

The Landscape of Candidate Driver Genes Differs between Male and Female Breast Cancer

Ida Johansson*, Markus Ringnér, Ingrid Hedenfalk

Division of Oncology, Department of Clinical Sciences, Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

Abstract

The rapidly growing collection of diverse genome-scale data from multiple tumor types sheds light on various aspects of the underlying tumor biology. With the objective to identify genes of importance for breast tumorigenesis in men and to enable comparisons with genes important for breast cancer development in women, we applied the computational framework COpy Number and EXpression In Cancer (CONEXIC) to detect candidate driver genes among all altered passenger genes. Unique to this approach is that each driver gene is associated with several gene modules that are believed to be altered by the driver. Thirty candidate drivers were found in the male breast cancers and 67 in the female breast cancers. We identified many known drivers of breast cancer and other types of cancer, in the female dataset (e.g. *GATA3*, *CCNE1*, *GRB7*, *CDK4*). In contrast, only three known cancer genes were found among male breast cancers; *MAP2K4*, *LHP*, and *ZNF217*. Many of the candidate drivers identified are known to be involved in processes associated with tumorigenesis, including proliferation, invasion and differentiation. One of the modules identified in male breast cancer was regulated by *THY1*, a gene involved in invasion and related to epithelial-mesenchymal transition. Furthermore, men with *THY1* positive breast cancers had significantly inferior survival. *THY1* may thus be a promising novel prognostic marker for male breast cancer. Another module identified among male breast cancers, regulated by *SPAG5*, was closely associated with proliferation. Our data indicate that male and female breast cancers display highly different landscapes of candidate driver genes, as only a few genes were found in common between the two. Consequently, the pathobiology of male breast cancer may differ from that of female breast cancer and can be associated with differences in prognosis; men diagnosed with breast cancer may consequently require different management and treatment strategies than women.

Citation: Johansson I, Ringnér M, Hedenfalk I (2013) The Landscape of Candidate Driver Genes Differs between Male and Female Breast Cancer. PLoS ONE 8(10): e78299. doi:10.1371/journal.pone.0078299

Editor: Irina U Agoulnik, Florida International University, United States of America

Received: June 25, 2013; **Accepted:** September 17, 2013; **Published:** October 23, 2013

Copyright: © 2013 Johansson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the Swedish Cancer Society, the G Nilsson Cancer Foundation, the Mrs. B Kamprad Foundation, the King Gustaf V's Jubilee Foundation, the Lund University Hospital Research Foundation and governmental funding of clinical research within the national health services (ALF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: Ida.Johansson@med.lu.se

Introduction

Male breast cancer (MBC) is a rare cancer form that has not been well studied, and there is hence limited knowledge of the disease at the genetic and molecular levels. A few molecular profiling studies of MBC have been performed by us and others [1–4], but the rarity of the disease makes the collection of large sample cohorts difficult. Although MBC shares many similarities with female breast cancer (FBC), distinct differences have been reported in age distribution, levels of hormone receptors, prognosis and survival [5–7], and on the transcriptional and genomic levels [1–4]. No large comprehensive studies of the etiology of the disease or randomized trials for optimizing patient management have been conducted to date, and MBC is often likened to hormone receptor-positive disease in postmenopausal women [8]. Thus,

recommendations for managing MBC have been based on available knowledge and guidelines for FBC. Several studies have nevertheless reported various differences between MBC and FBC based on gene expression (GEX) [2,4], copy number (using array comparative genomic hybridization, aCGH) [1,3], microRNA [9,10] and SNP levels [11]. Among these, we have shown in two previous studies that MBC, like FBC, is a heterogeneous disease, and many differences between MBC and FBC were revealed [1,2]. In a transcriptional profiling study we described two new subgroups of breast cancer (Luminal M1 and M2, respectively) that did not resemble any of the intrinsic subgroups reported in FBC, and as such may be unique to males. Luminal M1 tumors (70% of the MBC tumors) seemed to be more aggressive and were associated with worse prognosis and also appeared to have a less activated estrogen receptor (ER) module, while Luminal M2 tumors (30% of the

MBC tumors) displayed an up-regulated immune response and a more activated ER module [2]. These findings suggest that MBCs are in fact different from FBCs, and that men diagnosed with breast cancer may require other treatment strategies than women.

Many methods are currently available for producing various types of genome-scale data, and the amount of such data from human cancers is growing rapidly. There is however a need to combine the different types of data, reflecting various aspects of tumor biology, to gain a better and deeper understanding of the underlying biology [12]. In this context, one major challenge is to combine copy number and transcriptional data in a biologically meaningful fashion. Various methods for integrating copy number and transcriptional data are available, all aimed at differentiating between driver and passenger genes [13–15]. In this study, we used the computational framework COpy Number and EXpression In Cancer (CONEXIC), which integrates aCGH and GEX data to identify cancer drivers among all passenger genes aberrantly expressed in tumors. This approach not only identifies candidate drivers, but also associates each driver with gene modules that are believed to be altered by the driver [13]. In contrast, most previously described network-based methods have focused only on a single type of genome-scale data [16,17].

The main aim of this study was to identify potential candidate driver genes that drive proliferation and metastasis, thereby distinguishing them from passenger genes, in MBC, an approach that has not been comprehensively explored before. Secondly, we aimed to compare the landscape of potential candidate drivers between breast cancers diagnosed in male and female patients to explore differences as well as similarities.

Materials and Methods

Ethics statement

The study was approved by the regional Ethics Committee in Lund (2012/89), waiving the requirement for informed consent for the present study.

Patient material

The fresh frozen MBC tissues were obtained from The Southern Sweden Breast Cancer Group's tissue bank at the Department of Oncology, Skåne University Hospital, from Uppsala University Hospital, and Örebro Hospital, all in Sweden. Patients were diagnosed between 1983 and 2009. All cases with good quality aCGH [1] and GEX [2] data from our previous studies were included in the present study ($n=53$). A physician reviewed all patient charts and compiled all clinico-pathological data. A pathologist graded all tumors to current pathological standard when paraffin blocks were available; all histological grades were represented. ER, progesterone receptor (PR) and HER2 were re-evaluated (see 7,18 for further details). The patients had received different combinations of adjuvant treatment, including hormonal, chemotherapy and radiation treatment (see Table 1 for further information).

Table 1. Patient and tumor characteristics of the 53 MBC cases.

| Clinico-pathological characteristics | | N | % |
|--------------------------------------|--|-------|----|
| Age at diagnosis | Mean | 68 | |
| | Range | 42-92 | |
| ER status | Positive | 41 | 77 |
| | Negative | 3 | 6 |
| | N/A | 9 | 17 |
| PR status | Positive | 36 | 68 |
| | Negative | 8 | 15 |
| | N/A | 9 | 17 |
| HER2 status | Positive | 2 | 4 |
| | Negative | 27 | 51 |
| | N/A | 24 | 45 |
| BRCA2 mutation status | Positive | 3 | 6 |
| | Negative | 6 | 11 |
| | N/A | 44 | 83 |
| Histology | DCIS | 1 | 2 |
| | Invasive cancer in combination with DCIS | 9 | 17 |
| | Invasive cancer | 38 | 72 |
| | N/A | 5 | 9 |
| Histological grade | I | 2 | 4 |
| | II | 22 | 42 |
| | III | 11 | 21 |
| | N/A | 18 | 34 |

doi: 10.1371/journal.pone.0078299.t001

The FBC cohort included 359 tumors, representing all intrinsic subgroups of FBC and is described in [19].

Bioinformatic analyses

The computational framework CONEXIC, that integrates aCGH and GEX data for detecting candidate drivers, was used. CONEXIC uses an integrative Bayesian scoring approach [13] and is inspired by Module Networks [20]. The basic assumptions for CONEXIC are:

- 1 Drivers should modulate the gene modules' expression.
- 2 The driver's expression controls the malignant phenotype rather than the copy number, since also other mechanisms can cause altered expression of the driver.

