 doi:10.20944/preprints202105.0386.v1

Article

Meta-analysis of microdissected breast tumors reveals genes regulated in the stroma but hidden in bulk analysis

Aurora Savino [1,*], Niccolò de Marzo [1], Paolo Provero [2,3] and Valeria Poli [1,*]

[1] Molecular Biotechnology Center, Department of Molecular Biotechnology and Health Sciences, University of Turin, Via Nizza 52, 10126 Turin, Italy;
[2] Department of Neurosciences "Rita Levi Montalcini", University of Turin, Corso Massimo D'Azeglio 52, 10126 Turin, Italy;
[3] Center for Omics Sciences, Ospedale San Raffaele IRCCS, Via Olgettina 60, 20132 Milan, Italy

* Correspondence: aurora.savino@unito.it, valeria.poli@unito.it

Simple Summary:

Gene expression profiling of tumors is an essential approach for biomarkers' selection and to investigate cancer's molecular mechanisms, but transcriptomic results are often difficult to reproduce due to technical biases, samples' heterogeneity or small sample size. Combining many datasets can help reducing artefacts and improve statistical power, therefore we aimed at creating a comprehensive resource of transcriptomic datasets investigating breast cancers, focusing on microdissected tumors, which allow distinguishing the contribution of tumor microenvironment from cancer cells. We define robust lists of differentially expressed genes and describe their relationship with clinical features in each cellular compartment, identifying clinically relevant markers that are retrieved only by directly measuring their expression in the tumor microenvironment.

Abstract:

Background: transcriptome data provide a valuable resource for the study of cancer molecular mechanisms, but technical biases, samples' heterogeneity and small sample sizes result in poorly reproducible lists of regulated genes. Additionally, the presence of multiple cellular components contributing to cancer development complicate the interpretation of bulk transcriptomic profiles. Methods: we collected 48 microarray datasets of laser capture microdissected breast tumors, and performed a meta-analysis to identify robust lists of genes differentially expressed in these tumors. We created a database with carefully harmonized metadata to be used as a resource for the research community. Results: combining the results of multiple datasets improved the statistical power, and the analysis of stroma and epithelium separately allows identifying genes with different contribution in each compartment. Conclusions: our database can profitably help biomarkers' discovery and is readily accessible through a user-friendly web interface (https://aurorasavino.shinyapps.io/metalcm/).

Keywords: tumor microenvironment; meta-analysis; tumor stroma; breast cancer; LCM; microdissection; transcriptomics; microarray; database.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

1. Introduction

High-throughput analyses of gene expression hold great promise for the identification of biomarkers of clinical status, with the potential of predicting outcome, response to therapy or informing researchers about molecular mechanisms underpinning disease onset and progression [1]. Nevertheless, lists of candidate genes obtained through transcriptome-based studies have proved difficult to reproduce [2–6], raising a note of caution about conclusions that can be driven by single experiments. Sample collection and processing methods, protocols and platforms may impact on the resulting gene signatures, making them non-overlapping between studies [7]. Additional variability may be introduced by patient heterogeneity, not sufficiently represented in small samples.

To solve these issues, the vast amount of information present in gene expression databases such as Gene Expression Omnibus, ArrayExpress and EGA [8–10] can be integrated to improve the quality of gene signatures. The advantage is twofold: on one side, a larger sample size allows increasing statistical power; on the other side, merging data obtained with different experimental settings facilitates removing single-experiments' biases, improving robustness [11,12]. Meta-analyses serve this scope, providing a quantitative approach to combine the results of studies investigating the same biological system [5]. Several methods, based on different statistics, have been proposed [13,14]: i) aggregating gene lists based on p-value [15]; ii) effect size [16], and iii) gene rankings [17].

Cancer biology is an extremely prolific field, often with tens of independent studies analyzing the same biological or clinical question in different cohorts. Hence, meta-analyses here show the most of benefit. In their simplest, yet significant, form, meta-analyses have been applied to assess the reliability of specific genes as diagnostic and prognostic markers [18,19], while whole transcriptomic datasets have been employed to the unbiased evaluation and refinement of prognostic signatures [20–25], to identify patients' subgroups [26–30], markers of metastatic tumors [31] and of resistance to treatments [32].

The tumor microenvironment is an important player in determining tumor growth, disease progression and drug resistance [33–35]. It is a composite environment comprising growth factors, cytokines and other non-cancerous cells such as fibroblasts, endothelial, and immune cells [36]. Each of these components can support or inhibit tumor growth, and influence other cell types' behavior in supporting cancer cells. Indeed, cancer associated fibroblasts (CAFs) and immune cells in the tumor microenvironment are being studied as therapeutic targets [37,38].

Understanding the biology of each component of the tumor milieu is necessary to obtain a complete picture of tumors, but obtaining compartment-specific gene expression profiles is laborious and therefore most high-throughput datasets are based on bulk tissues. Nevertheless, relying on samples composed of cell admixtures may hide cell type specific signals, and create confounding effects. For example, tumor composition due to sampling variation significantly impacts on genomic data [39] and on tumor subtype definition [40]. Moreover, differences in the prognostic role of the same gene when measured in different compartments have been reported [41–43].

To overcome these limitations, a number of approaches have been introduced: i) laser capture microdissection (LCM) is commonly employed to separate cell compartments that are histologically well defined [44]; ii) single-cell techniques have allowed distinguishing transcriptomic profiles of different cell types within a tumor [45] and to dissect CAFs' transcriptional heterogeneity [46]; iii) computational methods have been designed to deconvolve *in silico* the contribution of each cell type to the final bulk gene expression profile [47]. Despite all methods being extremely valuable, it must be noted that single-cell techniques are affected by loss of information resulting in dropouts and zero-inflation and, due to the cost, they are usually only applied to screen a few tumors in a single study, impeding the correlation of gene expression profiles with clinical features. Deconvolution methods, on the other side, rely on strong assumptions and depend on the quality of the cell type specific signatures applied as input in the model. Finally, LCM, despite not allowing single-cell resolution, represents a good compromise to disentangle the specific

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

contribution of tumor epithelium and microenvironment, and collect information on many clinically distinct samples.

Here, we performed a meta-analysis of 48 transcriptomic datasets from LCM breast tumor samples, studying the specific contribution of epithelium and stroma to tumors' gene expression profiles. We identified genes robustly changing their expression in each compartment with respect to normal breast, and selected categories of genes with compartment-specific regulation and correlation with clinical features. Importantly, we make the whole database and the harmonized metadata available, and we provide a web-based interface to facilitate its interrogation (https://aurorasavino.shinyapps.io/metalcm/).

2. Materials and Methods

Search of datasets

Transcriptomic datasets of breast tumors analyzed in their stromal compartment were searched on Gene Expression Omnibus (GEO) on December 20th 2020, using the search terms "Breast cancer" AND "lcm" or "Breast cancer" AND "stroma" or "Breast cancer" AND "microdissect*" and selecting as study type "Expression profiling by array" and "Expression profiling by high-throughput sequencing". They were then individually screened to discard datasets not comprising untreated tumor samples. The whole list of datasets comprised in the final database is available in Suppl Table 1. Original works describing each dataset can be found at [48–92].

Data download and pre-processing

GEO datasets were downloaded using the GEOquery package [93]. Normalized (FPKM) breast cancer data from the TCGA data were obtained through TCGAbiolinks [94], and METABRIC transcriptome data were obtained from synapse.org (syn2160410). Clinical and biological annotations were obtained with the same methods.

Mapping of probes to gene symbols was obtained from respective platforms' information in GEO for each dataset, and, in case of multiple probes mapping to the same gene symbol, the probe with the highest mean expression across the dataset's samples was chosen.

Whenever data were not already log transformed, we applied log2 transformation, adding an offset of 1 when data minimum value was 0.

Replicates were merged calculating the average of their expression signals before log transformation. In GSE4823, expression values of dye-swap replicates (not log transformed) were inverted before averaging. In GSE8977, to allow variance stabilization, negative values were removed prior to log transformation.

Database metadata

After defining the biological and clinical annotations to be gathered, AS and NdM independently collected them from GEO. Discrepancies were then individually checked and resolved. Moreover, for datasets GSE14548, GSE16873, GSE20437, GSE21947, GSE22513, GSE26910, GSE33692, GSE35019, GSE38959, GSE5764 and GSE72644, additional clinical annotations were obtained from the original manuscripts' tables.

To allow comparing different datasets, annotations were harmonized as much as possible. When multiple samples from the same subject were available and matched by the subject of origin, all clinical annotations were assigned to all samples, including the histologically normal ones.

Details about single annotations are provided in Appendix A.

Differential expression

Differential expression was calculated with limma, and the information about the subject was added as a factor when samples were subject-matched.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

Comparison of logFC for different datasets was done by taking the genes, by their gene symbols, present in the arrays of all compared datasets and computing pairwise Spearman's correlation. For experiments using mouse systems, human orthologs were obtained from biomaRt [95].

Relationship with clinical features

Relationship between genes' expression and clinical features (grade, age at diagnosis, size) was obtained computing Spearman's correlation and corresponding p-values with the rcorr function from the Hmisc package (https://CRAN.R-project.org/package=Hmisc).

Relationship with survival in the METABRIC cohort was determined by dividing patients in two groups by the median expression value of the gene of interest, and computing their difference in disease free survival with the Kaplan-Meier method through the survival package (https://CRAN.R-project.org/package=survival).

Collapsing p-values

Uncorrected p-values for each gene and each dataset, obtained either from differential expression or correlation with clinical features, were collapsed with the Fisher's method [96] from the metap package (https://CRAN.R-project.org/package=metap). Two different tests were performed separately for testing coherent up- or down-regulation (positive or negative correlation) and taking logFC signs (correlation signs) into account. P-values were then cut at $2.2*10^{-16}$. Since all tests were two-sided, one-sided p-values were obtained with the two2one function from the metap package before applying the Fisher's method. Resulting p-values were corrected for multiple testing with the p.adjust function from R stats, with the default Holm method.

Definition of DEG categories

Collapsed and adjusted p-value cutoff was set to 0.05, while no evidence of differential expression was called for nominal p-values>0.05. Thus, for example, the genes significantly up-regulated only in tumor stroma, and not differentially expressed in tumor epithelium, are those with stromal up-regulation p-value (adjusted)<0.05 and both epithelial up- and down-regulation nominal p-values>0.05.

Enrichment for functional categories

Gene Ontology enrichment was calculated with the enrichGO function from the clusterProfiler package [97], using "Biological Process" GO categories and default parameters.

Epithelial, stromal and vascular scores

Scores for epithelial, stromal and vascular signatures' expression were calculated with a single sample GSEA (ssGSEA) via the GSVA package [98] using signatures obtained from the current meta-analysis comparing tumor and normal gene expression in samples of epithelial, stromal or vascular origin.

Stromal and epithelial markers were obtained comparing epithelial and stromal gene expression profiles from the same dataset, and merging the p-values with the Fisher method as described above. To achieve a higher stringency, we retained only DEGs with |average logFC|>1.

Multivariate Cox models were fit with the coxph function from the survival package (https://CRAN.R-project.org/package=survival).

Plots and statistical analyses

All statistical analyses were performed with R 4.0.4 [99].

