Research Paper

Role of germline variants in the metastasis of breast carcinomas

Ángela Santonja[1,2,*], Aurelio A. Moya-García[2,3,*], Nuria Ribelles[4,5], Begoña Jiménez-Rodríguez[4], Bella Pajares[4], Cristina E. Fernández-De Sousa[1,2], Elísabeth Pérez-Ruiz[6], María del Monte-Millán[5,7], Manuel Ruiz-Borrego[8], Juan de la Haba[5,9], Pedro Sánchez-Rovira[10], Atocha Romero[11], Anna González-Neira[12], Ana Lluch[5,13,14] and Emilio Alba[2,4,5]

[1]Instituto de Investigación Biomédica de Málaga (IBIMA), Hospitales Universitarios Regional y Virgen de la Victoria de Málaga, Spain

[2]Laboratorio de Biología Molecular del Cáncer, Centro de Investigaciones Médico-Sanitarias (CIMES), Universidad de Málaga, Málaga, Spain

[3]Departmento de Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, Spain

[4]Unidad de Gestión Clínica Intercentro de Oncología, Instituto de Investigación Biomédica de Málaga (IBIMA), Hospitales Universitarios Regional y Virgen de la Victoria de Málaga, Málaga, Spain

[5]Centro de Investigación Biomédica en Red de Oncología, CIBERONC-ISCIII, Madrid, Spain

[6]Medical Oncology Service, Hospital Costa del Sol, Marbella, Málaga, Spain

[7]Instituto de Investigación Sanitaria Gregorio Marañón, Universidad Complutense, Madrid, Spain

[8]Medical Oncology Service, Hospital Virgen del Rocío, Sevilla, Spain

[9]Biomedical Research Institute, Complejo Hospitalario Reina Sofía, Córdoba, Spain

[10]Department of Oncology, Complejo Hospitalario de Jaén, Jaén, Spain

[11]Molecular Oncology Laboratory, Hospital Clínico San Carlos, IdISSC, Madrid, Spain

[12]Human Genotyping-CEGEN Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

[13]Department of Oncology and Hematology, Hospital Clínico Universitario, Valencia, Spain

[14]INCLIVA Biomedical Research Institute, Universidad de Valencia, Valencia, Spain

[*]These authors contributed equally to this work

Correspondence to: Aurelio A. Moya-García, **email**: amoyag@uma.es

ABSTRACT

Most cancer-related deaths in breast cancer patients are associated with metastasis, a multistep, intricate process that requires the cooperation of tumour cells, tumour microenvironment and metastasis target tissues. It is accepted that metastasis does not depend on the tumour characteristics but the host's genetic makeup. However, there has been limited success in determining the germline genetic variants that influence metastasis development, mainly because of the limitations of traditional genome-wide association studies to detect the relevant genetic polymorphisms underlying complex phenotypes. In this work, we leveraged the extreme discordant phenotypes approach and the epistasis networks to analyse the genotypes of 97 breast cancer patients. We found that the host's genetic makeup facilitates metastases by the dysregulation of gene expression that can promote the dispersion of metastatic seeds and help establish the metastatic niche—providing a congenial soil for the metastatic seeds.

INTRODUCTION

The metastatic dissemination of the disease causes the overwhelming majority of cancer-related deaths, yet this enormously complex process remains poorly understood. The different models and unifying concepts that explain metastasis cannot yet explain the adaptive programs that allow tumour cells to thrive in distant tissues and, therefore, the biological and clinical observations associated with metastasis [1, 2].

The metastatic cascade requires the cooperation of tumour cells with different cells of the rest of the organism. Carter et al., showed how the genetic background (i.e., the inherited polymorphisms carried in the germline) could influence the somatic evolution of a tumour in at least two ways: by determining the site of tumourigenesis and modifying the likelihood of acquiring mutations in specific cancer genes [3]. Furthermore, genes are not expressed in isolation but promote the expression of other genes in a coordinated pattern. Therefore, the tendency of a tumour to metastasise may be determined by coordinated changes in gene expression. Studies based on many prognosis signatures [4–7] show that the coordinated gene expression in most cells present early in tumourigenesis— i.e., gene regulation—often determines tumour biology.

Animal models and epidemiological studies suggest that the risk of developing metastasis after breast cancer diagnosis depends on the characteristics of the tumour and germline gene variants [8, 9]. Furthermore, Xu et al., have shown how germline variants of natural killer cells in the tumour immune microenvironment can sway metastasis risk in several cancers [10]; they have also identified germline genomic patterns that contribute to cancer progression and metastasis [11].

Germline variants predate and complement tumour cells' somatic variants (mutations). The tumour acquisition of further mutations empowers its cells to disseminate and proliferate in a distant tissue (i.e., metastatic seeds), and the host's genetic makeup promotes metastasis by providing a congenial soil [12]. Germline genetic variants that facilitate metastasis tend to be spread across the genome and interact with one another. In complex traits such as metastasis, multiple genetic influences are responsible for moderate differences in survival. Variants with a single, potent effect on the phenotype are rare. Therefore, the metastatic phenotype depends on accumulating weak effects on a substantial fraction of the genes that comprise the regulatory pathways driving metastasis (Boyle et al., 2017).

Detecting these collectives of interacting variants, each with modest effect size, requires substantially greater research effort than the individual strong-effect variants usually studied [13]. The extreme discordant phenotypes approach based on comparing high-risk healthy individuals—and therefore likely to bear the genetic variants that protect them from disease—with sick, low-risk individuals likely to bear the genetic variants that predispose them to that disease. This approach assumes that the patients at both ends of the disease spectrum are the most informative and, therefore, requires fewer patients to genotype and increases the statistical power of gene association studies [14].

Epistasis networks constitute a novel technique to identify genetic variants associated with a disease that accounts for the heritability of complex traits—traits for which the interactions among many genes control the variations between individuals [13, 15]. Metastasis is such a complex process that metastasis susceptibility is probably due to complex allelic combinations of germline variants [16]. Therefore, metastasis is the perfect ground to deploy epistasis networks since they rest on the idea that the synergistic interactions among many genetic variants, each with a moderate individual effect, determine the disease susceptibility. Epistasis emphasises that the synergistic interactions among genetic variants determine their effect on the phenotype; that is, the effect of one genetic variant on a given trait depends on the genotype of many other variants affecting the trait [17]. Therefore, epistasis has the potential to characterise the network of interactions among genetic variants that shapes the genetic architecture of metastasis [18].

Agarwal et al., proposed that germline variants complement somatic mutations as breast cancer drivers— although germline polymorphisms affect critical biological processes required for breast carcinogenesis, their action is not enough for cancer initiation. Pre-existing germline variants could determine the subsequent somatic mutations required for cancer initiation [19]. Our work is based on a similar hypothesis: what constitutes a somatic event that drives metastasis in breast cancer is conditional on the collection of germline variations in the patient. Therefore, in this work, we have analysed the host factors (i.e., germline variants) that contribute to the susceptibility to metastasis in breast cancer patients. According to the extreme discordant phenotypes approach, our analysis framework is based on genome-wide genotyping of a small set of patients at the extreme of the metastasis susceptibility distributions—low-risk individuals who unexpectedly relapse within five years of follow-up and high-risk patients without relapse. We performed an epistasis network analysis to detect the variants that take part in metastasis by modulating the effect of other variants. We selected the genes that harbour these germline variants based on their role in the regulation of metastasis. We found several gene candidates through which the host's genetic makeup contributes to metastasis.

RESULTS AND DISCUSSION

Characteristics of the study and patients

We performed genome-wide genotyping in a cohort of 97 breast cancer patients that showed metastatic

Table 1: Top SNPs ranked by SNPrank score

SNP ID	Gene	Chromosome	SNP Rank score	SNP location	MAF
rs11139965	RN7SKP242	9	0.5898	intergenic	0.170
rs67242866	LINC01362	1	0.5784	intergenic	0.028
rs199830092	PIK3C2B	1	0.5749	intron variant	NA
rs72951131	ZAP70	2	0.5587	intron variant	0.183
rs34000182	MIR5689HG	6	0.5485	intron variant	0.053
rs61776380	RNU6-830P	1	0.5444	intergenic	0.065
rs778902	ISCA1P2	1	0.5378	intergenic	0.098
rs2501357	C1orf204	1	0.5360	intron variant	0.306
rs742635	ABTB2	11	0.5349	intron variant	0.085
rs16927008	CLVS1	8	0.5324	intron variant	0.027
rs4849127	IL1B	2	0.5257	downstream gene variant	0.125
rs41263676	C1orf21	1	0.5251	3'UTR variant	0.131
rs11692741	MYT1L	2	0.5211	intron variant	0.180
rs77142354	MIR4300HG	11	0.5165	intron variant	0.039
rs77162747	ZNF767P	7	0.5144	intergenic	0.072
rs77169575	MAML3	4	0.5142	intron variant	0.198
rs9994379	ELOVL6	4	0.5136	intron variant	0.034
rs13421497	PRKCE	2	0.5134	intergenic	0.073
rs1197934	LINGO2	9	0.5128	intron variant	0.112
rs75394800	EML6	2	0.5115	intron variant	0.033
rs3799142	VIP	6	0.5039	downstream gene variant	0.172

Abbreviation: MAF: minor allele frequency.

extreme discordant phenotypes: 34 good prognosis cases (with tumours smaller than 2 cm and no lymph nodes affected who relapsed within five years after surgery) and 63 poor prognosis cases (patients with more than ten lymph nodes affected who did not relapse within five years after surgery). This design has the advantage of reducing the phenotypic heterogeneity of the cohort, resulting in a potential enrichment of germline variants associated with metastasis predisposition. The median age at diagnosis was 50 years (range 29–89), and about half of the patients were postmenopausal at diagnosis (55%). Patients had mainly tumours with histological grade 2 (50%), hormone receptor-positive (72%) and HER2 negative (70%). Most of them received adjuvant therapy: chemotherapy (71%), hormonotherapy (68%), or radiotherapy (78%). None of the patients stopped the treatment unless they had progression. Supplementary Table 1 shows the characteristics of patients in the good and poor prognosis groups. We found a similar proportion of immunohistochemical subtypes in both cohorts (~55% luminal, ~20% HER2+ and ~15% triple-negative). Therefore, the prognosis risk in each group was likely due to clinical size and lymph node involvement and not enrichment in different biological subtypes. Patients from the poor prognosis cohort received more adjuvant treatment than patients from the good prognosis group,

as expected by their clinicopathological characteristics (tumour size and lymph node involvement).

Epistatic interactions unveil genes contributing to metastasis susceptibility

We obtained 2016 SNPs from our genome-wide genotyping of extreme discordant phenotypes ranked by SNPRank (see SNPs with SNPRank score > 0.5 in Table 1 and the 2016 SNPs in Supplementary Table 2). According to their minor allele frequencies (MAF) obtained from the 1000 Genomes Project, these SNPs are frequent in the population (91% of the SNPs had MAF greater than 5% and thus can be considered very common) and not different from what is expected by chance (Chi-square test, p-value = 0.76). However, SNPs at the top of the rank (i.e., those in Table 1) have lower MAF than the complete list of SNPs (median 0.09; IQR: 0.05, 0.17; Welch's t-test p-value < 0.0001). Some of these top SNPs are in or near genes associated with either enhanced metastatic dissemination or with reduced metastatic ability–see, for instance, *PIK3C2B* [20], *ZAP70* [21] and *VIP* [22].