For a detailed description of the CONEXIC method, see Akavia et.al. [13]. Briefly, the learning algorithm of CONEXIC consists of three main steps:

- 1 Selection of candidate drivers from commonly aberrant regions of the genome.
- 2 A Single Modulator step that finds the initial association between the candidate drivers and gene modules.
- 3 Improvement of the Single Modulator modules by an iterative network learning step.

The 53 MBCs and 359 FBCs with paired GEX and aCGH data were analyzed in the same manner. First, JISTIC (identification of significant targets of cancer) was used to identify commonly aberrant regions (i.e. more often than what

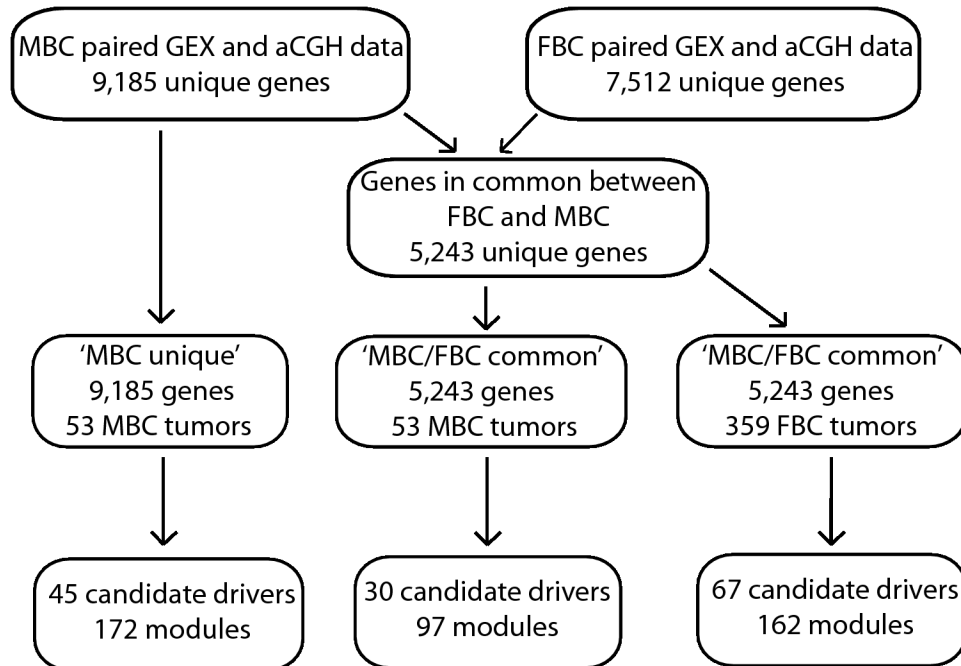


Figure 1. Flowchart outlining the steps in the CONEXIC analysis.

doi: 10.1371/journal.pone.0078299.g001

would be expected by chance) of the genome in the tumors [21]. JISTIC is an improvement of the GISTIC (Genomic Identification of Significant Targets In Cancer) algorithm [22], and uses both the frequency and magnitude of the alterations in the genome to calculate a statistical score. GISTIC has problems in identifying sub-regions when large regions are amplified or deleted, while JISTIC includes an algorithm with the capability to find these regions [21].

The GEX datasets were normalized as described before [2,19]. Probes with a standard deviation <0.25 on a \log_2 scale were removed and probe sets for genes were merged (inconsistent genes were removed). Genes without aCGH data were filtered out, resulting in a gene set of 9,185 unique genes for the MBCs and a gene set of 7,512 unique genes for the FBCs. Genes in common between the MBC and FBC datasets were then selected, resulting in a final set of 5,243 genes. CONEXIC was applied to the MBCs based on both the 'MBC/FBC common' (5,243 genes) and the 'MBC unique' (9,185 genes) gene sets (Figure 1).

The candidate driver lists were used in the Single Modulator step to establish an initial association to gene modules; non-parametric bootstrapping was used [23] and repeated 100 times. The candidate drivers that were selected in at least 90% of the runs were then used as candidate gene lists for the final run of the Single Modulator step. The result file was then split into two modules per candidate driver, one module each for genes positively and negatively correlated with the candidate driver. Genes that were not assigned to any gene module were clustered using K-means; 20 stable K-means clusters were found in the MBC dataset and five in the FBC dataset. The final modules from Single Modulator (after the split) and the K-

means modules were used as a starting point in the Module Network step, and then by using an iterative approach each gene was re-assigned into a module that best modeled its behavior. The Bayesian score is improved at each iteration and also here a non-parametric bootstrap was used and repeated 100 times. Candidate drivers that were selected in at least 40% of the runs were used as the candidate gene lists for the final run of the Module Network step. To rank the candidate drivers from the final list, CONEXIC scores were calculated. The candidate driver with the highest CONEXIC score is the most probable driver. The CONEXIC score was calculated with a Normal Gamma score; a score is calculated for each module and the scores for all modules regulated by a candidate driver were summed up to generate the CONEXIC score.

Each driver is associated with gene modules and a threshold expression level (split value) of the driver for each gene module divides the tumors into two groups, one with tumors that express the driver below the threshold and one with tumors that express the driver above the threshold.

An automated procedure, Literature Vector Analysis (LitVAn), was used to find over-represented terms associated with the genes in the modules in scientific papers from the NCBI Gene database [13].

CONEXIC was also applied separately to FBCs in the respective intrinsic subgroups (101 Luminal A, 65 Luminal B, 78 Basal-like and 42 HER2-enriched, respectively) and it was run in the same manner as described above. Unfortunately, the limited sample size precluded this possibility in the two previously identified subgroups of MBC (Luminal M1 and M2) [2].

Finally, a ranked-based module activity score was calculated for each MBC tumor for eight gene modules reflecting biologically relevant transcriptional programs found in FBC as previously described [24].

The GEX and aCGH data have been deposited in NCBI's Gene Expression Omnibus (GEO) and are available through GEO Series accession numbers [GSE31259] and [GSE50512], respectively.

Statistical analyses

All statistical calculations and figures were generated in R [25]. The survival and survcomp packages were used for the survival with distant metastasis free survival (DMFS) as endpoint. All P-values are two-sided.

Results

JISTIC results

A JISTIC analysis was performed to find significantly aberrant regions among the MBC tumors, whereby 67 significant genomic aberrations were identified (51 gains and 16 losses). The regions contain 1,223 genes (769 gained and 454 lost). By comparison, when GISTIC was applied to the same tumors, only 39 regions (25 gains and 14 losses) were found [1]. Correspondingly, 208 significant regions of genomic aberrations were found (89 gained and 119 lost), harboring 924 genes (450 gained and 474 lost) when JISTIC was applied to the FBC tumors.

CONEXIC results in MBC and FBC

Thirty candidate drivers were found among the MBC tumors using the 'MBC/FBC common' gene set (Table 2 and Table S1) and 67 among the FBC tumors (Table 3 and Table S2). The 30 candidate drivers identified in the MBC tumors regulated 97 modules and the 67 candidate drivers identified in the FBC tumors regulated 162 modules (Figure 1). Male and female breast cancers displayed remarkably different landscapes of candidate drivers, as only two candidate drivers were found in common (*TAF4* and *CD164*). When CONEXIC was applied separately to the intrinsic subgroups of FBC, only three more candidate drivers in common with MBC were found (*ARHGAP30*, *COG3* and *SPAG5*). This is in line with our previous results on transcriptional and copy number profiling, where we showed that MBCs appeared to differ from the intrinsic subgroups of FBC, and furthermore that MBCs could be divided into novel, distinct subgroups [1,2].

