Literature Lab™ analysis of pathway associations relevant to prognostic gene signatures for breast, bone and lung cancer metastasis

Damon Anderson, PhD, Applications Scientist and Paul Martinez, President and CEO, Acumenta Biotech, Westminster, MA.
Co-responding author: Damon Anderson, PhD, danderson@acumenta.com, 412-901-7785

The landmark breast cancer metastasis study by van't Veer, et al. 2002. *Nature* (415) 530-536 identified a 70-gene signature derived from microarray analysis of 117 patients that was strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumor cells in local lymph nodes at diagnosis (lymph node negative). We applied Literature Lab™ analysis on this gene signature to identify significantly associated pathways, thus offering a modern perspective based on the current PubMed content and Literature Lab™ analytics. Published gene signatures derived from microarray data from lung and bone metastatic tumors were then analyzed using Literature Lab™ and significant pathways were identified and compared. The Id signaling pathway showed a strong association with lung while the CXCR4 pathway was strongly associated with both lung and bone, results well supported in the literature. Importantly, Literature Lab™ identified the IL1 pathway, which was exclusively associated with lung, and the CCR5 pathway. which was significantly associated with lung and bone. The unexpected associations of these immune factor pathways and their roles in the lung and bone metastatic microenvironments indicate potential novel areas for future investigations. Moreover, these results and others showcase the powerful capabilities of Literature Lab™ as a genetic analysis platform that can offer validation and uncover meritorious functional associations missed by other genetic analysis tools.

Diseases-MeSH/Pathways Subset:a Analysis
Based on PubMed abstracts from 01/01/1990 through 02/28/2015

Click on column headers or buttons to sort columns or rows. Click again to reverse sort order. Enter the first few letters of a term to find a cell.
Cells contain the LPF for the terms. Click on a cell to view the abstracts for the terms. Right Click for Annotations.

Visualizer | Close this Subset | Row Term: | Column Term: | Find

Term	Abstract Count	View Genes	Breast Cancer	ERBB2	ErbB Signaling	Estrogen Signaling	Antigen	E-cadherin Signaling	BRCA1	mRNA Surveillance	EGF	RNA Polymerase	Basal Cell Carcinoma
Abstract Count			39669	8930	9937	45592	310335	7566	6634	4555	37456	92157	3307
View Genes													
Carcinoma, Ductal, Breast	12261		-2.4598	-3.0156	-3.0919	-3.3238	-4.0610	-4.2173	-4.3255	-4.4215	-4.4965	-4.5620	-4.6252
Carcinoma, Non-Small-Cell Lun	32779		-4.7004	-3.8775	-3.6533	-5.3683	-4.0492	-4.2147	-4.9749	-3.7391	-2.8713	-4.0073	-4.2074
Disease Models, Animal	275869		-3.4349	-3.9188	-3.9293	-3.7172	-2.3996	-4.3978	-4.9954	-4.5236	-3.8653	-3.1677	-4.8540
Bronchial Neoplasms	39034		-4.7393	-3.9359	-3.7100	-5.4122	-4.0738	-4.2559	-5.0328	-3.7890	-2.9324	-4.0099	-4.1928
Inflammatory Breast Neoplasn	215		-3.5684	-4.2005	-4.4212	-4.9085	-5.2222	-4.6550	-	-5.9909	-5.2158	-6.3427	-
Breast Neoplasms, Male	2304		-4.0240	-4.8524	-4.8988	-4.9584	-5.8716	-7.2413	-4.4238	-	-6.0275	-7.3728	-6.8819
Small Cell Lung Carcinoma	6083		-5.0022	-5.1775	-4.7986	-5.7582	-4.4801	-4.9781	-6.4018	-4.7192	-4.0470	-4.5009	-4.2406
Hereditary Breast and Ovarian	144		-4.5289	-6.1092	-6.1556	-5.6131	-6.6960	-	-3.4226	-	-6.7319	-	-
Pulmonary Sclerosing Hemang	146		-	-6.1152	-6.1616	-6.8232	-6.4521	-	-	-	-6.7379	-7.1289	-
Palatal Neoplasms	1180		-	-6.4207	-6.4671	-	-7.1658	-	-	-	-7.6454	-8.0364	-6.5913
Nose Neoplasms	8588		-	-7.2827	-7.3291	-	-5.9450	-6.6086	-7.7557	-6.6381	-6.8172	-6.4902	-6.0554
Skull Base Neoplasms	2526		-	-7.3533	-7.3997	-	-7.4963	-7.2813	-	-	-7.9760	-8.3670	-
Solitary Pulmonary Nodule	2690		-	-7.3806	-7.4270	-	-7.7175	-	-	-	-8.0033	-8.3943	-
Mandibular Neoplasms	4379		-8.2398	-7.5922	-7.6386	-	-7.9291	-	-	-	-	-8.0038	-5.7629
Spinal Neoplasms	7362		-6.3071	-	-	-8.5259	-7.4503	-	-	-	-7.8385	-8.8315	-7.3864

The Literature Lab™ basic viewer is shown above with co-occurrences tabulated between diseases (breast, lung, bone neoplasms; y-axis) and pathways (x-axis). The pathways are ranked by the strength of their associations in the literature with Carcinoma, Ductal, Breast from highest strength association according to LPF (log of product of frequency). Blue dots hyperlink to the genes associated in the literature with the disease or pathway. All other data is hyperlinked directly to the PubMed literature.

Nature **415**, 530-536 (31 January 2002) doi:10.1038/415530a

The log of product of frequency (LPF) is a quantitative measure of the strength of association based on the fractional co-occurrence between a term/gene and a term. The Literature Lab™ basic LPF is defined as Log(X/T1 x X/T2), where X/T1 is the percentage of abstracts that mention the first term and X/T2 is the percentage of abstracts that mention the second term. The Literature Lab™ PLUS LPF is defined as Log(X/G x X/T), where X/G is the percentage of gene abstracts that mention the term and X/T is the percentage of term abstracts that mention the given gene. The LPF is not sensitive to abstract volume and the closer to zero the stronger the association.

