



<b>Title</b>	Multi-Gene Prognostic Signatures and Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in ER-Positive, HER2-Negative Breast Cancer Patients
<b>Authors(s)</b>	Mazo, Claudia, Barron, Stephen, Mooney, Catherine, Gallagher, William M.
<b>Publication date</b>	2020-05-01
<b>Publication information</b>	Mazo, Claudia, Stephen Barron, Catherine Mooney, and William M. Gallagher. "Multi-Gene Prognostic Signatures and Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in ER-Positive, HER2-Negative Breast Cancer Patients." MDPI, May 1, 2020. <a href="https://doi.org/10.3390/cancers12051133">https://doi.org/10.3390/cancers12051133</a> .
<b>Publisher</b>	MDPI
<b>Item record/more information</b>	<a href="http://hdl.handle.net/10197/12179">http://hdl.handle.net/10197/12179</a>
<b>Publisher's statement</b>	This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited
<b>Publisher's version (DOI)</b>	<a href="https://doi.org/10.3390/cancers12051133">10.3390/cancers12051133</a>

Downloaded 2026-05-01 23:34:21

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa)



© Some rights reserved. For more information

Article

# Multi-Gene Prognostic Signatures and Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in ER-Positive, HER2-Negative Breast Cancer Patients

Claudia Mazo <sup>1,2,3</sup> , Stephen Barron <sup>3</sup>, Catherine Mooney <sup>1,†</sup>   
and William M. Gallagher <sup>3,4,\*</sup> 

<sup>1</sup> UCD School of Computer Science, University College Dublin, Dublin 4, Ireland; claudia.mazovargas@ucd.ie (C.M.); catherine.mooney@ucd.ie (C.M.)

<sup>2</sup> CeADAR: Centre for Applied Data Analytics Research, University College Dublin, Dublin 4, Ireland

<sup>3</sup> OncoMark Limited, NovaUCD, Dublin 4, Ireland; stephen.barron@oncomark.com

<sup>4</sup> UCD School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Dublin 4, Ireland

\* Correspondence: william.gallagher@ucd.ie

† These authors contributed equally to this work.

Received: 29 February 2020; Accepted: 27 April 2020; Published: 1 May 2020



**Abstract:** Determining which patients with early-stage breast cancer should receive chemotherapy is an important clinical issue. Chemotherapy has several adverse side effects, impacting on quality of life, along with significant economic consequences. There are a number of multi-gene prognostic signatures for breast cancer recurrence but there is less evidence that these prognostic signatures are predictive of therapy benefit. Biomarkers that can predict patient response to chemotherapy can help avoid ineffective over-treatment. The aim of this work was to assess if the OncoMasTR prognostic signature can predict pathological complete response (pCR) to neoadjuvant chemotherapy, and to compare its predictive value with other prognostic signatures: EndoPredict, Oncotype DX and Tumor Infiltrating Leukocytes. Gene expression datasets from ER-positive, HER2-negative breast cancer patients that had pre-treatment biopsies, received neoadjuvant chemotherapy and an assessment of pCR were obtained from the Gene Expression Omnibus repository. A total of 813 patients with 66 pCR events were included in the analysis. OncoMasTR, EndoPredict, Oncotype DX and Tumor Infiltrating Leukocytes numeric risk scores were approximated by applying the gene coefficients to the corresponding mean probe expression values. OncoMasTR, EndoPredict and Oncotype DX prognostic scores were moderately well correlated according to the Pearson's correlation coefficient. Association with pCR was estimated using logistic regression. The odds ratio for a 1 standard deviation increase in risk score, adjusted for cohort, were similar in magnitude for all four signatures. Additionally, the four signatures were significant predictors of pCR. OncoMasTR added significant predictive value to Tumor Infiltrating Leukocytes signatures as determined by bivariable and trivariable analysis. In this *in silico* analysis, OncoMasTR, EndoPredict, Oncotype DX, and Tumor Infiltrating Leukocytes were significantly predictive of pCR to neoadjuvant chemotherapy in ER-positive and HER2-negative breast cancer patients.

**Keywords:** breast cancer; multi-gene prognostic signature; neoadjuvant chemotherapy; breast cancer treatment; pathological complete response

## 1. Introduction

Breast cancer is the most frequently diagnosed cancer in women, with more than 2.1 million new diagnoses worldwide every year, and the second leading cause of cancer death. The American Cancer Society and the International Agency for Research on Cancer reported that mortality rates are on the decline in certain regions of the world as a result of earlier diagnosis and improved therapies [1,2]. Accurate tools to help with optimal treatment decisions for individual patients to improve their prognosis, survival and quality of life are needed, whilst also reducing associated healthcare costs [3]. Breast cancer patients who receive chemotherapy can experience several side effects and symptoms that have a negative effect on their quality of life during and after the completion of treatment [4]. Biomarkers that can predict patient response to chemotherapy can help identify which patients are likely to benefit from chemotherapy, thereby potentially reducing the adverse effects of over-treatment.

Multi-gene prognostic signatures may be used to estimate risk of recurrence following surgery and endocrine treatment to make decisions about the suitability of chemotherapy. A patient who is predicted to be at low risk of breast cancer recurrence can safely forego aggressive treatment plans such as chemotherapy. OncoMasTR (OM) is a new 6-gene assay (3 prognostic genes plus 3 reference genes) discovered using a novel transcriptional network analysis approach that identified genes—Master Transcriptional Regulators (MTRs)—that putatively regulate previously known prognostic signatures [5,6]. OM has been analytically validated [7] in terms of assay robustness and clinically validated [8,9] in terms of accurate risk stratification, providing significant additional prognostic value to clinical information. The Oncotype DX Recurrence Score (RS) is a 21-gene prognostic assay that is widely used to predict risk of recurrence [10]. The EndoPredict assay (EP) is a 12-gene prognostic assay which incorporates tumor size and nodal status [11]. The Tumor Infiltrating Leukocytes (TILs) signature [12] determines risk of recurrence using expression correlates of 60 genes representing 15 immune cell sub-populations. Elevated TILs levels have been associated with better survival in patients with either ER-positive, HER2-positive disease or Triple-Negative Breast Cancer (TNBC) [13–17]. Furthermore, high TILs levels are associated with increased sensitivity to chemotherapy, reflected by higher pCR rates to neoadjuvant chemotherapy [18,19]. Indeed, a critical question remains as to whether one can predict a particular benefit from chemotherapy, as opposed to solely prognostic information, in breast cancer patients using multi-gene signatures.