When CONEXIC was applied to MBCs using the 'MBC unique' gene set, 45 candidate drivers were identified (Figure 1) and 17 of these were shared with the 30 candidate drivers from the 'MBC/FBC common' gene set (17/30 (57%)). Reassuringly, among the top 15 candidate drivers from the 'MBC unique' gene set, 11/15 (73%) were also found in the 'MBC/FBC common' gene set (data not shown).

CONEXIC analyses of the intrinsic subgroups of FBC identified 28 candidate drivers in Luminal A tumors, 17 in Luminal B tumors, 22 in Basal-like tumors and 12 in HER2-enriched tumors (data not shown). Many of these candidate

Table 2. Top 15 candidate driver genes in male breast cancer.

| Candidate drivers | CONEXIC | | |
|-------------------|---------|----------|---|
| MBC* | score | Cytoband | LitVAN** |
| Amplified | | | |
| BLCAP | 2,544 | 20q11.23 | Hypoxia, vascular, invasion, mitosis, cyclin, notch |
| LAD1 | 1,209 | 1q32.1 | Vascular, hypoxia, notch, macrophages |
| CYC1 | 1,202 | 8q24.3 | Vascular, notch |
| DDX51 | 1,180 | 12q24.33 | Invasive, angiogenesis, collagen |
| ARHGAP30 | 1,085 | 1q23.3 | Invasive, collagen, tnf, MHC |
| SPAG5 | 823 | 17q11.2 | Mitosis, cyclin |
| TAF4 | 693 | 3p14.1 | Cyclin, p53 |
| Deleted | | | |
| ELAC2 | 1,960 | 17p11.2 | Vesicle |
| THY1 | 1,586 | 11q23.3 | Invasion, angiogenesis, collagen, mmp, integrin |
| LHFP | 1,309 | 13q12 | Mitosis, cyclin, mitosis, p53, tnf, MHC |
| CD164 | 1,016 | 6q21 | Collagen |
| POSTN | 815 | 13q13.3 | Invasive, angiogenesis, vascular, collagen |
| ELF1 | 802 | 13q13 | Hypoxia |
| FYN | 715 | 6q21 | Macrophage, tnf, MHC, collagen |
| LAMA4 | 607 | 6q21 | Invasion, vascular, collagen |

* Blue represents genes in regions with genomic losses and red represents genes in regions with genomic gains. ** LitVAN, Literature Vector Analysis.

doi: 10.1371/journal.pone.0078299.t002

drivers were also detected when all FBC tumors were analyzed together. As an example, *GATA3* was found as the top candidate driver among both Luminal A and Luminal B tumors. The landscapes of candidate drivers differed greatly between the intrinsic subgroups of FBC, and only *ARHGAP30* (Basal-like and HER2), *TAF4* (Basal-like), *SPAG5* (HER2 and Luminal A) and *COG3* (Luminal A) were shared with MBC.

CONEXIC identifies known drivers of FBC

In order to highlight known oncogenes and tumor suppressor genes in the candidate gene lists, these were compared with the cancer Gene Census list (COSMIC, Catalogue of Somatic Mutations In Cancer) and the Amplified and Overexpressed Genes In Cancer (AOGIC) list. COSMIC to date contains 487 genes and AOGIC contains 77 genes. Eight known cancer genes from these compiled gene lists were found among the FBC candidate drivers: *GATA3*, *CCNE1*, *PRCC*, *MALM2*, *CDK4*, *GRB7*, *NFIB* and *VTI1A*. Among these, *GATA3*, *CCNE1*, *CDK4* and *GRB7* are known drivers of breast cancer [26,27]. In contrast, only two known cancer genes, *ZNF217* and *LHFP*, were identified among the thirty MBC candidate driver genes. When CONEXIC was applied to MBCs using the extended 'MBC unique' gene set, an overall similar landscape

Table 3. Top 15 candidate driver genes in female breast cancer.

| Candidate drivers | CONEXIC | | |
|-------------------|---------|----------|--|
| FBC* | score | Cytoband | LitVAn** |
| Amplified | | | |
| GATA3 | 15,464 | 10p15 | Invasion, mcf, integrin |
| TIMP2 | 10,917 | 17q25 | Invasion, angiogenesis, metastasis, collagen |
| APOM | 9,563 | 6p21.33 | TNF, p53 |
| POLR2F | 9,346 | 22q13.1 | Lysine |
| NCAPG2 | 8,994 | 7q36.3 | Invasion, cyclin, checkpoint, notch, metastasis, collagen, p53 |
| CD4 | 8,299 | 12p13.31 | Invasion, MHC, notch |
| AIF1 | 7,723 | 6p21.3 | Invasion |
| KIFC1 | 6,012 | 6p21.3 | Mitochondrial |
| PRR7 | 5,767 | 5q35.3 | Cyclin, p53, |
| CSNK2B | 5,107 | 6p.21.3 | p53 |
| Deleted | | | |
| ARHGAP19 | 15,179 | 10q24.1 | Macrophage, MHC, cyclin, checkpoint, TNF |
| YIF1B | 10,200 | 19q13.2 | Invasion, collagen, notch, p53, vascular |
| DIAPH3 | 8,176 | 13q21.2 | Invasion, collagen, cyclin, checkpoint, p53 |
| TCF4 | 5,901 | 18q21.1 | Invasion, vascular, MHC, TNF |
| NISCH | 4,932 | 3p21.1 | Invasion, collagen, vascular |

* Blue represents genes in regions with genomic losses and red represents genes in regions with genomic gains. ** LitVAn, Literature Vector Analysis.

doi: 10.1371/journal.pone.0078299.t003

of candidate drivers for MBC was detected. In addition, *MAP2K4*, which was not included in the 'MBC/FBC common' gene set, was detected as a candidate driver in MBC (data not shown). *MAP2K4* has previously been reported to be over-expressed in pancreatic, breast and colorectal cancer, while *LHFP* expression has been reported in lipomas [26]. Interestingly, *GATA3* was not found to be a candidate driver for MBC in the present study, despite the fact that the majority of the MBC tumors were ER positive. *GATA3* was however identified as a candidate driver in the analyses of all FBCs, and within the Luminal A and Luminal B subtypes of FBC.

Biological processes and pathways regulated by the candidate drivers

Each potential driver is associated with gene modules (examples of two modules for MBC are shown in Figures 2-3) and the tumors are split into two groups for each candidate driver with a gene expression split value for that candidate driver. The two groups for each candidate driver display differential expression patterns of the genes in the modules, and these modules can help to explain the biological processes and pathways regulated by the candidate drivers. This information makes it possible to identify the most biologically and clinically interesting candidate drivers. LitVAn was used to

perform text mining on the 97 MBC modules. A number of over-represented terms were thus found for MBC, including 'cyclin', 'p53', 'tgf', 'invasion' and 'metastasis' (Tables 2-3 and Figures 4-5).

When gene set enrichments using gene ontology (GO) terms for the modules were studied to identify terms that were significantly enriched among MBCs, terms such as 'cell differentiation', 'cell cycle', 'division' and 'signal transduction' were found. Consequently, many of the candidate driver genes identified are known to be involved in processes associated with tumorigenesis, such as proliferation, motility, invasion, metastasis and cell differentiation.

