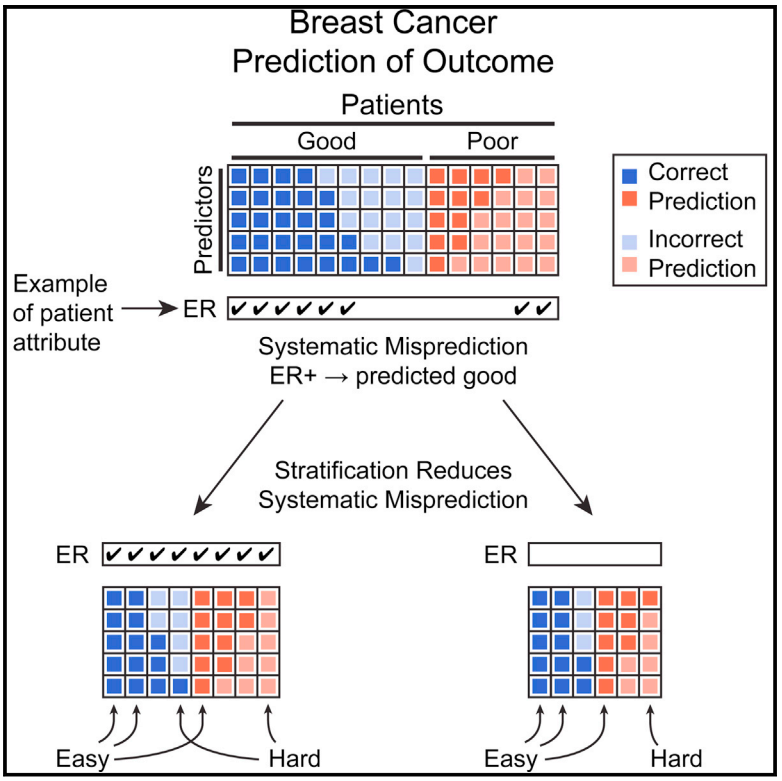
Article

Cell Reports

The Prognostic Ease and Difficulty of Invasive Breast Carcinoma

Graphical Abstract



Highlights

Predictors of outcome in breast cancer have been massively confounded

Subtype-specific prognostic genes/signatures exist but are rare and imperfect

About 7% of poor-outcome breast cancers are predicted as good by nearly all markers

Inherently difficult cases are prioritized for intratumoral heterogeneity studies

Authors

Ali Tofigh, Matthew Suderman, ..., Vanessa Dumeaux, Michael Hallett

Correspondence

michael.t.hallett@mcgill.ca

In Brief

Tofigh et al. perform a comprehensive and systematic comparison of existing prognostic signatures in breast cancer. The comparison establishes that the supposed ubiquity of prognostic genes and processes has been grossly overestimated and results from insufficient stratification by subtype and other clinicopathological variables. The study presents a refined subtyping scheme ablating these effects. Nevertheless, a set of patients is identified whose outcome appears inherently difficult to predict using all available information at time of diagnosis.

Accession Numbers

GSE58644

Tofigh et al., 2014, Cell Reports 9, 129–142
October 9, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.08.073

Cell Reports
Article

The Prognostic Ease and Difficulty of Invasive Breast Carcinoma

Ali Tofigh,[1,2,11] Matthew Suderman,[1,2,11] Eric R. Paquet,[1,2] Julie Livingstone,[1,2] Nicholas Bertos,[1] Sadiq M. Saleh,[1,2,3] Hong Zhao,[1] Margarita Souleimanova,[1] Sean Cory,[1,2] Robert Lesurf,[1,2,3] Solmaz Shahalizadeh,[2] Norberto Garcia Lopez,[6,9] Yasser Riazalhosseini,[6,10] Atilla Omeroglu,[7,9] Josie Ursini-Siegel,[4,5] Morag Park,[1,3,5] Vanessa Dumeaux,[5,8] and Michael Hallett[1,2,3,*]

[1]The Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, QC H3A1A3, Canada
[2]Centre for Bioinformatics, McGill University, Montreal, QC H3G0B1, Canada
[3]Department of Biochemistry, McGill University, Montreal, QC H3G1Y6, Canada
[4]Lady Davis Institute for Medical Research, McGill University, Montreal, QC H3T1E2, Canada
[5]Department of Oncology, McGill University, Montreal, QC H2W1S6, Canada
[6]Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada
[7]Department of Pathology, McGill University, Montreal, QC H3A 2B4, Canada
[8]Institute of Community Medicine, UiT the Arctic University of Norway, Tromso 9037, Norway
[9]McGill University Health Centre, McGill University, Montreal, QC H3A 1A1, Canada
[10]McGill University and Génome Québec Innovation Centre, Montreal, QC H3A 0G1, Canada
[11]Co-first author
*Correspondence: michael.t.hallett@mcgill.ca
http://dx.doi.org/10.1016/j.celrep.2014.08.073

SUMMARY

Breast carcinoma (BC) has been extensively profiled by high-throughput technologies for over a decade, and broadly speaking, these studies can be grouped into those that seek to identify patient subtypes (studies of heterogeneity) or those that seek to identify gene signatures with prognostic or predictive capacity. The sheer number of reported signatures has led to speculation that everything is prognostic in BC. Here, we show that this ubiquity is an apparition caused by a poor understanding of the interrelatedness between subtype and the molecular determinants of prognosis. Our approach constructively shows how to avoid confounding due to a patient's subtype, clinicopathological profile, or treatment profile. The approach identifies patients who are predicted to have good outcome at time of diagnosis by all available clinical and molecular markers but who experience a distant metastasis within 5 years. These inherently difficult patients (~7% of BC) are prioritized for investigations of intratumoral heterogeneity.

INTRODUCTION

There has been a sustained effort to identify markers of prognosis in women diagnosed with invasive breast cancer (IBC), a highly prevalent disease that accounts for 14% of all cancer deaths in women (Jemal et al., 2011). The estimation of prognosis at time of diagnosis relies primarily upon clinicopathological parameters such as tumor size, histological grade, stage, lymph node (LN) infiltrate, and molecular properties including expression of the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) (Reis-Filho and Pusztai, 2011). Prognostic insight can in turn provide predictive insight with respect to patient benefit from chemo-, endocrine, and targeted therapies.

The paradigm shift promised by technologies that measure expression of genomic features of IBC in a massively parallel fashion has inarguably occurred but is tempered by the fact that few prognostic gene signatures have found clinical use (Hornberger et al., 2012). The primary contribution of breast cancer genomics to date has been a deeper appreciation of IBC heterogeneity (Weigelt et al., 2010). Although the four clinical subtypes defined by ER and HER2 status have long been recognized as distinct forms of the disease, early genomic studies underscored their vast differences at the molecular level (Gruvberger et al., 2001; Perou et al., 2000) and stimulated work to identify other markers that capture proliferation, progenitor cell properties, androgen-receptor-related signaling, and other biologies (Desmedt et al., 2008; Guedj et al., 2012; Rakha et al., 2010). Moreover, unbiased bioinformatic analyses of profiles generated the so-called intrinsic subtyping scheme, consisting of two subtypes enriched for ER+ tumors (luminal A, B), a HER2+-enriched subtype (her2-enriched), a ER−/HER2−-enriched subtype (basal-like), and a so-called normal-like subtype (Perou et al., 2000; Sørlie et al., 2001). The intrinsic subtyping scheme has been refined (Parker et al., 2009) and extended to include the claudin-low class of tumors that display a high frequency of metaplastic and medullary differentiation (Prat et al., 2010). Other genomics-based subtyping schemes have been proposed, including the Cartes d'identité des tumeurs (CIT) scheme (Guedj et al., 2012), a "triple negative" specific scheme (Lehmann et al., 2011), a scheme based on joint DNA and RNA copy number (Curtis et al., 2012), and others (Haibe-Kains et al., 2012; Jönsson et al., 2010; Wirapati et al., 2008).

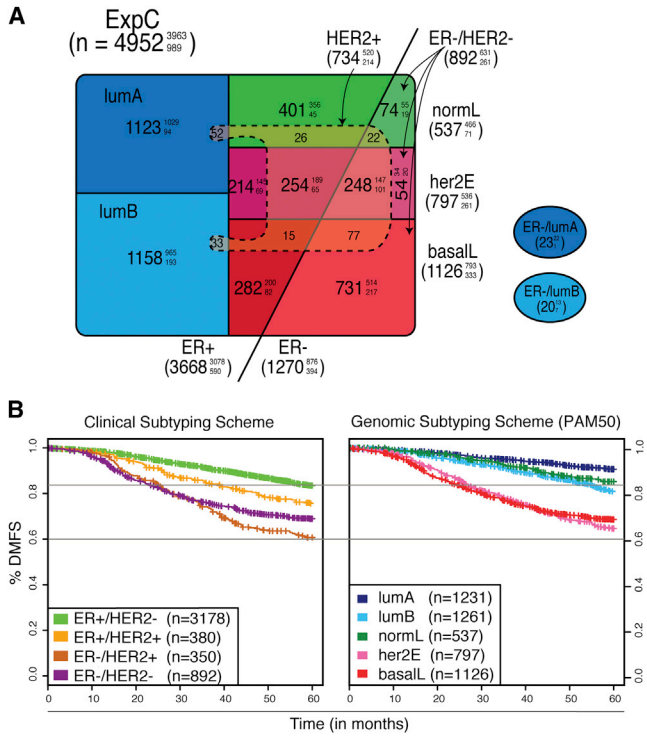


Figure 1. Characteristics of the ExpC and Subtyping Schemes

(A) The compendium of IBC gene expression profiles by intrinsic and clinical schemes. Total number of subjects with observed good (superscript) and poor (subscript) outcome, respectively.

