LUND UNIVERSITY

Transcriptional profiling of breast cancer metastases identifies liver metastasis-selective genes associated with adverse outcome in luminal A primary breast cancer.

Kimbung, Siker; Johansson, Ida; Danielsson, Anna; Veerla, Srinivas; Brage, Suzanne Egyhazi; Frostvik Stolt, Marianne; Skoog, Lambert; Carlsson, Lena; Einbeigi, Zakaria; Lidbrink, Elisabet; Linderholm, Barbro Kristina; Loman, Niklas; Malmström, Per; Söderberg, Martin; Walz, Thomas M; Fernö, Mårten; Hatschek, Thomas; Hedenfalk, Ingrid

Link to publication

Transcriptional profiling of breast cancer metastases identifies liver metastasis-selective genes associated with adverse outcome in luminal A primary breast cancer

Siker Kimbung[1,2], Ida Johansson[1,2], Anna Danielsson[3], Srinivas Veerla[1,2], Suzanne Egyhazi Brage[4], Marianne Frostvik Stolt[4], Lambert Skoog[4], Lena Carlsson[5], Zakaria Einbeigi[3], Elisabet Lidbrink[4], Barbro Linderholm[3], Niklas Loman[6], Per-Olof Malmström[6], Martin Söderberg[6], Thomas M. Walz[7], Mårten Fernö[1], Thomas Hatschek[4] and Ingrid Hedenfalk[1,2] in collaboration with the TEX study group.

[1]Division of Oncology and Pathology, Department of Clinical Sciences, Lund, Lund University, Sweden.

[2]CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden.

[3]Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

[4]Department of Oncology and Pathology, Karolinska Institutet and Karolinska University Hospital, Sweden.

[5]Department of Oncology, Sundsvall Hospital, Sundsvall, Sweden

[6]Department of Oncology, Skåne University Hospital, Lund/Malmö, Sweden.

[7]Division of Oncology, Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Linköping, Sweden.

Corresponding author: Associate Professor Ingrid Hedenfalk, Division of Oncology and Pathology, Department of Clinical Sciences, Lund, Lund University, Medicon Village, SE-22381 Lund, Sweden. Phone: +46-46-2220652. Fax: +46-46-147327. E-mail: Ingrid.Hedenfalk@med.lu.se

Running title: Liver metastasis genes can predict breast cancer survival.

Translational relevance

Although metastasis is the principal cause of cancer-related deaths, the scarcity of clinical breast cancer metastases has impeded their characterization in large genomic and transcriptomic studies. While metastases may be genetically similar to their seeding primary tumors, distinct differences which could be exploited to improve disease control may nonetheless exist. We performed global transcriptional profiling of 91 clinical breast cancer metastases, aiming to identify genes associated with liver metastases, given the inferior outcome associated with liver recurrence. We identified a set of 17 liver metastasis-selective genes of prognostic relevance in early breast cancer. Importantly, this signature showed an independent ability of identifying patients at higher risk of recurrence and death within the luminal A molecular subtype. These patients may benefit from closer disease monitoring and may in addition be amenable to enrollment into clinical trials investigating novel anti-neoplastic therapeutics targeting features other than increased proliferation.

Abstract

Purpose: The complete molecular basis of the organ-specificity of metastasis is elusive. This study aimed to provide an independent characterization of the transcriptional landscape of breast cancer metastases with the specific objective to identify liver metastasis-selective genes of prognostic importance following primary tumor diagnosis.

Experimental design: A cohort of 304 women with advanced breast cancer was studied. Associations between the site of recurrence and clinico-pathological features were investigated. Fine-needle aspirates of metastases (n=91) were subjected to whole genome transcriptional profiling. Liver metastasis-selective genes were identified by significance analysis of microarray (SAM) analyses and independently validated in external datasets. Finally, the prognostic relevance of the liver metastasis-selective genes in primary breast cancer was tested.

Results: Liver relapse was associated with estrogen receptor (ER) expression (P=0.002), luminal B subtype (P=0.01), and was prognostic for an inferior post-relapse survival (P=0.01). The major variation in the transcriptional landscape of metastases was also associated with ER expression and molecular subtype. However, liver metastases displayed unique transcriptional fingerprints, characterized by down-regulation of extracellular matrix (*i.e.* stromal) genes. Importantly, we identified a 17-gene liver metastasis-selective signature, which was significantly and independently prognostic for shorter relapse-free (P<0.001) and overall (P=0.001) survival in ER positive tumors. Remarkably, this signature remained independently prognostic for shorter relapse-free survival (P=0.001) among luminal A tumors.

3

Conclusions: Extracellular matrix (stromal) genes can be used to partition breast cancer by site of relapse and may be used to further refine prognostication in ER positive primary breast cancer.

Keywords: breast cancer metastasis, transcriptional profiling, liver metastasis-selective genes, stroma, luminal A, prognosis

Introduction

Metastasis is a significant clinical and socio-economic problem, accounting for over 90% of cancer-related deaths (1). After diagnosing metastatic breast cancer (MBC), the site of recurrence is an important feature for estimating the patient's prognosis. Liver metastasis is associated with the poorest survival relative to loco-regional, bone and lung colonization (2-7). Noteworthy, the diagnosis of liver metastases is on the rise (8, 9), suggesting that available adjuvant therapies may have limited efficacy in preventing liver colonization compared to metastases at other sites. Consequently, the increasing numbers of patients presenting with these adverse events warrants a better understanding of the molecular attributes of site-specific metastases to enable the identification of novel biomarkers to guide surveillance and improve personalization of therapy.

The selection of metastatic sites is not a random process. Once disseminated, circulating tumor cells exhibit tissue specific tropisms beyond what can be explained by normal circulatory patterns. Tissue selectivity for breast cancer metastatic colonization has been associated with primary tumor pathological characteristics such as estrogen receptor (ER) expression and tumor molecular subtypes (10, 11). However, a marked redundancy of

metastatic site selectivity prevails between these molecularly heterogeneous groups, limiting their accuracy as site-specific predictive markers.

Conventionally, at time of primary breast cancer diagnosis, the prognosis for a favorable outcome and decision for the exemption from chemotherapy is based on a combination of factors including ER positivity, negative nodal status, small tumor size and low histological grade (4). Tumors displaying these favorable prognostic factors are significantly enriched within the luminal A intrinsic subtype. However, intrinsic or acquired resistance to hormonal therapy and disease recurrence to distant sites, including the liver, may eventually occur in a clinically relevant number of patients with luminal A tumors, underlining the heterogeneity even within this favorable subtype. Metastases remain the main cause of breast cancer-related mortality. It is therefore necessary to identify better prognostic biomarkers, and if possible subtype-specific prognostic biomarkers to improve individualization of therapy.

