Literature Lab™ interrogation of gene sets relevant to gastric cancer: Comparison to NCI-Pathway Interaction Database based analysis.

Four gene sets were analyzed, specifically, those oncogenes identified in the four gastric cancer subtypes: Epstein-Barr positive tumors (EBV), Microsatellite unstable tumors (MSI, Genomically Stable tumors (GS), Chromosomal Instability tumors (CIN). Comparisons between Literature Lab and the Integrated Pathway Analysis methodology employed in the paper identified some similar trends in pathway associations. However, there were significant differences in pathways associated with TCPTP, ARF6, and Rho motility factors among others. These differences are summarized and may lead the way to novel hypotheses into the pathways relevant to gastric cancer subtype characterization and treatment. While the Integrated Pathway Analysis methodology compares experimental gene lists to a curated database (NCI-Pathway Interaction Database), Literature Lab treats each gene list as unique, and identifies and scores the strengths and significance of pathway associations mined in the PubMed literature. The two approaches are orthogonal and therefore complimentary; Literature Lab functions in a completely novel way offering insights that may be overlooked using traditional methods.

Comprehensive molecular characterization of gastric adenocarcinoma

The Cancer Genome Atlas Research Network. 2014. Nature. 513: Sept 11. 202-209.

Gastric cancer is a leading cause of cancer deaths, but analysis of its molecular and clinical characteristics has been complicated by histological and aetiological heterogeneity. Here we describe a comprehensivemolecular evaluation of 295 primary gastric adenocarcinomas as part of The Cancer Genome Atlas (TCGA) project. We propose a molecular classification dividing gastric cancer into four subtypes: **tumours positive for Epstein-Barr virus (EBV)**, which display recurrent PIK3CA mutations, extreme DNA hypermethylation, and amplification of JAK2, CD274 (also known as PD-L1) and PDCD1LG2 (also known as PD-L2); **microsatellite unstable tumours MSI)**, which show elevated mutation rates, including mutations of genes encoding targetable oncogenic signalling proteins; **genomically stable tumours (GS**), which are enriched for the diffuse histological variant and mutations of RHOA or fusions involving RHO-family GTPase-activating proteins; and **tumours with chromosomal instability (CIN)**, which show marked aneuploidy and focal amplification of receptor tyrosine kinases. Identification of these subtypes provides a roadmap for patient stratification and trials of targeted therapies.

Integrated pathway analysis - We integrated SCNA and mutation data to characterize genomic alterations in known signalling pathways, including candidate therapeutic targets (Fig. 5a, b). We focused on alterations in receptor tyrosine kinases (RTKs) and RAS and PI(3)-kinase signalling. EBV-positive tumours contained *PIK3CA* mutations and recurrent *JAK2* and *ERBB2* amplifications. Although MSI cases generally lacked targetable amplifications, mutations in *PIK3CA*, *ERBB3*, *ERBB2* and *EGFR* were noted, with many mutations at 'hotspot' sites seen in other cancers. Although the genomically stable subtype exhibited recurrent *RHOA* and *CLDN18* events, few other clear treatment targets were observed. In CIN tumours, we identified genomic amplifications of RTKs, many of which are amenable to blockade by therapeutics in current use or in development. Recurrent amplification of the gene encoding ligand VEGFA was notable given the gastric cancer activity of the VEGFR2 targeting antibody ramucirumab. Additionally, frequent amplifications of cell cycle mediators (*CCNE1*, *CCND1* and *CDK6*) suggest the potential for therapeutic inhibition of cyclin-dependent kinases.

Nature 513, 202–209 (11 September 2014) doi:10.1038/nature13480

We compared expression within each subtype to that of the other subtypes, and to non-malignant gastric tissue (n = 29). We computed an aggregate score for each pathway of the NCI pathway interaction database and determined statistical significance by comparison with randomly generated pathways.

