Breast cancer prognosis predicted by nuclear receptor-coregulator networks

Tram B. Doan\textsuperscript{a}, Natalie A. Eriksson\textsuperscript{b}, Dinny Graham\textsuperscript{a}, John W. Funder\textsuperscript{c}, Evan R. Simpson\textsuperscript{c}, Elizabeth S. Kuczek\textsuperscript{d}, Colin Clyne\textsuperscript{c}, Peter J. Leedman\textsuperscript{e}, Wayne D. Tilley\textsuperscript{f}, Peter J. Fuller\textsuperscript{c}, George E.O. Muscat\textsuperscript{b,\*}, Christine L. Clarke\textsuperscript{a,\*}

\textsuperscript{a}Westmead Millennium Institute, Sydney Medical School -- Westmead, University of Sydney, NSW, Australia
\textsuperscript{b}Institute for Molecular Bioscience, University of Queensland, St. Lucia, QLD, Australia
\textsuperscript{c}Prince Henry's Institute of Medical Research, Clayton, VIC, Australia
\textsuperscript{d}Sydney Medical School, University of Sydney, Westmead, NSW, Australia
\textsuperscript{e}Laboratory for Cancer Medicine, Centre for Medical Research, Western Australian Institute for Medical Research and School of Medicine and Pharmacology, the University of Western Australia, Perth, WA, Australia
\textsuperscript{f}Dame Roma Mitchell Cancer Research Laboratories, Discipline of Medicine, Hanson Institute, University of Adelaide, Adelaide, SA, Australia

\textbf{ARTICLE INFO}

Article history:
Received 20 December 2013
Received in revised form
20 February 2014
Accepted 24 March 2014
Available online 4 April 2014

Keywords:
Breast
Nuclear receptors
Transcriptome
Prognosis
Gene signature

\textbf{ABSTRACT}

Although molecular signatures based on transcript expression in breast cancer samples have provided new insights into breast cancer classification and prognosis, there are acknowledged limitations in current signatures. To provide rational, pathway-based signatures of disrupted physiology in cancer tissues that may be relevant to prognosis, this study has directly quantitated changed gene expression, between normal breast and cancer tissue, as a basis for signature development. The nuclear receptor (NR) family of transcription factors, and their coregulators, are fundamental regulators of every aspect of metazoan life, and were rigorously quantified in normal breast tissues and ER\textsubscript{\alpha} positive and ER\textsubscript{\alpha} negative breast cancers. Coregulator expression was highly correlated with that of selected NR in normal breast, particularly from postmenopausal women. These associations were markedly decreased in breast cancer, and the expression of the majority of coregulators was down-regulated in cancer tissues compared with normal. While in cancer the loss of NR-coregulator associations observed in normal breast was common, a small number of NR (Rev-ERB\textsubscript{\beta}, GR, NOR1, LRH-1 and PGR) acquired new associations with coregulators in cancer tissues. Elevated expression of these NR in cancers was associated with poorer outcome in large clinical cohorts, as well as suggesting the activation of ER\textsubscript{\alpha} -related, but ER\textsubscript{\alpha} -independent, pathways in ER\textsubscript{\alpha} negative breast cancers. In addition, the combined expression of small numbers of NR and coregulators in breast cancer was identified as a signature predicting outcome in ER\textsubscript{\alpha} negative breast cancer patients, not linked to proliferation and with predictive power superior to existing signatures containing many more genes. These findings highlight the power of predictive signatures derived from the quantitative determination of altered gene expression between normal breast and breast cancers. Taken together, the findings of this study identify networks of NR-coregulator...
1. Introduction

Breast cancer is one of the most common malignancies worldwide, affecting millions of women annually in developed and developing communities. Its prevalence and increasing incidence has galvanized significant advances in the treatment options available, and continued advances in understanding of its biological underpinnings have revealed that breast cancer is both clinically and biologically heterogeneous. Considerable effort has been expended to identify tumor features that are prognostic of patient outcome, and/or predictive of responsiveness to particular treatment regimens. In this regard, a defining feature of breast cancers is the presence of estrogen receptor alpha (ERα), where ERα positive and ERα negative cancers are characterized by major differences in biology, sensitivity and response to treatment, and patient outcome.

While ERα, progesterone receptor (PR) and epidermal growth factor receptor 2 (HER 2) are the histopathological markers that underpin current clinical management of breast cancer (Weigel and Dowsett, 2010), the imperative for increasingly refined and personalized predictive tools has driven the inclusion of molecular profiling in breast cancer prognostication, and there is now a significant body of evidence that molecular profiling of transcripts expressed in breast cancer samples can contribute to classification of breast cancers and to prognostic signatures (Weigelt et al., 2010a), now being tested in clinical trials (www.agendia.com/pages/mindact; Zujewski and Kamin, 2008). Although molecular-based signatures have revealed a range of breast cancer subtypes with distinct responses to treatment and outcome, and can identify patients at higher risk of early recurrence, there are acknowledged limitations in the current signatures (Geyer et al., 2012). These include the striking lack of overlap between signatures derived from different sample sets; their primary utility in ERα positive breast cancers and the lack of signatures that are predictive in both ERα positive and ERα negative cancers; the contribution of proliferation-related genes to signature performance (Venet et al., 2011); and the complicating influence of transcripts present in the variable proportion of normal tissue admixed with the cancer component of samples used in signature discovery (Weigelt et al., 2010b). In addition, signatures based on the combined complement of genes expressed in cancer tissue at the time of sampling fail to provide insight into biological pathways or processes involved in carcinogenesis.