A few studies have shown that primary tumors and their metastases generally share similar copy number aberrations (12, 13) and gene expression profiles (14, 15), but these studies were under-powered by the scarcity of metastatic biopsies, limiting the identification of differences between these matched tumor pairs. By utilizing experimental mouse models and a limited series of clinical metastatic biopsies, genes associated with the propensity of breast cancer relapse to the bone (16), lung (17, 18), and brain (19) have been published. Furthermore we (7) and others (20, 21) have shown an association between claudin-2 expression and liver recurrence. However, because experimental mouse models incompletely capture the relevant genetic complexity of tumor progression within the human host, studies using patient-derived biopsies from metastases may reveal additional clinically relevant site-specific attributes to complement and/or validate these preliminary reports.

The aim of this study was to provide an independent characterization of the transcriptional landscape of breast cancer metastases with the specific objective to identify genes selective for breast cancer liver metastases with prognostic potential at time of primary tumor diagnosis.

Materials and Methods

Patients and tumors

The study cohort consisted of 304 women diagnosed with locally advanced (inoperable) or MBC, enrolled in a randomized phase III trial (TEX) conducted between 2002 and 2007 in Sweden. As first line treatment for metastatic disease, patients received a combination of epirubicin and paclitaxel alone (ET) or with the addition of capecitabine (TEX). Patients presenting with brain metastases, approved for first-line HER2-targeted therapy, or diagnosed with other malignancies within five years of the trial commencement were exempted. Complete clinical and pathological data were recorded in a central clinical trial database. The median follow-up for post-recurrence survival was 45 months (range 9-135 months) for patients alive at last update (July 2013). Detailed information regarding the design and outcome of the trial has been published (22). Fine-needle aspirates (FNA) of at least one metastatic lesion were collected before commencement of treatment whenever possible. In addition, archival formalin-fixed paraffin-embedded primary tumor blocks were collected for tissue microarray (TMA) construction and central re-assessment of biomarkers by immunohistochemistry and *in situ* hybridization techniques where applicable. Table 1 and Supplementary Table 1 show the distribution of clinico-pathological factors in the cohort.

Ethics Statement

This sub-study was approved by all the regional ethics committees at the participating hospitals [Karolinska Institutet Stockholm (KI 02-205 & 02-206); Sahlgrenska University Hospital Gothenburg (M090-02 & M091-02); Linköping University Hospital (02-519 & 02-339); Örebro University Hospital (308/02 & 308/03); Umeå University Hospital (Um 02-336 & Um 03-03) and Lund University Hospital (LU 290-02 & LU 291-02)]. All patients provided written informed consent to participate in the clinical trial and translational studies. This study adheres to the REMARK guidelines for reporting prognostic biomarker studies (23).

RNA extraction and gene expression microarrays

Tumor cellularity of FNAs was assessed by a cytologist (LS) on Giemsa stained, ethanol-fixed, cytospin preparations and total RNA was extracted from samples with high (>50%) tumor cell content using Qiagen RNA Mini kit (Qiagen, Valencia, CA) following the manufacturer's recommendations. RNA quantity and integrity were analyzed on the NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) respectively, and cDNA was generated and biotin-labelled using the NuGen 50ng amplification protocol (Covance Genomics Laboratory, Princeton, NJ). Labelled cDNA was hybridized onto custom-made whole genome Affymetrix HuRSTA-2a520709 gene chips following the GeneChip Hybridization, Wash, and Stain Kit protocol (Affymetrix, Santa Clara, CA). Data pre-processing and normalization were performed using the robust multichip average (RMA) algorithm. After normalization, a presence filter was applied to select only features present in ≥90% of assays, and features with low intensities (below the median intensity for Y chromosome gene probes) were filtered out. The data were log2 transformed and only transcripts showing high variance across assays were selected (variance filter sd±1), leaving a final dataset with 8,339 features representing 5,232 unique gene variants for further analyses. All processes were performed using packages

in R (24) and the TM4 microarray software suite (25). The final dataset included 91 samples from 85 patients [liver (n=16), bone (n=5), lung (n=2), lymph node (n=39), local [breast (n=11) and skin (n=17), and ascites (n=1)]. The distribution of baseline clinico-pathological features in the original study cohort (n=304) and the subpopulation included in the transcriptional profiling study (n=85) is presented in Supplementary Table 1. Raw and processed data have been deposited in the Gene Expression Omnibus (GSE46141).

Multi-variable data analyses

Unsupervised analyses

Principal component analysis (PCA) was performed using SIMCA P version 13.0.2 software package (Umetrics AB, Umeå, Sweden). The dataset was mean-centered across rows (genes), unit variance scaled and model complexity was estimated by leave-one-out cross-validation. Unsupervised hierarchical clustering (HCL) was performed using the Pearson correlation distance metric and average linkage.

Supervised analyses

The intrinsic molecular subtypes of the metastases were determined using the research-based PAM50 algorithm as previously described (26). A two-class significance analysis of microarray (SAM; (27)) analysis was performed to identify significant differentially expressed genes in liver metastases compared to metastases from other sites. The liver-selectivity of the identified genes was verified in an external dataset of 36 breast cancer metastases (GSE14018, (28)). The biological processes and pathways enriched among the liver metastases-selective genes were uncovered by gene ontology analysis using the DAVID (29, 30) database. Furthermore, the activity of eight gene expression-based modules representing relevant breast cancer-specific biological processes (stroma, lipid, immune

response, mitotic progression, mitotic checkpoint, basal, early response and steroid response; (31) was assessed in the metastases.

In a final step, candidate liver metastasis-selective genes which may serve as biomarkers for predicting the liver metastatic potential of a primary tumor were identified using an external dataset of 192 primary breast tumors (GSE12276) (15) and associations between these candidate genes and outcome in early breast cancer were independently tested using Gene expression-based Outcome for Breast cancer Online (GOBO; (32)), an online tool for validation of the prognostic value of single genes or sets of genes in primary breast cancer (n=1,881).

Survival analyses

Kaplan-Meier plots were generated and the log-rank test was used to check for statistically significant differences between target groups. Cox-proportional hazards models were used to evaluate the independent prognostic significance of biomarkers, adjusting for conventional prognostic factors. *P*-values correspond to two-sided statistical tests and values <0.05 were considered significant.

