Evolution of metastases-associated fibroblasts in the lung microenvironment is driven by stage-specific transcriptional plasticity

Ophir Shani[1*], Yael Raz[1,2*], Or Megides[3], Hila Shacham[3], Noam Cohen[1], Dana Silverbush[4], Lea Monteran[1], Roded Sharan[4], Ilan Tsarfaty[3] and Neta Erez[1#].

[1]Department of Pathology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.
[2]Department of Obstetrics and Gynecology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel.
[3]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.
[4]Balavatnik School of Computer Science, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel.
* These two authors contributed equally to this study.

#Corresponding Author contact information:
Neta Erez, Department of Pathology, Sackler Faculty of Medicine,
Tel Aviv University, Tel Aviv 69978, Israel
Email: netaerez@tauex.tau.ac.il , Tel: +972-3-6408689

Running Title: Evolution of metastases-associated fibroblasts

Summary

Mortality from breast cancer is almost exclusively a result of tumor metastasis, and lungs are one of the main metastatic sites. Cancer-associated fibroblasts (CAFs) are prominent players in the microenvironment of breast cancer. However, their role in the metastatic niche is largely unknown. In this study, we profiled the transcriptional co-evolution of lung fibroblasts isolated from transgenic mice at defined stage-specific time points of metastases formation. Employing multiple knowledge-based platforms of data analysis provided powerful insights on functional and temporal regulation of the transcriptome of fibroblasts. We demonstrate that fibroblasts in lung metastases are transcriptionally dynamic and plastic, and reveal stage-specific gene signatures that imply functional tasks, including extracellular matrix remodeling, stress response and shaping the inflammatory microenvironment. Furthermore, we identified *Myc* as a central regulator of fibroblast rewiring and found that stromal upregulation of *Myc* transcriptional networks is associated with worse survival in human breast cancer.

Keywords
Brest cancer, metastasis, fibroblasts, microenvironment.

Introduction

Breast cancer continues to be one of the leading causes of cancer related death in women, and mortality is almost exclusively a result of tumor metastasis. Advanced metastatic cancers are mostly incurable and available therapies generally prolong life to a limited extent. It is increasingly appreciated that in addition to tumor cell-intrinsic survival and growth programs, the microenvironment is crucial in supporting metastases formation (Erez and Coussens, 2011; Obenauf and Massague, 2015; Quail and Joyce, 2013). Nevertheless, while years of research have revealed the central role of the microenvironment in supporting tumor growth and response to therapy at the primary tumor site (Albini et al., 2018; Hanahan and Coussens, 2012; Quail and Joyce, 2013), the role of the metastatic microenvironment and the molecular crosstalk between stromal cells, including fibroblasts and immune cells at the metastatic niche are poorly characterized.

Preparation of secondary sites to facilitate subsequent tumor cell colonization has been described for multiple cancers (Peinado et al., 2017). Secreted factors and extracellular vesicles from tumor and stromal cells were reported to instigate a permissive pre-metastatic niche by influencing the recruitment and activation of immune cells (Coffelt et al., 2015; Deng et al., 2012; Peinado et al., 2011; Qian et al., 2011; Quail et al., 2017), and by modifying the composition of the extracellular matrix (ECM) (Erler et al., 2009; Malanchi et al., 2012; Nielsen et al., 2016; Oskarsson et al., 2011; Oskarsson and Massague, 2012). Each metastatic microenvironment exerts specific functions that support or oppose colonization by disseminated tumor cells (Nguyen et al., 2009; Peinado et al., 2017). Therefore, understanding distinct organ-specific mechanisms that enable metastatic growth is of crucial importance.

Lungs are one of the most common sites of breast cancer metastasis. Various immune cell populations were shown to be functionally important in facilitating breast cancer pulmonary metastasis (Albrengues et al., 2018; Coffelt et al., 2015; DeNardo et al., 2009; Fridlender et al., 2015; Jablonska et al., 2017). However, very little is known about the role of fibroblasts during the complex process of metastases formation.
Cancer-associated fibroblasts (CAFs) are a heterogeneous population of fibroblastic cells found in the microenvironment of solid tumors. In some cancer types, including breast carcinomas, CAFs are the most prominent stromal cell type, and their abundance correlates with worse prognosis (Liu et al., 2016). We previously demonstrated a novel role for CAFs in mediating tumor-promoting inflammation in mouse and human carcinomas (Erez et al., 2013; Erez et al., 2010). We further characterized the origin, heterogeneity and function of CAFs in breast cancer (Cohen et al., 2017; Raz et al., 2018; Sharon et al., 2015). Importantly, we found profound changes in the expression of pro-inflammatory genes in fibroblasts isolated from metastases-bearing lungs (Raz et al., 2018). However, comprehensive profiling of metastases-associated fibroblasts was not previously done. Based on the central role of CAFs in supporting tumor growth at the primary tumor site (Kalluri, 2016), we hypothesized that transcriptional reprogramming of lung fibroblasts is an important factor in the formation of a hospitable metastatic niche that supports breast cancer metastasis.

In this study, we set out to characterize the dynamic co-evolution of fibroblasts during pulmonary metastasis. To achieve this goal, we utilized novel transgenic mice that enable visualization, tracking, and unbiased isolation of fibroblasts from spontaneous lung metastases. Here we demonstrate the profiling and analysis of the dynamic evolution of fibroblast transcriptome at distinct disease stages, including early and late metastatic disease.

Results

Fibroblasts are activated in the lung metastatic niche and acquire tumor promoting traits

We previously demonstrated that fibroblasts at the primary tumor microenvironment are reprogrammed to obtain a pro-inflammatory and tumor-promoting phenotype (Cohen et al., 2017; Erez et al., 2010; Sharon et al., 2015). Moreover, we found that fibroblasts are also modified at the lung metastatic niche (Raz et al., 2018). In this study, we set out to characterize the changes in lung fibroblasts that mediate the formation of a hospitable niche in breast cancer metastasis.

Fibroblasts at the primary tumor site were previously demonstrated to be reprogrammed by tumor cell-derived paracrine signaling (Jin et al., 2017; Sharon et al., 2015). Therefore, we initially asked whether fibroblasts at the metastatic microenvironment are activated in response to tumor-secreted factors. We isolated primary lung fibroblasts from lungs of normal female mice (normal lung fibroblasts-NLF), incubated them with conditioned media (CM) from Met-1, a PyMT-derived breast carcinoma cell line (Borowsky et al., 2005), and analyzed the effect of tumor-derived factors on multiple tasks associated with activation of fibroblasts. One of the central physiological roles of activated myofibroblasts is to facilitate wound healing, where their motility and contractility are crucial for wound closure (Rouabhia et al., 2013). We therefore performed an *in vitro* wound closure assay. Quantification of the wound area indicated that fibroblasts incubated with tumor cell secreted factors had significantly enhanced motility and scratch closure ability as compared with control fibroblasts (Figure 1A, B). Activated myofibroblasts are known to have enhanced matrix adhesions, leading to increased contraction of collagen gel matrices (Cooke et al., 2000). Thus, we next analyzed the effect of tumor-secreted factors on the ability of lung fibroblasts to mediate collagen contraction. Fibroblasts were suspended in collagen matrices and incubated with Met-1 conditioned media or in serum-free conditions. We found that tumor-secreted factors significantly enhanced the capacity of NLF to contract collagen gels as compared with control non-activated fibroblasts (Figure 1C, D). One of the prerequisites for metastatic progression is the ability of tumor cells to migrate to distant sites (Nguyen et al., 2009). Therefore, we next examined whether reprogrammed lung fibroblasts can enhance tumor cell migration. To that end, we incubated tumor cells with secreted factors derived from activated lung fibroblasts (ALF CM) or with normal controls (NLF CM). Transwell migration assays indicated that tumor cells incubated with ALF-derived factors exhibited increased migratory capacity as compared with controls (Figure 1E, F). Thus, normal lung fibroblasts are reprogrammed by signaling from breast cancer cells, resulting in acquisition of tumor-promoting properties.

We next set out to analyze the activation of metastases-associated fibroblasts *in vivo*. To that end, we utilized MMTV-PyMT mice, a transgenic mouse model of mammary carcinogenesis and spontaneous lung metastasis (Lin et al., 2003). We initially assessed the expression of αSMA, a known marker of activated fibroblasts (Kalluri and Zeisberg, 2006), in normal lungs and in metastases-bearing lungs. We found that the abundance of αSMA expressing fibroblasts was significantly increased in metastatic lungs as compared with normal lungs (Figure 1G, H). In order to be able to track fibroblasts in an unbiased manner, we employed a transgenic mouse model of genetically labeled fibroblasts (Raz et al., 2018). Flow cytometric analysis of normal lungs as compared with lungs of tumor-bearing mice revealed significantly increased numbers of fibroblasts in the metastatic lung (Figure 1I), indicating that activation of fibroblasts at the metastatic niche includes changes in their abundance, as well as in their activation state. Taken together, these results demonstrate that breast cancer lung metastasis is associated with fibroblast activation in the lung metastatic microenvironment.

