together these result in an AC-WGAN-GP  
554 architecture. After the conversion of patient transcriptome profiles into images, 4/5 of the MB  
555 dataset was used to train the GAN's discriminator. After 1000 epochs, the trained  
556 discriminator was used as a standalone classifier to separate the remaining 1/5 patients of the  
557 dataset into low and high risk categories. (B) Kaplan-Meier curves separating low vs. high  
558 risk patients as predicted with the T-GAN-D (iteration 1 of the 5-fold CV shown as  
559 representative). (C) Kaplan-Meier curves generated pooling the category predictions obtained  
560 for all patients of the MB dataset after five independent CV runs. (D) Separation of low vs.  
561 high risk patients predicted with a classical CNN on the same subset used in B and (E)  
562 comparison obtained pooling the predictions of five independent CV runs. The area between  
563 the curves (ABC) between Low risk (blue dashed line) and Predicted low risk (solid blue  
564 line), Predicted low risk and Predicted high risk (solid red line), Predicted high risk and High  
565 risk groups (dashed red line) are shown top to bottom in D and E.

566

567 **Fig. 3. Introducing the independent TCGA cohort improves MB patient classification**

568 (A) Schematic representing the training strategy: rescaled data from the entire TCGA cohort  
569 were merged with 4/5 of the MB cohort to train the T-GAN-D, which was subsequently used  
570 to predict the risk class of the remaining 1/5 of MB patients. The process was iterated 5 times.  
571 (B) Kaplan-Meier curves based on the pooled predictions of the T-GAN-D trained on both  
572 cohorts. (C) Kaplan-Meier curves separating low vs. high risk patients predicted with the

573 CNN that was trained after merging the MB and the TCGA cohorts. The area between the  
574 curves (ABC) between Low risk (blue dashed line) and Predicted low risk (solid blue line),  
575 Predicted low risk and Predicted high risk (solid red line), Predicted high risk and high risk  
576 groups (dashed red line) are shown top to bottom in **B** and **C**.

577

578 **Fig. 4 The T-GAN-D outperforms classical biomarkers after merging the MB and**

579 **TCGA cohorts and significantly stratifies early stage MB patients**

580 (A) Comparison of the hazard ratios (Cox model, univariate) of a multi-transcript signature  
581 (ROR-P) and established prognostic biomarkers (ER, HER2, PR) vs. the CNN and the T-  
582 GAN-D before and after cohort merging. (B) Multivariate Cox hazard ratio of the T-GAN-D  
583 compared to ROR-P and receptor status and (C) disease stage. (D) Kaplan -Meier curves of  
584 Stage I and (E) Stage II patients stratified by the T-GAN-D into low and high risk categories.

585

586 **Fig. 5 The T-GAN-D stratifies TCGA patients despite these being scarcely represented**

587 **in the merged training set**

588 (A) Schematic representing the training strategy: rescaled data from the entire MB cohort  
589 were merged with 4/5 of the TCGA cohort to train the T-GAN-D, which was subsequently  
590 used to predict the risk class of the remaining 1/5 of TCGA patients. The process was iterated  
591 5 times. (B) Stratification of the TCGA patients by T-GAN-D trained on the merged dataset  
592 and (C) the MB dataset alone. Kaplan-Meier curves were generated pooling the predictions of  
593 all iterations of the 5-fold CV. The area between the curves (ABC) between Low risk (blue  
594 dashed line) and Predicted low risk (solid blue line), Predicted low risk and Predicted high  
595 risk (solid red line), Predicted high risk and High risk groups (dashed red line) are shown top  
596 to bottom in **B** and **C**.



**Fig. 1**

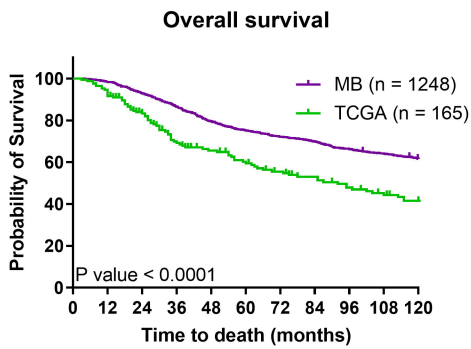
**A**

METABRIC sub-cohort (n = 1248)		
Age at diagnosis (years)	# cases	%
<40	83	6.7
40-49	188	15.1
50-59	288	23.1
60-69	392	31.4
70-79	235	18.8
80+	62	5.0
Disease Stage	# cases	%
Stage I	303	24.3
Stage II	514	41.2
Stage III	88	7.1
Stage IV	8	0.6
Not available	335	26.8
Survival (months)	Median	
Overall survival	163.2	
Relapse or disease specific death	160.2	