Results

Associations between primary tumor clinico-pathological factors and the first site(s) of recurrence

Because many patients with metastatic breast cancer present with relapses in more than one anatomical site at time of first metastasis diagnosis, we classified patients into four metastatic categories reflecting the most advanced site affected at first clinical presentation. These categories were: loco-regional (locally advanced or regional metastases in the lymph nodes or skin), bone (skeletal metastases with or without loco-regional disease), lung (lung parenchymal/plural metastases with or without bone and loco-regional disease), and liver

(hepatic metastases with or without lung, bone or loco-regional metastases). Associations between primary tumor clinico-pathological factors and the first site(s) of recurrence are shown in Table 1. ER positivity was found to be associated with bone and liver recurrences, while negative ER status correlated with loco-regional and lung relapses (Fisher's exact P=0.002). Liver recurrence was also common among patients with HER2 positive tumors (8/17), but this association was not statistically significant (probably due to the limited number of HER2 positive tumors in the study). Loco-regional and bone metastases were often detected as oligo-metastases, while liver and lung metastases were often diagnosed in parallel with deposits at other sites (P<0.001). Furthermore, low histological grade (grades 1 and 2) was associated with bone and liver recurrences, while high grade (grade 3) correlated with loco-regional and lung relapses (P=0.03). However, no significant association between histological grade and metastatic site was observed when ER positive tumors were analyzed separately (P=0.58). When the surrogate (IHC-based) molecular subtype of the primary tumor was considered, bone and hepatic recurrences were found to be associated with luminal-like (A and B) tumors, while relapses to the lung and loco-regional sites were associated with the triple-negative subtype (P=0.01). Sub-analyses within ER positive tumors revealed a borderline association of bone metastases with the luminal A-like and liver metastases with the luminal B-like subtypes, respectively (P=0.05). Overall, these results confirm that conventional tumor pathological biomarkers provide important insights into a primary tumor's metastatic propensity, with liver relapse commonly associated with poor prognostic pathological features. Nevertheless, in this cohort, a remarkably high prevalence of liver metastases was noted among patients who presented with primary tumors with favorable prognostic features; 40% of luminal A and 53% of histological grade 1 and 2 tumors progressed to the liver.

Liver-only relapse is associated with a relatively better outcome

We recently reported that liver relapse was associated with inferior survival after recurrence in the present study cohort (7). However, some studies suggest that patients with liver-only metastatic disease may experience longer survival compared to patients harboring liver metastases in parallel with metastases in other organs (5, 9). We found a similar trend in this cohort [Figure 1, log-rank P=0.01, Multivariable Cox model P<0.001 adjusting for age (>50 years or ≤50 years), metastasis-free interval (≤2 years or >2 years), nodal status, adjuvant endocrine therapy, and adjuvant chemotherapy], emphasizing the significance of tumor burden in addition to metastatic site for post relapse survival.

Identification of shared and distinct transcriptional portraits of site-specific metastases

The transcriptional landscape of breast cancer metastases has generally been inferred from primary tumors due to scarcity of clinical biopsies from metastases to perform independent studies. PCA analyses revealed that the first three principal components partitioned breast cancer metastases into groups which were strongly associated with primary tumor ER expression (Figure 2A) and the intrinsic molecular subtypes of the metastases (Figure 2B). Remarkably, liver metastases were the only class that was tightly clustered in the PCA score plot (Figure 2C), indicating a transcriptional distinction relative to other metastases. A similar tight clustering pattern for liver metastases was observed by unsupervised hierarchical clustering of the samples using the top 3,000 most variable probes (Figure 2D). Of note, all biological replicates (independent metastatic biopsies from the same patient) clustered together pair-wise and adjacent to each other in the sample dendrogram, confirming that transcriptional profiles of intra-individual tumors are more similar than inter-individual profiles.

11

To provide more insight into the biology of site-specific metastases, the activity of eight gene modules representing key biological aspects associated with breast cancer (31) were compared between the specific metastatic sites. Four modules were found to be significantly differentially expressed between the metastatic sites (Figure 3A-D). Liver metastases displayed a significantly lower expression of the 'stroma' (relative to the skin and lymph nodes, adjusted P=0.01 and P=0.001, respectively); 'basal' (relative to skin, P=0.043) and 'early response' (relative to bone, P=0.005) modules, and a higher expression of the 'steroid response' module relative to metastases in the skin (P=0.003). Considering that the transcriptional profiles of independent tumors from the same individual are highly similar, the activities of the eight gene modules were next compared between samples classified according to the four metastatic categories as previously defined in Table 1. Similarly, differential expression of the same four modules was observed (Figure 3E-H). Low expression of the 'stroma' module was observed in the liver category relative to bone (adjusted P=0.015). In addition, the 'basal' module was elevated in the lung category relative to the liver (P=0.018), while 'steroid response' was higher in the liver and bone categories relative to the lung category (P=0.013 and P=0.017, respectively). These results further confirm the association between the metastatic site and ER expression or molecular subtype since the 'basal' and 'steroid response' modules were shown to be strongly associated with the basal-like (ER-) and luminal subtypes (ER+), respectively (31).

Identification of liver metastasis-selective genes

To further dissect the transcriptional distinctiveness of liver metastases, a two-class SAM analysis was performed comparing liver *vs*. other metastases. This analysis was restricted to ER positive primary tumors, since the liver metastases were mainly of this phenotype (12/16; with 3 missing ER status, Figure 2). We found 358 genes to be differentially expressed (309 up-regulated and 49 down-regulated, FDR=0.1; Supplementary Table 2); henceforth referred

to as 'breast cancer liver metastasis-selective genes'. HCL of an independent set of 36 breast cancer metastases (GSE14018) (28) using only these 358 genes revealed a similar expression pattern in liver metastases (Supplementary Figure 1), thus confirming their liver selectivity.

Gene set enrichment analysis (29, 30) showed significant up-regulation of biological processes including endopeptidase inhibitor activity, complement activation, blood coagulation, immune response and steroid metabolism in liver metastases (Supplementary Table 3). Conversely, processes associated with extracellular matrix, biological adhesion, skeletal system development, and blood vessel development were enriched among the down-regulated genes in liver metastases (Supplementary Table 3). To ascertain that the enriched up-regulated biological processes, which are also common biological processes occurring in normal liver, was not a reflection of normal tissue contamination, we performed unsupervised HCL of all samples in the test cohort using previously reported normal breast and liver tissue-specific genes (33), as well as breast cancer-selective genes (34), respectively. Reassuringly, even though the liver metastases formed a distinct cluster in the sample dendrogram when clustered using the normal liver genes (Supplementary Figure 2A), no separation of the samples based on metastatic site was seen upon clustering with normal breast (Supplementary Figure 2B) or breast cancer-specific genes (Supplementary Figure 2C). Instead, clustering correlated with other biological characteristics, such as ER expression and molecular subtype. These results suggest that breast cancer liver metastases maintain a transcriptional profile consistent with the site of origin of the tumor cells (breast), and in addition adopt other transcriptional features associated with the metastatic microenvironment (liver) which may be important for their survival at this foreign site.