Shani et al. Figure 1

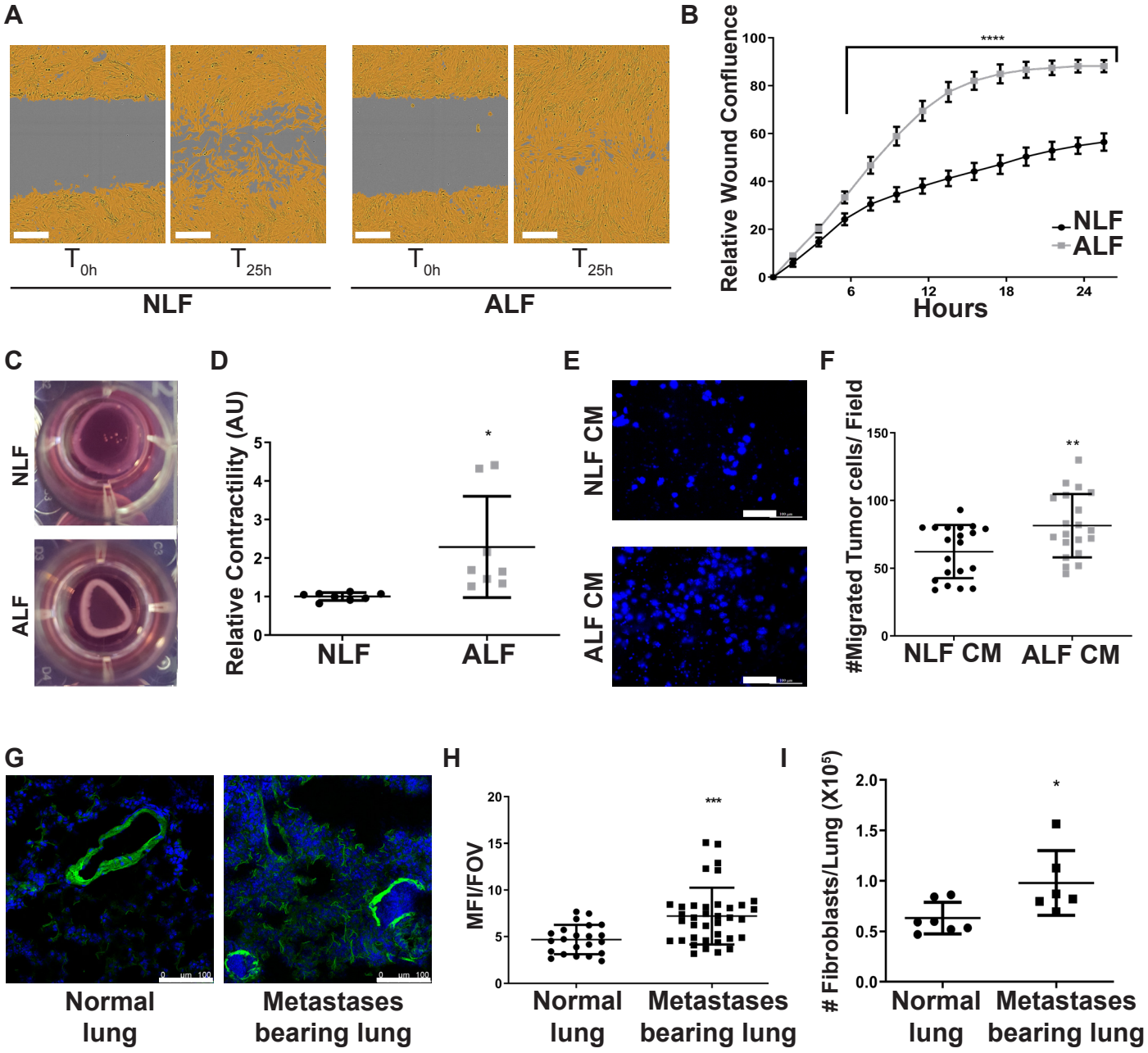


Figure 1. Fibroblasts are activated in the lung metastatic niche and acquire tumor promoting traits. (A) Representative images of scratch closure assay at 0 hours and 24 hours following scratch. Lung fibroblasts were incubated with SFM (NLF-normal lung fibroblasts), or with Met-1 CM (ALF-activated lung fibroblasts), scale bar: 300μM. **(B)** Quantification of scratch closure shown in (A), ****p<0.0001, Two-way ANOVA multiple comparisons, Data are represented as mean ± SD, n=3. **(C)** Representative images of collagen contraction assay at 24 hours. Lung fibroblasts were embedded in collagen gel and incubated with SFM (NLF) or in Met-1 CM (ALF). **(D)** Quantification of collagen contraction shown in (C), *p<0.05, Data are represented as mean ± SD, n=8. **(E)** Representative images of transwell migration assay of Met-1 tumor cells incubated with NLF CM or ALF CM. scale bar: 100μM. **(F)** Quantification of (E). Migrated tumor cells were counted using ImageJ in multiple fields of view (FOV). **p<0.01, Data are represented as mean ± SD, n=3. **(G)** Representative images of αSMA staining in normal lungs isolated from FVB/n mice, (n=3), and metastases-bearing lungs isolated from MMTV-PyMT mice (n=3). Scale bar: 100μM **(H)** Quantification of αSMA mean fluorescent intensity (MFI) in multiple fields of view (FOV) from 3 slides per mouse. ***p<0.001. Data are represented as mean ± SD **(I)** Number of fibroblasts (CD45-CD31-EpCAM-YFP+) in total lungs from FVB/n;*Col1a1*-YFP mice (n=7) and PyMT; *Col1a1*-YFP mice (n=6) were counted using flow cytometric analysis. *p<0.05. Data are represented as mean ± SD.

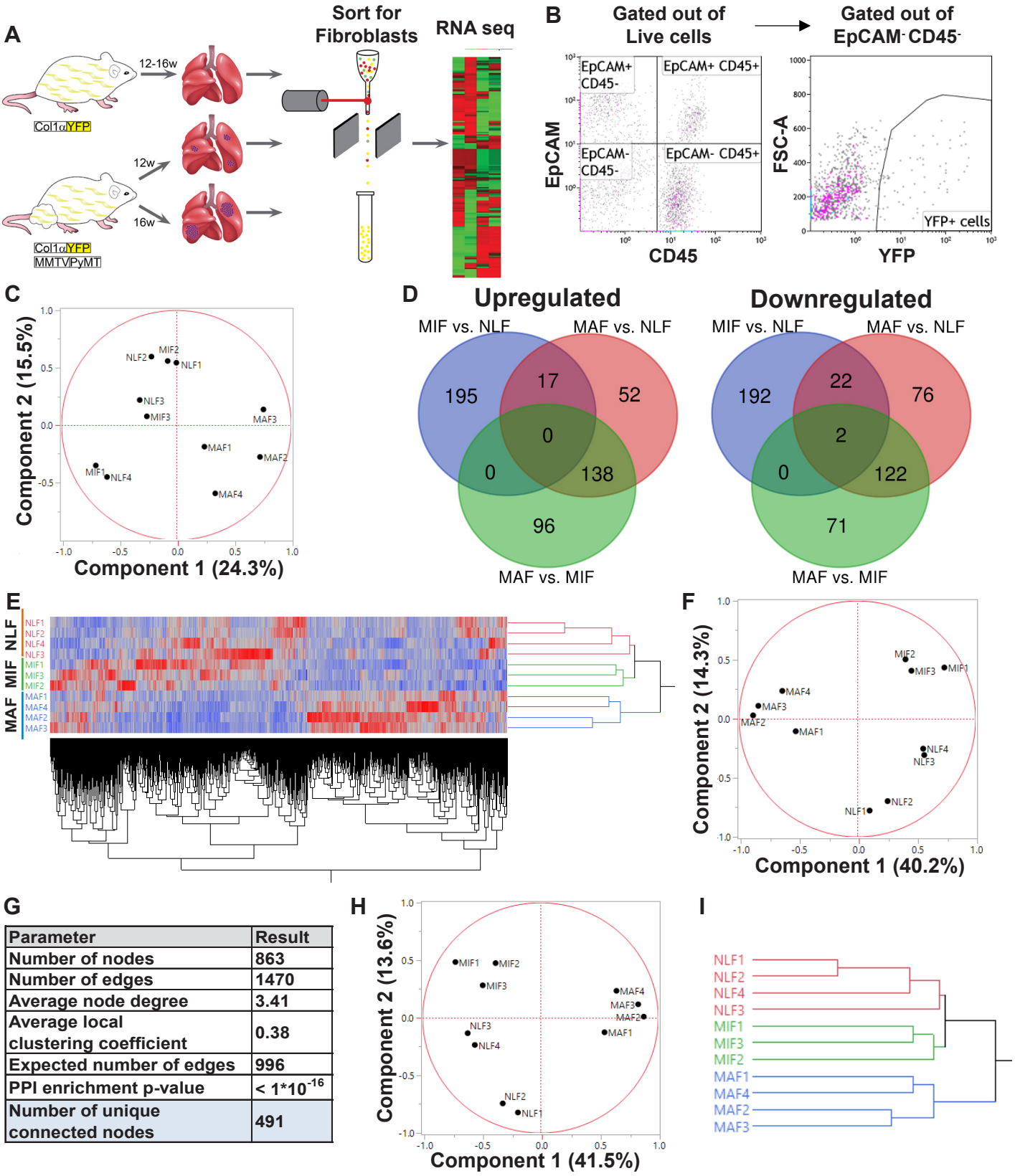
Transcriptome profiling of metastases-associated fibroblasts reveals dynamic stage-specific changes in gene expression.

To analyze the dynamic co-evolution of fibroblasts at the lung metastatic niche we performed unbiased transcriptome profiling by RNA-seq of fibroblasts isolated from lungs at distinct metastatic stages. To enable visualization, tracking, and isolation of fibroblasts, we established transgenic mouse models of breast cancer with fibroblast-specific reporter genes: transgenic mice that express the fluorescent reporter YFP under the Collagen-1α promoter (*Col1a1*-YFP) were crossed with MMTV-PyMT mice to create PyMT*;Col1a1*-YFP transgenic mice, in which all fibroblasts are fluorescently labeled (Raz et al., 2018). To explore the temporal dynamic changes in functional gene networks, we profiled fibroblasts ($EpCAM^-CD45^-YFP^+$ cells) isolated from normal lungs, and from lungs with micro- or macrometastases (Figure 2A, B). Micrometastases were defined as the latest time point in which metastases are not detectable macroscopically, or by CT imaging.

Initial data analysis indicated that fibroblasts isolated from lungs with macrometastases (macrometastases-associated fibroblasts- MAF) were strikingly different from NLF as well as from fibroblasts isolated from lungs with micrometastases (micrometastases-associated fibroblasts- MIF) (Figure 2C, Figure S1). Notably, since fibroblasts were isolated from entire lungs, rather than from specific metastatic niches, the MIF fraction contained also normal, non-metastases-associated fibroblasts. As a result, initial data analysis did not reveal significant differences between NLF and MIF (Figure 2C, Figure S1). To focus on the genes that are differentially expressed between the different stages, we selected the top 2% most upregulated and downregulated genes in each comparison (NLF vs. MIF, NLF vs. MAF, MIF vs. MAF), yielding a total of 871 genes. Notably, Venn Diagram analysis showed that only a small number of these genes were shared across the different stages, suggesting that each stage is defined by its own specific gene signature (Figure 2D). Unsupervised methods including Principal Component Analysis (PCA) and hierarchical clustering validated that each metastatic stage clustered separately and was defined by a unique gene signature (Figure 2E, F). Thus, although the transcriptional changes in the fibroblasts isolated from micro-metastases may have been masked by the presence of NLF in this fraction, analyses confirmed that MIF represent a bona fide functionally distinct stage.

Aiming to delineate the stage-specific gene signatures and the molecular mechanisms operative in metastases-associated fibroblasts, and to identify the most relevant functional pathways, we performed protein-protein interaction (PPI) analysis using the STRING platform (Szklarczyk et al., 2017). We found that the 871 unique altered genes had significantly more interactions than expected (p-value$<1*10^{-16}$). Specifically, out of 871 genes, 491 genes were inter-connected with other genes from the list (Figure 2G, Figure S2A). Furthermore, unsupervised hierarchical clustering and PCA analyses on the connected 491 genes indicated that their expression pattern maintained the distinct segmentation of the three fibroblast populations: NLF, MIF and MAF (Figure 2H, I). Interestingly, the remaining unconnected genes did not maintain the segmentation of distinct stages (Figure S2B), suggesting that the stage-specific gene expression signatures represent PPI connected pathways that are functional in metastases-associated fibroblasts. Therefore, we designated these 491 genes “stage-specific gene signature” and further focused our subsequent analyses on their functional roles.

Shani et al. Figure 2



Parameter	Result
Number of nodes	863
Number of edges	1470
Average node degree	3.41
Average local clustering coefficient	0.38
Expected number of edges	996
PPI enrichment p-value	$< 1*10^{-16}$
Number of unique connected nodes	491

Figure 2. Transcriptome profiling of metastases-associated fibroblasts reveals dynamic stage-specific changes in gene expression. (A) Workflow illustration of fibroblast isolation (CD45-EpCAM-YFP+) from normal FVB/n; *col1a1*-YFP mice (NLF), and of Micro- or Macro- metastases associated fibroblasts from PyMT; *col1a1*-YFP mice (MIF and MAF). **(B)** Flow cytometry gating strategy for the isolation of fibroblasts. **(C)** Principal Component Analysis (PCA) of 11,115 protein coding genes identified in the RNA-seq data. See also Figure S1. **(D)** Venn diagram of top 2% upregulated or downregulated genes (871 genes) in the different comparisons (MIF vs. NLF, MAF vs. NLF, MAF vs. MIF). **(E)** Two-way hierarchical clustering of top 2% genes based on Euclidian distance and average linkage. **(F)** PCA of top 2% genes. **(G)** Protein-protein interaction analysis of the 871 top 2% genes performed in STRING platform. Interconnected genes were selected for subsequent analysis. See also Figure S2A (H) PCA of 491 genes selected in (G). **(I)** Hierarchical clustering dendogram of 491 genes selected in (G) based on Euclidian distance and average linkage. See also Figure S2B.

Fibroblast metastases-promoting tasks are driven by functional gene signatures related to stress response, inflammation, and ECM remodeling.

We next asked whether the stage-specific gene signature in fibroblasts represents specific metastases-promoting tasks. To address this question, we performed further analysis of the selected genes in each stage by using the over-representation enrichment analysis of the Consensus Path DB (CPDB) platform (Kamburov et al., 2011). Our focus in these analyses was based on three different databases: GO (Ashburner et al., 2000; The Gene Ontology, 2019), KEGG (Kanehisa et al., 2017; Kanehisa and Goto, 2000), and Reactome (Fabregat et al., 2018). For our analysis, we selected terms that represent biological processes enriched in at least two databases, with a relative overlap of at least 0.2 and at least 2 shared entities (Figure 3A). Data analysis revealed significant and stage-specific changes in functional pathways including cellular stress response, extracellular matrix (ECM) remodeling and inflammation (Figure 3B, Table S1).

