

Topological and Functional Discovery in a Gene Coexpression Meta-Network of Gastric Cancer

Amit Aggarwal,¹ Dong Li Guo,³ Yujin Hoshida,⁵ Siu Tsan Yuen,³ Kent-Man Chu,⁴ Samuel So,⁶ Alex Boussioutas,⁸ Xin Chen,⁷ David Bowtell,⁸ Hiroyuki Aburatani,⁵ Suet Yi Leung,³ and Patrick Tan^{1,2}

¹Cellular and Molecular Research, National Cancer Centre of Singapore; ²Genome Institute of Singapore, Singapore; Departments of ³Pathology and ⁴Surgery, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong, Republic of China; ⁵Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; ⁶Department of Surgery, Stanford University School of Medicine, Stanford; ⁷Department of Biopharmaceutical Sciences, University of California, San Francisco, California; and ⁸Peter MacCallum Cancer Centre, East Melbourne Victoria, Australia

Abstract

Gastric cancer is a leading cause of global cancer mortality, but comparatively little is known about the cellular pathways regulating different aspects of the gastric cancer phenotype. To achieve a better understanding of gastric cancer at the levels of systems topology, functional modules, and constituent genes, we assembled and systematically analyzed a consensus gene coexpression meta-network of gastric cancer incorporating >300 tissue samples from four independent patient populations (the “gastrome”). We find that the gastrome exhibits a hierarchical scale-free architecture, with an internal structure comprising multiple deeply embedded modules associated with diverse cellular functions. Individual modules display distinct subtopologies, with some (cellular proliferation) being integrated within the primary network, and others (ribosomal biosynthesis) being relatively isolated. One module associated with intestinal differentiation exhibited a remarkably high degree of autonomy, raising the possibility that its specific topological features may contribute towards the frequent occurrence of intestinal metaplasia in gastric cancer. At the single-gene level, we discovered a novel conserved interaction between the *PLA2G2A* prognostic marker and the *EphB2* receptor, and used tissue microarrays to validate the *PLA2G2A/EphB2* association. Finally, because *EphB2* is a known target of the Wnt signaling pathway, we tested and provide evidence that the Wnt pathway may also similarly regulate *PLA2G2A*. Many of these findings were not discernible by studying the single patient populations in isolation. Thus, besides enhancing our knowledge of gastric cancer, our results show the broad utility of applying meta-analytic approaches to genome-wide data for the purposes of biological discovery. (Cancer Res 2006; 66(1): 232-41)

Introduction

Genome-wide expression profiling technologies such as microarrays and SAGE have contributed towards elucidating the molecular pathways of various cancers (1), the definition of molecular signatures for patient prognosis and treatment response

(2–5), and the identification of new therapeutic compounds (6). In most previous studies, the processes of sample collection, processing, and data generation have typically been centralized and done at single centers, to control for potential sources of technical variation such as surgical handling, experimental protocols, and histopathologic scoring. Although successful, such single-center approaches can also be associated with possible limitations. For example, relying on a single profiling platform can potentially introduce systematic errors of analysis (e.g., the mistaken genomic assignment of an array probe, or a particular data normalization protocol) that might affect downstream biological and clinical interpretation. More significantly, particularly for clinical studies, it is frequently difficult for single centers to amass sufficiently large numbers of tissue samples to robustly validate initial findings derived from smaller sample sets. Reflecting this difficulty, there are relatively few single-center studies in the biomedical literature in which the number of profiled clinical samples exceeded 100 to 200. In response to this challenge, there has recently been a growing interest in the use of meta-analytic approaches to analyze data from multiple data sets, to validate patterns of gene expression by examining their consistency of behavior across different studies (7). However, in addition to validating known results, there is also the intriguing possibility that such meta-approaches could potentially play a powerful role in biological discovery by generating novel insights and biological relationships. Although some studies of this kind have been reported, most of these reports have typically focused on pooling expression data derived from a wide variety of tissue and tumor types (8, 9). In contrast, comparatively fewer reports have specifically focused on cancers from a single site of origin to investigate tissue-specific aspects of cancer.

