



Published in final edited form as:

Cytogenet Genome Res. 2007 ; 118(2-4): 116–129. doi:10.1159/000108292.

Genomic signatures of breast cancer metastasis

V. Urquidi^a and S. Goodison^b

^a *Department of Medicine, University of Florida, Jacksonville, FL (USA)*

^b *Department of Surgery, University of Florida, Jacksonville, FL (USA)*

Abstract

Despite significant advances in the treatment of primary cancer, the ability to predict the metastatic behavior of a patient's cancer, as well as to detect and eradicate such recurrences, remain major clinical challenges in oncology. While many potential molecular biomarkers have been identified and tested previously, none have greatly improved the accuracy of specimen evaluation over routine histopathological criteria and they predict individual outcomes poorly. However, the recent introduction of high-throughput microarray technology has opened new avenues in genomic investigation of cancer, and through application in tissue-based studies and appropriate animal models, has facilitated the identification of gene expression signatures that are associated with the lethal progression of breast cancer. The use of these approaches has the potential to greatly impact our knowledge of tumor biology, to provide efficient biomarkers, and enable development towards customized prognostication and therapies for the individual.

Breast cancer is one of the most common causes of cancer-related deaths worldwide, largely due to the recurrence of therapeutically resistant disseminated disease. Despite significant advances in the treatment of primary cancer, the ability to predict the metastatic behavior of a patient's cancer, as well as to detect and eradicate such recurrences, remains the greatest clinical challenge in oncology. Currently, the majority of lymph node (LN)-negative breast cancer patients undergo systemic adjuvant therapy because of the inability to accurately determine an individual's risk of recurrence. Since only a minority of node-negative patients will actually develop disease recurrence, there is a critical need to stratify patients with respect to risk of breast cancer recurrence. While many potential molecular biomarkers have been identified and tested previously, none have greatly improved the stratification of disease and they predict individual outcomes poorly. Consequently, few prognostic molecular markers are currently in clinical use as standard prognostic or predictive factors.

Recent advancements in powerful new high-performance screening technologies have revolutionized the ways in which researchers can study the pathogenesis of disease. This offers great potential to improve our knowledge of tumor molecular biology, and to provide pivotal information for clinical evaluation of breast cancer progression. Micro-arrays have been employed to identify gross chromosomal changes in breast cancer (Albertson, 2003) and gene expression patterns that distinguish molecular subtypes of breast cancer, and that are predictive of metastatic relapse (Sorlie et al., 2001, 2003; van de Vijver et al., 2002; Weigelt et al., 2003, 2005a). This approach has also been used in animal models of metastasis and signatures of gene expression are beginning to take shape (Kang et al., 2003; Eckhardt et al., 2005; Goodison et al., 2005b; Kluger et al., 2005; Minn et al., 2005). The combination of these approaches could lead to the identification of genes playing a critical role in cancer

development and progression, which, in turn, will provide more efficient biomarkers, and ideally result in the avoidance of over- or under-treatment of individual patients through development towards customized prognostication and therapies. In this review, we focus on the rapidly evolving signatures of tumor progression, particularly those that correlate with the occurrence of distant metastases, and discuss some of the expected future developments in the field.

Gross chromosomal aberrations

The development of cancer is generally believed to require the accumulation of multiple genetic aberrations. These aberrations range from single nucleotide mutations to cytogenetically detectable numerical and structural chromosomal alterations (Devilee and Cornelisse, 1994; Bieche and Lidereau, 1995). While human tissue specimens and immortal breast cancer cell lines represent an excellent resource for genomic studies, chromosomal alterations have not been well characterized until relatively recently, largely because of limitations inherent to conventional chromosome banding techniques. The development of new techniques which enable the accurate, high throughput analysis of specimens has enabled substantial progress in this field. Comparative genomic hybridization (CGH), introduced in 1992, has been the technique of choice for mapping DNA copy number changes in human tumors. The technique uses hybridization to compare abundance of specific genome sequences in tumor cell DNA relative to normal reference genomes (Kallioniemi et al., 1994). This opened new avenues in genomic investigation because it obviated the need to culture cells before their chromosomes could be analyzed. The latest generation of CGH analysis, array-CGH or matrix-CGH, uses ordered arrays of genomic DNA sequences and further increases the potential of CGH to provide insight into chromosomal aberrations present in cancer by enabling single-gene resolution. Recently developed multi-color chromosome imaging techniques such as spectral karyotyping (SKY) (Shrock et al., 1996) and multiplex-FISH (M-FISH) (Speicher et al., 1996) utilize simultaneous visualization of each human chromosome with specific fluorochrome combinations and enable a far more rapid and detailed karyotypic analysis of solid tumors and cell lines (Macville et al., 1997; Kawai et al., 2002). The chromosomal aberrations in the most commonly used human breast carcinoma cell lines have recently been investigated using these techniques (Kytola et al., 2000; Xie et al., 2002; Watson et al., 2004; Goodison et al., 2005a; Shadoo and Lam, 2006). Karyotypic analysis has shown breast tumor cell lines to be either near diploid with simple rearrangements or highly aneuploid with multiple, complex rearrangements (Morris et al., 1997; Kytola et al., 2000).

The most frequent gains detected by CGH in breast tumor cell lines are 1q, 8q, 20q, 7, 11q13, 17q, 9q, and 16p, whereas losses were most common at 8p, 11q14→qter, 18q, and Xq (Kytola et al., 2000). The comparison of CGH data from cell lines to CGH studies from primary breast tumors (Nishizaki et al., 1997; Tirkkonen et al., 1998) reveals that the most common gains and losses are the same. In breast tumor specimens, the most common chromosomal imbalances detected by CGH are gains of 1q, 8q, 16p, 17q, and 20q and losses involving 8p, 13q, 18q, and 16q (Isola et al., 1995; Ried et al., 1995). Although literally hundreds of articles have been published describing the pattern of copy number alterations in cancer, very few of the genes affected have been identified. Consequently, investigators are currently attempting to correlate CGH data with gene expression. Hyman et al. (2002) combined CGH with cDNA arrays to analyze breast tumor cell lines and found that a significant correlation does exist between gene amplification and gene expression across the genome. The same group has focused on the 17q21→q23 amplicon, which includes *ERBB2* and is a common region of amplification in breast cancers with poor prognosis (Kallioniemi et al., 1994; Kauraniemi et al., 2001). These analyses are examples of the power achieved by combining genomic and expression approaches, and have identified several genes that are consistently overexpressed in breast cancer cell lines and advanced breast cancer.

SKY analysis of breast tumor cell line karyotypes reveals that the chromosomes most frequently involved in translocations are 8, 1, 17, 16, and 20. These chromosomes contain gene amplicons, for example, at 8q24, 17q11→q12, and 20q12→q13, that are present in up to 30% of breast carcinomas (Devilee and Cornelisse, 1994; Bieche and Lidereau, 1995; Tanner et al., 1996; Barlund et al., 1997). Furthermore, studies using conventional cytogenetics report that primary and metastatic breast carcinomas contain aberrations of chromosome 8 in up to 40% of cases (Adeyinka et al., 2000; Popescu and Zimonjic, 2002). Databases that make data from SKY/M-FISH and CGH studies in cancer available have been created by the NCI (Knutsen et al., 2005). The SKY/M-FISH and CGH Database (<http://www.ncbi.nlm.nih.gov/projects/sky/>) enables investigators to submit and analyze clinical and research cytogenetic data. The Cancer Chromosomes database integrates the SKY/M-FISH & CGH Database with the Mitelman Database of Chromosome Aberrations in Cancer (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>) and the Recurrent Chromosome Aberrations in Cancer database (<http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations>).

Relatively few studies have explored whether specific cytogenetic abnormalities can be used to stratify breast tumors with clinical course, namely metastatic relapse. Blegen et al. (2003) performed CGH on tumors from patients who had early relapse and from patients who remained free from distant metastases for more than ten years. Tumors in relapse patients showed a higher average number of chromosomal copy alterations compared to the long-term survivors, including gains of chromosome 3q, 9p, 11p and 11q and loss of 17p (Blegen et al., 2003). This aligns well with a CGH study that evaluated 76 LN-negative breast carcinomas (median follow-up 46 months) and found that a gain of 3q is a stronger predictor of recurrence than grade, tumor size, and estrogen receptor status (Janssen et al., 2003). In a CGH study of 39 invasive breast carcinomas with a mean follow-up period of 99 months, Aubele et al. (2002) identified an independent prognostic value for chromosomal gains on 11q13, 12q24, 17 and 18p. More recently, an array-CGH study of ER-positive breast cancer tissues reported that the most significant chromosomal alterations found more often in the group with metastatic recurrence within five years were loss of 11p15.5→p15.4, 1p36.33, 11q13.1, and 11p11.2 (Han et al., 2006). Array-CGH was also used by Yao et al. (2006) to show that the overall frequency of copy number alterations in regions 1q, 8q24, 11q13, 12p13, 17q21→q23, 16p13 and 20q13, correlated with the advancing nature of the tumors. Amplification of candidate loci was confirmed by quantitative PCR, and expression level analysis of genes present in these loci identified many putative target genes (Yao et al., 2006). These CGH results have significant overlap and indicate that malignant breast adenocarcinomas are characterized by specific chromosomal copy number changes.

The high divergence of karyotype data makes it difficult to define more specific aberrations involved in breast cancer etiology or progression, however, the study of copy number aberrations has proved to lead to the development of useful diagnostic markers. The tight correlation of amplification of *ERBB2* with increased expression of the gene product provides the basis for the tests that measure the DNA copy number of *ERBB2* for predicting response to Herceptin (Pauletti et al., 1996). Furthermore, combinations of FISH probes for regions of recurrent copy number aberration in other tumor types have proved useful for monitoring disease status (Sokolova et al., 2000) and for distinguishing benign from malignant skin lesions (Bastian et al., 1999).