Interestingly, we found that gene expression signatures in fibroblasts isolated from the micro-metastatic stage were highly and specifically enriched for functions related to cellular response to stress, including *HSF1* activation, heat shock response and response to unfolded protein (Table S1). Upregulated genes in MIF that were related to stress and protein folding included several heat shock proteins: *Hspa8, Hsp90aa1, Hspd1, Hspe1* and others (Figure 3C). ECM remodeling terms were enriched in both MIF and MAF (Figure 3B), indicating the central importance of ECM modifications in facilitating metastasis. Notably, while ECM remodeling was operative throughout the metastatic process, the specific genes related to ECM remodeling in the different metastatic stages were distinct (Figure 3D).

Gene expression signatures in fibroblasts isolated from macro-metastases were highly enriched for inflammation-related gene sets (Figure 3B, Table S1). Notably, while there was a slight decrease in inflammatory related signaling in MIF, it is based on a small number of genes, as compared with the plethora of genes related to inflammation and chemokine and cytokine activities upregulated in MAF (Figure 3E).
Taken together, these findings imply that metastases-associated fibroblasts assume distinct functional roles during the process of lung metastasis.

Encouraged by these findings, we next set out to obtain further insights on functional pathways that were modified in fibroblasts isolated from different metastatic stages. To that end, we performed Gene Set Enrichment Analysis (GSEA) of the whole dataset (Subramanian et al., 2005). We focused our analysis on the H collection: Hallmark gene sets that summarize specific well-defined biological states or processes based on multiple datasets (Liberzon et al., 2015). Similarly to the results obtained in our previous analyses, we found several functions related to inflammatory responses enriched in MAF (Figure 3F, Table S2). Interestingly, we found that *Myc* target genes were the most highly and significantly enriched in both metastatic stages (Figure 3G, Table S2), suggesting that this transcription factor may play a central role in the functional molecular co-evolution of metastases-associated fibroblasts.
Taken together, these findings imply that the transcriptome of lung fibroblasts is rewired during metastatic progression, driving changes in the expression of distinct molecular pathways. Moreover, the transcriptional changes in ECM remodeling and stress response, which are potential metastases-promoting tasks, are evident at early stages of metastases formation, suggest that fibroblasts play an important role already at the onset of the metastatic process.

Shani et al. Figure 3

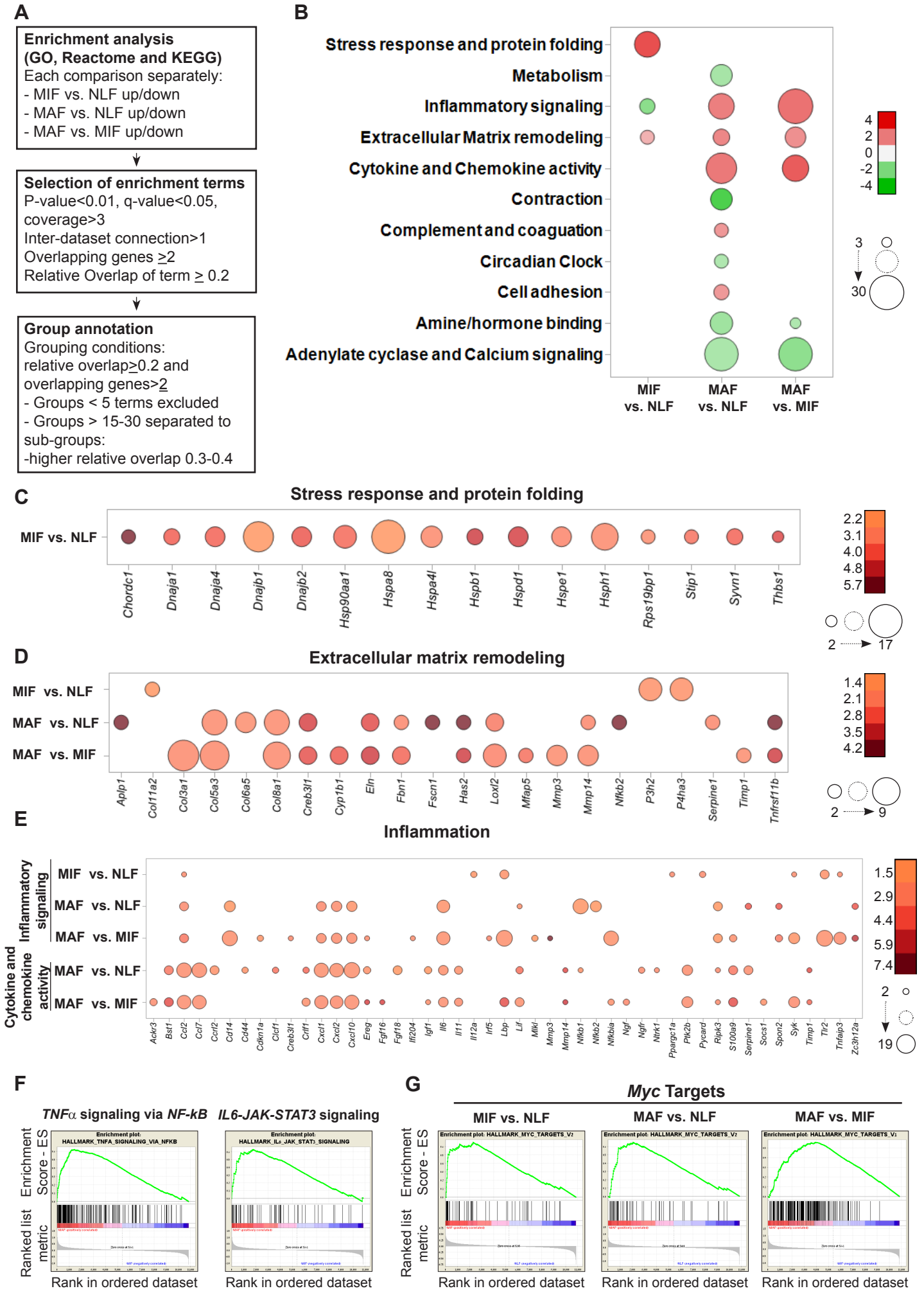


Figure 3. Metastases-promoting tasks are driven by functional gene signatures related to stress response, inflammation, and ECM remodeling. (A) Flow chart of the pathway enrichment over-representation analyses based on GO, Reactome and KEGG using the CPDB platform. **(B)** Bubble graph heat map based on the number of specific enrichment terms and their average Log-transformed q-value per group. Circle sizes denote number of terms included in a group, color indicates the average Log-transformed q-value. Enrichments based on downregulated genes are presented as negative values. See also Table S1. **(C-E)** Bubble graph heat map presenting genes in selected group annotations. Circle size denotes the number of terms that included the gene, color indicates the average Log-transformed q-value of these terms. Only genes found in at least 2 different terms are presented. (C) "Stress response and protein folding" enriched genes. (D) "Extracellular Matrix remodeling" enriched genes. (E) "Inflammatory signaling" and/or "Cytokine and Chemokine activity" enriched genes, genes with an average q-value>0.01 were excluded. **(F)** Gene Set Enrichment Analysis (GSEA) results for hallmark datasets upregulated in MAF vs. NLF related to inflammatory signaling, FDR<0.05, NES>2. See also Table S2. **(G)** GSEA results for "*Myc* targets" hallmark dataset that were upregulated in all comparisons (MIF vs. NLF, MAF vs. NLF, MAF vs. MIF), FDR<0.05 NES>2. See also Table S2.

Multiple gene network analyses identify Myc as a central transcription factor in the rewiring of metastases-associated fibroblasts.

To further characterize the regulatory nodes that govern the functional changes in fibroblasts, we hypothesized that these changes may be driven by the transcription factors that were identified by the CPDB and GSEA analyses (Figure 3). Specifically, we identified five candidate transcription factors (TFs) that were enriched in at least one analysis and in at least one metastatic stage: *Hif1a, Hsf1*, *Myc, Nfkb1 and Stat3.*

To rank these TFs, we performed knowledge-based multiple analyses examining their centrality in the selected gene signatures. First, we examined in how many different comparisons each TF was enriched. We found that *Hsf1* was only enriched in the micro-metastatic stage vs. normal lungs, and *Hif1a* was enriched only in the macro-metastatic stage. *Nfkb1* and *Stat3* were enriched in the macro-metastatic stage, compared with both normal and micro-metastases. Notably, only *Myc* was enriched in all three comparisons (Figure 4A, blue, Table S3). We next examined the protein-protein interactions (PPIs) of these TFs utilizing the STRING platform, and counted the number of direct connections of each TF with the stage-specific gene signature. *Stat3* had the largest number of connections, closely followed by *Myc* (Figure 4A, orange, Table S3). In addition to STRING, we examined PPIs using ANAT (Advanced Network Analysis Tool) (Yosef et al., 2011). In this platform, the inference is based on setting all the candidate TFs as anchors and the selected genes as targets in a network of protein-protein interactions (PPI), and searching for a putative compact sub-network that connects them. We analyzed the results according to four parameters: the number of direct connections of each TF, the total number of connected genes (including non-directly connected genes), the characteristic path length, and network centralization. Analysis of the results revealed that *Myc* had the largest number of direct connections, as well as the largest number of total connections and is overall connected to the fibroblast stage-specific gene signature with the shortest path and with the highest centrality (Figure 4A, yellow, Figure 4B, Fig S3, Table S3). These results suggested that *Myc* plays a central role in mediating the transcriptional rewiring of fibroblasts in the lung metastatic niche.

We next examined the specific connection of each TF as a regulator in the metastases-associated gene network. To that end, we utilized the RegNetwork tool (Liu et al., 2015), a knowledge-based database of gene regulatory networks. We found that *Myc* had the greatest number of targets, followed by *Stat3* and *Nfkb1* (Figure 4A, green, Table S3). Finally, we analyzed the correlation of the metastases-associated gene network with each candidate TF using the VarElect tool (Stelzer et al., 2016). This tool enables prioritization of genes related to a specific query term by using a direct and indirect relatedness score. We analyzed the scores of the stage specific signature genes with each candidate TF, and the number of directly related genes. The TFs were ranked based on the number and average score for the directly related genes, and the average score of the indirectly related genes. In agreement with previous analyses, *Myc* had the highest number of connections and the highest average score for both directly and indirectly related genes (Figure 4A, pink, Figure 4C, Table S3). To consolidate these comprehensive gene network analyses, we performed a comparative analysis on the TF bioinformatics measurements listed in Figure 4A. The results indicated that *Myc* achieved significantly higher scores than all other TFs (Figure 4D,E). These results implicate the centrality of *Myc* in the dynamic transcriptional changes that govern the function of metastases-associated fibroblasts in lung metastasis.