(B) Kaplan-Meier plots of observed patient outcome defined as distant metastasis-free survival across the ExpC in both the clinical and intrinsic schemes.

See also Figures S1 and S2.

In parallel with, but largely disjoint from, this work, many (>100) gene signatures have been reported to have prognostic capacity in IBC. The signatures were typically derived either by manual curation of specific molecular processes (e.g., antigen presentation and processing [APP] pathway), from experimental perturbations in cell lines and transgenic mouse models of the disease (e.g., Ursini-Siegel et al., 2010), or directly from gene expression profiles of IBC samples by contrasting observed good and poor outcome patients (van 't Veer et al., 2002). A classification of good prognosis by a signature suggests that an individual will respond positively to standard of care for their tumor type, whereas a prediction of poor prognosis suggests the need for an alternative regimen. Clinical utility for a few signatures has been established and translated to the clinic (Hornberger et al., 2012), including, for example, OncotypeDX, which assists in determining which ER+/LN− patients may benefit from additional adjuvant chemotherapy (Paik et al., 2004).

The marginal overlap and the differences in the underlying biological processes polled by these signatures have generated criticisms regarding experimental design and lack of standardized bioinformatics techniques (Ioannidis et al., 2009), leading to speculation that almost all genes and processes are prognostic in IBC (Venet et al., 2011). However, the root cause of the myriad of dissimilar signatures may be primarily due to deep, structural interdependencies between a patient's clinicopathological profile, tumor subtype, and prognosis (Iwamoto and Pusztai, 2010). Here, we clarify the nature and ubiquity of this poorly understood confounding.

The prognostic predictions made by published gene signatures are compared en masse patient by patient, in order to identify if and where additional progress is possible. Using essentially all available data, this investigation establishes that there is large confounding in existing signatures between clinicopathological variables, subtype, and clinical outcome. A de novo subtyping scheme is presented that ablates the majority of these effects. The resultant scheme highlights approximately 20% of patients whose prognosis appears inherently difficult to predict at time of diagnosis.

RESULTS

A Comprehensive Examination of Prognosis in Breast Cancer

Publicly available gene expression data sets for IBC were evaluated according to technical quality, clinical attributes, treatment regimen, data set size, and patient outcome (Supplemental Experimental Procedures, 2.1). Data sets deemed sufficient were harmonized as much as possible (Supplemental Experimental Procedures, 2.1.1) to measure patient outcome as event-free survival at 5 years (good outcome) versus the existence of distant metastases within the same time interval (poor outcome). The resultant compendium, referred to as the ExpC, contains approximately 5,000 patients from over 11 data sets using seven microarray and RNA-sequencing (RNA-seq) platforms (Table S1), including our de novo effort (McGill Genome Quebec [MGGQ] n = 314, GSE58644).

Samples of ExpC were labeled according to five subtyping schemes, where possible (Figure 1A; Table S7). ER and HER2 status were used to determine the so-called clinical subtypes (ER+/HER2+, ER+/HER2−, ER−/HER2+, and ER−/HER2−). The PAM50 gene set (Parker et al., 2009) was used for the intrinsic subtypes of luminal A (lumA), luminal B (lumB), normal-like (normL), basal-like (basalL), and her2-enriched (her2E). Patients were also labeled according to the IntClust (Curtis et al., 2012), CIT (Guedj et al., 2012), and triple-negative breast cancer (TNBC) (Lehmann et al., 2011) schemes (Tables S1 and S7). Although there are enrichments between subtypes of different schemes, all five are distinct (Table S2; see also Tables S4 and S5 and Figure S1). As expected, the vast majority (96.5%) of luminals (lumA and lumB) is ER+; however, over 25% of basalL and 50% of her2E patients are ER+ (Figure 1A).

We collected gene signatures from the literature reported to have prognostic capacity in IBC (n = 106; Table S3). Together, the signatures use one-third (~6.4K) of all human genes with few appearing in multiple signatures (Table S9). Not surprisingly then, these signatures are enriched for a wide range of biological processes (Table S3). Using a standardized technique with each prognostic signature, a naive Bayes' classifier was trained for each patient subtype within each subtyping scheme when there were a sufficient number of good- and poor-outcome

individuals (Supplemental Experimental Procedures, 2.6). This procedure was carried out systematically, and each signature was evaluated in all subtype stratifications irrespective of the subtype in which the signature was originally derived. We also developed de novo classifiers using the same methodology for training (Figure S2E; Supplemental Experimental Procedures, 2.7). The performance of both the de novo and literature-derived classifiers was thoroughly investigated (Figures S2F–S2I). This signature collection is referred to as SigC (n = 122). All analyses are available at http://www.bci.mcgill.ca/bresect.

Prediction of Outcome Is Confounded in Pan-IBC Analysis

When patient subtype is not considered in the training and application of classifiers, almost every signature in SigC appears capable of predicting patient outcome (89% with log-rank test $p < 0.05$; "unstratified" in Figure 2A). This suggests that an extremely broad range of biological processes represented by signatures in SigC including proliferation (Vanvliet-2008), microenvironmental factors (Finak-2008), and immunological responses (Rody-2009) all have prognostic capacity, an observation consistent with previous reports (Venet et al., 2011; Weigelt et al., 2010).

To investigate this apparent ubiquitous prognostic signal, we plotted the predictions of individual SigC classifiers across all patients stratified by outcome (Figure 3). The classifiers are ordered from top to bottom by decreasing performance in both good and poor outcome (Supplemental Experimental Procedures, 2.6.1). Only ~56% of SigC signatures perform better than random signatures built by choosing 25 arbitrary genes (Figure 3F). Those classifiers that did not perform better than random (Figure 3G) tend to systematically predict all patients as good outcome, likely because the majority (79%) of patients with IBC have good outcome. Such classifiers offer neither clinical utility nor insight into the molecular mechanisms of the disease.

The classifiers that outperform 99% of random gene signatures exhibited a high degree of agreement in their predictions. That is, either all signatures predict the patient to be good outcome (Figures 3B and 3E) or poor outcome (Figures 3C and 3D). Figures 3B and 3C then represent patients that are easy to predict correctly, whereas Figures 3D and 3E represent patients that appear inherently difficult to predict.

We asked if the surprisingly concordant incorrect predictions for the patients in Figures 3D and 3E could be due to the fact that the classifiers systematically confuse patient outcome with intrinsic subtype or other clinicopathological variables. To do this, a de novo statistical method entitled the Systematic MisPrediction (SMP) test was developed that identifies when predictions of outcome made by the classifiers en masse are highly associated with a variable of interest (e.g., ER status, treatment) within both the good- and poor-outcome subcohorts (Supplemental Experimental Procedures, 2.6.3). In essence, significance for the SMP test suggests that the classifiers have learnt to predict the variable of interest rather than patient prognosis per se.

For example, across all IBC samples, the SMP test finds that ER status is highly significant ($p < 0.001$; Table 1): nearly all ER+ patients are classified as good outcome, and nearly all ER− are assigned poor outcome. This represents systematic misprediction, since there are a significant number of both poor-outcome ER+ patients and good-outcome ER− patients; these patients are almost never predicted as such by the SigC classifiers. Intrinsic subtypes were similarly found associated with systematic misprediction of outcome; patients were predicted as good outcome if and only if their subtype was lumA, lumB, or normL (Table 1; Figure 3H). In summary, although the SigC classifiers were trained to predict outcome, they have instead learnt (clinical or intrinsic) subtype.

Prediction of Outcome Is Confounded in ER and HER2 Defined Cohorts

We asked if prediction of outcome within cohorts restricted to only ER+ tumors would ablate confounding. Here, 77% of the SigC classifiers still remain significant under survival analysis (log rank, $p < 0.05$; Figure 2A), and the SMP test establishes that the classifiers consistently assign prognosis according to intrinsic subtype: lumA and normL patients as good outcome, and her2E and basalL patients as poor outcome ($p < 0.001$ for each intrinsic subtype; Table 1; Figure S3A). Similar problems were identified for ER−, HER2−, and HER2+ cohorts (Table 1; Figures S3B and S3C). In fact, instead of prognosis, the classifiers built for ER-defined subtypes tend to learn HER2 status and HER2 cohorts learn ER status. The differential in rates of poor outcome between ER+ and ER− cohorts explains their apparent prognostic capacity (similarly for HER2), suggesting that the prognostic classifiers have no power to predict outcome above and beyond the information implicit in the patient's subtype.

Limitations of the Clinical and Intrinsic Schemes

We next built SigC classifiers for each subtype of the clinical scheme. No signature had significant prognostic capacity in all four cohorts (Figure 2A). The ER+/HER2− cohort contained the greatest degree of systematic misprediction where patients were still predicted based on their intrinsic subtype (Table 1; Figure S3C). The ER−/HER2− cohort contains a high number of claudin-low (CL) patients ($p < 0.0001$; Table S2) with evidence of systematic misprediction of CL patients as good prognosis (SMP test, $p < 0.0001$; Table 1). We investigated the refinement of the ER−/HER2− subtype into CL and non-CL cohorts for prognostic prediction but identified only 12 weakly differentially expressed genes (linear models for microarray data [limma], false discovery rate < 0.5; Table S10).

With respect to the intrinsic subtypes (Figure 2B), patients with high-grade tumors were systematically mispredicted as poor outcome within the lumB subtype (SMP test, $p < 0.01$; Table 1; Figure S3D). Since high-grade tumors often receive chemotherapy, it is not surprising that patients who received chemotherapy were also systematically mispredicted as poor outcome ($p < 0.01$). As per the ER−/HER2− clinical subtype, basalL was confounded by CL status with good prognosis assigned to CL individuals (SMP test, $p < 0.01$; Figure S3D).