Associations between breast cancer liver metastasis-selective genes and primary tumor clinico-biological factors and clinical outcome

Robust tissue-specific metastasis biomarkers may be detectable in the primary tumors of patients who eventually develop metastases in the corresponding target organ. Gene signatures with the potential to predict breast cancer metastasis to the lung, bone and brain (16, 17, 19) have been reported. Using an external primary breast cancer dataset including only patients with metastatic disease and for whom the annotation of the site(s) of metastasis was recorded (15, 19), we performed a restricted analysis using only the 358 liver metastasis-selective genes. Only ER positive tumors (n=119) were interrogated. 347 of the 358 genes could be mapped across datasets. We found 17 genes to be significantly (FDR<0.05) differentially over-expressed in tumors relapsing in the liver. This list was enriched for genes involved in cadherin and integrin signaling pathways, as well as in skeletal system development. Of note, 6/17 (*CDH11, COL11A1, FBN1, MFAP5, SFRP4, SPON1*) genes overlapped with the previously described 'stroma' module (Spearman correlation coefficient 0.7). Figure 4 shows the expression of these 17 genes in both the test and the validation cohorts. Surprisingly, while all 17 genes were up-regulated in primary tumors with liver metastatic potential, 14/17 genes were down-regulated in liver metastases in both the test and the validation (GSE14018) metastasis datasets.

Finally, in GOBO (32), a database containing 1,881 annotated primary breast tumors, we aimed to identify relevant associations between the 17-gene signature and other primary tumor pathological features and prognosis. The expression of the signature was heterogeneous between different molecular subtypes and histological grades (Figures 5A-D, Anova *P*<0.00001), with a significantly lower expression in luminal B and basal-like tumors compared to the other subtypes (adjusted *P*<0.0001 for all pairwise comparisons of luminal B or basal tumors *vs*. other subtypes). In addition, low expression was significantly correlated

with high histological grade (Figure 5C-D, adjusted $P<0.0001$ for pairwise comparisons between grade 3 tumors vs. grade 1 and 2). Exploratory analyses revealed significant differential expression of the 17-gene signature across the recently described IntClust subgroups (35, 36) (Supplementary Figure 3A). Decreased expression was observed in IntClust subgroups 10, 1, and 9 relative to IntClust 3 and 4 (adjusted $P<0.0001$ for all pairwise comparisons). Furthermore, among ER positive tumors, a significantly lower expression was also noted in subgroups 7 and 8 relative to subgroups 3 and 4 ($P<0.0001$).

Remarkably, low expression of the 17-gene signature was significantly associated with shorter recurrence-free survival (RFS; Figure 5E, log-rank P= $3x10^{-5}$; Supplementary Table 4, multivariable Cox model HR=1.5, $P=0.001$) and overall survival (OS; Figure 5G, log-rank P=0.00927; Supplementary Table 5, multivariable Cox model HR=1.4, $P=0.026$) in patients with ER positive tumors. More importantly, the 17-gene signature remained significantly and independently prognostic for RFS when the subset of luminal A tumors was analyzed separately (Figure 5F, log-rank $P=0.00097$; Supplementary Table 5, multivariable Cox model HR=2.2, $P=0.004$). A trend toward poor OS for patients with luminal A tumors with low expression of the 17-gene signature was observed in univariable analysis (Figure 5H, log-rank $P=0.083$; Supplementary Table 4, multivariable Cox model HR=1.4, $P=0.29$). In sub analyses restricted to tumors in IntClust subgroups 3, 7 and 8 (those highly enriched for luminal A tumors), low expression of the 17-gene signature was associated with an inferior RFS (Supplementary Figure 3B, Log-rank $P=0.001$) and OS (Log-rank $P=0.06$). Exploratory analyses confirmed the association of the signature with poor prognosis when all tumors were included in the analysis, irrespective of ER status (Supplementary Figure 3C-D, RFS $P=0.00012$, OS $P=0.01872$).

Discussion

In this study we identified a 17-gene signature enriched for extracellular matrix or stroma genes, the majority of which were selectively down-regulated in breast cancer liver metastases. Furthermore, down-regulation in primary tumors, irrespective of site of relapse, was associated with aggressive tumor biological features and inferior prognosis. Liver metastases are deleterious, leading to the early demise of MBC patients (2-6, 37). We observed significant positive associations between liver recurrence and poor tumor biological characteristics, including luminal B subtype, high histological grade and large tumor burden. However, despite the statistically significant associations, the prevalence of liver relapses was notable in all subgroups, indicating a low specificity and sensitivity of these factors for accurate metastatic site prediction. Since liver relapse is indicative of inferior post-recurrence survival, there is a need for more specific and independent biomarkers to identify patients at risk. Recently, we demonstrated that CLDN2, which is significantly up-regulated in liver metastases, is an independent prognostic factor for early liver recurrence in breast cancer (37). Here, we show that down-regulation of various genes involved in cell adhesion is characteristic of liver metastases.

Patients with liver-only metastatic disease had a better post-recurrence survival compared to those harboring liver metastases in parallel to metastases in other organs. This finding corroborates results from other studies (5, 9). There is great interest in evaluating local treatment options such as surgery or stereotactic radiotherapy in patients with oligo-metastases in the liver but randomized studies are needed to evaluate the efficacy of these treatment options.

Transcriptional profiling has increased our understanding of the biology of organ-specific metastases and has led to the identification of site-specific metastasis genes and signatures

(16, 17, 19, 20). PCA and unsupervised HCL analyses reported herein revealed that the major variation across breast cancer metastases was strongly associated with ER status and molecular subtype, an observation consistent with the conventional understanding of breast cancer biology. This similarity underscores that primary tumor molecular traits are conserved across stages of tumor progression. Interestingly, we observed minor but significant site-specific differences at the transcriptional level, which reflects additional alterations acquired by breast cancer cells to thrive and evolve into overt metastases in the foreign milieu. Interestingly, our data suggest that mimicry of 'normal processes' of the new microenvironment may be a necessary adaptation. An enrichment of genes and biological processes commonly observed in normal liver was noted among up-regulated genes. Most of these genes code for signaling peptides commonly found in the extracellular space, further highlighting the importance of the microenvironment in metastatic colonization. The deregulation of genes which mimic target organ functions has previously been observed in other studies investigating the organ-specificity of metastases. Differential expression of genes important for ossification in bone metastases (16, 38), brain metabolism in brain metastases (19), pulmonary function in lungs (17, 18) and liver function in liver metastases (20) have been reported. This phenomenon can be interpreted within the confinements of the "seed and soil theory" of tumor invasion and metastatic colonization (39). Of note, mimicry of target-organ properties was observed even when pure tumor cell line populations displaying distinct site-specific preferences were studied (16, 17, 19, 20), suggesting that part of this expression profile is indeed intrinsic to the tumor cells. Furthermore, we did not observe any segregation of our samples according to metastatic site when subjected to HCL on normal breast (33) or breast cancer (34) selective genes, confirming that all samples were enriched for breast cancer cells and that the transcriptional profiles observed are most likely mainly tumor cell intrinsic. Nonetheless, the possibility of normal tissue contamination cannot

be completely ruled out. On the other hand, down-regulation of extracellular matrix genes and genes involved in cell adhesion and the development of blood vessels and the skeletal system, which are all processes that have been linked with invasion and metastasis in breast cancer (40) was seen in liver metastases. Of note, the top down-regulated gene was the epithelial mesenchymal transition inducer *PRRX1*, recently reported to play an important role in metastatic colonization through repression of its expression to favor reversion of the mesenchymal phenotype which is necessary for the outgrowth of metastases (41). However, analyses performed in the external dataset which included many more lung metastases and in addition included brain metastases, suggested that down-regulation of these genes may also be a trait of lung and brain, but not bone metastases. Further studies are necessary to investigate this phenomenon.