Hierarchical clustering of samples and pathways (Fig. 5c) revealed several notable patterns, including elevated expression of mitotic network components such as AURKA/B and E2F, targets of MYC activation, FOXM1 and PLK1 signalling and DNA damage response pathways across all subtypes, but to a lesser degree in genomically stable tumours. In contrast, the genomically stable subtype exhibited elevated expression of cell adhesion pathways, including the B1/B3 integrins, syndecan-1 mediated signalling, and angiogenesis-related pathways. These results suggest additional candidate therapeutic targets, including the aurora kinases (AURKA/B) and Polo-like (PLK) family members. The strength of IL-12 mediated signalling signatures in EBV-positive tumours suggests a robust immune cell presence. When coupled with evidence of PD-L1/2 overexpression, this finding adds rationale for testing immune checkpoint inhibitors in EBV-positive gastric cancer.

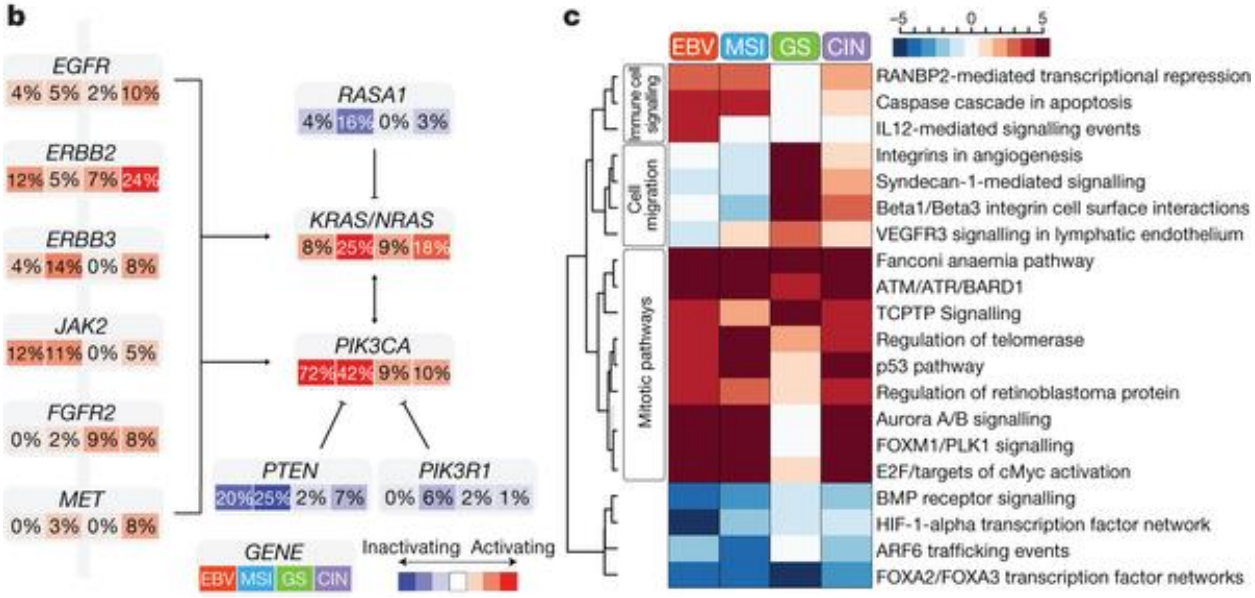


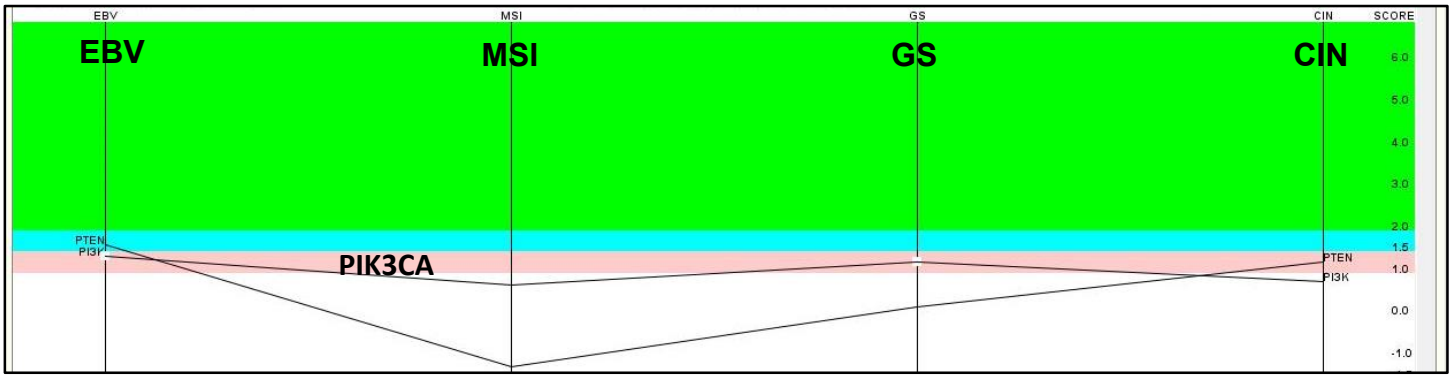
Fig 5 a, (not shown). Mutations, copy-number changes and translocations for select genes are shown across samples organized by molecular subtypes. Mutations that are recurrent in this data set or in the COSMIC repository are distinguished by colour. Alteration frequencies are expressed as a percentage of all cases.

Fig 5 b (shown). Alterations in RTK/RAS and RTK/PI(3)K signalling pathways across molecular subtypes. Red denotes predicted activation; blue denotes predicted inactivation.

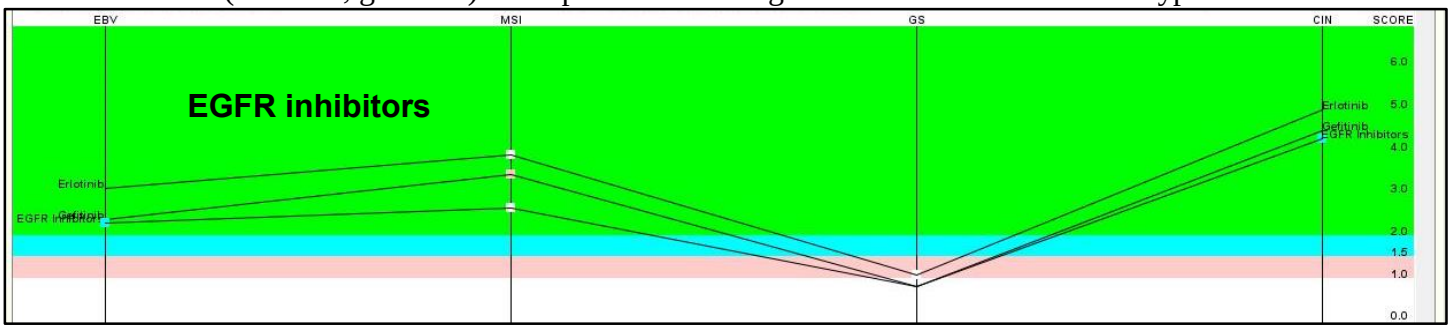
Fig 5 c (shown). The heatmap shows NCI-PID pathways that are significantly elevated (red) or decreased (blue) in each of the four subtypes as compared with non-malignant gastric mucosa.