The SMP test did not identify significant confounding by ER or HER2 status for the normL, basalL, or her2E subtypes, suggesting that these subtypes represent a sufficiently homogeneous

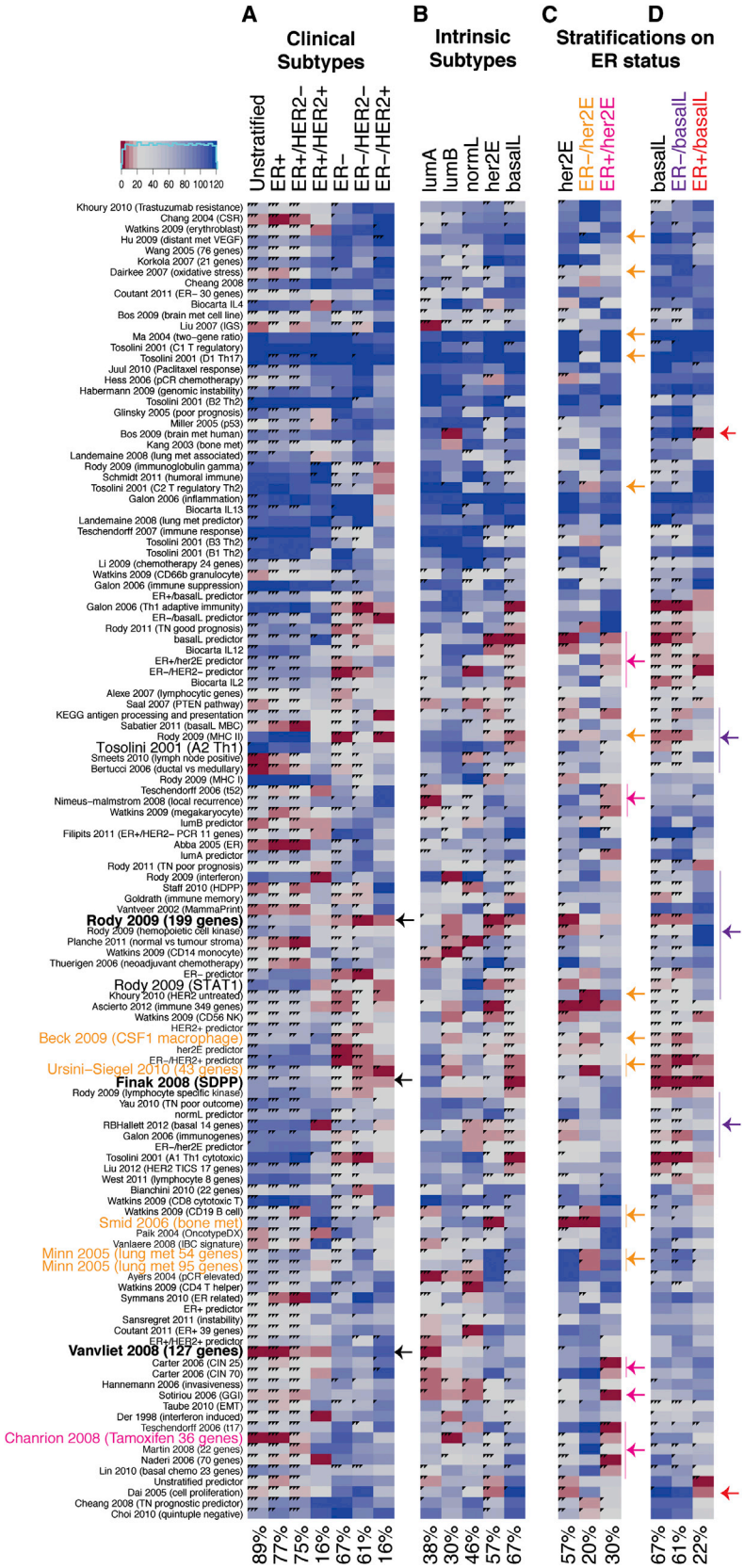


Figure 2. Subtype-Specific Performance of Prognostic Classifiers

Colors are proportional to the rank of the classifier within the specific patient cohort, with red representing the highest-performing classifiers relative to the remaining SigC members. Ticks represent the level of significance of the classifier (log-rank test, $p < 0.05$, 0.01, 0.001, respectively). Bottom row contains percentage of SigC significant at $p < 0.05$. Major signatures of interest have been highlighted. See also Figure S2.

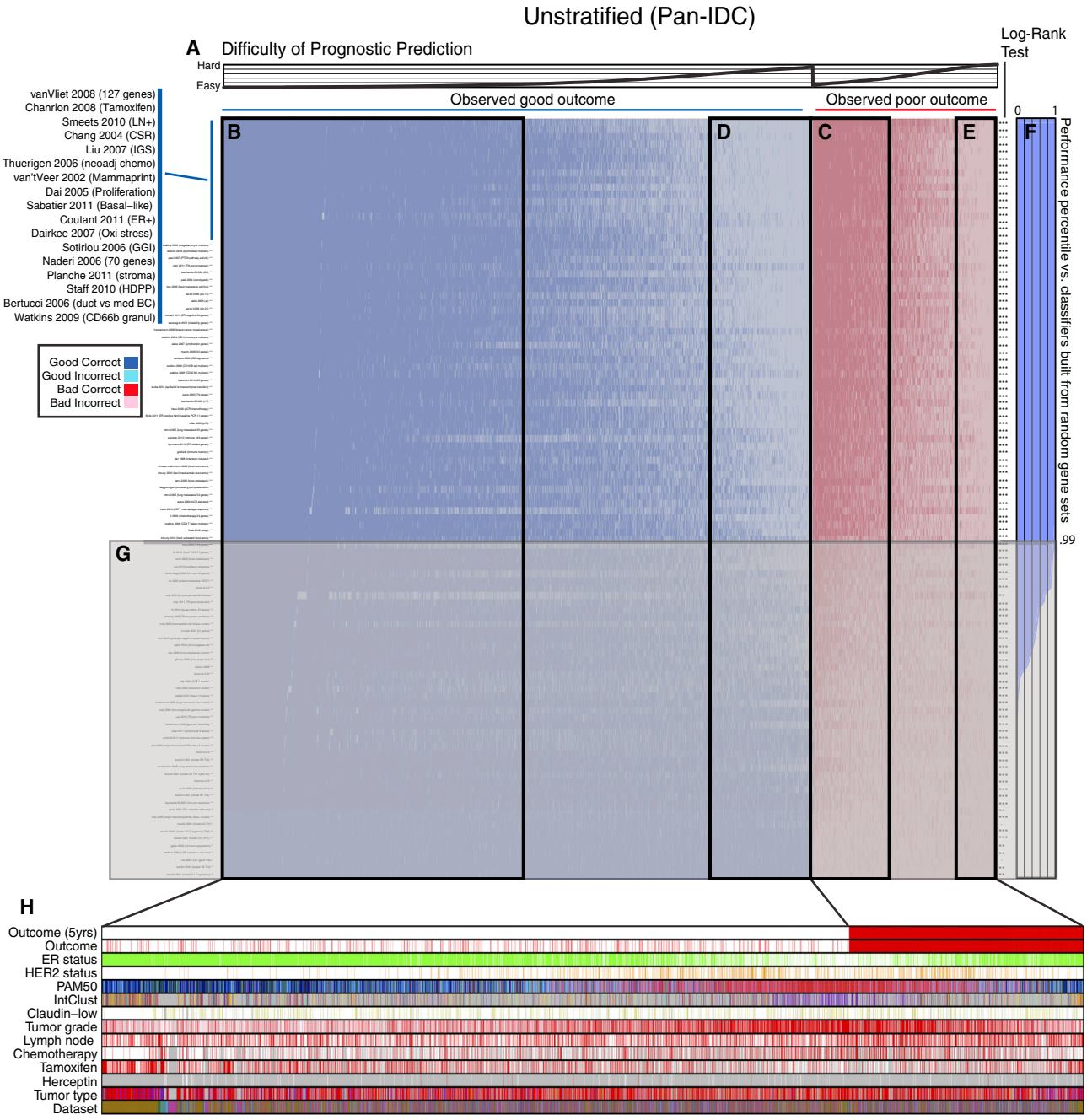


Figure 3. Patient-Signature Heatmap of Classifiers in Unstratified Analysis of IBC

Blue and red shaded columns correspond to good- and poor-outcome patients, respectively. The classifiers (rows) from SigC are ordered by their ability to predict patient outcome. Asterisks correspond to significance by survival analysis (log-rank test). Dark and light shades correspond to correct and incorrect predictions, respectively. Patients (columns) are ordered by the degree of agreement of predictions across all members of SigC across good and bad outcome.

(A) Difficulty score of predicting outcome.

(B and C) A subset of patients that almost every classifier predicts correctly.

(D and E) Patients mispredicted by almost every classifier.

(F) The percentage of trials (n = 100K) whereby the specific classifier (row) outperformed a classifier built from a random set of k = 25 genes.

(G) The box shades the subset of classifiers from SigC that does not outperform classifiers built from a random set of k = 25 genes over 99% of the trials.

(H) Clinicopathological, treatment, and subtype attributes per patient. A red tick in "Outcome (5 yrs)" refers to an observed distant metastasis within 5 years. "Outcome" does not place a restriction on length of follow-up. Color-coding for PAM50 and IntClust follows original publications. Color-coding for data sets is given in Table S1.

See also Figure S3.