Tissue-based signatures of breast cancer progression

Large-scale microarray analyses of human tissue specimens are building consensus gene expression profiles of various tumors, including breast (Sorlie et al., 2001, 2003, 2006; van 't Veer et al., 2002; van de Vijver et al., 2002; Ma et al., 2003, 2004; Weigelt et al., 2003; Hu et

al., 2006; Sotiriou et al., 2006). Breast cancer is clearly a heterogeneous disease and microarray analysis has been successfully used to derive distinct patterns of gene expression that correlate with molecular subtypes of breast cancers. Expression profiles that are associated with estrogen-receptor status, HER2 (*ERBB2*) expression, *BRCA1* or *BRCA2* mutations and with 'basal type' or 'luminal types' have been identified using unsupervised analyses (Sorlie et al., 2001; Hedenfalk et al., 2003; Mackay et al., 2003; Kristensen et al., 2005). These breast cancer subtypes also represent clinically distinct subgroups of patients, with differences in disease progression and overall survival. For example, ER-positive tumors tend to have the best outcome, whereas HER2+ tumors have a bad prognosis. The genes which show high variance across different tumors have been termed 'intrinsic genes' and a set of 1300 such genes has been reported to accurately subdivide the molecular subtypes across multiple array platforms and independent microarray studies (Hu et al., 2006). Studies using reduced lists of genes are beginning to be used in independent validation studies, and as few as 40 genes have been used to stratify subtypes of breast cancer using quantitative PCR assays (Perreard et al., 2006).

In a study conducted in 2002, a supervised classification analysis of DNA-microarray data predicted prognosis better than clinical prognostic indicators such as grade, stage, and nodal status. The investigators queried samples from a cohort of young (<53) breast cancer patients with LN-negative tumors. Reduction of the data identified a 70-gene signature that distinguished patient groups with good or poor prognosis with respect to the likelihood of the later development of clinical metastases (van 't Veer et al., 2002). Subsequent testing of the 70-gene prognosis gene set in an independent cohort of 295 patients confirmed the accuracy of the signature, regardless of LN status, in predicting the occurrence of distant metastases within five years of treatment (van de Vijver et al., 2002). The resulting gene-expression profile was a far more powerful predictor of the outcome of disease in young patients with breast cancer than the currently used St. Gallen or NIH consensus criteria based on clinical and histological characteristics. Importantly, the predictive power of the 70-gene signature was based upon metastasis to non-lymphatic tissues. More recently, a similar approach identified a 76-gene signature that was also successfully used to predict distant metastasis in patients with LN-negative primary breast cancer (Wang et al., 2005). The signature showed 93% sensitivity and 48% specificity when tested in an independent set of 171 LN-negative patients. As with the 70-gene signature described above, this signature outperformed the St Gallen criteria. Another study that included 159 samples obtained from both treated and untreated patients, with and without LN involvement, derived a 64-gene signature set that identified genes associated with the occurrence of distant metastasis or death within five years (Pawitan et al., 2005). Once again, the signature was validated in an independent set of 289 patients and was found to outperform clinical criteria in the stratification of risk and overall survival.

A number of studies have investigated the hypothesis that molecular programs of wound healing might be reactivated in cancer metastasis (Bissell and Radisky, 2001). Chang et al. identified consistent features in the transcriptional response of normal fibroblasts to serum, and used this so called 'wound-response signature' to reveal links between this phenomena and cancer progression in a variety of common epithelial tumors, including breast (Chang et al., 2004). The same group subsequently tested the accuracy of the wound-response signature in the same 295 patients with early-stage breast cancer used to identify and validate the 70-gene prognostic signature (van 't Veer et al., 2002). It was found that this signature could identify approximately 90% of patients who developed metastases, independently of clinical or pathological risk factors (Chang et al., 2005).

The suggestion that the risk of metastasis for breast cancer patients can be predicted by the gene expression profile of its heterogeneous primary tumor has come as something of a surprise. The prevailing idea has been that metastatic potential is acquired relatively late in the multistep process of tumor progression (Fidler and Kripke, 1977), however, the recent

microarray findings have reopened the debate on this topic by suggesting that the ability to metastasize to distant sites may be an early and inherent genetic property of breast tumors. An early study reported the derivation of a gene expression signature that distinguished primary from metastatic adenocarcinomas. Ramaswamy et al. (2003) compared expression profiles of a range of 64 primary adenocarcinomas (breast, prostate, lung, colon, uterus and ovary) to 12 unmatched metastases resulting from adenocarcinomas from the same spectrum of sites, but resected from a variety of end-organs. This comparison identified an expression pattern of 17 genes that best distinguished primary and metastatic adenocarcinomas. Notably, components of the protein translation apparatus were heavily represented in the 17-gene pan-metastasis signature. By re-applying this metastasis-associated gene expression pattern to data on 279 primary solid tumors of diverse types, including 78 stage I primary breast carcinomas, it was found that primary tumors carrying the 17-gene expression signature were more likely to be associated with metastasis and poor clinical outcome (Ramaswamy et al., 2003). Conversely, in more recent studies utilizing sets of matched tissue specimens, Weigelt et al. (2003, 2005a) found that gene expression profiles in human primary breast carcinoma are preserved in the associated distant metastasis, with respect to both subtype profile and poor prognosis signature, even if metastases develop after a long interval.

In breast cancer, it is the axillary LNs that are most often the first sites in which disseminated tumor cells can be detected (Stacker et al., 2002) and their presence or absence is currently one of the most important factors for disease course prediction for breast cancer patients (McGuire, 1987). While prognosis signatures for distant metastasis occurrence have been achieved, the same research groups have not been able to identify a classifier predicting the LN status of primary breast tumors. The comparison of gene expression profiles of 15 primary breast carcinomas and their matching LN metastases revealed no common subset of differentially expressed genes, and in the analysis of the Dutch 295 primary breast tumor profile dataset, no classifier predicting LN metastasis could be developed (Weigelt et al., 2005b). However, a study by Huang et al. (2003) did identify patterns of expression that were associated with the LN status of 89 Taipei breast cancer patients. These analyses were based on biopsy material and resulted in a predictive accuracy of approximately 90% for LN metastasis and relapse. None of the genes implicated in the overall recurrence of disease in this study were common to the 70-gene prognosis signature of van 't Veer et al. (2002), and cross-validation with a previous US study did not identify significant numbers of overlapping genes, presumably due to the different racial genetic backgrounds, but the authors concluded that LN metastasis and disease recurrence are distinct biologically. Collectively, these data suggest that proximal LN metastasis occurs independently of distant metastasis (Weigelt et al., 2005b), but it remains unclear whether metastasis to more distant sites proceeds sequentially from LN metastasis or in parallel via a hematogenous route (Chambers et al., 2000).

Current limitations and future prospects

Improved prognostic markers are clearly needed in order to better stratify patients with respect to their risk for developing metastases, and important advances towards this have been made. However, many limitations with microarray-based tumor classifications exist. The subtype and prognosis gene sets remain relatively broad and no quantitative information is available. Subsequent tests based on the qualitative presence or absence of a single gene, or even multiple genes are unlikely to be accurate enough for clinical assay development. Due to economic considerations, the number of samples tested to date remains very small relative to the feature dimensionality present on microarray chips. Consequently, as more data are added the current signatures can change dramatically, and another round of adjustment to tissue-based signatures will occur now that arrays that provide complete genome coverage have become available. Furthermore, studies that assay grossly homogenized tissues define an average gene expression signature that does not account for variations in tumor complexity, heterogeneity or non-tumor

cell contributions. Gene expression signatures may reflect genetic and/or epigenetic phenomena, effects of growing in different tissue environments, or simply the composition of cell types. These problems have caused some debate regarding the timing of the next phase of current gene signature clinical testing (Brenton et al., 2005; Loi et al., 2005), but it is clear that independent validation of these signatures is the next step. Indeed, multi-institutional studies are being organized (TransBIG in Europe (Tuma, 2004) and NCI PACCT in the US) to facilitate the evaluation of breast cancer genomic signatures in far greater numbers of clinical specimens.

At this stage, microarray data in itself does not provide definitive answers. The inclusion of a gene in a prognostic list that is determined by a supervised classification method does not indicate the importance of that gene in cancer biology, nor does it provide functional insight into the underlying mechanisms of disease. However, these analyses have provided a wealth of candidate genes and pathways for further study. In order to investigate the role of these genes in the mechanisms of metastasis animal models are needed.