Packages used for plotting are R base graphics, ggplot2 [100] and ggsignif (https://CRAN.R-project.org/package=ggsignif), ggvenn (https://CRAN.R-project.org/package=ggvenn), survminer (https://CRAN.R-project.org/package=survminer) and pheatmap (https://CRAN.R-project.org/package=pheatmap).

 doi:10.20944/preprints202105.0386.v1

Web app

The web app (https://aurorasavino.shinyapps.io/metalcm/) was built with the shiny package (https://CRAN.R-project.org/package=shiny), and makes use of rintrojs [101], shinybusy (https://CRAN.R-project.org/package=shinybusy), shinythemes (https://CRAN.R-project.org/package=shinythemes) and shinyWidgets (https://CRAN.R-project.org/package=shinyWidgets).
The data that can be easily interrogated with the app are datasets of primary invasive breast cancers, excluding inflammatory and micropapillary cancers. The user can choose between two conditions to compare based on compartment (stroma or epithelium), disease status (invasive BC, normal or normal counterpart – "Counterpart") and PAM50 subtype. The analysis pipeline applied is the same used in this work and detailed above. Additionally, the enrichment of user-defined gene lists for DEGs can be assessed in the second tab, displaying the result of a one-tailed Fisher test (fisher.test function from R stats). To be noted, when the users inputs a gene list, the correction for multiple testing, used to determine the list of DEGs shown in the first tab, is applied to those genes only. The p-adjustment is applied to all genes when obtaining DEGs for the Fisher test.

Stat3 signatures

Signatures of Stat3 activity were obtained from: Azare et al. [102], Dauer et al. [103], IL6 and Jak/STAT from MSigDB [104], Alvarez et al. [105], Tell and Horvath [106], Sonnenblick et al. [107].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

3. Results

Database construction

We collected 48 transcriptomic datasets of breast tumors or breast hyperplasias deposited in the Gene Expression Omnibus (GEO) database, selecting experiments where different cellular compartments were separated prior to RNA extraction. Most of the datasets (43) derive from laser capture microdissected primary tumors, while a minority measure gene expression of cancer associated fibroblasts (CAFs) grown *in vitro*, either derived from primary tumors or from mouse models. Overall, we collected 2144 samples, 2048 of which derive from primary tumors. The complete list of datasets is available in Suppl. Table 1.

Table 1. Summary of samples' clinical features (complete list in Suppl. Table 2)

Compartment	**# of samples**		**PAM50 subtype**	**# of samples**
Epi	1230		Basal-like	120
Stroma	664		Her2+	71
Vessels	64		LumA	194
Adipose	16		LumB	140
			Normal-like	35
Disease status	**# of samples**			
Invasive BC	990		**Age**	
Tumor (other)	296		Median (range)	55 (27-94)
Normal counterpart	370		# annotated samples	756
Normal	326			
			Size (in mm)	
ER status	**# of samples**		Median (range)	24 (4-161)
Positive	502		# annotated samples	378
Negative	419			
			Grade	**# of samples**
PR status			I	50
Positive	306		II	189
Negative	435		III	312
Her2 status	**# of samples**		**LN positivity**	**# of samples**
Positive	302		Positive	309
Negative	661		Negative	228

To facilitate the comparability of different experiments, we mapped the probes used for each specific experiment to gene symbols, and we did an extensive and careful harmonization of biological and clinical annotations, as detailed in the Methods. Specifically, we gathered information about cellular compartment, disease status, receptors' status (Estrogen receptor, Progesteron receptor and HER2 amplification), PAM50 subtype, tumor histology, size, grade, TNM and overall pathological stage, node positivity, recurrence, response to treatment and patient's age at diagnosis, ethnicity, and menopause status (Table 1 and Suppl. Table 2). Moreover, wherever possible and appropriate, we predicted the PAM50 breast cancer (BC) subtype from gene expression, obtaining a good concordance between subtypes and receptors' expression (Suppl. Fig. 1), and inferred clinical variables from other available clinical annotations (Methods and Appendix A). Importantly, for 11 of the 48 datasets, we found clinical annotations that were made available by the authors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

in the original publication, but not accompanying the corresponding dataset in GEO. Including this information in our database, we significantly improved clinical annotations, initially relatively scarce. For example, we increased the number of samples annotated for age from 530 to 756, and for size from 122 to 378. The complete database comprising gene expression data and metadata is available in Suppl. Tables 3-4, while Suppl. Table 2 comprises the complete summary of available clinical features.

Primary and in vitro datasets are only slightly correlated

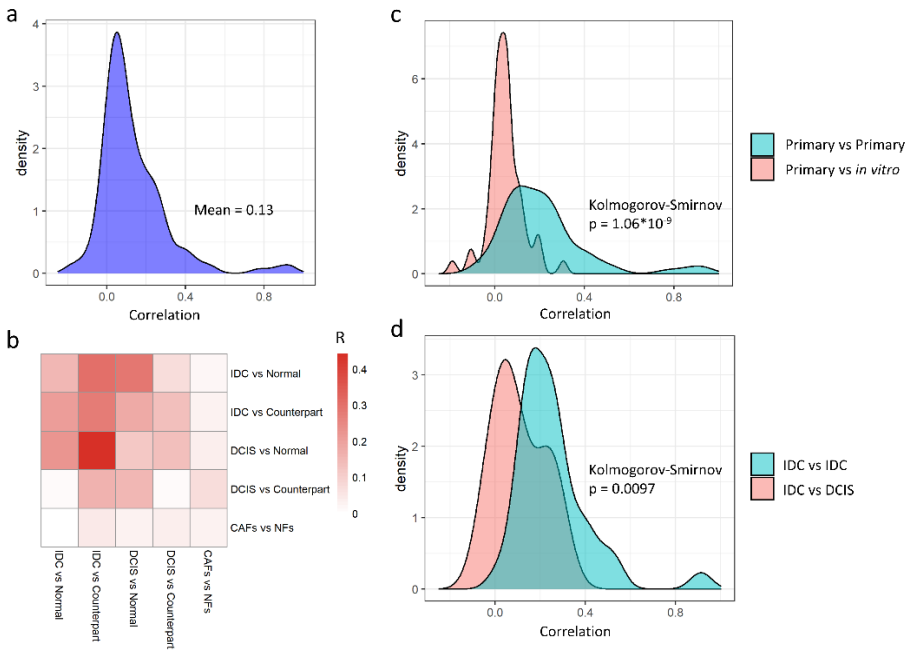

Figure 1. Similarity between stroma datasets of detected genes changing between normal and cancer samples. a) Distribution of Pearson's correlation between log2 fold changes (logFC) of genes between cancerous/non-cancerous conditions in each dataset, showing an average correlation of 0.13. b) Correlation of logFC between groups of datasets comparing Invasive Ductal Carcinoma (IDC) and normal breast tissue from healthy donors (Normal) or histologically normal tissue adjacent to tumor (Counterpart), Ductal Carcinoma In Situ (DCIS) vs Normal or Counterpart, Cancer Associated Fibroblasts (CAFs) and Normal Fibroblasts (NFs) grown in vitro. Red indicates high average positive correlation. c) Distribution of Pearson's correlations between logFC obtained from datasets sampling primary tumors or comparing logFC obtained from primary tumors or from in vitro experiments. d) Distribution of Pearson's correlations between logFC obtained from datasets sampling IDCs or comparing logFC obtained from IDCs and from DCIS.

First, we quantified gene expression changes between normal and tumor stroma in primary Invasive Ductal Carcinomas (IDC), Ductal Carcinoma *In Situ* (DCIS) or CAFs grown *in vitro*. As reference condition, we used samples from cancer patients labelled as histologically normal ("Normal counterpart" or simply "Counterpart"), or, where available, normal breast tissue from reduction mammoplasty ("Normal"). To avoid batch effects, we analyzed each dataset separately and then compared gene expression fold changes to assess similarities and differences between datasets. The correlations between pairs of comparisons are globally, but slightly, positive (Figure 1a, mean correlation of 0.13), indicating intrinsic differences between datasets. Moreover, different classes of datasets display different degrees of similarity (Figure 1b). Comparing primary tumors and *in vitro* samples, it appears that they are only slightly positively correlated (Figure 1b-c, ϱ=0.05). Similarly, IDC and DCIS samples behave differently (Figure 1d), motivating us to keep them separated for meta-analytic purposes in order to limit the biological heterogeneity.

Non-redundant information is obtained by separating different tissues

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

We then took advantage of the tissue specificity of the collected datasets to compare tumor epithelium and stroma gene expression behaviors. We thus calculated differentially expressed genes (DEGs) for each dataset comprising normal and tumor samples, analyzing epithelium and stroma separately. Then, we collapsed the differential gene expression statistics to obtain a global measure of the reliability of gene expression changes across all available datasets. As microarray platform and pre-processing can impact on the measure of fold change in differential gene expression, we chose to apply the widely employed Fisher's method, summing the log-transformed p-values obtained from independent studies [96]. To limit the heterogeneity of input data, for this analysis we employed invasive BC samples only. In total, we could perform 9 comparisons for epithelium and 9 for stroma. Strikingly, comparing the average fold change for each gene in tumor stroma or epithelium, we observed that genes behave similarly across compartments (Figure 2a), suggesting coordinated gene expression reorganization between tumor and surrounding cells. Alternatively, it is possible that, despite the use of LCM, the two compartments have not been perfectly separated, resulting in shared DEGs.

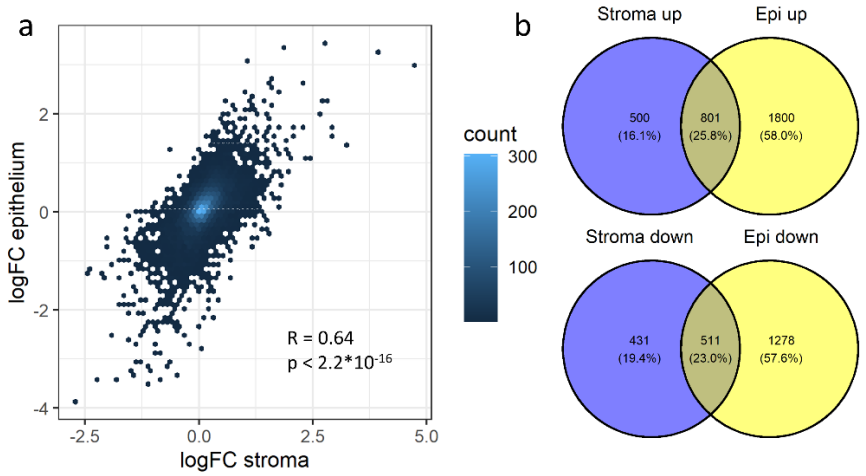

Figure 2. DEGs' comparison between stroma and epithelium. a) Cross-datasets average logFC for genes measuderd in normal/cancerous stroma (x axis) or in normal/cancerous epithelium (y axis). Pearson's correlation and p-value are indicated; the colour indicates the number of overlapping dots. b) Venn diagrams comparing significant DEGs (p-adjusted<0.05) detected in the stroma and in the epithelium.