Table 1. Table of Associations from the Systematic MisPrediction Test

	IHC		Treatment		Patient		Tumor			PAM50					Claudin-Low
	ER	HER2	Chemotherapy	Tamoxifen	Age	LN	Size	Stage	Grade	lumA	lumB	normL	her2E	basalL	CL
Unstratified	++++	−−−−	−−−−	++++	++++				−−−−	++++	++++	++++	−−−−	−−−−	
ER+	NA	−−−−	−−−						−−−−	++++		++++	−−−−	−−−−	
ER−	NA	−−−								NA			−−−−	++	++++
HER2+	++	NA		++											
HER2−	++++	NA	−−−−	++++	++++				−−−−	++++		++++		−−−−	
ER+/HER2+	NA	NA										NA		NA	
ER+/HER2−	NA	NA	−−−						−−−−	++++		++++	−−−−	−−−−	
lumA	NA									NA	NA	NA	NA	NA	
lumB									−−	NA	NA	NA	NA	NA	
normL										NA	NA	NA	NA	NA	
her2E										NA	NA	NA	NA	NA	
basalL										NA	NA	NA	NA	NA	++
ER−/her2E	NA									NA	NA	NA	NA	NA	
ER−/basalL	NA									NA	NA	NA	NA	NA	++
ER+/lumA	NA									NA	NA	NA	NA	NA	
ER+/lumB									−−	NA	NA	NA	NA	NA	
ER+/normL	NA	NA								NA	NA	NA	NA	NA	
ER+/her2E	NA									NA	NA	NA	NA	NA	
ER+/basalL	NA	NA								NA	NA	NA	NA	NA	
ER−/HER2+	NA	NA								NA	NA				
ER−/HER2-	NA	NA								NA	NA				++++

Each entry represents the extent to which consistent misclassification with a specific subtype (row) is associated with a specific clinicopathological or patient attribute. "+" indicates that patients positive for that attribute are systematically assigned good prognosis, and "−" indicates the reverse. The number of ticks (two, three, or four) corresponds to p values < 0.01, 0.001, and 0.0001, respectively. NA indicates insufficient data.

group of patients suitable for prognostic studies. The homogeneity suggests that a biological process that has prognostic capacity in an ER+/her2E tumor will also have capacity in an ER−/her2E tumor, and vice versa. However, a comparison of the SigC classifiers between the her2E, ER+/her2E, and ER−/her2E cohorts shows a marked difference (Figure 2C; Figure S4A). The most significant classifiers for ER+/her2E poll biological processes shown previously to be prognostic in ER+ tumors (e.g., Chanrion-2008). In contrast, several classifiers are only prognostic in ER−/her2E including metastasis to bone (Smid-2006) and lung (Minn-2005) signatures and diverse immune-related signals (Beck-2009, Rody-2009, Tosolini-2001, and Ursini-Siegel-2010). Lastly, the ER+ and ER− subcohorts of her2E have distinct survival characteristics, suggesting that her2E is a very heterogeneous subtype (Figure S2B; log-rank test, $p < 0.0001$; Cox proportional hazards: 1.86, 95% confidence interval: [1.34, 2.59]).

Similarly for the basalL related comparisons (Figure S4B), relatively few (22%) SigC classifiers are significant for ER+/basalL tumors compared to ER−/basalL (Figure 2D). Although ER+ and ER− basalL tumors displayed similar survival characteristics (Figure S2C), ER+/basalL were less likely to have received chemotherapy (Table S2).

Together, these differences suggest that joint clinical and intrinsic subtyping may be necessary to move beyond the strong, dominant ER- and HER2-related signals found in IBC.

Identification of a De Novo Hybrid Subtyping Scheme

We computationally searched across all possible combinations of the clinical and intrinsic subtyping schemes for a hybrid approach maximizing prognostic capacity while minimizing systematic misprediction. The most complicated hybrid scheme would partition tumors into 20 subtypes: 2 (ER±) × 2 (HER2±) × 5 (intrinsic subtypes). A single gene expression-based classifier would be created for each subtype (Supplemental Experimental Procedures, 2.7; Figures S2E and S2F; Table S7). For example, an ER+, HER2+, and basalL tumor would be evaluated with a classifier learnt previously in a training set specific to this cohort.

This search identified a scheme that partitions IBC into seven hybrid subtypes (Figure 4A). For ER+ tumors, intrinsic subtyping is used for further refinement. For ER− tumors, the search recommends the use of only HER2 status. Figure 5 depicts the performance of the SigC classifiers across the hybrid subtypes (Ward's method, Euclidean distance) with Figures 5A–5C depicting performance, overlap between signatures, and the MSigDB molecular processes enriched in these gene sets (Subramanian et al., 2005; Supplemental Experimental Procedures, 2.4). Although none of the SigC classifiers were significant across all hybrid subtypes, the most universal (area labeled "Immune SigC") or significant ("Top") classifiers were mainly immune related: Ursini-Siegel-2010, Finak-2008, and Rody-2009 signatures (Figure 5A).

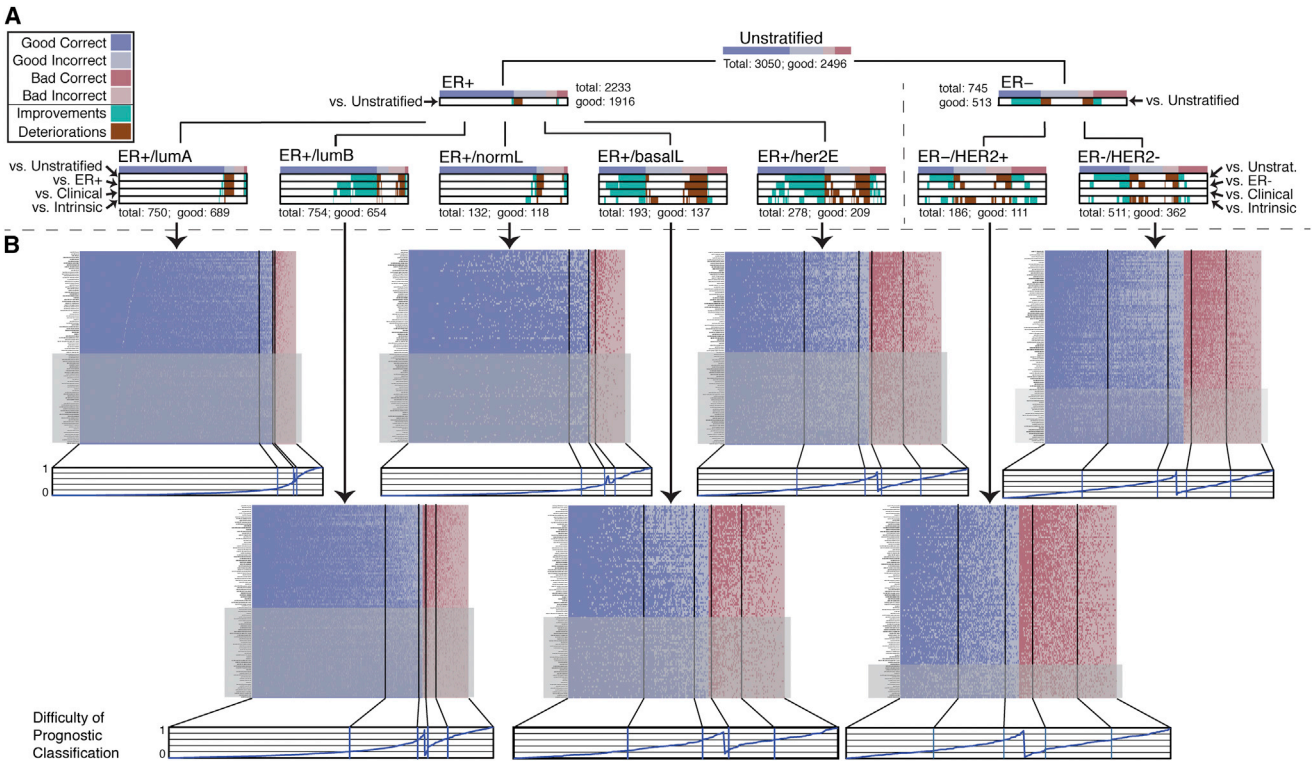


Figure 4. The Hybrid Subtyping Scheme

(A) A decision tree where the root corresponds to all patients. ER+ samples move left in the tree. The row labeled "vs. Unstratified" is a comparison of the predictions made by a classifier trained in unstratified analysis versus the classifier trained in an ER+-restricted cohort. Green entries correspond to samples where the ER+-restricted classifier is correct but the parental classifier is incorrect (improvements due to additional stratification). Brown corresponds to incorrect predictions by the ER+ classifier but correct predictions by the unstratified classifier (deteriorations). The row labeled "Clinical" at each node of the tree compares classifiers built with the hybrid scheme versus classifiers built using the clinical scheme (similarly for row "Intrinsic"). For instance, for an ER+/HER2−/basalL sample, the row labeled "vs Intrinsic" compares the prediction made by a ER+/basalL classifier with a basalL classifier, while the row labeled "vs Clinical" compares the predictions made by a ER+/basalL classifier with a ER+/HER2− classifier.

(B) Patient-signature heatmap for each hybrid subtype as per Figure 2. Vertical black lines delimit the inherently easy and difficult cases within both good-outcome (blue) and poor-outcome (red) portions.

See also Figure S4.

The coclustering of ER+/lumA and ER+/lumB in Figure 5A suggest that roughly the same signatures (labeled "Sub-ER+" and "ER+ SigC") have prognostic capacity in both cohorts, in turn suggesting that both cohorts share the same prognostic biological processes primarily related to proliferation and cell cycle (Figures 5B and 5C, pink region). However, some immune-related signatures appear to have more prognostic capacity in ER+/lumB than in ER+/lumA (e.g., Ursini-Siegel-2010 and Beck-2009).