The emerging evidence for increased immune response activity as a predictor of improved prognosis has prompted a focus on signatures that incorporate immune pathways. These are proving to be useful in ERα negative breast cancers (Nagalla et al., 2013), but identification of proliferation-independent signatures for ERα positive breast cancers, or signatures predictive in both ERα positive and ERα negative cancers, remains challenging. To overcome some of these limitations, and also to provide rational, pathway-based signatures of disrupted physiology in cancer tissues that may be relevant to prognosis, this study has directly quantitated changed gene expression between normal breast, and cancer tissue, as a basis for signature development. Our emphasis was two-fold: first, to base the discovery on accurately quantitated differences between normal and cancer tissue, and second, to focus on components of a gene family that serves as fundamental regulators of metazoan biology and are therefore active in every aspect of human physiology. The nuclear receptor (NR) superfamily of transcription factors are crucial in reproduction, development, growth, metabolism and homeostasis. In terms of pathophysiology, they play key roles in the cardiovascular and immune systems, the central nervous system and the musculoskeletal system, and in the genesis and progression of cancer (Conzen, 2008). The NR family are expressed in the normal breast, and pan-repression of the majority of NR is associated with carcinogenesis (Muscat et al., 2013), highlighting the rationale for a role for NR in development and progression of the disease.

The cell- and context-specificity of NR action occurs through their interaction with defined groups of coregulators, proteins that interact with NR to either activate or repress the transcription of specific genes (Lanz et al., 2007; McKenna and O’Malley, 2010). The transcriptional outcome regulated by NR is dependent on the complement of coregulators present in the target cell. Using human breast specimens (pre-and post-menopausal normal breast, and ERα positive and ERα negative breast cancers), we have quantitated all known NR coregulators in the human breast, and identified NR-coregulator interactions altered in cancer. We have developed molecular signatures independent of proliferation that outperform existing discriminators, and of demonstrated utility in both ERα positive and ERα negative breast cancers.

2. Materials and Methods

2.1. Human breast tissue cohorts and mRNA profiling by TaqMan low-density array (TLDA)

The breast tissue cohorts profiled in this study and the TLDA experimental procedures have been previously published (Muscat et al., 2013). The present study employs the same experimental procedures to profile the expression of 238 coregulators (listed in Supplementary Figure 1) in the same human breast tissue cohorts. Briefly, we custom-designed micro-fluidic cards, TaqMan Low Density arrays from ABI (Applied BioSystems), which included 238 coregulators (as defined by the Nuclear Receptor Signaling Atlas (NURSA: www.nursa.org)) and 16 internal
controls for normalization (these comprised 18S-Eukaryotic ribosomal RNA; ACTB – Actin, beta; 2M – Beta-2-microglobulin; GAPDH – Glyceraldehyde 3-phosphatedehydrogenase; GUSB – Glucuronidase, beta; HBMS – Hydroxyethylbilirubin synthase; HPRT1 – Hypoxanthine phosphoribosyltransferase 1; IPO8 – Importin 8; PGK1-Phosphoglycerate kinase 1; POLR2A – Polymerase (RNA) II (DNA directed) polypeptide A, 220 kDa; PPIA – Peptidylprolyl isomerase A (cyclophilin A); RPLP0 – Ribosomal protein, large, P0; TBP – TATA box binding protein; TFRC – Transferrin receptor (p90, CD71); UBC – Ubiquitin C; YWHAZ – Tyrosine 3-monoxygenase/tetrathopan 5-monoxygenase activation protein, zeta polyptide) against which to normalize the expression of the coregulators in normal and neoplastic breast tissue. These controls include 18S rRNA, GAPDH and RPLP0, validated real-time PCR controls in NURSA-supported nuclear receptor profiling studies (Holbeck et al., 2010). The geNorm software embedded within the ABI/Integromics StatMiner V4.1 software package was used to compute least expression variation and select the most appropriate, stable and robust combination of internal control genes (MRPL19, PGK1, PPIA, TFRC and UBC) against which to normalise the expression data (against the median of the most stable controls).

For each sample, 1.5 μg of total RNA was reverse transcribed with random hexamers and SuperScript III reverse transcriptase (Invitrogen) in a total volume of 45 μl. A total of 100 μl reaction mixture containing 50 μl cDNA template (333 ng) in RNase-free water and an equal volume of TaqMan® universal master mix (Applied Biosystems, Foster City, CA) was added to each TLDA fill reservoir. Four reservoirs per sample were filled. The TLDA includes all coregulators and endogenous controls in triplicate. After sealing the plate, it was run on an ABI 7900HT Real Time instrument (Applied Biosystems).

We made use of the delta Ct (ΔCt) values previously obtained for the 48 NR in (Muscat et al., 2013) to perform combined analyses of the expression correlation pattern of nuclear receptors and coregulators in this study.

### 2.2. Normalization of TaqMan low density arrays

TLDAs were analyzed by the relative quantification method of ΔCt. The geNorm (Vandesompele et al., 2002) algorithm in the Integromics StatMiner software package was used to select the most stable housekeeping genes to be used as reference for normalization. The Ct values (Supplementary Table 1) of each assayed genes were then normalized against the median of the selected housekeeping genes to obtain the ΔCt values.

Supplementary Figure 1 shows a boxplot of the ΔCt values of all coregulators in each of the 4 cohorts.

### 2.3. Filtering of outlier samples

Ct value measurements above 35 cycles were considered inaccurate. Samples for which more than 40% of genes had Ct values above 35 were deemed to be outliers and removed from subsequent analyses. Three samples (ERneg02TB003, ERneg08RMH456, ERpos07RMH188) were removed from both the nuclear receptor and coregulator TLDA data.

### 2.4. Principal Component Analysis

Principal component analysis of coregulator ΔCt values was performed in R with genes as variables to investigate the coregulators that best characterize the different patient cohorts. Three dimensional scatter plots of samples against the first 3 principal components were generated by the Scatterplot3d (Ligges and Machler, 2003) package.

### 2.5. Detection of differentially expressed genes

A moderated t-statistic test based on the limma framework implemented in the HTqPCR (Dvinge and Bertone, 2009) software package was used to detect differentially expressed genes. Genes having a Benjamini and Hochberg adjusted P-value less than 0.05 and at least 1.5 fold up or down-regulated were termed differentially expressed (Supplementary Table 2A).