Predicting the future metastatic site(s) of a primary breast cancer is multifaceted and challenging. In their recent study aimed at unraveling how bone-specific metastatic traits arise in the primary tumor, Zhang and colleagues (38) showed that stromal signals resembling those of the distant target organ play important roles at the primary tumor site to prime cells for colonizing of a specific metastatic niche. Also, three independent gene modules enriched for extracellular matrix (*i.e*. stroma) genes were among the 11 gene modules recently identified to shape the transcriptional landscape of primary breast cancer (42). Interestingly, in this study (42), only expression of the ECM modules showed significant associations with the site of recurrence, although liver metastases were not annotated in this study. Our 17-gene signature was enriched for stroma-related genes and was significantly correlated to the stroma module described by Fredlund *et al*. (31). Consistent with our results, they found that low expression of the stroma module was associated with shorter distant metastasis free survival among patients with luminal A primary tumors (31). Furthermore, an independent study by Bergamaschi and colleagues (43) identified four extracellular matrix gene modules (ECM1 –

ECM4) with prognostic significance in ER positive (luminal) breast cancer, but their survival analyses were not stratified to assess differences between the luminal subtypes. Of note, down-regulation of several genes in our signature was characteristic of the ECM1 module (43), which was associated with the poorest outcome. Taken together, these studies highlight the possibility of harnessing the heterogeneity in the expression of extracellular matrix (stroma) genes to improve prognostication in hormone receptor positive disease.

Currently, prediction of the prognosis in ER positive breast cancer at the transcriptional level is limited to the expression of proliferation-related genes, but high proliferative rate alone is not sufficient to account for all the recurrences observed among patients with ER positive breast cancer, especially among patients with luminal A tumors which are generally of a low proliferating phenotype. Down-regulation of the 17-gene signature was indirectly associated with high proliferation, since features such as high histological grade and luminal B subtype are common to proliferative tumors. Consequently, low expression was independently prognostic of shorter time to recurrence and shorter overall survival among patients with ER positive tumors. Remarkably, the 17-gene signature and tumor size were the only independently prognostic factors for early recurrence among patients with (low proliferative) luminal A tumors in multivariable analyses. Importantly, the luminal A tumors in this cohort were mostly of histological grades 1 and 2. The significantly lower expression of the 17-gene signature in IntClust subgroups 3, 7, and 8, which are predominantly comprised of luminal A tumors, confirms that the IntClust subtypes may also be used to further stratify luminal A tumors into groups with distinct outcome. InClust 3 is mainly characterized by low genomic instability, while IntClust 7 and 8 harbor the characteristic ("luminal") 16p gain/16q loss and 1q gain/16q loss aberrations, respectively. Interestingly, the 17-gene signature captures the diversity in prognosis even within these well-characterized subgroups. Metastases remain the main cause of death from cancer. The goal of individualizing therapy for breast cancer can

only be achieved if all patients at risk can be accurately identified. The prognostic relevance of the 17-gene signature in luminal A breast cancer holds great promise in this context and needs to be independently validated.

The fact that all 17 genes in our signature were found to be over-expressed in the group of primary tumors from patients who subsequently developed liver metastases is surprising since the majority of the genes showed low expression in the liver metastases. The SAM analysis comparing metastases from specific organs, *i.e*. liver *vs*. other sites, disregards the fact that the same patient from whom the liver metastasis was collected may have metastases in other organs. Also, in this study we confirm that paired tumors from the same individual have highly similar global transcriptional profiles. Taken together, the high concordance in global transcription and the fact that SAM analysis only detects differences in levels of gene expression between groups and not an absolute presence or absence thereof, argues that the genes we identified are more liver-selective and therefore likely not uniquely liver-specific *per se*. Furthermore, searching for the expression of site-selective genes in primary tumors, which represent a heterogeneous mix of clones with diverse site-specific metastatic propensities is also complex. In the primary tumor cohort used to identify the subset of liver-selective genes differentially expressed at this early time point during tumor progression, the sub-categorization of patients was also confounded by intra-individual overlap of several metastatic sites. Nonetheless, the inverse correlation in the direction of expression of many of the genes between primary tumors and metastases is intriguing and requires further functional investigation. However, importantly, low expression as observed in the liver metastases was prognostic of an inferior outcome. Since decreased expression of most of the genes (as observed in the liver metastases) is associated with inferior outcome, we hypothesize that the lower expression may be a stronger marker of overall inferior prognosis rather than only a marker for liver-specific recurrence. This is in line with the understanding that liver

metastasis is an indicator of poor prognosis. The scarcity of datasets with annotations for the metastatic site(s) hindered an independent evaluation of the ability of this signature to specifically predict breast cancer liver recurrence.

Diagnosis of liver-only metastases in breast cancer is not common and liver metastases are frequently diagnosed in tandem with other sites as can be seen in our patient cohort where only 19/133 (14%) patients presented with liver-only disease at first diagnosis of metastatic disease. This suggests that liver metastases and tumor burden are strongly associated and our signature may to some extent be associated with tumor burden. Of importance however, the liver metastases clustered together and displayed similar transcriptional profiles regardless of whether they were diagnosed as oligo-metastases or in parallel to other known metastatic deposits, supporting the liver selectivity of the identified gene signature. Identification of independent site-specific signatures would therefore require a well-annotated and sufficiently large cohort of patients with oligo-metastatic disease, which is challenging given the scarcity of patients presenting with oligo-metastases as well as the fact that biopsies are seldom taken from patients presenting with oligo-metastatic disease. Biopsies of metastases are now routinely collected whenever possible for reassessment of biomarkers to guide treatment for metastatic breast cancer. Ultimately, the gap of scarcity of these samples will be bridged and larger collections of metastases will become available for research purposes, enabling *e.g.* validation of the data presented herein. Notwithstanding this limitation, our analysis pipeline enabled us to identify a biologically important gene set, the clinical relevance of which was independently validated in a large cohort of primary breast cancer.

In conclusion, we have identified a 17-gene signature enriched for genes selectively under-expressed in breast cancer liver metastases, with a remarkable ability to independently identify patients with luminal A primary breast cancers who may benefit from closer disease monitoring and may in addition be candidates for enrollment into clinical trials investigating

novel targeted therapies. Further studies are warranted to validate our results especially in more recently diagnosed patient series to adjust for modern advances in adjuvant breast cancer management.