The hybrid scheme separates classification of ER+/her2E tumors from ER−/HER2+ tumors. Consistent with this is the observation that very few SigC signatures are significant in both cohorts (e.g., immune related signatures Ascierto-2012, APP, and a chromosomal instability signature, Carter-2006).

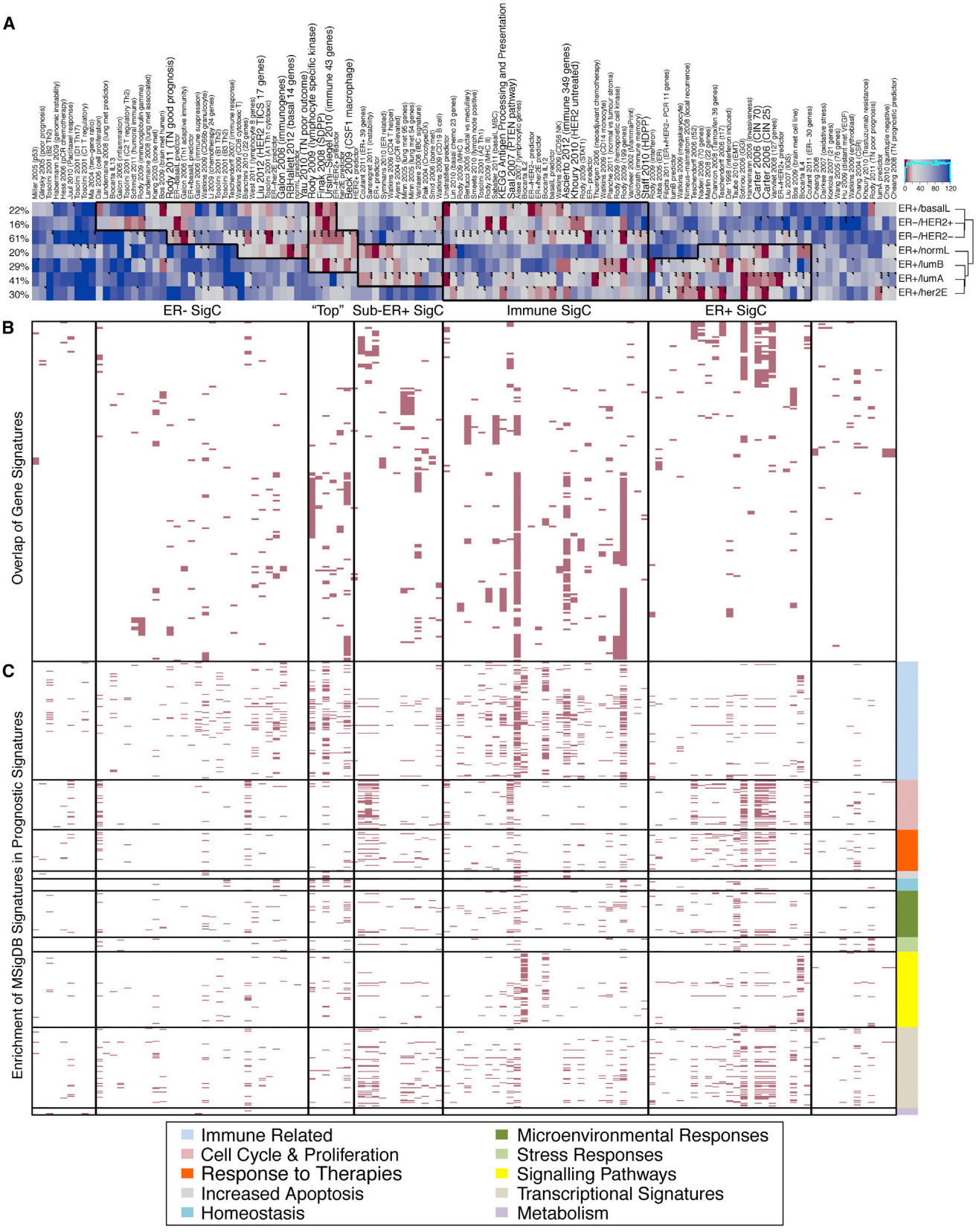
The scheme recommends separate classifiers for the ER+/normL cohort, whereas ER−/normL tumors should be classified using ER−/HER2− (Figure 4A). The ER+/normL subtype is an outgroup within the ER+ subtree containing lumA, lumB, and her2E (Figure 5A), and some of the most prognostic signatures for this subtype are shared with ER− cohorts (e.g., Yau-2010 and Hallett-2012).

Finally, we observe that the ER+/basalL subtype is most similar to ER− related cohorts. However, some ER+/basalL signatures do remain significant within the ER+/luminal subtypes (e.g., Saal-2007) but are not significant in the ER−/basalL subtype, suggesting that ER+ and ER− basalL tumors are significantly different from a prognostic perspective (Figure S4B).

The Hybrid Scheme versus Alternatives

The hybrid scheme was then compared against alternative subtyping schemes including IntClusters (IC), CIT, and the TNBC subtypes. For each hybrid subtype, we asked (1) if it contains a surprising number of patients from a subtype of an alternative scheme (Table S2) and (2) if the SMP test identified a subtype from an alternative scheme that confounds prognostic predictions (Table S8).

ER+/lumB contains surprisingly many tumors of subtype IC8 and IC9 (both $p < 0.0001$; Table S2). The IC9 subtype corresponds to ER+ tumors with 8q *cis*-acting/20q-amplification events (Curtis et al., 2012). Within ER+/lumB, they are systematically mispredicted as poor outcome ($p < 0.01$). The IC8 tumors



(legend on next page)

are characterized by classical 1q gain/16q loss. In ER+/lumB, these tumors are systematically mispredicted as good outcome ($p < 0.01$).

With respect to ER−/HER2− tumors, we observed systematic misprediction of IC4 tumors (characterized by extensive lymphocytic infiltration) as good outcome and IC10 tumors (characterized as primarily basalL tumors by Curtis et al.) as poor outcome (both $p < 0.0001$).

Our attempts to develop prognostic classifiers for refinements of the hybrid subtypes that include IC subtypes were limited by the fact that IC subtypes are currently only available within the Curtis et al. data set (Curtis et al., 2012). Notably, when we trained prognostic classifiers for each of the 10 IC subtypes alone, their performance did not achieve statistical significance in the validation portion of their data set.

There were also surprisingly many CIT-"lumC" tumors within the ER+/basalL hybrid subtype (Table S2; $p < 0.0001$) that were systematically predicted as good outcome (Table S8; $p < 0.001$). This is surprising, since "lumC" is characterized as a highly proliferative subtype with relatively poor prognosis (Guedj et al., 2012). This suggests that the restrictive ER+/basalL subtype still contains latent confounding due to the expression of proliferation genes.

A Trade-Off between Systematic Misprediction and Performance

The hybrid scheme identifies a partitioning of IBC that reduces as best possible the systematic misprediction of patient outcome. Somewhat counterintuitively, however, the ablation of systematic misprediction may appear to hinder our ability to predict survival times. The removal of confounding variables and subsequent increased homogeneity of the hybrid subtypes should allow classification based on actually biological processes that associate with patient outcome. We plotted the improvements and deteriorations between the hybrid, clinical and intrinsic schemes (Figure 4A). Here, each patient is first subtyped by each of the three schemes, and then outcome is predicted via a classifier specific for each of the assigned subtypes (Supplemental Experimental Procedures, 2.7). All classifiers were developed with the same methodology. Indeed, the overall performance of the hybrid scheme is only slightly better than previous subtyping schemes (Figure S2D for survival; +3% product of accuracy). The performance within the ER+/lumA subtype is inferior but increases in the ER+/lumB, ER+/basalL, and ER+/her2E hybrid subtypes.

The Inherent Complexity of Tumors

The performance of the SigC classifiers across the hybrid subtypes is plotted in Figure 5A (see also Figures S3C–S3E). In every hybrid subtype, we observe a set of patients who are correctly predicted to have good outcome by almost every SigC classifier and likewise a set of patients consistently correctly predicted as bad outcome (Figure 4B). We term these patients "inherently easy." Given the diversity of biology that the different classifiers poll to make predictions, this agreement suggests that many distinct biological processes expressed in these tumors have prognostic capacity. In contrast, we also observe patients that appear to be systematically misclassified by almost every member of SigC. The outcome of these patients appears to be "inherently difficult" to predict. The black vertical bars delimit the inherently easy/difficult good/poor outcome patients based upon our de novo statistic used quantify the difficulty of predicting outcome (Supplemental Experimental Procedures, 2.6.2). The difficulty score for each patient is a function of how many of the SigC signatures incorrectly predict its outcome, but weighting the contribution of each signature by its overall performance.

The existence of inherently easy and difficult patients was also witnessed in analyses without stratification and with alternative subtyping schemes; however, these inherent classes were found to be the result of confounding due to underlying clinicopathological variables (Table S8). Here with the hybrid scheme, there is minimal systematic misprediction with known variables, implying that the inherent easy and difficult classes remain unexplained. We verified that the inherent ease and difficulty was not caused by technical reasons such as the choice of classifier method; partition of ExpC into training, testing, and validation; cellularity of samples (Table S6; e.g., fibroblasts); and other variables (Supplemental Experimental Procedures, 2.14).