In the meta-analytic setting, we defined lists of robust DEGs. Specifically, we found 4390 DEGs in the epithelium (2601 up- and 1789 down-regulated in tumors) and 2243 DEGs in the stroma (1301 up- and 942 down-regulated) (Figure 2b, Suppl. Table 5). The full list of fold changes and p-values obtained for each dataset and condition is available in Suppl. Table 6 and GO categories enriched for each class are listed in Suppl. Table 7. Of note, combining the information present in several datasets we are able to identify 229 DEGs that would not be identified in any individual dataset analyzed separately.
Comparing changes in gene expression in the tumor stroma and epithelium, we observed that they are globally correlated (Figure 2a). Nevertheless, for 17% of the DEGs there was statistically significant evidence of differential expression only in stroma/epithelium. Interestingly, 104 genes show significant differential expression in both tumor stroma and epithelium, but with opposite signs (Suppl. Table 8). We posit that genes regulated in one compartment only or with opposite regulation between compartments are potentially relevant in tumor progression, but their regulation in tumorigenesis is hidden in bulk datasets since the regulation in one compartment is confounded by a different regulation in the other.

To test this hypothesis, we selected 5 classes of genes: 1) genes up-regulated in both tumor stroma and epithelium (UpBoth), 2) genes with evidence of up-regulation only in one of the two compartments (up-regulated in the epithelium – UpEpi – or in the stroma –UpStr-), 3) genes with opposite sign for differential expression in the two compartments

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

(up-regulated in stroma and down-regulated in epi – StrEpi – or down-regulated in stroma and up-regulated in epi – EpiStr -) , 4) genes with evidence of down-regulation only in one of the two compartments (down-regulated in the epithelium – DnEpi – or in the stroma –DnStr-), 5) genes down-regulated in both tumor stroma and epithelium (DnBoth) (Suppl. Table 8). For each gene, we measured the average expression fold change between normal breast and breast tumor in bulk samples obtained from the TCGA [108]. The classes with the highest fold changes are those comprising the genes that are differentially expressed in both compartments, while the remaining classes show average fold changes closer to zero (Figure 3a), indicating that their genes are not coherently differentially expressed in bulk tumors. Similarly, using the extensive clinical annotations of the METABRIC dataset, the same classes show graded relationships with patients' overall survival, tumor grade and size (Figure 3b-d). A more detailed picture of DEGs' classes indicates that genes with opposite regulation in the stroma/epithelium show correlations with clinical features in line with their regulation in the epithelium: the EpiStr class increases its expression in bulk tumors when compared to normal breast (Suppl. Fig. 2a), it is enriched for genes correlated with poor prognosis (Suppl. Fig. 2b), and correlates with higher tumor grade (Suppl. Fig. 2c) and size (Suppl. Fig. 2d). The class of genes down-regulated in the epithelium and up-regulated in the stroma shows the opposite trend for its expression in bulk tumors and its correlation with size, while it is not related with patients' survival or tumor grade (Suppl. Fig. 2). These observations fit with higher epithelium than stromal content in bulk samples, which hides the signal originating from the tumor stroma. Indeed, the top DEGs in the category of genes up in tumor stroma and down in tumor epithelium are relatively coherently up-regulated in tumor stroma across available datasets (Figure 4a), but they all appear significantly down-regulated in tumor samples from the TCGA (Figure 4b), confirming that their regulation in the stroma is not detected in bulk.

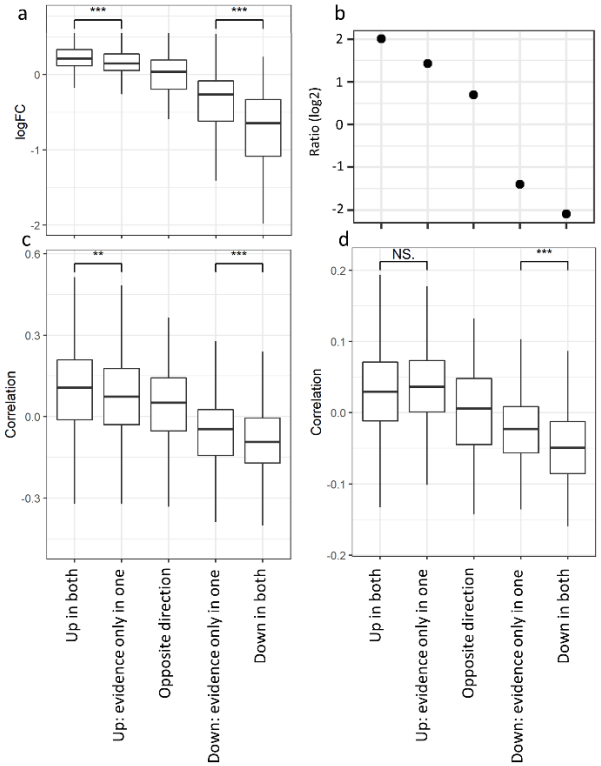

Figure 3. The 5 classes of DEGs in bulk samples. Genes regulated in both tumor epithelium and stroma, significantly regulated only in one of the two compartments and with opposite regulation in the two compartments show different degrees of: a) differential expression between tumor and normal samples in the bulk samples of the TCGA cohort; b) ratio between the number of genes significantly correlated with poor prognosis and the number of genes significantly correlated with good prognosis in the METABRIC cohort. High values indicate that many genes are correlated with poor prognosis, while, inversely, negative values

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

indicate that many genes are correlated with good prognosis; Spearman's correlation with tumor grade (c) and size (d). *** Wilcoxon test p<0.001.

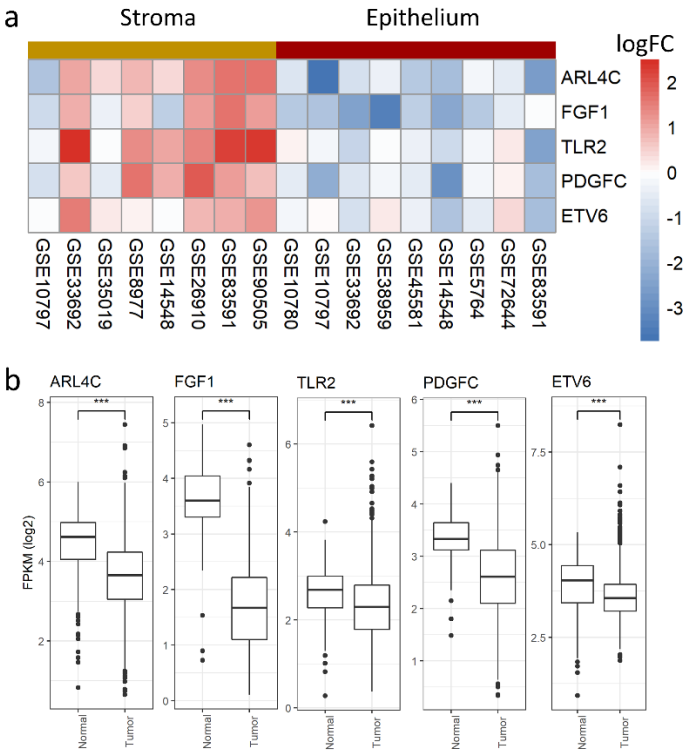

Figure 4. The opposite regulation of genes in tumor stroma or epithelium is hidden in bulk samples. a) Top 5 significantly differentially expressed genes with higher expression in tumor stroma and lower expression in tumor epithelium when compared to respective compartments in normal breast tissue. Rows correspond to the 5 selected genes, columns to GEO IDs of datasets where the comparisons were possible and that were merged in the meta-analysis. On the left, datasets comparing normal and tumor stroma are shown, while on the right there are datasets comparing normal and tumor epithelium. The color in the heatmap indicates logFC value for the corresponding gene and dataset. b) Gene expression changes detected in bulk samples from the TCGA dataset for the 5 selected genes tested with the Wilcoxon test (* <0.05, **<0.01, ***<0.001).

To test the potential relevance of genes that are differentially expressed in one compartment only as compartment-specific clinical markers, we computed their correlation with tumor grade or age at onset when measured in the stroma or in the epithelium of invasive BC. Indeed, genes up-regulated in the stroma are more strongly correlated with higher grade and earlier onset when measured in the stroma than in the epithelium, while genes down-regulated in the stroma show the opposite trend (Figure 5) , supporting the hypothesis of their potential compartment-specific clinical relevance, which can not be assessed without separating them. Similar trends are observed with node positivity and size, even if some tests do not reach statistical significance (Figure 5). Nevertheless, we did not observe this consistent relationship for genes regulated in the epithelium (Suppl fig. 3a). Accordingly, the classes of genes with opposite regulation in stroma and epithelium display opposite relationship with clinical features when measured either in the stroma or in the epithelium (Suppl. Fig. 3b), confirming that also these classes of DEGs might be relevant for tumor progression but differentially regulated in the two compartments.

Indeed, with our compartment-specific gene expression database and the use of meta-analysis, we can identify relationships between genes and clinical features that could not be identified otherwise. As an example, we took the top 50 genes most down-regulated in tumor stroma when compared to normal stroma in DnStr. Twelve of these genes are also significantly negatively correlated with tumor grade when measured in the stroma, and most of them (75%) display stronger relationship with grade in LCM data

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

than in bulk. We identified UPB1 as negatively correlated with grade in tumor stroma but not in bulk (Figure 6) and we note that, despite the relatively small number of samples annotated for clinical features and the high within- and between-dataset variability, we are able to improve statistical power by combining multiple datasets (Figure 6, left panel). In one case (the HSD11B2 gene), the correlation in stroma and in bulk shows opposite trends (Suppl. Fig. 4), that could be due to the confounding effect of multiple cell types present in mixed samples. Conversely, we identified NECAP2 as up-regulated in tumor stroma and correlated with higher tumor grade when measured directly in the stroma (p-value=0.01, average ϱ=0.30), but only slightly when measured in bulk (p-value=0.03, ϱ=0.05).

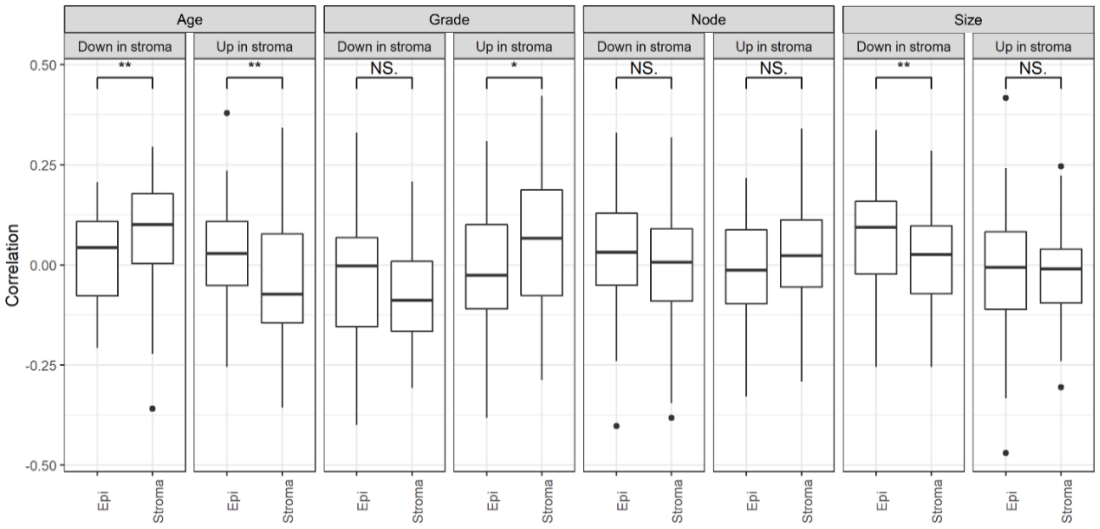

Figure 5. Compartment-specific relationship with clinical features for genes regulated in tumor stroma. Spearman's correlation between genes' expression and age at onset, tumor grade, lymph node status, and size for the classes of DEGs regulated only in tumor stroma, calculated for each dataset separately and then averaged. The correlation computed when their expression is measured directly in the stroma or in the epithelium is compared with the Wilcoxon test (* <0.05, **<0.01, ***<0.001).