Figure legends

Figure 1. Post-recurrence survival according to metastatic category. Patients were categorized according to the most advanced metastatic site (loco-regional, locally advanced or regional metastases in the lymph nodes or skin; bone, skeletal metastases with or without loco-regional disease; lung, lung parenchymal/pleural metastases with or without skeletal and loco-regional metastases; liver, hepatic metastases with or without lung, skeletal or loco-regional metastases). In addition, patients with liver recurrences were further stratified into two groups based on the number of sites involved (oligo, n=1 and multiple, n>1). A significantly inferior survival was observed for patients with liver metastases occurring parallel with metastatic deposits in other organs.

Figure 2. Unsupervised analyses of global transcriptional similarities and differences between breast cancer metastases. PCA analyses showing associations with (**A**) ER status of the primary tumor, (**B**) intrinsic subtype of the metastasis and (**C**) specific site of the metastatic biopsy profiled. The contributions of the first three components in explaining the observed variation in the data were: PC1=t(1)=15.1%, PC2=t(2)=8.52%, and PC3=t(3)=4.38%. (Overall Model coefficients: R2X=variation in X=0.512 and Q2=variation from cross-validation=0.261). **D**) Dendrogram showing HCL of metastases using the top 3,000 most variable probes. Highlighted samples in the tree represent pair-wise independent metastases from the same patient.

Figure 3. Associations between key breast cancer-specific biological gene modules and the site of metastasis. **A-D** represent comparisons between site-specific metastatic biopsies and **E-**

H represent comparisons between patient metastatic categories. Statistical significance was evaluated with Kruskal-Wallis tests. The open circles and asterisks represent mild and extreme outliers respectively for each group in each comparison. All statistical tests are two-sided.

Figure 4. Heatmaps from two independent datasets, showing the expression of the 17 liver metastasis-selective genes found to be differentially expressed in primary tumors with a predilection to metastasize to the liver compared to other sites. The heatmap in (**A**) represents our study cohort and (**B**) an external dataset of breast cancer metastases (GSE14018). Red corresponds to up-regulated genes and green corresponds to down-regulated genes. The color scale represents the mean centered log2 expression.

Figure 5. Associations between the 17-gene signature and primary breast cancer pathological features and prognosis. The boxplots in A-D illustrates the median expression of the 17 liver metastasis-selective genes in primary breast tumors. Tumors were stratified according to the PAM50 intrinsic subtypes: **A**) all tumors and **B**) ER positive tumors; and tumor histological grade: **C)** all tumors and **D)** ER positive tumors. *P*-values are from Anova tests. Associations with survival are shown in E-H. **E**) RFS for all ER positive tumors, **F**) RFS for luminal A (PAM50) tumors only, **G**) OS for all ER positive tumors, and **H**) OS for luminal A tumors only. Log rank tests were used for comparison. All statistical tests were two sided and *P*<0.05 was considered to be significant.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Funding

References

1.   Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2012;23 Suppl 7:vii11-9.

2.   Goldhirsch A, Gelber RD, Castiglione M. Relapse of breast cancer after adjuvant treatment in premenopausal and perimenopausal women: patterns and prognoses. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1988;6:89-97.

3.   Imkampe A, Bendall S, Bates T. The significance of the site of recurrence to subsequent breast cancer survival. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2007;33:420-3.

4.   Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V, et al. Prognostic factors in 1,038 women with metastatic breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2008;19:2012-9.

5.   Pentheroudakis G, Fountzilas G, Bafaloukos D, Koutsoukou V, Pectasides D, Skarlos D, et al. Metastatic breast cancer with liver metastases: a registry analysis of clinicopathologic, management and outcome characteristics of 500 women. Breast cancer research and treatment. 2006;97:237-44.

6.   Yardley DA. Visceral disease in patients with metastatic breast cancer: efficacy and safety of treatment with ixabepilone and other chemotherapeutic agents. Clinical breast cancer. 2010;10:64-73.

7.   Kimbung S, Kovacs A, Bendahl PO, Malmstrom P, Ferno M, Hatschek T, et al. Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. Molecular oncology. 2014;8:119-28.

8.   Yerushalmi R, Woods R, Kennecke H, Speers C, Knowling M, Gelmon K. Patterns of relapse in breast cancer: changes over time. Breast cancer research and treatment. 2009;120:753-9.

9.   Atalay G, Biganzoli L, Renard F, Paridaens R, Cufer T, Coleman R, et al. Clinical outcome of breast cancer patients with liver metastases alone in the anthracycline-taxane era: a retrospective analysis of two prospective, randomised metastatic breast cancer trials. European journal of cancer. 2003;39:2439-49.

10.   Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28:3271-7.

11.   Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, et al. Subtypes of breast cancer show preferential site of relapse. Cancer research. 2008;68:3108-14.

12.   Desouki MM, Liao S, Huang H, Conroy J, Nowak NJ, Shepherd L, et al. Identification of metastasis-associated breast cancer genes using a high-resolution whole genome profiling approach. J Cancer Res Clin Oncol. 2010;137:795-809.

13.   Wang C, Iakovlev VV, Wong V, Leung S, Warren K, Iakovleva G, et al. Genomic alterations in primary breast cancers compared with their sentinel and more distal lymph node metastases: an aCGH study. Genes, chromosomes & cancer. 2009;48:1091-101.

14.   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. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:15901-5.

15.   Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, et al. Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. Breast cancer research and treatment. 2012.

16.   Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer cell. 2003;3:537-49.

17.   Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;436:518-24.

18.   Landemaine T, Jackson A, Bellahcene A, Rucci N, Sin S, Abad BM, et al. A six-gene signature predicting breast cancer lung metastasis. Cancer research. 2008;68:6092-9.

19.   Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature. 2009;459:1005-9.

20.   Tabaries S, Dong Z, Annis MG, Omeroglu A, Pepin F, Ouellet V, et al. Claudin-2 is selectively enriched in and promotes the formation of breast cancer liver metastases through engagement of integrin complexes. Oncogene. 2011;30:1318-28.

21.   Tabaries S, Dupuy F, Dong Z, Monast A, Annis MG, Spicer J, et al. Claudin-2 promotes breast cancer liver metastasis by facilitating tumor cell interactions with hepatocytes. Mol Cell Biol. 2012;32:2979-91.

22.   Hatschek T, Carlsson L, Einbeigi Z, Lidbrink E, Linderholm B, Lindh B, et al. Individually tailored treatment with epirubicin and paclitaxel with or without capecitabine as first-line chemotherapy in metastatic breast cancer: a randomized multicenter trial. Breast cancer research and treatment. 2012;131:939-47.

23.   McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). Nature clinical practice Urology. 2005;2:416-22.

24.   **The R Project for Statistical Computing**. www.r-project.org.