The Inherent Complexity of ER−/HER2− Tumors Is Reflected at the DNA Level

To investigate whether the inherent complexity of tumors observed in RNA profiles was conserved at the DNA level, we selected three ER−/HER2− samples from patients determined as easy good, difficult good, and difficult poor outcome in the MGGQ data set. The tumors were subjected to massively parallel exome sequencing (Nextera; Supplemental Experimental Procedures, 2.8). In total, 232,000,000 read pairs were uniquely aligned to the human genome (NCBI, hg19) with an average sequencing depth of 400–600 reads per exonic site. CoNIFER (Krumm et al., 2012) was used to identify regions of deletions and amplification in each sample while excluding regions polymorphic in normal, healthy samples (Supplemental Experimental Procedures, 2.8). In total, 118 loci contained 280 distinct genes with differing copy number in at least one of the three individuals were identified (Figure 6). A correlation of 0.8 with respect to the copy number for these loci was observed between the inherently easy good-outcome and the inherently difficult poor-outcome individuals (Spearman, $p < 10^{-15}$). In contrast, no significant correlation was observed between the inherently

Figure 5. The Prognostic Capacity of Each Classifier from SigC across the Hybrid Subtypes

(A) Colors are proportional to the rank of the signature within the specific patient cohort, with red representing the highest performance. Ticks represent the level of significance of the classifier in the cohort (log-rank test, 0.05, 0.01, 0.001). Leftmost column is the percentage of SigC significant at $p < 0.05$. Rightmost column contains a hierarchical clustering of the hybrid subtypes. Major signatures of interest have been highlighted.

(B) Rows correspond to genes that appear as members of at least three SigC signatures.

(C) Rows correspond to signatures from MSigDB that have surprisingly large overlap with at signatures in SigC (Fisher's exact test, $p < 0.01$). Rows grouped by molecular function.

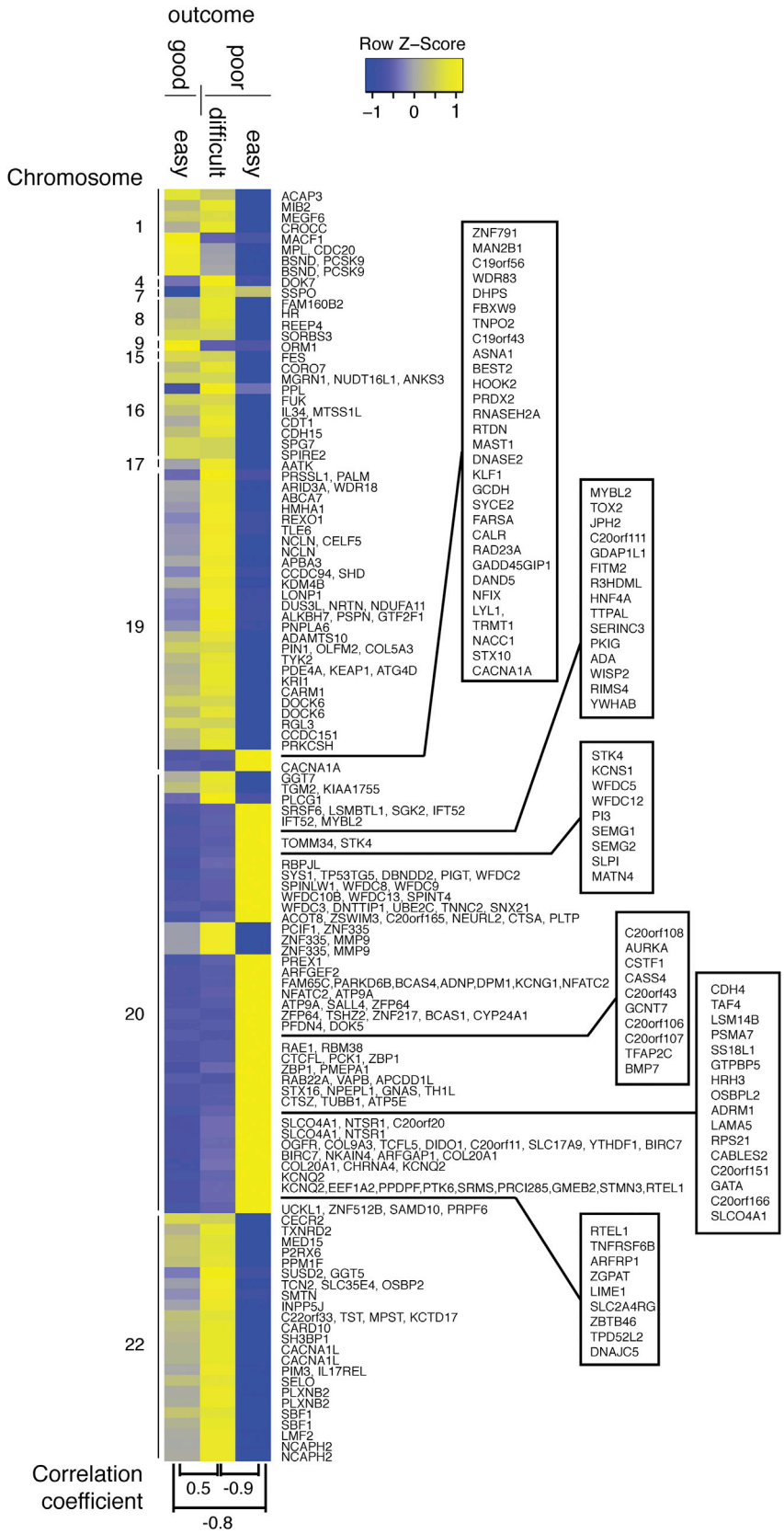


Figure 6. A Comparison of Exonic Differences between Three Samples that Differ in Outcome and Difficulty of Prediction

Columns correspond to an inherently easy, good-outcome patient (left); an inherently difficult, poor-outcome patient (middle); and an inherently easy, poor-outcome patient (right). Each row represents a contiguous series of exons that had significantly different DNA copy number labeled by genes contained in the regions. Yellow represents increased copy number in comparison to blue (Z transformed).

easy and difficult poor-outcome samples (Spearman correlations < 0.2, p = 0.06).

DISCUSSION

Subtyping Reshapes the Landscape of Prognosis

Signatures reported to have prognostic capacity in IBC were systematically compared across a large compendium of expression profiles. When all IBCs are considered, or when liberal subtypes defined only by ER or HER2 status are used, most classifiers had statistically significant ability to predict patient outcome. This ubiquity, which has been reported previously (Sotiriou and Pusztai, 2009; Venet et al., 2011; Weigelt et al., 2010), is difficult to comprehend given the great diversity of biological processes that are polled by the individual signatures. Using our de novo SMP test, we are able to explain such behavior by quantitatively identifying clinicopathological variables that confuse the prediction of outcome. In unstratified analysis, the classifiers only predict ER status. Since there are differential rates of poor outcome between ER+ and ER− tumors (Parl et al., 1984), the classifiers have a prognostic capacity proportional to this differential. For variables such as ER status, the molecular differences between ER+ and ER− subtypes are so large that almost any gene or gene set is likely to exhibit differential expression between the subtypes (Gruvberger et al., 2001), implying the magnitude of the transcriptional fingerprint of ER is more predominant than signals directly associated with patient prognosis. The confusion between prognostic and classification markers explains the apparition that "almost everything" is prognostic in breast cancer.

The Insufficiency of the Clinical Scheme

The clinical subtypes do not remove all systematic misprediction of prognosis. For example, in the ER+/HER2− cohort, the classifiers predicted nearly all lumA and normL tumors to have good outcome, while basalL and her2E were assigned poor outcome. This is due possibly to the fact that pathologic determination of ER positivity in the clinic is intentionally permissive ranging from few (1%) to many (10%) ER+ cells (Iwamoto et al., 2012), since all such patients may benefit from antiestrogen therapies (Harbeck and Rody, 2012). In cases where few ER+ cells are present in the tumor bed, the molecular signal related to ER may be weak in the expression profile and tools such as PAM50 for intrinsic subtyping may identify more substantive her2E or basalL signals, as observed by others (Deyarmin et al., 2013). The ER+/HER2− cohort is heterogeneous (Aparicio and Caldas, 2013), and our results establish that this heterogeneity impairs the prediction of outcome.

The Insufficiency of the Intrinsic Scheme

With respect to the intrinsic scheme, there are notable issues surrounding the her2E, normL, and basalL subtypes. In all three cases, these intrinsic subtypes have a significant number of both ER+ and ER− tumors (ER+: her2E 58.7%, basalL 26.3%, normL 79.5%). For her2E, the two ER-defined subcohorts show significant differences in disease-free survival, and prognosis is predicted by a distinct set of classifiers. Surprisingly, no signature was observed to be significant for both ER+ and ER− HER2-related cohorts, with few immune-related exceptions that appear to be prognostic in most subtypes (not necessarily HER2 related). The paucity of HER2-specific signatures is surprising given that some (e.g., Khoury-2010, Staff-2010, and Liu-2012) were built explicitly for this purpose. Although HER2 amplification is a good predictive marker for response to anti-HER2 therapy across all IBCs, these results establish that it does not function as prognostic marker within the HER2 subcohort.

For basalL, the classifiers most significant for ER+/basalL differ from those significant for ER−/basalL. Several signatures significant for ER−/basalL patients were not significant within ER+/basalL patients, including several immune-related signatures (Galon-2006 and Rody-2011) and the APP pathway.

The Hybrid Subtyping Scheme

In order to ablate misprediction and confusion, we considered all possible subtyping schemes that can arise from combining the clinical and intrinsic subtypes. The focus was placed on these two schemes since they are the most clinically feasible (Harbeck et al., 2014). This produced the decision tree of Figure 4A.

For ER− tumors, the search identified only HER2 status as significant, which is a surprising lack of refinement given the attention subtyping within ER− tumors has received. The ER−/lumA and ER−/lumB subtypes studied by Prat et al. (Prat et al., 2013) were too infrequent in the ExpC (n = 23, one poor outcome; n = 20, seven poor outcome, respectively) for deeper statistical evaluation. We compared classifiers for ER−/HER2− tumors with the more restricted cohort of ER−/HER2−/basalL but failed to identify a difference in survival characteristics, standard of care, or significant SigC classifiers, suggesting that basalL is not a proper subset of ER−/HER2− tumors.