Nature 513, 202–209 (11 September 2014) doi:10.1038/nature13480

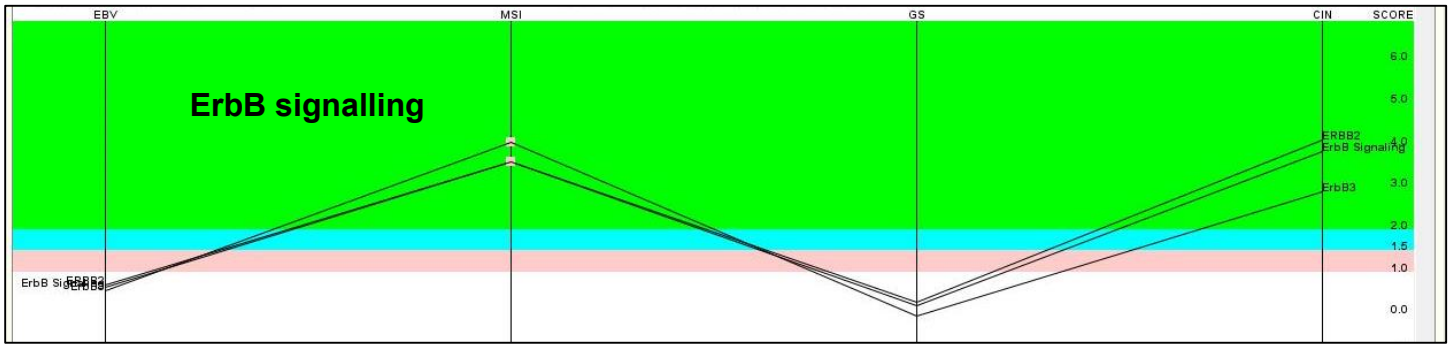
Literature Lab analysis of the four gene sets from the gastric cancer subtypes. In these data, the focus is on pathway associations using the gene set comparison feature (green – strong, blue – moderate, pink – positive associations). PIK3CA encodes p110 catalytic subunit of PI3K, which is opposed by the tumor suppressor PTEN. PI3KCA demonstrates positive association across the four subtypes.



EGFR Inhibitors (erlotinib, gefitinib) show positive to strong associations across four subtypes.



ErbB signaling pathway strongly associates in the genomic unstable subtypes (MSI, CIN) with near positive association in EBV and GS.



Analysis:

The PIK3CA signaling pathway component is positively associated across the four cancer subtypes, a similar observation as the paper although to differing degrees. The EGFR Inhibitors associations indicate a significant correlation between EGFR and the cancer subtypes, a result identified in the paper. The ErbB associations

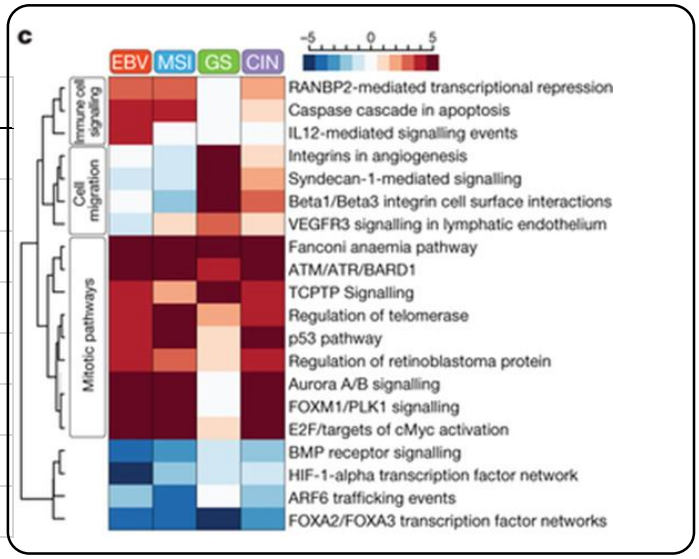
Nature 513, 202–209 (11 September 2014) doi:10.1038/nature13480

again are observed in the paper although to differing degrees. An important overall result is that, although associations are shared between the two methods, the degree or magnitude of these associations are quite different. These differences may be critical in the analysis of pathway targets for drug intervention.

'Heatmaps'(on the left, top) representing several pathways from clustering analysis via NCI-PID in the paper; IL12 mediated signaling events, TCPTP signaling, and ARF6 trafficking events. The same pathways were identified in Literature Lab, the Scores and P-values can be seen in the lower table. These Scores were used to produce a 'heatmap' based on the scale provided in Figure 5 c (cropped, on the right).