In this report, we present a systematic and comprehensive meta-analysis of the gene coexpression networks in gastric cancer, the second highest cause of global cancer mortality next to lung cancer (10). To assemble a collection of clinical samples with sufficient breadth and diversity to characterize the gene expression interactions associated with this complex disease, we formed an international Gastric Cancer Consortium whose founding members from Australia, Hong Kong, Japan, and Singapore, had previously reported analyses of single-center gastric cancer data sets (11–14). We pooled the data from these independent studies to create a gene expression database of gastric cancer, comprising >300 human samples profiled at various histologic stages of gastric tumorigenesis, ranging from normal gastric tissue, chronic gastritis, intestinal metaplasia, to overt carcinoma. Employing rank-based statistics, we identified gene coexpression interactions conserved across the four centers and used these interactions to

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Requests for reprints: Patrick Tan, Cellular and Molecular Research, National Cancer Centre of Singapore, 11 Hospital Drive, Singapore 169610. Fax: 65-6-226-5694; E-mail: cmrtan@nccs.com.sg.

©2006 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-05-2232

construct a consensus gene coexpression meta-network of gastric cancer (the “gastrome”). The resulting meta-network supports a scale-free hierarchical architecture, containing several deeply embedded functional modules associated with distinct biological functions. Although some of these modules were previously known, other modules were previously unreported, demonstrating the ability of the meta-approach to reveal novel biological information. The modules were diverse in their overall connectivity—some modules (e.g., cellular proliferation), were highly integrated with other modules, whereas others (e.g., ribosomal biosynthesis) were isolated and autonomous. Strikingly, a module associated with intestinal differentiation exhibited a remarkably high degree of autonomy, raising the possibility that the specific topological features of this module may functionally contribute towards the frequent appearance of intestinal metaplasia in gastric cancer. Finally, the gastrome analysis identified biological relationships that were less apparent from the single data sets—specifically, we found that *PLA2G2A*, previously identified as a potential prognostic marker for gastric cancer (15), exhibited a conserved coexpression relationship with the *EphB2* receptor and subsequently validated this association using tissue microarrays. Motivated by previous reports that *EphB2* is a target of the Wnt signaling pathway, we obtained further experimental evidence that the Wnt pathway may also regulate *PLA2G2A*. Taken collectively, our findings show that meta-analytic approaches can successfully identify novel systems-level features as well as subtle but potentially important biological relationships in tumor biology. Such strategies may thus prove useful and applicable to many other disease types, for the purposes of both biological validation and discovery.

Materials and Methods

Gene expression data sets and data preprocessing. The gastric cancer expression database derives from four independent studies (refs. 11–14; Supplementary Information S1). Each data set was individually normalized, preprocessed, and deposited into a centralized server for further integration and preprocessing. 2,252 unique Unigenes overlapped between the four centers’ data sets. Multiple probes mapping to the same Unigene were averaged after a \log_2 transformation of the data. Histopathologic and clinical information associated with the tissue samples are described in the original reports and summarized in Supplementary Information S2. The expression data sets, collectively termed a “meta-data set,” can be downloaded from <http://www.omniarray.com/GCGC/>. The primary data sets are available at <http://www.ncbi.nlm.nih.gov/geo/>.

Identification of conserved coexpression interactions. For each data set, we calculated Pearson correlation coefficients between every gene-gene pair, and for each gene, ranked all other genes (from 1 to 2251) by its correlation coefficient to the former, rank 1 being the least correlated and rank 2,251 the most correlated. Notably, although the absolute correlation value of $(A \rightarrow B) = \text{correlation}(B \rightarrow A)$, the rank of gene A with respect to gene B may not be the same as the rank of gene B with respect to gene A (see Supplementary Methods SM1 for an example). To identify gene pairs exhibiting consistently high ranks across multiple data sets, we calculated the joint probability of observing a particular sequence of ranks: briefly, for a random sample X of size “ n ” drawn from a uniform distribution $[1 \dots N]$ ($n = 4$ and $N = 2,252$ in this case), if the “ n ” numbers are sorted such that $X_{(1)}$ and $X_{(n)}$ are the smallest and largest numbers, respectively; then $X_{(k)}$ follows a β distribution with parameters $(k, n - k + 1)$, and $p(X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)})$ or the joint probability “ p ” of observing a sequence of ranks $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ can be calculated (Supplementary Methods SM2). Because the $X_{(k)}$ s correspond to an observed sequence of ranks for a particular gene pair (A, B) across the four data sets, we can define the null hypothesis H_0 that for a gene pair (A, B) the sequence of $X_{(k)}$ s or ranks is random, and the alternative hypothesis H_1 that the observed ranks across