A

Platform	Analysis	*Hif1a*	*Hsf1*	*Myc*	*Nfkb1*	*Stat3*
Enrichments	# comparisons	1	1	3	2	2
STRING	# 1st neighbour	66	39	134	80	144
ANAT	# 1st neighbour	14	4	152	4	84
	total connections	416	447	465	412	438
	characteristic path length	3.93	4.83	3.222	4.566	3.677
	network centralization	0.111	0.217	0.419	0.104	0.231
RegNetwork	# targets	12	10	122	58	46
VarElect	directly related	293	226	458	332	408
	average score directly related	1.01	1.15	1.91	1.5	1.92
	average score indirectly related	0.54	0.84	2.83	1.12	1.47

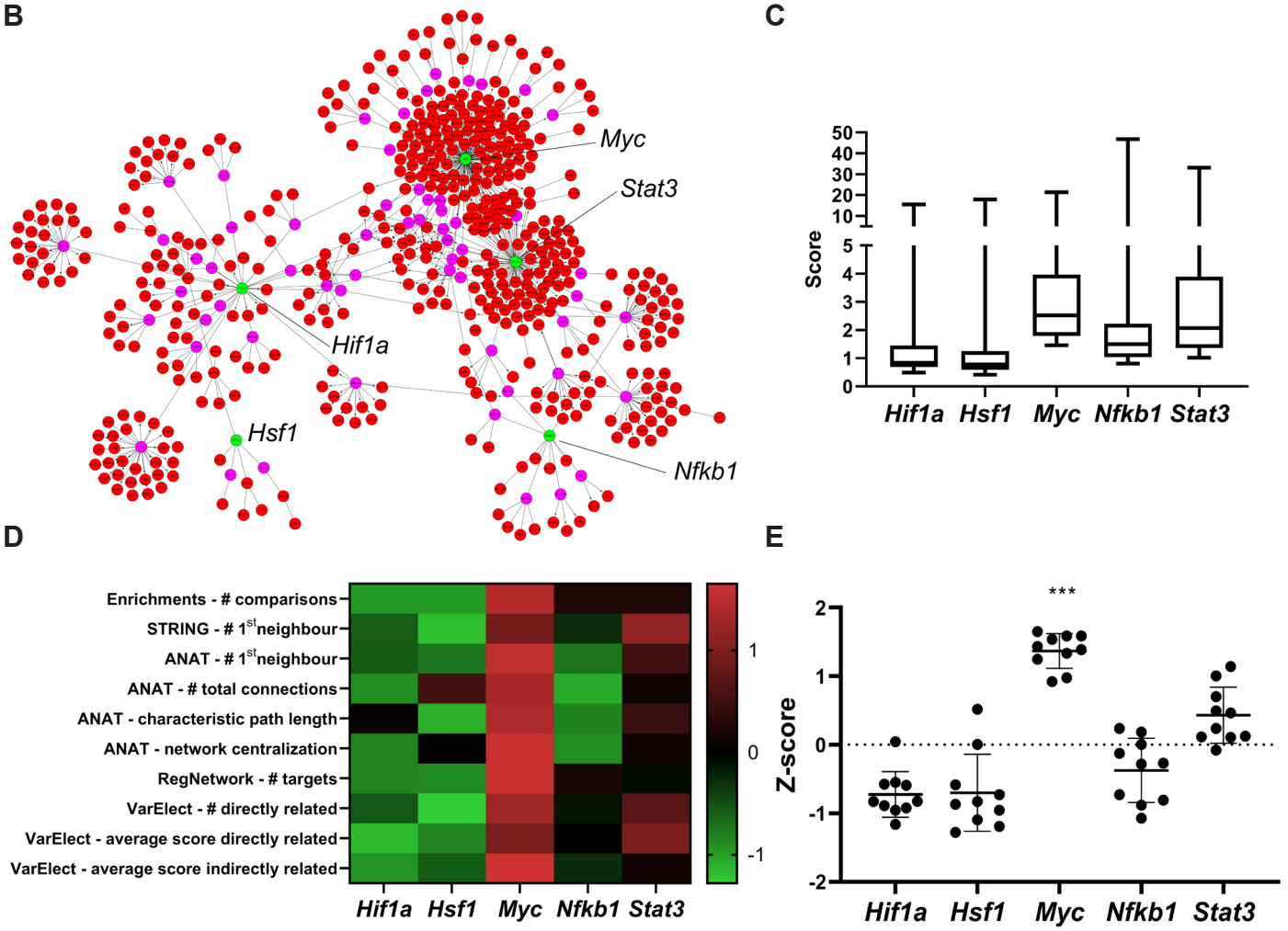


Figure 4. Multiple gene network analyses identify *Myc* as a central transcription factor in the rewiring of metastases-associated fibroblasts. (A) Table of ranking parameters and analyses performed to identify the centrality of the five candidate transcription factors (TFs): *Hif1a, Hsf1, Myc, Nfkb1, Stat3*. Blue - number of comparisons that identified each TF. Orange – STRING PPI analysis results. Yellow - ANAT pathway analysis results. Green - RegNetwork analysis of connectivity between target genes and TFs. Pink - VarElect analysis results. See also Table S3. **(B)** Representative ANAT protein-protein network, using all TFs as anchors (green) and the stage-specific signature as target genes (red). (Presenting only confidence >0.6). See also Figure S3. **(C)** Box Plot of VarElect scores for directly related genes to each TF (Presenting top 200 per TF). **(D,E)** Heat map and graph of Z-scores of the results described in (A), *** $p<0.001$, one-way ANOVA with Tukey correction for multiple comparisons. Data are represented as mean ± SD.

Myc is a central regulator in metastases-associated fibroblasts and contributes to their acquisition of tumor-promoting traits

Myc (myelocytomatosis oncogene) is a transcription factor involved in many biological processes, including cell growth and proliferation, cell stemness, and metabolism. *Myc* is deregulated in many human cancers, and is known to play an important role in the pathogenesis of cancer, particularly in cancer cells (Dang, 2012; Poole and van Riggelen, 2017).

To validate the ranking results, we analyzed by qRT-PCR the expression of *Myc* in fibroblasts isolated from normal lungs, or from lungs with micro- and macrometastases. Analysis of the results indicated that *Myc* is significantly upregulated in macrometastases-associated fibroblasts (Figure 5A). In addition, we assessed the expression of central *Myc* targets that were found to be upregulated in metastases-associated fibroblasts, including *Hspe1*, *Hsp90aa1*, *Odc1* and *Fosl1* (Belinky et al., 2015; Chakravorty et al., 2017). We found that these *Myc* targets were upregulated in fibroblasts isolated from lungs with metastases (Figure 5B). Interestingly, the qRT-PCR results validated our findings in the RNA-seq data, in which *Hsp90aa1* was upregulated in MIF, whereas the other *Myc* targets were upregulated in MAF (Figure 5B). Importantly, fibroblasts highly upregulated the expression of *Myc* in response to tumor cell secreted factors (Figure 5C, left bars). To elucidate the functional importance of *Myc* in mediating lung fibroblast reprogramming, we targeted its expression by a specific *Myc* targeting siRNA in primary lung fibroblasts. Abrogation of *Myc* expression by siMyc resulted in significant inhibition of *Myc* expression as compared with control fibroblasts. Moreover, *Myc* inhibition abrogated the upregulation of *Myc* in response to tumor cell secreted factors in activated fibroblasts (Figure 5C, right bars). We next assessed whether inhibition of *Myc* affected the upregulation of selected *Myc* target genes in activated lung fibroblasts. Analysis of the results indicated that the expression of *Myc* is central to the upregulation of its known targets (Figure 5D). Finally, we examined the importance of *Myc* for functional activation of fibroblasts in response to tumor cell-secreted factors, by analyzing hallmark functions of activated lung fibroblasts, including enhanced migration and matrix contraction. To that end, we performed a wound healing assay and collagen contraction assay with tumor-activated lung fibroblasts that were transfected with siMyc or with siCtrl. We found that siMyc fibroblasts were less contractile and exhibited significantly attenuated migration capacity as compared with controls (Figure 5E-H). Taken together, our findings imply that *Myc* has a central role in enhancing fibroblast activation and in mediating their acquisition of metastasis-promoting functions.

Shani et al. Figure 5

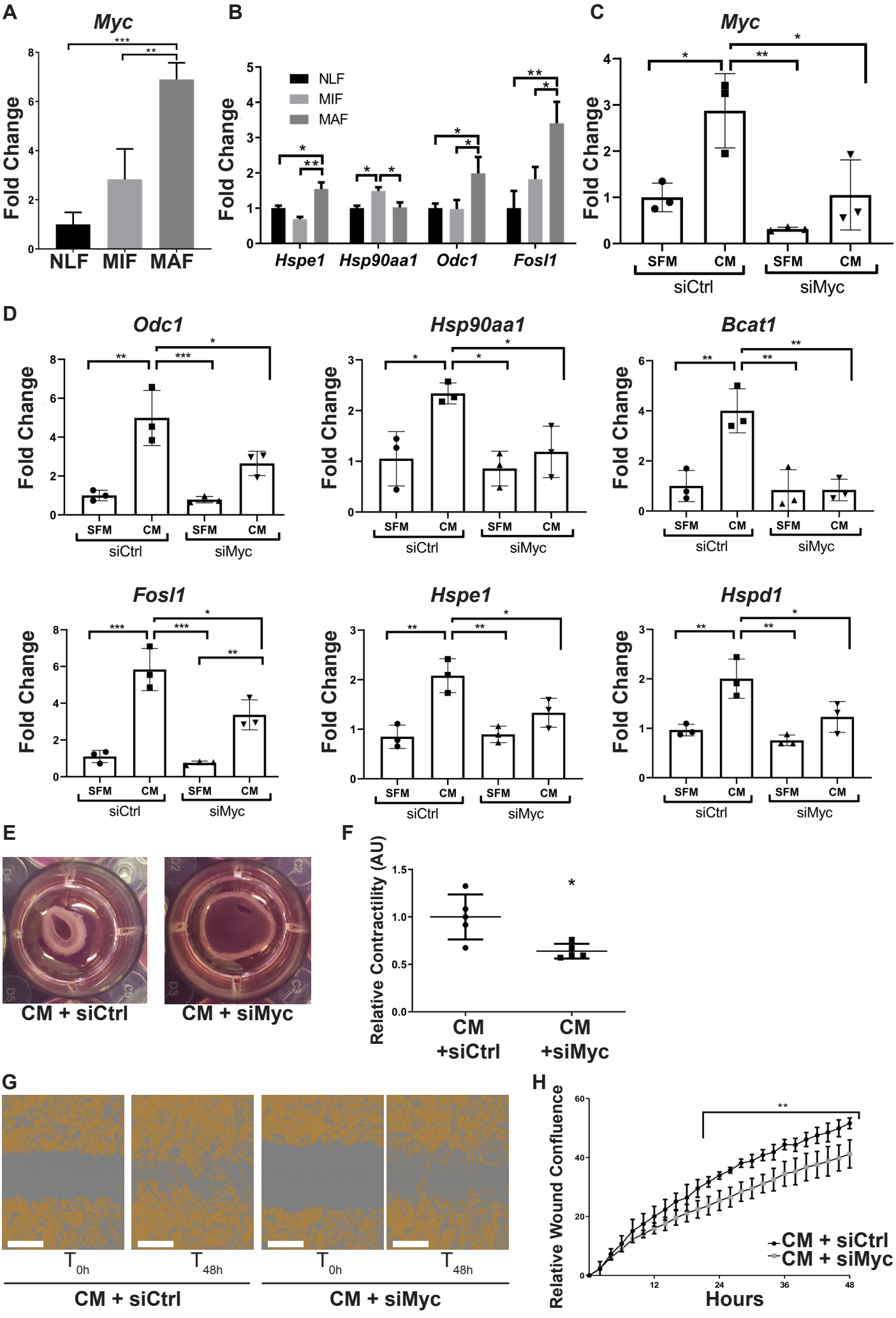


Figure 5. *Myc* is a central regulator of metastases-associated fibroblasts and contributes to their acquisition of tumor promoting traits. (A) qRT-PCR analysis of *Myc* expression in sorted NLF, MIF and MAF. **p<0.01, Data are represented as mean ± SD, n=3 per group. **(B)** qRT-PCR analysis in sorted NLF, MIF and MAF. Relative expression of *Myc* target genes that were found to be differentially expressed in the RNA-seq. *p<0.05, Data are represented as mean ± SD, n=3 per group. **(C)** *Myc* targeting by siRNA: *Myc* expression in NLF transfected with siRNA targeting *Myc* or with control siRNA (siMyc or siCtrl). Following transfection, cells were incubated with SFM or with Met-1 CM supplemented with the same siRNA for additional 24h. Data are represented as mean ± SD, n=3. **(D)** qRT-PCR analysis of *Myc* targets following treatment as in (C). Data are represented as mean ± SD, n=3. **(E,F)** Representative images and quantification of collagen contraction assay of fibroblasts transfected with siMyc or siCtrl, incubated with Met-1 CM. *p-value<0.05, Data are represented as mean ± SD, n=5. **(G,H)** Representative images and quantification of scratch closure assay at 0h and 48h of NLF transfected with siMyc or siCtrl and incubated with Met-1 CM. Scale bar: 400µm. Two-way ANOVA with multiple comparisons, ***p-value<0.001, Data are represented as mean ± SD, n=5.