Previous studies that tried identifying germline variants associated with breast cancer survival did not find any that reached genome-wide significance [23–25]. In line with these results, we did not find any

germline variants that showed a strong single-effect on the predisposition to metastasis at the required GWA statistical significance (p-value $< 5 \times 10^{-8}$). Similarly, when looking for these 2016 SNPs in dbSNP, only ten were previously associated with breast cancer two were associated with breast cancer risk, and none were associated with metastasis. This underannotation could be because dbSNP includes SNPs found in tumour/normal tissue, i.e., somatic variants, while we are investigating germline variants that have not been described previously. Another possibility is that none of these variants has a strong single-effect, as shown by previous GWAS. Thus, our genotype analysis suggests that germline variants do not affect the susceptibility to metastasis by acting individually on a few genes; they act in coordination over many genes.

We looked for germline variants that affect the susceptibility to metastasis through epistatic interactions to study this pervasive action of germline variants over many genes. These are statistical interactions between loci in their effect on a trait such that the impact of a particular single-locus genotype depends on the genotype at other loci. Therefore, the germline variants in our cohort might affect genes that are associated with regulatory networks and pathways, driving the susceptibility to develop metastasis. We modelled the gene epistasis network that encodes the susceptibility to develop metastasis in our cohort and identified the core genes that direct the susceptibility to metastasis by the community centrality measure.

The epistasis network contains 1428 genes and ca. 5600 links among them. It is a large and dense network (i.e., there are many links among genes), which further emphasises the polygenic nature of the germline contribution to a complex trait such as metastasis. It is also a small-world network, meaning that most genes can be reached from every other gene in a few steps and that genes are tightly interconnected, forming communities. The sheer number of nodes and edges obscure essential features such as the nodes that support the integrity of connections—around which the network is organised. The centrality of these nodes is an indicator of a particular node's relevance to the network's large-scale structure, which helps us prioritise nodes and identify the essential genes in the epistasis network. The epistasis network is modular, i.e., it is formed of communities that group together highly interconnected nodes representing related genes that work together. The interactions within a community are somehow autonomous from interactions in other communities; thus, in the epistasis network, each community embodies a different aspect of the susceptibility to metastasis. In such a network, the nodes that participate in several communities partake in most interactions throughout the network, connecting different communities otherwise isolated and maintaining the global network structure. These are the core genes expected to play a direct role in metastasis by their influence on many other genes—which germline variants might also perturb—that either promote the migration of tumour cells or favour the seeding of cells disseminated from the primary tumour in target tissues. Figure 1 shows the epistasis network and illustrates the community centrality using AR and TSHZ2, two examples of genes influencing the susceptibility to metastasis because they connect several communities in the epistasis network.

We found that the top 10% of community-central genes are overrepresented in KEGG pathways (over-representation test; multiple comparisons adjusted and false discovery rate controlled), such as the interaction with extracellular matrix receptors (KEGG: hsa04512; q-value = 0.018) and the establishment of cell-extracellular matrix contact points (KEGG: hsa04510; q-value = 0.017).
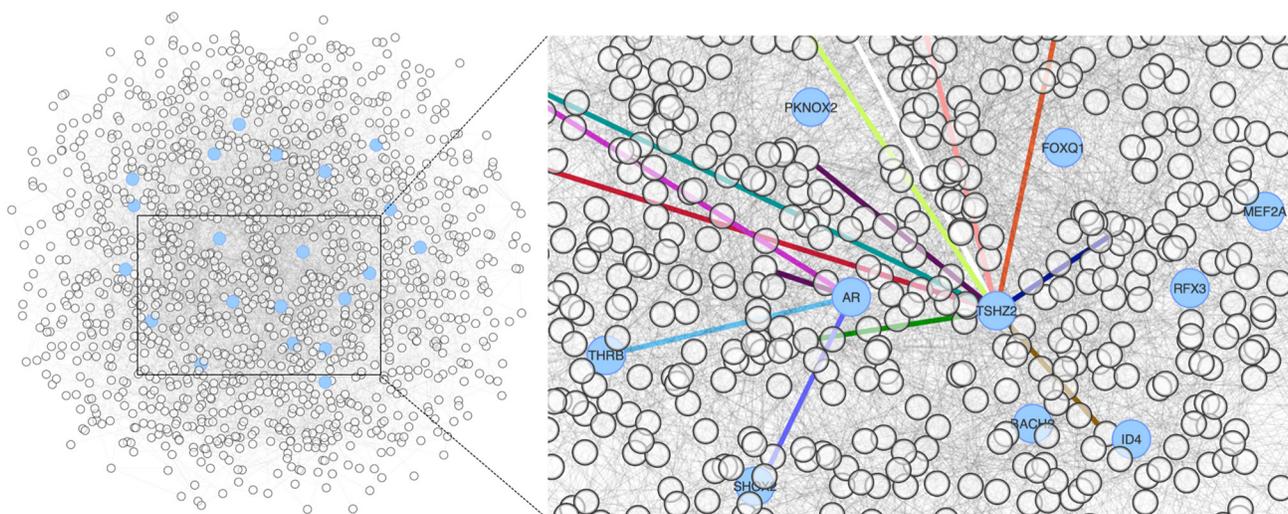
Figure 1: Epistasis network encoding the susceptibility to metastasis in our cohort. The genes with high community centrality are represented in blue. The right panel highlights the participation of two community-central genes in several communities by the colour of their links.

This implies that the genetic variants of breast cancer patients who tend to develop metastasis affect genes mechanistically involved in metastasis.

The preliminary analysis of our genotyping data suggests that the community-central genes integrate the communities effectively in the topology of the epistasis network. Therefore, the community-central genes contribute extensively to metastasis by influencing many other genes implicated in metastasis.

Genes influence breast cancer metastasis through gene regulation

According to the seed and soil hypothesis, we postulate that the community-central genes partake in metastasis either by being part of the metastatic seeds—expressed in the tumour or its microenvironment—or by priming the congenial soil—i.e., they are expressed in non-tumour tissue. Our analysis of the epistatic interactions suggests that germline variants affect genes expressed either in the tumour and its microenvironment or target tissues, making the primary tumour more prone to develop metastasis. Since a tumour's capability to metastasise depends on the coordinated gene expression present early in tumour development, we also postulate that the genes harbouring germline variants will be critical players in gene regulation. To analyse the regulatory role of the community-central genes, we have modelled a gene regulatory network for metastasis in breast cancer.

We modelled a metastasis gene regulatory network by expanding a set of metastasis genes on a breast cancer gene regulatory network that contained 254 TFs, 3178 target genes and 9414 TF-target interactions. These metastasis genes are dysregulated in metastatic tumours and responsible for the dedifferentiation to cancer stem cells, for stem cells are crucial in establishing the premetastatic niche—they mobilise and eventually arrive in and manipulate the secondary microenvironment in sites that will become metastases [26, 27]. The network comprises 142 transcription factors that regulate the expression of 373 target genes. Although the network is dense, connections are not evenly distributed, for the metastasis regulatory network is scale-free. That means that there are a few nodes with many connections, and most nodes have very few connections. For example, the median number of connections for a gene in the metastasis regulatory network is 4, but 9 genes (all of them are transcription factors) have more than ten times the median number of connections and 81% of the genes have lower than twice the median number of connections. Since all the connections come from TF and end up in target genes, the TF regulating many target genes have the most outgoing connections. In contrast, genes subjected to extensive regulation have many incoming connections in our metastasis regulation network. This handful of TF and target genes holds a privileged position in gene regulation.

The metastasis regulatory network is modular. That means that genes interact more closely within the community than with other parts of the network. Communities in a network tend to be associated with a biological function. Therefore, in our metastasis regulatory network, the communities are associated with processes and biological functions related to metastasis.

Figure 2 illustrates the composition of the network in communities. We can see that the yellow community (31 genes) is enriched in genes regulated by E2F—which promotes metastasis in breast cancer [28].

The genes regulated by MYC form the blue community (16 genes), whereas the dark yellow (43 genes) and purple (67 genes) communities are formed by genes associated with adipogenesis and cell cycle checkpoints, respectively. Both processes are relevant in metastasis. Adipogenesis is related to metastasis in triple-negative breast cancer through different mechanisms (Oshi et al., 2021). The dysregulation of genes involved in the cell cycle checkpoints is associated with aggressive cellular behaviour, including invasion- and metastasis-associated changes [29]. The genes involved in the response to Interferon gamma (IFN-γ) signalling arrange in the small light blue community (9 genes). The intensity of the IFN-γ signalling can describe the pro-metastatic role of tumours since tumours treated with low-dose IFN-γ acquired metastatic properties while those infused with high dose led to tumour regression [30].

Chromosome instability is another tumour feature that leads to the metastatic phenotype [31–33]. Our metastasis regulatory network captures chromosome instability in the green community (72 genes). Finally, the two processes are more clearly associated with metastasis, the epithelial-mesenchymal transition (EMT) and the extracellular matrix's remodelling. These two processes are fundamental for the metastasis hallmark of motility and invasion (Welch and Hurst, 2019). They are present in our regulatory network across the red (55 genes) and orange (48 genes) communities, respectively.

More (smaller) communities on the network are not highlighted in Figure 2 because they are loosely related to metastasis. These results show that we have modelled a gene regulatory network that encodes the complex regulation of breast cancer metastasis.

Since inherited genetic variants tend to associate with complex phenotypes through the regulation of gene expression [34], we assessed whether the genes with germline variants might regulate the expression of tumour genes, thus influencing metastasis. We have used the gene regulatory network focused on metastasis in breast cancer to determine genes harbouring germline variants of metastasis susceptibility that effectively participate in metastasis. We mapped our community-central genes onto the metastasis regulatory network and analysed their relevance in the network topology, which indicates their importance in metastasis. Thirty-nine genes out of the

1428 genes present on the epistasis network map to the metastasis regulatory network. Sixteen of them are TFs that target 23 genes in the metastasis regulatory network. These transcription factors tend to have larger regulons in the breast cancer gene regulatory network than the rest of the transcription factors in the network (average regulon size of 40 and 30, respectively). The 23 target genes are not more extensively regulated than the rest of the target genes in the network; on average, three transcription factors regulate each gene.

The 39 genes that have germline variants associated with metastasis and that participate in the regulation of metastasis tend to be in the communities highlighted in Figure 2. Half of the genes are in the communities associated with the EMT and the extracellular matrix reorganisation; six genes are in the community associated with the alterations in cell cycle checkpoints. These processes are essential for the dispersion of the metastatic seeds and for establishing the pre-metastatic niche.

The TFs and the targets of regulation that bear germline variants could participate in breast cancer metastasis by regulating many other genes. Therefore, they are representative of the host genetic makeup that makes some breast cancer patients more susceptible to develop metastasis. We postulate that these regulators and regulated genes influence the predisposition to develop metastasis in breast cancer patients, and thus we termed them metastasis influence genes. The rest of the paper focuses on how these 39 genes partake in metastasis.