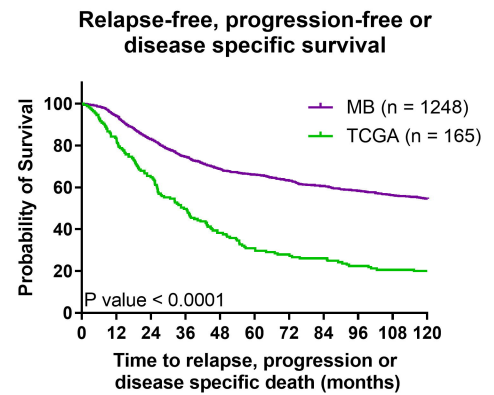
**B**

BRCA-TCGA sub-cohort (n = 165)		
Age at diagnosis (years)	# cases	%
<40	20	12.1
40-49	38	23.0
50-59	29	17.6
60-69	44	26.7
70-79	22	13.3
80+	12	7.3
Disease Stage	# cases	%
Stage I	22	13.3
Stage II	72	43.6
Stage III	50	30.3
Stage IV	14	8.5
Not available	7	4.2
Survival (months)	Median	
Overall survival	92	
Progression or disease specific death	35.5	

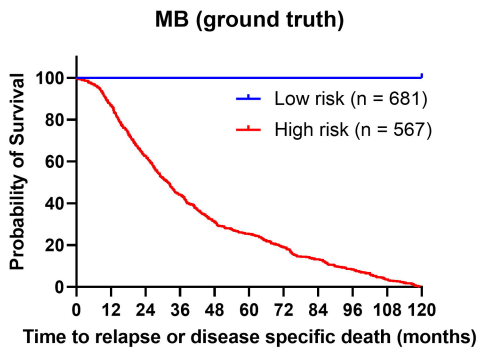
**C**



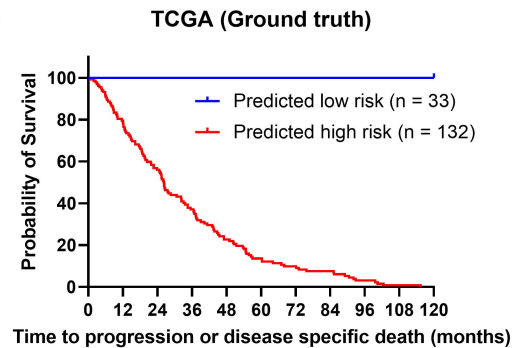
**D**



**E**

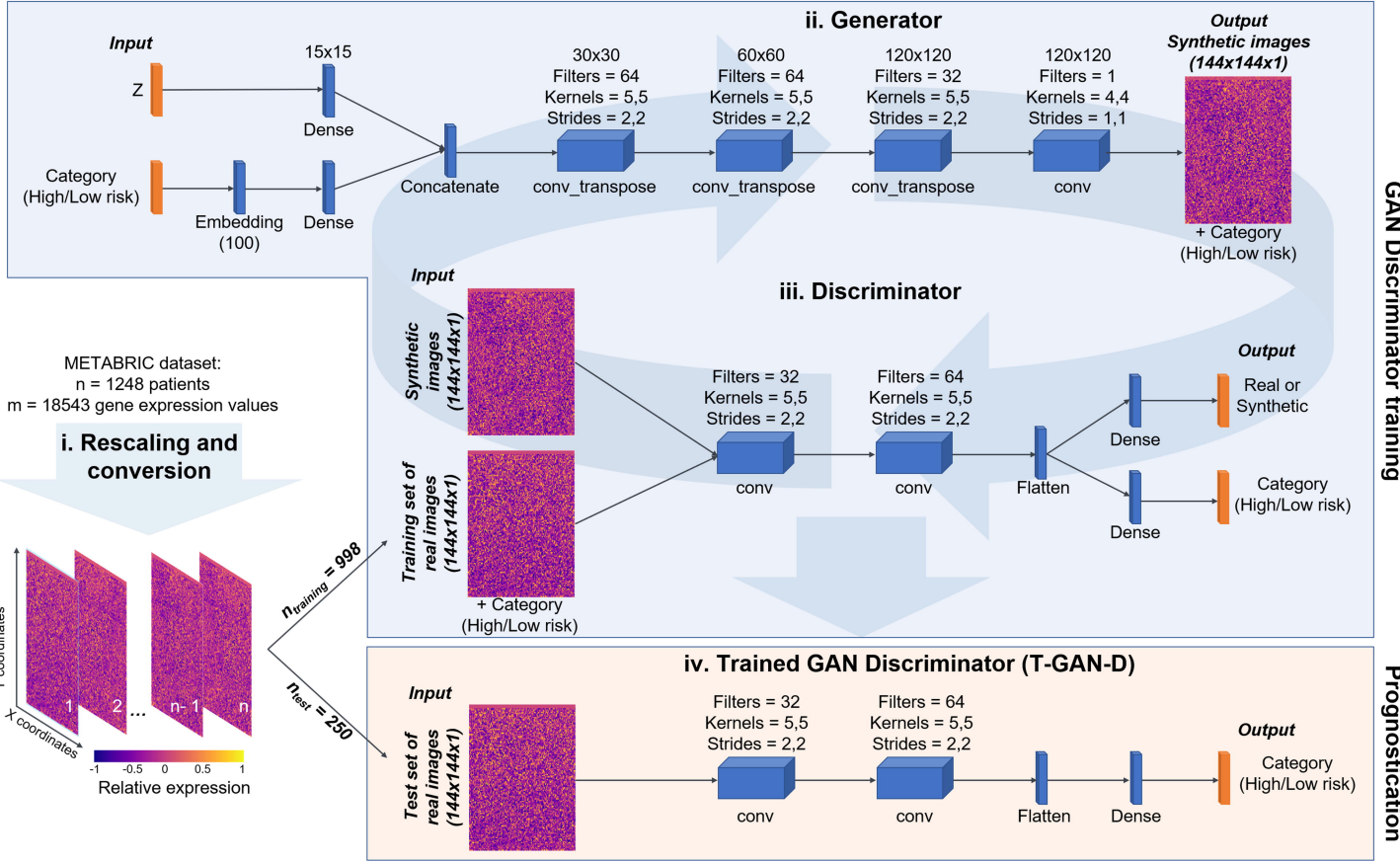


**F**

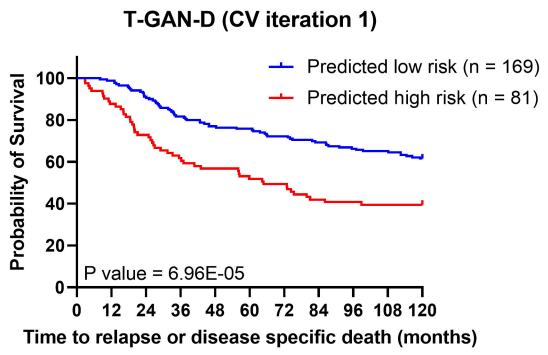


**Fig. 2**

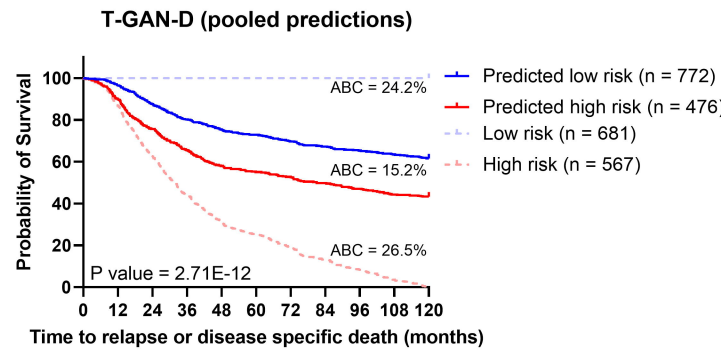