The focus of this work is on ER-positive, HER2-negative breast cancer patients as the pCR to neoadjuvant chemotherapy is normally quite low in these patients compared to other breast cancer subtypes. We present an approach to expand on previous *in silico* analyses of other multi-gene prognostic signatures [20]. We assessed if OM could predict pCR of early-stage breast cancer patients to neoadjuvant chemotherapy and subsequently compared its predictive performance with EP, RS and TILs. Additionally, we assessed if TILs added significant predictive value to RS, EP and OM using univariable, bivariable and trivariable logistic regression analysis.

## 2. Method

First, a systematic search was carried out to select the datasets for this study. Following this, the relationship between OM, RS, EP, and TILs standardised risk scores was assessed by the Pearson's correlation coefficient [21,22]. Finally, logistic regression [23] was used to estimate the associations between the signatures and pCR.

### 2.1. Systematic Search

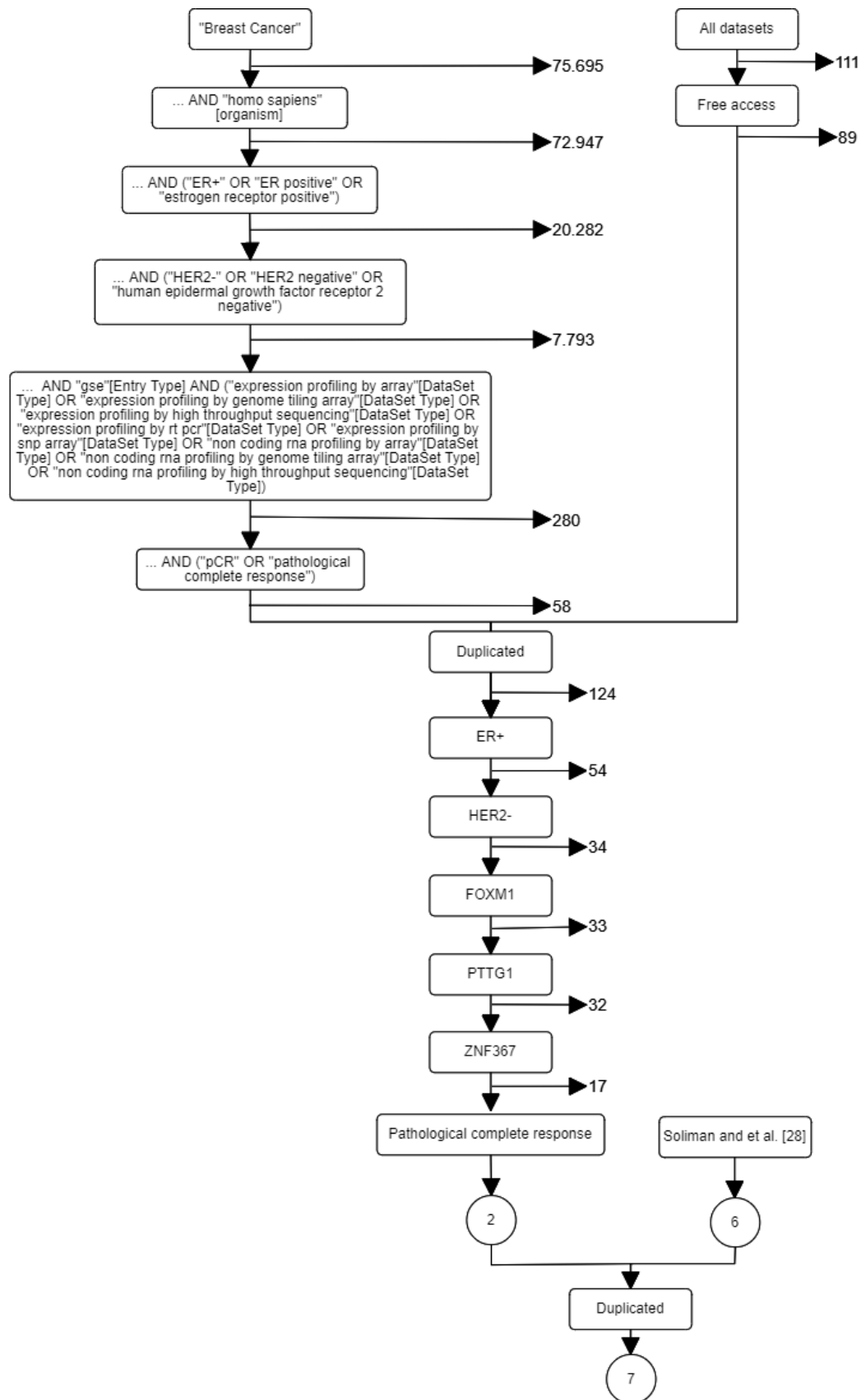
The following sources were used to identify relevant breast cancer datasets: (i) dataset engines such as the Gene Expression Omnibus (GEO) repository [24], ArrayExpress [25], GDC data portal [26] and EGAS [27]; (ii) related papers and similar studies, as well as peer reviewed papers that performed a systematic search for gene expression datasets in breast cancer.