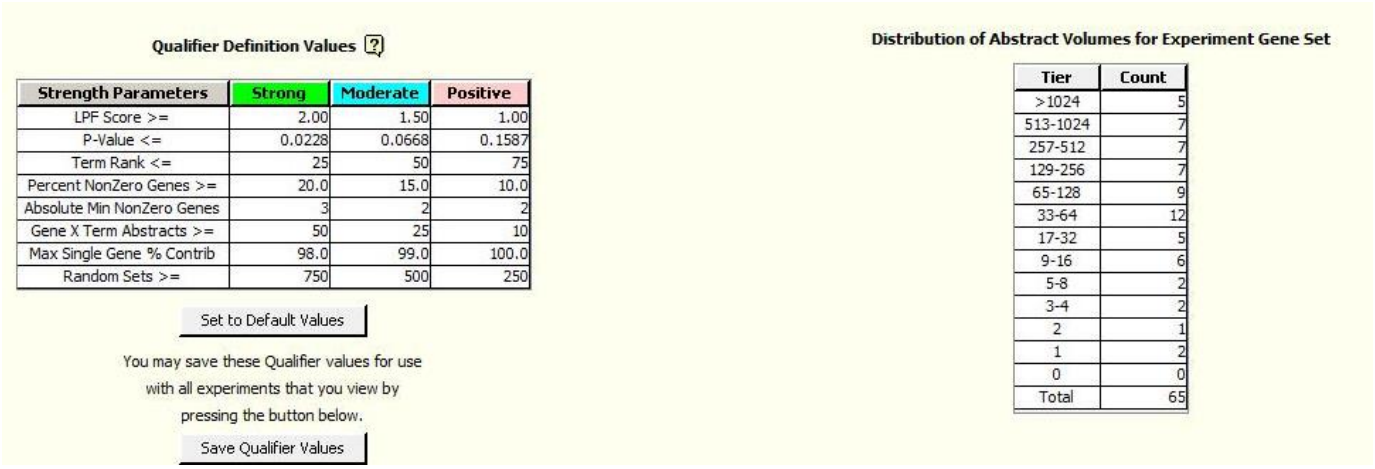
Literature Lab™ PLUS was employed to analyze a 70-gene signature derived by DNA microarray analysis on primary breast tumors of 117 young patients. This gene expression signature is strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumor cells in local lymph nodes at diagnosis (lymph node negative).

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer, et al.*Nature***415**, 530-536 (31 January 2002)

Breast cancer patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70–80% of patients receiving this treatment would have survived without it. None of the signatures of breast cancer gene expression reported to date allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of *BRCA1* carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and angiogenesis. This gene expression profile will outperform all currently used clinical parameters in predicting disease outcome. Our findings provide a strategy to select patients who would benefit from adjuvant therapy.

Statistically significant associations are scored according to qualifiers as shown below:



Qualifier Definition Values

Strength Parameters	Strong	Moderate	Positive
LPF Score >=	2.00	1.50	1.00
P-Value <=	0.0228	0.0668	0.1587
Term Rank <=	25	50	75
Percent NonZero Genes >=	20.0	15.0	10.0
Absolute Min NonZero Genes	3	2	2
Gene X Term Abstracts >=	50	25	10
Max Single Gene % Contrib	98.0	99.0	100.0
Random Sets >=	750	500	250

Set to Default Values

You may save these Qualifier values for use with all experiments that you view by pressing the button below.

Save Qualifier Values

Distribution of Abstract Volumes for Experiment Gene Set

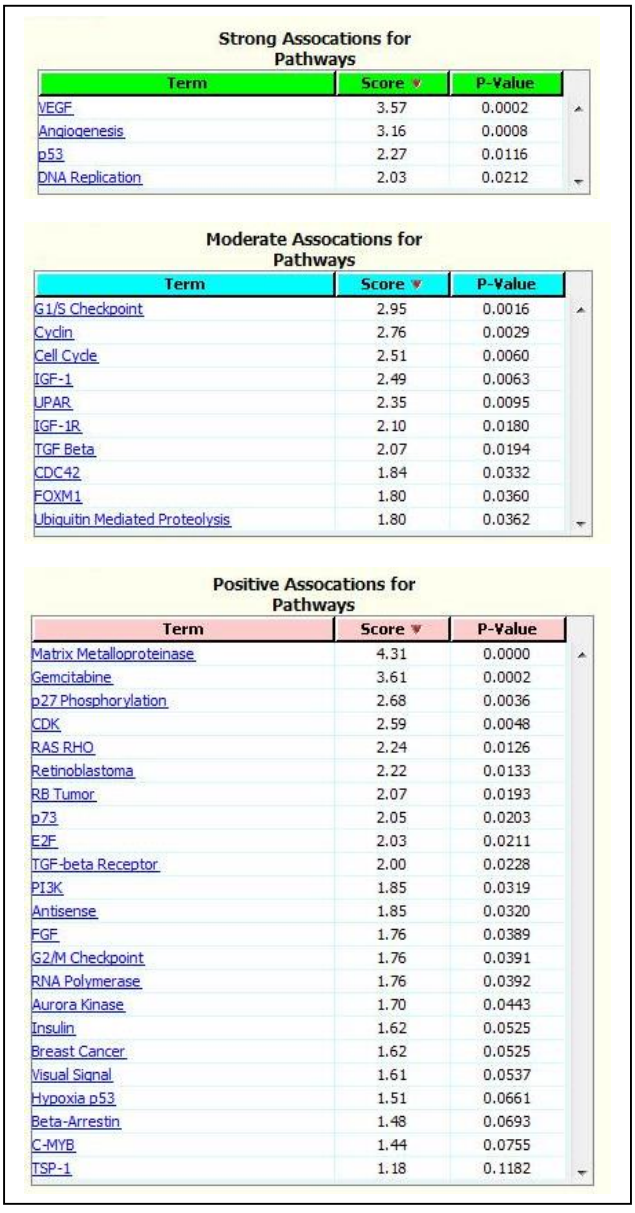
Tier	Count
>1024	5
513-1024	7
257-512	7
129-256	7
65-128	9
33-64	12
17-32	5
9-16	6
5-8	2
3-4	2
2	1
1	2
0	0
Total	65

The P-value and other measures of statistical significance are used to derive the strength of association. The heuristics involved reward associations for representation of more genes in a set and downgrade associations driven by sparse data. The qualifier criteria can be adjusted according to experimental goals but typically default values are used for straight forward analysis. Importantly, Literature Lab™ is the only platform that treats each gene list (and gene) as unique and finds significant associations in literature based on LPF and statistical qualifiers. Other methods search for associations between an experimentally derived gene list and pathway or domain information within curated databases, without regard for the uniqueness of each gene set.