EBV	MSI	GS	CIN	NCI-PID Heatmap data; Fig 5c
				IL12 mediated signalling events
				TCPTP signalling
				ARF6 trafficking events

EBV	MSI	GS	CIN	Literature Lab data using scale in Fig 5c
				IL12 mediated signalling events
				TCPTP signalling
				ARF6 trafficking events



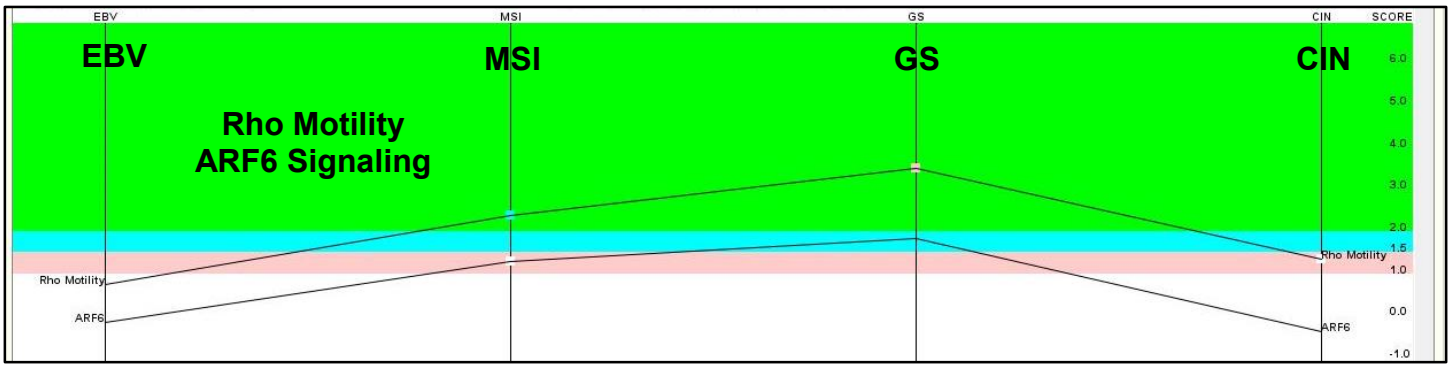
Literature Lab analysis: Score, P-value; color-coded to match association strengths in Literature Lab

EBV	MSI	GS	CIN	
2.62, 0.0044	1.40, 0.0814	0.29, 0.3845	-0.48, 0.3142	**IL12 mediated signalling events**
3.26, 0.0006	-1.84, 0.0330	-1.14, 0.1277	0.37, 0.3567	**TCPTP signalling**
-0.15, 0.4410	1.30, 0.0971	1.84, 0.0330	-0.38, 0.3513	**ARF6 trafficking events**
Strong	Moderate	Positive	Score, P-val	**Legend**

Analysis:

Literature Lab identified moderate and positive associations in IL12 mediated signaling events and TCPTP signaling, respectively, within the EBV subtype, a result overall similar to the NCI-PID method in the paper. Interestingly, an IL12 moderate association also was identified in the MSI subtype, a result absent in the paper. In contrast, there are distinct differences in associations between Literature Lab and the NCI-PID method. While the NCI-PID method identified considerable associations with TCPTP signaling across all four subtypes, these associations were absent in Literature Lab (with exception to EBV as described above). Conversely, ARF6 trafficking events were not positively associated with any of the subtypes via NCI-PID method. Literature Lab, however, identified associations within the MSI and GS subtypes, a result further exemplified in the comparison feature shown below.