High expression of MYC and the metastases-associated fibroblast gene signature is associated with worse survival in human breast cancer

Encouraged by these findings, we next asked if the metastases-associated gene signature is operative in human breast cancer. Since there are no available transcriptomic datasets of lung metastases we analyzed patient data from breast tumors. We initially analyzed the expression of genes that we found to be upregulated in murine metastases-associated fibroblasts (upMAG-upregulated metastases-associated genes) in the stromal and epithelial compartments of human breast tumors, utilizing a publicly available dataset (Ma et al., 2009). This analysis revealed that while some of the genes are more expressed in the epithelial compartment (cancer cells, E), 40% of the genes were significantly more expressed in the stromal compartment (S) of human breast tumors (Figure 6A). Motivated by this result, we then performed survival analysis based on the genes that were significantly overexpressed in the stroma, and found that high expression of 21 of these genes was correlated with worse prognosis in breast cancer patients, including: *SERPINE1, PLAUR, FGFR4* and *BCAT1* (Figure 6B, Table S4). Notably, these results implicate stromal gene expression in predicting patient survival, and suggest that the expression of genes identified in metastases-associated fibroblasts contributes to worse prognosis in human breast cancer.

Since we showed that *MYC* is a central regulator of fibroblast rewiring during metastatic progression in mice, we next asked whether stromal expression of upMAG in human breast cancer is related to *MYC*. Interestingly, analysis of *MYC* expression in the stromal compartment of breast cancer patients revealed that *MYC* is upregulated in breast cancer stroma in correlation with disease progression, as reflected by pathological grade: expression of *MYC* was elevated in grade 3 disease, compared with stroma isolated from more differentiated tumors (Figure 6C). Moreover, to assess the centrality of *MYC* in the stromal compartment of breast tumors, we performed correlation analysis between the expression of *MYC* and the expression of the upMAG signature. We found that stromal expression of *MYC* was positively correlated with stromal expression of multiple genes that we found to be upregulated in fibroblasts (Figure 6D). Furthermore, the number of *MYC* positively correlated genes was significantly higher than the number of negatively correlated genes (Figure 6E). Notably, both *STAT3* and *NFKB1* had significantly less positive correlations as compared with *MYC,* further supporting our findings that *MYC* is a central regulator of stromal rewiring in breast cancer (Figure 6F).

We next stratified patients based on the expression of *MYC* and the expression of each gene from the upMAG signature into 4 groups: high-high, high-low, low-high, low-low, respectively, and performed survival analysis. Strikingly, in 80 different genes, including the *MYC* downstream genes validated in murine fibroblasts in Figure 5, patients with combined overexpression (high-high) had the worst survival, while patients with combined lower expression had the best survival (Figure 6G-H, Table S4). Taken together, these results indicate that the activation of *MYC* transcriptional networks in the stroma of breast tumors plays a role in tumor progression and survival in human breast cancer.

Shani et al. Figure 6

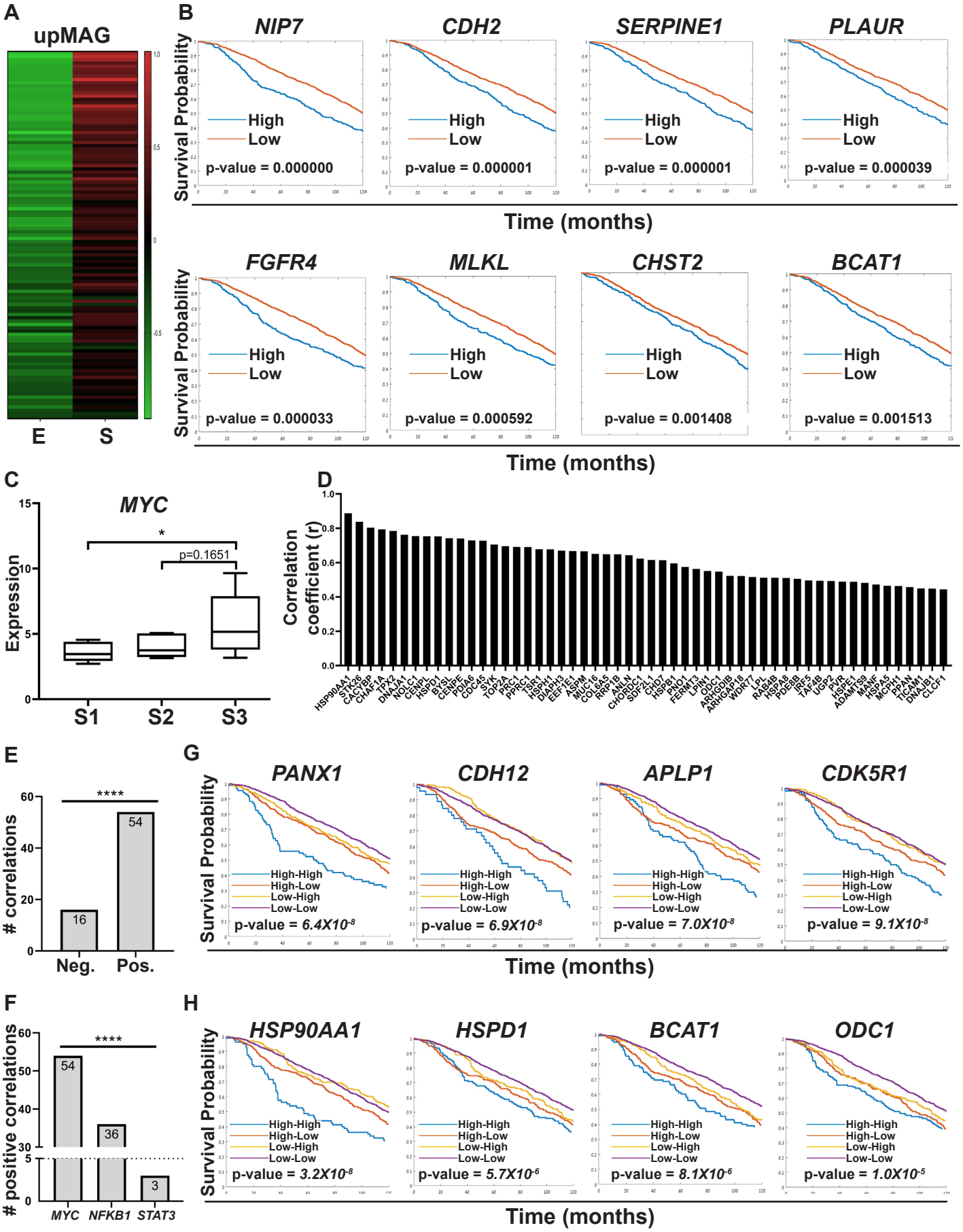


Figure 6. High expression of MYC and the metastases-associated fibroblast gene signature is associated with worse survival in human breast cancer. (A) Heat Map presenting upregulated metastases-associated genes (upMAG) that were significantly overexpressed in the stroma of human breast cancer (S) as compared with the epithelial (cancer cell) tissue (E), based on GSE14548, q-value<0.05. **(B)** Kaplan-Meier analyses for overall survival in breast cancer patients in the METABRIC cohort based on the expression of genes found to be overexpressed in the stroma in (A). High=top 20%, Low=bottom 80%, p-value<0.05. See also Table S4. **(C)** Box-plot of *MYC* expression in tumor associated-stroma (S) from the GSE14548 by disease grade (Stroma-grade 1: S1; Stroma-grade 2: S2; Stroma-grade 3: S3). One-way ANOVA with Tukey correction for multiple comparisons, *p-value<0.05. **(D)** Positive correlations between the expression of *MYC* and the expression of upMAG in tumor-associated stroma based on GSE14548. P-value<0.05. **(E)** Comparison between the number of significant positive and significant negative correlations between upMAG and *MYC* in tumor-associated stroma based on GSE14548. Chi-square test, ****p-value<0.0001 **(F)** Comparison between the number of significant positive correlations between upMAG and *MYC, NFKB1* and *STAT3* in tumor-associated stroma based on GSE14548. Chi-square test, ****p-value<0.0001 **(G-H)** Kaplan-Meier analyses for overall survival in breast cancer patients in the METABRIC cohort based on the expression of *MYC* in combination with upMAG. High-High: top 20% MYC and top 20% upMAG gene, High-Low: top 20% *MYC* and bottom 80% upMAG gene, Low-High: bottom 80% MYC and top 80% upMAG gene, Low-Low: bottom 80% *MYC* and bottom 80% upMAG gene. P-value<0.0001. (G) Top 4 most significant genes. See also Table S4. (H) Genes validated in Figure 5.

Discussion

In this study we set out to elucidate the dynamic changes in the stromal compartment that facilitate the formation of a hospitable metastatic niche during breast cancer metastasis to lungs. We utilized a unique model of transgenic mice that enabled unbiased isolation of fibroblasts from spontaneous lung metastasis and performed comprehensive analysis of the transcriptome of fibroblasts isolated from normal lungs, and lungs with micro- or macrometastases. By employing multiple platforms of data analysis, we integrated ontology analyses with data on protein interactions and functional pathways from knowledge-based databases to identify the relevant and stage-specific gene signatures that imply functional tasks of metastases-associated fibroblasts.

Importantly, we performed the analysis on fibroblasts isolated directly from fresh tissues, with no additional culture step that may affect gene expression. Our findings indicated that ECM remodeling programs were instigated early in micrometastases, and persisted to be functional throughout the metastatic pathway, while other signaling pathways were activated in a stage-specific manner. Activation of the cellular stress response was associated with micrometastases, and inflammatory signaling was instigated in fibroblasts isolated from advanced metastases, suggesting that fibroblasts are transcriptionally dynamic and plastic, and that they adapt their function to the evolving microenvironment (Figure 7).

Initial analysis of the RNA-seq data revealed distinct gene signatures associated with early and advanced metastatic disease. By performing step by step analysis, utilizing the STRING and CPDB platforms combined with GSEA, we unraveled multiple pathways operative in fibroblasts in different metastatic stages, relying not only on altered gene expression but also on mechanistic significance and interaction of genes.

Interestingly, this multi-layered analysis indicated that fibroblasts isolated from micrometastases instigated the expression of genes related to cellular response to stress, including the transcriptional regulator *Hsf1*. *Hsf1* was previously shown to be upregulated in CAFs in early stage breast and lung cancers and to drive a stromal tumor-promoting transcriptional program that correlated with worse prognosis (Scherz-Shouval et al., 2014). Moreover, *Hsf1* was recently implicated in activating stromal expression of Dickkopf-3 (*Dkk3*) in mammary fibroblasts, which is a central modulator of their pro-tumorigenic activity, thus linking *Hsf1* with reprogramming of stromal fibroblasts to CAFs (Ferrari et al., 2019). Our findings expand these observations, and show that activation of *Hsf1* transcriptional regulation in lung fibroblasts occurs during the early stages of metastasis and may play an important role in instigating tumor-promoting tasks in metastases-associated fibroblasts.

In addition to stress response, our findings indicated that ECM remodeling is a central task of metastases-associated fibroblasts throughout the metastatic cascade. Indeed, ECM components and remodeling were demonstrated to facilitate breast cancer metastasis to lungs, and pancreatic cancer metastasis to liver (Cox et al., 2013; Malanchi et al., 2012; Nielsen et al., 2016; Oskarsson et al., 2011;Yuzhalin et al., 2018). We show that transcriptional rewiring of fibroblasts to mediate tasks such as collagen synthesis and ECM organization is a central function of metastases-associated fibroblasts, which is instigated early during the metastatic process and persists during advanced metastatic disease.