The 'Highlights' tab displays the MeSH term domains that were used in the analysis (left column of left panel) and the number and strength of associations that were identified based on the qualifiers (right side of left panel).The right panel displays the identity of the associations, the score, and the p-value. Strong (green), Moderate (blue), and Positive (pink) pathway associations are shown:

Domains & Association Counts

	Strong	Moderate	Positive	Visualizer
Pathways	4	10	23	
Diseases-MeSH	1	9	45	
PathologicalCond-MeSH	5	5	9	
PharmacoAction-MeSH	0	1	9	
PharmacoSubst-MeSH	2	3	23	
ChemicalActions-MeSH	2	4	9	
ChemicalsDrugs-MeSH	15	46	179	
Anatomy-MeSH	2	6	24	
Physiology-MeSH	3	12	16	
CellPhysiology-MeSH	8	7	7	
CellTypes-MeSH	0	5	8	
CellStructures-MeSH	7	7	4	
CellLines	2	5	10	
TissueTypes-MeSH	1	5	11	
Metabolism-MeSH	2	0	3	
Biogenetics-MeSH	4	16	7	
OtherBiology-MeSH	5	17	17	
Organisms-MeSH	0	1	11	
Substances-MeSH	3	8	34	
Psychology-MeSH	0	1	0	



Strong Assocations for Pathways

Term	Score	P-Value
VEGF	3.57	0.0002
Angiogenesis	3.16	0.0008
p53	2.27	0.0116
DNA Replication	2.03	0.0212

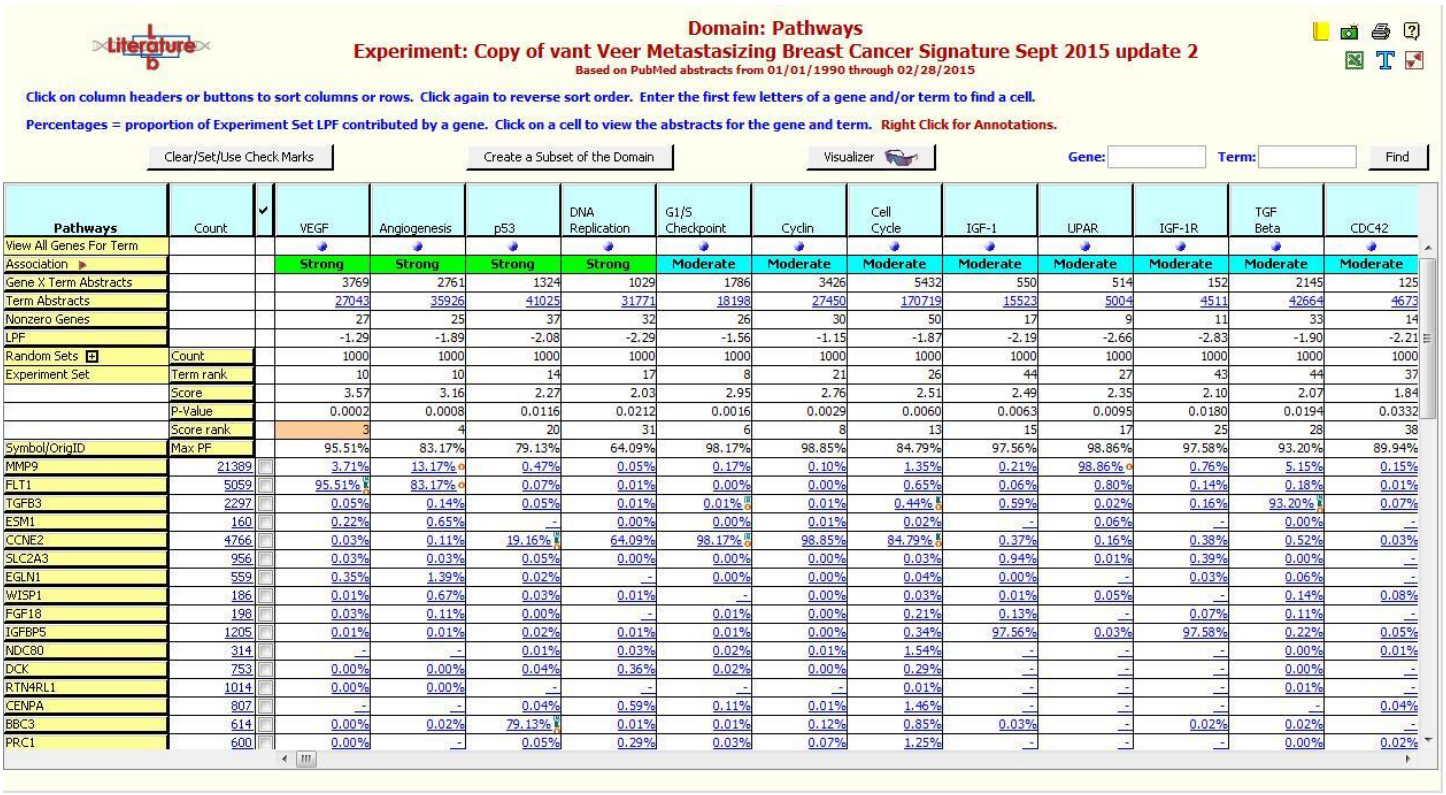
Moderate Assocations for Pathways

Term	Score	P-Value
G1/S Checkpoint	2.95	0.0016
Cyclin	2.76	0.0029
Cell Cycle	2.51	0.0060
IGF-1	2.49	0.0063
UPAR	2.35	0.0095
IGF-1R	2.10	0.0180
TGF Beta	2.07	0.0194
CDC42	1.84	0.0332
FOXM1	1.80	0.0360
Ubiquitin Mediated Proteolysis	1.80	0.0362

Positive Assocations for Pathways

Term	Score	P-Value
Matrix Metalloproteinase	4.31	0.0000
Gemcitabine	3.61	0.0002
p27 Phosphorylation	2.68	0.0036
CDK	2.59	0.0048
RAS RHO	2.24	0.0126
Retinoblastoma	2.22	0.0133
RB Tumor	2.07	0.0193
p73	2.05	0.0203
E2F	2.03	0.0211
TGF-beta Receptor	2.00	0.0228
PI3K	1.85	0.0319
Antisense	1.85	0.0320
FGF	1.76	0.0389
G2/M Checkpoint	1.76	0.0391
RNA Polymerase	1.76	0.0392
Aurora Kinase	1.70	0.0443
Insulin	1.62	0.0525
Breast Cancer	1.62	0.0525
Visual Signal	1.61	0.0537
Hypoxia p53	1.51	0.0661
Beta-Arrestin	1.48	0.0693
C-MYB	1.44	0.0755
TSP-1	1.18	0.1182

Associations include a number of pathways well documented in cancer growth and development including: Cell Cycle/Cyclins, DNA replication/G1/S Phase, etc. Other pathway associations have well documented ties to metastasis including: Angiogenesis/VEGF, Matrix Metalloproteinase, Beta-Arrestin, etc. Pathway associations representing interesting areas of investigation in the literature are also uncovered. For instance, the Insulin pathway has been the target of research and therapeutic discovery over the past few years. However, the complex cross talk between factors and receptors has made it elusive in the treatment of breast carcinoma.