Animal models of metastasis

Signatures of breast cancer metastasis

The multistep nature of metastasis poses difficulties in both design and interpretation of experiments to unveil the mechanisms causing the process. Studies on excised human tissues are complicated by the variance of genetic background between individuals and by the cellular heterogeneity of a complex tissue mass. Critical to the experimental analysis of metastasis has been the isolation of human tumor cell lines and the ability to study their behavior *in vivo* by inoculation into immune-compromised mice. Several established human breast cancer cell lines with varying documented abilities of invasiveness and/or migration *in vitro* are available, and some are capable of spontaneous metastasis *in vivo*, i.e. dissemination from a primary tumor and proliferation in a distal site (Price et al., 1990). This xenograft model represents an experimental system in which the role of specific genes can be screened and tested. However, many breast cell lines, especially those isolated from pleural effusions, are polyclonal in nature and composed of cell populations that are heterogeneous in metastatic phenotype, thus in order to use cell lines as models in studies seeking to define genes causing metastasis it is optimal to isolate single cell progenies from the parental cell line source. Through *in vivo* selection of monoclonal cultures of the MDA-MB-435 breast tumor cell line we were able to characterize a pair of single cell progenies (M4A4 and NM2C5) which differ markedly in their ability to complete the metastatic process (Urquidi et al., 2002). When orthotopically inoculated into athymic mice, both cell lines form primary tumors, but only M4A4 is capable of metastasis to the lungs and lymph nodes (Urquidi et al., 2002; Goodison et al., 2003). These isogenic cell lines of opposite metastatic propensity constitute a stable and accessible model for the identification of genes involved in the process of tumor metastasis. We have performed multiple comparative analyses of these paired cell lines, including cytogenetic analyses and evaluation of the expression of a number of gene products previously implicated in cellular transformation and metastasis (Urquidi et al., 2002; Agarwal et al., 2003; Goodison et al., 2003, 2005a). To further elucidate the extent of the molecular changes associated with acquisition of the metastatic phenotype in this model, we recently employed a genome-wide expression profiling approach. Intensity modeling and hierarchical clustering analysis revealed a subset of 85 genes (>2-fold change) whose expression was statistically correlated with metastatic phenotype (Goodison et al., 2005b). Some genes in this group have been implicated in invasion, tumor cell proliferation and/or metastasis previously, but GTPase signaling components were one of the most-well-represented functional groups. Restoration of the expression of deleted in liver cancer-1 (DLC-1), a Rho-GTPase-activating protein, in metastatic M4A4 cells resulted in the inhibition of migration and invasion *in vitro* and a significant reduction in the ability of these cells to form pulmonary metastases in athymic mice. DLC-1 has specific GTPase activating protein functions for RhoA and Cdc42 (Wong et al.,

2003), members of the Rho family that are consistently overexpressed in breast tumors (Fritz et al., 1999). The finding that *DLC1* can act as a 'metastasis-suppressor gene' supports an influential role for GTPase signaling in tumor progression. Metastasis suppressor genes are potential candidates for marker development because, by definition, their loss should be associated with the acquisition of metastatic potential (Shevde and Welch, 2003).

A similar methodological approach was used by Kang et al. (2003) to study breast tumor cell homing to specific organ sites. The investigators were able to derive monoclonal lines from the MDA-MB-231 breast tumor cell line that had differing degrees of ability to form tumor deposits in murine bone. The experimental system was necessarily different to that used in a lung metastasis model, in that cells need to be inoculated directly into the heart in order to get sufficient cell numbers to the arterial side of the murine body and to the skeleton. Using comparative microarray analyses of the parental MDA-MB-231 cell line and a variant selected for bone colonization, the investigators identified a bone metastasis gene expression signature of 102 genes. Interestingly, in line with our lung metastasis model, *DLC-1* was found to be downregulated in breast cell populations which were highly metastatic to bone (Kang et al., 2003) suggesting that some signaling pathways may be pivotal to metastatic efficiency regardless of the target organ. Transfection of combinations of genes confirmed some of these genes (*IL11*, *OPN/SPPI*, *CTGF* and *CXCR4*) as being functionally involved in the efficiency of MDA-MB-231 cell bone colonization. Both *IL11* and *CTGF* are known to be activated by TGF β , suggesting a potential prometastatic role for this cytokine in bone. The same group continued to derive monoclonal lines from MDA-MB-231 and test their propensity to colonize other organs. Comparison of transcriptional profiles of cell populations highly or weakly metastatic to the murine lung resulted in a 54-gene lung metastasis signature (Minn et al., 2005). This gene set was distinct from the bone metastasis signature overall (Kang et al., 2003), but some genes were common to both MDA-MB-231 models, including *MMP1* and *CXCR4* (Minn et al., 2005). The MDA-MB-231 lung metastasis signature gene set was also compared to expression profiles obtained from a cohort of 82 therapeutically excised primary breast carcinomas from patients with known metastatic status. This revealed that expression profiles in the primary tumors from patients with subsequent lung metastases, but not bone metastases, correlated to some extent with the lung metastasis signature (Minn et al., 2005).

Another variation on clonal models of metastasis has been characterized by Lev et al. (2003). This model consists of a more aggressive metastatic clone (GILM2) derived from the weakly metastatic GI101A human breast cancer cell line which metastasizes to the lung and lymph nodes when inoculated orthotopically (Hurst et al., 1993). Microarray-based comparison with the parental line identified a list of 106 genes that were differentially expressed (>2.5 fold) in the highly metastatic GILM2 variant (Kluger et al., 2005). Immunohistochemical confirmation in human breast tissue specimens of the expression of three markers, heat shock protein 70 (HSP70), chemokine ligand 1 (CXCL1), and secretory leukocyte protease inhibitor (SLPI), revealed that the expression of all three genes was correlated with lymph node involvement, and the expression of HSP70 and CXCL1 was associated with decreased overall patient survival. Interestingly, we have identified and functionally proven the role of SLPI in an invasion-independent model of metastasis (Sugino et al., 2004). A wholly murine model of spontaneous breast cancer metastasis to multiple sites has been characterized and used to identify genes involved in metastatic progression. Several syngeneic tumor lines with a spectrum of metastatic phenotypes were isolated from a spontaneous mammary tumor in a BALB/cfC3H mouse (Lelekakis et al., 1999). When inoculated orthotopically, the resulting primary tumors are either nonmetastatic or produce spontaneous metastases to lymph node, lung and/or bone (Eckhardt et al., 2005). To identify metastasis-related genes, the investigators grouped the expression profiles of a weakly meta-static group and compared these with the profiles of a highly metastatic group using mouse cDNA arrays. A metastasis signature of 216 genes was derived, of which 125 were known genes. A significant proportion of genes

belonging to extracellular matrix protein families had elevated expression in the highly metastatic tumors. The role of one of these genes, *POEM* (nephronectin), was further investigated using RNA interference technology, and decreased *POEM* expression in 4T1 tumors significantly inhibited spontaneous metastasis to the lung, kidney and bone. *POEM* is a secreted ECM molecule that has been reported to be involved in kidney morphogenesis and the development of bone (Brandenberger et al., 2001; Morimura et al., 2001), and so it is logical to assume that the expression of *POEM* by the tumor cells in the murine model is critical for the establishment of metastases in these organs.

While genetic studies of clinical specimens will continue to be informative, they provide only a snapshot of a complicated disease state, and there are few experimental opportunities in such analyses. Thus, the study of breast cancer progression requires experimental models for the investigation of the links between genetic profiles and a more aggressive tumor phenotype. Animal models provide a powerful resource for the identification and investigation of genes essential in distinct steps of the metastatic cascade, in site-specific homing, in complex tumor-host interactions, and enable the identification of targets that are optimal for therapeutic perturbation. The unique advantage of the clonal metastasis models described above is the ability to profile cells of opposing metastatic phenotype that originate from a common genetic background. The human origin of the cells and the spontaneous acquisition of the distinct phenotypes make them optimal for investigations into genetic changes that correlate with metastatic sufficiency. Investigators can alter the expression or activity of single, or multiplexes of candidate genes in the clonal lines with known metastatic characteristics and monitor which specific mechanisms are perturbed by comparison with isogenic controls. Many of these candidate genes will come from tissue-based analyses, such as those described above. There are certainly limitations regarding extrapolation from studies in a murine host to the human clinical situation, perhaps particularly acute when investigating tumor-stroma interactions, but *in vivo* functional analyses can distinguish which genes are essential in tumor progression, and better delineate pathway and signaling network interactions in tumor cells, and thus may in turn aid the prioritization of genes in prognostic signatures. In this way the two fields can overlap and mutually benefit.

Signature overlap

Given the differences in experimental design used in the production of the nine signatures described in this review, the presence of specific genes in multiple tissue or model signatures, or better yet, in both, might implicate pathways or mechanisms for further research focus. The nine signatures contain a total of 683 non-redundant genes for which a unique gene identifier (Entrez Gene ID) is available to all signatures, and of these, 47 genes were found to appear in at least two of the nine signatures. The details of the distribution of these signature intersect genes is depicted in Table 1. Grouping genes by cellular location revealed that the most prevalent (34%) localization was extracellular. This is a common theme revealed in metastasis studies due to the complex interactions that need to occur with the secondary tissue environment and the requirement for extracellular matrix remodeling during the development of metastasis. This was reinforced by the fact that 23% of common genes were localized to the plasma membrane. The majority of genes were common to only two signatures, just a few appeared in three signatures.

The overlap between tissue-based studies was weak (signature gene range 17–83). Only five genes appeared in two signatures resulting from tissue-based microarray studies (Table 1). There was no overlap between the 17-gene metastasis signature of Ramaswamy et al. (2003) with any other tissue-based study. Of the microarray studies discussed in this review, those performed by van 't Veer et al. (2002) and Wang et al. (2005) are the most aligned in many ways. While there were many differences between the studies with respect to population,

clinical information availability, platform, and statistical treatment, both studies evaluated lymph-node negative samples and the platforms had thousands of genes in common, yet only one gene, cyclin E2 (*CCNE2*), was common to both signatures. Three genes (*LOC51203*, *PRC1* and *L2DTL*) of the 64 genes in the study by Pawitan et al. (2005) were present in the 70-gene prognostic signature identified by van 't Veer and colleagues, and one gene (*MLF1IP*) was present in both the Pawitan and the Wang signatures.

Although there are clear differences between studies, including selection of patients according to different inclusion and exclusion criteria, different gene expression platforms, and mathematical analysis of the data, thousands of genes were common to all studies and the overall goals were similar. The lack of convergence of breast cancer prognostic signatures to date makes it difficult to discern which signatures are the most accurate and optimal for prognosis and/or potential therapeutic development going forward.