**A**



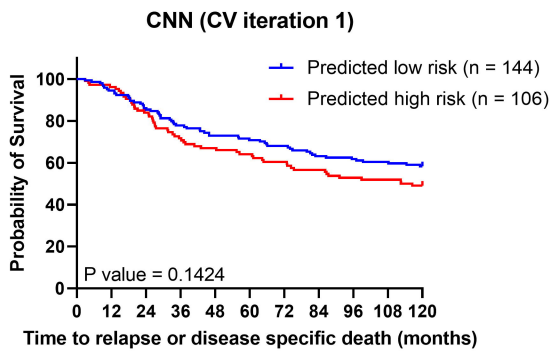
**B**



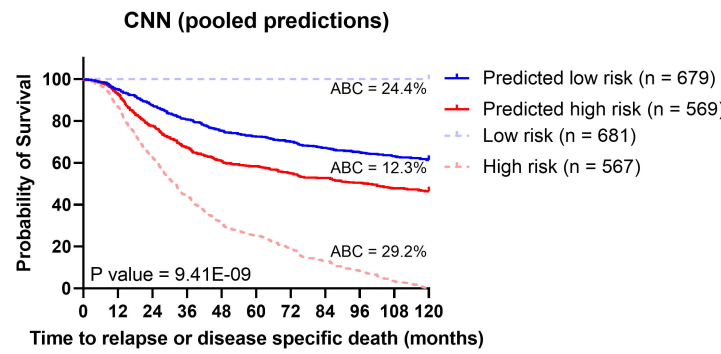
**C**



**D**

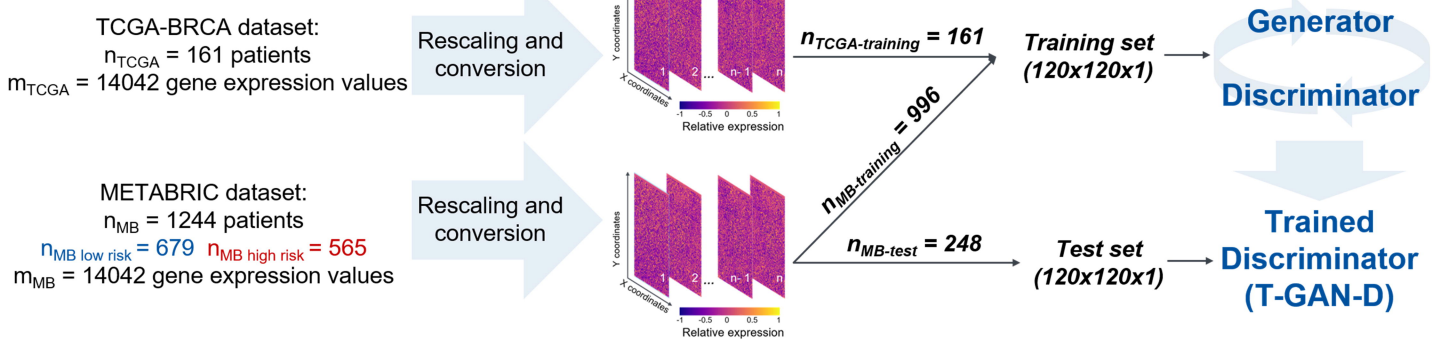


**E**

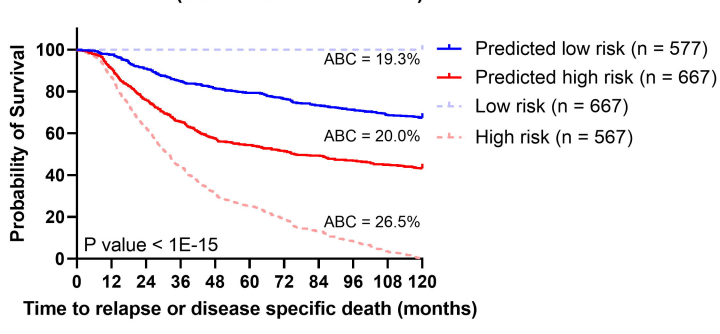


**Fig. 3**

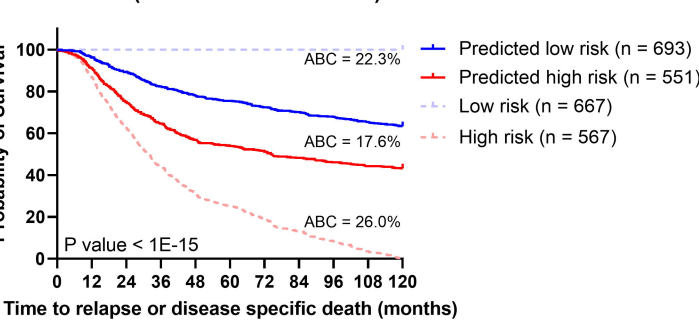
**A**



**B**



**C**



**Fig. 4**