Domain: Pathways
Experiment: Copy of vant Veer Metastasizing Breast Cancer Signature Sept 2015 update 2
Based on PubMed abstracts from 01/01/1990 through 02/28/2015

Click on column headers or buttons to sort columns or rows. Click again to reverse sort order. Enter the first few letters of a gene and/or term to find a cell.

Percentages = proportion of Experiment Set LPF contributed by a gene. Click on a cell to view the abstracts for the gene and term. Right Click for Annotations.

Clear/Set/Use Check Marks | Create a Subset of the Domain | Visualizer | Gene: | Term: | Find

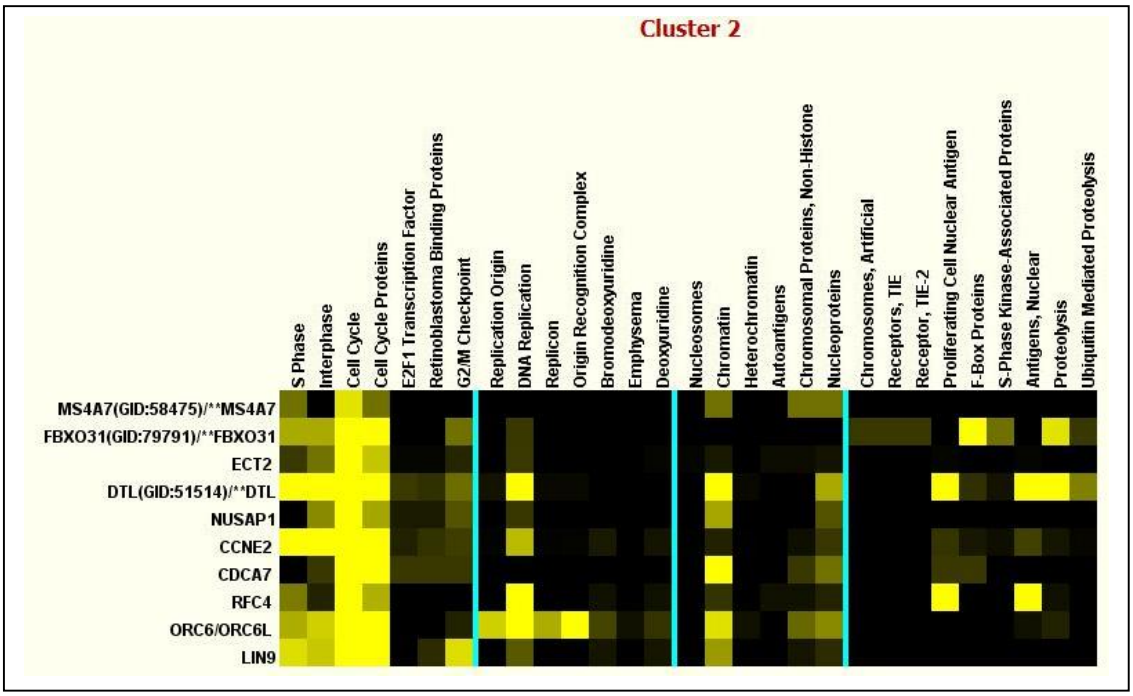
Pathways	Count	VEGF	Angiogenesis	p53	DNA Replication	G1/S Checkpoint	Cyclin	Cell Cycle	IGF-1	UPAR	IGF-1R	TGF Beta	CDC42
View All Genes For Term													
Association		Strong	Strong	Strong	Strong	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Gene X Term Abstracts		3769	2761	1324	1029	1786	3426	5432	550	514	152	2145	125
Term Abstracts		27043	35926	41025	31771	18198	27450	170719	15523	5004	4511	42664	4673
Nonzero Genes		27	25	37	32	26	30	50	17	9	11	33	14
LPF		-1.29	-1.89	-2.08	-2.29	-1.56	-1.15	-1.87	-2.19	-2.66	-2.83	-1.90	-2.21
Random Sets	Count	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Experiment Set	Term rank	10	10	14	17	8	21	26	44	27	43	44	37
	Score	3.57	3.16	2.27	2.03	2.95	2.76	2.51	2.49	2.35	2.10	2.07	1.84
	P-Value	0.0002	0.0008	0.0116	0.0212	0.0016	0.0029	0.0060	0.0063	0.0095	0.0180	0.0194	0.0332
	Score rank	3	4	20	31	6	8	13	15	17	25	28	38
Symbol/OrigID	Max PF	95.51%	83.17%	79.13%	64.09%	98.17%	98.85%	84.79%	97.56%	98.86%	97.58%	93.20%	89.94%
MMP9	21389	3.71%	13.17%	0.47%	0.05%	0.17%	0.10%	1.35%	0.21%	98.86%	0.76%	5.15%	0.15%
FLT1	5059	95.51%	83.17%	0.07%	0.01%	0.00%	0.00%	0.65%	0.06%	0.80%	0.14%	0.18%	0.01%
TGFB3	2297	0.05%	0.14%	0.05%	0.01%	0.01%	0.01%	0.44%	0.59%	0.02%	0.16%	93.20%	0.07%
ESM1	160	0.22%	0.65%	-	0.00%	0.00%	0.01%	0.02%	-	0.06%	-	0.00%	-
CCNE2	4766	0.03%	0.11%	19.16%	64.09%	98.17%	98.85%	84.79%	0.37%	0.16%	0.38%	0.52%	0.03%
SLC2A3	956	0.03%	0.03%	0.05%	0.00%	0.00%	0.00%	0.03%	0.94%	0.01%	0.39%	0.00%	-
EGLN1	559	0.35%	1.39%	0.02%	-	0.00%	0.00%	0.04%	0.00%	-	0.03%	0.06%	-
WISP1	186	0.01%	0.67%	0.03%	0.01%	-	0.00%	0.03%	0.01%	0.05%	-	0.14%	0.08%
FGF18	198	0.03%	0.11%	0.00%	-	0.01%	0.00%	0.21%	0.13%	-	0.07%	0.11%	-
IGFBP5	1205	0.01%	0.01%	0.02%	0.01%	0.01%	0.00%	0.34%	97.56%	0.03%	97.58%	0.22%	0.05%
NDC80	314	-	-	0.01%	0.03%	0.02%	0.01%	1.54%	-	-	-	0.00%	0.01%
DCK	753	0.00%	0.00%	0.04%	0.36%	0.02%	0.00%	0.29%	-	-	-	0.00%	-
RTN4RL1	1014	0.00%	0.00%	-	-	-	-	0.01%	-	-	-	0.01%	-
CENPA	807	-	-	0.04%	0.59%	0.11%	0.01%	1.46%	-	-	-	-	0.04%
BBC3	614	0.00%	0.02%	79.13%	0.01%	0.01%	0.12%	0.85%	0.03%	-	0.02%	0.02%	-
PRC1	600	0.00%	-	0.05%	0.29%	0.03%	0.07%	1.25%	-	-	-	0.00%	0.02%