### 2.6. Validation of combining NR and coregulator data by investigation of TLDAs cross-card batch effect

As NR and coregulator data were derived from distinct TLDAs, we determined the validity of a combined analysis of NR and coregulator data by comparing expression of the housekeeping control genes on both the NR and coregulator TLDAs. In addition, unsupervised hierarchical clustering of samples based on raw Ct values of all NR and coregulator genes was performed to determine if samples clustered in a batch (array) specific manner. We found expression of the control genes to be highly consistent between arrays (r² value 0.98) with no evidence of cross-card batch effects in clustering of samples (results not shown).

### 2.7. Correlation-based analysis of nuclear receptors and coregulators interactions

Pairwise Spearman Rank Correlation coefficients were calculated for all possible pairs of nuclear receptors and coregulators in a cohort-specific manner. Two genes were considered highly correlated if the Spearman Rank Correlation coefficient was ≥0.7 or ≤−0.7.

### 2.8. Pathways and functional analysis

Differentially expressed coregulators were separated into two groups based on the direction of change in expression (up- or down-regulated). The Ingenuity Pathways Analysis (IPA) Package was used to detect over-represented functional annotations and canonical pathways associated with up-regulated or down-regulated coregulators. Given the very small number of genes that were up-regulated, we assessed only enriched functional annotations for this group and retained annotations with enrichment P-value of less than 0.05. For down-regulated coregulators, over-represented canonical pathways with Bonferroni adjusted P-values of less than 0.05 were retained.
2.9. Breast cancer microarray datasets

In order to establish nuclear receptors and coregulators associated with prognosis, we compiled a large cohort of breast cancer gene expression microarrays with associated clinical data from previously published work. In total, we compiled 2227 breast cancer cases (of which 2189 were annotated with survival data) from multiple datasets profiled on the Affymetrix Hg133A platform and an additional 1992 breast cancer cases (of which 1853 were annotated with survival data) from the METABRIC study (Curtis et al., 2012) profiled on the Illumina HT-12 v3 platform. Supplementary Table 3 lists the microarrays datasets included in our analysis.

fRMA (Frozen Robust Multiarray Analysis) normalization (McCall et al., 2010) was performed on the Affymetrix microarrays to enable combined analysis of the multiple datasets. This algorithm utilizes pre-computed estimates of probe-specific effects and variances obtained from large publicly available databases. These pre-computed values are then used in concert with information from new array(s) for normalization and summarization, allowing separately processed datasets to be compared. fRMA normalized expression values were retrieved from inSilicoDb via its Biocductor package. Illumina HT-12 v3 array data (EAGD00100000210 and EAGD00010000211) were downloaded from the European Genome Phenome Archive and the normalized expression values as published in the METABRIC publication were used.

To identify genes correlated with ERα expression, breast cancer microarrays profiled on Affymetrix Human Hg133A arrays (listed in Supplementary Table 3) were divided into ERα positive and ERα − subsets based on the associated IHC ERα status. Pairwise Spearman Rank Correlation coefficients were calculated for each gene profiled on the array with each of the nuclear receptors: ESR1, REV-ERα Bα, GR, NOR1, LRH-1 and PR. Spearman Rank Correlation coefficient ≥ 0.7 or ≤ −0.7 were used to filter for genes showing high expression correlation with these nuclear receptors in either ERα positive or ERα negative breast cancers.

2.10. Identification of nuclear receptors and coregulators associated with prognosis

We employed the process outlined in Supplementary Figure 2 to identify nuclear receptors and coregulators, expression of which associates with patient survival. In this process, samples from each of the microarray platforms were randomly divided into two groups, one used for training and the other for validation. The Training and Validation sets were of comparable size, as was the ratio of patients with events to those without events. A bootstrap approach was employed to identify genes significantly associated with breast cancer survival in the Training Set. Specifically, we randomly sampled with replacement from the Training Sets and performed Univariate Cox Regression on the randomly selected subgroup. This process (sampling with replacement followed by Univariate Cox Regression) was repeated 1000 times. We then defined an r-index for each gene as the ratio of times out of the 1000 analyses that the gene showed significant association with survival (Adjusted Cox Regression P-value ≤ 0.05). We used distant metastasis free survival (DMFS) for samples where that information is available. For samples without associated DMFS information, the following survival measures in order of decreasing preference were used: relapse free survival (RFS), disease free survival (DFS) and disease specific survival (DSS).

For the Affymetrix Training samples, we identified 80 coregulators and 23 nuclear receptors with an r-index ≥ 0.75; 43 coregulators and 5 nuclear receptors were identified in the Illumina Training Set. Genes with an r-index of at least 0.75 on both Affymetrix and Illumina platforms were then selected for Multivariate Cox Regression adjusting for ERα status and node status. This process identified 3 nuclear receptors and 16 coregulators, expression of which significantly associates with patient survival independent of ERα status and node status on both microarray platforms (Supplementary Table 4).

2.11. Validation of NR and coregulator centered gene signature

Breast cancer microarray samples from the Validation Sets were used to assess whether the nuclear receptors and coregulators identified can predict patient survival. We assessed the performance of 3 gene signatures: (i) 19-gene signature (19 nuclear receptors and coregulators listed in Supplementary Table 4); (ii) 3-NR signature (PPARδ, PGR and GR); and (iii) 3 NR plus cytoskeletal gene signature (PPARδ, PGR, GR, CFL1 and GSN).

For each signature, samples from the Validation Sets were assigned risk scores calculated on the signed average of the signature genes expression (by the “sig.score” function from the genefu R package (Haibe-Kains et al., 2011)). In this method, the sample risk score is calculated as the average of the weighted expression of the genes in each signature. We used the sign and magnitude of the Univariate Cox Regression coefficient of expression of each gene obtained from all samples of the Discovery Sets as the input coefficients into sig.score. Samples were then classified into 3 different risk groups based on the assigned risk scores: low (bottom tertile), intermediate (middle tertile) and high (top tertile). The performance of each gene signature was then assessed on Kaplan Meier and a Concordance Index [as detailed in (Harrell et al., 1996)]. The Concordance Index represents the probability that, for a pair of randomly chosen comparable patients, the patient with the higher risk prediction will experience an event before the lower risk patient.