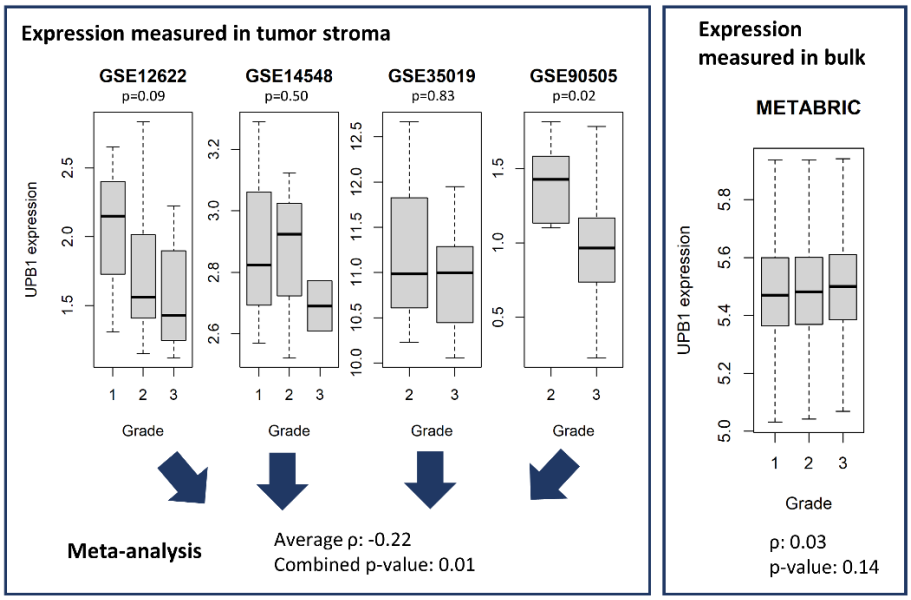

Figure 6. UPB1 is correlated with lower grade in the stroma, but not in bulk. Left panel, UPB1 expression in the 4 datasets with annotation for grade in the tumor stroma. The p-value for the correlation between UPB1 expression and tumor grade is indicated above each boxplot. Collapsing p-values with a meta-analysis, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

statistical power increases and the p-value reaches significance. Overall, the correlation between UPB1 expression and tumor grade in stroma is significantly negative. Right panel, no significant relationship between UPB1 expression and tumor grade in bulk, in the METABRIC cohort.

Considering the enrichment for Gene Ontology categories, genes in the UpBoth class show enrichment for mitochondrion-related categories, ECM and antigen processing, while UpEpi genes are enriched for cell cycle and DNA repair. No categories are enriched for UpStr genes. Counter-intuitively, DnBoth genes are enriched for angiogenesis-related categories. DnEpi genes belong to the "cornification" category, while no biological processes are over-represented in DnStr genes. Interestingly, genes with opposite regulation and over-expressed in tumor stroma are enriched for cytokine secretion and Toll like receptor 2 signaling. Full GO lists are available in Suppl. Table 9.

We repeated the meta-analysis to identify robust DEGs in tumor blood vessels, obtaining 13 up- and 1 down-regulated genes (Suppl. Table 5).

Despite the difficulty of accurately detecting in bulk the signal deriving from specific cellular compartment, as discussed above, we showed that our epithelial and vascular signatures are independent predictors of patients' disease-free survival (DFS) (Figure 7) also in bulk samples. This, again, supports the importance of cell type specific analysis of gene expression in cancer biology. Indeed, higher expression of genes up-regulated in the epithelium or in blood vessels independently correlate with poor prognosis in the METABRIC BC cohort (Figure 7a), while genes down-regulated in blood vessels define the only significant signature of good prognosis (Figure 7b). We could not detect any relationship between stromal signatures and patients' DFS. This result can be attributed to the presence of multiple cell types in bulk samples, confounding compartment-specific signals. Nevertheless, it is also possible that the stromal signatures we defined, despite being correlated with tumor grade and age at onset, are not correlated with patients' survival.

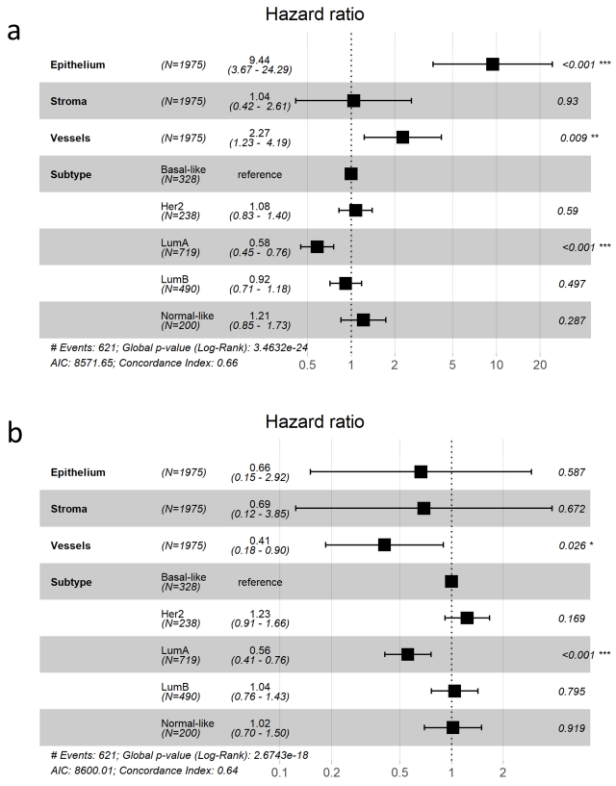

Figure 7. Survival models with epithelial, stromal and vessels' DEGs. Forest plots of multivariate Cox models of disease free survival with PAM50 subtypes and expression levels of genes in our signatures of genes (a) up- or (b) down-regulated in tumor epithelium, stroma or vessels when compared with corresponding normal tissues (* <0.05, **<0.01, ***<0.001).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 May 2021 doi:10.20944/preprints202105.0386.v1

Stromal and epithelial markers to impute cell proportions from bulk samples

Computational methods to estimate cell type proportions in bulk transcriptomes often require gene expression signatures of the cell types of interest. Amongst the first proposed and most commonly applied methods there is ESTIMATE [109], based on single sample GSEA (ssGSEA) of stromal and immune signatures to infer their proportions from transcriptomes of cell admixtures. We computed ssGSEA on primary BC transcriptomes of the METABRIC cohort based on stromal and epithelial markers obtained through a meta-analysis comparing tumor epithelium and stroma from our gene expression database. We observed that the epithelial signature is positively correlated with cellularity (Spearman's rho= 0.12, p-value= $2.8*10^{-7}$) and the stromal signature is negatively correlated with cellularity (Spearman's rho= -0.25, p-value < $2.2*10^{-16}$) (Figure 8), indicating that obtained marker lists are representative of the corresponding cell compartment and appropriate to infer cell proportions. Moreover, our stromal signature shows stronger correlation with cellularity than the estimated tumor purity obtained with the original ESTIMATE signatures (for which Spearman's rho= -0.15, p.value= $1.4*10^{-11}$).

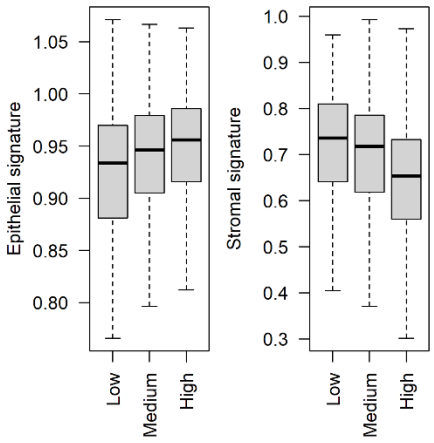

Figure 8. Correlation between epithelial or stromal markers and cellularity in the METABRIC cohort.

Potential use of the database and web platform

Given the number of datasets and of annotation categories in our database, many combinations of variables can be selected for comparison, to address specific questions. To facilitate the choice of the datasets to analyze, we indicate the annotations available for each dataset (Suppl. Fig. 5-6) and some of the comparisons that can be made using a specific dataset's samples (Suppl. Fig. 7-8). For example, there are 10 different datasets with estrogen receptor status' annotations in stroma samples (Suppl. Fig. 5), three datasets that can be used to compare the stroma of IDC and DCIS (Suppl. Fig. 7b) and three datasets allowing comparing Basal and LumA stroma in invasive BC (Suppl. Fig. 8a).
To ease the interrogation of the whole collection of datasets (available in Supplementary Tables 3 and 4) we created a web app providing a user-friendly interface, which allows generating lists of DEGs between two conditions, and testing the enrichment for user-provided gene lists (https://aurorasavino.shinyapps.io/metalcm/). As use-case example, we tested the enrichment of genes up-regulated in the tumor stroma for five signatures of the pro-oncogenic transcription factor Signal Transducer and Activator of Transcription (STAT) 3 activation. We set the parameters for comparing tumor stroma and normal stroma of any BC subtype, with a p-value threshold of 0.05. We then loaded the gene lists corresponding to the five signatures, one at a time, and obtained the Fisher test enrichment p-value in the "Enrichment" tab. We observe that the enrichment is significant for 3/5 of them when tested in the stroma, while none show enrichment for genes up-regulated in the tumor epithelium. This observation is suggestive of different roles of STAT3 in the two compartments, linking tumorigenesis and its up-regulation in the stroma. A lack of significance in the epithelium due to higher variability in the available datasets

 doi:10.20944/preprints202105.0386.v1

cannot be excluded, but it is unlikely given that the number of DEGs detected in tumor epithelium was higher than in tumor stroma (Figure 2b).

4. Discussion

Breast cancer is a heterogeneous disease, with several cellular components playing specific roles in its development and clinical course. In particular, the microenvironment has been shown to either counteract or promote tumor progression depending on the specific conditions, and important players such as immune cells or cancer associated fibroblasts are object of intense study. Importantly, the analysis of bulk tumor samples, comprising cell admixtures, complicates disentangling specific behaviors of each cell component. Laser capture microdissection can help separating the contribution of different cell compartments, and still present some relevant advantages with respect to single cell techniques. Indeed, the cell type separation does not rely on set of markers, but it is directly based on histological features. Moreover, the relatively contained cost allows sampling ranges of tumors with different characteristics and to assess relationships between gene expression and clinical features.

As an additional complication, breast cancer heterogeneity is hardly captured in the small sample sizes of most microarray studies, but the strong research effort dedicated to this biological system has led to a rich collection of independent datasets that can be combined to improve robustness and statistical power.