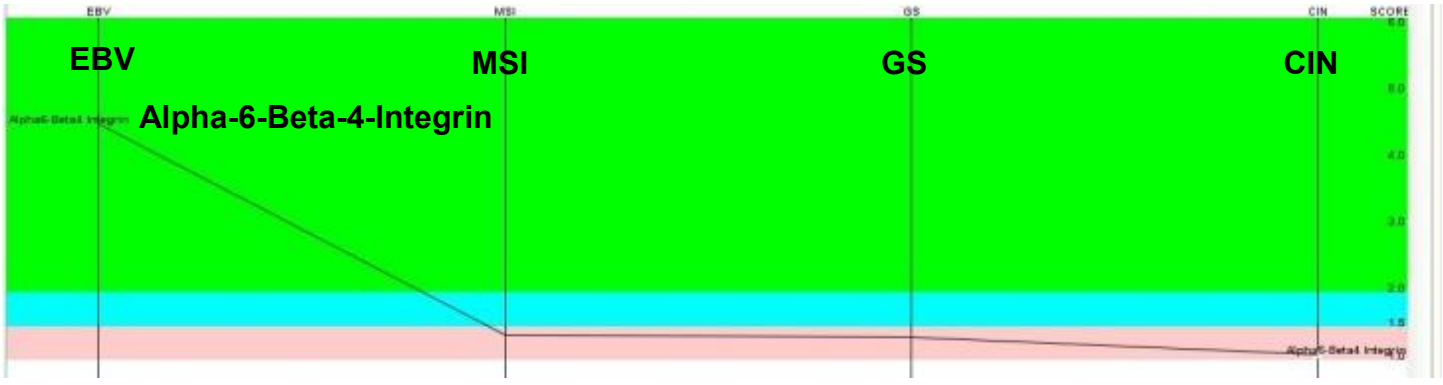
The Interaction Pathway Analysis (NCI-PID) method used in the paper showed ARF6 trafficking events were significantly decreased in the four subtypes. However, Literature Lab found positive and moderate associations in MSI and GS subtypes, while EBV and CIN fall just out of range.



The GS subtype was noted in the paper for elevated adhesion markers and also as the only subtype exhibiting recurrent Rho mutations. The combination of adhesion and Rho leads to Rho Motility and may be an indicator of metastasis. Interestingly, Rho Motility associates (positive, moderate, strong) across the four subtypes (the EBV subtype shows near positive association), despite only being mentioned in the paper regarding the GS subtype.

EBV	MSI	GS	CIN	Rho motility
				Integrative pathway analysis
				Literature Lab analysis

A heat map demonstrating the differences between the Integrated Pathway Analysis (NCI-PID) method used in the paper and Literature Lab with respect to Rho motility.



Literature Lab identified an important pathway (alpha-6-beta-4-integrin) that showed strong association with EBV and positive associations with the MSI and GS subtypes (CIN showed a near-positive association). Notably, no associations were found via the Integrated Pathway Analysis (NCI-PID) method in the paper.

Nature 513, 202–209 (11 September 2014) doi:10.1038/nature13480

Summary:

The Literature Lab gene set comparison feature allows the user to visualize the magnitude of associations across lists, in this case gastric cancer subtypes. Associations shown for comparison are: the PIK3CA signaling pathway, EGFR inhibitors (erlotinib, gefitinib), and the ErbB signaling pathway. The observation that associations also are seen using the Integrated Pathway Analysis (NCI-PID) method used in the paper, demonstrates the basic complementarity of the two gene set analysis platforms.

However, there are differences between these platforms as exemplified by the observation that strengths of the associations are somewhat different. For example, the NCI-PID method in the paper indicates that activation of ERBB2 and ERBB3 signaling is associated with only 5% and 14% of MSI tumors, whereas Literature Lab demonstrates the ERB signaling pathway is strongly associated with the MSI subtype.

Three additional pathways were chosen for comparison and contrast: IL12 mediated signaling, TCPTP signaling, and ARF6 trafficking. The IL12 pathway has immune and antiangiogenic properties and, in EBV linked cancer, components of the IL-12 pathway are differentially modulated, (*e.g.* increased expression (10-100 fold) of the IL12 p40 isotype is related to modulation of local immune responses in EBV-associated tumors). Whereas T-cell protein tyrosine phosphatase (TCPTP) is a ubiquitously expressed non-receptor protein tyrosine phosphatase that is involved in the negative regulation of many cellular signaling pathways. Finally, ADP-ribosylation factor 6 (ARF6) is a member of the ADP ribosylation factor family of GTP-binding proteins. This pathway plays a role in membrane trafficking and actin remodeling. Literature Lab values for score, log of product frequency (LPF), and P-value were used to generate heat maps, which were subsequently compared to heat maps for the Integrated Pathway Analysis (NCI-PID) method in the paper.