Notably, analyzing the central pathways in fibroblasts that were isolated from advanced metastases, indicated that metastases-associated fibroblasts upregulated pro-inflammatory pathways including multiple cytokines and chemokines. CAFs are known to play a central role in mediating tumor-promoting inflammation at the primary tumor site (Servais and Erez, 2013). Importantly, in recent years activation of inflammation has also emerged in shaping of

the metastatic microenvironment (Albrengues et al., 2018; Coffelt et al., 2015; Quail et al., 2017), but the role of fibroblasts in mediating inflammation at the metastatic site was not previously shown. Our findings implicate CAF-derived inflammatory signaling in promoting the growth of lung metastases. Further studies are required to determine which pro-inflammatory factors upregulated in metastases-associated fibroblasts are functionally important in formation of a hospitable metastatic niche in lungs.

We further characterized the molecular mechanisms operative in metastases-associated fibroblasts, by identifying the central transcription factors that drive the metastases-associated gene programs upregulated in lung fibroblasts. Our analyses revealed several central regulators that are operative in metastases-associated fibroblasts, including the well-known modulators of CAF activity *Nfkb1* (Erez et al., 2013; Erez et al., 2010) and *Stat3* (Chakraborty et al., 2017; Li et al., 2018).
Surprisingly, the most prominent regulator in the metastasis-associated fibroblasts network was the transcription factor *Myc*. While the importance of *Myc* in promoting cell transformation and tumorigenesis is well established (Poole and van Riggelen, 2017), its role in the tumor stroma is largely uncharacterized. *Myc* expression in tumor cells was shown to drive an inflammatory and immunosuppressive microenvironment (Kortlever et al., 2017). Moreover, the expression of *Myc* in the stromal compartment was suggested to mediate metabolic and transcriptional reprogramming of fibroblasts (Minciacchi et al., 2017; Yan et al., 2018). Our study identifies *Myc* as a central regulator in the transcriptional plasticity of metastases-associated fibroblasts. Indeed, inhibition of *Myc* attenuated tumor promoting functions of fibroblasts, confirming that *Myc* functionally contributes to their acquisition of tumor promoting traits. Importantly, validation of these findings in human breast cancer patients revealed that stromal expression of *Myc* and its downstream genes is highly correlated with worse survival in breast cancer patients. Stromal gene expression was previously found to be associated with bad prognosis in colon cancer (Calon et al., 2015). Our findings implicate activation of *Myc* and stromal gene expression in breast cancer patient survival. Taken together, these findings indicate that in addition to its known role in driving carcinogenesis in tumor cells, *Myc* functions in stromal rewiring in the tumor microenvironment in both primary tumors and metastases of breast cancer.

In summary, we show that integration of multiple analytical platforms of gene expression, connectivity and function provided a powerful insight on functional and temporal regulation of the dynamic transcriptome of fibroblasts in lung metastasis. We uncovered central molecular pathways that drive the activation of growth-promoting tasks in fibroblasts via known regulators of CAF tumor-promoting activities including *Stat3*, *Nfkb1*, and *Myc*, a novel regulator of fibroblast metastases-promoting tasks. Our findings elucidate for the first time the dynamic transcriptional co-evolution of fibroblasts during the multi-stage process of breast cancer metastasis.

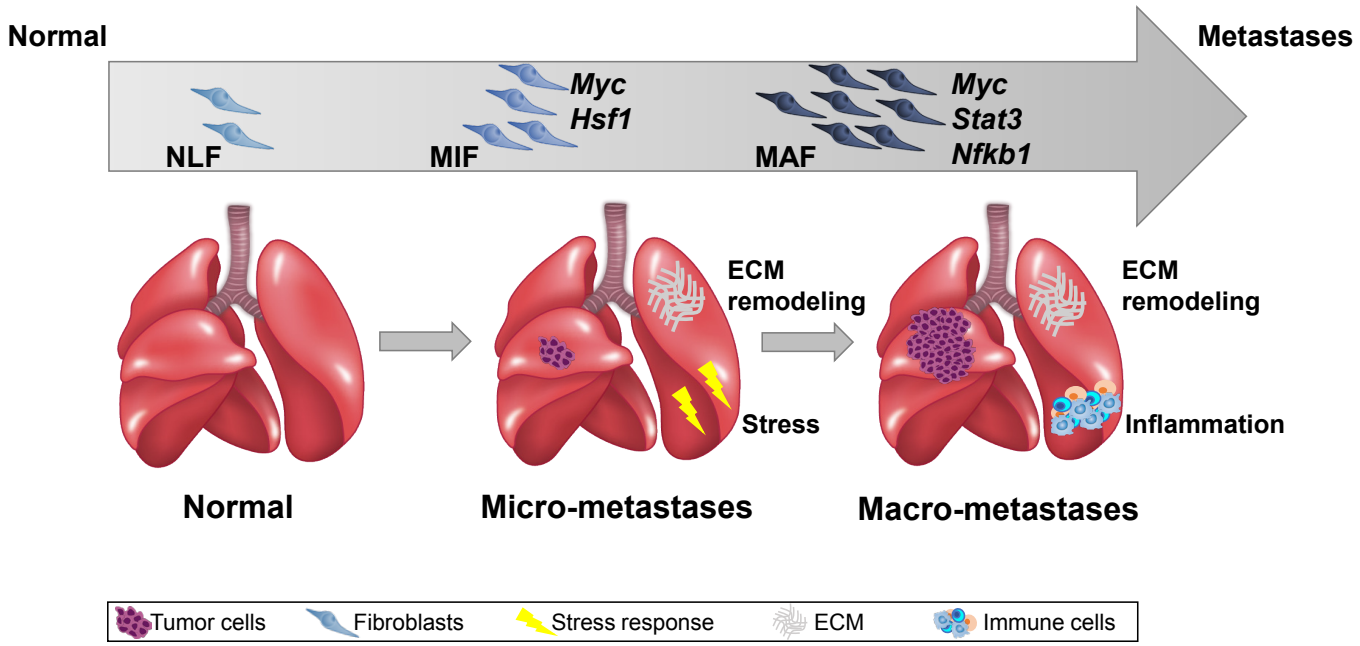


Figure 7. Summary scheme: The co-evolution of lung fibroblasts at the metastatic microenvironment is driven by stage-specific transcriptional plasticity that activates growth-promoting tasks including stress response, ECM remodeling and instigation of inflammatory signaling.

Acknowledgments

This research was supported by grants to NE from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 637069 MetCAF), from the Israel Science Foundation (#813/12), the Israel Cancer Association (ICA), and The Israel Cancer Research Fund (*Research Career Development Award and Project Grant*). IT and OM were supported in part by a research grant from the Breast Cancer Research Foundation (BCRF). The authors would like to thank Dr. Ran Elkon for his help with data analysis.

Author Contributions

OS, YR conceived and carried out experiments; NC and LM participated in experiments, OS analyzed data and generated the figures; OM, DS and HS analyzed data, RS provided essential resources; IT was involved in data analysis and discussions; NE designed and supervised the study; OS and NE wrote the manuscript.

Declaration of Interests

The authors declare no conflict of interests.

Methods

Mice

All experiments were performed using 6-8 weeks old female mice, unless otherwise stated. All experiments involving animals were approved by the Tel Aviv University Institutional Animal Care and Use Committee. FVB/n *Col1a1*-YFP mice were a kind gift from Dr. Gustavo Leone. FVB/N-Tg (MMTV-PyMT) 634Mul/J were backcrossed with FVB/n;*Col1a1*-YFP mice to create PyMT;*Col1a1*-YFP double-transgenic mice as described previously (Raz et al., 2018). All animals were maintained within the Tel Aviv University Specific Pathogen Free (SPF) facility.

Cell cultures

Met-1 mouse mammary gland carcinoma cells: Met-1 cells were a gift from Prof. Jeffrey Pollard. Met-1 cells were plated on 100mm plastic dishes and cultured with DMEM medium supplemented with 10% FCS, 1% penicillin-streptomycin and 1% Sodium-pyruvate (Biological Industries).

Primary lung fibroblasts cultures: Lungs were isolated from 6-8 weeks old FVB/n female mice. Single cell suspensions were prepared as previously described (Sharon et al., 2013). Single cell suspensions were seeded on 6-well plates pre-coated with Rat tail collagen (Corning; 354236). Cells were grown in DMEM media supplemented with 10% FCS, and maintained at 37°C with 5% CO_2.

Conditioned media

Tumor cell conditioned media (Met-1 CM): cells were cultured as described above. When cells reached 80% confluency, plates were washed twice with PBS and fresh serum free DMEM medium (SFM) was applied. After 48h, medium was collected, filtered through 0.45μm filters under aseptic conditions, flash-frozen in liquid nitrogen and stored at -80°C. DMEM SFM supplemented as above was used as control.

Normal lung fibroblasts (NLF) or Activated lung fibroblasts (ALFs) conditioned media: NLF were plated as described above. CM was prepared by incubating NLF with either SFM (for NLF CM) or Met-1 CM (for ALF CM) for 24 hours. After 24h, plates were washed twice with

PBS and cells were incubated for additional 24h with fresh SFM. After 48h, medium was collected, filtered through 0.45μm filters under aseptic conditions, flash-frozen in liquid nitrogen and stored at -80°C.

Scratch assay

NLF were plated in a 96-well IncuCyte® imageLock plate (Essen BioSciense). SFM was applied for 16h. Wells were then washed three times with PBS and a scratch was made using the IncuCyte® WoundMaker (EssenBiosciense). Wells were washed three times with PBS and Met-1 CM or SFM were applied. The plate was placed in the IncuCyte® system (Essen BioSciense) for 48 hours. Images were analyzed using the IncuCyte® software.

Collagen contraction

NLF were plated as mentioned above and incubated with SFM for 16h. Following, Cells were detached from dishes with trypsin and counted. A total of 1.5×10^5 fibroblasts were suspended in a medium and collagen mixture [Met-1 CM or SFM mixed with High Concentration Rat Tail Collagen, type 1 (BD bioscience)], and allowed to set at 37°C for 45 min. Met-1 CM or SFM were applied, gels were released and incubated for 24 hours. Gels were photographed at various time points. ImageJ software was used to measure gel area and assess collagen contraction.

Migration assay

Met-1 (5×10^4) cells were placed into the upper chamber of 24 Transwell inserts, with pore sizes of 8μm, in 300μl NLF CM or ALF CM. Following 24h incubation, the upper side of the apical chamber was scraped gently with cotton swabs to remove non-migrating cells, fixed with methanol and stained with DAPI. Migrated cells were documented under a fluorescence microscope. ImageJ software was used to quantify migration.

Immunostaining

Lungs were harvested, shortly washed in PBS and embedded in Optimal Cutting Temperature compound (O.C.T, Tissue-Tek) on dry ice. Serial sections were obtained to ensure equal sampling of the examined specimens (8μm trimming). Tissue sections were incubated for 1 hour at room temperature with anti-mouse *αSMA* antibody (Sigma-Aldrich; F3777). Sections were mounted with DAPI Fluoromount-G (Southern Biotech; 0100-20). Slides were visualized using Leica SP5 microscope. Quantitative analyses were performed using ImageJ Software.

Flow cytometry analysis and cell sorting

Single cell suspensions of Lungs isolated from FVB/n;*Col1α1*-YFP or PyMT;*Col1α1*-YFP mice were stained using the following antibodies: anti-EpCAM-APC (eBioscience, 17-5791), anti-CD45-PerCP-Cy5.5 (eBioscience; 45-0451), anti-CD31-PE-Cy7 (eBioscience; 25-0311). DAPI was used to exclude dead cells (Molecular Probes; D3571). Flow cytometric analysis was performed using CytoFLEX Flow Cytometer (Beckman Coulter). Cell sorting was performed using BD FACSAria II or BD FACSAria Fusion (BD bioscience). Data analysis was performed with the Kaluza Flow Analysis software (Beckman Coulter).