THY1, a candidate prognostic invasion marker in MBC

We used the results from the LitVAn text mining and GO enrichment analysis for the selection of candidate driver genes of biological and clinical relevance for further validation. One of the most interesting modules was regulated at the first split by *THY1* (Thy-1 cell surface antigen, or *CD90*) and contained 41 genes; the text mining revealed over-represented terms including 'collagen', 'invasion', 'mmp', 'integrin', 'angiogenesis', and 'tgf' (Figures 2 and 4). The *THY1* module was enriched for GO terms including cell adhesion and collagen. Taken together, these findings suggest that *THY1* plays a role in invasion and metastasis in MBC. The *THY1* positive group in the present study was significantly associated with the previously described transcriptionally derived Luminal M1 group, and the *THY1* negative group with the Luminal M2 group (Fisher's exact test, $P=0.0027$). Moreover, the Luminal M1 group had a significantly higher module score for the invasion and metastasis module represented by *PLAU*, further supporting that *THY1* is involved in invasion and metastasis in MBC [2]. *THY1* maps to chromosome 11q22.3 and is a glycoposphatidyl-inositol (GPI)-anchored cell surface glycoprotein with a number of biological functions involving cell-cell and cell-matrix interactions and cell motility [28]. Furthermore, *THY1* is included as an up-regulated gene in the EMT core signature defined by Taube et al., and the module that it regulates contains seven of these up-regulated EMT genes. The genes up-regulated in the EMT core signature were also significantly up-regulated in the *THY1* positive MBCs, while the down-regulated genes were not differentially expressed across the MBC tumors (Figure S1) [29].

The LitVAn text mining also revealed 'tgf' as an over-represented term in MBC and TGF β signaling is known to play an important role in EMT [30], further supporting that the *THY1* module is involved in EMT. Fredlund et al. identified an EMT-induced stroma module correlating with aggressive disease in the Basal-like subtype of FBC [31]. The *THY1* module identified among the MBCs in the present study contained 50% of the genes in this stroma module and the *THY1* positive MBC tumors displayed significantly higher activity for the stroma module (Figure S2). Furthermore, the *THY1* negative tumors displayed low activity for the steroid response module, which is also in agreement with our earlier results, where the Luminal M1 tumors displayed low activity for the ER module [2].

EMT has been shown to be closely associated with the claudin-low subtype of breast cancer in earlier studies [29,32].

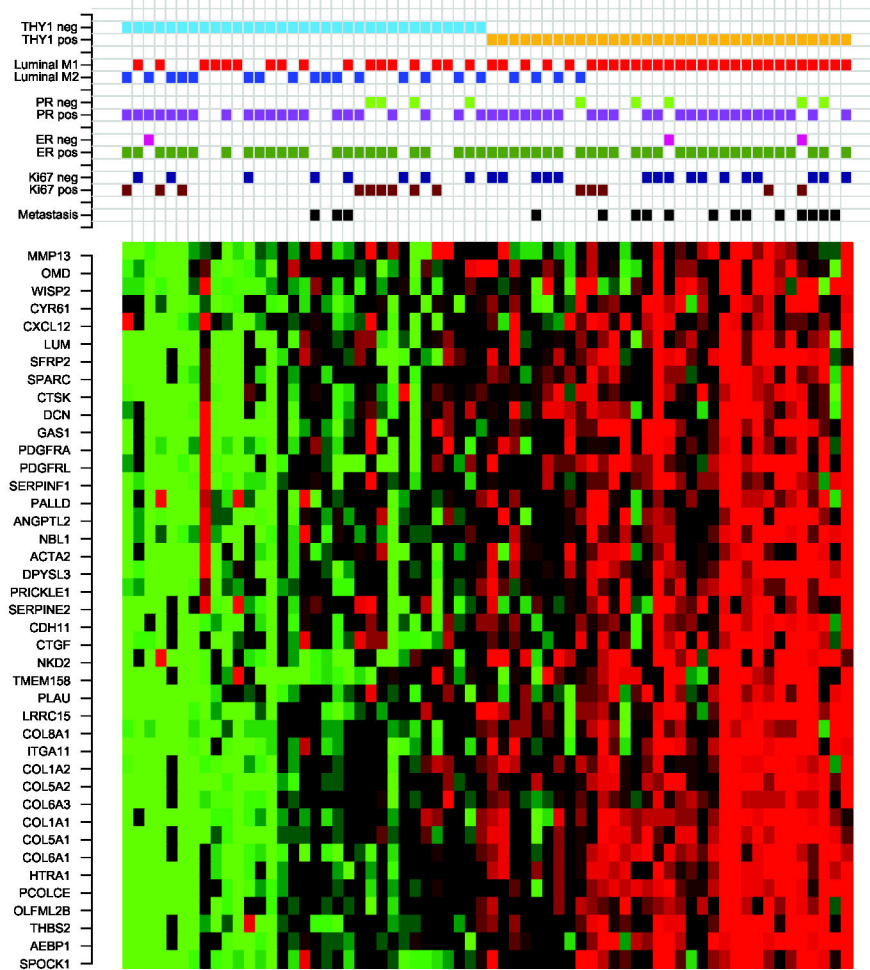


Figure 2. Heatmap of the genes in the module regulated by *THY1* in MBC. Red corresponds to up-regulation and green to down-regulation. The 66 MBC tumors are sorted according to their gene expression level of *THY1*.

doi: 10.1371/journal.pone.0078299.g002

A claudin-low signature was therefore used to cluster the MBC tumors in the present study, but no correlation with the claudin-low subtype was found (data not shown) [33]. Next, the 66 MBC tumors included in our original transcriptional profiling study were split into two groups using the gene expression split value for *THY1*, resulting in two equally sized groups containing 33 patients each [2]. The 13 MBC tumors not included in the CONEXIC analysis in the present study displayed the same pattern as the 53 tumors included in the CONEXIC analysis (Figure 2). Importantly, *THY1* was found to be significantly associated with survival among MBC patients, with high expression of *THY1* predicting a significantly shorter time to DMFS ($P=0.0087$; Figure 4B).

In order to validate the association between *THY1* expression and the genes in its module, an independent dataset with 37 MBC tumors was used [4]. The MBC tumors in the validation cohort were sorted according to increasing expression of *THY1*, whereupon a comparison of the heatmaps for the two datasets displayed highly similar patterns (Figure

S3), adding support for *THY1* as a potential candidate driver in MBC. Moreover, the genes in the *THY1* module were significantly correlated to *THY1* compared to all other genes in both datasets (Wilcoxon test, $P=2.2e-16$ and $P=6.0e-11$, respectively).

SPAG5, a potential driver of cell proliferation in MBC

Another biologically and clinically interesting module was regulated at the first split by *SPAG5* (sperm-associated antigen 5) and contained 48 genes (Figure 3). LitVAn found 'cyclin', 'mitot', 'mitosis', 'spindl', 'checkpoint', 'g2', 'nocodazol', 'g1', 'p53' and 'cdk2' to be overrepresented terms (Figure 5A). The *SPAG5* module had enrichment of GO terms including cell cycle, cell division, M phase, mitosis, spindle and regulation of cell cycle. These results indicate that the *SPAG5* module is involved in the mitotic checkpoint and progression of MBC, and that *SPAG5* is a potential driver of proliferation in MBC. *SPAG5* maps to chromosome 17q11.2 and encodes a protein, also called astrin, associated with the mitotic spindle machinery