the four data sets is nonrandom. Using a metric termed the “log-likelihood ratio” (LLR) = $\log_{10} [p(H_0)/p(H_1)]$, H_0 is rejected iff $\text{LLR} \geq \text{LLR}_{\text{crit}}$, where LLR_{crit} is user-defined. As the LLR is based on ranks, $\text{LLR}(A \rightarrow B) \neq \text{LLR}(B \rightarrow A)$, hence we define $\text{LLR}(A \leftrightarrow B) = \max[\text{LLR}(A \rightarrow B), \text{LLR}(B \rightarrow A)]$, and two genes A and B are called “coexpressed” iff $\text{LLR}(A \leftrightarrow B) \geq \text{LLR}_{\text{crit}}$. Only 0.94% and 0.27% of possible interactions show a LLR difference >1.5 and 2 between $\text{LLR}(A \rightarrow B)$ and $\text{LLR}(B \rightarrow A)$, respectively (A. Aggarwal, data not shown). The LLR score is also blind to gene pairs being positively or negatively correlated; thus, this technique does not segregate on the basis of gene-pairs being positively or negatively coexpressed. In addition to the LLR, a false discovery rate (FDR) was used to reflect the fraction of false-positives present in hypotheses deemed significant above a given LLR_{crit} . To estimate the FDR, we generated 50 randomly permuted meta-data sets in which the rank order of the genes in the single-center data sets were shuffled, and calculated the number of links found significant at a given LLR. This randomization process was then repeated 50 independent times and the results averaged. For a given LLR_{crit} , $\text{FDR} = \frac{\#\{\text{LLR} \geq \text{LLR}_{\text{crit}} | H_0\}}{\#\{\text{LLR} \geq \text{LLR}_{\text{crit}} | H_1\}}$.

Clustering coefficient. The clustering coefficient (16) C_i of a node “ i ” with “ k_i ” neighbors is $C_i = 2 \times C / k_i(k_i - 1)$ where C is the number of connections between the “ k_i ” neighbors of node “ i .” The C_i provides a quantification of the extent to which the coexpression neighbors of a gene are also connected to each other. The clustering coefficient for the network is given by $C_{\text{No.}} = 1/\text{No.} \sum C_i$ where “No.” is the number of nodes in the network with more than one coexpression partner. For comparative analysis of the clustering coefficient, we simulated “equivalent” scale-free networks using a previously described preferential attachment model (ref. 17; Supplementary Methods SM3).

Assembly of expression communities and functional modules. To generate expression communities, we identified all coexpressed gene pairs of $\text{LLR} \geq \text{LLR}_{\text{crit}}$ and connected these genes by “chain-linking”: assume gene A’s maximal link occurs with gene B, and gene B’s maximal link is to gene C (and not necessarily gene A). Gene A is then connected to gene B, and gene B to gene C. This process continues, forming a gene chain, until a terminator pair is encountered which is reciprocally maximally linked (i.e., gene A \rightarrow gene B $\rightarrow \dots \rightarrow$ gene I \leftrightarrow gene J). Chains of length >3 and which end in the same chain terminator pair are pooled to form a “scaffold.” The coexpression partners of the genes in the “scaffold” are subsequently aggregated to form an expression community. Communities of size $> S_{\text{min}}$, where S refers to the number of genes, are considered for further analysis. To assess the similarity of any two communities, we calculated the probability of overlap: two communities of gene sizes S and T are chosen out of N (2,252 in this case), with the probability of the two groups having M genes in common being $\frac{C(N,M) C(N-M,S-M) C(N-S,T-M)}{C(N,S) C(N,T)}$, where $C(X,Y)$ is the number of ways sets of size Y can be selected from a set of size X ($X \geq Y$). To assemble functional modules from the communities, we ranked the communities in order of increasing size. Starting with the smallest community; its highest overlapping community was identified and if 50% or more of the members in the former were also present in the larger community, then the smaller-sized community was merged with the larger. The above steps were repeated until all the expression communities were partitioned into distinct functional modules. To confirm their modularity, we compared the numbers of internal to external expression links of each module to an equivalent number of randomly selected genes drawn from the original 588 gene set (see Results). Across 2,000 runs, the modularity of the former always exceeded the random data, thus assigning an empirical $P < 0.0005$. We note that in constructing the final network, all samples (nonmalignant and cancerous) were examined irrespective of phenotype. However, subnetworks that are present in both nonmalignant and malignant tissues (e.g., the ribosomal network) do not seem to exhibit significantly higher statistical associations compared with subnetworks present in either cancer or nonmalignant tissues only (Supplementary Methods SM4).