RNA isolation and qRT-PCR

RNA from sorted cells was isolated using the EZ-RNAII kit (20-410-100, biological industries) according to the manufacturer's protocol. RNA from *in vitro* experiments was isolated using the PureLink™ RNA Mini Kit (Invitrogen; 12183018A). cDNA synthesis was conducted using qScript cDNA Syntesis kit (Quanta; 95047-100). Quantitative real-time PCRs (qRT-PCR) were conducted using PerfeCTa SYBR Green Fastmix ROX (Quanta; 95073-012). In all

analyses expression results were normalized to *Gusb or Gapdh* and to control cells. RQ ($2^{-\Delta\Delta Ct}$) was calculated.

Transfection of primary fibroblasts

NLF were cultured in DMEM supplemented with 10% FCS. At 70% confluency, cells were transfected with Accell Delivery Media (GE Dharmacon; B-005000) supplemented with 1μM Accell SMARTpool mouse *Myc* siRNA (Dharmacon; E-040813) or Accell Control Pool non-targeting siRNA (Dharmacon; D-001910) for 96h. Accell SMARTpool contains a mixture of four siRNAs targeting one gene, and provide extended duration of gene knockdown with only minimal effects on cell viability and the innate immune response. The efficiency of *Myc* siRNA knockdown was analyzed by qRT-PCR.

RNA-seq

$CD45^{-}EpCAM^{-}YFP^{+}$ Fibroblasts were isolated by cell sorting from Normal FVB/n; *Col1a1*-YFP mice (n=4), PyMT;*Col1a1*-YFP Micro-metastases bearing mice (n=3) and PyMT;*Col1a1*-YFP Macro-metastases bearing mice (n=4). Micro-metastases were defined as visible mammary tumors, the absence of visible macro-metastases and the presence of $EpCAM^{+}$ cells in lungs. Cells were collected into Trizol LS reagent (Life Technologies; 10296-028) and RNA was isolated according to the manufacturer's instructions. Transcriptomic sequencing of RNA was performed using NEBNext® rRNA Depletion Kit (New England Biolabs, Inc.; E6310S) and SMARTer Stranded Total RNA-Seq Kit - Pico Input (Clontech; 635005) and sequenced on the Illumina HiSeq 2500 sequencer (Illumina, Inc.) at the Technion Genome Center. Sequenced reads were aligned to the mouse genome (mm9) using TopHat2 (Kim et al., 2013). Gene expression counts were calculated using HTseq-count (Anders et al., 2015) using Gencode annotations. Only genes that got at least 20 counts in at least 3 replicate samples were included in subsequent analysis (12,105 genes). Gene expression counts were normalized using quantile normalization (Bolstad et al., 2003). Levels below 20 were then set to 20 to reduce inflation of fold-change estimates for lowly expressed genes. Preliminary differential expression analysis was carried out using DESeq2 (Love et al., 2014). For subsequent analyses, only protein coding genes were included. In addition, coefficient of variance (CV) was calculated per group (NLF, MIF, MAF) and the top 1% most in-group deviated genes (top 1% CV) were excluded, leaving a total of 11,115 genes.

Stage-specific signature analysis

The top 2% most altered genes from each comparison (up/down) were selected (871 genes in total) as a primary signature for the different stages (NLF, MIF, MAF). Data was Z-scored per gene. Venn diagram was generated using Bioinformatics & Evolutionary Genomics website (http://bioinformatics.psb.ugent.be/webtools/Venn/). All hierarchical clustering (based on Euclidian distance and average linkage) and principal component analyses were performed using JMP 14 software.

Gene selection based on network connectivity

All 871 selected genes were subjected to protein-protein interactions analysis using the STRING platform (Szklarczyk et al., 2017). The minimum confidence of interaction was defined based on the upper quartile of all interactions (confidence$\geq$0.3) and connections based on text-mining were excluded. Groups of under 4 genes were excluded, leaving a total of 491 genes ("stage-specific gene signature").

Pathway enrichment

For functional annotation, pathway and enrichment analysis, each comparison was analyzed

separately, to a total of 6 comparisons (MIF vs. NLF up, MIF vs. NLF down, MAF vs. NLF up, MAF vs. NLF down, MAF vs. MIF up, MAF vs. MIF down). Over-representation analysis was performed using the ConsensusPath DataBase CPDB, (Kamburov et al., 2011) (http://cpdb.molgen.mpg.de/MCPDB) platform for GO-molecular function (MF) and GO-biological process (BP), Reactome, and KEGG. Terms larger than 500 genes were excluded. Results were considered significant with a p-value<0.01, q-value<0.05 and the top 75% highest coverage (≥3%). The CPDB platform utilizes multiple databases, enables visualization of relationships between different enrichments, and allows comparison and integration between them (Herwig et al., 2016). To increase the specificity of the enriched terms, we compared the relative overlap and the number of shared entities between the enriched terms from the three different databases (GO, KEGG and Reactome). Selected terms with at least 2 shared entities and a relative overlap ≥ 0.2 were grouped and annotated based on a common enriched function. Groups containing more than 15-20 terms were divided to sub-groups based on a higher relative overlap criteria. Groups smaller than 5 terms were excluded. These steps enabled the selection of the ~10% most highly and significantly connected terms.

For the generation of bubble plot heat maps all q-values were log transformed [$-Log_{10}$(q-value)]. For the group annotation heat map, log-transformed q-values of all the terms included in a specific group were averaged. For terms enriched in a group of downregulated genes, the value of the average log transformed q-value was transformed to a negative value by duplicating the average log transformed value by (-1). For the generation of gene bubble graph heat maps per selected groups, the average of log-transformed q-values was calculated per gene, based on the enrichment terms that included it.

Gene Set Enrichment Analysis (GSEA)

The GSEA Java plug-in was used to probe log-transformed normalized expression data (Subramanian et al., 2005) http://software.broadinstitute.org/gsea/index.jsp). Settings for the analysis were defined as the follows: Gene set database – Hallmark gene sets only, Number of permutations -1000, comparisons – each separately (MIF vs. NLF, MAF vs. NLF, MAF vs. MIF), Permutation type - gene_set, minimum size - 5, maximum size - 500. Significant results were considered for False Discovery Rate (FDR) <0.05 and normalized enrichment score (NES) > |2|.

Transcription Factor Ranking

Selection of transcription factors

Transcription factors (*Hif1a, Hsf1, Myc, Nfkb1, Stat3*) that were shown to be enriched in the pathway enrichment and/or GSEA analyses described above were selected as candidates and subjected to subsequent analyses. The Transcription Factors (TFs) were ranked as candidates for their relevance to the transcriptional rewiring of metastases-associated fibroblasts.

Selection of transcription factors

Transcription factors (*Hif1a, Hsf1, Myc, Nfkb1, Stat3*) that were shown to be enriched in the pathway enrichment and/or GSEA analyses described above were selected as candidates and subjected to subsequent analyses. The Transcription Factors (TFs) were ranked as candidates for their relevance to the transcriptional rewiring of metastases-associated fibroblasts.

STRING

All five candidate TFs were subjected to protein-protein interaction analysis in combination

with the "stage-specific gene signature" using the STRING platform (Szklarczyk et al., 2017). The minimum confidence of interaction was defined based on the median confidence value of all interactions (confidence >0.2). For the ranking of each TF, the number of separate interactions for each TF was counted.

ANAT

The ANAT (Advanced Network Analysis Tool) application (Yosef et al., 2011) was used as an add-in to Cytoscape (version 7 and up) software. We performed the analysis for each TF separately and for all of the TFs combined. The TFs were defined as anchors in the list, and the target genes were the stage-specific list of genes. An HTML report of all possible pathways between the anchor and each gene in the target genes list was generated. The minimum confidence for a connection was defined based on the median confidence value of all interactions (confidence >0.2). An anchor could be connected to a target directly, or indirectly. For the ranking of each TF, we calculated several parameters of the protein-protein network: 1) The number of stage-specific genes connected with each TF (directly and 2-4 shells); 2) The average shortest path for each TF; 3) The centrality of the network. Parameters 2 and 3 were calculated using the network analysis tool of the Cytoscape software.

RegNetwork

Each TF was defined as a regulator in the RegNetwork database (Liu et al., 2015) and all non-microRNA results were downloaded. For the ranking of each TF, the number of genes out of the stage-specific gene signature that are registered as targets were counted.

VarElect

Using the VarElect platform (Stelzer et al., 2016), we analyzed the relation of the stage-specific genes with each TF. Each gene from the list received a score according to its relation to the TF. For the ranking of each TF, we analyzed several parameters: 1) the number of directly related genes; 2) The average score of related genes; 3) The average score of indirectly related genes.

Ranking

To assess the relative relevance of the candidate TFs, ranking analysis was performed. Ranking parameters described above were Z-scored per parameter. For "Characteristic path length" results were first transformed with a (-1) power. Statistical analysis was performed using One-Way ANOVA with Tukey correction for multiple comparisons.

Human breast cancer data

Stromal and epithelial expression profiles and correlation analyses

The expression of upMAG (upregulated metastases-associated genes) and *Myc*, *Nfkb1* and *Stat3* were analyzed in human breast cancer stroma and epithelial tissues based on a publicly available dataset GSE14548 (Ma et al., 2009). Correlation analysis was performed on normalized expression values using Pearson correlation. P-value below 0.05 was considered significant.

Kaplan-Meier analyses

Overall survival analysis was performed using patient gene expression profiles generated by the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study (Curtis et al., 2012). Kaplan-Meier analysis was performed using MATLAB on patients stratified by gene expression (high=top 20% expression, low=bottom 80% expression). For

the combination with Myc patients were divided into 4 groups: Myc high (top 20%) and upMAG gene high (top 20%), Myc high and upMAG gene low (bottom 80%), Myc low and upMAG gene high, and Myc low and upMAG gene low. p-value below 0.05 was considered significant.

Statistical analysis

Statistical analyses were performed using GraphPad Prism software and JMP pro 14 software. For two groups, statistical significance was calculated using t-test with Welch correction unless otherwise stated. For more than two comparisons, One-Way ANOVA with Tukey correction for multiple comparisons was applied unless otherwise stated. All tests were two-tailed. P-value of ≤0.05 was considered statistically significant unless otherwise stated. Bar graphs represent mean and standard deviation (SD) unless otherwise stated. All experiments represent at least 3 separate biological repeats.

Data Access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE128999.

Supplementary Tables

Table S1. Related to Figure 3. Detailed enrichment results for all comparisons based on selection criteria.
Table S2. Related to Figure 3. Full GSEA results for all comparisons, FDR<0.05, NES>|2|.
Table S3. Related to Figure 4. Full TF network analysis results.
Table S4. Related to Figure 6. List of significant KM results.

References

Albini, A., Bruno, A., Noonan, D.M., and Mortara, L. (2018). Contribution to Tumor Angiogenesis From Innate Immune Cells Within the Tumor Microenvironment: Implications for Immunotherapy. Frontiers in immunology *9*, 527.

Albrengues, J., Shields, M.A., Ng, D., Park, C.G., Ambrico, A., Poindexter, M.E., Upadhyay, P., Uyeminami, D.L., Pommier, A., Kuttner, V.*, et al.* (2018). Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice. Science *361*.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny Stein, T., Safran, M., and Lancet, D. (2015). PathCards: multi-source consolidation of human biological pathways. Database (Oxford) *2015*.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Borowsky, A.D., Namba, R., Young, L.J., Hunter, K.W., Hodgson, J.G., Tepper, C.G., McGoldrick, E.T., Muller, W.J., Cardiff, R.D., and Gregg, J.P. (2005). Syngeneic mouse mammary carcinoma cell lines: two closely related cell lines with divergent metastatic behavior. Clinical & experimental metastasis *22*, 47-59.

Calon, A., Lonardo, E., Berenguer-Llergo, A., Espinet, E., Hernando-Momblona, X., Iglesias, M., Sevillano, M., Palomo-Ponce, S., Tauriello, D.V., Byrom, D.*, et al.* (2015). Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet *47*, 320-329.