The metastasis regulatory network has a bow-tie structure, i.e., a structure in which the genes that form a tightly interconnected inner core facilitate the effective communication between the genes in the periphery of the network, the TFs and the genes under their regulation. The metastasis influence transcription factors are significantly overrepresented on the bow-tie core of the network (one proportion z-test, p-value < 0.05). Of the five transcription factors that harbour germline variants of
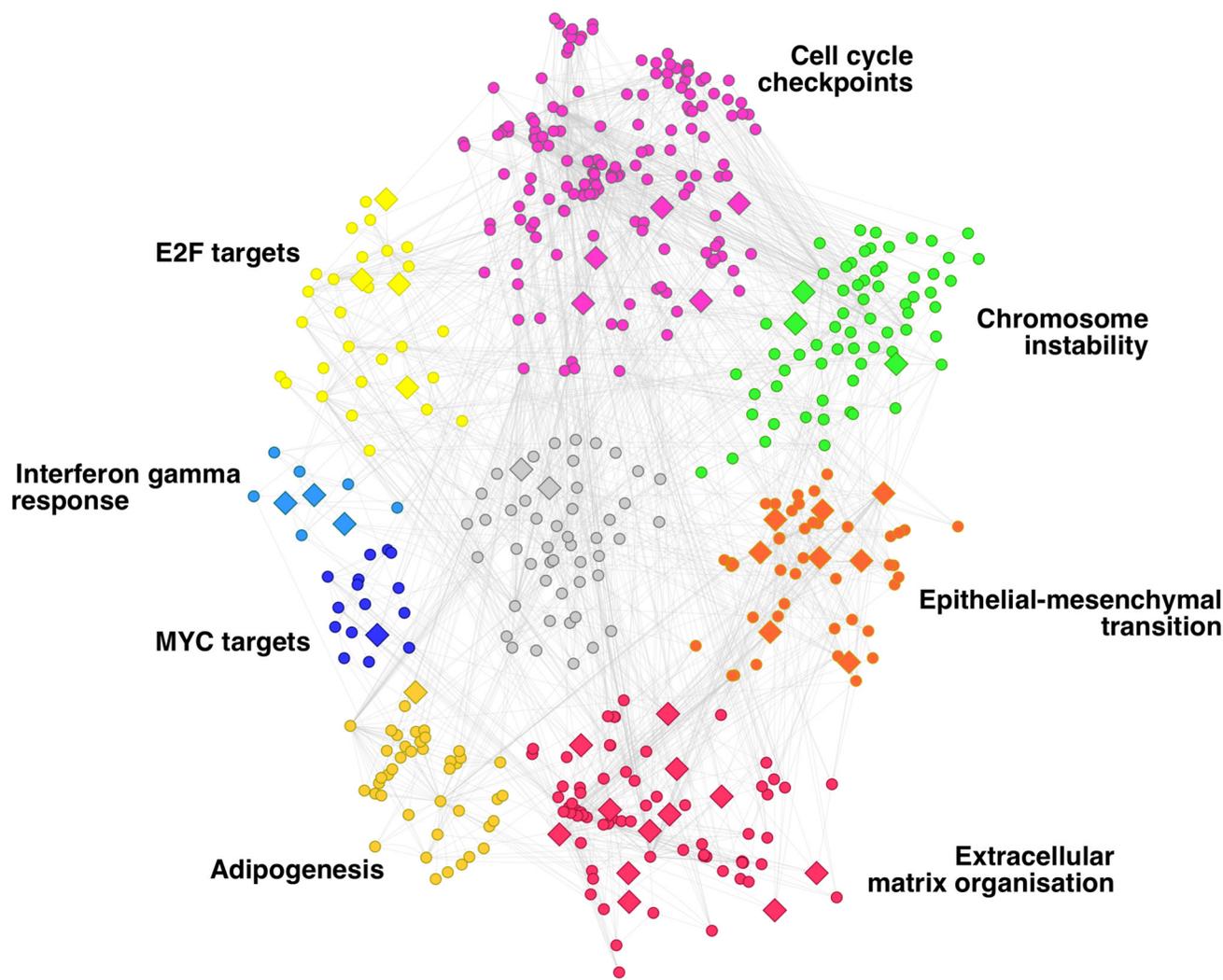
Figure 2: Gene regulatory network of breast cancer metastasis. Network communities are depicted in different colours and annotated according to the enriched functions of their genes.

metastasis susceptibility, TSHZ2 is an important regulator that participates in breast cancer and metastasis [35]. Our metastasis influence genes are significantly central (one proportion z-test, p-value < 10^{-6}), which means that they tend to be relevant in the network, i.e., the network revolves around these genes. The relevance in the network translates to importance in the phenotype encoded in the network [36]. Therefore, our network analysis suggests that we have selected the genes that contribute most effectively to metastasis among those that have germline variants associated with metastasis susceptibility.

Metastasis influence genes are expressed across all breast cancer subtypes and in metastatic breast cancer cell lines

Breast cancer molecular subtypes are associated with survival and patterns of distant metastasis. For example, the luminal A subtype is associated with the longest survival times, followed by luminal B, HER2-enriched and Basal-like [7, 37]. Therefore, the expression in breast tumours of our metastasis influence genes could be directed by the breast cancer subtype.

We have tested the expression of the 39 metastasis influence genes in the breast cancer cohort of the TCGA. This cohort has gene expression data for 1100 tumour samples and 112 control samples; we used both tumour and control samples to assess whether each community-central gene is expressed in breast cancer. We calculated the proportion of tumour samples for each gene in which the gene is differentially expressed compared with control samples. The gene is expressed in a subtype if the tumour expression index is higher than 0.35 in that PAM50 subtype (see Methods).

The metastasis influence genes are expressed in all the molecular subtypes, except for the transcription factors EN1 and AR and the regulated gene SMARCD3. EN1 and AR are expressed only in basal-like tumours, and SMARCD3 is not expressed in luminal-A and normal-like tumours (Supplementary Tables 3 and 4 have the Kaplan-Meier p-values for the metastasis influence genes and their regulons, respectively). Even though the molecular subtypes can have different tendencies to produce metastasis and have an assorted pattern of association with distant metastasis-free survival, our results apply to breast cancer regardless of its subtype since all the genes of interest are expressed in all the subtypes—with the caveat of EN1, AR and SMARCD3 mentioned above.

This result agrees with the idea that tumour size and lymph node involvement have a more substantial effect on metastasis than the molecular subtype of the tumour. These two factors are the most relevant in the prognosis of localized breast cancer even when the hormone receptors and HER2 status were not assessed. Carter et al., [38] showed that 5278 patients with tumours smaller than 2 cm and no lymph node involvement had a 5-year survival of 96.3%, suggesting that the metastasis-free survival was even higher. Furthermore, the prognosis of patients with more than ten metastatic axillary lymph nodes had a five-year disease-free survival of 30–39%, independently of the adjuvant treatment received and their ER status [39]. Therefore, the predictive value of the subtype (and its importance in developing metastasis) is less relevant than the tumour size and the number of lymph nodes involved, which are the criteria we have used to design our cohort of patients. That is why the metastasis influence genes we have found are evenly expressed in all the subtypes.

We compared the expression of the metastasis influence genes in metastatic vs. non-metastatic cell lines and metastatic vs. healthy mammary epithelium cell lines. The transcription factors EN2, NFE2L3 and SALL4, are upregulated in metastasis in both assessments. However, AR is upregulated in metastasis compared with the healthy mammary epithelium cell line, and genes such as EBF1 and LHX2 are upregulated in metastasis only when compared with the non-metastatic breast cancer cell line MDA-MB-468GFP.

We also assessed the differential expression of the metastasis influence genes in metastatic tumours compared with healthy tissue and non-metastatic tumours in MMTV-Wnt1 transgenic mice. Together with the expression data from cell lines, this data provides further validation of the participation in metastasis for 20 metastasis influence genes (see Table 2).

Germline variants in metastasis influence genes are associated with somatic events in cancer genes

Carter et al., [3] leveraged genotype, clinical, copy-number variation, and somatic mutation data from TCGA to search for germline variants that either: (i) predict the tissue of origin of the tumour (i.e., cancer type across the 22 types compiled in the TCGA); or (ii) are associated with somatic events (i.e., both somatic mutations and somatic copy-number changes) in cancer genes. They found 232 genes harbouring germline variants that predict breast cancer and 364 genes with germline variants associated with somatic events in cancer genes.

Since the focus of Carter's work is significantly different from ours, we do not expect a high coincidence between the germline variants associated with the origin of breast cancer or associated with somatic alterations in cancer genes and the germline variants associated with the susceptibility to develop metastasis we are looking for in this study. Nevertheless, we find substantial overlap between the genes that affect somatic events from Carter et al., and our metastasis susceptibility genes (see Table 2).

This result further supports the idea that germline variants work in collaboration with somatic alterations to promote tumour development and progress, as well as provide additional validation of the implication of our metastasis influence genes.

Table 2: Role of the metastasis influence genes on the metastasis regulatory network, their expression in models of metastatic breast tumours and their association with distant metastasis-free survival

Gene	Regulon size	Central in the metastasis regulatory network	Bow-tie core of the metastasis regulatory network	Expression in BC subtypes	Differentially expressed in metastatic tumours	Germline-somatic interaction	Associated with DMFS	Regulon associated with DMFS	Stemness phenotype	Implicated in metastasis
AR	5	yes		BL	cell line 2		–	NKI; METABRIC		[42, 79]
BACH2	3			BL, HER2, LUM, NL	mouse		–			
CALN1	NA			BL, HER2, LUM, NL	–			NA		
CDCA8	NA			BL, HER2, LUM, NL	cell line 1		NKI; METABRIC; UNT	NA	yes	[80, 81]
CLEC14A	NA	yes		BL, HER2, LUM, NL	–		–	NA		[82]
COL10A1	NA			BL, HER2, LUM, NL	cell line 2		VDX	NA	yes	[83–86]
COMP	NA			BL, HER2, LUM, NL	–	yes	TRANSBIG	NA	yes	[87, 88]
EBF1	13		yes	BL, HER2, LUM, NL	cell line 1; mouse	yes	–	MAINZ	yes	
EN1	3	yes		BL	cell line 1	yes	–	MAINZ		[89, 90]
EN2	2			BL, HER2, LUM, NL	cell line 1; cell line 2	yes	–	NKI; METABRIC		[91–93]
EXO1	NA			BL, HER2, LUM, NL	–		MAINZ; UNT	NA	yes	[94, 95]
FLI1	4			BL, HER2, LUM, NL	–		MAINZ; UNT	MAINZ; METABRIC; UNT; VDX		[96–98]
GNA14	NA			BL, HER2, LUM, NL	cell line 2	yes	–	NA		[99]
GPIHBP1	NA	yes		BL, HER2, LUM, NL	mouse		–	NA	yes	
GRM7	NA			BL, HER2, LUM, NL	–	yes	–	NA		
L3MBTL4	6		yes	BL, HER2, LUM, NL	–		UNT	–	yes	
LHX2	3			BL, HER2, LUM, NL	cell line 1		TRANSBIG; METABRIC	METABRIC		[100]
LRP1B	NA			BL, HER2, LUM, NL	–	yes	–	NA		[101–104]
LRRC4B	NA			BL, HER2, LUM, NL	–	yes	–	NA		
MEF2A	4			BL, HER2, LUM, NL	–	yes	–	TRANSBIG		[105]
METTL11B	NA	yes		BL, HER2, LUM, NL	–		–	NA	yes	
NEIL3	NA	yes		BL, HER2, LUM, NL	cell line 1		TRANSBIG; METABRIC	NA	yes	[106]
NEK2	NA	yes		BL, HER2, LUM, NL	cell line 1; mouse		NKI	NA	yes	[107, 108]
NFE2L3	3	yes		BL, HER2, LUM, NL	cell line 1; cell line 2		–	–		[109, 110]
NMNAT3	NA			BL, HER2, LUM, NL	–		–	NA		
NR3C1	5		yes	BL, HER2, LUM, NL	cell line 2; mouse	yes	–	NKI; TRANSBIG; METABRIC	yes	[111–113]
RP9P	NA			BL, HER2, LUM, NL	–		–	NA		