ER−/HER2−/claudin-low (CL) tumors have significantly better prognosis than non-CL counterparts and were systematically mispredicted as good outcome by ER−/HER2− classifiers. We were, however, unable to identify markers of prognosis between good- and poor-outcome ER−/HER2−/CL patients. Together, this suggests that although the ER−/HER2−/CL subtype has interesting molecular and pathological properties (Prat et al., 2010), they do not assist in the prediction of outcome. IC4 of IntClusters is highly enriched in ER−/HER2− tumors, and the patient overlap between IC4 and CL is extremely high, suggesting near equivalence.

The hybrid scheme recommends the partitioning of ER+ tumors by intrinsic subtype. In particular, ER+ and ER− HER2+ tumors are segregated for prognostic treatment. The distinctiveness of ER+/her2E is in agreement with randomized clinical trials that have found benefit in dual targeted treatment combining endocrine and anti-HER2 treatment (Montemurro et al., 2013).

In the ExpC, the ER+/basalL patients appear to have been treated solely as ER+ tumors receiving tamoxifen but were less likely to have received chemotherapy in comparison to ER−/basalL, especially for LN+ patients. This is likely due to the fact that intrinsic subtyping is not clinically available. Although ER+ tumors may be more resistant to chemotherapy than ER− tumors (Rouzier et al., 2005), ER+/basalL cases may represent candidates that stand to benefit from broader use of chemotherapy. Consideration of a novel treatment strategy for the

ER+/basalL hybrid subtype may have clinical relevance, since they have the highest rate of recurrence within ER+ tumors, at double the frequency of poor-outcome cases as lumB.

The Prognostic Axis across All Signatures

The clustering in Figure 5A highlights the relationships between the members of the SigC and the hybrid subtypes. At one end of the axis ("ER-SigC"), the signatures have prognostic capacity almost exclusively within ER−-related cohorts and poll aspects of the immune response. At the other end of the axis ("ER+SigC"), the signatures have the most capacity in ER+ tumors and poll processes such as proliferation, cell cycle, and regulation of transcription. Some ER−-related signatures have some prognostic capacity within ER+ cohorts, including two of the highest-performing signatures across all hybrid subtypes (Finak-2008 and Ursini-Siegel-2010). Neither of these signatures was originally learnt in gene expression profiles from bulk (epithelial enriched) clinical samples of IBC but instead from stroma-microdissected clinical samples and transgenic mouse models of the disease, respectively. Both have significant adaptive immune and microenvironmental components. This leads to a hypothesis that gene signatures built in contexts where ER-related signaling is ablated or irrelevant tend to be the most universal across IBC.

The Inherent Difficulty of Some Tumors

Although the degree to which a biological process can predict outcome varies according to subtype, the existence of the inherently difficult and easy classes of tumors (Figure 4B) does suggest that almost every biological process has some degree of prognostic ability in every subtype. This would suggest that the molecular profile of, for example, an inherently easy individual contains a clear and universal signal of prognosis that is "encoded" in almost every biological process, regardless of tumor grade, stage, LN status, age, subtype, or any other patient/tumor property. In essence, for good-outcome individuals, standard of care was sufficient for the individual given their exposures, lifestyle choices, genotypic polymorphisms, tumor colony structure, and other variables.

For poor-outcome inherently easy individuals, this implies that at time of diagnosis before treatment the molecular profile contained a ubiquitous signal that the tumor was likely to progress under standard of care. Factors such as low-penetrant resistance subcolonies and intratumoral complexity do not play a role.

Almost no genes or pathways are differentially expressed between inherently difficult, good-outcome and inherently easy, poor-outcome individuals. There are many possible reasons why two individuals who have essentially identical transcriptomes may have differing outcome, such as lifestyle, exposures (Poole et al., 2013), genotypic variation (Landmark-Høyvik et al., 2013), and other variables largely ignored to date in genomic studies. It is possible that an epigenetic mark, posttranslational modification, expressed microRNA, or other genomic features not measured in transcriptional studies could distinguish these individuals. However, we comment that such an alternative mark would have to have had no feedback on transcriptional levels. For example, the hypothetical epigenetic mark must not have disrupted the expression of transcripts; otherwise, its fingerprint would be detected in these studies. Polymorphisms or other genomic features may exist in these individuals who do not exert changes in the transcriptome until the tumor is challenged with therapy. Alternatively, changes at the transcriptional level may be too small to detect, especially when the signal originates from only a marginal population of cells within the tumor (progenitor cells or a low penetrant colony). This suggests the need for greater integration of physiologic, epidemiological, and environment data into genomic studies and underscores the need for tumor progression studies.

Lastly, the inherently difficult, poor-outcome individuals correspond to patients for whom almost every classifier predicts incorrectly as good outcome. One of the possible explanations for the inherent difficulty is the existence of low-penetrant subcolonies or rare populations of atypical progenitor cells capable of forming resistance to standard of care. Likely undetectable in bulk profiling, these classifiers would not be able to differentiate such tumors. If tumor heterogeneity is the primary cause of resistance to therapy and subsequent poor outcome, it is these individuals who are most likely to harbor the intratumoral complexity.

A preliminary study examining a series of ER−/HER2− tumors from patients determined as easy good, difficult good, and difficult bad prognosis by exomic deep sequencing (400–600 average depth) suggests that the chromosomal aberrations of the inherently difficult poor-outcome sample is structurally most similar to the inherently easy, good-outcome sample. At least in this pilot study, the inherent complexity of patients is conserved at the DNA level, although all three tumors harbor both distinct and shared mutations.

Conclusions

Given the limited nature of genomic studies in breast cancer to date, no prognostic classifier should achieve a perfect success rate at time of diagnosis, since this would imply that lifestyle, exposures, tumor heterogeneity, and genotype do not play a role in determining disease course. Our findings suggest that markers of patient outcome at time of diagnosis exist, although they are subtype specific, relatively rare, and imperfect.

We identified 20% of IBCs whose prognosis appears inherently difficult to predict, with approximately one-third of these patients of poor outcome. This latter cohort should be the priority for studies via massively deep sequencing, since it is only these patients who likely harbor clinically relevant intratumoral heterogeneity affecting disease progression.

Our approach and website provide the community with a resource to cross-compare findings in the continual stream of new data sets, subtyping schemes, and signatures for breast cancer. This will make more precise the true nature of a proposed biomarker after detangling the effects of confounding factors, a problem that has plagued breast cancer informatics to date.

EXPERIMENTAL PROCEDURES

Expression Data

We compiled publicly available gene expression profiles of IBC (n = ~10K) and evaluated each data set according to criteria including quality and technical

disparities of the microarray/sequencing technology, availability of clinical and histopathological information, nonoverlap of patients, and overall size. We harmonized as much as possible differences in follow-up time and defined poor outcome as an observed distant metastasis within 5 years of diagnosis (Supplemental Experimental Procedures, 2.1). The procedure produced a compendium of 4,952 patients (ExpC) with expression measured with seven distinct technologies including several microarray and RNA-seq platforms (Table S1). Approval for the de novo MGGQ dataset was received from the McGill ethical review board (#A10-M92-10A).

Subtyping Schemes

We used ER status as measured by IHC reported and HER2 status measured by fluorescence in situ hybridization or immunohistochemistry where available. When unavailable, HER2 status was determined using gene expression of members of the HER2 amplicon as previously done (Staaf et al., 2010). Since many of the data sets lacked information on the PR status, we used only ER and HER2 status either in isolation or in combination to define eight possible cohorts (ER+, ER−, HER2+, HER2−, and the four clinical subtypes using ER/HER2). All ExpC samples were labeled according to four additional subtyping schemes from the literature: IntClusters (Curtis et al., 2012), CIT (Guedj et al., 2012), TNBC (Lehmann et al., 2011), and intrinsic subtypes via PAM50 (Parker et al., 2009) (Figure 1A; Table S7).

Previously Reported Prognostic Signatures

Our goal was a systematic cross-comparison of all signatures (sets of genes) reported to have prognostic capacity for IBC in the literature. In total, the signature collection (SigC) contains n = 106 gene sets ranging in size from three to 886 (Table S3).

Construction of Prognostic Classifiers

A unique classifier was built for each signature in SigC within each stratification defined by clinical variables (e.g., ER+, HER2−) and intrinsic subtypes (lumA, lumB, normL, her2E, and basalL). Combinations of the clinical and intrinsic subtyping schemes (e.g., ER+/basalL) were also considered. The naive Bayes' classifier (NBC) was trained under leave-one-out cross-validation (1) for each signature, (2) for each patient cohort, and (3) within each individual data set of the ExpC, for which there were sufficient numbers of event (distant metastasis within 5 years; poor-outcome) and event-free (good-outcome) individuals (Figure 1A; Supplemental Experimental Procedures, 2.7). NBCs were chosen as they provide a simple, transparent, and uniform technique to cross-evaluate the signatures, although some signatures were originally developed using other techniques (detailed in Table S3). In addition to the gene signatures of the SigC, the prognostic capacity of clinical attributes including grade, stage, LN status, and age were considered. A small number of signatures in SigC have been further developed into tools to aid during treatment decision making (e.g., Oncotype DX and MammaPrint). Here, we classify patients according to the output from NBCs and not according to the methods of those tools. Lastly, the prognostic capacity of each gene was evaluated across each data set of the ExpC with respect to every possible subtype (Supplemental Experimental Procedures, 2.2 and 2.9).