There was significantly more overlap between model signatures (signature gene range 85–143) than that observed between tissue signatures. A total of 42 genes were found to be present in at least two model-based signatures. Although each model was derived from a different original tumor source, the clonality of the models represents a far simpler situation than that found in excised tumors consisting of complex mixtures of tumor and non-tumor cell components. The largest overlap (ten genes) was between the MDA-MB-231 bone and lung metastasis signatures. This is likely due to the fact that both signatures are derived from the same parental cell line, MDA-MB-231. However, there was also significant overlap (seven genes) between the MDA-MB-435 model and the GILM2 model. There was also considerable overlap (five genes) between the MDA-MB-435 model and the MDA-MB-231-bone signature, and this is in line with our recent observations that the M4A4 metastatic counterpart of the MDA-MB-435 model is also aggressively metastatic to bone when inoculated via an intra-cardiac route (article in preparation).

Even though there was a relatively high intersection between the model signatures, there are major differences between the five models described in this review. The MDA-MB-231 model is a xenograft model of 'experimental' metastasis involving the injection of human breast tumor cells directly into the circulatory system of immunocompromised mice, resulting in metastases in specific organs. This model has been valuable for analyzing the final stages of metastasis, and can identify genes that regulate the colonization of specific tissues (Kang et al., 2003; Minn et al., 2005), but does not encompass the initial stages of metastasis. Conversely, the MDA-MB-435 and the GILM2 model are xenograft models of 'spontaneous' metastasis involving the formation of a primary tumor in an orthotopic site, in this case breast, from which the cells must disseminate naturally. These spontaneous metastasis models recapitulate all of the steps of metastatic efficiency, from escape from the primary tumor to the establishment of distant metastases, and perhaps mimic more accurately the progression of breast cancer observed clinically. However, the MDA-MB-231, MDA-MB-435 and the GILM2 models involve human tumor cells forming tumors in a mouse host and, therefore, may not best mimic important tumor-host interactions. The 4T1 mouse model has the advantage of being syngeneic and so may provide more insight into such interactions, but on the other hand, the genes identified as being associated with the tumor cell phenotype may not be functionally equivalent in human cells. Rather than being a hindrance to forming converging hypotheses of metastasis, the testing of candidate metastasis genes across these models with their inherent differences will provide powerful validation of specific gene functionality.

A total of seven genes were found to be common to at least one tissue-based signature and one model-based signature (Table 1). Three to four such intersect genes were present in each of the tissue-based signatures, except for the 70-gene signature of van 't Veer et al. which had no intersect with any model signature. All model signatures were represented in the tissue vs.

model intersect. A total of four genes were present in at least three signatures. Three of these (*HLA-DPBI*, *ARHGDI1B* and *C14orf139*) were present in two model signatures and one tissue signature, whereas the *CTGF* gene was present in three model signatures but no tissue signature.

Chromosomal location and alignment with CGH studies

Alignment of the genes in the signature intersect list with chromosomal location reveals that they are randomly spread across the genome with extremes being an absence of any genes on chromosome 5, and seven genes being located on chromosome 6. Yet, per chromosome the intersect genes appear to be tightly clustered in many cases. For example, five of the seven genes on chromosome 6 are located within 6p22→p21, both genes on chromosome 7 are located at 7q22, and all three genes on chromosome 11 are located at 11q22→q23. The *NUSAPI* gene located at 15q15.1 is one of the few genes that were present in two of the tissue-based analyses. Interestingly, alignment of our 171-gene metastasis signature using 'Genome View' revealed that three of the genes more highly expressed in metastatic M4A4 cells were grouped together at 15q15 (Goodison et al., 2005b). Moreover, this region was revealed to be involved in cell-specific chromosomal rearrangements in the molecular cytogenetic analysis of MDA-MB-435 sub-lines (Goodison et al., 2005a). Although specific genes may not make the overlap lists, future meta-analyses of gene expression profiles with respect to chromosomal location may reveal metastasis-associated DNA hot-spots.

Notably, many of the intersect genes are located at loci revealed by CGH to be perturbed in breast cancer. The one gene on chromosome 3 that was present in at least two signatures is *TNFSF10*, a member of the tumor necrosis factor ligand superfamily, located at 3q26. This region was found to be a strong predictor of breast cancer recurrence by CGH (Janssen et al., 2003). Furthermore, four clusters of intersect genes align very closely with recent reports of copy number alterations which correlate with the advancing nature of breast tumors. These include two genes located at 8q22→q24, two genes at 12p13.1→p12.3, one gene at 16p13, and the two genes on chromosome 17 at 17q21→q22. These four regions were all identified through the use of array-CGH to be amplified in late-stage breast cancer (Yao et al., 2006). The considerable alignment between the CGH and the gene expression microarray data suggests that the combination of DNA- and RNA-based approaches can inform and guide the search for essential cancer-associated genetic events.

Chemokines

A common theme throughout the cell line model metastasis signatures are components of the human chemokine system. The chemokines and their receptors are a family of small secreted molecules that regulate the migration around the body of cells in the lymphoid system, a process that shares many characteristics with the successful dissemination of tumor cells in the body. Chemokine receptors are G-protein-coupled cell surface proteins expressed on leukocytes and many nonhematopoietic cells. These receptors bind various chemokines that are constitutively expressed in distinct tissue microenvironments. The chemokine CXCL1 and the chemokine receptor CXCR4 are present in the signature intersect gene list (Table 1). CXCR4 is the most widely expressed chemokine receptor in many different cancers. Primary breast tumors express the CXCR4 receptor, and target sites of breast cancer metastases, such as lung and bone, express more CXCL12, the CXCR4 ligand, than other organs (Muller et al., 2001). Antibody, or peptide-mediated blocking of the CXCR4 receptor decreased breast cancer cell invasiveness in vitro and also the bulk of metastases in the MDA-MB-231 xenograft mouse model (Muller et al., 2001; Liang et al., 2004; Lapteva et al., 2005). Smith et al. (2004) used stable RNAi to reduce expression of CXCR4 in murine 4T1 cells, a highly metastatic mammary cancer cell line that is a model for advanced human breast cancer. The knockdown of CXCR4 significantly limited the growth of orthotopically transplanted breast cancer cells in all mice and even

prevented primary tumor formation in some. Most importantly, all mice transplanted with CXCR4 RNAi cells survived without developing macroscopic metastases (Smith et al., 2004). Recent studies have also identified a role for CXCL12 as a proliferation factor for breast tumor cells and in recruiting progenitor endothelial cells required for angiogenesis (Orimo et al., 2005). CXCR4 expression is also a prognostic marker in various types of cancer, including breast carcinoma. In a recent histological study of over 2000 breast carcinoma cases, CXCR4 was found to be expressed in the majority of cancers, and cytoplasmic expression was associated with parameters of tumor aggression, including grade and lymph node status, and had prognostic value (Salvucci et al., 2006). G protein-coupled receptors are considered among the most desirable targets for drug development (Li et al., 2005), thus, CXCR4 antagonists may become effective agents for the treatment of various malignancies.

CCN family

The connective tissue growth factor gene (*CTGF*) was present in three signatures, and the closely related *CYR61* gene was present in two signatures. These genes encode proteins that belong to the CCN family of secreted regulatory factors involved in angiogenesis, chondrogenesis, and wound healing (Brigstock, 1999; Perbal, 2004). CTGF and CYR61 also promote endothelial cell growth, migration, adhesion, and survival in vitro, and many of these actions are mediated at least partly through interactions with integrins (Lau and Lam, 1999). The proangiogenic activity of CTGF and CYR61 supports a role in the establishment and functioning of the vasculature in metastasis. Recent studies have revealed an influence of CTGF in osteolytic bone metastases and its expression shown to be regulated by PTHrP (Shimo et al., 2006). In a large immunochemical study of human breast tumor specimens, high levels of CYR61, but low levels of CTGF were associated with poor prognosis and meta-static disease (Jiang et al., 2004). Interestingly, both genes are transcriptionally regulated by TGF β (Bartholin et al., 2007), a multifunctional growth factor, which is thought to promote breast tumor metastasis and invasiveness (Beisner et al., 2006).

ARHGD1B

Another gene that was present in three metastasis signatures was the Rho GDP dissociation inhibitor (GDI) beta (*ARHGD1B*) gene, a member of the Rho GDI family which are major regulators of Rho GTPases. Evidence for the influence of Rho GTPase signaling in tumor metastasis is growing. The expression of the RhoC molecule was identified as being correlated with metastatic propensity in an increasingly metastatic series of the melanoma A375 cell line and the Rho-GTPases, RhoGD12 and Rac1, have been associated with aggressive phenotypes in experimental metastasis assays when testing bladder and breast cell line models respectively (Bourguignon et al., 2000; Seraj et al., 2000). In our own work, restoration of one of these GTPases, deleted in liver cancer-1 (DLC-1), in metastatic M4A4 cells resulted in the inhibition of migration and invasion in vitro and a significant reduction in the ability of these cells to form pulmonary metastases in athymic mice. Although RhoGDIs are known to inhibit Rho activities, recent studies indicate that RhoGDIs can also act as positive regulators through their ability to target Rho GTPases to specific sub-cellular membranes or to protect the GTPases from degradation by caspases. This dual role makes them promising cellular targets for novel anticancer drugs (Fritz and Kaina, 2006).