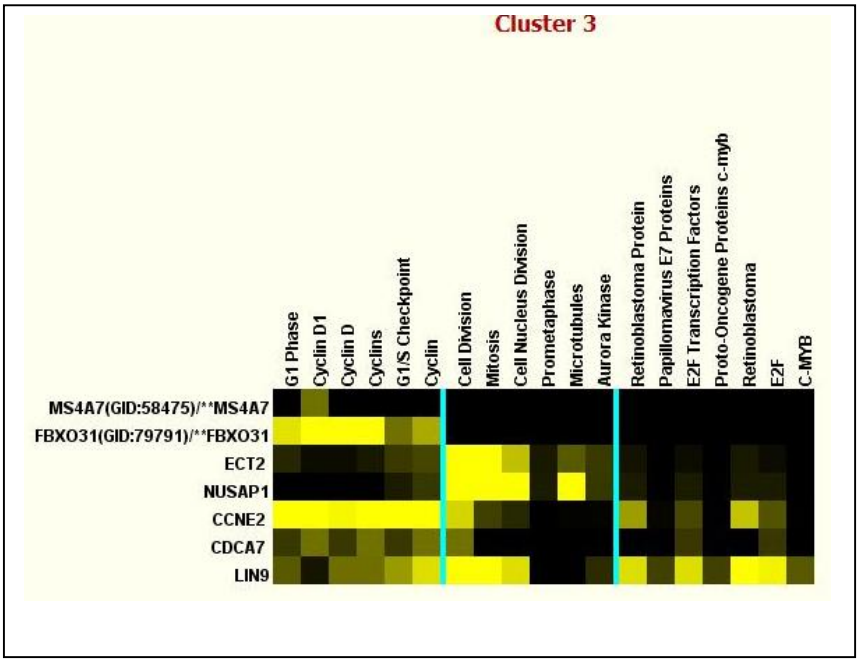
Shown above are the pathway associations in the 'Details' tab of the Literature Lab™ PLUS viewer. Down the y-axis are the qualifiers used to establish the strength of associations including: the number of gene abstracts, the number of term abstracts, the number of non-zero genes, the log of product of frequency, the number of random sets, the term rank, the score, the p-value, etc. Also shown is the list of genes and their relative contribution to the LPF with hyperlinks directly to the literature. Across the x-axis are the associations, which can be ordered by any of the rows on the left column. (Note: shown is a truncated list of genes and pathway associations).

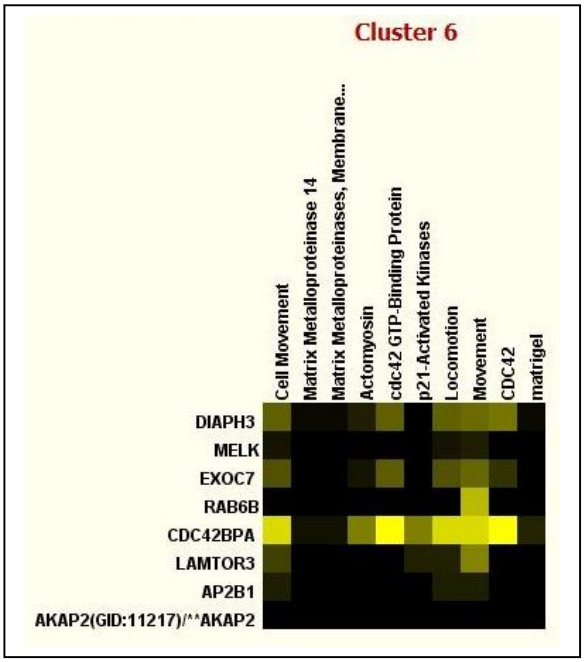
The paper discussed functional annotation for the genes in the 70-gene signature, which provided insight into the underlying biological mechanism leading to rapid metastases. Genes involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction are significantly upregulated in the poor prognosis signature (for example cyclin E2, *MCM6*, metalloproteinases *MMP9* and *MP1*, *RAB6B*, *PK428*, *ESM1*, and the VEGF receptor *FLT1*.

Literature Lab™ PLUS found a number of associations between the signature and these gene pathways (*e.g.* VEGF, Metalloproteinases, etc.). Interestingly, some pathways are not significantly associated as indicated by the qualifier scores (*e.g.* MCM6, RAB6B). In the paper, the authors do not detail a comprehensive annotation of all of the genes in the signature. A powerful aspect of Literature Lab™ PLUS is that it treats each gene in the signature as unique and derives a statistically significant association of the entire signature based on scoring parameters. Another important aspect to note is the difference in dates of this published work (1/2002) and the current build of the Literature Lab™ database (3/2015). Undoubtedly, there has been a progression in published papers following the role of certain genes in specific pathways, thereby leading to an increase in the strength of association during this time. Likewise, other gene/pathway associations have not been as fruitful and therefore the strength has likely diminished.

Clustering analysis on the significant terms and the genes in the70-gene signature was then explored using the Literature Lab™ PLUS 'Term/Gene Cluster' tab. Eight clusters of genes with related function were identified and three of these and their associated heat maps are shown below.







Clusters 2 and 3 are related to cell cycle associated pathways including G1 phase, cyclins, mitosis, microtubules, as well as signaling pathways well documented with tumor proliferation including, aurora kinase, retinoblastoma, E2F, and C-MYB, etc. Cluster 6 is related to pathways associated with tumor metastasis including, matrix metalloproteases, CDC42, etc. The heat maps are an indicator of strength of association of each clustered gene with a pathway, and each box is hyperlinked directly to the PubMed literature.