After the resources were specified, the following key words were selected: breast cancer, early, *Homo sapiens*, ER+, and HER2-. Particularly in GEO, there are many keywords and hierarchical orders that may be used in a systematic search resulting in different results. Furthermore, including the synonyms of the keywords or GEO's query fields may alter the results obtained. Additional constraints were applied according to our project's need: (i) including FOXM1, PTTG1, and ZNF367 genes; (ii) endpoint or outcome (e.g., pCR, distant recurrence, death, response to therapy); (iii) number of endpoint events; (iv) lymph node status; (v) therapy/treatment (e.g., none, endocrine, chemotherapy) (vi) platform; and (vii) public access. Figure 1 shows the flow diagram according to the project's objectives after different search criteria. Finally, joining the results from the systematic search and the reference-work presented in [28], seven datasets (GSE16716, GSE20271, GSE25066, GSE32646, GSE34138, GSE41998, and GSE22226) were identified (see Table 1).

**Table 1.** GEO datasets included in the analysis \*.

GEO Dataset	Platform	Patients (N)	pCR (N)	Missing Genes			
				EP	RS	TILs	OM
GSE16716	Affymetrix Human Genome U133A Array	140	7			PTRPC, KLRK1, EOMES, KIR3DL2, XCL2, CD8B	ZNF367
GSE20271	Affymetrix Human Genome U133A Array	89	6			PTRPC, KLRK1, EOMES, KIR3DL2, XCL2, CD8B	ZNF367
GSE25066	Affymetrix Human Genome U133A Array	278	30			PTRPC, KLRK1, EOMES, KIR3DL2, XCL2, CD8B	ZNF367
GSE32646	Affymetrix Human Genome U133A Plus 2.0 Array	55	5			PTRPC, KLRK1, KIR3DL2, XCL2	
GSE34138	Illumina Human WG 6 v3.0 expression bead chip	119	4		MYBL2	PTRPC, KLRK1, TPSB2, XCL2, NCR1, FOXP3	ZNF367
GSE41998	Affymetrix Human U133A 2.0 Array	93	10			PTRPC, KLRK1, EOMES, KIR3DL2, XCL2, CD8B	ZNF367
GSE22226 GPL1708	Agilent 012391 Whole Human Genome Oligo Microarray G4112A (Feature Number version)	39	4	CCNB1	MYBL2	PTRPC, EOMES, TPSB2, TPSB1, MS4A2, KIR3DL2, CD3E	
		813	66				

\* OM corresponds to the OncoMasTR score, RS corresponds to the Oncotype DX Recurrence Score, EP corresponds to the EndoPredict score and TILs corresponds to the Tumor Infiltrating Leukocytes signature.



**Figure 1.** A systematic search for breast cancer datasets. Left path corresponds to GEO search. Right path corresponds to related papers search.

## 2.2. Dataset

Seven GEO dataset were used—GSE16716, GSE20271, GSE25066, GSE32646, GSE34138, GSE41998, and GSE22226. OM, EP and RS numeric risk scores were approximated by applying the signatures' gene coefficients to the mean of the corresponding probe expression values. TILs scores is computed as the simple average *log* expression of their marker genes. Risk scores were standardised within each dataset to have a mean of 0 and a standard deviation of 1. This allowed us to compare different risk scores from different datasets on a similar scale. A total of 813 patients with 66 pCR events (8.1%) were analysed. All the datasets except GSE32646 were missing at least one gene from at least one signature (see Table 1).

## 2.3. Predicting Probability of pCR Using Logistic Regression Analysis

Logistic regression is a statistical model that uses a logistic function to model a binary dependent variable. Logistic regression is used for different scenarios; in this case to estimate the probability that an event will occur, the outcome, using information thought to influence or be related to such events called predictors [23]. The outcome is a binary variable (1/0, Yes/No, True/False). The odds are defined as the probability that the event will occur divided by the probability that the event will not occur. Odds > 1 shows an association or correlation between risk scores and pCR, in our case. The 95% Confidence Interval (CI) is used to estimate the precision of the odds. Additionally, a *p*-value is used as probability measure. The smaller the *p*-value (*p*) is the more significant the result is considered to be. 0.05 is chosen as an arbitrary threshold where *p* > 0.05 is considered not to be significant. Univariable analysis was used to examine the relationship between pCR and each risk score while multivariable analysis was used to examine the relationship between pCR and each risk score, adjusted for cohort/GSE dataset or signature.

## 3. Results

### 3.1. Correlation Analysis

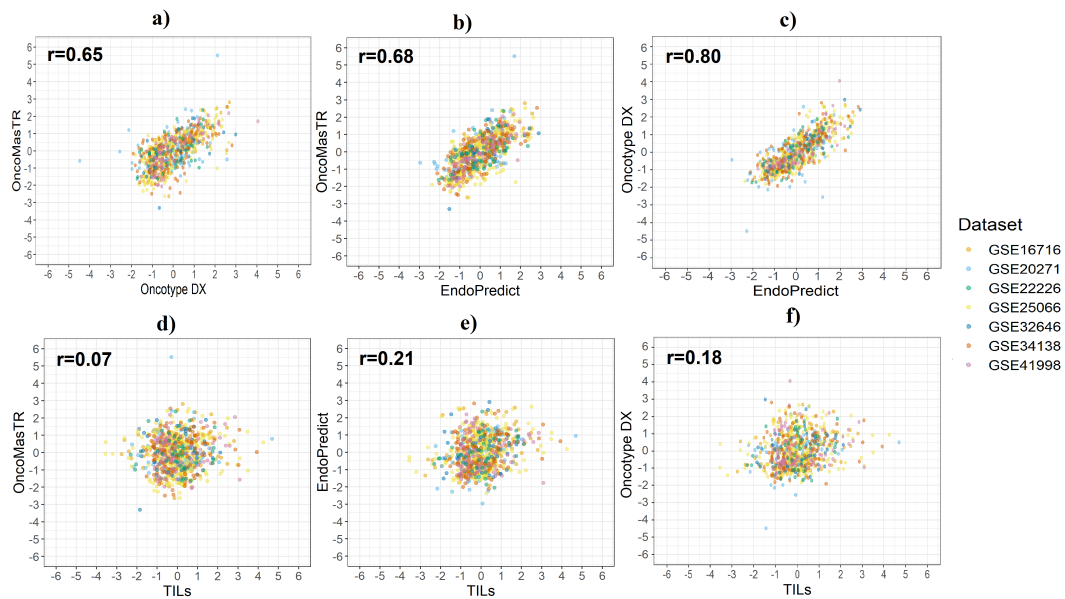
A Pearson's correlation coefficient was used to assess the linear relationship between the standardised scores from the four signatures. We evaluated the correlation between six pairs of the four signatures: OM vs. RS, OM vs. EP, RS vs. EP, TILs vs. OM, TILs vs. EP, and TILs vs. RS for the seven datasets combined (Table 2).