Construction of De Novo Prognostic Classifiers per Subtype

In addition to existing prognostic signatures, we also constructed de novo prognostic signatures for each subtype (Figure S2). For some subtypes, such as the hybrid ER+/basalL, these represent the only available classifiers. Toward this end, the ExpC was tripartitioned into learning (n = 897; van Vliet et al., 2008), training (n = 819), and validation (n = 2,412) data sets (Tables S1 and S7). The training and validation data sets were chosen so that each underlying technology (microarray or sequencing platform) was present. Using the learning data set, our approach samples a set of k genes from the n most prognostic genes in univariate analysis (absolute value of Cox-PH coefficient). This sampling is repeated m times and each such sample is used to construct an NBC in each data set of the learning partition. We experimented with parameterizations of this approach and found that k = 25, n = 100, and m = 25K provided small but high-performing classifiers (Figure S2F). Only the classifier that achieves the maximum observed performance across the m = 25K samples in the learning and training data sets is tested in the validation data set. This procedure is repeated for each of the target subtypes. A wide range of measures were used to evaluate performance on the validation data sets including the log-rank test, Cox proportional hazard, area under the curve, Matthews correlation coefficient, Fisher's exact test, and de novo permutation tests (Supplemental Experimental Procedures, 2.7; Figures S2G and S2H).

Further analyses are available at http://www.bci.mcgill.ca/bresect.

ACCESSION NUMBERS

Microarray data for the MGGQ data set have been deposited to the NCBI Gene Expression Omnibus under accession number GSE58644.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and ten tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.08.073.

AUTHOR CONTRIBUTIONS

A.T., M. Suderman, V.D., M.H., and N.B. designed the experiments. A.T., M. Suderman, V.D., and M.H. wrote the manuscript. Software development was done by M. Suderman and A.T. A.T., M. Suderman, E.P., R.L., J.U.S., M.P., V.D., and M.H. were involved in bioinformatics and biostatistical analysis. N.B., H.Z., M. Souleimanova, A.O., N.B.L., Y.R., and M.P. were responsible for sample collection and preparation. Sample processing and quality control were performed by J.L., N.B., S.C., A.T., and M. Suderman. J.L., S.M.S., R.L., S.C., S.S., A.T., and M. Suderman collected data sets and compendiums.

ACKNOWLEDGMENTS

The authors acknowledge the financial support provided by Génome Québec (to M.H., M.P., and J.U.S.), technical support from the McGill University and Génome Québec Innovation Centre, and infrastructure support and technical assistance from the Breast Cancer Functional Genomics Group, partially funded by the Terry Fox New Frontiers Program (to M.P.). V.D. is supported by the European Research Council grant ERC-2008-AdG 232997. The tissue and data bank at the MUHC is supported by funding from the Database and Tissue Bank Axis of the Réseau de Recherche en Cancer of the Fonds de Recherche du Québec-Santé and the Quebec Breast Cancer Foundation (to M.P.). The authors gratefully acknowledge the assistance provided by D. Fleiszer, A. Loutfi, A. Meguerditchian, S. Meterissian, C. Milne, F. Tremblay, M. Wexler (Surgery); R. Amre, D. Haegert, Y. Kanber, R. Michel, G. Omeroglu-Altinel, K. Watters (Pathology); and the MUHC Anaesthesia Department. This work is dedicated to the memory of Rosalind Goodman.

Received: January 10, 2014
Revised: July 18, 2014
Accepted: August 27, 2014
Published: October 2, 2014

REFERENCES

Aparicio, S., and Caldas, C. (2013). The implications of clonal genome evolution for cancer medicine. N. Engl. J. Med. *368*, 842–851.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin. Cancer Res. *14*, 5158–5165.

Deyarmin, B., Kane, J.L., Valente, A.L., van Laar, R., Gallagher, C., Shriver, C.D., and Ellsworth, R.E. (2013). Effect of ASCO/CAP guidelines for determining ER status on molecular subtype. Ann. Surg. Oncol. *20*, 87–93.

Gruvberger, S., Ringnér, M., Chen, Y., Panavally, S., Saal, L.H., Borg, A., Fernö, M., Peterson, C., and Meltzer, P.S. (2001). Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. *61*, 5979–5984.

Guedj, M., Marisa, L., de Reynies, A., Orsetti, B., Schiappa, R., Bibeau, F., MacGrogan, G., Lerebours, F., Finetti, P., Longy, M., et al. (2012). A refined molecular taxonomy of breast cancer. Oncogene *31*, 1196–1206.

Haibe-Kains, B., Desmedt, C., Loi, S., Culhane, A.C., Bontempi, G., Quackenbush, J., and Sotiriou, C. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. J. Natl. Cancer Inst. *104*, 311–325.

Harbeck, N., and Rody, A. (2012). Lost in translation? Estrogen receptor status and endocrine responsiveness in breast cancer. J. Clin. Oncol. *30*, 686–689.

Harbeck, N., Sotlar, K., Wuerstlein, R., and Doisneau-Sixou, S. (2014). Molecular and protein markers for clinical decision making in breast cancer: today and tomorrow. Cancer Treat. Rev. *40*, 434–444.

Hornberger, J., Alvarado, M.D., Rebecca, C., Gutierrez, H.R., Yu, T.M., and Gradishar, W.J. (2012). Clinical validity/utility, change in practice patterns, and economic implications of risk stratifiers to predict outcomes for early-stage breast cancer: a systematic review. J. Natl. Cancer Inst. *104*, 1068–1079.

Ioannidis, J.P., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Culhane, A.C., Falchi, M., Furlanello, C., Game, L., Jurman, G., et al. (2009). Repeatability of published microarray gene expression analyses. Nat. Genet. *41*, 149–155.

Iwamoto, T., and Pusztai, L. (2010). Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? Genome Med. *2*, 81.

Iwamoto, T., Booser, D., Valero, V., Murray, J.L., Koenig, K., Esteva, F.J., Ueno, N.T., Zhang, J., Shi, W., Qi, Y., et al. (2012). Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J. Clin. Oncol. *30*, 729–734.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. CA Cancer J. Clin. *61*, 69–90.

Jönsson, G., Staaf, J., Vallon-Christersson, J., Ringnér, M., Holm, K., Hegardt, C., Gunnarsson, H., Fagerholm, R., Strand, C., Agnarsson, B.A., et al. (2010). Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res. *12*, R42.

Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E.; NHLBI Exome Sequencing Project (2012). Copy number variation detection and genotyping from exome sequence data. Genome Res. *22*, 1525–1532.

Landmark-Høyvik, H., Dumeaux, V., Nebdal, D., Lund, E., Tost, J., Kamatani, Y., Renault, V., Børresen-Dale, A.L., Kristensen, V., and Edvardsen, H. (2013). Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. Genomics *102*, 278–287.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. *121*, 2750–2767.

Montemurro, F., Di Cosimo, S., and Arpino, G. (2013). Human epidermal growth factor receptor 2 (HER2)-positive and hormone receptor-positive breast cancer: new insights into molecular interactions and clinical implications. Ann. Oncol. *24*, 2715–2724.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. *351*, 2817–2826.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. *27*, 1160–1167.

Parl, F.F., Schmidt, B.P., Dupont, W.D., and Wagner, R.K. (1984). Prognostic significance of estrogen receptor status in breast cancer in relation to tumor stage, axillary node metastasis, and histopathologic grading. Cancer *54*, 2237–2242.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Poole, E.M., Shu, X., Caan, B.J., Flatt, S.W., Holmes, M.D., Lu, W., Kwan, M.L., Nechuta, S.J., Pierce, J.P., and Chen, W.Y. (2013). Postdiagnosis supplement use and breast cancer prognosis in the After Breast Cancer Pooling Project. Breast Cancer Res. Treat. *139*, 529–537.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68.

Prat, A., Adamo, B., Cheang, M.C., Anders, C.K., Carey, L.A., and Perou, C.M. (2013). Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. Oncologist *18*, 123–133.

Rakha, E.A., Reis-Filho, J.S., and Ellis, I.O. (2010). Combinatorial biomarker expression in breast cancer. Breast Cancer Res. Treat. *120*, 293–308.

Reis-Filho, J.S., and Pusztai, L. (2011). Gene expression profiling in breast cancer: classification, prognostication, and prediction. Lancet *378*, 1812–1823.

Rouzier, R., Perou, C.M., Symmans, W.F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K.R., Stec, J., Ayers, M., Wagner, P., et al. (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin. Cancer Res. *11*, 5678–5685.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

Sotiriou, C., and Pusztai, L. (2009). Gene-expression signatures in breast cancer. N. Engl. J. Med. *360*, 790–800.

Staaf, J., Jönsson, G., Ringnér, M., Vallon-Christersson, J., Grabau, D., Arason, A., Gunnarsson, H., Agnarsson, B.A., Malmström, P.O., Johannsson, O.T., et al. (2010). High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. Breast Cancer Res. *12*, R25.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Ursini-Siegel, J., Cory, S., Zuo, D., Hardy, W.R., Rexhepaj, E., Lam, S., Schade, B., Jirstrom, K., Bjur, E., Piccirillo, C.A., et al. (2010). Receptor tyrosine kinase signaling favors a protumorigenic state in breast cancer cells by inhibiting the adaptive immune response. Cancer Res. *70*, 7776–7787.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

van Vliet, M.H., Reyal, F., Horlings, H.M., van de Vijver, M.J., Reinders, M.J., and Wessels, L.F. (2008). Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. BMC Genomics *9*, 375.

Venet, D., Dumont, J.E., and Detours, V. (2011). Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput. Biol. *7*, e1002240.

Weigelt, B., Baehner, F.L., and Reis-Filho, J.S. (2010). The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J. Pathol. *220*, 263–280.

Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schütz, F., et al. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. *10*, R65.