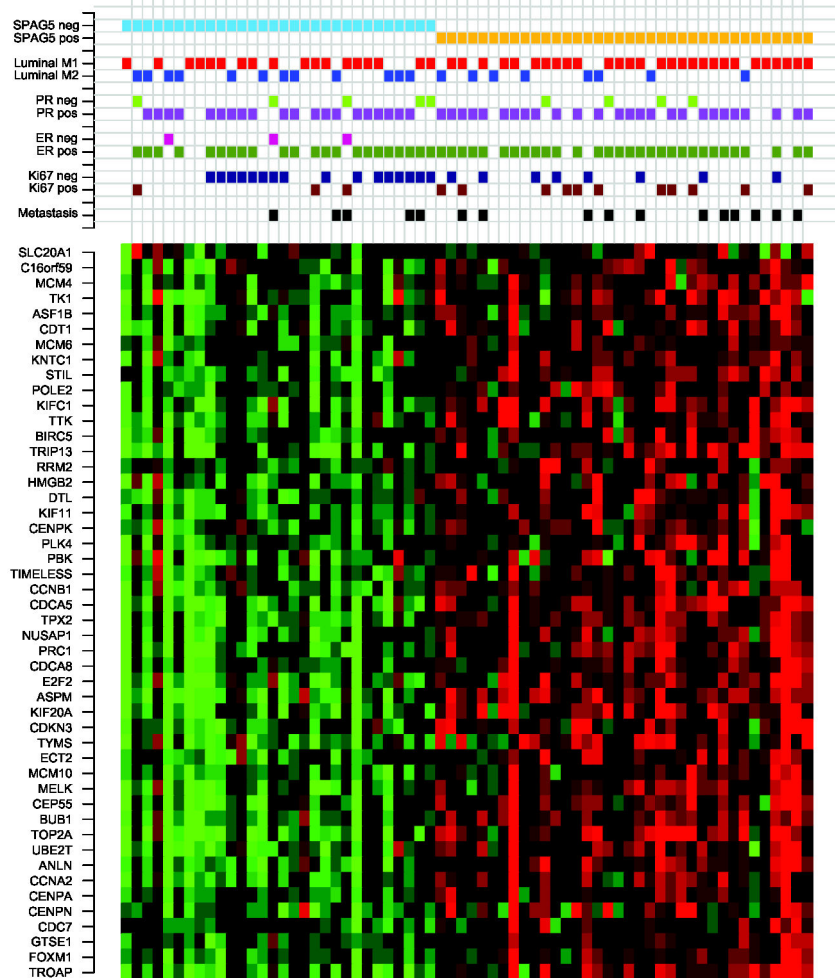


Figure 3. Heatmap of the genes in the module regulated by *SPAG5* in MBC. Red corresponds to up-regulation and green to down-regulation. The 66 MBC tumors are sorted according to their gene expression level of *SPAG5*.

doi: 10.1371/journal.pone.0078299.g003

[34]. Fredlund et al. identified two modules involved in proliferation in breast cancer, the mitotic checkpoint module and the mitotic progression module [31], and the *SPAG5* module identified herein among MBCs contained 70% of the mitotic checkpoint genes and 100% of the mitotic progression genes. The highly expressing *SPAG5* tumors displayed a significantly higher activity for the mitotic checkpoint and mitotic progression modules, while there were no differences for the other modules (Figure S4), further supporting that this module is involved in the proliferation of MBC. *SPAG5* has been shown to be prognostic in ER positive, but not in ER negative breast cancer in women, while it has been shown to be predictive for chemotherapy response in the latter group [35]. Analysis of *SPAG5* in 1,379 FBCs in GOBO (Gene expression-based Outcome for Breast cancer Online) [36] confirmed that it was prognostic for distant disease free survival in ER positive, but not ER negative FBC (data not shown).

Next, the 66 MBC tumors included in our original transcriptional profiling study were split into two groups using

the gene expression split value for *SPAG5*, resulting in one group containing 30 tumors with low expression of *SPAG5* and one group containing 36 tumors with high expression of *SPAG5*. The 13 MBC tumors not included in the CONEXIC analysis in the present study showed the same pattern as the 53 tumors included in the CONEXIC analysis. The *SPAG5* groups were however not associated with the transcriptional Luminal M1 and M2 groups (Fisher's exact test, $P=0.18$) [2]. Tumors in the *SPAG5* high group had significantly increased levels of Ki67 compared to tumors in the *SPAG5* low group (Wilcoxon test, $P<0.05$). However, when comparing survival between the groups, no significant difference could be seen (Figure 5B), which is in line with our earlier study where no prognostic value of Ki67 was seen in MBC [37].

Discussion

The development of high throughput and massively parallel techniques has moved forward rapidly and today many

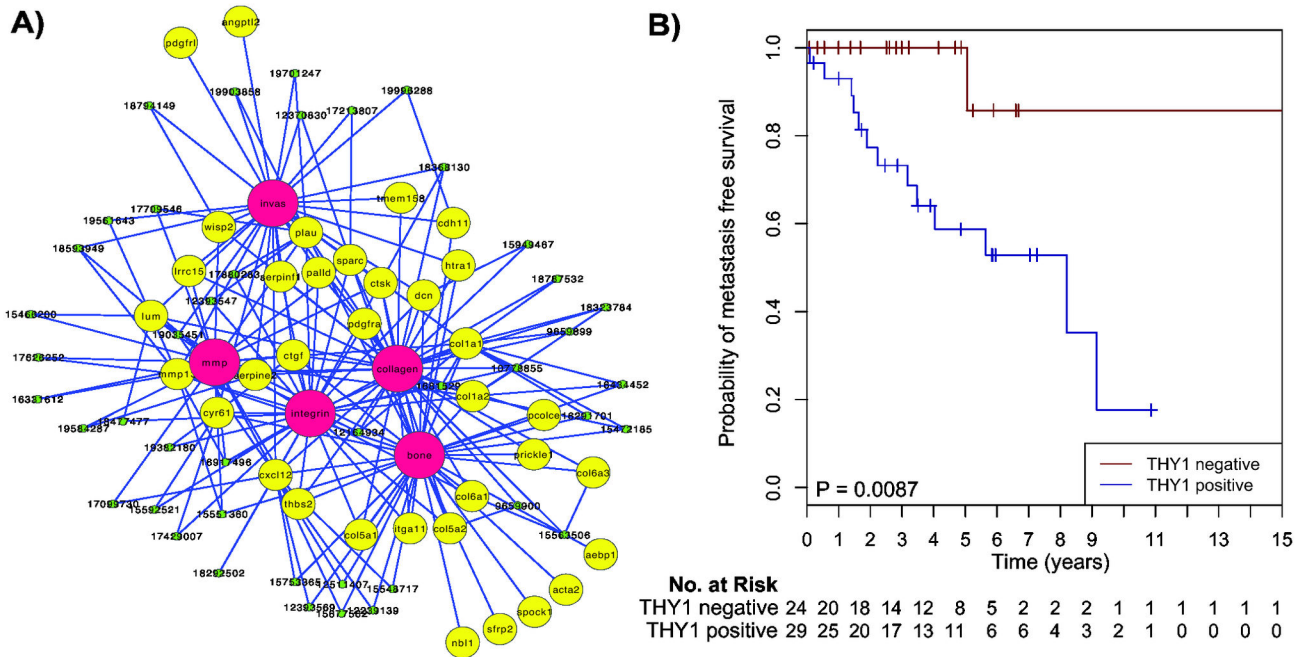


Figure 4. Graphic output of the LitVAN result for the MBC *THY1* module and corresponding Kaplan-Meier survival analysis. **A)** Significantly over-represented terms are represented as red circles and their association (graph edges) with multiple genes (yellow circles) from the module. The green dots represent the publication that significantly associates between the gene and the term, and the numbers in the green dots are the PubMed IDs for the respective publications. **B)** Distant metastasis free survival of the 66 MBC patients stratified by *THY1* gene expression. The numbers below the plots indicate the number of patients at risk in each group at the given time points.

doi: 10.1371/journal.pone.0078299.g004

different methods for generating various different types of genome-scale data are available. As these techniques improve in resolution and scope, the complexity and heterogeneity of tumors is becoming increasingly evident. Cancer cells are equipped with multiple regulatory networks, enabling them to rapidly adapt to new contexts and to quickly grow and spread. This makes the task of eradicating tumor cells difficult and complex, and puts high demands on the development of therapy regimes. Mutations and chromosomal aberrations are central characteristics of tumors, and epithelial cancers harbor large numbers of such aberrations. It is therefore important to be able to distinguish between genetic changes that drive tumor progression and those that may be considered passenger events. An encouraging approach is to integrate different types of genome-scale data, which each shed light on different aspects of the underlying biology, to better understand which genes drive tumorigenesis, and the networks that they regulate. This may help us to better understand the complexity of tumors and move us closer to truly personalized cancer treatment.