Nature **415**, 530-536 (31 January 2002) doi:10.1038/415530a

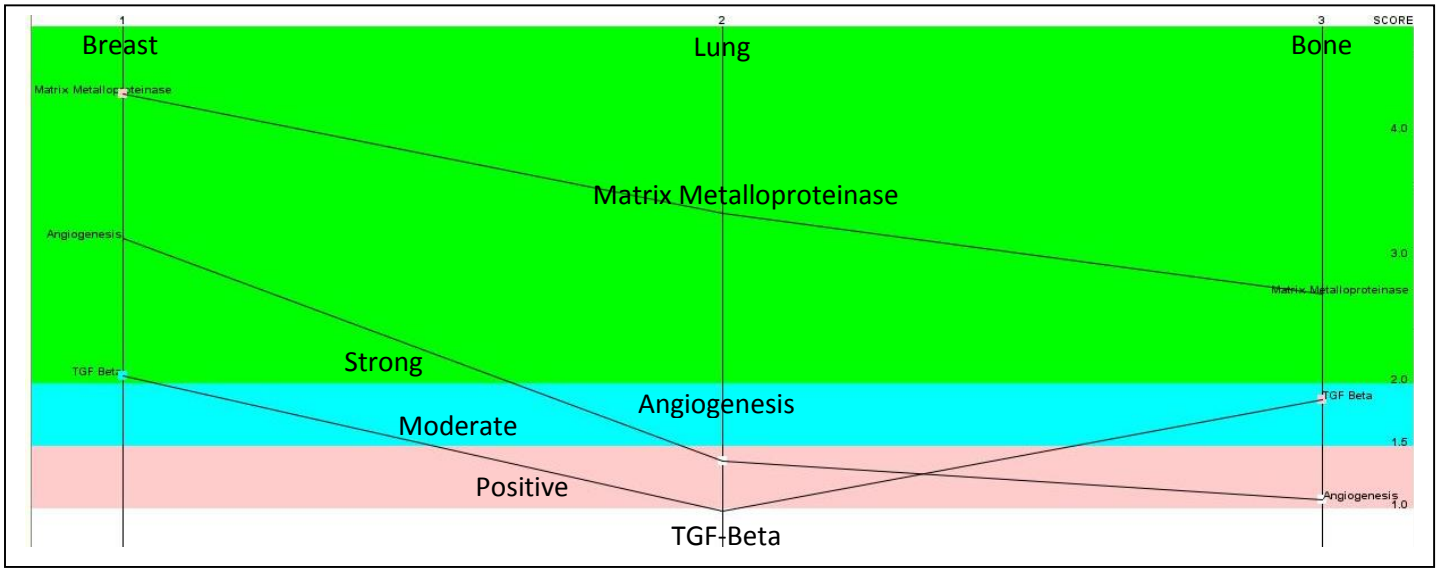
The landmark paper authored by van't Veer *et. al.* defined a prognostic 70-gene signature that identifies the metastatic potential of breast tumor cells. This signature was originally associated with several cell growth, metastatic, and angiogenesis pathways.

We next wanted to compare pathway associations between this 70-gene metastatic signature and two signatures linked with metastasis targeted to two organ systems: 1) lung and 2) bone. The goal was to identify pathways that are related to the distinction between basic metastatic potential and metastasis targeted to specific tissues, *i.e.* lung and bone.

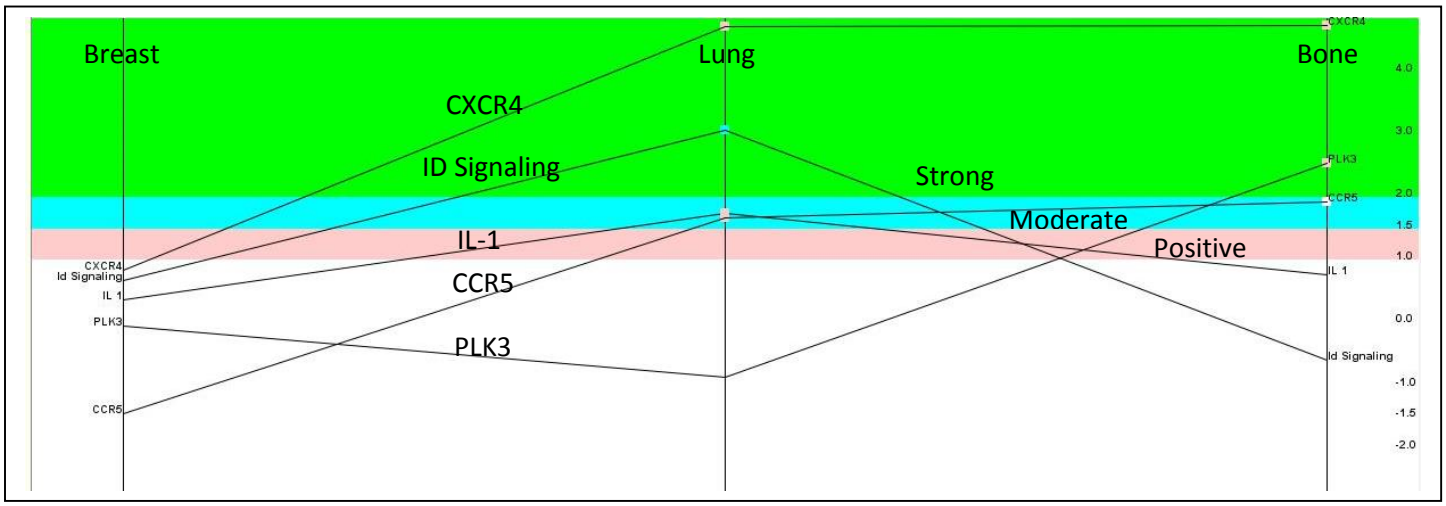
In the paper entitled **Genes that mediate breast cancer metastasis to lung** (Andy J. Minn, *et al*. *Nature* **436**, 518-524. July 28 2005), the authors define a 95-gene signature indicative of aggressive lung metastatic behavior by means of transcriptomic microarray analysis of highly and weakly lung-metastatic cell populations. The genes in this signature were largely distinct from those identified in bone metastatic isolates derived from the same parental line (see below).A 54-gene subset of this signature, which was more similar to the lung metastatic populations selected *in vivo* and postulated to serve specialized lung-related function, was selected as a refined signature. Included are genes associated with: EGF and HER/ErbB signaling, MMP1 and MMP2 matrix metalloproteases, and the IL13Ra2 and VCAM1 receptors, the latter indicating specific roles in the lung tumor microenvironment.

In the paper entitled **A multigenic program mediating breast cancer metastasis to bone** (Yibin Kang, *et al.* *Cancer Cell* **3,** 537-549. June 2003), the authors define a 102-gene signature indicative of osteolytic bone metastasis by means of microarray analysis of highly and weakly bone-metastatic cell populations. Most of these genes encode osteolytic and angiogenic factors and interestingly, none of the genes in the signature overlap with those in the van't Veer signature.