While the signatures obtained from different microarray platforms might reveal different gene sets, they may be reporting related biological processes. Annotation of the 47 signature intersect genes using GenMapp (<http://www.genmapp.org/>) and KEGG (<http://www.genome.jp/kegg/>) pathway information summaries (Doniger et al., 2003; Kanehisa et al., 2006) revealed only one significant grouping. Three genes (*FNI*, *LAMB1* and *COL1A1*) belong to the inflammatory response pathway. This finding perhaps reinforces the concept that gene expression programs related to normal physiological responses, as described above with respect to wound healing,

may confer an increased risk of metastasis (Chang et al., 2004). Investigation of potential interaction between the gene products of the 47 signature intersect genes was queried using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (<http://string.embl.de/>) (von Mering et al., 2005). Evidence for protein associations derived from experimental data and from the mining of databases and the literature was sought using the Protein Mode for maximum specificity. With the exception of *C14orf139*, *DTL*, and *MLF1IP* genes, and the pseudogene *ANXA2P1*, all gene products had annotations in the STRING database. Without entering into the detailed specific interactions here, it is worth noting that the analysis revealed that the transcription factor Sp1 may be a unifying factor for many of the proteins present in the signature intersect list. Sp1 was present in the node that connected a group that included CTGF, CYR61, MMP1, and several collagens with the chemokines CXCR4 and CXCL1. Sp1 protein expression is known to be elevated in breast carcinomas (Zannetti et al., 2000), and several studies show that Sp1 transcriptionally regulates many genes previously implicated in cancer progression, including *VEGF* (Ryuto et al., 1996; Abdelrahim et al., 2004), *CCND1*, *FOS*, and the anti-apoptosis *BCL2* gene (Safe and Abdelrahim, 2005). Furthermore, Sp proteins have been shown to regulate the expression of TGF β receptors (Ji et al., 1997; Jennings et al., 2001) and thus will impact TGF β signaling, which in turn regulates the CCN proteins discussed above. While additional research is required to determine direct linkages between the expression of Sp1 and its downstream factors in various tumor types, as a protein high in the regulatory hierarchy, Sp1 may be an important prognostic factor and therapeutic target for metastatic disease. If a comparison of the intersect of just the nine signatures described in this review can implicate some common themes, a comprehensive meta-analysis of all available, unprocessed, metastasis-associated gene expression microarray data with more of a focus on biochemical and signaling pathways than gene identities may reveal much about the mechanisms of metastatic efficiency.

Future directions

Genes revealed in analysis of multiple models, and those that are part of any overlap between models and tissue-based studies, will be a logical focus for functional investigation. As described above, relevant models for such investigations do exist, but it would be beneficial to develop more such models in order to identify those genes and pathways that are implicated across a number of platforms and are thus likely to be pivotal to metastatic efficiency. Likewise, more tissue-based studies are needed, both for discovery and validation of existing candidate genes. As discussed above, little overlap between tissue-based studies has been evident to date, presumably due to differences in tissue sources and processing and availability of consistent clinical information. The initiation of large-scale consortia such as TransBig and PACT will greatly facilitate this through standardization of all parameters. While there are economical and ease-of-use reasons to reduce the number of genes in diagnostic and prognostic signatures for routine clinical use, global gene expression profiling can have more specific roles. For example, based upon the expression of drug targets, and/or the expression of genes that metabolize these drugs, gene expression profiling may guide adjuvant therapy for individuals. Two studies that highlight this approach have reported an association in breast cancer patients between gene expression signatures and drug sensitivity to docetaxel or to a combination regimen containing paclitaxel, fluorouracil, doxorubicin and cyclophosphamide (Brown et al., 2004; Villeneuve et al., 2006). This use of gene expression profiling is likely to have a more rapid role in tailoring individual patient treatment.

Prognostic signatures gleaned from gene expression data alone have been achieved through normalized microarray data comparison across literally hundreds of specimens. However, the heterogeneity of breast cancer as a disease, a fact confirmed by gene expression studies, plus the variation in patient genetic background makes the goal of predicting the outcome for an individual patient far from achievable at this stage. One way to improve upon this situation is

to further stratify patients using clinical measures and current understanding of disease pathology. For example, the expression of the estrogen receptor (ER) and/or the Her2-neu gene at diagnosis is currently used to subdivide patients. ER and Her2-neu expression have a marked influence on the expression of many of the genes associated with breast cancer, and are thought to have an important impact on survival and are used as independent prognostic factors (Hu et al., 2006; Nicolini et al., 2006). Investigators have begun to use clinical parameters to subdivide patients in order to further refine the prognostic classifiers for specific circumstances. For example, by combining expression data with ER status and age as clinical variables, Dai et al. (2005) identified a subgroup of patients in which a set of proliferation-associated genes was a strong predictor of poor outcome and occurrence of metastasis.

The majority of expression studies to date have attempted to develop genetic marker-based prognostic systems that might replace the existing clinical criteria, rather than incorporating the valuable clinical information contained in established clinical markers. Given the complexity of breast cancer prognosis, a more promising strategy may be to combine both clinical and genetic marker information that may be complementary (Brenton et al., 2005). Along these lines, Pittman et al. (2004) have described a comprehensive modeling approach based on statistical classification tree models that evaluate the contributions of multiple forms of data, both clinical and genomic, in order to define multiple risk factors that associate with clinical outcome. A study of primary breast cancer recurrence demonstrated that models using 'metagenes' derived from microarray expression data combined with traditional clinical risk factors improved the accuracy of prediction on an individual basis over genomic or clinical data alone (Pittman et al., 2004). Furthermore, we have recently performed a computational study using the Dutch 70-gene prognosis signature and associated clinical information. The recently proposed I-RELIEF algorithm was used to identify a 'hybrid signature' through the combination of both genetic and clinical markers. The hybrid signature performed significantly better, with respect to specificity and odds-ratio, than the 70-gene signature, clinical markers alone and the St Gallen and NIH consensus criteria (Sun et al., 2007).

One of the major differences between gene expression-profile studies performed to date is in data processing. The advent of microarray technology has spawned a whole new field of bioinformatics which has the common goal of optimizing the accuracy and the applicability of the information gleaned from the wealth of data now available, but individual studies have often used quite different approaches which do not always facilitate inter-study comparisons. A number of groups have tackled this by designing algorithms that enable meta-analysis of publicly available data, including analysis across multiple array platforms (Bammler et al., 2005; Rhodes et al., 2005; Segal et al., 2005). These analyses seek to not only identify inter-study overlap but also to identify multidimensional interaction networks which in turn may implicate molecular hierarchies and regulatory mechanisms involved in cancer gene expression. Rhodes et al. (2004) analyzed compiled multiple data sets from some 3,700 cancer samples and developed an analytical strategy to assess the intersection between profiles. This approach identified components of transcriptional profiles which are similar across many cancer types, including breast. Specific to breast cancer, Shen et al. (2004) applied a Bayesian modeling strategy to analyze four independent microarray studies to derive a 'meta-signature' associated with breast cancer prognosis. The meta-signature had better prognostic performance than any of the classifiers of survival in each study, and which had minimal overlap with each other. These studies reveal that there are many legitimate ways to analyze array data sets, and it is hoped that the formation of consortia to monitor large-scale tissue-based studies will employ a standardized arsenal of bioinformatic and statistical analyses. Ongoing computational developments that enable inter-study comparison and the incorporation of distinct forms of data, including genomic and clinical data, will provide platforms for the refinement of cancer-related gene signatures and lead to more economical and accurate prognostic systems that may facilitate personalized patient evaluation and treatment decisions. These integrated analyses

also have the potential to highlight pivotal genes and pathways that are likely part of the biological driving force of metastasis.

Another way to significantly improve the quality of data resulting from tissue profile analyses is to provide more accurate histopathological details of each sample upfront. Concordance between two pathologists has been investigated and found to range from 50% to 85% (Robbins et al., 1995), and inaccuracy in sample evaluation may account for considerable error when array data is obtained from partial tissue specimens and subsequently related to parameters such as disease staging and histological grade. Furthermore, solid tumors are complex entities composed of malignant cells mixed and interacting with nonmalignant cells such as normal or benign epithelia, stromal counterparts and lymphocytic infiltrate. Thus, molecular analyses by standard gene expression profiling are limited when tissues are crudely homogenized causing loss of information on non-tumor cell types in sample preparation. Gene expression differences derived from such tumor samples may primarily reflect varying proportions of the non-neoplastic tumor components. In order to try to overcome these limitations we have previously employed a regression-based informatics approach for the identification of cell-type-specific patterns of gene expression in prostate cancer (Stuart et al., 2004). Through intense histological evaluation, we scored 88 prostate specimens for relative content of neoplastic and non-neoplastic components. The proportions of these cell types were then linked in silico to gene expression levels determined by microarray analysis, revealing unique cell-specific profiles. Gene expression differences for malignant and nonmalignant epithelial cells could be identified without being confounded by contributions from stroma that dominate many samples. Validation of selected cell-specific expression patterns confirmed that this analysis allowed segregation of molecular markers into more discrete and informative groups. This investigative approach is applicable to tumor marker discovery in any solid tumor and is an example of how more detailed evaluation of the tissue specimen can further refine expression profile data.

Another way to separate the cells of a complex tissue from each other is through microdissection. This technique has most often been used to evaluate only several genes per sample (Wang-Rodriguez et al., 2003), but the advent of semi-automated, laser capture microdissection (LCM) systems now makes it feasible to use this procedure to obtain enough material for microarray analysis (Nishidate et al., 2004; Dahl et al., 2006; Schuetz et al., 2006; Yang et al., 2006). The combination of laser microdissection and gene expression profiling has been used to explore the gene expression changes that are associated with the early stages of breast cancer progression and ER-regulated genes (Ma et al., 2003; Yang et al., 2006). In a study by Dahl et al. the investigation of matched pairs of invasive ductal breast cancer and corresponding benign breast tissue by LCM and cDNA array profiling ultimately led to the identification of karyopherin alpha2 (*KPNA2*) expression as being associated with shorter overall survival and recurrence-free survival (Dahl et al., 2006).

Finally, it has to be stated that genomic approaches are unlikely to be adequate as a sole prognostic and predictive platform in breast cancer. Most transcriptome array analyses implicitly treat mRNA expression as a surrogate for protein activity level, an assumption that does not account for processes such as mRNA stability, protein degradation and post-translational modifications. The key proteins driving tumor progression may very well undergo abnormal post-translational modifications. While the animal models of metastasis described above lend themselves to proteomic analysis (Kreunin et al., 2004), proteomic analysis of complex tumor tissues is far more difficult. Proteomic techniques are far more limited with respect to high-throughput screening platforms. Current techniques cannot yet achieve high proteome coverage rates and they need large amounts of material given that there is no method for amplification. Due to the vastly superior gene screening advantages of microarrays, it is logical that transcriptome data should guide proteomics, for example, validation of array data should be performed at the protein level wherever possible.