Again, similar associations were identified between Literature Lab and the NCI-PID method; for instance the IL12 and TCPTP signaling pathways within the EBV subtype. Importantly, there are a number of differences as well. Literature Lab identified an IL12 association in the MSI subtype whereas the NCI-PID method did not, indicating that perhaps some component(s) of the IL12 pathway is involved in MSI tumor biology, through immune or angiogenic activities. The NCI-PID method identified associations with TCPTP signaling across all four subtypes, in contrast to Literature Lab which found no significant associations across MSI, GS, and CIN. These Literature Lab results imply that, although the TCPTP pathway may be of importance in EBV tumors, it may not have as prominent a role in the other subtypes and therefore may not be an ideal therapeutic target.

Conversely, the NCI-PID method found no significant associations in ARF6 trafficking across the subtypes, whereas Literature Lab identified associations within the MSI and GS subtypes. ARF6 is a member of the RAS superfamily and the pathway's role in MSI and GS cancer subtypes may involve cell remodeling and metastasis. Differences also were evident in Rho motility, another indicator of metastasis, where the NCI-PID method identified an association with the GS subtype only. Literature Lab, on the other hand, found strong, moderate, and positive associations across all four subtypes, implying a pronounced role of this pathway in tumor proliferation and metastasis. All of these differences may point the way to untapped information and insight into novel lines of investigation.

A very interesting result in the Literature Lab analysis was the identification of the alpha-6-beta-4-integrin pathway associations which were absent in the NCI-PID method in the paper. There were strong associations in the EBV subtype and positive associations among the MSI and GS subtypes (CIN showed near positive association). Expression of alpha-6-beta-4-integrin is increased in many epithelial tumors, implicating its involvement in tumor malignancy. Moreover, this integrin activates several key signaling molecules in carcinoma cells thereby regulating tumor behavior. The Literature Lab associations indicate a prominent role

for alpha-6-beta-4-integrin and its targets in gastric cancer biology in general. The fact that the NCI-PID method did not identify the pathway given the same gene sets points to a broader discussion.

The differences in pathway associations between the Integrated Pathway Analysis (NCI-PID) method used in the paper and Literature Lab arise from the fundamentally different methods for gene-based literature interrogation. The NCI-PID method compares experimentally derived gene lists to a curated database of gene list information (NCI-PID) and draws inferences to literature-based pathways. Literature Lab considers each gene list as unique, and mines the literature directly measuring the strength of the association between a gene list (as a list) and a domain (pathway) using the strength of co-occurrence or log of product frequency (LPF), the contribution of individual genes to the LPF, and the significance of the association based on 1000 random gene sets (P-value). Both platforms identified some similar pathway associations in this work, and in this way the two can be thought of as complementary. However, Literature Lab's unique approach to gene set enrichment may uncover associations and connections that simply cannot be found otherwise, and therefore has the potential to be an invaluable tool in the search for gene function and biological relevance.

In addition, with Literature Lab the investigator is always one click away from the source material from which the domain (pathway) associations are derived. Literature Lab is not a black box and one doesn't have to just "accept" what it shows. One can read the abstracts behind a gene/term association if they want to explore the source material. With other approaches, the user is put in a position of having to accept what the tool is saying. Literature Lab allows the user to analyze the data, the literature, directly from the driver's seat.

Joel Jacobs PhD contributed to the data analysis and composition of this document.

Damon Anderson, PhD
Application Scientist
Acumenta Biotech
412-901-7785
danderson@acumenta.com