**Table 2.** Correlation coefficient *r* among the four signatures \* \*\*.

Signatures	Overall <i>r</i>	Lowest <i>r</i> (Dataset)	Highest <i>r</i> (Dataset)
OM vs. RS	0.65	0.34 (GSE20271)	0.79 (GSE41998)
OM vs. EP	0.68	0.44 (GSE20271)	0.82 (GSE41998)
RS vs. EP	0.80	0.55 (GSE20271)	0.90 (GSE32646)
OM vs. TILs	0.07	0.01 (GSE20271)	0.17 (GSE32646)
EP vs. TILs	0.21	0.09 (GSE41998)	0.33 (GSE20271)
RS vs. TILs	0.18	0.12 (GSE41998)	0.28 (GSE20271)

\* Overall corresponds to the average of the seven datasets; Lowest and Highest correspond to the minimum and maximum, respectively. \*\* OM corresponds to the OncoMasTR score, RS corresponds to the Oncotype DX Recurrence Score, EP corresponds to the EndoPredict score and TILs corresponds to the Tumor Infiltrating Leukocytes signature.

OM, RS and EP were moderately well correlated. However, there was very low correlation between TILs and the other three signatures. The GSE20271 dataset provided the lowest correlation in most cases. Figure 2 shows the results of the correlation analysis between risk score in the seven datasets combined.



**Figure 2.** Correlation between risk scores by dataset. (a) OM vs. RS; (b) OM vs. EP; (c) RS vs. EP; (d) OM vs. TILs; (e) EP vs. TILs; (f) RS vs. TILs.

### 3.2. Predicted Probability of pCR

Univariable, bivariable and trivariable logistic regression was used to estimate the relationship of OM, RS, EP and TILs with pCR. Table 3 contains a summary of the odds ratio (for a 1 standard deviation in risk score), 95% confidence intervals, and p-values for the relationship between risk scores and pCR.

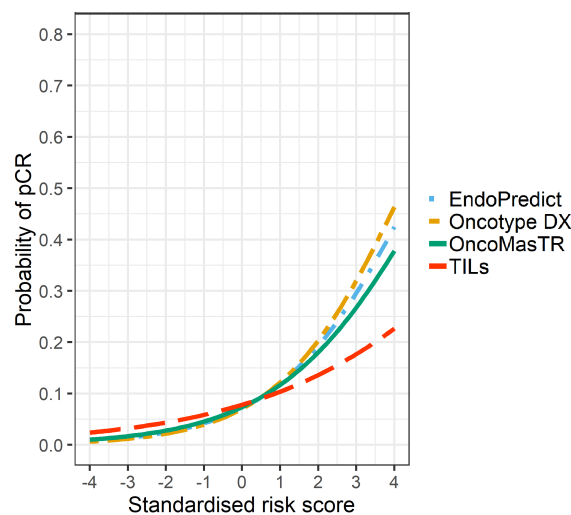
Table 3 shows that the odds ratios of the signatures were similar in magnitude. OM, RS, EP and TILs were significant predictors of pCR, with TILs being less predictive. OM had an odds ratio of 1.66 with a 95% confidence interval of 1.29 to 2.16. EP had an odds ratio of 1.76 with a 95% confidence interval of 1.37 to 2.27. RS had an odds ratio of 1.84 with 95% confidence interval of 1.44 to 2.35. TILs had an odds ratio of 1.36 with a 95% confidence interval of 1.07 to 1.72.

The univariable odds ratios for OM, RS, EP and TILs were similar when controlled for dataset and/or other signatures (bivariable and trivariable models). Accordingly, we based our predicted probabilities on univariable odds ratio for clarity of presentation. Plots of the predicted probability (Figures 3–5) are derived from the univariable odds ratios in Table 3 and show that the probabilities of pCR increase as the risk scores increase, indicating that the risk scores predict response to chemotherapy. Figure 4 shows that signatures with fewer genes are more likely to be impacted by missing genes. As hypothesised, the predictive performance of OM is better in datasets with complete OM genes (Figure 5).

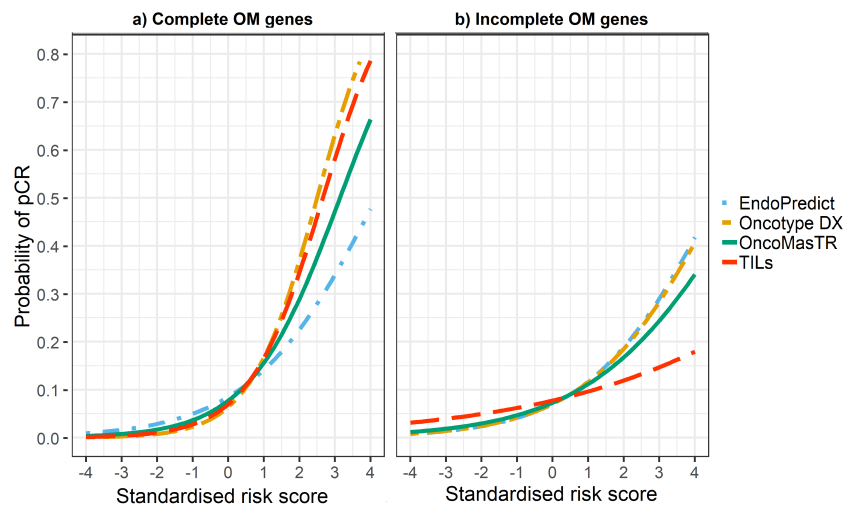
**Table 3.** Odds ratio (95% confidence intervals) for pCR by risk score \* \*\*.