Further, using the Validation samples (Supplementary Figure 2) of METABRIC dataset, we tested the prognostic power of the 3NR, 3NR+cytoskeletal genes and the 19 genes in breast cancer subgroups identified by immunohistochemical marker staining. Subtype classification as assigned in the METABRIC dataset was used to categorize samples into subgroups (Normal-like, Basal, Luminal A, Luminal B, HER2s). Then the risk prediction method as described above is used to test the prognostic power of the signatures in each of the subtypes. In this analysis, we estimated input coefficients into sig.score in a subtype specific manner instead of using all Discovery samples.

Prognostic performance of the NR and coregulator based signatures identified was assessed in comparison to 48 other published signatures (Venet et al., 2011) with the same sample cohorts and outcome association method detailed above.
3. Results

3.1. Expression of coregulators differs between normal and breast cancer tissues

Supervised hierarchical cluster analysis demonstrates that a majority of coregulators were expressed differently in breast cancers compared with normal breast and that the expression of many coregulators also differs between ER\textsubscript{a} positive and ER\textsubscript{a} negative cancers (Figure 1A).

This observation is supported by unsupervised Principal Component Analysis (PCA), with normal and cancer samples clustering into two relatively well-defined subgroups reflecting the change in coregulator expression in cancer samples compared with normal tissue (Figure 1B). Expression levels of coregulators showed greater variation among cancer samples, evidenced by the greater spread observed on the PCA plot. In addition, expression profiles of coregulators also differ between ER\textsubscript{a} positive and ER\textsubscript{a} negative breast cancers. This difference is more evident when PCA was performed on the cancer samples independently of the normal samples (Figure 1C).

The changed expression of the majority of coregulators between normal and cancer cohorts supports the view that the separation of cohorts based on coregulator expression reflects...
multiple coregulator changes, each with an individually small contribution to the phenotype, rather than a few coregulators with a large effect on the phenotype. This is reflected in the small proportion of variance of each gene captured by each principal component in the PCA analysis (Supplementary Figure 3A). The first principal component only accounted for ~55% of the total variance in the dataset with each subsequent principal component capturing small increases in cumulative variance (Supplementary Figure 3B).

3.2. Pan-repression of coregulators in neoplastic breast tissues

The majority of coregulators showed decreased expression in cancer compared with normal tissue, with coregulator repression more pronounced in ERα negative than ERα positive cancers (Figures 1A and 2).

Overall, ERα negative breast cancers are associated with more extensive disruptions in the coregulator transcriptome,
with an increased number of genes differentially expressed plus an increase in the magnitude of the change in expression (Figure 2, Supplementary Table 2A). Cluster A in Figure 2 highlights coregulators that are down-regulated in both ERα positive and ERα negative breast cancers compared with normal breast. Cluster B in Figure 2 shows coregulators with increased expression in either ERα positive or ERα negative cancers; cluster C in Figure 2 highlights coregulator expression that differs between ERα positive and ERα negative tumors.

Using a moderated t-test implemented in the HTqPCR package, we identified differentially expressed coregulators. On the criterion of at least 1.5 fold up or down-regulation and Benjamini Hochberg adjusted P-value ≤0.05, we found 101 (42%) of the 238 coregulators profiled in ERα positive cancer samples are differentially expressed compared with normal, of which 11 genes are up-regulated and 90 down-regulated. In ERα negative cancers compared with normal, 203 (85%) of the 240 coregulators are differentially expressed, of which only 3 coregulators are up-regulated and the other 200 down-regulated (Figure 2, Supplementary Table 2).

Although a minority of coregulators are up-regulated in either ERα positive or ERα negative breast cancers compared with normal breast, most (10/12, 83%) of these are associated with Cellular Growth and Proliferation, and predicted to be activated, according to functional annotation on IPA (Fisher Exact P-value = 1.03E-05, activation z-score = 2.307, Supplementary Table 5A). Of the up-regulated coregulators, only CNNE1 is specifically up-regulated in ERα negative breast cancers while CNNA2 and TRIP13 are up-regulated in both ERα positive and ERα negative cancers; the other 9 coregulators are up-regulated specifically in ERα positive cancers (Figure 2B, Supplementary Table 2B).

In contrast with the specific association of coregulators up-regulated in breast cancer with NR pathways involved in proliferation and disease progression, coregulators that were down-regulated in either ERα positive or ERα negative breast cancers compared with normal were associated with a large range of NR pathways (Supplementary Table 5B). Among the most significantly down-regulated pathways in breast cancer are Estrogen Receptor Signaling; RAR activation; Glucocorticoid Receptor Signaling, Aryl Hydrocarbon Receptor Signaling; TR-RXR Activation; and PPAR signaling. As noted above, the numbers of down-regulated pathways in cancer compared with normal was greater for ERα negative than for ERα positive breast cancers (Supplementary Table 5B).

3.3. Relationships between expression of NR and coregulators in normal breast and breast cancer

To explore the potential functional parallels between NR and coregulators, we determined the correlation between NR and coregulator expression in each of the four breast cohorts (ERα positive, ERα negative, premenopausal normal, postmenopausal normal). This was done by combining the coregulator expression data of this study with previously published expression data for NR (Muscat et al., 2013) measured in the same cohort of samples. We first examined the pairwise expression correlation between each nuclear receptor and coregulator. Figure 3A shows the numbers of coregulators with expression highly correlated (Spearman Rank Correlation ≥0.7 or ≤-0.7) with that of each nuclear receptor, with NR ranked by decreasing NR-coregulator associations in normal postmenopausal breast samples, where the most abundant NR-coregulator associations are observed. Transcript levels of NR such as RARγ, RORγ, VDR, RARβ, HNF4γ, RXRβ, COUP-TF1/2, MR and PPARγ are correlated with the largest number of coregulators in these tissues. Although fewer associations are evident in normal premenopausal samples, and in either ERα positive or ERα negative breast cancers, cancer-related gains and losses in NR-coregulator association are observed (Figure 3A). For example, expression of RARβ correlates strongly with 97 coregulators specifically in postmenopausal normal but not in any of the other three sample cohorts. Likewise, expression of RXXRs shows strong inter-correlation with multiple coregulators specifically in ERα positive cancer samples. In contrast, AR expression correlates with fewer coregulators and the correlation is only observed in ERα positive cancer samples. All nuclear receptors display positive correlation with coregulators except for PPARγ, expression of which strongly and negatively correlates with more than 30 coregulators only in normal breast tissues (Supplementary Table 6).