The strategy employed by CONEXIC of identifying networks that the candidate drivers regulate has multiple advantages. Firstly, it gives a better understanding of the physiological context of the driver. Not all genes are targetable, thus targeting individual drivers may be impossible, but targeting suitable components of the networks might be a more

amenable approach. A number of methods combining aCGH and GEX data for the detection of driver genes have been published to date [38-40], and they all define an important driver gene by simultaneous copy number aberration and differential expression. This assumption may however not always be valid for defining a cancer driver, as copy number alteration is only one of many means of altering the expression level of a gene. CONEXIC does not make this assumption [13], and may hence identify important driver genes that the other methods do not detect. Nonetheless, CONEXIC does not identify all potential cancer drivers, since it only detects drivers among amplified or deleted regions that pass the stringent statistical tests and will thus miss drivers arising from for example point mutations [13]. Recent attempts to identify mutations in driver genes in FBC have shown marked heterogeneity between individual tumors, as only 3-7 genes have been found to be mutated in more than 10% of FBCs [41,42]. The task of identifying candidate drivers in breast cancer is hence not easy. Thus, to identify all driver genes in MBC a much larger patient cohort would be required, as well as information regarding e.g. mutations and epigenetic changes. Nevertheless, our cohort constitutes the most comprehensive and well-annotated collection of MBC tumors, and we believe that the CONEXIC analysis has revealed several interesting and novel potential candidate drivers of importance for breast cancer progression and aggressiveness

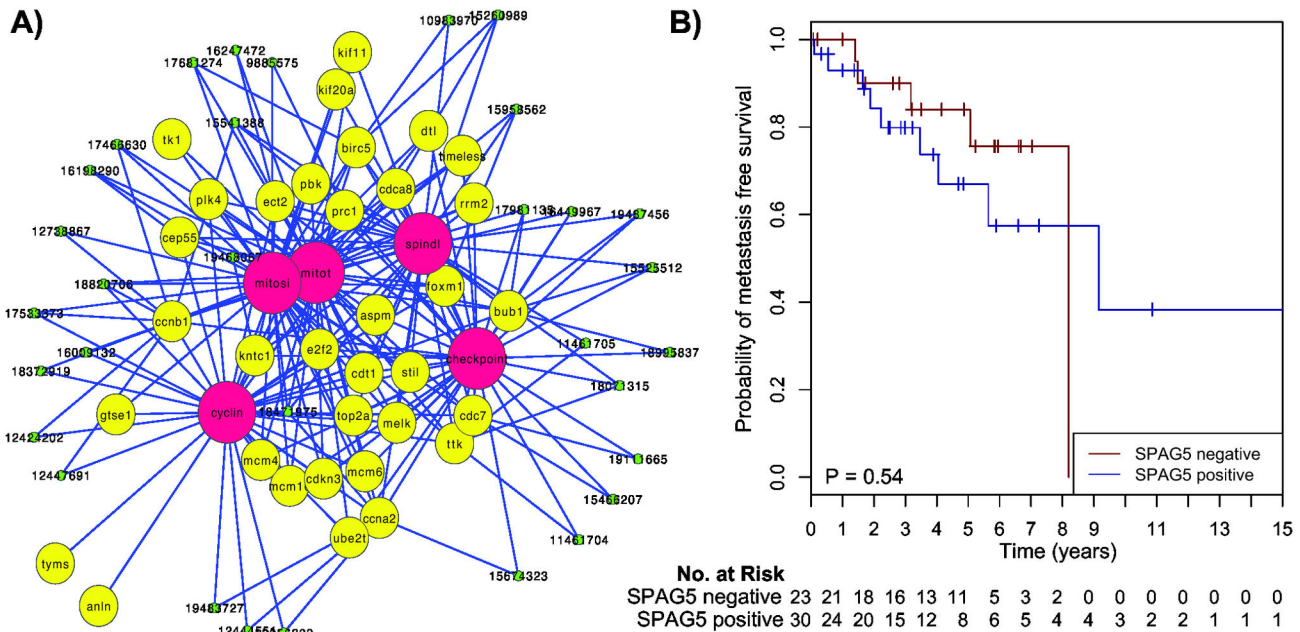


Figure 5. Graphic output of the LitVAN result for the MBC SPAG5 module and corresponding Kaplan-Meier survival analysis. **A)** Significantly over-represented terms are represented as red circles and their association (graph edges) with multiple genes (yellow circles) from the module. The green dots represent the publication that significantly associates between the gene and the term, and the numbers in the green dots are the PubMed IDs for the respective publications. **B)** Distant metastasis free survival of the 66 MBC patients stratified by SPAG5 gene expression. The numbers below the plots indicate the number of patients at risk in each group at the given time points.

doi: 10.1371/journal.pone.0078299.g005

in men. Notwithstanding, we anticipate the outcome of the ongoing international effort aimed at further enhancing our knowledge of the underpinnings of this malignancy in a much larger cohort [8]. Furthermore, deep sequencing of large numbers of MBC tumors will be required to establish the involvement of driver genes implicated in MBC tumorigenesis by mutational events.

In this study, CONEXIC was applied to 53 MBC tumors and 359 FBC tumors. CONEXIC identified 30 candidate drivers in the MBC tumors and 67 in the FBC tumors, of which only two candidate driver genes were in common (*TAF4* and *CD164*). Three additional candidate driver genes (*ARHGAP30*, *COG3* and *SPAG5*) in common with MBC were found when CONEXIC was applied separately to the intrinsic subtypes of FBC. This illustrates that male and female breast tumors are very different and is in line with previous results showing differences between MBC and FBC on the genomic and transcriptional levels [1-4]. We recently reported two new subgroups of MBC (Luminal M1 and M2, respectively) that differed from the intrinsic subtypes of FBC and as such appeared to occur only among males [2], further supporting these differences.

The MBC candidate drivers included two known cancer genes, *LHFP* and *ZNF217*. In addition, when CONEXIC was applied to the 'MBC unique' gene set, *MAP2K4* was detected as a candidate driver. *MAP2K4* was not included in the 'MBC/FBC common' dataset. *MAP2K4* is involved in the JNK pathway and loss of this gene results in defective apoptosis in

response to stress [43]. Mutations in *MAP2K4*, which is a substrate for *MAP3K1*, have been seen as a rare event in Luminal FBCs, and pathway analysis showed that it produces similar pathway perturbations as *MAP3K1* mutations [44]. Furthermore, *MAP2K4* deletions were detected predominantly in ER positive tumors in a comprehensive study of 2,000 FBC tumors [45], suggesting that MBC tumors share some features with ER positive FBC.

Many known drivers of breast cancer, including *GATA3*, *CCNE1*, *CDK4* and *GRB7*, as well as known drivers for other tumor types [26,27] were identified among the FBC candidate drivers. Ellis et al. also found *GATA3* to be a driver of FBC and they suggested that *GATA3* mutations might be a positive predictive marker for response to aromatase inhibitors [44]. In another study, Dutta et al. found *GATA3* to be a driver specifically for ER positive FBC. We identified *GATA3* as the top candidate driver among all FBC tumors and also for Luminal A and Luminal B FBC tumors. *GATA3* was however not identified among the candidate drivers in the MBCs in the present study, despite the majority of these tumors being ER positive. Taken together, these findings suggest that MBC tumors appear to share features with both ER positive and ER negative FBC tumors.

In all, the LitVAN study and GO enrichment analysis of the gene modules showed that many of the candidate drivers identified in male as well as female breast cancer are known to be involved in processes associated with tumorigenesis.