Gene	Community centrality	Central in network	Bow-tie core	Subtype	Cell line	Germline-somatic interaction	Associated with DMFS	Associated with DMFS (regulon)	Stemness phenotype	Implicated in metastasis
RPS6KA2	NA			BL, HER2, LUM, NL	–		–	NA		[114]
SALL4	6			BL, HER2, LUM, NL	cell line 1; cell line 2	yes	–	MAINZ	yes	[115,116]
SMAD3	3			BL, HER2, LUM, NL	cell line 1		–	–		[117–119]
SMARCD3	NA			BL, HER2, LUMB	–		–	NA		[120,121]
SMYD3	6		yes	BL, HER2, LUM, NL	cell line 1		–	–		[122–124]
SPARCL1	NA	yes		BL, HER2, LUM, NL	cell line 2		NKI	NA	yes	[125–127]
STARD8	NA	yes		BL, HER2, LUM, NL	–		–	NA		[128]
TMEM132C	NA			BL, HER2, LUM, NL	–		–	NA	yes	
TNS1	NA	yes		BL, HER2, LUM, NL	cell line 1; mouse		–	NA	yes	[129,130]
TSHZ2	28		yes	BL, HER2, LUM, NL	–	yes	–	MAINZ; NKI; METABRIC	yes	[35,131]
TUBA1C	NA			BL, HER2, LUM, NL	mouse		METABRIC	NA	yes	[132,133]
ZNF385D	3			BL, HER2, LUM, NL	–	yes	–	METABRIC	yes	

The gene is central in the metastasis regulatory network if its community centrality (a measure of its importance by the number of network communities to which the gene belongs) is in the top 20%. The transcription factor is in the bow-tie core if its bow-tie score is in the top 10%. Abbreviations: BL, basal-line; HER2, HER2-enriched; LUM, luminal; LUMB, luminal-B; NL, normal-like. Cell line 1, the gene is upregulated in the metastatic breast cancer cell line MDA-MB-468GFP vs. the poorly metastatic breast cancer cell line MDA-MB-468LN; cell line 2, the gene is upregulated in the metastatic breast cancer cell line MCF7 vs. the mammary epithelium cell line MCF10. Germline-somatic interaction, the gene harbours germline variations associated with somatic events. Stemness phenotype, the gene has an expression profile across the TCGA breast tumour samples significantly correlated with the mRNA stemness index (see Methods). Associated with DMFS, the gene is significantly more associated with DMFS than random genes in the indicated expression datasets. Implicated in metastasis reference of the studies that show how the gene participates in the molecular mechanisms of metastasis.

Metastasis influence genes correlate with distant metastasis-free survival

To validate the role of the metastasis influence genes in modulating metastasis, we tested their involvement in breast cancer outcomes. Since metastasis determines the clinical outcome and survival of breast cancer patients, we used DMFS as a proxy to analyse the impact of our metastasis influence genes. We first investigated whether the expression of our metastasis influence genes was significantly altered in breast cancer patients and if their expression profiles were significantly more correlated with DMFS than random genes. We found that the association between the expression profile of a gene and the outcome is highly dependent on the cohort analysed and thus inconsistent among different gene expression datasets analysed. That means that a gene significantly associated with survival in a particular cohort will probably not be associated with survival in a different cohort. This instability and study-dependency of prognostic genes in cancer and its implication on the reliability of gene expression signatures have been studied before [40]. Venet et al., [41] showed that random gene signatures could be significantly associated with breast cancer outcome, being better outcome predictors than published signatures. We tested the 70-gene prognostic signature [5] in four gene expression data sets and found that it is significantly associated with breast cancer outcome in only two of them. This phenomenon has important implications for our work: since our metastasis influence genes do not result from analysing any gene expression dataset, they will compare unfavourably with random genes in any gene expression dataset. To provide an unbiased assessment of the association with DMFS our metastasis influence genes might have, we have tested them across several gene expression datasets (Supplementary Tables 3 and 4).

Table 2 reports the gene expression datasets in which our metastasis influence genes (or their regulons) are significantly associated with DMFS. Twenty out of 39 metastasis influence genes are associated with DMFS either by their expression or by the genes they regulate. Eleven of them are also upregulated in metastatic tumours. The androgen receptor (AR) activation regulates several pathways leading to different processes like proliferation, migration and invasiveness [42]. AR correlates with a good prognosis in ER-positive breast cancer patients and controls progression and drug resistance in ER-negative [43]. This agrees with our result that AR is upregulated in metastatic tumours, and its regulon is implicated in metastasis. The downregulation of Engrailed-2 (EN2) suppresses prostate cancer cell survival and metastasis [44], which is in concordance with our results, for EN2 is upregulated in metastatic breast cancer cell lines and regulates genes that influence metastasis. These are just two examples of how our metastasis influence genes can affect metastasis. Some of them were known to participate

in metastasis, either in breast cancer or other tumours, as shown in Table 2. Approximately half of the metastasis influence genes participate in establishing the stemness phenotype in the tumour. Since cancer stem cells have a unique role in establishing the pre-metastatic niche [26, 45, 46], this result indicates that the metastasis influence genes could be involved in how the primary tumour prepares its future metastatic niche. Our analysis shows that the metastasis influence genes participate in metastasis not just because they are altered in the tumour but also because they have germline variants that make them prone to contribute to metastasis development.

In conclusion, our work moves onward with the seed and soil hypothesis—i.e., the host's genetic background contributes to the development of metastasis. We unveiled several genes altered by germline variants that influence metastasis through their synergistic interaction with many other genes in our epistasis network. Therefore, we suggest that women who harbour specific sequence variants in the metastasis influence genes will deploy gene expression patterns that favour metastasis should they develop breast cancer.

The metastasis influence genes could affect the susceptibility to develop metastasis in two ways: either they are dysregulated in breast tumours and partake in the mechanisms of metastasis, or they regulate genes that form part of these mechanisms. Furthermore, they favour the dissemination of metastatic seeds and contribute to the congenial soil by priming the pre-metastatic niche.

MATERIALS AND METHODS

Patients and study design

We collected patients diagnosed in eight Spanish Hospitals that met the inclusion criteria. These criteria included female patients over 18 years old with histologically confirmed invasive breast cancer who had undergone surgery and had at least five years of follow-up. Patients with bilateral breast cancer and second primary tumours were excluded. All patients participating in the study gave their informed consent and protocols were approved by institutional ethical committees (Comité Coordinador de Ética de la Investigación Biomédica de Andalucía).

According to our extreme discordant phenotypes framework, we selected patients with a low risk of developing metastasis who nevertheless relapsed (good prognosis cases) and patients with a high risk of developing metastasis who did not develop a disseminated disease (poor prognosis cases). In the recently published 8th Edition Cancer Staging Manual from the American Joint Committee on Cancer, patients with tumours smaller than 2 cm and without lymph node involvement have an excellent prognosis, independently of their biology. Likewise, tumours with more than ten positive lymph nodes (pN3) are classified as stage IIIC, independently of the primary tumour size, hormone receptor and HER2 status. These patients had a 5-year disease-free survival (DFS) of 30-39%, independently of the adjuvant treatment received and their oestrogen receptor (ER) status [39]; in this set of patients, the absence of relapse within five years was a rare event. Therefore, our cohort encompasses breast cancer patients that, regardless of their histological subtype, are either good prognosis cases, i.e., patients with tumours smaller than 2 cm and no lymph nodes affected who relapsed within five years after surgery; or poor prognosis cases, i.e., patients with more than ten lymph nodes affected regardless their tumour size who did not relapse within five years after surgery.

Samples and genotyping

Genomic DNA was extracted from 3 mL of peripheral blood using the QiaAmp DNA Blood Mini Kit (Qiagen). The Human Genotyping Unit-CeGen CNIO conducted the genome-wide genotyping using the Illumina Infinium LCG Quad Assay protocol with the HumanOmni5-Quad Beadchip (Illumina). This chip contains ca. 4.3 million SNPs selectively distributed and separated by an average distance of 0.68 kb. The scanned signal raw intensities from all SNPs in the assay were analysed using the GenomeStudio software (Illumina). We filtered the data for quality control using the open-source tool PLINK [47]. We did not exclude any patient from the study due to low genotyping (call-rate < 90%). SNPs were excluded if they had a call-rate < 90% or a Hardy-Weinberg equilibrium p-value $< 10^{-6}$.

Epistasis network analysis

We looked for germline variants with a robust individual effect on susceptibility to metastasis (i.e., with a p-value $< 5 \times 10^{-8}$) by performing an association analysis between SNPs in the good and poor prognosis cases with the PLINK library within the Encore pipeline [48].

We modelled and analysed a genetic interaction network from our genetic population data with the Encore pipeline. Figure 3 depicts the Encore workflow to model the epistasis network from genotyping data. Encore is an open-source tool for the analysis of biological data with the power to detect genetic variants relevant to a phenotype using genetic epistasis; it discovers variants without a substantial individual effect but whose relevance to the phenotype comes from their multiple interactions [49]. Encore focuses on common and rare variants to identify susceptibility hubs or groups of variants with numerous connections that influence the phenotype. To characterise these hubs, Encore computes a genetic association interaction matrix (reGAIN matrix) that ranks the variants according to their connection with other variants with the algorithm SNPrank [48]. Therefore, we

obtain a list of ordered variants based on their importance to the phenotype of interest, which in our study is susceptibility to metastasis. With this list of SNPs and the reGAIN matrix, we identified the genes that harbour the most relevant SNPs (top SNPs). We modelled the gene epistasis network for the susceptibility to metastasis, keeping only significant epistatic interactions (Benjamini-Hochberg false discovery rate corrected p-value < 0.01).

According to network theory, a network revolves around a set of nodes termed central nodes. These central nodes capture the information flow represented in the network, and there are many ways to determine them [52]. In our gene epistasis network, central genes contain most of the metastasis susceptibility information through their interactions. To identify these central genes, we used the community centrality, which measures the importance of a gene by the number of network communities to which the gene belongs. We performed all the network analyses with the iGraph package [53] and the R platform for statistical computing.

Gene expression in breast tumour samples, animal models and cell lines

We downloaded the RNASeq normalised gene expression dataset for breast cancer from the TCGA [54] with the R library RTCGAToolbox [55]. We transformed the RNASeq data to Z-scores so that per each tumour sample for each gene, we measured how many standard deviations (sd) away from the mean that gene expression is. We considered those with $Z > 1.96$ (roughly p-value < 0.05 or more than 2 sd away) to be differentially expressed. We considered the tumour t and control c samples in the

calculation of Z-scores for each gene g using the following equation:

$$Z_t^g = \frac{E_t^g - \langle E_c^g \rangle}{\sigma_c^g}$$

We thus compare the expression of gene g in the tumour sample t with the average and standard deviation of g in control samples. The gene's tumour expression index is the proportion of tumour samples in which the gene is differentially expressed (i.e.) $|Z| > 1.96$. We established a threshold for the tumour expression index using a random model of 10000 genes: less than 5% of random genes have a tumour expression index higher than 0.35. Therefore, we considered a gene expressed in tumour samples if its tumour expression index was higher than 0.35.

We performed differential expression analyses in animal models and breast cancer cell lines. We used microarray expression data from the Gene Expression Omnibus (GEO) dataset GSE84917 for MMTV-Wnt1 transgenic mice to compare the expression profiles of metastatic versus non-metastatic mammary tumours and metastatic mammary tumours versus healthy mammary tissue. We considered genes dysregulated if their logFC > 1 and their Benjamini and Hochberg false discovery rate (FDR) adjusted p-value < 0.01 with the limma 3.46 library on R 4.0.2 [56].