Signature	Odds Ratio (95% CI)	p-Value	Model
Univariable Analysis			
OM	1.66 (1.29–2.16)	0.0001	OM
RS	1.84 (1.44–2.35)	<0.0001	RS
EP	1.76 (1.37–2.27)	<0.0001	EP
TILs	1.36 (1.07–1.72)	0.0099	TILs
Bivariable Analysis (adjusted for dataset or TILs)			
OM	1.68 (1.30–2.18)	<0.0001	OM + Dataset
RS	1.85 (1.45–2.37)	<0.0001	RS + Dataset
EP	1.77 (1.37–2.30)	<0.0001	EP + Dataset
TILs	1.37 (1.08–1.73)	0.0095	TILs + Dataset
OM	1.63 (1.26–2.12)	0.0002	OM + TILs
TILs	1.32 (1.04–1.66)	0.0226	
RS	1.80 (1.40–2.31)	<0.0001	RS + TILs
TILs	1.26 (0.98–1.60)	0.0630	
EP	1.69 (1.31–2.20)	<0.0001	EP + TILs
TILs	1.23 (0.96–1.56)	0.0886	
Trivariable Analysis (adjusted for dataset and TILs)			
OM	1.65 (1.28–2.14)	0.0002	OM + TILs + Dataset
TILs	1.33 (1.04–1.69)	0.0199	
RS	1.81 (1.41–2.33)	<0.0001	RS + TILs + Dataset
TILs	1.27 (0.99 – 1.62)	0.0546	
EP	1.71 (1.32–2.23)	<0.0001	EP + TILs + Dataset
TILs	1.25 (0.97–1.59)	0.0748	

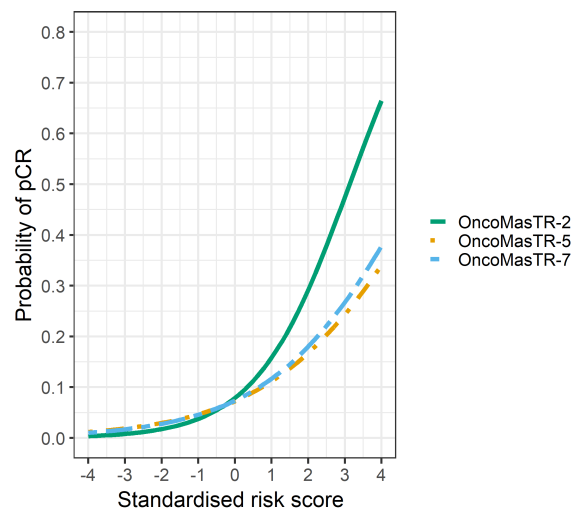
\* Odds ratio is for a 1 standard deviation increase in risk score. \*\* OM corresponds to the OncoMasTR score, RS corresponds to the Oncotype DX Recurrence Score, EP corresponds to the EndoPredict score and TILs corresponds to the Tumor Infiltrating Leukocytes signature.



**Figure 3.** Predicted probability of pCR by risk score in 7 datasets (5 datasets missing ZNF367).



**Figure 4.** Predicted probability of pCR by risk score in (a) datasets with complete OM genes (2 datasets); (b) datasets with incomplete OM genes (5 datasets missing ZNF367).



**Figure 5.** Predicted probability of pCR by OM risk score using different dataset groups. Green line complete OM genes (2 datasets); Orange line incomplete OM genes (5 datasets missing ZNF367); Blue line combined datasets (7 datasets).

Residual deviance was used to assess how much TILs added to the other three risk scores. Residual deviance is a measure of goodness of fit of a model, a lower deviance means a better fit model and a 0 deviance is a perfect model. TILs added significant predictive value if the residual deviance decreased by  $>3.84$  when TILs was added to a model ( $p(\text{chi-sq} > 3.84, df = 1) < 0.05$ ) (see Table 4).

TILs added significant predictive value to OM and vice versa and the significant predictive values are increased when the model is adjusted by dataset. TILs does not add significant predictive value to RS and EP ( $p = 0.0630$  and  $0.0886$  respectively). However, RS and EP added significant predictive value to TILs. Controlling for the differences between the datasets further decrease the residual deviance of OM + TILs, RS + TILs and EP + TILs. The results show that the models adjusted for dataset are a better fit than unadjusted models. Finally, the best fit model was RS + TILs + Dataset with a residual deviance of 419.60.

**Table 4.** Deviance statistic by model \*.

Model	Null Deviance	df Null	LogLik	AIC	BIC	Deviance	df Residual
Univariable Analysis							
OM	457.95	812	−221.16	446.33	455.73	442.33	811
RS	457.95	812	−216.89	437.79	447.19	433.79	811
EP	457.95	812	−219.27	442.53	451.93	438.53	811
TILs	457.95	812	−225.80	455.59	464.99	451.59	811
Bivariable Analysis (adjusted for dataset or TILs)							
OM + Dataset	457.95	812	−215.84	447.68	485.29	431.68	805
RS + Dataset	457.95	812	−211.59	439.18	476.78	423.18	805
EP + Dataset	457.95	812	−213.94	443.89	481.49	427.89	805
TILs + Dataset	457.95	812	−220.55	457.09	494.70	441.09	805
OM + TILs	457.95	812	−218.65	443.30	457.41	437.30	810
RS + TILs	457.95	812	−215.22	436.44	450.55	430.44	810
EP + TILs	457.95	812	−217.86	441.71	455.82	435.71	810
Trivariable Analysis (adjusted for dataset and TILs)							
OM + TILs + Dataset	457.95	812	−213.21	444.43	486.73	426.43	804
RS + TILs + Dataset	457.95	812	−209.80	437.60	479.90	419.60	804
EP + TILs + Dataset	457.95	812	−212.40	442.80	485.11	424.80	804

\* OM corresponds to the OncoMasTR score, RS corresponds to the Oncotype DX Recurrence Score, EP corresponds to the EndoPredict score and TILs corresponds to the Tumor Infiltrating Leukocytes signature.