THY1 was one of the most biologically and clinically interesting candidate drivers for MBC, and it seems to be a driver of invasion related to EMT. Interestingly, the *THY1* groups were highly correlated with the previously described Luminal M1 and M2 groups, despite being identified independently of these groups. This finding thus further validates that the transcriptional MBC groups represent two stable subgroups of MBC [2]. There was a significant difference in survival between the *THY1* groups, where high expression of *THY1* correlated to poor prognosis. Interestingly, this phenotype also corresponds to the claudin-low subtype of triple-negative breast cancers, but MBC tumors are not classified as claudin-low. Furthermore, *THY1* expression levels were significantly correlated to the target genes in the module compared to all other genes, supporting a role for *THY1* in MBC. The module regulated by *THY1* was highly correlated to the stroma module described by Fredlund et al. in FBC, and high stroma module activity correlated with poor outcome in Basal-like FBC tumors [31]. The high expressing *THY1* group of MBCs also displayed a significantly decreased activity for the steroid response module, a feature common to ER negative breast cancers, further indicating that these tumors behave more like ER negative than ER positive FBC, despite the majority of the MBCs being ER positive. This is in line with our earlier study where the Luminal M1 group, which correlates with the high expressing *THY1* group, showed decreased activity for the ER module. However, the Luminal M2 group, while most resembling Luminal A FBCs, was also found to differ from the conventional intrinsic subgroups of FBC. Here, we show that the *THY1* negative group, containing the majority of the Luminal M2 tumors, had low activity for the stroma module, which is one of the highest scoring modules for Luminal A tumors [2]. This further confirms that the heterogeneity observed among MBCs on the transcriptional and genomic levels is not readily captured with the intrinsic subtypes of FBC.

Of interest, *THY1* has been shown to have contradictory functions in different tumor types, where it functions as a tumor suppressor in ovarian [46] and nasopharyngeal cancer [47], yet it promotes migration and metastasis in melanoma [48] and hepatocarcinomas [49].

Another potential driver of biological and clinical interest for MBC identified in the present study was *SPAG5*, the expression of which likely explains the differentially expressed genes involved in mitotic progression and mitotic checkpoint control, and illustrating that proliferation is an important feature of MBC. Proliferation is one of the most important determinants separating Luminal A from Luminal B tumors among ER positive FBC tumors, and patients with highly proliferative Luminal B tumors have a significantly inferior outcome compared to patients with Luminal A tumors [50]. Despite the majority of the MBC tumors being ER positive, they did not behave like ER positive FBC tumors, as no difference in survival between the two *SPAG5* groups was detected although the *SPAG5* module was linked to proliferation. Most probably the *SPAG5* negative tumors harbor other features than high proliferative that render them aggressive. A recent study showed that *SPAG5* was prognostic in ER positive, but

not in ER negative FBC, while it was predictive for response to chemotherapy in ER negative FBC [35]. Unfortunately, we were not able to stratify the *SPAG5* groups on chemotherapy, as only 5 patients received chemotherapy. *SPAG5* may be predictive for chemotherapy response also in MBC, in which case more MBC patients may benefit from chemotherapy.

In the context of MBC this is a large collection of fresh frozen MBC tumors, but we need to keep in mind that the study is based on a small number of patients and further studies are needed to validate these findings. An international consortium for studying MBC has recently been formed, and clinical information as well as tumor material from a large number of MBC patients around the world is currently being collected with the aim to enhance our understanding of MBC [8].

In summary, the combination of copy number and gene expression data revealed vast differences in the landscapes of candidate drivers between male and female breast cancers. Furthermore, these results suggest that although the vast majority of MBCs are ER positive they share many features with ER negative FBC. Consequently, the pathobiology of MBC may be very distinct from that of FBC, and men diagnosed with breast cancer may therefore require different management and treatment strategies than women. Furthermore, *THY1* is a promising prognostic invasion marker for MBC, and the chemotherapy treatment predictive value of *SPAG5* in MBC warrants further investigation.

Supporting Information

Figure S1. The mean expression values of genes up-regulated or down-regulated by the EMT core signature in the *THY1* groups. The up-regulated genes from the EMT core signature were significantly up-regulated for the *THY1* positive tumors, while the down-regulated genes were down-regulated for all MBC tumors. (EPS)

Figure S2. The module activity of the eight gene modules reflecting biologically relevant transcriptional programs found in FBC in the *THY1* groups. *THY1* positive tumors had significantly higher module activity for the stroma module and significantly lower activity for the steroid response module. (EPS)

Figure S3. Validation of the *THY1* module in an external MBC dataset. The heatmaps of the *THY1* module genes revealed similar transcriptional profiles in our dataset (A) and the external validation dataset (B). Red corresponds to up-regulation and green to down-regulation. The genes are sorted in the same order and the MBC tumors are sorted according to increasing expression of *THY1*. (EPS)

Figure S4. The module activity of the eight gene modules reflecting biologically relevant transcriptional programs found in FBC in the *SPAG5* groups. *SPAG5* positive tumors had significantly higher module activity for the mitotic checkpoint and mitotic progression modules.

(EPS)

Table S1. All MBC candidate drivers.

(DOCX)

Table S2. All FBC candidate drivers.**References**

- Johansson I, Nilsson C, Berglund P, Strand C, Jönsson G et al. (2011) High-resolution genomic profiling of male breast cancer reveals differences hidden behind the similarities with female breast cancer. *Breast Cancer Res Treat* 129: 747–760. doi:10.1007/s10549-010-1262-8. PubMed: 21113657.
- Johansson I, Nilsson C, Berglund P, Lauss M, Ringnér M et al. (2012) Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. *Breast Cancer Res* 14: R31. doi:10.1186/bcr3116. PubMed: 22333393.
- Tommasi S, Mangia A, Iannelli G, Chiarappa P, Rossi E et al. (2010) Gene copy number variation in male breast cancer by aCGH. *Anal Cell Pathol (Amst)* 33: 113–119. doi:10.1007/s13402-011-0041-9. PubMed: 2154757721045282.
- Callari M, Cappelletti V, Cecco L, Musella V, Miodini P et al. (2010) Gene expression analysis reveals a different transcriptomic landscape in female and male breast cancer. *Breast Cancer Res Treat* 127: 601–610. doi:10.1007/s10549-010-1015-8. PubMed: 20625818.
- Giordano SH (2005) A Review of the Diagnosis and Management of Male Breast Cancer. *Oncologist* 10: 471–479. doi:10.1634/theoncologist.10-7-471. PubMed: 16079314.
- Agrawal A, Ayantunde AA, Rampaul R, Robertson JFR (2006) Male breast cancer: a review of clinical management. *Breast Cancer Res Treat* 103: 11–21. doi:10.1007/s10549-006-9356-z. PubMed: 17033919.
- Nilsson C, Holmqvist M, Bergkvist L, Hedenfalk I, Lambe M et al. (2011) Similarities and differences in the characteristics and primary treatment of breast cancer in men and women – a population based study (Sweden). *Acta Oncol* 50: 1083–1088. doi:10.3109/0284186X.2011.602114. PubMed: 21830994.
- Korde LA, Zujewski JA, Kamin L, Giordano S, Domchek S et al. (2010) Multidisciplinary Meeting on Male Breast Cancer: Summary and Research Recommendations. *J Clin Oncol* 28: 2114–2122. doi:10.1200/JCO.2009.25.5729. PubMed: 20308661.
- Lehmann U, Streichert T, Otto B, Albat C, Hasemeier B et al. (2010) Identification of differentially expressed microRNAs in human male breast cancer. *BMC Cancer* 10: 109. doi:10.1186/1471-2407-10-109. PubMed: 20331864.
- Fassan M, Baffa R, Palazzo JP, Lloyd J, Crosariol M et al. (2009) MicroRNA expression profiling of male breast cancer. *Breast Cancer Res* 11: R58. doi:10.1186/bcr2348. PubMed: 19664288.
- Orr N, Lemnrau A, Cooke R, Fletcher O, Tomczyk K et al. (2012) Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. *Nat Genet* 44: 1182–1184. doi:10.1038/ng.2417. PubMed: 23001122.
- Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. *Nat Rev Genet* 11: 476–486. doi:10.1038/nrg2795. PubMed: 20531367.
- Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D et al. (2010) An Integrated Approach to Uncover Drivers of Cancer. *Cell* 143: 1005–1017. doi:10.1016/j.cell.2010.11.013. PubMed: 21129771.
- Dutta B, Pusztai L, Qi Y, André F, Lazar V et al. (2012) A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. *Br J Cancer* 106: 1107–1116. doi:10.1038/bjc.2011.584. PubMed: 22343619.
- Solvang HK, Lingjærde OC, Frigessi A, Børresen-Dale A-L, Kristensen VN (2011) Linear and non-linear dependencies between copy number aberrations and mRNA expression reveal distinct molecular pathways in breast cancer. *BMC Bioinformatics* 12: 197. doi:10.1186/1471-2105-12-197. PubMed: 21609452.
- Chuang H-Y, Lee E, Liu Y-T, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. *Mol Syst Biol* 3: 140. doi:10.1038/msb4100180. PubMed: 17940530.
- Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C et al. (2009) Dynamic modularity in protein interaction networks predicts breast