MDA-MB-468GFP is a poorly metastatic cell line; however, it has a variant (MDA-MB-468LN) with high metastatic ability. We compared the expression profiles of non-metastatic vs. metastatic tumours using the microarray expression data for these two cell lines in the GEO dataset GSE11683. We performed a differential expression

1. Linkage disequilibrium prunning	2. Filter by evaporative cooling	3. Build a reGAIN matrix and filter with SNPrank	4. Associate SNPs with genes	5. Ranking genes by community centrality
⇩	⇩	⇩	⇩	⇩
Removes highly correlated variants	Removes SNPs irrelevant to the phenotype	Matrix with SNP-SNP interactions to determine the most relevant SNPs	Identify genes in/ near the most relevant SNPs	The most important genes participate in more communities

Figure 3: A pipeline of the epistasis network modelling with Encore. We used as input .bim/.bed/.bam files from PLINK. 1) The linkage disequilibrium pruning step removes highly correlated (i.e. low informative) SNPs. 2) Evaporative cooling is a machine learning method that integrates multiple importance scores while removing irrelevant genetic variants. In this step, we kept the 10000 most relevant SNPs, which constitutes a significant reduction from the initial ~ 4.3 million. 3) After filtering, Encore calculates the pairwise interaction for the 10000 SNPs with a generalised linear model. It computes a matrix of epistatic interactions among SNPs with Benjamini-Hochberg false discovery rate corrected p-values (reGAIN matrix). From that matrix, SNPs are ranked and filtered with SNPrank; we kept 2016 SNPs. 4) We obtained the names of the genes in or near (1 MB) the most relevant SNPs with the R library PostGWAS [50]. 5) Finally, we ranked the most relevant genes by their community centrality (using link communities [51]); genes are important if they participate in many communities.

analysis (logFC > 1, Benjamini and Hochberg FDR adjusted *p*-value < 0.01) with the limma library in R 4.0.2.

We also compared the expression profiles of metastatic tumours and healthy mammary epithelium using the cell lines MCF7 and MCF10A, respectively. MCF7 is a transformed breast cancer cell line derived from a metastatic site, and MCF10A is a normal-like mammary epithelial cell line. We performed a differential expression analysis (logFC > 1, Benjamini and Hochberg FDR adjusted *p*-value < 0.01) with the expression data from the GEO dataset GSE71862 [57], which contains RNA-seq data for these two cell lines. We calculated the differential gene expression using the DESeq2 version 1.30.1 [58] library in R 4.0.2.

Map of gene regulation in metastasis

To represent the map of gene regulatory interactions in breast cancer metastasis, we have modelled a transcriptional regulatory network focused on metastasis. We started by building a broad gene regulatory network for breast cancer from the collection of 1612 transcription factors (TF) compiled in [59] and the genes controlled by those TFs, which we obtained from the TCGA (RNASeq breast cancer gene expression dataset). We used the RTN pipeline [60] to reconstruct transcriptional regulatory networks.

A transcriptional regulatory network consists of a collection of TFs and their regulated target genes. Each TF guides the expression of a set of genes, which forms a regulon. Therefore, TFs are regulators that either activate or repress the expression of the target genes. The RTN pipeline first computes the interactions between each TF and all potential target genes through the mutual information between a regulator and all potential targets— i.e., the mutual dependence between the expression profiles of the TF and their targets. Then, it performs a bootstrapping analysis to remove non-significant and unstable TF-target interactions. Each target gene may be linked to many TFs at this stage because regulation can occur through direct interactions between a transcription factor and a target gene and indirect interactions (TF-TF-target). The final step in the RTN pipeline is the ARACNe algorithm [61] to remove the weakest interaction in any triplet formed by two TFs and a common target gene, preserving the dominant TF-target pair.

From this broad network, we wanted to model a subnetwork centred on metastasis, that is, the part of the general network that contains the TFs and their regulons involved in the regulation of metastasis. Based on the idea that the network neighbourhood of a set of genes contains information about the biological processes in which the genes participate [62], we started from a set of genes involved in metastasis and characterised its network neighbourhood to obtain the gene regulatory network for metastasis in breast cancer.

We obtained the genes implicated in breast cancer metastasis from three sources: genes differentially expressed in metastasis samples, genes involved in the stemness phenotype, and genes dysregulated in metastasis through the metastasis expression index.

We compared the expression of metastatic samples (from both local and distant metastasis) with the expression of healthy control tissue from the TCG to obtain 65 differentially expressed genes (logFC ≥ 5; *p*-value < 0.0001). These genes were enriched in KEGG pathways related to metastasis, such as ECM-receptor interaction, IL-17 signalling and PPAR signalling.

Cancer stem cells are responsible for recurrence, relapse and metastasis [63]. Breast cancer metastasis involves acquiring stem-cell-like features characterised by the expression of markers that contribute toward a stemness phenotype [64]. Malta et al., [65] found that the stemness phenotype was generally most prominent in metastatic tumours and developed a stemness index for assessing this phenotype. We have used their mRNA stemness index and the weighted gene correlation analysis [66, 67] to find 75 genes whose expression profiles were significantly correlated with the stemness phenotype (*p*-value < 10^{-10}). As expected, these genes are enriched in cell cycle-related biological processes and pathways related to metastasis and breast cancer progression, such as the oestrogen-responsive protein and Sonic Hedgehog signalling.

We developed the metastasis expression index analogous to the tumour expression index to obtain additional genes implicated in breast cancer metastasis. In this case, we obtained a Z-score for each gene in each breast metastasis sample from the TCGA RNASeq gene expression dataset by comparing the gene expression in the metastasis sample with the average and standard deviation of the gene expression in tumour samples. Thus, we obtained 121 genes dysregulated in metastatic breast tumours that were enriched in processes related to metastasis, such as the activation of epithelial cell proliferation and Wnt signalling.

We mapped the 261 genes related to metastasis onto the general breast cancer gene regulatory network. We used the DIAMOnD network diffusion algorithm [68] to obtain the network neighbourhood of these genes. DIAMOnD evaluates the significance of the connections that the initial set of genes has in the network to incorporate those genes better connected with the initial set. Therefore, with enough iterations of the algorithm (200 iterations), we obtained the subnetwork that comprises the initial set of 261 genes and their network neighbourhood, resulting in a gene regulatory network of breast cancer metastasis.

The metastasis regulatory network is modular. Each community of genes that interact more closely among them than with the rest of the network tends to encode a particular feature of the phenotype encoded in the network. We have highlighted those communities associated with metastatic processes through the over-representation test

(multiple false discovery rate controlled; q-values < 0.001) on the Gene Ontology, REACTOME and the MSigDB hallmark gene set collection [69].

Gene regulatory networks often exhibit a bow-tie topology [70]. The presence of a robust interconnected core characterises the topology of these networks; this core is essentially a set of genes that communicate the fan-in component of source nodes (the transcription factors) with the fan-out component of sink nodes (i.e., the target genes). The core of the bow-tie structure reduces the number of genes and connections required to connect the transcription factors with the target genes, decreasing perturbation and noise [71]. We can identify the genes that belong to the core of the network by the bow-tie score [72]:

$$b(v) = \frac{S_v T_v}{S\ T}$$

Where S_v is the number of source nodes (i.e., transcription factors) that can reach the gene v, T_v is the number of target genes that v can reach, S and T are the total number of source and target nodes, respectively. We have implemented the bow-tie score to characterise the topology of the metastasis gene regulatory network.

Survival analysis

We identified the transcription factors among our genes of interest and their regulons on our gene regulatory map. We tested the association of our genes and their regulons (when they are transcription factors) with breast cancer outcomes using the Kaplan-Meier survival analysis with the SigCheck R library [73]. For each gene and each regulon, we tested whether they were more significantly associated with distant metastasis-free survival (DMFS) than random genes (comparative survival analyses with random genes), on six gene expression datasets of breast cancer: NKI, 319 samples [5]; METABRIC, 1422 samples [74]; TRANSBIG, 198 samples [75]; MAINZ, 200 samples [76]; UNT, 125 samples [77]; and VDX, 344 samples [78]. The algorithm computes the mean expression value for each sample across the regulon (or the gene) in each independent survival analysis, which allows dividing the samples into a high expression group and a low expression group. Comparing the survival curves of these two groups results in a p-value that indicates the confidence that the samples are separable into groups with distinct survival outcomes. The comparative survival analysis produces an empirical p-value of the performance of genes or regulons against random genes in 1000 independent survival analyses.

Author contributions

EA supervised the study. AS, AAMG and EA conceived the study. NR, BJR, BP, EPR, MdMM, MRB, JdlH, PSR, AR and AL recruited the patients and acquired their clinical data. AS, CFDS and AGN performed the experiments. AS, AGN and AAMG collated and analysed the data. AAMG, BP, AS and EA wrote this paper. All authors approved the paper.

CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

REFERENCES

1. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. Cell. 2017; 168:670–91. https://doi.org/10.1016/j.cell.2016.11.037. [PubMed]

2. Hunter K. The role of individual inheritance in tumor progression and metastasis. J Mol Med (Berl). 2015; 93:719–25. https://doi.org/10.1007/s00109-015-1299-6. [PubMed]

3. Carter H, Marty R, Hofree M, Gross AM, Jensen J, Fisch KM, Wu X, DeBoever C, Van Nostrand EL, Song Y, Wheeler E, Kreisberg JF, Lippman SM, et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. Cancer Discov. 2017; 7:410–23. https://doi.org/10.1158/2159-8290.CD-16-1045. [PubMed]

4. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002; 415:530–36. https://doi.org/10.1038/415530a. [PubMed]

5. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002; 347:1999–2009. https://doi.org/10.1056/NEJMoa021967. [PubMed]

6. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res. 2005; 65:9155–58. https://doi.org/10.1158/0008-5472.CAN-05-2553. [PubMed]

7. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001; 98:10869–74. https://doi.org/10.1073/pnas.191367098. [PubMed]

8. Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, Hunter KW. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. Int J Cancer. 1998; 77:640–44. https://doi.org/10.1002/(sici)1097-0215(19980812)77:4<640::aid-ijc26>3.0.co;2-8. [PubMed]

9. Hemminki K, Ji J, Försti A, Sundquist J, Lenner P. Survival in breast cancer is familial. Breast Cancer Res Treat. 2008; 110:177–82. https://doi.org/10.1007/s10549-007-9692-7. [PubMed]

10. Xu X, Li J, Zou J, Feng X, Zhang C, Zheng R, Duanmu W, Saha-Mandal A, Ming Z, Wang E. Association of Germline Variants in Natural Killer Cells With Tumor Immune Microenvironment Subtypes, Tumor-Infiltrating Lymphocytes, Immunotherapy Response, Clinical Outcomes, and Cancer Risk. JAMA Netw Open. 2019; 2:e199292. https://doi.org/10.1001/jamanetworkopen.2019.9292. [PubMed]

11. Xu X, Zhou Y, Feng X, Li X, Asad M, Li D, Liao B, Li J, Cui Q, Wang E. Germline genomic patterns are associated with cancer risk, oncogenic pathways, and clinical outcomes. Sci Adv. 2020; 6:eaba4905. https://doi.org/10.1126/sciadv.aba4905. [PubMed]

12. Ribelles N, Santonja A, Pajares B, Llácer C, Alba E. The seed and soil hypothesis revisited: current state of knowledge of inherited genes on prognosis in breast cancer. Cancer Treat Rev. 2014; 40:293–99. https://doi.org/10.1016/j.ctrv.2013.09.010. [PubMed]

13. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. https://doi.org/10.1038/nature08494. [PubMed]