25.   Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. BioTechniques. 2003;34:374-8.

26.   Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27:1160-7.

27.   Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:5116-21.

28.   Zhang XH, Wang Q, Gerald W, Hudis CA, Norton L, Smid M, et al. Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer cell. 2009;16:67-78.

29.   Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1-13.

30.   Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4:44-57.

31.   Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A, Ringner M. The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. Breast cancer research : BCR. 2012;14:R113.

32.   Ringner M, Fredlund E, Hakkinen J, Borg A, Staaf J. GOBO: gene expression-based outcome for breast cancer online. PloS one. 2011;6:e17911.

33.   Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, et al. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics. 2005;86:127-41.

34.   Axelsen JB, Lotem J, Sachs L, Domany E. Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:13122-7.

35.   Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486:346-52.

36.   Ali HR, Rueda OM, Chin SF, Curtis C, Dunning MJ, Aparicio SA, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome biology. 2014;15:431.

37.   Kimbung S, Kovacs A, Bendahl PO, Malmstrom P, Ferno M, Hatschek T, et al. Claudin-2 is an independent negative prognostic factor in breast cancer and specifically predicts early liver recurrences. Molecular oncology. 2013.

38.   Zhang XH, Jin X, Malladi S, Zou Y, Wen YH, Brogi E, et al. Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. Cell. 2013;154:1060-73.

39.   Talmadge JE, Fidler IJ. AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective. Cancer research. 2010;70:5649-69.

40.   Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144:646-74.

41.   Ocana OH, Corcoles R, Fabra A, Moreno-Bueno G, Acloque H, Vega S, et al. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer cell. 2012;22:709-24.

42.   Wolf DM, Lenburg ME, Yau C, Boudreau A, van 't Veer LJ. Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. PloS one. 2014.

43.   Bergamaschi A, Tagliabue E, Sorlie T, Naume B, Triulzi T, Orlandi R, et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. The Journal of pathology. 2008;214:357-67.

Table 1. Associations between the first site(s) of metastasis and patients and tumor pathological features.

Primary tumor Charateristic	All tumors						ER positive tumors					
		Metastatic category						Metastatic category				
	N	Loco-regional	Bone	Lung	Liver	P value	N	Loco-regional	Bone	Lung	Liver	P value
ER Status												
Negative	68	20	7	19	22	**0.002**						
Positive	213	29	43	40	101							
PR Status												
Negative	81	18	10	18	35	0.21	44	8	5	7	24	0.32
Positive	110	16	25	21	48		107	16	25	20	46	
HER2 Status												
Negative	179	32	32	37	78	0.76	143	23	29	24	67	0.39
Positive	17	2	2	5	8		8	0	1	3	4	
Number of metastatic sites												
Oligo (n=1)	76	25	23	9	19	**<0.001**	57	17	19	5	16	**<0.001**
Multiple (n>1)	226	25	33	54	114		156	12	24	35	85	
Histological grade												
Grade 1/2	80	9	17	12	42	**0.03**	72	9	15	10	38	0.58
Grade 3	105	24	15	26	40		67	10	13	15	29	
Adjuvant Endocrine therapy												
No	147	27	30	37	53	0.05	73	8	19	17	29	0.17
Yes	154	22	26	26	80		140	21	24	23	72	
Adjuvant Chemotherapy												
No	152	24	30	32	66	0.93	112	15	22	21	54	0.99

Yes	148	25	25	31	67		101	14	21	19	47	
Age at primary diagnosis												
< 50 years	152	17	28	37	70	0.08	107	12	19	24	52	0.38
≥ 50 years	149	32	28	26	63		106	17	24	16	49	
Metastasis-free interval												
≤ 24 months	80	15	15	16	34	0.91	55	7	11	11	26	0.99
> 24 months	221	34	41	47	99		158	22	32	29	75	
Nodal Status												
N0	91	14	17	22	38	0.69	65	7	12	16	30	0.47
N+	202	35	37	37	93		145	22	30	23	70	
Tumor size												
≤ 20 mm	119	14	26	21	58	0.21	82	9	20	12	41	0.47
> 20 mm	178	33	30	40	75		128	18	23	27	60	
Molecular subtype *												
Luminal A-like	65	9	19	11	26	**0.01**	65	9	19	11	26	0.05
Luminal B-like	81	13	9	16	43		81	13	9	16	43	
HER2 positive	9	2	1	2	4							
Triple negative	24	8	2	9	5							