A major cancer-associated shift in the expression correlation patterns of nuclear receptors and coregulators is supported by Figure 3B, which illustrates the proportion of strongly correlated NR-coregulator pairs that are observed only in normal breast, ERα positive or ERα negative samples as well as those commonly observed in multiple cohorts. Only 36% of all highly correlated NR-coregulator pairs are common to both normal and cancer samples (Figure 3B: ERα positive and Normal; ERα negative and Normal; ERα positive and ERα negative and Normal) while 31% are observed only in normal samples (Figure 3B: Normal only) and 32% are acquired correlations observed only in cancer samples (Figure 3B: ERα positive only; ERα negative only; ERα positive and ERα negative).

The proportion of highly correlated NR-coregulator pairs in ERα negative cancer samples not observed in normal samples is much higher (Figure 3B: ERα negative only, 26%) compared to that observed in ERα positive cancer samples (Figure 3B: ERα positive only, 3%). This suggests that overall, ERα negative breast cancers are associated with increased dysregulation of the NR-coregulator interaction network compared with ERα positive cancers.

Several nuclear receptors displayed pan-ERα status changes in expression correlation with coregulators. Of particular interest are the nuclear receptors ERβ, LXRα and PPARγ which show either no or negative correlation with most coregulators in normal breast samples but strong positive correlation with coregulators in cancer (Figure 3C). Other nuclear receptors display ERα status-specific shifts in expression correlation with coregulators (Figure 3D). AR, ESRRβ, RORγ, RXRα and RXXRγ are highly correlated with multiple coregulators in ERα positive cancers and these correlations are disrupted in ERα negative cancers. In contrast, REV-ERβ, GR, NOR1, LRH-1 and PGR displayed the opposite pattern and acquire strong correlations with multiple coregulators specifically in ERα negative cancers.
Figure 3 — Expression correlation patterns of nuclear receptors and coregulators. A) Bar plot of the numbers of coregulators highly correlated (Spearman Rank Correlation ≥ 0.7) with each nuclear receptor in each of the 4 cohorts, ranked by NR with greatest numbers of associations in post-menopausal normal samples. B) Pie chart showing the proportions of highly correlated NR-coregulator gene pairs that are observed uniquely or in multiple cohorts as a fraction of all highly correlated gene pairs observed. C) lists nuclear receptors displaying pan-cancer changes in correlation pattern with coregulators. D) lists coregulators showing ERα status specific correlation pattern with coregulators. For each panel of C) and D), cohorts are listed in rows and all coregulators are tiled across columns. Heatmap cells are colored according to the Spearman Rank Correlation coefficients of each NR-coregulator pair.
3.4. NR-coregulator interactions acquired in cancer have stronger prognostic power than NR-coregulator interactions disrupted between normal and cancer

To determine whether nuclear receptors that become strongly correlated with coregulators in ERα-negative breast cancer (Figure 3D panel 1) are associated with DMFS, we made use of published microarray datasets with associated clinical information (Validation samples, refer to Supplementary Figure 2 and Materials and Methods). The prognostic value of those nuclear receptors with disrupted correlations with multiple coregulators in ERα-negative cancers (Figure 3D panel 2) was also explored.
Breast cancer cases were classified into low, intermediate and high risk-groups based on risk scores calculated from the expression of the nuclear receptors in each of Panel 1 and Panel 2 in Figure 3D. Kaplan Meier plots of the probability of DMFS of the risk-groups for ER$\alpha$ positive and ER$\alpha$ negative breast cancers show that the expression of nuclear receptors displaying acquired correlations with multiple coregulators in ER$\alpha$ negative cancers (Group 1: Rev-ERβ, GR, NOR1, LRH-1 and PGR) significantly segregates patients at low from those at high risk (Figure 4), and this is confirmed in an independent dataset (Supplementary Figure 4). The expression of nuclear receptors showing disrupted interactions (Group 2: AR, ESRR$\gamma$, ROR$\gamma$, RXR$\alpha$ and RXR$\gamma$) has no predictive power for patient survival in ER$\alpha$ negative cancers while having modest predictive power in ER$\alpha$ positive cancers (Figure 4 and Supplementary Figure 4).

Given that Group 1 (Rev-ERβ, GR, NOR1, LRH-1 and PGR) included NR linked with ERs expression and/or function, we asked whether the genes that acquired correlations with Group 1 NR in ER$\alpha$ negative cancers were ER$\alpha$-associated, perhaps reflecting ER$\alpha$-independent signaling in these cancers. Genes most strongly correlated with ERs were identified in the Affymetrix datasets (921 genes; Pearson correlation coefficient ≥0.7 or ≤−0.7), and their correlation with the Group 1 NR determined. While GR and LRH-1 were correlated with few/no ER$\alpha$-correlated genes in ER$\alpha$ positive cancers, they acquired such correlations in ER$\alpha$ negative cancers (Figure 5), supporting the possibility that these NR may be involved in ER$\alpha$-associated but ER$\alpha$-independent pathways. Although correlations were detected between PGR and these genes in ER$\alpha$ negative cancers, the significance level did not pass the cut-off threshold of 0.7 (not shown).

3.5. Prognostic value of nuclear receptors and coregulators in breast cancer

The prognostic power of the small number of nuclear receptors that acquired correlation with multiple coregulators in breast cancers (Figure 4) raised the question of whether coregulators, alone or in combination with nuclear receptors, were associated with breast cancer outcome. To address this question, we performed a comprehensive survey of the association of all nuclear receptor and coregulator expression levels with patient survival in large cohort of publicly available breast cancer microarray datasets. In total, we compiled gene expression data from over 4000 breast cancer tissue microarrays profiled on two microarray platforms: Affymetrix Hg133A (2189 samples) and Illumina HT-12 v3 (1853 samples) (Supplementary Table 3). Given the multiplicity of markers already linked to clinical outcome in breast cancer, and the challenges in identifying prognostic gene signatures that perform better than chance alone (Venet et al., 2011), we employed the rigorous analytical approach outlined in Supplementary Figure 2 (Materials & Methods). Using the Affymetrix Training samples, as previously noted, we identified 80 coregulators and 23 NR for the Affymetrix datasets (921 genes; Pearson correlation coefficient ≥0.750, and 43 coregulators and 5 NR for the Illumina Training Set.