14. Pérez-Gracia JL, Gúrpide A, Ruiz-Ilundain MG, Alfaro Alegría C, Colomer R, García-Foncillas J, Melero Bermejo I. Selection of extreme phenotypes: the role of clinical observation in translational research. Clin Transl Oncol. 2010; 12:174–80. https://doi.org/10.1007/s12094-010-0487-7. [PubMed]

15. Hu T, Sinnott-Armstrong NA, Kiralis JW, Andrew AS, Karagas MR, Moore JH. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. BMC Bioinformatics. 2011; 12:364. https://doi.org/10.1186/1471-2105-12-364. [PubMed]

16. Hunter K, Welch DR, Liu ET. Genetic background is an important determinant of metastatic potential. Nat Genet. 2003; 34:23–24. https://doi.org/10.1038/ng0503-23b. [PubMed]

17. Mackay TF, Moore JH. Why epistasis is important for tackling complex human disease genetics. Genome Med. 2014; 6:124. https://doi.org/10.1186/gm561. [PubMed]

18. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet. 2014; 15:722–33. https://doi.org/10.1038/nrg3747. [PubMed]

19. Agarwal D, Nowak C, Zhang NR, Pusztai L, Hatzis C. Functional germline variants as potential co-oncogenes. NPJ Breast Cancer. 2017; 3:46. https://doi.org/10.1038/s41523-017-0051-5. [PubMed]

20. Domin J, Harper L, Aubyn D, Wheeler M, Florey O, Haskard D, Yuan M, Zicha D. The class II phosphoinositide 3-kinase PI3K-C2beta regulates cell migration by a PtdIns3P dependent mechanism. J Cell Physiol. 2005; 205:452–62. https://doi.org/10.1002/jcp.20478. [PubMed]

21. Fu D, Liu B, Zang LE, Jiang H. MiR-631/ZAP70: A novel axis in the migration and invasion of prostate cancer cells. Biochem Biophys Res Commun. 2016; 469:345–51. https://doi.org/10.1016/j.bbrc.2015.11.093. [PubMed]

22. Vacas E, Arenas MI, Muñoz-Moreno L, Bajo AM, Sánchez-Chapado M, Prieto JC, Carmena MJ. Antitumoral effects of vasoactive intestinal peptide in human renal cell carcinoma xenografts in athymic nude mice. Cancer Lett. 2013; 336:196–203. https://doi.org/10.1016/j.canlet.2013.04.033. [PubMed]

23. Azzato EM, Tyrer J, Fasching PA, Beckmann MW, Ekici AB, Schulz-Wendtland R, Bojesen SE, Nordestgaard BG, Flyger H, Milne RL, Arias JI, Menéndez P, Benítez J, et al, and Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. Association between a germline OCA2 polymorphism at chromosome 15q13.1 and estrogen receptor-negative breast cancer survival. J Natl Cancer Inst. 2010; 102:650–62. https://doi.org/10.1093/jnci/djq057. [PubMed]

24. Pirie A, Guo Q, Kraft P, Canisius S, Eccles DM, Rahman N, Nevanlinna H, Chen C, Khan S, Tyrer J, Bolla MK, Wang Q, Dennis J, et al, and kConFab Investigators, and NBCS Investigators. Common germline polymorphisms associated with breast cancer-specific survival. Breast Cancer Res. 2015; 17:58. https://doi.org/10.1186/s13058-015-0570-7. [PubMed]

25. Scannell Bryan M, Argos M, Andrulis IL, Hopper JL, Chang-Claude J, Malone K, John EM, Gammon MD, Daly M, Terry MB, Buys SS, Huo D, Olopade O, et al. Limited influence of germline genetic variation on all-cause mortality in women with early onset breast cancer: evidence from gene-based tests, single-marker regression, and whole-genome prediction. Breast Cancer Res Treat. 2017; 164:707–17. https://doi.org/10.1007/s10549-017-4287-4. [PubMed]

26. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, Psaila B, Kaplan RN, Bromberg JF, Kang Y, Bissell MJ, Cox TR, Giaccia AJ, et al. Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer. 2017; 17:302–17. https://doi.org/10.1038/nrc.2017.6. [PubMed]

27. Sanmartin MC, Borzone FR, Giorello MB, Pacienza N, Yannarelli G, Chasseing NA. Bone marrow/bone pre-metastatic niche for breast cancer cells colonization: The role of mesenchymal stromal cells. Crit Rev Oncol Hematol. 2021; 164:103416. https://doi.org/10.1016/j.critrevonc.2021.103416. [PubMed]

28. Hollern DP, Swiatnicki MR, Rennhack JP, Misek SA, Matson BC, McAuliff A, Gallo KA, Caron KM, Andrechek ER. E2F1 Drives Breast Cancer Metastasis by Regulating the Target Gene FGF13 and Altering Cell Migration. Sci Rep. 2019; 9:10718. https://doi.org/10.1038/s41598-019-47218-0. [PubMed]

29. Adikes RC, Kohrman AQ, Martinez MAQ, Palmisano NJ, Smith JJ, Medwig-Kinney TN, Min M, Sallee MD, Ahmed OB, Kim N, Liu S, Morabito RD, Weeks N, et al. Visualizing the metazoan proliferation-quiescence decision in vivo. Elife. 2020; 9:e63265. https://doi.org/10.7554/eLife.63265. [PubMed]

30. Jorgovanovic D, Song M, Wang L, Zhang Y. Roles of IFN-γ in tumor progression and regression: a review. Biomark Res. 2020; 8:49. https://doi.org/10.1186/s40364-020-00228-x. [PubMed]

31. Gao C, Su Y, Koeman J, Haak E, Dykema K, Essenberg C, Hudson E, Petillo D, Khoo SK, Vande Woude GF. Chromosome instability drives phenotypic switching to metastasis. Proc Natl Acad Sci U S A. 2016; 113:14793–98. https://doi.org/10.1073/pnas.1618215113. [PubMed]

32. Roschke AV, Glebov OK, Lababidi S, Gehlhaus KS, Weinstein JN, Kirsch IR. Chromosomal instability is associated with higher expression of genes implicated in epithelial-mesenchymal transition, cancer invasiveness, and metastasis and with lower expression of genes involved in cell cycle checkpoints, DNA repair, and chromatin maintenance. Neoplasia. 2008; 10:1222–30. https://doi.org/10.1593/neo.08682. [PubMed]

33. Novikov NM, Zolotaryova SY, Gautreau AM, Denisov EV. Mutational drivers of cancer cell migration and invasion. Br J Cancer. 2021; 124:102–14. https://doi.org/10.1038/s41416-020-01149-0. [PubMed]

34. Bai L, Yang HH, Hu Y, Shukla A, Ha NH, Doran A, Faraji F, Goldberger N, Lee MP, Keane T, Hunter KW. An Integrated Genome-Wide Systems Genetics Screen for Breast Cancer Metastasis Susceptibility Genes. PLoS Genet. 2016; 12:e1005989. https://doi.org/10.1371/journal.pgen.1005989. [PubMed]

35. Uribe ML, Dahlhoff M, Batra RN, Nataraj NB, Haga Y, Drago-Garcia D, Marrocco I, Sekar A, Ghosh S, Vaknin I, Lebon S, Kramarski L, Tsutsumi Y, et al. TSHZ2 is an EGF-regulated tumor suppressor that binds to the cytokinesis regulator PRC1 and inhibits metastasis. Sci Signal. 2021; 14:eabe6156. https://doi.org/10.1126/scisignal.abe6156. [PubMed]

36. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001; 411:41–42. https://doi.org/10.1038/35075138. [PubMed]

37. Brasó-Maristany F, Paré L, Chic N, Martínez-Sáez O, Pascual T, Mallafré-Larrosa M, Schettini F, González-Farré B, Sanfeliu E, Martínez D, Galván P, Barnadas E, Salinas B, et al. Gene expression profiles of breast cancer metastasis according to organ site. Mol Oncol. 2022; 16:69–87. https://doi.org/10.1002/1878-0261.13021. [PubMed]

38. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989; 63:181–87. https://doi.org/10.1002/1097-0142(19890101)63:1<181::aid-cncr2820630129>3.0.co;2-h. [PubMed]

39. Montero AJ, Rouzier R, Lluch A, Theriault RL, Buzdar AU, Delaloge S, Bermejo B, Le M, Kau SW, Dunant A, Arriagada R, Spielmann M, Garcia-Conde J, et al. The natural history of breast carcinoma in patients with > or = 10 metastatic axillary lymph nodes before and after the advent of adjuvant therapy: a multiinstitutional retrospective study. Cancer. 2005; 104:229–35. https://doi.org/10.1002/cncr.21182. [PubMed]

40. Califano A, Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. Nat Rev Cancer. 2017; 17:116–30. https://doi.org/10.1038/nrc.2016.124. [PubMed]

41. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol. 2011; 7:e1002240. https://doi.org/10.1371/journal.pcbi.1002240. [PubMed]

42. Zhou HC, Liu CX, Pan WD, Shang LR, Zheng JL, Huang BY, Chen JY, Zheng L, Fang JH, Zhuang SM. Dual and opposing roles of the androgen receptor in VETC-dependent and invasion-dependent metastasis of hepatocellular carcinoma. J Hepatol. 2021; 75:900–11. https://doi.org/10.1016/j.jhep.2021.04.053. [PubMed]

43. Pizon M, Lux D, Pachmann U, Pachmann K, Schott D. Influence of endocrine therapy on the ratio of androgen receptor (AR) to estrogen receptor (ER) positive circulating epithelial tumor cells (CETCs) in breast cancer. J Transl Med. 2018; 16:356. https://doi.org/10.1186/s12967-018-1724-z. [PubMed]

44. Li Q, Lu S, Li X, Hou G, Yan L, Zhang W, Qiao B. Biological function and mechanism of miR-33a in prostate

cancer survival and metastasis: via downregulating Engrailed-2. Clin Transl Oncol. 2017; 19:562–70. https://doi.org/10.1007/s12094-016-1564-3. [PubMed]

45. Liu Y, Cao X. Characteristics and Significance of the Pre-metastatic Niche. Cancer Cell. 2016; 30:668–81. https://doi.org/10.1016/j.ccell.2016.09.011. [PubMed]

46. Ursini-Siegel J, Siegel PM. The influence of the pre-metastatic niche on breast cancer metastasis. Cancer Lett. 2016; 380:281–88. https://doi.org/10.1016/j.canlet.2015.11.009. [PubMed]

47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. https://doi.org/10.1086/519795. [PubMed]

48. Davis NA, Lareau CA, White BC, Pandey A, Wiley G, Montgomery CG, Gaffney PM, McKinney BA. Encore: Genetic Association Interaction Network centrality pipeline and application to SLE exome data. Genet Epidemiol. 2013; 37:614–21. https://doi.org/10.1002/gepi.21739. [PubMed]

49. McKinney BA, Lareau C, Oberg AL, Kennedy RB, Ovsyannikova IG, Poland GA. The Integration of Epistasis Network and Functional Interactions in a GWAS Implicates RXR Pathway Genes in the Immune Response to Smallpox Vaccine. PLoS One. 2016; 11:e0158016. https://doi.org/10.1371/journal.pone.0158016. [PubMed]

50. Hiersche M, Rühle F, Stoll M. Postgwas: advanced GWAS interpretation in R. PLoS One. 2013; 8:e71775. https://doi.org/10.1371/journal.pone.0071775. [PubMed]

51. Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. Nature. 2010; 466:761–64. https://doi.org/10.1038/nature09182. [PubMed]

52. Newman MEJ. Networks: An introduction. Oxford University Press; 2010. Available from http://www.worldcat.org/title/networks-an-introduction/oclc/964511577.

53. Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research. Inter J. 2005; Complex Systems:1695.

54. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. https://doi.org/10.1038/nature11412. [PubMed]

55. Samur MK. RTCGAToolbox: a new tool for exporting TCGA Firehose data. PLoS One. 2014; 9:e106397. https://doi.org/10.1371/journal.pone.0106397. [PubMed]

56. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat. 2016; 10:946–63. https://doi.org/10.1214/16-AOAS920. [PubMed]

57. Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, Browne G, van Wijnen AJ, Lian JB, Stein JL, Dekker J, Imbalzano AN, Stein GS. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. Genome Biol. 2015; 16:214. https://doi.org/10.1186/s13059-015-0768-0. [PubMed]

58. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. https://doi.org/10.1186/s13059-014-0550-8. [PubMed]

59. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The Human Transcription Factors. Cell. 2018; 172:650–65. https://doi.org/10.1016/j.cell.2018.01.029. [PubMed]

60. Castro MA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, Tilley WD, Markowetz F, Ponder BA, Meyer KB. Regulators of genetic risk of breast cancer identified by integrative network analysis. Nat Genet. 2016; 48:12–21. https://doi.org/10.1038/ng.3458. [PubMed]

61. Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A. Reverse engineering cellular networks. Nat Protoc. 2006; 1:662–71. https://doi.org/10.1038/nprot.2006.106. [PubMed]

62. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. Cell. 2011; 144:986–98. https://doi.org/10.1016/j.cell.2011.02.016. [PubMed]

63. Prasad S, Ramachandran S, Gupta N, Kaushik I, Srivastava SK. Cancer cells stemness: A doorstep to targeted therapy. Biochim Biophys Acta Mol Basis Dis. 2020; 1866:165424. https://doi.org/10.1016/j.bbadis.2019.02.019. [PubMed]

64. Zhao W, Li Y, Zhang X. Stemness-Related Markers in Cancer. Cancer Transl Med. 2017; 3:87–95. https://doi.org/10.4103/ctm.ctm_69_16. [PubMed]

65. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, et al, and Cancer Genome Atlas Research Network. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell. 2018; 173:338–54.e15. https://doi.org/10.1016/j.cell.2018.03.034. [PubMed]

66. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559. [PubMed]

67. Zhang B, Horvath S, Carlson M, Dong J, Drake T, Geschwind D, Lusis J, Li A, Mischel P, Nelson S, Yip A. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol. 2005; 4. https://doi.org/10.2202/1544-6115.1128.

68. Ghiassian SD, Menche J, Barabási AL. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Comput Biol. 2015; 11:e1004120. https://doi.org/10.1371/journal.pcbi.1004120. [PubMed]

69. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015; 1:417–25. https://doi.org/10.1016/j.cels.2015.12.004. [PubMed]

70. Ghosh Roy G, He S, Geard N, Verspoor K. Bow-tie architecture of gene regulatory networks in species of varying complexity. J R Soc Interface. 2021; 18:20210069. https://doi.org/10.1098/rsif.2021.0069. [PubMed]

71. Csermely P, London A, Wu LY, Uzzi B. Structure and dynamics of core/periphery networks. J Complex Networks. 2013; 1:93–123. https://doi.org/10.1093/comnet/cnt016.

72. Supper J, Spangenberg L, Planatscher H, Dräger A, Schröder A, Zell A. BowTieBuilder: modeling signal transduction pathways. BMC Syst Biol. 2009; 3:67. https://doi.org/10.1186/1752-0509-3-67. [PubMed]

73. Stark R, Norden J. SigCheck: Check a gene signature's prognostic performance against random signatures, known signatures, and permuted data/metadata. 2016.

74. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, et al, and METABRIC Group. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–52. https://doi.org/10.1038/nature10983. [PubMed]

75. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, et al, and TRANSBIG Consortium. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res. 2007; 13:3207–14. https://doi.org/10.1158/1078-0432.CCR-06-2765. [PubMed]

76. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrmann M. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. 2008; 68:5405–13. https://doi.org/10.1158/0008-5472.CAN-07-5206. [PubMed]

77. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006; 98:262–72. https://doi.org/10.1093/jnci/djj052. [PubMed]

78. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005; 365:671–79. https://doi.org/10.1016/S0140-6736(05)17947-1. [PubMed]

79. Liu P, Xu Y, Zhang W, Li Y, Tang L, Chen W, Xu J, Sun Q, Guan X. Prohibitin promotes androgen receptor activation in ER-positive breast cancer. Cell Cycle. 2017; 16:776–84. https://doi.org/10.1080/15384101.2017.1295193. [PubMed]

80. Peng R, Wang Y, Mao L, Fang F, Guan H. Identification of Core Genes Involved in the Metastasis of Clear Cell Renal Cell Carcinoma. Cancer Manag Res. 2020; 12:13437–49. https://doi.org/10.2147/CMAR.S276818. [PubMed]

81. Ci C, Tang B, Lyu D, Liu W, Qiang D, Ji X, Qiu X, Chen L, Ding W. Overexpression of CDCA8 promotes the malignant progression of cutaneous melanoma and leads to poor prognosis. Int J Mol Med. 2019; 43:404–12. https://doi.org/10.3892/ijmm.2018.3985. [PubMed]

82. Su C, Shi K, Cheng X, Han Y, Li Y, Yu D, Liu Z. Methylation of CLEC14A is associated with its expression and lung adenocarcinoma progression. J Cell Physiol. 2019; 234:2954–62. https://doi.org/10.1002/jcp.27112. [PubMed]

83. Yang W, Wu X, Zhou F. Collagen Type X Alpha 1 (COL10A1) Contributes to Cell Proliferation, Migration, and Invasion by Targeting Prolyl 4-Hydroxylase Beta Polypeptide (P4HB) in Breast Cancer. Med Sci Monit. 2021; 27:e928919. https://doi.org/10.12659/MSM.928919. [PubMed]

84. Liang Y, Xia W, Zhang T, Chen B, Wang H, Song X, Zhang Z, Xu L, Dong G, Jiang F. Upregulated Collagen COL10A1 Remodels the Extracellular Matrix and Promotes Malignant Progression in Lung Adenocarcinoma. Front Oncol. 2020; 10:573534. https://doi.org/10.3389/fonc.2020.573534. [PubMed]

85. Li T, Huang H, Shi G, Zhao L, Li T, Zhang Z, Liu R, Hu Y, Liu H, Yu J, Li G. TGF-β1-SOX9 axis-inducible COL10A1 promotes invasion and metastasis in gastric cancer via epithelial-to-mesenchymal transition. Cell Death Dis. 2018; 9:849. https://doi.org/10.1038/s41419-018-0877-2. [PubMed]

86. Zhang M, Chen H, Wang M, Bai F, Wu K. Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer. Biosci Rep. 2020; 40:BSR20193286. https://doi.org/10.1042/BSR20193286. [PubMed]

87. Liu TT, Liu XS, Zhang M, Liu XN, Zhu FX, Zhu FM, Ouyang SW, Li SB, Song CL, Sun HM, Lu S, Zhang Y, Lin J, et al. Cartilage oligomeric matrix protein is a prognostic factor and biomarker of colon cancer and promotes cell proliferation by activating the Akt pathway. J Cancer Res Clin Oncol. 2018; 144:1049–63. https://doi.org/10.1007/s00432-018-2626-4. [PubMed]

88. Englund E, Bartoschek M, Reitsma B, Jacobsson L, Escudero-Esparza A, Orimo A, Leandersson K, Hagerling C, Aspberg A, Storm P, Okroj M, Mulder H, Jirström K, et al. Cartilage oligomeric matrix protein contributes to the development and metastasis of breast cancer. Oncogene. 2016; 35:5585–96. https://doi.org/10.1038/onc.2016.98. [PubMed]

89. Peluffo G, Subedee A, Harper NW, Kingston N, Jovanović B, Flores F, Stevens LE, Beca F, Trinh A, Chilamakuri CSR, Papachristou EK, Murphy K, Su Y, et al. *EN1* Is a Transcriptional Dependency in Triple-Negative Breast

Cancer Associated with Brain Metastasis. Cancer Res. 2019; 79:4173–83. https://doi.org/10.1158/0008-5472. CAN-18-3264. [PubMed]

90. Bell D, Bell A, Roberts D, Weber RS, El-Naggar AK. Developmental transcription factor EN1--a novel biomarker in human salivary gland adenoid cystic carcinoma. Cancer. 2012; 118:1288–92. https://doi.org/10.1002/cncr.26412. [PubMed]

91. Li Y, Liu H, Lai C, Su Z, Heng B, Gao S. Repression of engrailed 2 inhibits the proliferation and invasion of human bladder cancer in vitro and in vivo. Oncol Rep. 2015; 33:2319–30. https://doi.org/10.3892/or.2015.3858. [PubMed]

92. Lin X, Liu X, Gong C. Expression of engrailed homeobox 2 regulates the proliferation, migration and invasion of non-small cell lung cancer cells. Oncol Lett. 2018; 16:536–42. https://doi.org/10.3892/ol.2018.8693. [PubMed]

93. Zhou Y, Ji Z, Yan W, Zhou Z, Li H. The biological functions and mechanism of miR-212 in prostate cancer proliferation, migration and invasion via targeting Engrailed-2. Oncol Rep. 2017; 38:1411–19. https://doi.org/10.3892/or.2017.5805. [PubMed]

94. Luo F, Wang YZ, Lin D, Li J, Yang K. Exonuclease 1 expression is associated with clinical progression, metastasis, and survival prognosis of prostate cancer. J Cell Biochem. 2019. https://doi.org/10.1002/jcb.28415. [Epub ahead of print.] [PubMed]

95. Dai Y, Tang Z, Yang Z, Zhang L, Deng Q, Zhang X, Yu Y, Liu X, Zhu J. EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients. Cell Cycle. 2018; 17:2386–97. https://doi.org/10.1080/15384101.2018.1534511. [PubMed]

96. Yan X, Yu Y, Li L, Chen N, Song W, He H, Dong J, Liu X, Cui J. Friend leukemia virus integration 1 is a predictor of poor prognosis of breast cancer and promotes metastasis and cancer stem cell properties of breast cancer cells. Cancer Med. 2018; 7:3548–60. https://doi.org/10.1002/cam4.1589. [PubMed]

97. Chen N, Zhao G, Yan X, Lv Z, Yin H, Zhang S, Song W, Li X, Li L, Du Z, Jia L, Zhou L, Li W, et al. A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. Genome Biol. 2018; 19:218. https://doi.org/10.1186/s13059-018-1594-y. [PubMed]

98. Sakurai T, Kondoh N, Arai M, Hamada J, Yamada T, Kihara-Negishi F, Izawa T, Ohno H, Yamamoto M, Oikawa T. Functional roles of Fli-1, a member of the Ets family of transcription factors, in human breast malignancy. Cancer Sci. 2007; 98:1775–84. https://doi.org/10.1111/j.1349-7006.2007.00598.x. [PubMed]

99. Song G, Zhu X, Xuan Z, Zhao L, Dong H, Chen J, Li Z, Song W, Jin C, Zhou M, Xie H, Zheng S, Song P. Hypermethylation of GNA14 and its tumor-suppressive role in hepatitis B virus-related hepatocellular carcinoma. Theranostics. 2021; 11:2318–33. https://doi.org/10.7150/thno.48739. [PubMed]

100. Kuzmanov A, Hopfer U, Marti P, Meyer-Schaller N, Yilmaz M, Christofori G. LIM-homeobox gene 2 promotes tumor growth and metastasis by inducing autocrine and paracrine PDGF-B signaling. Mol Oncol. 2014; 8:401–16. https://doi.org/10.1016/j.molonc.2013.12.009. [PubMed]