#### 4. Discussion

Multi-gene expression prognostic assays are commonly used to aid clinical decision-making in early-stage ER-positive, HER2-negative breast cancer as they provide complementary prognostic information to clinico-pathologic features [3]. There are many assays available, including RS, EP, TILs, MammaPrint [29], PAM50 [30], Breast Cancer Index (BCI) [31], along with the IHC4 test that is immunohistochemically based [32]. The amount of prognostic information provided in relation to risk of recurrence at an early (0–5 years) and late (beyond five years) time-point varies across these tests [33]. While the majority of data regarding these signatures focuses on prognostic value in the context of a background of adjuvant hormone therapy, there is comparatively little data available relating to their use in predicting benefit to neoadjuvant chemotherapy.

In this paper, we evaluated a new multi-gene prognostic signature, OM. We compared three other multi-gene expression prognostic signatures (RS, EP and TILs) with OM and evaluated the correlation between the four signatures (OM, RS, EP and TILs). Furthermore, we assessed if OM can predict pCR to neoadjuvant chemotherapy, and we compared the OM predictive performance with RS, EP and TILs. Finally, we evaluated if TILs added significant predictive value to OM, RS and EP.

Our findings suggest that OM, RS, and EP are moderately well correlated across the seven datasets assessed, with an average Pearson's correlation coefficient  $r \geq 0.65$  between each pair of signatures (OM vs RS, OM vs EP, and RS vs EP). TILs demonstrated a lower correlation for all datasets with an average  $r \leq 0.21$ . Our results suggest that OM, RS, EP and TILs were significant predictors of pCR to neoadjuvant chemotherapy in ER+, HER2- breast cancer, with odds ratios  $\geq 1.36$ .

TILs added significant predictive value to OM with a residual deviance decrease. OM, RS and EP added significant predictive value to TILs and controlling for the differences between the datasets further decreased the residual deviance. The best fit model was RS+TILs+Dataset demonstrating that including additional features provide a better fit to the data.

However, the predicted probability of pCR of the four signatures is affected by missing genes. It is important to highlight that OM is a significant pCR predictor even when only two of its prognostic genes are used (excluding missing genes). Our hypothesis is that if more datasets with complete genes were available then the performance of OM could be improved.

## 5. Conclusions and Future Work

OM has been clinically validated in the TransATAC cohort and a subset of the TAILORx cohort and has shown superior prognostic performance to RS [7–9]. Here, we show that the signature is predictive of treatment response (i.e., pCR response to neoadjuvant chemotherapy treatment). Some of the main differences between OM, RS, EP, and TILs are the number of genes in the assay (60–genes TILs, 21–genes RS, 12–genes EP, 6–genes OM, when normalization genes are included) and the resulting cost. Some limitations of this study are the approximation of risk scores in micro-array data, the small number of pCR events in the dataset, and the missing genes effect.

Biomarkers that predict patient response to neoadjuvant chemotherapy offer the opportunity for personalised care, improved therapy response rates, reduced adverse effects and decreased costs of unnecessary treatment. In the future, we will extend this work by following two lines of investigation: (i) using more datasets thereby increasing the number of subjects evaluated and reducing the number of missing genes; and (ii) exploring new techniques which combine images with genetic information.

**Author Contributions:** Conceptualization, C.M. (Claudia Mazo), S.B., C.M. (Catherine Mooney), and W.M.G.; Data curation, C.M. (Claudia Mazo), S.B., C.M. (Catherine Mooney), and W.M.G.; Formal analysis, C.M. (Claudia Mazo), and S.B.; Funding acquisition, C.M. (Claudia Mazo), C.M. (Catherine Mooney), and W.M.G.; Investigation, C.M. (Claudia Mazo), and S.B.; Methodology, C.M. (Claudia Mazo) and S.B.; Project administration, C.M. (Claudia Mazo), C.M. (Catherine Mooney), and W.M.G.; Resources, C.M. (Claudia Mazo), S.B., C.M. (Catherine Mooney), and W.M.G.; Supervision, S.B., C.M. (Catherine Mooney), and W.M.G.; Validation, C.M. (Claudia Mazo), S.B., C.M. (Catherine Mooney), and W.M.G.; Writing—original draft, C.M. (Claudia Mazo); Writing—review and editing, C.M. (Claudia Mazo), S.B., C.M. (Catherine Mooney), and W.M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from Enterprise Ireland (EI) and from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713654. This publication has been supported in part by TACTIC—Tailoring the Communication of risk To Individual breast Cancer patients, Irish Research Council (IRC) grant CLNE/2018/1338, CAROLINE Postdoctoral Fellowship. This publication emanated from research supported in part by research grants from Science Foundation Ireland (SFI) under Grant Numbers 15/IA/3104 and 18/SPP/3522. This material is also based upon works supported by the Irish Cancer Society Collaborative Cancer Research Centre BREAST-PREDICT Grant CCRC13GAL.

**Conflicts of Interest:** William M. Gallagher (W.M.G.) is Chief Scientific Officer, shareholder and Director of OncoMark, Stephen Barron (S.B.) is an employee of OncoMark.