There were 3 nuclear receptors (GR, PGR, PPAR$\beta$) and 16 coregulators with expression profiles robustly associated with patient survival, independent of ER$\alpha$ status or node status on both microarray platforms (listed in Supplementary Table 4). These genes have functional roles relevant to cancer biology including: cell cycle regulation (CDKN1C, CCNA2, CCNE1, PA2G4), cytoskeletal organisation (CLF1, GSN), chromatin modifiers (SMARCA2, HMGB2), known oncogenes or cancer-related genes (RPL7A, NSD1, PRAME, FUS), and also include 4 (33%) of the 12 genes that are increased in cancer compared with normal breast (Figure 2: TRIP13, CCNE3, CCNA2, PRAME). Figure 6A represents a heatmap of the r-index of each of the 19 genes which represent the number of times out of 1000 bootstrap Cox regression analyses of different sample subgroups in which the expression of the
Figure 6 — Nuclear receptors and coregulators with prognostic values. A) Heatmap of r-index values of the 19-gene signature. The r-index is a measure of the robustness of a gene’s association with survival and represents the number of analyses out of 1000 Cox Regression analyses of randomly sampled subgroups of samples in which the gene is found to be significantly associated with patient survival. B) Kaplan Meier plot showing significant segregation of samples (1094 validation samples profiled on Affymetrix Hg133a arrays) classified into Low, Intermediate and High risk-groups based on the expression of the 19 genes. C) Similar to B) but using a second set of 927 validation samples profiled on Illumina
gene is found to be associated with survival. The r-index thus represents the robustness of a gene’s association with patient survival.

The combined expression of the 19 genes identified can segregate low-risk from high-risk breast cancers. Samples from the Validation Sets (1094 samples profiled on Affymetrix Hg133A and 927 samples profiled on Illumina HT-12 v3) were classified into 3 different risk groups based on risk scores calculated from the expression of the 19 genes: low (bottom tertile), intermediate (middle tertile) and high (top tertile). Kaplan Meier analysis showed significant segregation of the three risk-groups (Figures 6B and C). Furthermore, classification of samples into low, intermediate and high-risk-groups on expression profiles of just the 3 nuclear receptors (GR, PGR, PPARα) also results in significant segregation of low, intermediate and high-risk samples (Figure 6D). We also observed that addition of the two genes known to be involved in cytoskeletal organization (CFL1 and GSN) into the survival model together with the 3 NR resulted in further separation of the risk groups (Figure 6E).

We also tested the prognostic power of the 3NR, 3NR-cytoskeletal genes and the 19-gene signature in breast cancer subgroups identified by immunohistochemical marker staining. The signatures had little or no predictive power in Basal subtypes, but identified subgroups of patients with different prognosis in the Luminal A, Luminal B and Normal-like subtypes (not shown).

3.6. Performance comparison

A limitation of current breast cancer signatures is that most have been shown to be no more strongly associated with outcome than random sets of genes, although signatures containing over 100 genes are more significant (Venet et al., 2011). We compared the performance of NR and coregulator centered gene signatures with 48 other published breast cancer signatures (compiled from Venet et al. (2011)). Using samples from our Validation Sets, we calculated sample risk scores based on the expression of the genes in each signature. The association of the calculated risk scores with patient survival was then tested by Univariate Cox Regression and Concordance Index.

Figure 7 shows the Concordance Index for each gene signature calculated for (i) all samples in the Validation Sets (Figure 7A) and (ii) ERα negative samples of the Validation Sets (Figure 7B). The 19-gene signature performance is comparable to signatures of larger sizes, in all cases (Figure 7A) and more particularly in ERα negative cases (Figure 7B). The signatures for the 5 NR that acquired correlations in ERα negative cancer (from Figure 4); and the 3 NR Plus Cytoskeletal gene signatures do not rank well when all samples of the Validation Sets are used (Figure 7A), they are, however, able to predict patient survival for ERα negative samples better than signatures with larger gene numbers (Figure 7B). We also observed that prediction accuracy of gene signatures varies, depending on the microarray platform used, with some signatures performing better on one microarray platform than the other. In ERα negative breast cancers, there are two published signatures that have comparable or fewer numbers of genes than the NR-coregulator signatures identified in this study, namely the Yu et al. (2007) (14 genes) and Buffa et al. (2010) (3 genes) signatures (Figure 7B). There was no overlap between the 19-gene NR-coregulator signature identified in this study, and the genes being contained within either the Yu or Buffa signatures, nor did functional analysis reveal any overlap, with the Yu signature being associated with cell surface receptor linked signal transduction, immune response and cell motility (not shown).

It is known that most of the genes in breast cancer signatures are correlated with proliferation, and this is a major confounder in outcome prediction (Venet et al., 2011). To assess the impact of proliferation on the prognostic power of the NR-coregulator gene signatures that we identified, we adjusted for the effect of proliferation by adjusting for the expression of the meta-PCNA signature, which comprises the genes most strongly correlated with Proliferating Cell Nuclear Antigen (PCNA) gene expression (Venet et al., 2011). We found that while for ERα positive cancers, meta-PCNA adjustment reduced the prognostic power of most of the gene signatures we examined, it had minimal impact for ERα negative cancers and the 19-gene signature; the 5 NR that acquired correlations in cancer signature (from Figure 4); and the 3 NR plus cytoskeletal gene signature remained predictive of outcome in this cohort (Supplementary Figure 5).