Here, we gathered 48 transcriptomic datasets of microdissected breast tumors, where stroma and epithelium were separated prior to RNA extraction, to study the distinct behavior of different cellular compartments. We carefully collected and harmonized corresponding biological and clinical annotations to facilitate data integration.

With this tool in hand, we identify genes changing their expression robustly and coherently in breast tumors when compared to normal tissue, either in the stroma or in the epithelium. Analyzing these lists separately, we detect increased expression of cell cycle related genes in the tumor epithelium, and of immune-related categories in the tumor stroma. Also, the non-canonical Wnt/PCP pathway is over-expressed in both stroma and epithelium, involved in breast cancer progression [110], and synergizing with the STAT3 pathway in contributing to its aggressiveness [111]. Moreover, we observe a decrease in lipid catabolism in the tumor stroma, consistent with a potential metabolic coupling between cancer cells and the microenvironment, with stromal components reducing their consumption to release lipids and feed cancer cells' growth [112].

Comparing differentially expressed genes in the two compartments, genes up-regulated in both are enriched in mitochondrial-related and extracellular matrix Gene Ontology categories. Indeed, the extracellular matrix can act as a reservoir of growth factors and its remodeling has been associated with metastatic spread [113], and the role of high oxidative phosphorylation in tumors is being increasingly recognized [114,115]. Counter-intuitively, down-regulated genes are enriched for angiogenesis. A similar trend had already been observed [116] and could be explained by the higher resistance of tumor cells to apoptosis under hypoxic conditions, especially of advanced tumors that also show lower microvessel densities than normal tissues [117].

Moreover, we find groups of genes significantly regulated only in one compartment, and with a corresponding compartment-specific relationship with clinical features. For example, genes up-regulated only in tumor stroma correlate with higher grade and earlier age at onset when measured directly in the stroma, while their relationship with clinical features is weaker when their expression is measured in the epithelium. Among regulated genes we find UPB1 and HSD11B2, the enzyme that converts cortisol in cortisone, which is down-regulated in tumor stroma and correlated with lower grade. Its decrease across tumor progression might be responsible for high cortisol levels, which have been associated with higher severity and mortality [118]. UPB1 codes for the last enzyme in the pyrimidine degradation pathway, and its down-regulation might lead to dihydropyrimidine accumulation, linked with EMT [119]. Similarly, NECAP2 is amongst the top up-regulated

 doi:10.20944/preprints202105.0386.v1

genes in tumor stroma, and is also correlated with higher grade. NECAP2 plays a role in endocytic recycling [120], suggesting a potential role in regulating surface proteins and hence cell-cell communication. Interestingly, we find genes with opposite regulation in tumor epithelium and tumor stroma, which compartment-specific regulation would have not been found with bulk data: genes up-regulated in tumor stroma and down-regulated in tumor epithelium are enriched for cytokine secretion and Toll-like receptor 2 signaling, suggesting that the microenvironment produces different sets of cytokines than the tumor itself, possibly reflecting different chemoattraction and hence different immune cell proportions depending on tumor proximity. Indeed, immune cells' distribution has been found to be of clinical relevance in cancer [121–123].

To the best of our knowledge, this is the first meta-analysis performed specifically on LCM transcriptomic data. We show how different platforms can be successfully integrated to reveal robust differential expression patterns and increase statistical power, identifying differentially expressed genes that would not be identified otherwise. We note, however, that some of our results need to be cautiously interpreted, given that the definition of non-differential genes cannot be given in a statistically rigorous way. Nevertheless, this caveat applies only to a limited part of our work, and the lists DEGs in tumor epithelium, stroma and vessels that we provide are indeed robust.

From the analysis of our database we can conclude that, although the behavior of epithelium and stroma at the gene expression level is globally similar, separating compartments allows identifying gene regulation patterns that could not be detected in bulk. Moreover, integrating many different datasets, we were able to improve statistical power and, despite the small sample size of each dataset, identify genes correlated with clinical features in a compartment-specific manner.

We show additional use-cases of our database, such as the selection of epithelial and stromal markers, which better correlate with cellularity than previously published signatures, and the study of STAT3 signaling pathway regulation in tumors, resulting to be particularly relevant in tumor stroma. Specific questions can be addressed by performing the wide variety of comparisons allowed by the conditions represented in our database. A particularly interesting application will be the construction of compartment-specific gene regulatory networks. Indeed, cancer gene networks built from bulk transcriptome data are affected by the presence of multiple cell types and often include microenvironment-related gene sub-networks, confounding the identification of cancer cells' gene interactions [124]. Therefore, the use of the LCM data collected here will be a valuable resource to build more specific and robust gene co-expression network.

Finally, we make the whole database available as a resource for other researchers to explore and, to ease its interrogation, we provide a simple web platform allowing differential gene expression and enrichment analyses (https://aurorasavino.shinyapps.io/metalcm/).

5. Conclusions

By collecting and harmonizing multiple datasets of LCM breast tumors we generated a resource that can be profitably used to discover biomarkers, investigate cancer molecular mechanisms or test specific research-driven hypotheses in a robust setting. We envision several applications for our database, from the meta-analytic comparison of the several biological conditions and clinical statuses there annotated, to the construction of compartment-specific co-expression networks, hopefully helping the formulation of robust and specific research hypotheses.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Markers' expression for breast cancer PAM50 subtypes, Figure S2: The 8 classes of DEGs in bulk samples , Figure S3: Compartment-specific relationship with clinical features for genes regulated in tumor epithelium, Figure S4: HSD11B2 is correlated with lower grade in the stroma, but not in bulk, Figure S5: Summary of annotations available for each dataset , Figure S6: Summary of histological

types available for each dataset , Figure S7: Summary of possible comparisons between disease status and histological types that can be performed in each dataset , Figure S8: Summary of possible comparisons between PAM50 subtypes that can be performed in each dataset, Table S1: List of all datasets comprised in the database , Table S2: Statistics for the annotations of primary tumors in the database, Table S3: Database's metadata, Table S4: Database's expression data, Table S5: Differentially expressed genes between normal and invasive BC, Table S6: Full list of genes measured across all compared datasets, Table S7: Gene ontology categories enriched in each DEG list , Table S8: Genes belonging to the 8 classes of DEGs, Table S9: Gene ontology categories enriched in each DEG class.

Author Contributions: Conceptualization, A.S, V.P. and P.P.; methodology, A.S and P.P.; software, A.S and NdM.; formal analysis, A.S, P.P and NdM.; investigation, A.S, V.P and P.P.; resources, V.P.; data curation, A.S and NdM.; writing—original draft preparation, A.S, V.P, P.P and NdM.; writing – review & editing, A.S, V.P, P.P and NdM; visualization, A.S and NdM.; supervision, V.P. and P.P.; project administration, V.P.; funding acquisition, V.P. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by the Italian Cancer Research Association (AIRC, IG16930 and IG24851 to V.P.); the Italian Ministry of University and Research (MIUR PRIN *20175L9H7H* to V.P.); the Truus and Gerrit van Riemsdijk Foundation, Liechtenstein, donation to V.P.; Piedmont Region (Deflect F.S.E. 2014/2020). The APC was funded by AIRC, IG24851 to V.P.

Data Availability Statement: All the data used in this work are available in the Supplementary Materials, and easily accessible through a Shiny web app (https://auro-rasavino.shinyapps.io/metalcm/). Main functions used for data analyses can be retrieved from the Shiny app's code.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Detailed description of metadata categories and harmonization.

References

1. Rung, J.; Brazma, A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* **2013**, *14*, 89–99, doi:10.1038/nrg3394.
2. Kela, I.; Ein-Dor, L.; Getz, G.; Givol, D.; Domany, E. Outcome signature genes in breast cancer: is there a unique set? *Breast Cancer Res.* **2005**, *7*, P4.38, doi:10.1186/bcr1168.
3. Michiels, S.; Koscielny, S.; Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet (London, England)* **2005**, *365*, 488–492, doi:10.1016/S0140-6736(05)17866-0.
4. Shen, R.; Chinnaiyan, A.M.; Ghosh, D. Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. *BMC Med. Genomics* **2008**, *1*, 1–7, doi:10.1186/1755-8794-1-28.
5. Ioannidis, J.P.A.; Ntzani, E.E.; Trikalinos, T.A.; Contopoulos-Ioannidis, D.G. Replication validity of genetic association studies. *Nat. Genet.* **2001**, *29*, 306–309, doi:10.1038/ng749.
6. Cui, W.; Xue, H.; Wei, L.; Jin, J.; Tian, X.; Wang, Q. High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis. *Hum. Genomics* **2021**, *15*, 7, doi:10.1186/s40246-021-00308-5.
7. Shedden, K.; Taylor, J.M.G.; Enkemann, S.A.; Tsao, M.-S.; Yeatman, T.J.; Gerald, W.L.; Eschrich, S.; Jurisica, I.; Giordano, T.J.; Misek, D.E.; et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* **2008**, *14*, 822–827, doi:10.1038/nm.1790.
8. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **2013**, *41*, D991–D995, doi:10.1093/nar/gks1193.
9. Lappalainen, I.; Almeida-King, J.; Kumanduri, V.; Senf, A.; Spalding, J.D.; ur-Rehman, S.; Saunders, G.; Kandasamy, J.; Caccamo, M.; Leinonen, R.; et al. The European Genome-phenome Archive of human data consented for biomedical research. *Nat. Genet.* **2015**, *47*, 692–695, doi:10.1038/ng.3312.