Using the gene list comparison tool, Literature Lab™ PLUS identified several pathways that are associated with all three of the gene lists, including those involved in angiogenesis, cell adhesion, and apoptosis.



We then looked at specific pathways that were identified in the previous papers as uniquely indicative of either lung or bone metastasis. Previous work identified gene combinations that act synergistically to promote lung metastasis including: ID1, SPARC, IL13Ra2, VCAM1, MMP2, MMP1, CXCL1, EREG, and COX2.Whereas, genes that were identified as promoting bone metastasis include: IL11, OPN, CXCR4 and CTGF.



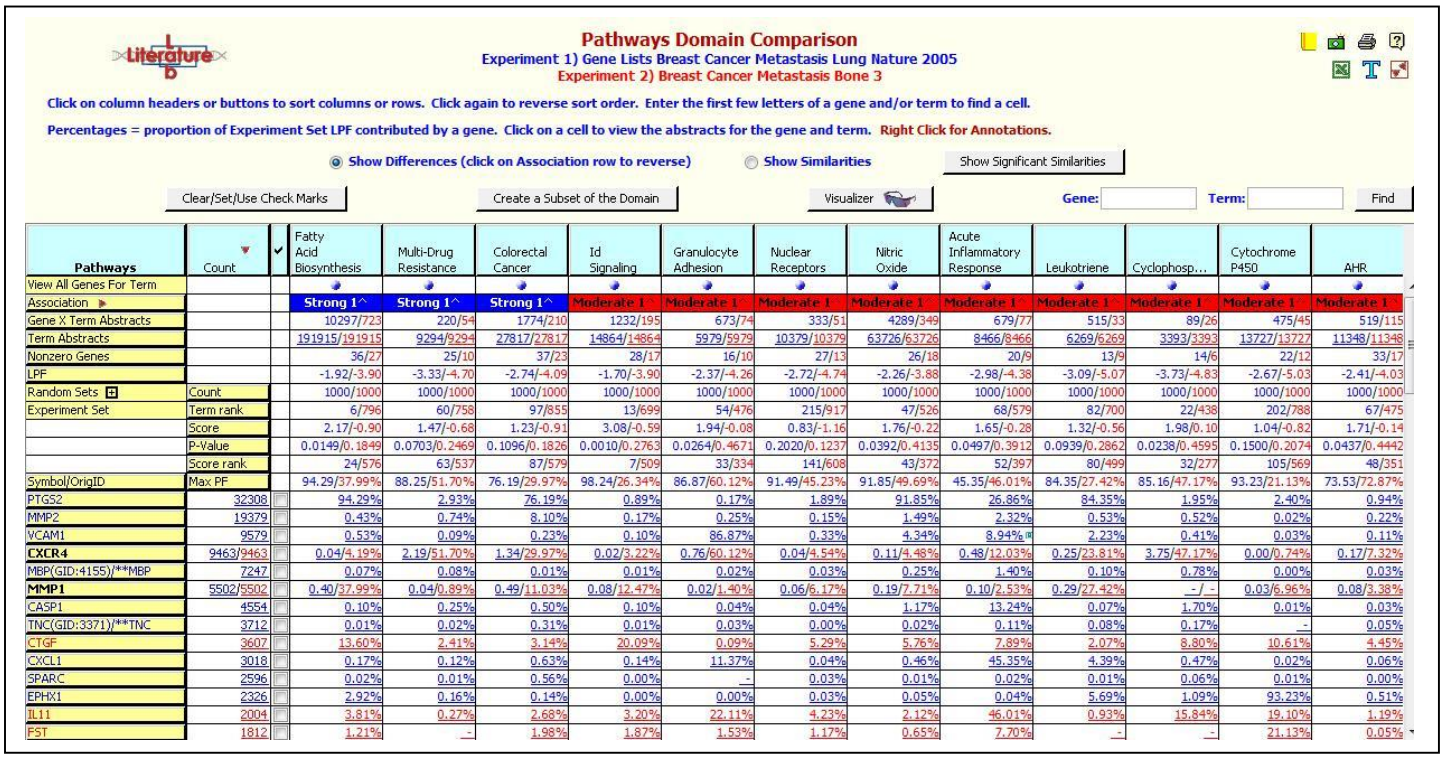
The Id Signaling pathway (ID1, inhibitor of DNA binding), previously identified as being significantly upregulated in lung metastasis, shows a predictably strong association with the lung metastasis gene set here. Interestingly, while IL1 is an inflammatory cytokine expressed by macrophages typically associated with infection, it shows a moderate association exclusively with the lung metastasis gene set. "Educated" macrophages often arise in response to the tumor microenvironment, and the exclusive moderate association between IL1 and lung metastasis may indicate a promising potential area of investigation.

Recent research has also pointed to the critical role that CXCR4 receptor and its ligand CXCL12 play in the metastasis of various types of cancer. Lung, bones and lymph nodes all secrete high levels of CXCL12, which acts as a chemoattractant that drives CXCR4-positive primary breast tumor cells towards these secondary metastatic sites. Therefore, it is not unexpected that the lung and bone gene lists show strong associations with CXCR4. It is however interesting that CCR5 shows a moderate association with these gene lists.CCR5 is predominantly expressed on T cells, macrophages, dendritic cells, eosinophils and microglia, and it is likely that CCR5 plays a role in inflammatory responses to infection, though its exact role in normal immune function is unclear. CCR5 may be another emerging player in the tumor metastasis microenvironment area that warrants further investigation.

PLK3 (Polo like kinase 3) is a cytokine inducible serine/threonine kinase that has been implicated in a number of cancers including breast cancer. The fact that it is strongly associated with the bone metastasis gene set, to a significantly greater degree than the other two gene sets, implies that there is a focus on the role of PLK3 in bone metastatic cancer. Likewise, PLK3 may be yet another promising target for further investigation.

Nature **415**, 530-536 (31 January 2002) doi:10.1038/415530a

We next used the Literature Lab™ gene set comparison feature to compare the lung and the bone gene signatures and to identify pathways that are significantly different between the two.ID Signaling is moderately different according to the statistical data and qualifiers, consistent with our earlier result. Interestingly, there are a number of pathway differences (*e.g.* Fatty Acid Biosynthesis, Multi-Drug Resistance, Colorectal Cancer, Nitric Oxide, etc.), which may be useful in further characterizing these metastatic subtypes and helping to drive hypotheses and future studies.



Pathways Domain Comparison

Experiment 1) Gene Lists Breast Cancer Metastasis Lung Nature 2005

Experiment 2) Breast Cancer Metastasis Bone 3

Click on column headers or buttons to sort columns or rows. Click again to reverse sort order. Enter the first few letters of a gene and/or term to find a cell.