4. Discussion

This study comprises the first combined profiling of nuclear receptors and coregulators in the human breast. Notable strengths include the robust quantitation of coregulator transcripts in cohorts of curated human breast tissues, both normal and cancer, plus validation cohorts comprising of all publicly available breast cancer microarray datasets across all platforms. In this regard the study provides significant pathophysiologically relevant insights to supplement studies based on cell line models (Kittler et al., 2013). Our findings reveal an almost pan-repression of coregulator expression in neoplastic breast tissue and show that ERα negative cancers are associated with more widespread disruptions of the coregulator transcriptome. In addition, the study highlights important cancer-associated shifts in nuclear receptor and coregulator expression patterns. We then complemented the expression profiling of nuclear receptors and coregulators with a comprehensive survey of the prognostic value of NR and coregulators in published breast cancer microarrays to identify nuclear receptors and coregulators with high prognostic significance.

The coregulator transcriptome was strikingly disrupted in cancer compared with normal breast, with a small number of coregulators up-regulated and the majority repressed. Up-regulation was overwhelmingly associated with proliferation,
consistent with different regulatory networks being operative in both ERα positive and ERα negative breast cancers; differential expression of the two cyclins CCND1 and CCNE1 in ERα positive and ERα negative cancers suggest different mechanisms for controlling progression of G1 to S phase of the cell cycle in these two subgroups. CCND1 is up-regulated in ERα positive and down-regulated in ERα negative cancers, consistent with published reports that ERα can induce CCND1 expression (Cicatiello et al., 2010). It is possible that in ERα negative cancers progression from G1 to S phase is mediated through CCNE1, which is specifically up-regulated in ERα negative cancers in contrast with the decreased level of CCND1. In addition, the coregulators VAV3, a Rho GTPase guanine nucleotide exchange factor, and SPDEF, the SAM pointed domain containing ETS transcription factor, are up-regulated specifically in ERα positive breast cancers and down-regulated in ERα negative breast cancers, consistent with previously published reports that VAV3 overexpression enhances ERα-mediated signaling (Doolan et al., 2008), and that SPDEF is associated with patient survival in ERα positive breast cancers (Sood et al., 2009). Up-regulation of both RARα and CRABP2 (cellular
retinoic binding protein 2) specifically in ERα positive cancers is further evidence for differential regulation, in the case of retinoic acid signaling, in ERα positive and ERα negative breast cancer subgroups.

Pan-repression of coregulator expression in neoplastic breast tissues was not unexpected given our earlier observation that breast cancer is associated with a pan-repression of nuclear receptors except for EAR2, AR, RARα and the NR4A subgroup (Muscat et al., 2013). The specific mechanisms whereby repression of nuclear receptors and coregulators is involved in breast cancer pathogenesis are unclear. In cancers multiple NR-mediated pathways associated with cell differentiation were repressed, eg Retinoic Acid Activation and Glucocorticoid Receptor Signaling pathways; with up-regulation of cell cycle related genes this would have a net effect of increased cell growth and proliferation.

In the absence of detailed genomic localization maps and protein–protein interaction data for most nuclear receptors and coregulators in breast cancer tissues, and the impossibility of experimentally deriving functional data from human tissue cohorts, we determined the correlation between NR and coregulator expression as a measure of functional relatedness. While setting very strong correlation thresholds for significance, we acknowledge that correlated expression indicates, but does not prove, functional co-operation. Nevertheless, this approach has revealed interesting insights into the complex regulatory networks of nuclear receptors in the human breast.

First, most of the strong NR-coregulator correlations occurred in the normal postmenopausal breast, with NR including RARγ, RORγ, VDR, RARβ, HNF4γ, RXRβ, COUP-TF1/2, MR and PPARγ associating with the largest number of coregulators, and pointing to the potential for active signaling networks around these receptors in normal breast. These NR were also associated with coregulators in pre-menopausal breast, but less frequently. Many if not all of these associations were lost in cancer, consistent with the overall down-regulation of both NR (Muscat et al., 2013) and coregulators (this study) in cancer compared with normal tissue. NR that acquired expression correlations (Rev-ERB, GR, NOR1, LRH-1, PR) specifically in ERα negative breast cancers predicted poorer outcome for those patients, suggesting the potential for clinically significant activation of new signaling networks associated with unrestrained cancer biology involving these NR specifically in ERα negative breast cancers. Interestingly, NR whose associations with coregulators are disrupted in ERα negative breast cancers are not as strongly associated with patient outcome. This finding is particularly interesting in the context of regulatory networks, given that cancer is not simply a disease with a genetic basis but is ultimately driven by perturbations at the network level (Creixell et al., 2012). During cancer development, it is likely that signaling processes are not simply disrupted but are also re-wired resulting in new signaling landscapes favorable to the development of cancer cells. It is therefore perhaps logical that those re-wired signaling processes favoring maintenance and progression of cancer will have a stronger impact on patient outcome.

Secondly, we observed a positive correlation between the expression of most nuclear receptors and coregulators, with the exception of PPARγ, ERβ and RXRγ, expression of which inversely correlates with multiple coregulators specifically in normal but not cancer tissues. It is arguably more logical for the expression of nuclear receptors to be positively correlated with that of coregulators for co-operational control of gene regulation. It is more unexpected to observe patterns of widespread inverse correlations, as in the case of PPARγ; PPARγ and its coregulator PPARγ1A are both strongly down-regulated in both ERα positive and ERα negative cancers, consistent with reports that the presence of PPARγ is associated with an anti-tumorigenic effect (Bonofigo et al., 2006; Lu et al., 2005; Suzuki et al., 2006). PPARγ positively correlates with very few genes, among which are the tumor suppressor CAV1, the testosterone 17-beta-dehydrogenase gene HSD17B13 and the aldo-keto reductase AKR1C1/AKR1C2 gene. HSD17B13 is involved in estrogen biosynthesis and responsible for converting estrone to estradiol-17-beta while AKR1C1/AKR1C2 catalyses the reduction of progesterone to the inactive 20α-hydroxy-progesterone. Both HSD17B13 and AKR1C1/AKR1C2 are strongly down-regulated, consistent with attenuated steroid metabolism, in both ERα positive and ERα negative cancer samples. The strong positive correlation of PPARγ with HSD17B13 and AKR1C1/AKR1C2 in normal breast only, its repression in cancer and the striking inverse correlation pattern between PPARγ and multiple coregulators suggest that dysregulation of this NR might contribute to altered steroid metabolism and thereby to aberrant control of estrogen and progesterone mediated signaling in breast cancer.