References

- Abdelrahim M, Smith R 3rd, Burghardt R, Safe S. Role of Sp proteins in regulation of vascular endothelial growth factor expression and proliferation of pancreatic cancer cells. *Cancer Res* 2004;64:6740–6749. [PubMed: 15374992]
- Adeyinka A, Kytola S, Mertens F, Pandis N, Larsson C. Spectral karyotyping and chromosome banding studies of primary breast carcinomas and their lymph node metastases. *Int J Mol Med* 2000;5:235–240. [PubMed: 10677562]
- Agarwal D, Goodison S, Nicholson B, Tarin D, Urquidi V. Expression of matrix metalloproteinase 8 (MMP-8) and tyrosinase-related protein-1 (TYRP-1) correlates with the absence of metastasis in an isogenic human breast cancer model. *Differentiation* 2003;71:114–125. [PubMed: 12641565]
- Albertson DG. Profiling breast cancer by array CGH. *Breast Cancer Res Treat* 2003;78:289–298. [PubMed: 12755488]
- Aubele M, Auer G, Braselmann H, Nahrig J, Zitzelsberger H, et al. Chromosomal imbalances are associated with metastasis-free survival in breast cancer patients. *Anal Cell Pathol* 2002;24:77–87. [PubMed: 12446957]
- Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2005;2:351–356. [PubMed: 15846362]
- Barlund M, Tirkkonen M, Forozan F, Tanner MM, Kallioniemi O, Kallioniemi A. Increased copy number at 17q22→q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer* 1997;20:372–376. [PubMed: 9408753]
- Bartholin L, Wessner LL, Chirgwin JM, Guise TA. The human CYR61 gene is a transcriptional target of transforming growth factor beta in cancer cells. *Cancer Lett* 2007;246:230–236. [PubMed: 16616811]
- Bastian BC, Wesslmann U, Pinkel D, Leboit PE. Molecular cytogenetic analysis of spitz nevi shows clear differences to melanoma. *J Invest Dermatol* 1999;113:1065–1069. [PubMed: 10594753]
- Beisner J, Buck MB, Fritz P, Dippon J, Schwab M, et al. A novel functional polymorphism in the transforming growth factor-beta2 gene promoter and tumor progression in breast cancer. *Cancer Res* 2006;66:7554–7561. [PubMed: 16885354]
- Bieche I, Lidereau R. Genetic alterations in breast cancer. *Genes Chromosomes Cancer* 1995;14:227–251. [PubMed: 8605112]
- Bissell MJ, Radisky D. Putting tumours in context. *Nat Rev Cancer* 2001;1:46–54. [PubMed: 11900251]
- Blegen H, Will JS, Ghadimi BM, Nash HP, Zetterberg A, et al. DNA amplifications and aneuploidy, high proliferative activity and impaired cell cycle control characterize breast carcinomas with poor prognosis. *Anal Cell Pathol* 2003;25:103–114. [PubMed: 12775914]
- Bourguignon LY, Zhu H, Shao L, Chen YW. Ankyrin-tiam1 interaction promotes Rac1 signaling and metastatic breast tumor cell invasion and migration. *J Cell Biol* 2000;150:177–191. [PubMed: 10893266]
- Brandenberger R, Schmidt A, Linton J, Wang D, Backus C, et al. Identification and characterization of a novel extracellular matrix protein nephronectin that is associated with integrin alpha8beta1 in the embryonic kidney. *J Cell Biol* 2001;154:447–458. [PubMed: 11470831]
- Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: Ready for clinical application? *J Clin Oncol* 2005;23:7350–7360. [PubMed: 16145060]
- Brigstock DR. The connective tissue growth factor/cysteine-rich 61/nephroblastoma overexpressed (CCN) family. *Endocr Rev* 1999;20:189–206. [PubMed: 10204117]
- Brown I, Shali K, McDonald SL, Moir SE, Hutcheon AW, et al. Reduced expression of p27 is a novel mechanism of docetaxel resistance in breast cancer cells. *Breast Cancer Res* 2004;6:R601–607. [PubMed: 15318941]
- Chambers AF, Naumov GN, Vantyghem SA, Tuck AB. Molecular biology of breast cancer metastasis. Clinical implications of experimental studies on metastatic inefficiency. *Breast Cancer Res* 2000;2:400–407. [PubMed: 11250733]

- Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, et al. Gene expression signature of fibro-blast serum response predicts human cancer progression: Similarities between tumors and wounds. *PLoS Biol* 2004;2:E7. [PubMed: 14737219]
- Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* 2005;102:3738–3743. [PubMed: 15701700]
- Dahl E, Kristiansen G, Gottlob K, Klamann I, Ebner E, et al. Molecular profiling of laser-microdissected matched tumor and normal breast tissue identifies karyopherin alpha2 as a potential novel prognostic marker in breast cancer. *Clin Cancer Res* 2006;12:3950–3960. [PubMed: 16818692]
- Dai H, van 't Veer L, Lamb J, He YD, Mao M, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 2005;65:4059–4066. [PubMed: 15899795]
- Devilee P, Cornelisse CJ. Somatic genetic changes in human breast cancer. *Biochim Biophys Acta* 1994;1198:113–130. [PubMed: 7819270]
- Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. Mappfinder: Using gene ontology and genmapp to create a global gene-expression profile from microarray data. *Genome Biol* 2003;4:R7. [PubMed: 12540299]
- Eckhardt BL, Parker BS, van Laar RK, Restall CM, Natoli AL, et al. Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. *Mol Cancer Res* 2005;3:1–13. [PubMed: 15671244]
- Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977;197:893–895. [PubMed: 887927]
- Fritz G, Kaina B. Rho GTPases: Promising cellular targets for novel anticancer drugs. *Curr Cancer Drug Targets* 2006;6:1–14. [PubMed: 16475973]
- Fritz G, Just I, Kaina B. Rho GTPases are overexpressed in human tumors. *Int J Cancer* 1999;81:682–687. [PubMed: 10328216]
- Goodison S, Kawai K, Hihara J, Jiang P, Yang M, et al. Prolonged dormancy and site-specific growth potential of cancer cells spontaneously disseminated from nonmetastatic breast tumors as revealed by labeling with green fluorescent protein. *Clin Cancer Res* 2003;9:3808–3814. [PubMed: 14506175]
- Goodison S, Viars C, Urquidi V. Molecular cytogenetic analysis of a human breast metastasis model: Identification of phenotype-specific chromosomal rearrangements. *Cancer Genet Cytogenet* 2005a;156:37–48. [PubMed: 15588854]
- Goodison S, Yuan J, Sloan D, Kim R, Li C, et al. The RhoGAP protein DLC-1 functions as a metastasis suppressor in breast cancer cells. *Cancer Res* 2005b;65:6042–6053. [PubMed: 16024604]
- Han W, Han MR, Kang JJ, Bae JY, Lee JH, et al. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* 2006;6:92. [PubMed: 16608533]
- Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, et al. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc Natl Acad Sci USA* 2003;100:2532–2537. [PubMed: 12610208]
- Hu Z, Fan C, Oh DS, Marron JS, He X, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96. [PubMed: 16643655]
- Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, et al. Gene expression predictors of breast cancer outcomes. *Lancet* 2003;361:1590–1596. [PubMed: 12747878]
- Hurst J, Maniar N, Tombarkiewicz J, Lucas F, Roberson C, et al. A novel model of a metastatic human breast tumour xenograft line. *Br J Cancer* 1993;68:274–276. [PubMed: 8394103]
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, et al. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 2002;62:6240–6245. [PubMed: 12414653]
- Isola JJ, Kallioniemi OP, Chu LW, Fuqua SA, Hilsenbeck SG, et al. Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. *Am J Pathol* 1995;147:905–911. [PubMed: 7573366]
- Janssen EA, Baak JP, Guervos MA, van Diest PJ, Jiwa M, Hermsen MA. In lymph node-negative invasive breast carcinomas, specific chromosomal aberrations are strongly associated with high mitotic activity and predict outcome more accurately than grade, tumour diameter, and oestrogen receptor. *J Pathol* 2003;201:555–561. [PubMed: 14648658]