## Abbreviations

The following abbreviations are used in this manuscript:

BCI	Breast Cancer Index
CI	Confidence Interval
EP	EndoPredict
GEO	Gene Expression Omnibus
MTRs	Master Transcriptional Regulators
OM	OncoMasTR
pCR	Pathological Complete Response
RS	Oncotype DX
TILs	Tumor Infiltrating Lymphocytes
EI	Enterprise Ireland
IRC	Irish Research Council
SFI	Science Foundation Ireland

## References

1. American Cancer Society (ACS). 2019. Available online: <https://www.cancer.org/cancer/breast-cancer.html> (accessed on 26 November 2019).
2. International Agency for Research on Cancer (IARC). 2019. Available online: <https://www.iarc.fr/> (accessed on 26 November 2019).

3. Mazo, C.; Kearns, C.; Mooney, C.; Gallagher, W.M. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers* **2020**, *12*, 369. [[CrossRef](#)] [[PubMed](#)]
4. Paraskevi, T. Quality of life outcomes in patients with breast cancer. *Oncol. Rev.* **2012**, *6*, e2. doi:10.4081/oncol.2012.e2. [[CrossRef](#)] [[PubMed](#)]
5. Lanigan, F.; Brien, G.L.; Fan, Y.; Madden, S.F.; Jerman, E.; Maratha, A.; Aloraifi, F.; Hokamp, K.; Dunne, E.J.; Lohan, A.J.; et al. Delineating transcriptional networks of prognostic gene signatures refines treatment recommendations for lymph node-negative breast cancer patients. *FEBS J.* **2015**, *282*, 3455–3473. [[CrossRef](#)] [[PubMed](#)]
6. Moran, B.; Rahman, A.; Palonen, K.; Lanigan, F.T.; Gallagher, W.M. Master transcriptional regulators in cancer: Discovery via reverse engineering approaches and subsequent validation. *Cancer Res.* **2017**, *77*, 2186–2190. [[CrossRef](#)]
7. Loughman, T.; Chan-Ju Wang, A.; Dynoodt, P.; Fender, B.; Lopez Ruiz, C.; Barron, S.; Stapleton, S.; O’Leary, D.; Fabre, A.; Quinn, C.; et al. 204P Analytical validation of OncoMasTR, a multigene test for predicting risk of distant recurrence in hormone receptor-positive early stage breast cancer. *Ann. Oncol.* **2018**, *29*, doi:10.1093/annonc/mdy270. [[CrossRef](#)]
8. Buus, R.; Sestak, I.; Barron, S.; Loughman, T.; Fender, B.; Ruiz, C.L.; Dynoodt, P.; Wang, C.J.A.; O’Leary, D.; Gallagher, W.M.; et al. Validation of the OncoMasTR Risk Score in Estrogen Receptor-Positive/HER2-Negative Patients: A TransATAC study. *Clin. Cancer Res.* **2020**, *26*, 623–631. [[CrossRef](#)]
9. O’Connor, D.; Kelly, C.M.; Crown, J.; Russell, N.; Barron, S.; Loughman, T.; Lynch, S.; O’Grady, A.; Sheehan, K.M.; Fay, J.; et al. Additional prognostic value of OncoMasTR multigene prognostic signature to clinicopathological information in patients with HR-positive, HER2-negative, lymph node-negative breast cancer from the TAILORx Tissue Bank, Ireland. *J. Clin. Oncol.* **2019**, *37*, 535–535. [[CrossRef](#)]
10. Paik, S.; Shak, S.; Tang, G.; Kim, C.; Baker, J.; Cronin, M.; Baehner, F.L.; Walker, M.G.; Watson, D.; Park, T.; et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **2004**, *351*, 2817–2826. [[CrossRef](#)]
11. Filipits, M.; Rudas, M.; Jakesz, R.; Dubsy, P.; Fitzal, F.; Singer, C.F.; Dietze, O.; Greil, R.; Jelen, A.; Sevelde, P.; et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **2011**, *17*, 6012–6020. [[CrossRef](#)]
12. Danaher, P.; Warren, S.; Dennis, L.; D’Amico, L.; White, A.; Disis, M.L.; Geller, M.A.; Odunsi, K.; Beechem, J.; Fling, S.P. Gene expression markers of tumor infiltrating leukocytes. *J. Immunother. Cancer* **2017**, *5*, 18. [[CrossRef](#)]
13. Bianchini, G.; Qi, Y.; Alvarez, R.H.; Iwamoto, T.; Coutant, C.; Ibrahim, N.K.; Valero, V.; Cristofanilli, M.; Green, M.C.; Radvanyi, L.; et al. Molecular Anatomy of Breast Cancer Stroma and Its Prognostic Value in Estrogen Receptor-Positive and -Negative Cancers. *J. Clin. Oncol.* **2010**, *28*, 4316–4323. doi:10.1200/JCO.2009.27.2419. [[CrossRef](#)] [[PubMed](#)]
14. Karn, T.; Pusztai, L.; Holtrich, U.; Iwamoto, T.; Shiang, C.Y.; Schmidt, M.; Miller, V.; Solbach, C.; Gaetje, R.; Hanker, L.; et al. Homogeneous Datasets of Triple Negative Breast Cancers Enable the Identification of Novel Prognostic and Predictive Signatures. *PLoS ONE* **2011**, *6*, e0028403. doi:10.1371/journal.pone.0028403. [[CrossRef](#)] [[PubMed](#)]
15. Adams, S.; Gray, R.J.; Demaria, S.; Goldstein, L.; Perez, E.A.; Shulman, L.N.; Martino, S.; Wang, M.; Jones, V.E.; Saphner, T.J.; et al. Prognostic Value of Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancers From Two Phase III Randomized Adjuvant Breast Cancer Trials: ECOG 2197 and ECOG 1199. *J. Clin. Oncol.* **2014**, *32*, 2959–2966. doi:10.1200/JCO.2013.55.0491. [[CrossRef](#)] [[PubMed](#)]
16. Loi, S.; Sirtaine, N.; Piette, F.; Salgado, R.; Viale, G.; Van Eenoo, F.; Rouas, G.; Francis, P.; Crown, J.P.; Hitre, E.; et al. Prognostic and Predictive Value of Tumor-Infiltrating Lymphocytes in a Phase III Randomized Adjuvant Breast Cancer Trial in Node-Positive Breast Cancer Comparing the Addition of Docetaxel to Doxorubicin With Doxorubicin-Based Chemotherapy: BIG 02-98. *J. Clin. Oncol.* **2013**, *31*, 860–867. doi:10.1200/JCO.2011.41.0902. [[CrossRef](#)] [[PubMed](#)]
17. Iwamoto, T.; Bianchini, G.; Booser, D.; Qi, Y.; Coutant, C.; Ya-Hui Shiang, C.; Santarpia, L.; Matsuoka, J.; Hortobagyi, G.N.; Symmans, W.F.; et al. Gene Pathways Associated With Prognosis and Chemotherapy Sensitivity in Molecular Subtypes of Breast Cancer. *JNCI: J. Natl. Cancer Inst.* **2010**, *103*, 264–272. doi:10.1093/jnci/djq524. [[CrossRef](#)] [[PubMed](#)]