The demonstration that NR-coregulator associations in normal breast are disrupted in cancer, and the acquisition of associations unique to cancer tissues, signal the potential for NR-coregulator networks not only to be implicated in the biology of breast cancer but also relevant to clinical outcome. Testing of such possibilities, in order to discover meaningful associations between gene signatures and clinical outcome, is fraught with pitfalls in that most current breast cancer signatures are not robust across datasets and microarray platforms, and moreover are no more strongly associated with outcome than random sets of genes (Venet et al., 2011). This is probably due to the extremely high individual variability of sampled tumor expression profiles resulting in passenger signals overwhelming the real cancer signature. To circumvent this limitation, we employed a bootstrap resampling approach on 2 different microarray platforms. There were only 19 genes (3 nuclear receptors (GR, PGR, PPARα) and 16 coregulators) with expression profiles robustly associated with patient survival, independent of ERα or node status on both microarray platforms. Low levels of expression of the signature genes identify patients with a significantly lower probability of distant metastases, across both the microarray platforms, whereas high expression is associated with poorer prognosis. Low expression of the 3 NR alone, or of the 3 NR with the cytoskeletal genes, also stratified patients with better prognosis. Two of the 3 NR in the 19-gene signature (GR, PPARα) have been identified by ChIP-chip to be central nodes in a regulatory network of NR and associated transcription factors in MCF-7 cells (Kittler et al., 2013), and their robust association with clinical outcome identified in this study highlights the importance of signaling networks specifically associated with these 3 NR in breast cancer.
To evaluate how well NR-coregulator-based signatures performed in comparison to other published breast cancer signatures, and mindful of the weak predictive power of published signatures containing fewer than 100 genes, the performance of the NR-coregulator-based signatures was compared to that of all published breast cancer signatures. The performance of the 19-gene NR-coregulator-based signature is comparable to many other signatures of much larger sizes in predicting outcome for all breast cancer samples (both ERα positive and ERα negative), highlighting the power of predictive signatures derived from the quantitative determination of altered gene expression between normal breast and breast cancers.

This study highlights major differences in coregulator expression and NR-coregulator interaction between ERα positive and ERα negative breast cancers that reflects divergent regulatory controls in these two breast cancer subtypes. Specifically, a majority of the profiled coregulators (81%) are differentially expressed between ERα positive and ERα negative breast cancers, with ERα negative cancers being associated with an increase in both the number of coregulators being repressed as well as an increase in the magnitude of the repression. The change in coregulator expression is associated with ERα-specific perturbations in NR-coregulator expression. Specifically, RXRα, TRα, COUP-TF2, RORγt showed strong associations with multiple coregulators in ERα positive breast cancers suggesting the increased importance of retinoic acid signaling in this subtype and consistent with proliferation being a major influence on patient outcome in ERα positive cancers. ERα negative cancers, on the other hand, are associated with acquired interactions between REV-ERBα, TR4, GR, PPARα, NOR1, LRH-1 and a large number of coregulators. NR-coregulator interactions acquired specifically in ERα negative cancers have the potential to impact a wide range of cellular pathways and processes including chromatin remodeling, glucocorticoid receptor signaling and cell cycle regulation. Overall, our results suggest that ERα negative breast cancer is associated with increased severity in perturbation of normal NR-coregulator interactions and that there exists significant differences in the way normal gene regulatory networks are altered in these breast cancers. As a consequence, some NR and coregulators have stronger influence on patient outcome in ERα negative or positive breast cancer compared to the other. We observed that the 3NR plus cytoskeletal genes obtained the highest Concordance Index for ERα negative cancer, with its performance in predicting outcome for ERα positive cancer less robust. In addition, PCNA-adjustment has minimal influence on the prognostic power of most gene signatures in ERα negative cancers. It is thus possible that while proliferation is a major influence on outcome association in ERα positive cancers, it is not the principal factor that influences outcome for ERα negative breast cancers. ERα negative breast cancer outcomes might be driven by factors including cell movement and metastasis, as suggested by the strong influence of cytoskeletal genes in predicting patient outcome for this subgroup. In addition, we demonstrated that NR showing acquired correlations with multiple coregulators specifically in ERα negative cancers have stronger influence on patient outcome in these cancers. This observation emphasizes the importance of examining not only “loss-of-function” disruptions but also “gain-of-function” perturbations in the search for therapeutic and prognostic markers for breast cancer.

In conclusion, this study has identified networks of nuclear receptor-coregulator interactions active in normal breast but disrupted in breast cancer, and moreover provides evidence that signatures based on NR networks disrupted in cancer can provide important prognostic information in breast cancer patients, including ERα negative patients, for whom existing prognostic measures are limited.

Acknowledgments

This research was supported by a Collaborative Program Grant (CG-08-03) from the National Breast Cancer Foundation Australia. GEOM, ERS, PJF and CLC are research fellows of the National Health and Medical Research Council of Australia; TBD is supported by an APA Scholarship. Prince Henry’s Institute is supported by the Victorian Government’s Operational Infrastructure Program. Breast cancer and normal tissues were provided by Australian Breast Cancer Tissue Bank (ABCTB), which is supported by the National Health and Medical Research Council of Australia (633291), the Cancer Institute NSW (12/RIG/1-16) and the National Breast Cancer Foundation (IF-12-05); or by the Victorian Cancer BioBank Australia, which is supported by the Victorian Government. ABCTB tissues and samples were made available to researchers on a non-exclusive basis. Normal breast biopsies were obtained from the Susan G. Komen for the Cure Tissue Bank at the IU Simon Cancer Center. We thank contributors to the Susan G. Komen for the Cure Tissue Bank, including Indiana University that collected the samples used in this study, as well as parents and families whose participation and help made this work possible.

Appendix A.
Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2014.03.017

REFERENCES