10. Athar, A.; Füllgrabe, A.; George, N.; Iqbal, H.; Huerta, L.; Ali, A.; Snow, C.; Fonseca, N.A.; Petryszak, R.; Papatheodorou, I.; et al. ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res.* **2019**, *47*, D711–D715, doi:10.1093/nar/gky964.
11. Ramasamy, A.; Mondry, A.; Holmes, C.C.; Altman, D.G. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* **2008**, *5*, 1320–1332, doi:10.1371/journal.pmed.0050184.
12. Taminau, J.; Lazar, C.; Meganck, S.; Nowé, A. Comparison of Merging and Meta-Analysis as Alternative Approaches for Integrative Gene Expression Analysis. *ISRN Bioinforma.* **2014**, *2014*, 1–7, doi:10.1155/2014/345106.
13. Chang, L.C.; Lin, H.M.; Sibille, E.; Tseng, G.C. Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. *BMC Bioinformatics* **2013**, *14*, doi:10.1186/1471-2105-14-368.
14. Walsh, C.; Hu, P.; Batt, J.; Santos, C. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays* **2015**, *4*, 389–406, doi:10.3390/microarrays4030389.
15. Rhodes, D.R.; Yu, J.; Shanker, K.; Deshpande, N.; Varambally, R.; Ghosh, D.; Barrette, T.; Pandey, A.; Chinnaiyan, A.M. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9309–9314, doi:10.1073/pnas.0401994101.
16. Choi, J.K.; Yu, U.; Kim, S.; Yoo, O.J. Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* **2003**, *19 Suppl 1*, i84-90, doi:10.1093/bioinformatics/btg1010.
17. Hong, F.; Breitling, R.; McEntee, C.W.; Wittner, B.S.; Nemhauser, J.L.; Chory, J. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **2006**, *22*, 2825–2827, doi:10.1093/bioinformatics/btl476.
18. Cai, K.-T.; Feng, C.-X.; Zhao, J.-C.; He, R.-Q.; Ma, J.; Zhong, J.-C. Upregulated miR-203a-3p and its potential molecular mechanism in breast cancer: A study based on bioinformatics analyses and a comprehensive meta-analysis. *Mol. Med. Rep.* **2018**, *18*, 4994–5008, doi:10.3892/mmr.2018.9543.
19. Bozovic-Spasojevic, I.; Zardavas, D.; Brohée, S.; Ameye, L.; Fumagalli, D.; Ades, F.; de Azambuja, E.; Bareche, Y.; Piccart, M.; Paesmans, M.; et al. The Prognostic Role of Androgen Receptor in Patients with Early-Stage Breast Cancer: A Meta-analysis of Clinical and Gene Expression Data. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **2017**, *23*, 2702–2712, doi:10.1158/1078-0432.CCR-16-0979.
20. Tang, H.; Wang, S.; Xiao, G.; Schiller, J.; Papadimitrakopoulou, V.; Minna, J.; Wistuba, I.I.; Xie, Y. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. *Ann. Oncol.* **2017**, *28*, 733–740, doi:10.1093/annonc/mdw683.
21. Wang, J.; Gao, P.; Song, Y.; Sun, J.; Chen, X.; Yu, H.; Wang, Y.; Wang, Z. Prognostic value of gastric cancer-associated gene signatures: Evidence based on a meta-analysis using integrated bioinformatics methods. *J. Cell. Mol. Med.* **2018**, *22*, 5743–5747, doi:10.1111/jcmm.13823.
22. Sandhu, V.; Labori, K.J.; Borgida, A.; Lungu, I.; Bartlett, J.; Hafezi-Bakhtiari, S.; Denroche, R.E.; Jang, G.H.; Pasternack, D.; Mbaabali, F.; et al. Meta-Analysis of 1,200 Transcriptomic Profiles Identifies a Prognostic Model for Pancreatic Ductal Adenocarcinoma. *JCO Clin. cancer informatics* **2019**, *3*, 1–16, doi:10.1200/CCI.18.00102.
23. Yasrebi, H. Comparative study of joint analysis of microarray gene expression data in survival prediction and risk assessment of breast cancer patients. *Brief. Bioinform.* **2016**, *17*, 771–785, doi:10.1093/bib/bbv092.
24. Xu, L.; Tan, A.C.; Winslow, R.L.; Geman, D. Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinformatics* **2008**, *9*, doi:10.1186/1471-2105-9-125.
25. Long, N.P.; Jung, K.H.; Anh, N.H.; Yan, H.H.; Nghi, T.D.; Park, S.; Yoon, S.J.; Min, J.E.; Kim, H.M.; Lim, J.H.; et al. An integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer. *Cancers (Basel).* **2019**, *11*, doi:10.3390/cancers11020155.

26. Ma, S.; Ogino, S.; Parsana, P.; Nishihara, R.; Qian, Z.; Shen, J.; Mima, K.; Masugi, Y.; Cao, Y.; Nowak, J.A.; et al. Continuity of transcriptomes among colorectal cancer subtypes based on meta-analysis. *Genome Biol.* **2018**, *19*, 1–14, doi:10.1186/s13059-018-1511-4.
27. de Santiago, I.; Yau, C.; Heij, L.; Middleton, M.R.; Markowetz, F.; Grabsch, H.I.; Dustin, M.L.; Sivakumar, S. Immunophenotypes of pancreatic ductal adenocarcinoma: Meta-analysis of transcriptional subtypes. *Int. J. cancer* **2019**, *145*, 1125–1137, doi:10.1002/ijc.32186.
28. Qadir, F.; Lalli, A.; Dar, H.H.; Hwang, S.; Aldehlawi, H.; Ma, H.; Dai, H.; Waseem, A.; Teh, M.-T. Clinical correlation of opposing molecular signatures in head and neck squamous cell carcinoma. *BMC Cancer* **2019**, *19*, 830, doi:10.1186/s12885-019-6059-5.
29. Tan, T.Z.; Rouanne, M.; Tan, K.T.; Huang, R.Y.-J.; Thiery, J.-P. Molecular Subtypes of Urothelial Bladder Cancer: Results from a Meta-cohort Analysis of 2411 Tumors. *Eur. Urol.* **2019**, *75*, 423–432, doi:10.1016/j.eururo.2018.08.027.
30. Oztemur Islakoglu, Y.; Noyan, S.; Aydos, A.; Gur Dedeoglu, B. Meta-microRNA biomarker signatures to classify breast cancer subtypes. *Omi. A J. Integr. Biol.* **2018**, *22*, 709–716, doi:10.1089/omi.2018.0157.
31. Bell, R.; Barraclough, R.; Vasieva, O. Gene Expression Meta-Analysis of Potential Metastatic Breast Cancer Markers. *Curr. Mol. Med.* **2017**, *17*, 200–210, doi:10.2174/1566524017666170807144946.
32. Lee, Y.S.; Hwang, S.G.; Kim, J.K.; Park, T.H.; Kim, Y.R.; Myeong, H.S.; Choi, J.D.; Kwon, K.; Jang, C.S.; Ro, Y.T.; et al. Identification of novel therapeutic target genes in acquired lapatinib-resistant breast cancer by integrative meta-analysis. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **2016**, *37*, 2285–2297, doi:10.1007/s13277-015-4033-7.
33. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674, doi:10.1016/j.cell.2011.02.013.
34. Junttila, M.R.; De Sauvage, F.J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **2013**, *501*, 346–354, doi:10.1038/nature12626.
35. Plava, J.; Cihova, M.; Burikova, M.; Matuskova, M.; Kucerova, L.; Miklikova, S. Recent advances in understanding tumor stroma-mediated chemoresistance in breast cancer. *Mol. Cancer* **2019**, *18*, 1–10, doi:10.1186/s12943-019-0960-z.
36. Joyce, J.A.; Pollard, J.W. Microenvironmental regulation of metastasis. *Nat. Rev. Cancer* **2009**, *9*, 239–252, doi:10.1038/nrc2618.
37. Sahai, E.; Astsaturov, I.; Cukierman, E.; DeNardo, D.G.; Egeblad, M.; Evans, R.M.; Fearon, D.; Greten, F.R.; Hingorani, S.R.; Hunter, T.; et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev. Cancer* **2020**, *20*, 174–186, doi:10.1038/s41568-019-0238-1.
38. Lei, X.; Lei, Y.; Li, J.K.; Du, W.X.; Li, R.G.; Yang, J.; Li, J.; Li, F.; Tan, H.B. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Lett.* **2020**, *470*, 126–133, doi:10.1016/j.canlet.2019.11.009.
39. Aran, D.; Sirota, M.; Butte, A.J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **2015**, 1–12, doi:10.1038/ncomms9971.
40. Schwede, M.; Waldron, L.; Mok, S.C.; Wei, W.; Basunia, A.; Merritt, M.A.; Mitsiades, C.S.; Parmigiani, G.; Harrington, D.P.; Quackenbush, J.; et al. The impact of stroma admixture on molecular subtypes and prognostic gene signatures in serous ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **2020**, *29*, 509–519, doi:10.1158/1055-9965.EPI-18-1359.
41. Matikas, A.; Zerdes, I.; Lövrot, J.; Richard, F.; Sotiriou, C.; Bergh, J.; Valachis, A.; Foukakis, T. Prognostic implications of PD-L1 expression in breast cancer: Systematic review and meta-analysis of immunohistochemistry and pooled analysis of transcriptomic data. *Clin. Cancer Res.* **2019**, *25*, 5717–5726, doi:10.1158/1078-0432.CCR-19-1131.
42. Donnem, T.; Al-Saad, S.; Al-Shibli, K.; Delghandi, M.P.; Persson, M.; Nilsen, M.N.; Busund, L.T.; Bremnes, R.M. Inverse prognostic impact of angiogenic marker expression in tumor cells versus stromal cells in non-small cell lung cancer. *Clin. Cancer Res.* **2007**, *13*, 6649–6657, doi:10.1158/1078-0432.CCR-07-0414.
43. Sjöberg, E.; Augsten, M.; Bergh, J.; Jirström, K.; Östman, A. Expression of the chemokine CXCL14 in the tumour stroma is an independent marker of survival in breast cancer. *Br. J. Cancer* **2016**, *114*, 1117–1124, doi:10.1038/bjc.2016.104.
44. von Eggeling, F.; Hoffmann, F. Microdissection—An Essential Prerequisite for Spatial Cancer Omics. *Proteomics* **2020**, *20*,

doi:10.1002/pmic.202000077.

45. Chung, W.; Eum, H.H.; Lee, H.-O.; Lee, K.-M.; Lee, H.-B.; Kim, K.-T.; Ryu, H.S.; Kim, S.; Lee, J.E.; Park, Y.H.; et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* **2017**, *8*, 15081.
46. Bartoschek, M.; Oskolkov, N.; Bocci, M.; Lövrot, J.; Larsson, C.; Sommarin, M.; Madsen, C.D.; Lindgren, D.; Pekar, G.; Karlsson, G.; et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat. Commun.* **2018**, *9*, doi:10.1038/s41467-018-07582-3.
47. Steen, C.B.; Liu, C.L.; Alizadeh, A.A.; Newman, A.M. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol. Biol.* **2020**, *2117*, 135–157, doi:10.1007/978-1-0716-0301-7_7.
48. Chen, D.T.; Nasir, A.; Culhane, A.; Venkataramu, C.; Fulp, W.; Rubio, R.; Wang, T.; Agrawal, D.; McCarthy, S.M.; Gruidl, M.; et al. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast Cancer Res. Treat.* **2010**, *119*, 335–346, doi:10.1007/s10549-009-0344-y.
49. Casey, T.; Bond, J.; Tighe, S.; Hunter, T.; Lintault, L.; Patel, O.; Eneman, J.; Crocker, A.; White, J.; Tessitore, J.; et al. Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast Cancer Res. Treat.* **2009**, *114*, 47–62, doi:10.1007/s10549-008-9982-8.
50. Graham, K.; Ge, X.; De Las Morenas, A.; Tripathi, A.; Rosenberg, C.L. Gene expression profiles of estrogen receptor-positive and estrogen receptor-negative breast cancers are detectable in histologically normal breast epithelium. *Clin. Cancer Res.* **2011**, *17*, 236–246, doi:10.1158/1078-0432.CCR-10-1369.
51. Bauer, J.A.; Chakravarthy, A.B.; Rosenbluth, J.M.; Mi, D.; Seeley, E.H.; Granja-Ingram, N.D.M.; Olivares, M.G.; Kelley, M.C.; Mayer, I.A.; Meszoely, I.M.; et al. Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. *Clin. Cancer Res.* **2010**, *16*, 681–690, doi:10.1158/1078-0432.CCR-09-1091.
52. Planche, A.; Bacac, M.; Provero, P.; Fusco, C.; Delorenzi, M.; Stehle, J.C.; Stamenkovic, I. Identification of prognostic molecular features in the reactive stroma of human breast and prostate cancer. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0018640.
53. Lehmann, B.D.B.; Bauer, J. a J.; Chen, X.; Sanders, M.E.; Chakravarthy, a B.; Shyr, Y.; Pietenpol, J. a Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **2011**, *121*, 2750–2767, doi:10.1172/JCI45014DS1.
54. Hosein, A.N.; Livingstone, J.; Buchanan, M.; Reid, J.F.; Hallett, M.; Basik, M. A functional in vitro model of heterotypic interactions reveals a role for interferon-positive carcinoma associated fibroblasts in breast cancer. *BMC Cancer* **2015**, *15*, 1–11, doi:10.1186/s12885-015-1117-0.
55. Harvell, D.M.E.; Kim, J.; O'Brien, J.; Tan, A.C.; Borges, V.F.; Schedin, P.; Jacobsen, B.M.; Horwitz, K.B. Genomic Signatures of Pregnancy-Associated Breast Cancer Epithelia and Stroma and their Regulation by Estrogens and Progesterone. *Horm. Cancer* **2013**, *4*, 140–153, doi:10.1007/s12672-013-0136-z.
56. Knudsen, E.S.; Ertel, A.; Davicioni, E.; Kline, J.; Schwartz, G.F.; Witkiewicz, A.K. Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. *Breast Cancer Res. Treat.* **2012**, *133*, 1009–1024, doi:10.1007/s10549-011-1894-3.
57. Vargas, A.C.; Reed, A.E.M.C.; Waddell, N.; Lane, A.; Reid, L.E.; Smart, C.E.; Cocciardi, S.; Da Silva, L.; Song, S.; Chenevix-Trench, G.; et al. Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res. Treat.* **2012**, *135*, 153–165, doi:10.1007/s10549-012-2123-4.
58. Tchou, J.; Kossenkov, A. V.; Chang, L.; Satija, C.; Herlyn, M.; Showe, L.C.; Puré, E. Human breast cancer associated fibroblasts exhibit subtype specific gene expression profiles. *BMC Med. Genomics* **2012**, *5*, doi:10.1186/1755-8794-5-39.
59. Komatsu, M.; Yoshimaru, T.; Matsuo, T.; Kiyotani, K.; Miyoshi, Y.; Tanahashi, T.; Rokutan, K.; Yamaguchi, R.; Saito, A.; Imoto, S.; et al. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. *Int. J. Oncol.* **2013**, *42*, 478–506, doi:10.3892/ijo.2012.1744.