Chakraborty, D., Sumova, B., Mallano, T., Chen, C.W., Distler, A., Bergmann, C., Ludolph, I., Horch, R.E., Gelse, K., Ramming, A.*, et al.* (2017). Activation of STAT3 integrates common profibrotic pathways to promote fibroblast activation and tissue fibrosis. Nat Commun *8*, 1130.

Chakravorty, D., Jana, T., Das Mandal, S., Seth, A., Bhattacharya, A., and Saha, S. (2017). MYCbase: a database of functional sites and biochemical properties of Myc in both normal and cancer cells. BMC Bioinformatics *18*, 224.

Coffelt, S.B., Kersten, K., Doornebal, C.W., Weiden, J., Vrijland, K., Hau, C.S., Verstegen, N.J., Ciampricotti, M., Hawinkels, L.J., Jonkers, J.*, et al.* (2015). IL-17-producing gammadelta T cells and neutrophils conspire to promote breast cancer metastasis. Nature *522*, 345-348.

Cohen, N., Shani, O., Raz, Y., Sharon, Y., Hoffman, D., Abramovitz, L., and Erez, N. (2017). Fibroblasts drive an immunosuppressive and growth-promoting microenvironment in breast cancer via secretion of Chitinase 3-like 1. Oncogene *36*, 4457-4468.

Cooke, M.E., Sakai, T., and Mosher, D.F. (2000). Contraction of collagen matrices mediated by alpha2beta1A and alpha(v)beta3 integrins. Journal of cell science *113 (Pt 13)*, 2375-2383.

Cox, T.R., Bird, D., Baker, A.M., Barker, H.E., Ho, M.W., Lang, G., and Erler, J.T. (2013). LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. Cancer research *73*, 1721-1732.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y.*, et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346-352.

Dang, C.V. (2012). MYC on the path to cancer. Cell *149*, 22-35.

DeNardo, D.G., Barreto, J.B., Andreu, P., Vasquez, L., Tawfik, D., Kolhatkar, N., and Coussens, L.M. (2009). CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages. Cancer cell *16*, 91-102.

Deng, J., Liu, Y., Lee, H., Herrmann, A., Zhang, W., Zhang, C., Shen, S., Priceman, S.J., Kujawski, M., Pal, S.K*., et al.* (2012). S1PR1-STAT3 Signaling Is Crucial for Myeloid Cell Colonization at Future Metastatic Sites. Cancer cell *21*, 642-654.

Erez, N., and Coussens, L.M. (2011). Leukocytes as paracrine regulators of metastasis and determinants of organ-specific colonization. International journal of cancer Journal international du cancer *128*, 2536-2544.

Erez, N., Glanz, S., Raz, Y., Avivi, C., and Barshack, I. (2013). Cancer Associated Fibroblasts express pro-inflammatory factors in human breast and ovarian tumors. Biochemical and biophysical research communications *437*, 397-402.

Erez, N., Truitt, M., Olson, P., Arron, S.T., and Hanahan, D. (2010). Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner. Cancer cell *17*, 135-147.

Erler, J.T., Bennewith, K.L., Cox, T.R., Lang, G., Bird, D., Koong, A., Le, Q.T., and Giaccia, A.J. (2009). Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. Cancer cell *15*, 35-44.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B*., et al.* (2018). The Reactome Pathway Knowledgebase. Nucleic acids research *46*, D649-D655.

Ferrari, N., Ranftl, R., Chicherova, I., Slaven, N.D., Moeendarbary, E., Farrugia, A.J., Lam, M., Semiannikova, M., Westergaard, M.C.W., Tchou, J*., et al.* (2019). Dickkopf-3 links HSF1 and YAP/TAZ signalling to control aggressive behaviours in cancer-associated fibroblasts. Nat Commun *10*, 130.

Fridlender, Z.G., Albelda, S.M., and Granot, Z. (2015). Promoting metastasis: neutrophils and T cells join forces. Cell Res *25*, 765-766.

Hanahan, D., and Coussens, L.M. (2012). Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer cell *21*, 309-322.

Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nature protocols *11*, 1889-1907.

Jablonska, J., Lang, S., Sionov, R.V., and Granot, Z. (2017). The regulation of pre-metastatic niche formation by neutrophils. Oncotarget *8*, 112132-112144.

Jin, K., Pandey, N.B., and Popel, A.S. (2017). Crosstalk between stromal components and tumor cells of TNBC via secreted factors enhances tumor growth and metastasis. Oncotarget *8*, 60210-60222.

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. Nature reviews Cancer *16*, 582-598.

Kalluri, R., and Zeisberg, M. (2006). Fibroblasts in cancer. Nature reviews Cancer *6*, 392-401.

Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., and Herwig, R. (2011). ConsensusPathDB: toward a more complete picture of cell biology. Nucleic acids research *39*, D712-717.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research *45*, D353-D361.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research *28*, 27-30.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kortlever, R.M., Sodir, N.M., Wilson, C.H., Burkhart, D.L., Pellegrinet, L., Brown Swigart, L., Littlewood, T.D., and Evan, G.I. (2017). Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. Cell *171*, 1301-1315 e1314.

Li, A., Chen, P., Leng, Y., and Kang, J. (2018). Histone deacetylase 6 regulates the immunosuppressive properties of cancer-associated fibroblasts in breast cancer through the STAT3-COX2-dependent pathway. Oncogene *37*, 5952-5966.

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417-425.

Lin, E.Y., Jones, J.G., Li, P., Zhu, L., Whitney, K.D., Muller, W.J., and Pollard, J.W. (2003). Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. The American journal of pathology *163*, 2113-2126.

Liu, L., Liu, L., Yao, H.H., Zhu, Z.Q., Ning, Z.L., and Huang, Q. (2016). Stromal Myofibroblasts Are Associated with Poor Prognosis in Solid Cancers: A Meta-Analysis of Published Studies. PloS one *11*, e0159947.

Liu, Z.P., Wu, C., Miao, H., and Wu, H. (2015). RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database (Oxford) *2015*.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Ma, X.J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D.C. (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast cancer research : BCR *11*, R7.

Malanchi, I., Santamaria-Martinez, A., Susanto, E., Peng, H., Lehr, H.A., Delaloye, J.F., and Huelsken, J. (2012). Interactions between cancer stem cells and their niche govern metastatic colonization. Nature *481*, 85-89.

Minciacchi, V.R., Spinelli, C., Reis-Sobreiro, M., Cavallini, L., You, S., Zandian, M., Li, X., Mishra, R., Chiarugi, P., Adam, R.M.*, et al.* (2017). MYC Mediates Large Oncosome-Induced Fibroblast Reprogramming in Prostate Cancer. Cancer research *77*, 2306-2317.

Nguyen, D.X., Bos, P.D., and Massague, J. (2009). Metastasis: from dissemination to organ-specific colonization. Nature reviews Cancer *9*, 274-284.

Nielsen, S.R., Quaranta, V., Linford, A., Emeagi, P., Rainer, C., Santos, A., Ireland, L., Sakai, T., Sakai, K., Kim, Y.S.*, et al.* (2016). Macrophage-secreted granulin supports pancreatic cancer metastasis by inducing liver fibrosis. Nature cell biology *18*, 549-560.

Obenauf, A.C., and Massague, J. (2015). Surviving at a Distance: Organ-Specific Metastasis. Trends Cancer *1*, 76-91.

Oskarsson, T., Acharyya, S., Zhang, X.H., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massague, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nature medicine *17*, 867-874.

Oskarsson, T., and Massague, J. (2012). Extracellular matrix players in metastatic niches. The EMBO journal *31*, 254-256.

Peinado, H., Lavotshkin, S., and Lyden, D. (2011). The secreted factors responsible for pre-metastatic niche formation: old sayings and new thoughts. Seminars in cancer biology *21*, 139-146.

Peinado, H., Zhang, H., Matei, I.R., Costa-Silva, B., Hoshino, A., Rodrigues, G., Psaila, B., Kaplan, R.N., Bromberg, J.F., Kang, Y.*, et al.* (2017). Pre-metastatic niches: organ-specific homes for metastases. Nature reviews Cancer *17*, 302-317.

Poole, C.J., and van Riggelen, J. (2017). MYC-Master Regulator of the Cancer Epigenome and Transcriptome. Genes (Basel) *8*.

Qian, B.Z., Li, J., Zhang, H., Kitamura, T., Zhang, J., Campion, L.R., Kaiser, E.A., Snyder, L.A., and Pollard, J.W. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature *475*, 222-225.

Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nature medicine *19*, 1423-1437.

Quail, D.F., Olson, O.C., Bhardwaj, P., Walsh, L.A., Akkari, L., Quick, M.L., Chen, I.C., Wendel, N., Ben-Chetrit, N., Walker, J.*, et al.* (2017). Obesity alters the lung myeloid cell landscape to enhance breast cancer metastasis through IL5 and GM-CSF. Nature cell biology *19*, 974-987.

Raz, Y., Cohen, N., Shani, O., Bell, R.E., Novitskiy, S.V., Abramovitz, L., Levy, C., Milyavsky, M., Leider-Trejo, L., Moses, H.L.*, et al.* (2018). Bone marrow-derived fibroblasts are a functionally distinct stromal cell population in breast cancer. The Journal of experimental medicine *215*, 3075-3093.

Rouabhia, M., Park, H., Meng, S., Derbali, H., and Zhang, Z. (2013). Electrical stimulation promotes wound healing by enhancing dermal fibroblast activity and promoting myofibroblast transdifferentiation. PloS one *8*, e71660.

Scherz-Shouval, R., Santagata, S., Mendillo, M.L., Sholl, L.M., Ben-Aharon, I., Beck, A.H., Dias-Santagata, D., Koeva, M., Stemmer, S.M., Whitesell, L.*, et al.* (2014). The reprogramming of tumor stroma by HSF1 is a potent enabler of malignancy. Cell *158*, 564-578.

Servais, C., and Erez, N. (2013). From sentinel cells to inflammatory culprits: cancer-associated fibroblasts in tumour-related inflammation. The Journal of pathology *229*, 198-207.

Sharon, Y., Alon, L., Glanz, S., Servais, C., and Erez, N. (2013). Isolation of Normal and Cancer-associated Fibroblasts from Fresh Tissues by Fluorescence Activated Cell Sorting (FACS). J Vis Exp.

Sharon, Y., Raz, Y., Cohen, N., Ben-Shmuel, A., Schwartz, H., Geiger, T., and Erez, N. (2015). Tumor-derived osteopontin reprograms normal mammary fibroblasts to promote inflammation and tumor growth in breast cancer. Cancer research *75*, 963-973.

Stelzer, G., Plaschkes, I., Oz-Levi, D., Alkelai, A., Olender, T., Zimmerman, S., Twik, M., Belinky, F., Fishilevich, S., Nudel, R.*, et al.* (2016). VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. BMC Genomics *17 Suppl 2*, 444.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P.*, et al.* (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic acids research *45*, D362-D368.

The Gene Ontology, C. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic acids research *47*, D330-D338.

Yan, W., Wu, X., Zhou, W., Fong, M.Y., Cao, M., Liu, J., Liu, X., Chen, C.H., Fadare, O., Pizzo, D.P.*, et al.* (2018). Cancer-cell-secreted exosomal miR-105 promotes tumour growth

through the MYC-dependent metabolic reprogramming of stromal cells. Nature cell biology *20*, 597-609.

Yosef, N., Zalckvar, E., Rubinstein, A.D., Homilius, M., Atias, N., Vardi, L., Berman, I., Zur, H., Kimchi, A., Ruppin, E.*, et al.* (2011). ANAT: a tool for constructing and analyzing functional protein networks. Sci Signal *4*, pl1.

Yuzhalin, A.E., Lim, S.Y., Kutikhin, A.G., and Gordon-Weeks, A.N. (2018). Dynamic matrisome: ECM remodeling factors licensing cancer progression and metastasis. Biochim Biophys Acta Rev Cancer *1870*, 207-228.