(DOCX)

Author Contributions

Conceived and designed the experiments: IJ IH. Performed the experiments: IJ. Analyzed the data: IJ MR IH. Wrote the manuscript: IJ MR IH.

- Nat Biotechnol 27: 199–204. doi:10.1038/nbt.1522. PubMed: 19182785.
- Nilsson C, Johansson I, Ahlin C, Thorstenson S, Amini R-M et al. (2013) Molecular subtyping of male breast cancer using alternative definitions and its prognostic impact. *Acta Oncol* 52: 102–109. doi:10.3109/0284186X.2012.711952. PubMed: 22928693.
- Jönsson G, Staaf J, Vallon-Christersson J, Ringnér M, Holm K 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. doi:10.1186/bcr2596. PubMed: 20576095.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34: 166–176. doi:10.1038/ng1165. PubMed: 12740579.
- Sanchez-Garcia F, Akavia UD, Mozes E, Pe'er D (2010) JISTIC: identification of significant targets in cancer. *BMC Bioinformatics* 11: 189. doi:10.1186/1471-2105-11-189. PubMed: 20398270.
- Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA* 104: 20007–20012. doi:10.1073/pnas.0710052104. PubMed: 18077431.
- Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. *Ann Statist* 7: 1–26. doi:10.1214/aos/1176344552.
- Fredlund E, Ringnér M, Maris JM, Pählman S (2008) High Myc pathway activity and low stage of neuronal differentiation associate with poor outcome in neuroblastoma. *Proc Natl Acad Sci USA* 105: 14094–14099. doi:10.1073/pnas.0804455105. PubMed: 18780787.
- R Development Core Team (2013) R: A Language and Environment for Statistical Computing. 1 pp. Available: <http://www.r-project.org/>. Accessed 2013 October 5.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177–183. doi:10.1038/nrc1299. PubMed: 14993899.
- Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10: 59–64. doi:10.1038/nrc2771. PubMed: 20029424.
- Rege TA, Hagood JS (2006) Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. *FASEB J* 20: 1045–1054. doi:10.1096/fj.05-5460rev. PubMed: 16770003.
- Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S et al. (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci USA* 107: 15449–15454. doi:10.1073/pnas.1004900107. PubMed: 20713713.
- Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2: 442–454. doi:10.1038/nrc822. PubMed: 12189386.
- Fredlund E, Staaf J, Rantala JK, Kallioniemi O-P, Borg A et al. (2012) The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res* 14: R113. doi:10.1186/bcr3236. PubMed: 22839103.
- Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM et al. (2009) Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci USA* 106: 13820–13825. doi:10.1073/pnas.0905718106. PubMed: 19666588.
- Prat A, Parker JS, Karginova O, Fan C, Livasy C et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12: R68. doi:10.1186/bcr2635. PubMed: 20813035.
- Kersten FF, van Wijk E, Hettenschijt L, Bauß K, Peters TA et al. (2012) The mitotic spindle protein SPAG5/Astrin connects to the Usher protein

- network postmitotically. *Cilia* 1: 2. doi:10.1186/2046-2530-1-2. PubMed: 23351521.
35. Abdel-Fatah T, Ball G, Miles AK, Moseley P, Green A et al. (2012) Identification of Sperm Associated Antigen 5 (SPAG5) as a novel biological and predictive biomarker in Breast cancer. *Cancer Res* 72: Supplement3.
 36. Ringnér M, Fredlund E, Häkkinen J, Borg A, Staaf J (2011) GOBO: gene expression-based outcome for breast cancer online. *PLOS ONE* 6: e17911. doi:10.1371/journal.pone.0017911. PubMed: 21445301.
 37. Nilsson C, Koliadi A, Johansson I, Ahlin C, Thorstenson S et al. (2013) High proliferation is associated with inferior outcome in male breast cancer patients. *Mod Pathol* 26: 87–94. doi:10.1038/modpathol.2012.145. PubMed: 22918167.
 38. Lin WM, Baker AC, Beroukhir R, Winckler W, Feng W et al. (2008) Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res* 68: 664–673. doi: 10.1158/0008-5472.CAN-07-2615. PubMed: 18245465.
 39. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R et al. (2010) Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene* 29: 2013–2023. doi:10.1038/onc.2009.489. PubMed: 20101236.
 40. Salari K, Tibshirani R, Pollack JR (2010) DR-Integrator: a new analytic tool for integrating DNA copy number and gene expression data. *Bioinformatics* 26: 414–416. doi:10.1093/bioinformatics/btp702. PubMed: 20031972.
 41. TOBCC OSBREAC (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486: 400–404. doi: 10.1038/nature11017. PubMed: 22722201.
 42. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J et al. (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490: 61–70. doi:10.1038/nature11412. PubMed: 23000897.
 43. Wagner EF, Nebreda AR (2009) Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer* 9: 537–549. doi:10.1038/nrc2694. PubMed: 19629069.
 44. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ et al. (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486: 353–360. doi:10.1038/nature11143. PubMed: 22722193.
 45. METABRIC G (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486: 346–352. doi:10.1038/nature10983. PubMed: 22522925.
 46. Abeysinghe HR, Pollock SJ, Guckert NL, Veyberman Y, Keng P et al. (2004) The role of the THY1 gene in human ovarian cancer suppression based on transfection studies. *Cancer Genet Cytogenet* 149: 1–10. doi:10.1016/S0165-4608(03)00234-6. PubMed: 15104276.
 47. Lung HL, Cheung AKL, Cheng Y, Kwong FM, Lo PHY et al. (2010) Functional characterization of THY1 as a tumor suppressor gene with antiinvasive activity in nasopharyngeal carcinoma. *Int J Cancer* 127: 304–312. doi:10.1002/ijc.25047. PubMed: 19921696.
 48. Saalbach A, Wetzel A, Hausteil U-F, Sticherling M, Simon JC et al. (2005) Interaction of human Thy-1 (CD 90) with the integrin $\alpha\beta 3$ (CD51/CD61): an important mechanism mediating melanoma cell adhesion to activated endothelium. *Oncogene* 24: 4710–4720. doi: 10.1038/sj.onc.1208559. PubMed: 15897908.
 49. Cheng B-Q, Jiang Y, Li D-L, Fan J-J, Ma M (2012) Up-regulation of thy-1 promotes invasion and metastasis of hepatocarcinomas. *Asian Pac J Cancer Prev* 13: 1349–1353. doi:10.7314/APJCP.2012.13.4.1349. PubMed: 22799330.
 50. Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869–10874. doi:10.1073/pnas.191367098. PubMed: 11553815.