60. Santucci-Pereira, J.; Zeleniuch-Jacquotte, A.; Afanasyeva, Y.; Zhong, H.; Slifker, M.; Peri, S.; Ross, E.A.; López De Cicco, R.; Zhai, Y.; Nguyen, T.; et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* **2019**, *21*, 1–19, doi:10.1186/s13058-019-1128-x.
61. Lee, S.; Stewart, S.; Nagtegaal, I.; Luo, J.; Wu, Y.; Colditz, G.; Medina, D.; Allred, D.C. Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. *Cancer Res.* **2012**, *72*, 4574–4586, doi:10.1158/0008-5472.CAN-12-0636.
62. Woodward, W.A.; Krishnamurthy, S.; Yamauchi, H.; El-Zein, R.; Ogura, D.; Kitadai, E.; Niwa, S.-I.; Cristofanilli, M.; Vermeulen, P.; Dirix, L.; et al. Genomic and expression analysis of microdissected inflammatory breast cancer. *Breast Cancer Res. Treat.* **2013**, *138*, 761–772, doi:10.1007/s10549-013-2501-6.
63. Finak, G.; Sadekova, S.; Pepin, F.; Hallett, M.; Meterissian, S.; Halwani, F.; Khetani, K.; Souleimanova, M.; Zabolotny, B.; Omeroglu, A.; et al. Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. *Breast Cancer Res.* **2006**, *8*, 1–18, doi:10.1186/bcr1608.
64. Turashvili, G.; Bouchal, J.; Baumforth, K.; Wei, W.; Dziechciarkova, M.; Ehrmann, J.; Klein, J.; Fridman, E.; Skarda, J.; Srovnal, J.; et al. Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* **2007**, *7*, 1–20, doi:10.1186/1471-2407-7-55.
65. Reyngold, M.; Turcan, S.; Giri, D.; Kannan, K.; Walsh, L.A.; Viale, A.; Drobnjak, M.; Vahdat, L.T.; Lee, W.; Chan, T.A. Remodeling of the methylation landscape in breast cancer metastasis. *PLoS One* **2014**, *9*, 1–10, doi:10.1371/journal.pone.0103896.
66. Boersma, B.J.; Reimers, M.; Yi, M.; Ludwig, J.A.; Luke, B.T.; Stephens, R.M.; Yfantis, H.G.; Lee, D.H.; Weinstein, J.N.; Ambs, S. A stromal gene signature associated with inflammatory breast cancer. *Int. J. Cancer* **2008**, *122*, 1324–1332, doi:10.1002/ijc.23237.
67. Winslow, S.; Leandersson, K.; Edsjö, A.; Larsson, C. Prognostic stromal gene signatures in breast cancer. *Breast Cancer Res.* **2015**, *17*, 1–13, doi:10.1186/s13058-015-0530-2.
68. Oh, E.Y.; Christensen, S.M.; Ghanta, S.; Jeong, J.C.; Bucur, O.; Glass, B.; Montaser-Kouhsari, L.; Knoblauch, N.W.; Bertos, N.; Saleh, S.M.I.; et al. Extensive rewiring of epithelial-stromal co-expression networks in breast cancer. *Genome Biol.* **2015**, *16*, doi:10.1186/s13059-015-0675-4.
69. Abdalla, M.; Tran-Thanh, D.; Moreno, J.; Iakovlev, V.; Nair, R.; Kanwar, N.; Abdalla, M.; Lee, J.P.Y.; Kwan, J.Y.Y.; Cawthorn, T.R.; et al. Mapping genomic and transcriptomic alterations spatially in epithelial cells adjacent to human breast carcinoma. *Nat. Commun.* **2017**, *8*, doi:10.1038/s41467-017-01357-y.
70. Lee, S.; Medina, D.; Tsimelzon, A.; Mohsin, S.K.; Mao, S.; Wu, Y.; Allred, D.C. Alterations of gene expression in the development of early hyperplastic precursors of breast cancer. *Am. J. Pathol.* **2007**, *171*, 252–262, doi:10.2353/ajpath.2007.061010.
71. Gregory, K.J.; Roberts, A.L.; Conlon, E.M.; Mayfield, J.A.; Hagen, M.J.; Crisi, G.M.; Bentley, B.A.; Kane, J.J.; Makari-Judson, G.; Mason, H.S.; et al. Gene expression signature of atypical breast hyperplasia and regulation by SFRP1. *Breast Cancer Res.* **2019**, *21*, 1–18, doi:10.1186/s13058-019-1157-5.
72. Lehmann, B.D.; Jovanović, B.; Chen, X.; Estrada, M. V.; Johnson, K.N.; Shyr, Y.; Moses, H.L.; Sanders, M.E.; Pietenpol, J.A. Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. *PLoS One* **2016**, *11*, 1–22, doi:10.1371/journal.pone.0157368.
73. Liu, H.; Dowdle, J.A.; Khurshid, S.; Sullivan, N.J.; Bertos, N.; Rambani, K.; Mair, M.; Daniel, P.; Wheeler, E.; Tang, X.; et al. Discovery of Stromal Regulatory Networks that Suppress Ras-Sensitized Epithelial Cell Proliferation. *Dev. Cell* **2017**, *41*, 392-407.e6, doi:10.1016/j.devcel.2017.04.024.
74. Gruosso, T.; Gigoux, M.; Manem, V.S.K.; Bertos, N.; Zuo, D.; Perlitch, I.; Saleh, S.M.I.; Zhao, H.; Souleimanova, M.; Johnson, R.M.; et al. Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. *J. Clin. Invest.* **2019**,

129, 1785–1800, doi:10.1172/JCI96313.

75. Karnoub, A.E.; Dash, A.B.; Vo, A.P.; Sullivan, A.; Brooks, M.W.; Bell, G.W.; Richardson, A.L.; Polyak, K.; Tubo, R.; Weinberg, R.A. Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* **2007**, *449*, 557–563, doi:10.1038/nature06188.
76. Saleh, S.M.I.; Bertos, N.; Gruosso, T.; Gigoux, M.; Souleimanova, M.; Zhao, H.; Omeroglu, A.; Hallett, M.T.; Park, M. Identification of interacting stromal axes in triple-negative breast cancer. *Cancer Res.* **2017**, *77*, 4673–4683, doi:10.1158/0008-5472.CAN-16-3427.
77. Brouwers, B.; Fumagalli, D.; Brohee, S.; Hatse, S.; Govaere, O.; Floris, G.; Van den Eynde, K.; Bareche, Y.; Schöffski, P.; Smeets, A.; et al. The footprint of the ageing stroma in older patients with breast cancer. *Breast Cancer Res.* **2017**, *19*, 1–14, doi:10.1186/s13058-017-0871-0.
78. Tripathi, A.; King, C.; De La Morenas, A.; Perry, V.K.; Burke, B.; Antoine, G.A.; Hirsch, E.F.; Kavanah, M.; Mendez, J.; Stone, M.; et al. Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *Int. J. Cancer* **2008**, *122*, 1557–1566, doi:10.1002/ijc.23267.
79. Jones, D.T.; Lechertier, T.; Mitter, R.; Herbert, J.M.J.; Bicknell, R.; Jones, J.L.; Li, J.L.; Buffa, F.; Harris, A.L.; Hodivala-Dilke, K. Gene Expression Analysis in Human Breast Cancer Associated Blood Vessels. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0044294.
80. François, P.; Bertos, N.; Laferrière, J.; Sadekova, S.; Souleimanova, M.; Zhao, H.; Finak, G.; Meterissian, S.; Hallett, M.T.; Park, M. Gene-expression profiling of microdissected breast cancer microvasculature identifies distinct tumor vascular subtypes. *Breast Cancer Res.* **2012**, *14*, doi:10.1186/bcr3246.
81. Le, P.U.; Li, J.; Cantin, C.; Collins, C.; Nantel, A.; Tremblay, T.; Kelly, J.F.; Wang, E.; Connor-mccourt, M.D.O.; Eg, A. Molecular Characterization of the Vasculature of Patients with Infiltrating Ductal Carcinoma Generates a Gene Signature Predictive of Breast Cancer Survival. **2015**, *3*.
82. Russo, J.; Balogh, G.A.; Russo, I.H. Full-term pregnancy induces a specific genomic signature in the human breast. *Cancer Epidemiol. Biomarkers Prev.* **2008**, *17*, 51–66, doi:10.1158/1055-9965.EPI-07-0678.
83. Dhage, S.; Ernlund, A.; Ruggles, K.; Axelrod, D.; Berman, R.; Roses, D.; Schneider, R.J. A genomic ruler to assess oncogenic transition between breast tumor and stroma. *PLoS One* **2018**, *13*, 1–20, doi:10.1371/journal.pone.0205602.
84. Ershaid, N.; Sharon, Y.; Doron, H.; Raz, Y.; Shani, O.; Cohen, N.; Monteran, L.; Leider-Trejo, L.; Ben-Shmuel, A.; Yassin, M.; et al. NLRP3 inflammasome in fibroblasts links tissue damage with inflammation in breast cancer progression and metastasis. *Nat. Commun.* **2019**, *10*, doi:10.1038/s41467-019-12370-8.
85. Bauer, M.; Su, G.; Casper, C.; He, R.; Rehrauer, W.; Friedl, A. Heterogeneity of gene expression in stromal fibroblasts of human breast carcinomas and normal breast. *Oncogene* **2010**, *29*, 1732–1740, doi:10.1038/onc.2009.463.
86. Bianchini, G.; Qi, Y.; Alvarez, R.H.; Iwamoto, T.; Coutant, C.; Ibrahim, N.K.; Valero, V.; Cristofanilli, M.; Green, M.C.; Radvanyi, L.; et al. Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and -negative cancers. *J. Clin. Oncol.* **2010**, *28*, 4316–4323, doi:10.1200/JCO.2009.27.2419.
87. Calvo, F.; Ege, N.; Grande-Garcia, A.; Hooper, S.; Jenkins, R.P.; Chaudhry, S.I.; Harrington, K.; Williamson, P.; Moeendarbary, E.; Charras, G.; et al. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat. Cell Biol.* **2013**, *15*, 637–646, doi:10.1038/ncb2756.
88. Ma, X.J.; Wang, Z.; Ryan, P.D.; Isakoff, S.J.; Barmettler, A.; Fuller, A.; Muir, B.; Mohapatra, G.; Salunga, R.; Tuggle, J.T.; et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **2004**, *5*, 607–616, doi:10.1016/j.ccr.2004.05.015.
89. Marino, N.; German, R.; Rao, X.; Simpson, E.; Liu, S.; Wan, J.; Liu, Y.; Sandusky, G.; Jacobsen, M.; Stoval, M.; et al. Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. *npj Breast Cancer* **2020**, *6*, 1–13, doi:10.1038/s41523-020-00191-8.