18. Denkert, C.; Von Minckwitz, G.; Brase, J.C.; Sinn, B.V.; Gade, S.; Kronenwett, R.; Pfitzner, B.M.; Salat, C.; Loi, S.; Schmitt, W.D.; et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J. Clin. Oncol.* **2015**, *33*, 983–991. [[CrossRef](#)] [[PubMed](#)]
19. Wimberly, H.; Brown, J.R.; Schalper, K.; Haack, H.; Silver, M.R.; Nixon, C.; Bossuyt, V.; Pusztai, L.; Lannin, D.R.; Rimm, D.L. PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. *Cancer Immunol. Res.* **2015**, *3*, 326–332. [[CrossRef](#)]
20. Mazo, C.; Barron, S.; Mooney, C.; Gallagher, W. 257P Multi-gene prognostic signatures and prediction of pathological complete response of ER-Positive HER2-negative breast cancer patients to neo-adjuvant chemotherapy. *Ann. Oncol.* **2019**, *30*, doi:10.1093/annonc/mdz240.081. [[CrossRef](#)]
21. Altman, D.G. *Practical Statistics for Medical Research*; CRC Press: Boca Raton, FL, USA, 1990.
22. Campbell, M.J.; Swinscow, T.D.V. *Statistics at Square One*; John Wiley & Sons: New York, NY, USA, 2011.
23. Kirkwood, B.R.; Sterne, J.A. *Essential Medical Statistics*; John Wiley & Sons: New York, NY, USA, 2010.
24. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
25. Kolesnikov, N.; Hastings, E.; Keays, M.; Melnichuk, O.; Tang, Y.A.; Williams, E.; Dylag, M.; Kurbatova, N.; Brandizi, M.; Burdett, T.; et al. ArrayExpress update—Simplifying data submissions. *Nucleic Acids Res.* **2015**, *43*, D1113–D1116. [[CrossRef](#)]
26. Zhang, J.; Baran, J.; Cros, A.; Guberman, J.M.; Haider, S.; Hsu, J.; Liang, Y.; Rivkin, E.; Wang, J.; Whitty, B.; et al. International Cancer Genome Consortium Data Portal—A one-stop shop for cancer genomics data. *Database* **2011**, doi:10.1093/database/bar026. [[CrossRef](#)] [[PubMed](#)]
27. Lappalainen, I.; Almeida-King, J.; Kumanduri, V.; Senf, A.; Spalding, J.D.; Saunders, G.; Kandasamy, J.; Caccamo, M.; Leinonen, R.; Vaughan, B.; et al. The European Genome-phenome Archive of human data consented for biomedical research. *Nature Genet.* **2015**, *47*, 692–695. [[CrossRef](#)] [[PubMed](#)]
28. Soliman, H.H.; Robson, M.E.; Wagner, S.; Flake, D.D.; Schwartzberg, L.S.; Sharma, P.; Magliocco, A.M.; Kronenwett, R.; Lancaster, J.M.; Lanchbury, J.S.; et al. In silico evaluation of the 12-gene molecular score (EndoPredict) and the recurrence score (Oncotype DX) as predictors of response to neo-adjuvant chemotherapy in estrogen receptor positive (ER+), HER2 negative (HER2-) breast cancer. *J. Clin. Oncol.* **2018**, *36*, 539–539. [[CrossRef](#)]
29. Cardoso, F.; van't Veer, L.J.; Bogaerts, J.; Slaets, L.; Viale, G.; Delaloge, S.; Pierga, J.Y.; Brain, E.; Causeret, S.; DeLorenzi, M.; et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* **2016**, *375*, 717–729. [[CrossRef](#)] [[PubMed](#)]
30. Nielsen, T.O.; Parker, J.S.; Leung, S.; Voduc, D.; Ebbert, M.; Vickery, T.; Davies, S.R.; Snider, J.; Stijleman, I.J.; Reed, J.; et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **2010**, *16*, 5222–5232. [[CrossRef](#)] [[PubMed](#)]
31. Jerevall, P.L.; Ma, X.J.; Li, H.; Salunga, R.; Kesty, N.C.; Erlander, M.G.; Sgroi, D.; Holmlund, B.; Skoog, L.; Fornander, T.; et al. Prognostic utility of HOXB13: IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br. J. Cancer* **2011**, *104*, 1762–1769. [[CrossRef](#)] [[PubMed](#)]
32. Cuzick, J.; Dowsett, M.; Pineda, S.; Wale, C.; Salter, J.; Quinn, E.; Zabaglo, L.; Mallon, E.; Green, A.R.; Ellis, I.O.; et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J. Clin. Oncol.* **2011**, *29*, 4273–4278. [[CrossRef](#)]
33. Sestak, I.; Dowsett, M.; Zabaglo, L.; Lopez-Knowles, E.; Ferree, S.; Cowens, J.W.; Cuzick, J. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J. Natl. Cancer Inst.* **2013**, *105*, 1504–1511. [[CrossRef](#)]