* Molecular subtyping using immunohistochemical staining for ER, PR, HER2 and Ki67 according to the 2013 St Gallen concensus guidelines. Patients were categorized according to the most advanced metastatic site affected (loco-regional, locally advanced or regional metastases in the lymph nodes or skin; bone, skeletal metastases with or without loco-regional metastases; lung, plural metastases with or without skeletal and loco-regional metastases; liver, hepatic metastases with or without plural, skeletal or loco-regional metastases). P values are from Fisher's exact tests.
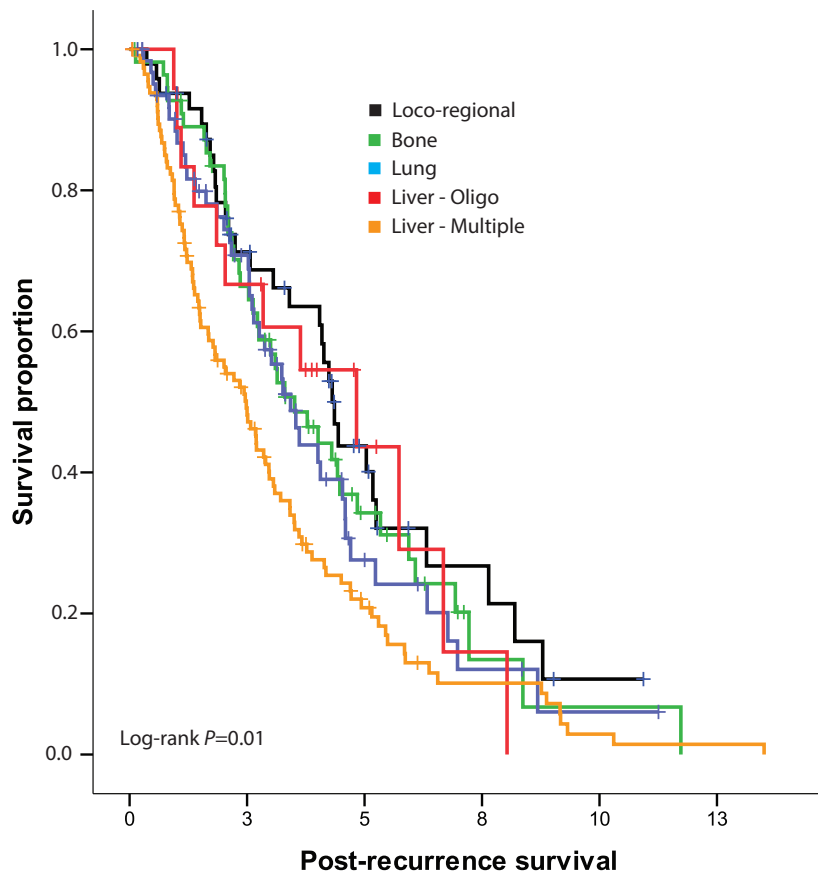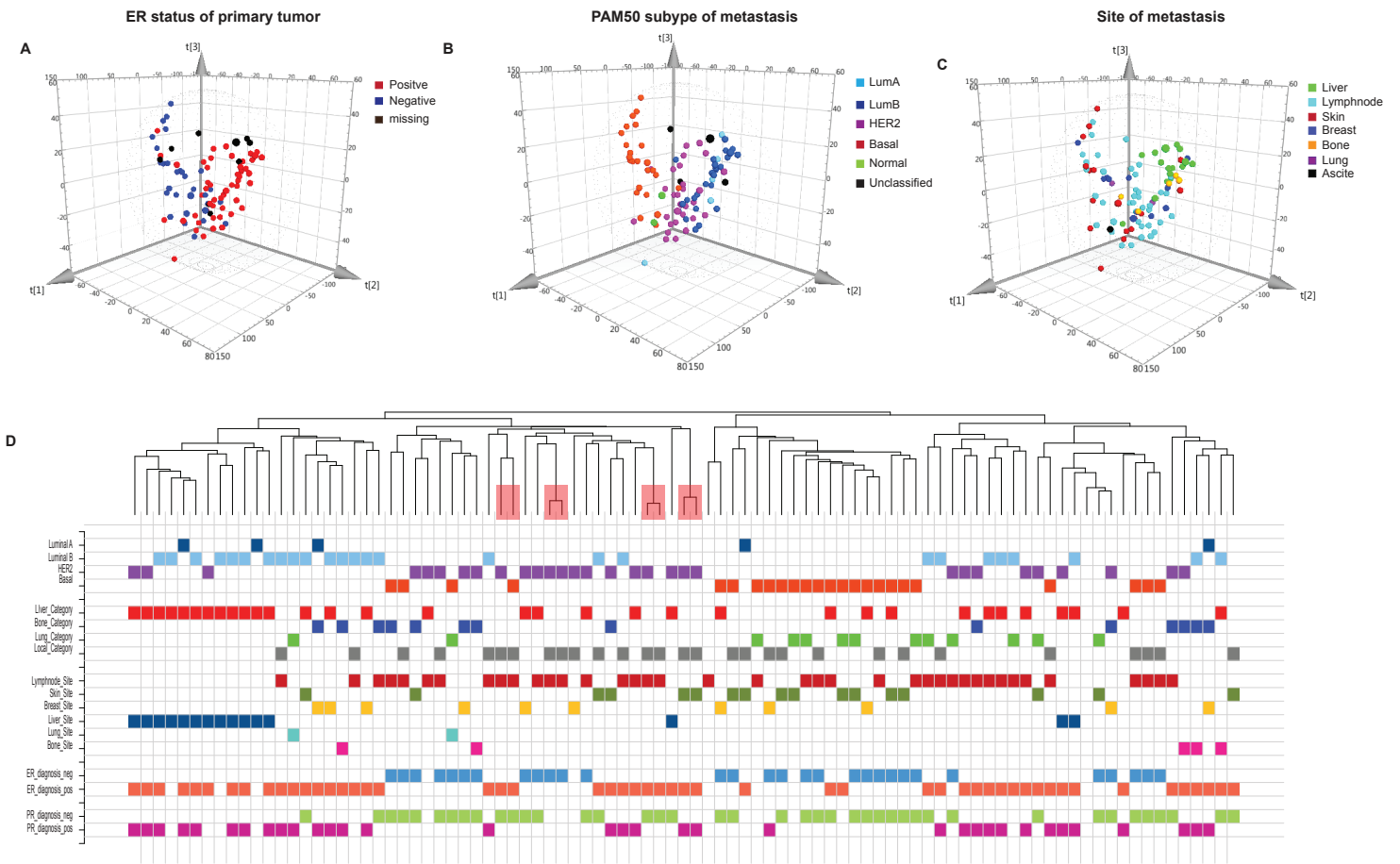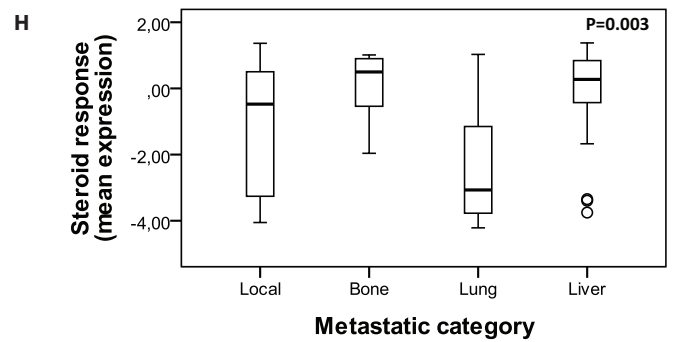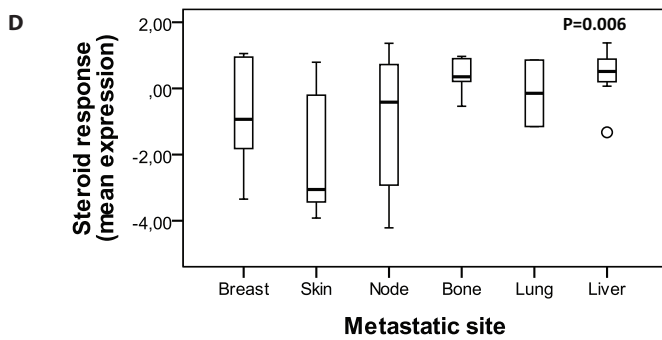
Figure 1

Figure 2

A ER status of primary tumor

Positve
Negative
missing

B PAM50 subype of metastasis

LumA
LumB
HER2
Basal
Normal
Unclassified

C Site of metastasis

Liver
Lymphnode
Skin
Breast
Bone
Lung
Ascite

D

Luminal A
Luminal B
HER2
Basal

Liver_Category
Bone_Category
Lung_Category
Local_Category

Lymphnode_Site
Skin_Site
Breast_Site
Liver_Site
Lung_Site
Bone_Site

ER_diagnosis_neg
ER_diagnosis_pos

PR_diagnosis_neg
PR_diagnosis_pos

Figure 3

Figure 4

A

Liver   Bone   Breast   Lymphnode   Skin   Lung

B

Liver   Brain   Bone   Lung

Figure 5

A All tumors

B ER positive tumors

C All tumors

D ER positive tumors

E All ER positive tumors

F Luminal A tumors

G All ER positive tumors

H Luminal A tumors