101. Ni S, Hu J, Duan Y, Shi S, Li R, Wu H, Qu Y, Li Y. Down expression of LRP1B promotes cell migration via RhoA/Cdc42 pathway and actin cytoskeleton remodeling in renal cell cancer. Cancer Sci. 2013; 104:817–25. https://doi.org/10.1111/cas.12157. [PubMed]

102. Xu SF, Guo Y, Zhang X, Zhu XD, Fan N, Zhang ZL, Ren GB, Rao W, Zang YJ. Somatic Mutation Profiling of Intrahepatic Cholangiocarcinoma: Comparison between Primary and Metastasis Tumor Tissues. J Oncol. 2020; 2020:5675020. https://doi.org/10.1155/2020/5675020. [PubMed]

103. Wang Z, Sun P, Gao C, Chen J, Li J, Chen Z, Xu M, Shao J, Zhang Y, Xie J. Down-regulation of LRP1B in colon cancer promoted the growth and migration of cancer cells. Exp Cell Res. 2017; 357:1–8. https://doi.org/10.1016/j.yexcr.2017.04.010. [PubMed]

104. Liu L, Ren M, Han S, Sun L, Zhu L. Expression level and clinical significance of low-density lipoprotein receptor-related protein 1B gene in cervical squamous cell carcinoma. Int J Clin Exp Pathol. 2018; 11:1701–6. [PubMed]

105. Xiao Q, Gan Y, Li Y, Fan L, Liu J, Lu P, Liu J, Chen A, Shu G, Yin G. MEF2A transcriptionally upregulates the expression of ZEB2 and CTNNB1 in colorectal cancer to promote tumor progression. Oncogene. 2021; 40:3364–77. https://doi.org/10.1038/s41388-021-01774-w. [PubMed]

106. Tran OT, Tadesse S, Chu C, Kidane D. Overexpression of *NEIL3* associated with altered genome and poor survival in selected types of human cancer. Tumour Biol. 2020; 42:1010428320918404. https://doi.org/10.1177/1010428320918404. [PubMed]

107. Rivera-Rivera Y, Marina M, Jusino S, Lee M, Velázquez JV, Chardón-Colón C, Vargas G, Padmanabhan J, Chellappan SP, Saavedra HI. The Nek2 centrosome-mitotic kinase contributes to the mesenchymal state, cell invasion, and migration of triple-negative breast cancer cells. Sci Rep. 2021; 11:9016. https://doi.org/10.1038/s41598-021-88512-0. [PubMed]

108. Zhang Y, Wang W, Wang Y, Huang X, Zhang Z, Chen B, Xie W, Li S, Shen S, Peng B. NEK2 promotes hepatocellular carcinoma migration and invasion through modulation of the epithelial-mesenchymal transition. Oncol Rep. 2018; 39:1023–33. https://doi.org/10.3892/or.2018.6224. [PubMed]

109. Ren Y, Wang Y, Hao S, Yang Y, Xiong W, Qiu L, Tao J, Tang A. NFE2L3 promotes malignant behavior and EMT of human hepatocellular carcinoma (HepG2) cells via Wnt/β-catenin pathway. J Cancer. 2020; 11:6939–49. https://doi.org/10.7150/jca.48100. [PubMed]

110. Sun J, Zheng Z, Chen Q, Pan Y, Lu H, Zhang H, Yu Y, Dai Y. NRF3 suppresses breast cancer cell metastasis and cell proliferation and is a favorable predictor of survival in breast cancer. Onco Targets Ther. 2019; 12:3019–30. https://doi.org/10.2147/OTT.S197409. [PubMed]

111. Xie F, Xiao X, Tao D, Huang C, Wang L, Liu F, Zhang H, Niu H, Jiang G. circNR3C1 Suppresses Bladder Cancer Progression through Acting as an Endogenous Blocker of BRD4/C-myc Complex. Mol Ther Nucleic Acids. 2020; 22:510–19. https://doi.org/10.1016/j.omtn.2020.09.016. [PubMed]

112. Zhang L, Jiang H, Zhang Y, Wang C, Xia X, Sun Y. GR silencing impedes the progression of castration-resistant prostate cancer through the JAG1/NOTCH2 pathway via up-regulation of microRNA-143-3p. Cancer Biomark. 2020; 28:483–97. https://doi.org/10.3233/CBM-191271. [PubMed]

113. Pan D, Kocherginsky M, Conzen SD. Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. Cancer Res. 2011; 71:6360–70. https://doi.org/10.1158/0008-5472.CAN-11-0362. [PubMed]

114. Yu G, Lee YC, Cheng CJ, Wu CF, Song JH, Gallick GE, Yu-Lee LY, Kuang J, Lin SH. RSK promotes prostate cancer progression in bone through ING3, CKAP2, and PTK6-mediated cell survival. Mol Cancer Res. 2015; 13:348–57. https://doi.org/10.1158/1541-7786.MCR-14-0384-T. [PubMed]

115. Liu C, Yao F, Mao X, Li W, Chen H. Effect of SALL4 on the Proliferation, Invasion and Apoptosis of Breast Cancer Cells. Technol Cancer Res Treat. 2020; 19:1533033820980074. https://doi.org/10.1177/1533033820980074. [PubMed]

116. Li J, Zhang Y, Tao X, You Q, Tao Z, Zhang Y, He Z, Ou J. Knockdown of *SALL4* inhibits the proliferation, migration, and invasion of human lung cancer cells *in vivo* and *in vitro*. Ann Transl Med. 2020; 8:1678. https://doi.org/10.21037/atm-20-7939. [PubMed]

117. Petersen M, Pardali E, van der Horst G, Cheung H, van den Hoogen C, van der Pluijm G, Ten Dijke P. Smad2 and Smad3 have opposing roles in breast cancer bone metastasis by differentially affecting tumor angiogenesis. Oncogene. 2010; 29:1351–61. https://doi.org/10.1038/onc.2009.426. [PubMed]

118. Tang PM, Zhou S, Meng XM, Wang QM, Li CJ, Lian GY, Huang XR, Tang YJ, Guan XY, Yan BP, To KF, Lan HY. Smad3 promotes cancer progression by inhibiting E4BP4-mediated NK cell development. Nat Commun. 2017; 8:14677. https://doi.org/10.1038/ncomms14677. [PubMed]

119. Qian Z, Zhang Q, Hu Y, Zhang T, Li J, Liu Z, Zheng H, Gao Y, Jia W, Hu A, Li B, Hao J. Investigating the mechanism by which SMAD3 induces PAX6 transcription to promote the development of non-small cell lung cancer. Respir Res. 2018; 19:262. https://doi.org/10.1186/s12931-018-0948-z. [PubMed]

120. Jordan NV, Prat A, Abell AN, Zawistowski JS, Sciaky N, Karginova OA, Zhou B, Golitz BT, Perou CM, Johnson GL. SWI/SNF chromatin-remodeling factor Smarcd3/Baf60c controls epithelial-mesenchymal transition by inducing Wnt5a signaling. Mol Cell Biol. 2013; 33:3011–25. https://doi.org/10.1128/MCB.01443-12. [PubMed]

121. Jiang M, Wang H, Chen H, Han Y. SMARCD3 is a potential prognostic marker and therapeutic target in CAFs. Aging (Albany NY). 2020; 12:20835–61. https://doi.org/10.18632/aging.104102. [PubMed]

122. Fenizia C, Bottino C, Corbetta S, Fittipaldi R, Floris P, Gaudenzi G, Carra S, Cotelli F, Vitale G, Caretti G. SMYD3 promotes the epithelial-mesenchymal transition in breast cancer. Nucleic Acids Res. 2019; 47:1278–93. https://doi.org/10.1093/nar/gky1221. [PubMed]

123. Zhang L, Jin Y, Yang H, Li Y, Wang C, Shi Y, Wang Y. SMYD3 promotes epithelial ovarian cancer metastasis by downregulating p53 protein stability and promoting p53 ubiquitination. Carcinogenesis. 2019; 40:1492–503. https://doi.org/10.1093/carcin/bgz078. [PubMed]

124. Zhu CL, Huang Q. Overexpression of the SMYD3 Promotes Proliferation, Migration, and Invasion of Pancreatic Cancer. Dig Dis Sci. 2020; 65:489–99. https://doi.org/10.1007/s10620-019-05797-y. [PubMed]

125. Shen C, Yin Y, Chen H, Wang R, Yin X, Cai Z, Zhang B, Chen Z, Zhou Z. Secreted protein acidic and rich in cysteine-like 1 suppresses metastasis in gastric stromal tumors. BMC Gastroenterol. 2018; 18:105. https://doi.org/10.1186/s12876-018-0833-8. [PubMed]

126. Xiang Y, Qiu Q, Jiang M, Jin R, Lehmann BD, Strand DW, Jovanovic B, DeGraff DJ, Zheng Y, Yousif DA, Simmons CQ, Case TC, Yi J, et al. SPARCL1 suppresses metastasis in prostate cancer. Mol Oncol. 2013; 7:1019–30. https://doi.org/10.1016/j.molonc.2013.07.008. [PubMed]

127. Zhao SJ, Jiang YQ, Xu NW, Li Q, Zhang Q, Wang SY, Li J, Wang YH, Zhang YL, Jiang SH, Wang YJ, Huang YJ, Zhang XX, et al. SPARCL1 suppresses osteosarcoma metastasis and recruits macrophages by activation of canonical WNT/β-catenin signaling through stabilization of the WNT-receptor complex. Oncogene. 2018; 37:1049–61. https://doi.org/10.1038/onc.2017.403. [PubMed]

128. Zhang S, Chang X, Ma J, Chen J, Zhi Y, Li Z, Dai D. Downregulation of STARD8 in gastric cancer and its involvement in gastric cancer progression. Onco Targets Ther. 2018; 11:2955–61. https://doi.org/10.2147/OTT.S154524. [PubMed]

129. Zhan Y, Liang X, Li L, Wang B, Ding F, Li Y, Wang X, Zhan Q, Liu Z. MicroRNA-548j functions as a metastasis promoter in human breast cancer by targeting Tensin1. Mol Oncol. 2016; 10:838–49. https://doi.org/10.1016/j.molonc.2016.02.002. [PubMed]

130. Duan J, Wang L, Shang L, Yang S, Wu H, Huang Y, Miao Y. miR-152/TNS1 axis inhibits non-small cell lung cancer progression through Akt/mTOR/RhoA pathway. Biosci Rep. 2021; 41:BSR20201539. https://doi.org/10.1042/BSR20201539. [PubMed]

131. Xie W, Du Z, Chen Y, Liu N, Zhong Z, Shen Y, Tang L. Identification of Metastasis-Associated Genes in Triple-Negative Breast Cancer Using Weighted Gene Co-expression Network Analysis. Evol Bioinform Online. 2020; 16:1176934320954868. https://doi.org/10.1177/1176934320954868. [PubMed]

132. Wang J, Chen W, Wei W, Lou J. Oncogene TUBA1C promotes migration and proliferation in hepatocellular carcinoma and predicts a poor prognosis. Oncotarget. 2017; 8:96215–24. https://doi.org/10.18632/oncotarget.21894. [PubMed]

133. Albahde MAH, Zhang P, Zhang Q, Li G, Wang W. Upregulated Expression of TUBA1C Predicts Poor Prognosis and Promotes Oncogenesis in Pancreatic Ductal Adenocarcinoma via Regulating the Cell Cycle. Front Oncol. 2020; 10:49. https://doi.org/10.3389/fonc.2020.00049. [PubMed]