90. Ma, X.-J.; Dahiya, S.; Richardson, E.; Erlander, M.; Sgroi, D.C. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res.* **2009**, *11*, R7, doi:10.1186/bcr2222.
91. Emery, L.A.; Tripathi, A.; King, C.; Kavanah, M.; Mendez, J.; Stone, M.D.; De Las Morenas, A.; Sebastiani, P.; Rosenberg, C.L. Early dysregulation of cell adhesion and extracellular matrix pathways in breast cancer progression. *Am. J. Pathol.* **2009**, *175*, 1292–1302, doi:10.2353/ajpath.2009.090115.
92. Graham, K.; De Las Morenas, A.; Tripathi, A.; King, C.; Kavanah, M.; Mendez, J.; Stone, M.; Slama, J.; Miller, M.; Antoine, G.; et al. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br. J. Cancer* **2010**, *102*, 1284–1293, doi:10.1038/sj.bjc.6605576.
93. Davis, S.; Meltzer, P.S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **2007**, *23*, 1846–1847, doi:10.1093/bioinformatics/btm254.
94. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71–e71, doi:10.1093/nar/gkv1507.
95. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–91, doi:10.1038/nprot.2009.97.
96. Fisher, R.A. Statistical methods for research workers. In *Breakthroughs in statistics*; Springer, 1992; pp. 66–70.
97. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. a J. Integr. Biol.* **2012**, *16*, 284–287.
98. Hänzelmann, S.; Castelo, R.; Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **2013**, *14*, 7, doi:10.1186/1471-2105-14-7.
99. R Core Team R: A language and environment for statistical computing 2018.
100. Wickham, H. ggplot2 – Elegant Graphics for Data Analysis. *J. Stat. Softw. April* **2017**, *77*, 3–5, doi:10.18637/jss.v077.b02.
101. Ganz, C. rintrojs: A Wrapper for the Intro.js Library. *J. Open Source Softw.* **2016**, *1*, 63, doi:10.21105/joss.00063.
102. Azare, J.; Leslie, K.; Al-Ahmadie, H.; Gerald, W.; Weinreb, P.H.; Violette, S.M.; Bromberg, J. Constitutively activated Stat3 induces tumorigenesis and enhances cell motility of prostate epithelial cells through integrin beta 6. *Mol. Cell. Biol.* **2007**, *27*, 4444–4453, doi:10.1128/MCB.02404-06.
103. Dauer, D.J.; Ferraro, B.; Song, L.; Yu, B.; Mora, L.; Buettner, R.; Enkemann, S.; Jove, R.; Haura, E.B. Stat3 regulates genes common to both wound healing and cancer. *Oncogene* **2005**, *24*, 3397–3408, doi:10.1038/sj.onc.1208469.
104. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425, doi:10.1016/j.cels.2015.12.004.
105. Alvarez, J. V; Febbo, P.G.; Ramaswamy, S.; Loda, M.; Richardson, A.; Frank, D.A. Identification of a Genetic Signature of Activated Signal Transducer and Activator of Transcription 3 in Human Tumors. *Cancer Res.* **2005**, *65*, 5054–5063.
106. Tell, R.W.; Horvath, C.M. Bioinformatic analysis reveals a pattern of STAT3-associated gene expression specific to basal-like breast cancers in human tumors. *Proc Natl Acad Sci U S A* **2014**, *111*, 12787–12792, doi:10.1073/pnas.1404881111.
107. Sonnenblick, A.; Brohée, S.; Fumagalli, D.; Vincent, D.; Venet, D.; Ignatiadis, M.; Salgado, R.; Van den Eynden, G.; Rothé, F.; Desmedt, C.; et al. Constitutive phosphorylated STAT3-associated gene signature is predictive for trastuzumab resistance in primary HER2-positive breast cancer. *BMC Med.* **2015**, *13*, 177, doi:10.1186/s12916-015-0416-2.
108. Network, C.G.A. Comprehensive molecular portraits of human breast tumours. *Nature* **2012**, *490*, 61–70, doi:10.1038/nature11412.
109. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Treviño, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, doi:10.1038/ncomms3612.
110. Klemm, F.; Bleckmann, A.; Siam, L.; Chuang, H.N.; Rietktter, E.; Behme, D.; Schulz, M.; Schaffrinski, M.; Schindler, S.;

Trümper, L.; et al. β-catenin-independent WNT signaling in basal-like breast cancer and brain metastasis. *Carcinogenesis* **2011**, *32*, 434–442, doi:10.1093/carcin/bgq269.

111. Monteleone, E.; Orecchia, V.; Corrieri, P.; Schiavone, D.; Avalle, L.; Moiso, E.; Savino, A.; Molineris, I.; Provero, P.; Poli, V. SP1 and STAT3 Functionally Synergize to Induce the RhoU Small GTPase and a Subclass of Non-canonical WNT Responsive Genes Correlating with Poor Prognosis in Breast Cancer. *Cancers (Basel).* **2019**, *11*, 101–116, doi:10.3390/cancers11010101.
112. Lopes-Coelho, F.; André, S.; Félix, A.; Serpa, J. Breast cancer metabolic cross-talk: Fibroblasts are hubs and breast cancer cells are gatherers of lipids. *Mol. Cell. Endocrinol.* **2018**, *462*, 93–106, doi:10.1016/j.mce.2017.01.031.
113. Yuzhalin, A.E.; Lim, S.Y.; Kutikhin, A.G.; Gordon-Weeks, A.N. Dynamic matrisome: ECM remodeling factors licensing cancer progression and metastasis. *Biochim. Biophys. Acta - Rev. Cancer* **2018**, *1870*, 207–228, doi:10.1016/j.bbcan.2018.09.002.
114. Ashton, T.M.; Gillies McKenna, W.; Kunz-Schughart, L.A.; Higgins, G.S. Oxidative phosphorylation as an emerging target in cancer therapy. *Clin. Cancer Res.* **2018**, *24*, 2482–2490, doi:10.1158/1078-0432.CCR-17-3070.
115. Davis, R.T.; Blake, K.; Ma, D.; Gabra, M.B.I.; Hernandez, G.A.; Phung, A.T.; Yang, Y.; Maurer, D.; Lefebvre, A.E.Y.T.; Alshetaiwi, H.; et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. *Nat. Cell Biol.* **2020**, *22*, 310–320, doi:10.1038/s41556-020-0477-0.
116. Boneberg, E.-M.; Legler, D.F.; Hoefer, M.M.; Ohlschlegel, C.; Steininger, H.; Füzesi, L.; Beer, G.M.; Dupont-Lampert, V.; Otto, F.; Senn, H.-J.; et al. Angiogenesis and lymphangiogenesis are downregulated in primary breast cancer. *Br. J. Cancer* **2009**, *101*, 605–614, doi:10.1038/sj.bjc.6605219.
117. Eberhard, A.; Kahlert, S.; Goede, V.; Hemmerlein, B.; Plate, K.H.; Augustin, H.G. Heterogeneity of angiogenesis and blood vessel maturation in human tumors: implications for antiangiogenic tumor therapies. *Cancer Res.* **2000**, *60*, 1388–1393.
118. Al Sorkhy, M.; Fahl, Z.; Ritchie, J. Cortisol and breast cancer: A review of clinical and molecular evidence. *Ann. Cancer Res. Ther.* **2018**, *26*, 19–25, doi:10.4993/acrt.26.19.
119. Shaul, Y.D.; Freinkman, E.; Comb, W.C.; Cantor, J.R.; Tam, W.L.; Thiru, P.; Kim, D.; Kanarek, N.; Pacold, M.E.; Chen, W.W.; et al. Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* **2014**, *158*, 1094–1109, doi:10.1016/j.cell.2014.07.032.
120. Chamberland, J.P.; Antonow, L.T.; Dias Santos, M.; Ritter, B. NECAP2 controls clathrin coat recruitment to early endosomes for fast endocytic recycling. *J. Cell Sci.* **2016**, *129*, 2625–2637, doi:10.1242/jcs.173708.
121. Zheng, X.; Weigert, A.; Reu, S.; Guenther, S.; Mansouri, S.; Bassaly, B.; Gattenlöhner, S.; Grimminger, F.; Pullamsetti, S.S.; Seeger, W.; et al. Spatial density and distribution of tumor-associated macrophages predict survival in non-small cell lung carcinoma. *Cancer Res.* **2020**, *80*, 4414–4425, doi:10.1158/0008-5472.CAN-20-0069.
122. Medrek, C.; Pontén, F.; Jirström, K.; Leandersson, K. The presence of tumor associated macrophages in tumor stroma as a prognostic marker for breast cancer patients. *BMC Cancer* **2012**, *12*, 1–9, doi:10.1186/1471-2407-12-306.
123. Jackute, J.; Zemaitis, M.; Pranys, D.; Sitkauskiene, B.; Miliauskas, S.; Vaitkiene, S.; Sakalauskas, R. Distribution of M1 and M2 macrophages in tumor islets and stroma in relation to prognosis of non-small cell lung cancer. *BMC Immunol.* **2018**, *19*, 1–13, doi:10.1186/s12865-018-0241-4.
124. Savino, A.; Provero, P.; Poli, V. Differential co-expression analyses allow the identification of critical signalling pathways altered during tumour transformation and progression. *Int. J. Mol. Sci.* **2020**, *21*, 1–23, doi:10.3390/ijms21249461.