Percentages = proportion of Experiment Set LPF contributed by a gene. Click on a cell to view the abstracts for the gene and term. Right Click for Annotations.

Show Differences (click on Association row to reverse) | Show Similarities | Show Significant Similarities

Clear/Set/Use Check Marks | Create a Subset of the Domain | Visualizer | Gene: | Term: | Find

Pathways	Count	Fatty Acid Biosynthesis	Multi-Drug Resistance	Colorectal Cancer	Id Signaling	Granulocyte Adhesion	Nuclear Receptors	Nitric Oxide	Acute Inflammatory Response	Leukotriene	Cyclophosp...	Cytochrome P450	AHR
View All Genes For Term													
Association		Strong 1^	Strong 1^	Strong 1^	Moderate 1	Moderate 1	Moderate 1	Moderate 1	Moderate 1	Moderate 1	Moderate 1	Moderate 1	Moderate 1
Gene X Term Abstracts		10297/723	220/54	1774/210	1232/195	673/74	333/51	4289/349	679/77	515/33	89/26	475/45	519/115
Term Abstracts		191915/191915	9294/9294	27817/27817	14864/14864	5979/5979	10379/10379	63726/63726	8466/8466	6269/6269	3393/3393	13727/13727	11348/11348
Nonzero Genes		36/27	25/10	37/23	28/17	16/10	27/13	26/18	20/9	13/9	14/6	22/12	33/17
LPF		-1.92/-3.90	-3.33/-4.70	-2.74/-4.09	-1.70/-3.90	-2.37/-4.26	-2.72/-4.74	-2.26/-3.88	-2.98/-4.38	-3.09/-5.07	-3.73/-4.83	-2.67/-5.03	-2.41/-4.03
Random Sets	Count	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000
Experiment Set	Term rank	6/796	60/758	97/855	13/699	54/476	215/917	47/526	68/579	82/700	22/438	202/788	67/475
	Score	2.17/-0.90	1.47/-0.68	1.23/-0.91	3.08/-0.59	1.94/-0.08	0.83/-1.16	1.76/-0.22	1.65/-0.28	1.32/-0.56	1.98/0.10	1.04/-0.82	1.71/-0.14
	P-Value	0.0149/0.1849	0.0703/0.2469	0.1096/0.1826	0.0010/0.2763	0.0264/0.4671	0.2020/0.1237	0.0392/0.4135	0.0497/0.3912	0.0939/0.2862	0.0238/0.4595	0.1500/0.2074	0.0437/0.4442
	Score rank	24/576	63/537	87/579	7/509	33/334	141/608	43/372	52/397	80/499	32/277	105/569	48/351
Symbol/OrigID	Max PF	94.29/37.99%	88.25/51.70%	76.19/29.97%	98.24/26.34%	86.87/60.12%	91.49/45.23%	91.85/49.69%	45.35/46.01%	84.35/27.42%	85.16/47.17%	93.23/21.13%	73.53/72.87%
PTGS2	32308	94.29%	2.93%	76.19%	0.89%	0.17%	1.89%	91.85%	26.86%	84.35%	1.95%	2.40%	0.94%
MMP2	19379	0.43%	0.74%	8.10%	0.17%	0.25%	0.15%	1.49%	2.32%	0.53%	0.52%	0.02%	0.22%
VCAM1	9579	0.53%	0.09%	0.23%	0.10%	86.87%	0.33%	4.34%	8.94%	2.23%	0.41%	0.03%	0.11%
CXCR4	9463/9463	0.04/4.19%	2.19/51.70%	1.34/29.97%	0.02/3.22%	0.76/60.12%	0.04/4.54%	0.11/4.48%	0.48/12.03%	0.25/23.81%	3.75/47.17%	0.00/0.74%	0.17/7.32%
MBP(GID:4155)/**MBP	7247	0.07%	0.08%	0.01%	0.01%	0.02%	0.03%	0.25%	1.40%	0.10%	0.78%	0.00%	0.03%
MMP1	5502/5502	0.40/37.99%	0.04/0.89%	0.49/11.03%	0.08/12.47%	0.02/1.40%	0.06/6.17%	0.19/7.71%	0.10/2.53%	0.29/27.42%	-/-	0.03/6.96%	0.08/3.38%
CASP1	4554	0.10%	0.25%	0.50%	0.10%	0.04%	0.04%	1.17%	13.24%	0.07%	1.70%	0.01%	0.03%
TNC(GID:3371)/**TNC	3712	0.01%	0.02%	0.31%	0.01%	0.03%	0.00%	0.02%	0.11%	0.08%	0.17%	-	0.05%
CTGF	3607	13.60%	2.41%	3.14%	20.09%	0.09%	5.29%	5.76%	7.89%	2.07%	8.80%	10.61%	4.45%
CXCL1	3018	0.17%	0.12%	0.63%	0.14%	11.37%	0.04%	0.46%	45.35%	4.39%	0.47%	0.02%	0.06%
SPARC	2596	0.02%	0.01%	0.56%	0.00%	-	0.03%	0.01%	0.02%	0.01%	0.06%	0.01%	0.00%
EPHX1	2326	2.92%	0.16%	0.14%	0.00%	0.00%	0.03%	0.05%	0.04%	5.69%	1.09%	93.23%	0.51%
IL11	2004	3.81%	0.27%	2.68%	3.20%	22.11%	4.23%	2.12%	46.01%	0.93%	15.84%	19.10%	1.19%
FST	1812	1.21%	-	1.98%	1.87%	1.53%	1.17%	0.65%	7.70%	-	-	21.13%	0.05%

Summary

The Literature Lab™ basic platform allows exploration of the PubMed database, highlighting co-occurences such as disease vs pathway, or co-morbidities *i.e.* disease vs disease. Breast, lung, and bone neoplasms were compared with pathways ranked according to strength of co-occurrence or LPF. The Literature Lab™ PLUS platform allows the identification of statistically significant associations between gene lists and term domains in the PubMed literature. In the 70-gene signature from the van't Veer study, strong, moderate, and positive associations were linked with several hallmark pathways of metastasis, a result shared with the lung and bone gene signatures.

Interestingly, the gene set comparison feature highlighted several pathways specific to either lung or bone, or shared by both. This indicates that there are pathways specifically associated with these tumor metastatic microenvironments that may suitable for targeted investigations and potential therapeutic development. Moreover, there were several other pathway associations not typically associated with metastasis that may be fruitful for future investigations. This latter result highlights the hypothesis generation potential of the Literature Lab™ platform.