- Jennings R, Alsarraj M, Wright KL, Munoz-Antonia T. Regulation of the human transforming growth factor beta type II receptor gene promoter by novel Sp1 sites. *Oncogene* 2001;20:6899–6909. [PubMed: 11687969]
- Ji C, Casinghino S, McCarthy TL, Centrella M. Multiple and essential Sp1 binding sites in the promoter for transforming growth factor-beta type I receptor. *J Biol Chem* 1997;272:21260–21267. [PubMed: 9261136]
- Jiang WG, Watkins G, Fodstad O, Douglas-Jones A, Mokbel K, Mansel RE. Differential expression of the CCN family members Cyr61, CTGF and Nov in human breast cancer. *Endocr Relat Cancer* 2004;11:781–791. [PubMed: 15613452]
- Kallioniemi A, Kallioniemi OP, Piper J, Tanner M, Stokke T, et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci USA* 1994;91:2156–2160. [PubMed: 8134364]
- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 2006;34:D354–357. [PubMed: 16381885]
- Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;3:537–549. [PubMed: 12842083]
- Kauraniemi P, Barlund M, Monni O, Kallioniemi A. New amplified and highly expressed genes discovered in the *ERBB2* amplicon in breast cancer by cDNA microarrays. *Cancer Res* 2001;61:8235–8240. [PubMed: 11719455]
- Kawai K, Viars C, Arden K, Tarin D, Urquidi V, Goodison S. Comprehensive karyotyping of the HT-29 colon adenocarcinoma cell line. *Genes Chromosomes Cancer* 2002;34:1–8. [PubMed: 11921276]
- Kluger HM, Chelouche Lev D, Kluger Y, McCarthy MM, Kiriakova G, et al. Using a xenograft model of human breast cancer metastasis to find genes associated with clinically aggressive disease. *Cancer Res* 2005;65:5578–5587. [PubMed: 15994930]
- Knutsen T, Gobu V, Knaus R, Padilla-Nash H, Augustus M, et al. The interactive online SKY/m-FISH & CGH database and the Entrez cancer chromosomes search database: Linkage of chromosomal aberrations with the genome sequence. *Genes Chromosomes Cancer* 2005;44:52–64. [PubMed: 15934046]
- Kreunin P, Urquidi V, Lubman DM, Goodison S. Identification of metastasis-associated proteins in a human tumor metastasis model using the mass-mapping technique. *Proteomics* 2004;4:2754–2765. [PubMed: 15352249]
- Kristensen VN, Sorlie T, Geisler J, Langerod A, Yoshimura N, et al. Gene expression profiling of breast cancer in relation to estrogen receptor status and estrogen-metabolizing enzymes: Clinical implications. *Clin Cancer Res* 2005;11:878s–883s. [PubMed: 15701881]
- Kytola S, Rummukainen J, Nordgren A, Karhu R, Farnebo F, et al. Chromosomal alterations in 15 breast cancer cell lines by comparative genomic hybridization and spectral karyotyping. *Genes Chromosomes Cancer* 2000;28:308–317. [PubMed: 10862037]
- Lapteva N, Yang AG, Sanders DE, Strube RW, Chen SY. CXCR4 knockdown by small interfering RNA abrogates breast tumor growth in vivo. *Cancer Gene Ther* 2005;12:84–89. [PubMed: 15472715]
- Lau LF, Lam SC. The CCN family of angiogenic regulators: The integrin connection. *Exp Cell Res* 1999;248:44–57. [PubMed: 10094812]
- Lelekakis M, Moseley JM, Martin TJ, Hards D, Williams E, et al. A novel orthotopic model of breast cancer metastasis to bone. *Clin Exp Metastasis* 1999;17:163–170. [PubMed: 10411109]
- Lev DC, Kiriakova G, Price JE. Selection of more aggressive variants of the gi101a human breast cancer cell line: A model for analyzing the metastatic phenotype of breast cancer. *Clin Exp Metastasis* 2003;20:515–523. [PubMed: 14598885]
- Li S, Huang S, Peng SB. Overexpression of G protein-coupled receptors in cancer cells: Involvement in tumor progression. *Int J Oncol* 2005;27:1329–1339. [PubMed: 16211229]
- Liang Z, Wu T, Lou H, Yu X, Taichman RS, et al. Inhibition of breast cancer metastasis by selective synthetic polypeptide against CXCR4. *Cancer Res* 2004;64:4302–4308. [PubMed: 15205345]
- Loi S, Desmedt C, Cardoso F, Piccart M, Sotiriou C. Breast cancer gene expression profiling: Clinical trial and practice implications. *Pharmacogenomics* 2005;6:49–58. [PubMed: 15723605]
- Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci USA* 2003;100:5974–5979. [PubMed: 12714683]

- Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004;5:607–616. [PubMed: 15193263]
- Mackay A, Jones C, Dexter T, Silva RL, Bulmer K, et al. cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene* 2003;22:2680–2688. [PubMed: 12730682]
- Macville M, Veldman T, Padilla-Nash H, Wangsa D, O'Brien P, et al. Spectral karyotyping, a 24-colour FISH technique for the identification of chromosomal rearrangements. *Histochem Cell Biol* 1997;108:299–305. [PubMed: 9387921]
- McGuire WL. Prognostic factors for recurrence and survival in human breast cancer. *Breast Cancer Res Treat* 1987;10:5–9. [PubMed: 3689982]
- Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005;436:518–524. [PubMed: 16049480]
- Morimura N, Tezuka Y, Watanabe N, Yasuda M, Miyatani S, et al. Molecular cloning of POEM: A novel adhesion molecule that interacts with alpha8beta1 integrin. *J Biol Chem* 2001;276:42172–42181. [PubMed: 11546798]
- Morris JS, Carter NP, Ferguson-Smith MA, Edwards PA. Cytogenetic analysis of three breast carcinoma cell lines using reverse chromosome painting. *Genes Chromosomes Cancer* 1997;20:120–139. [PubMed: 9331563]
- Muller A, Homey B, Soto H, Ge N, Catron D, et al. Involvement of chemokine receptors in breast cancer metastasis. *Nature* 2001;410:50–56. [PubMed: 11242036]
- Nicolini A, Carpi A, Tarro G. Biomolecular markers of breast cancer. *Front Biosci* 2006;11:1818–1843. [PubMed: 16368559]
- Nishidate T, Katagiri T, Lin ML, Mano Y, Miki Y, et al. Genome-wide gene-expression profiles of breast-cancer cells purified with laser micro-beam microdissection: Identification of genes associated with progression and metastasis. *Int J Oncol* 2004;25:797–819. [PubMed: 15375528]
- Nishizaki T, DeVries S, Chew K, Goodson WH 3rd, Ljung BM, et al. Genetic alterations in primary breast cancers and their metastases: Direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer* 1997;19:267–272. [PubMed: 9258662]
- Orimo A, Gupta PB, Sgroi DC, Arenzana-Seisdedos F, Delaunay T, et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* 2005;121:335–348. [PubMed: 15882617]
- Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* 1996;13:63–72. [PubMed: 8700555]
- Pawitan Y, Bjohle J, Amler L, Borg AL, Eghyazi S, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: Derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;7:R953–964. [PubMed: 16280042]
- Perbal B. CCN proteins: Multifunctional signalling regulators. *Lancet* 2004;363:62–64. [PubMed: 14723997]
- Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* 2006;8:R23. [PubMed: 16626501]
- Pittman J, Huang E, Dressman H, Horng CF, Cheng SH, et al. Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. *Proc Natl Acad Sci USA* 2004;101:8431–8436. [PubMed: 15152076]
- Popescu NC, Zimonjic DB. Chromosome and gene alterations in breast cancer as markers for diagnosis and prognosis as well as pathogenetic targets for therapy. *Am J Med Genet* 2002;115:142–149. [PubMed: 12407694]
- Price JE, Polyzos A, Zhang RD, Daniels LM. Tumorigenicity and metastasis of human breast carcinoma cell lines in nude mice. *Cancer Res* 1990;50:717–721. [PubMed: 2297709]
- Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54. [PubMed: 12469122]

- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* 2004;101:9309–9314. [PubMed: 15184677]
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM. Mining for regulatory programs in the cancer transcriptome. *Nat Genet* 2005;37:579–583. [PubMed: 15920519]
- Ried T, Just KE, Holtgreve-Grez H, du Manoir S, Speicher MR, et al. Comparative genomic hybridization of formalin-fixed, paraffin-embedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res* 1995;55:5415–5423. [PubMed: 7585611]
- Robbins P, Pinder S, de Klerk N, Dawkins H, Harvey J, et al. Histological grading of breast carcinomas: A study of interobserver agreement. *Hum Pathol* 1995;26:873–879. [PubMed: 7635449]
- Ryuto M, Ono M, Izumi H, Yoshida S, Weich HA, et al. Induction of vascular endothelial growth factor by tumor necrosis factor alpha in human glioma cells. Possible roles of Sp-1. *J Biol Chem* 1996;271:28220–28228. [PubMed: 8910439]
- Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. *Eur J Cancer* 2005;41:2438–2448. [PubMed: 16209919]
- Salvucci O, Bouchard A, Baccarelli A, Deschenes J, Sauter G, et al. The role of CXCR4 receptor expression in breast cancer: A large tissue microarray study. *Breast Cancer Res Treat* 2006;97:275–283. [PubMed: 16344916]
- Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, et al. Multicolor spectral karyotyping of human chromosomes. *Science* 1996;273:494–497. [PubMed: 8662537]
- Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res* 2006;66:5278–5286. [PubMed: 16707453]
- Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: Understanding cancer using microarrays. *Nat Genet* 2005;37 (Suppl):S38–45. [PubMed: 15920529]
- Seraj MJ, Harding MA, Gildea JJ, Welch DR, Theodorescu D. The relationship of *BRMS1* and *RHOGDI2* gene expression to metastatic potential in lineage related human bladder cancer cell lines. *Clin Exp Metastasis* 2000;18:519–525. [PubMed: 11592309]
- Shadeo A, Lam WL. Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res* 2006;8:R9. [PubMed: 16417655]
- Shen R, Ghosh D, Chinnaiyan AM. Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* 2004;5:94. [PubMed: 15598354]
- Shevde LA, Welch DR. Metastasis suppressor pathways—an evolving paradigm. *Cancer Lett* 2003;198:1–20. [PubMed: 12893425]
- Shimo T, Kubota S, Yoshioka N, Ibaragi S, Isowa S, et al. Pathogenic role of connective tissue growth factor (CTGF/CCN2) in osteolytic metastasis of breast cancer. *J Bone Miner Res* 2006;21:1045–1059. [PubMed: 16813525]
- Smith MC, Luker KE, Garbow JR, Prior JL, Jackson E, et al. CXCR4 regulates growth of both primary and metastatic breast cancer. *Cancer Res* 2004;64:8604–8612. [PubMed: 15574767]
- Sokolova IA, Halling KC, Jenkins RB, Burkhardt HM, Meyer RG, et al. The development of a multitarget, multicolor fluorescence in situ hybridization assay for the detection of urothelial carcinoma in urine. *J Mol Diagn* 2000;2:116–123. [PubMed: 11229514]
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001;98:10869–10874. [PubMed: 11553815]
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100:8418–8423. [PubMed: 12829800]
- Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: Gene expression analyses across three different platforms. *BMC Genomics* 2006;7:127. [PubMed: 16729877]

- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, et al. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–272. [PubMed: 16478745]
- Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multicolor FISH. *Nat Genet* 1996;12:368–375. [PubMed: 8630489]
- Stacker SA, Achen MG, Jussila L, Baldwin ME, Alitalo K. Lymphangiogenesis and cancer metastasis. *Nat Rev Cancer* 2002;2:573–583. [PubMed: 12154350]
- Stuart RO, Wachsman W, Berry CC, Wang-Rodriguez J, Wasserman L, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc Natl Acad Sci USA* 2004;101:615–620. [PubMed: 14722351]
- Sugino T, Yamaguchi T, Ogura G, Saito A, Hashimoto T, et al. Morphological evidence for an invasion-independent metastasis pathway exists in multiple human cancers. *BMC Med* 2004;2:9. [PubMed: 15066199]
- Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* 2007;23:30–37. [PubMed: 17130137]
- Tanner MM, Tirkkonen M, Kallioniemi A, Isola J, Kuukasjarvi T, et al. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res* 1996;56:3441–3445. [PubMed: 8758909]
- Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* 1998;21:177–184. [PubMed: 9523192]
- Tuma RS. A big trial for a new technology: Trans-BIG project takes microarrays into clinical trials. *J Natl Cancer Inst* 2004;96:648–649. [PubMed: 15126595]
- Urquidi V, Sloan D, Kawai K, Agarwal D, Woodman AC, et al. Contrasting expression of thrombospondin-1 and osteopontin correlates with absence or presence of metastatic phenotype in an isogenic model of spontaneous human breast cancer metastasis. *Clin Cancer Res* 2002;8:61–74. [PubMed: 11801541]
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536. [PubMed: 11823860]
- van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AA, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009. [PubMed: 12490681]
- Villeneuve DJ, Hembruff SL, Veitch Z, Cecchetto M, Dew WA, Parissenti AM. cDNA microarray analysis of isogenic paclitaxel- and doxorubicin-resistant breast tumor cell lines reveals distinct drug-specific genetic signatures of resistance. *Breast Cancer Res Treat* 2006;96:17–39. [PubMed: 16322897]
- von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, et al. STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 2005;33:D433–437. [PubMed: 15608232]
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671–679. [PubMed: 15721472]
- Wang-Rodriguez J, Urquidi V, Rivard A, Goodison S. Elevated osteopontin and thrombospondin expression identifies malignant human breast carcinoma but is not indicative of metastatic status. *Breast Cancer Res* 2003;5:R136–143. [PubMed: 12927044]
- Watson MB, Bahia H, Ashman JN, Berrieman HK, Drew P, et al. Chromosomal alterations in breast cancer revealed by multicolour fluorescence in situ hybridization. *Int J Oncol* 2004;25:277–283. [PubMed: 15254723]
- Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van 't Veer LJ. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci USA* 2003;100:15901–15905. [PubMed: 14665696]
- Weigelt B, Hu Z, He X, Livasy C, Carey LA, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* 2005a;65:9155–9158. [PubMed: 16230372]

- Weigelt B, Wessels LF, Bosma AJ, Glas AM, Nuyten DS, et al. No common denominator for breast cancer lymph node metastasis. *Br J Cancer* 2005b;93:924–932. [PubMed: 16189523]
- Wong CM, Lee JM, Ching YP, Jin DY, Ng IO. Genetic and epigenetic alterations of *DLC-1* gene in hepatocellular carcinoma. *Cancer Res* 2003;63:7646–7651. [PubMed: 14633684]
- Xie D, Jauch A, Miller CW, Bartram CR, Koeffler HP. Discovery of over-expressed genes and genetic alterations in breast cancer cells using a combination of suppression subtractive hybridization, multiplex fish and comparative genomic hybridization. *Int J Oncol* 2002;21:499–507. [PubMed: 12168092]
- Yang F, Foekens JA, Yu J, Sieuwerts AM, Timmermans M, et al. Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* 2006;25:1413–1419. [PubMed: 16261164]
- Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, et al. Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res* 2006;66:4065–4078. [PubMed: 16618726]
- Zannetti A, Del Vecchio S, Carriero MV, Fonti R, Franco P, et al. Coordinate up-regulation of Sp1 DNA-binding activity and urokinase receptor expression in breast carcinoma. *Cancer Res* 2000;60:1546–1551. [PubMed: 10749121]

Table 1 Genes common to at least two breast cancer metastasis signatures amongst five animal models and four tissue-based expression microarray studies

Gene name	Gene symbol	Chromosomal location	Entrez gene ^d	Subcell. localiz. ^b	Animal models					Human tumors									
					MDA-231-BONE					Human tumors									
					A	B	C	D	E	F	G	H	I						
Amnixin A2	ANXA2	15q21-q22	302	1															
Clusterin	CLU	8p21-p12	1191	1															
Collagen, type I, alpha 1	COL1A1	17q21.33	1277	1															
Collagen, type I, alpha 2	COL1A2	7q22.1	1278	1															
Collagen, type V, alpha 1	COL5A1	9q34.2-q34.3	1289	1															
Collagen, type VI, alpha 1	COL6A1	21q22.3	1291	1															
Connective tissue growth factor	CTGF	6q23.1	1490	1															
Fibronectin 1	FN1	2q34	2335	1															
Chemokine (C-X-C motif) ligand 1	CXCL1	4q21	2919	1															
Cysteine-rich, angiogenic inducer, 61	CYR61	1p31-p22	3491	1															
Laminin, beta 1	LAMB1	7q22	3912	1															
Lysyl oxidase-like 2	LOXL2	8p21.3-p21.2	4017	1															
Latent transforming growth factor beta binding protein 1	LTBP1	2p22-p21	4052	1															
Matrix Gla protein	MGP	12p13.1-p12.3	4256	1															
Matrix metalloproteinase 1 (interstitial collagenase)	MMP1	11q22.3	4312	1															
Serpine peptidase inhibitor	SERPINE2	2q33-q35	5270	1															
Amnixin A1	ANXA1	9q12-q21.2	301	2															
Major histocompatibility complex, class II, DP alpha 1	HLA-DPA1	6p21.3	3113	2															
Major histocompatibility complex, class II, DP beta 1	HLA-DPBI	6p21.3	3115	2															
Major histocompatibility complex, class II, DR beta 1	HLA-DRBI	6p21.3	3123	2															
Lymphocyte antigen 6 complex, locus E	LY6E	8q24.3	4061	2															
Periplakin	PPL	16p13.3	5493	2															
PTK7 protein tyrosine kinase 7	PTK7	6p21.1-p12.2	5754	2															
Chemokine (C-X-C motif) receptor 4	CXCR4	2q21	7852	2															
ATP-binding cassette, sub-family C (CFTR/MRP), member 3	ABCC3	17q22	8714	2															
Tumor necrosis factor (ligand) superfamily, member 10	TNFSF10	3q26	8743	2															
G protein-coupled receptor 56	GPR56	16q12.2-q21	9289	2															
Epoxide hydrolase 1, microsomal (xenobiotic)	EPHX1	1q42.1	2052	3															
Rho GDP dissociation inhibitor (GDI), beta	ARHGDI3	12p12.3	397	4															
Keratin, hair, basic, 1	KRT86	12q13	3887	4															
Tropomyosin 2 (beta)	TPM2	9p13.2-p13.1	7169	4															
Deleted in liver cancer 1	DLC1	8p22	10395	4															
Crystallin, alpha B	CRYAB	11q22.3-q23.1	1410	5															
Inhibitor of DNA binding 1	IDI	20q11	3397	5															
SRY (sex determining region Y)-box 4	SOX4	6p22.3	6659	5															
Histone 1, H2ac	HIST1H2AC	6p21.3	8334	5															

Gene name	Gene symbol	Chromosomal location	Entrez gene ^d	Subcell. localiz. ^b	Animal models						I								
					A	B	C	D	E	F		G	H						
Protein regulator of cytokinesis 1	<i>PRC1</i>	15q26.1	9055	5															
Cyclin E2	<i>CCNE2</i>	8q22.1	9134	5															
Aryl-hydrocarbon receptor nuclear translocator 2	<i>ARNT2</i>	15q24	9915	5															
Nucleolar and spindle associated protein 1	<i>NUSAP1</i>	15q15.1	51203	5															
G-2 and S-phase expressed 1	<i>GTSE1</i>	22q13.2-q13.3	51512	5															
Denticleless homolog (Drosophila)	<i>DTL</i>	1q32.1-32.2	51514	5															
SPANX family, members A1, A2, B1, B2, C	<i>SPANX</i>	Xq27.1	****	5															
MLF1 interacting protein	<i>MLF1IP</i>	4q35.1	79682	5															
Chromosome 10 open reading frame 116	<i>C10orf116</i>	10q23.2	10974	6															
Chromosome 14 open reading frame 139	<i>C14orf139</i>	14q32.13	79686	6															
Annexin A2 pseudogene 1	<i>ANXA2P1</i>	4q21-q31	303	NA															

Studies in References: A: Kang et al. (2003); B: Eckhardt et al. (2005); C: Goodison et al. (2005); D: Kluger et al. (2005); E: Minn et al. (2005); F: Wang et al. (2005); G: Ramaswamy et al. (2002); H: Pawitan et al. (2005); I: Van 't Veer et al. (2002).

**** = Various SPANX family members, Entrez Gene IDs: 30014; 64649; 64650; 64663; 64694.

Subcellular localization of gene products, 1: Extracellular space; 2: Plasma membrane; 3: Other membrane; 4: Cytoplasm; 5: Nucleus; 6: Other.

Homo sapiens homologous genes were considered for comparison (HomoloGene release 50.1).