Hierarchical clustering and other software sources. Average linkage hierarchical clustering was done using an uncentered correlation similarity metric. Median centering by genes was done prior to clustering each data set. Cluster and Treeview (<http://rana.lbl.gov/EisenSoftware.htm>) software were used for clustering and generating expression heatmaps. All other

methodologies were implemented in MatLab (<http://www.mathworks.com>) and the network plots visualized using Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

Immunohistochemistry. A gastric cancer tissue microarray of 343 gastric tumors was prepared using standard methods (ref. 18; see Supplementary Methods SM5 for clinical data associated with these arrays). *Villin1* (VIL1) expression was assessed using antibody ID2C3 (Immunotech, France) at a 1:200 dilution and streptavidin-biotin peroxidase after heat-mediated antigen retrieval. EphrinB2 (*EphB2*) expression was assessed using goat anti-*EphB2* (R&D Systems, Minneapolis, MN) at a 1:100 dilution and the DAKO (Carpinteria, CA) EnVision+ System Peroxidase (3,3'-diaminobenzidine) after heat-mediated antigen retrieval. Phospholipase AII group IIA (*PLA2G2A*) expression was assessed by nonradioactive *in situ* hybridization (19). Primers for generating the PCR-based riboprobe templates (minus T7 promoter sequences) were: *PLA2G2A* (forward) TTCTACGGCTGC-CACTGTGG; (reverse) GGAGGAGAGCAGTAGAAGGC. A β -actin riboprobe (Roche) was used to ensure RNA integrity. For cell line experiments, a mouse monoclonal anti- β -catenin antibody (Transduction Laboratory, Lexington, KY) was used at a 1:200 dilution. Gastric cell lines were cultured to 80% confluence, washed in PBS and fixed in 4% paraformaldehyde. Paraffin cell blocks were prepared. Immunohistochemical staining was done using the DAKO EnVision+ system, peroxidase (3,3'-diaminobenzidine) after heat-mediated antigen retrieval.

Results

The gastrome: a consensus gene coexpression meta-network of gastric cancer. We combined microarray expression data from four separate studies encompassing a total of 302 patient samples

reflecting different aspects of gastric cancer pathology, from normal mucosa, chronic gastritis, intestinal metaplasia, to carcinoma (Supplementary Information S1 and S2). Collectively termed a “meta-data set,” the four data sets were integrated as shown in Fig. 1A. First, we identified 2,252 common Unigene entities across the four data sets, and computed Pearson's correlation coefficients for all possible gene pairs within each data set. For each individual gene, we ranked all other 2,251 genes based on their strength of correlation to the former, producing for each data set a “rank matrix” of $\sim 5 \times 10^6$ ($2,252 \times 2,252$) pair wise ranks, and ultimately associating every gene pair (A, B) with a characteristic sequence of four ranks (one from each data set). For example, the rank order of proliferating cell nuclear antigen (*PCNA*) with respect to *TOP2A* is (2,224, 2,230, 2,242, and 2,244), indicating that *PCNA* is the 28th, 22nd, 10th, and 8th most positively correlated gene to *TOP2A* across the four sets. To identify pairs of genes whose expression was significantly correlated and conserved, we applied a rank-based statistic to calculate the LLR for each gene pair, in which a large LLR indicates a deviation from the null hypothesis (i.e., that the sequence of ranks associated with the gene pair is unlikely to be random). “Significant” gene pairs (henceforth termed “expression links”) were defined as those whose LLR exceeded a user-defined cutoff (LLR_{crit}). To reflect the balance between sensitivity and specificity, we computed FDRs to estimate the proportion of false-positive expression links associated with any particular LLR value—this was achieved by generating a series

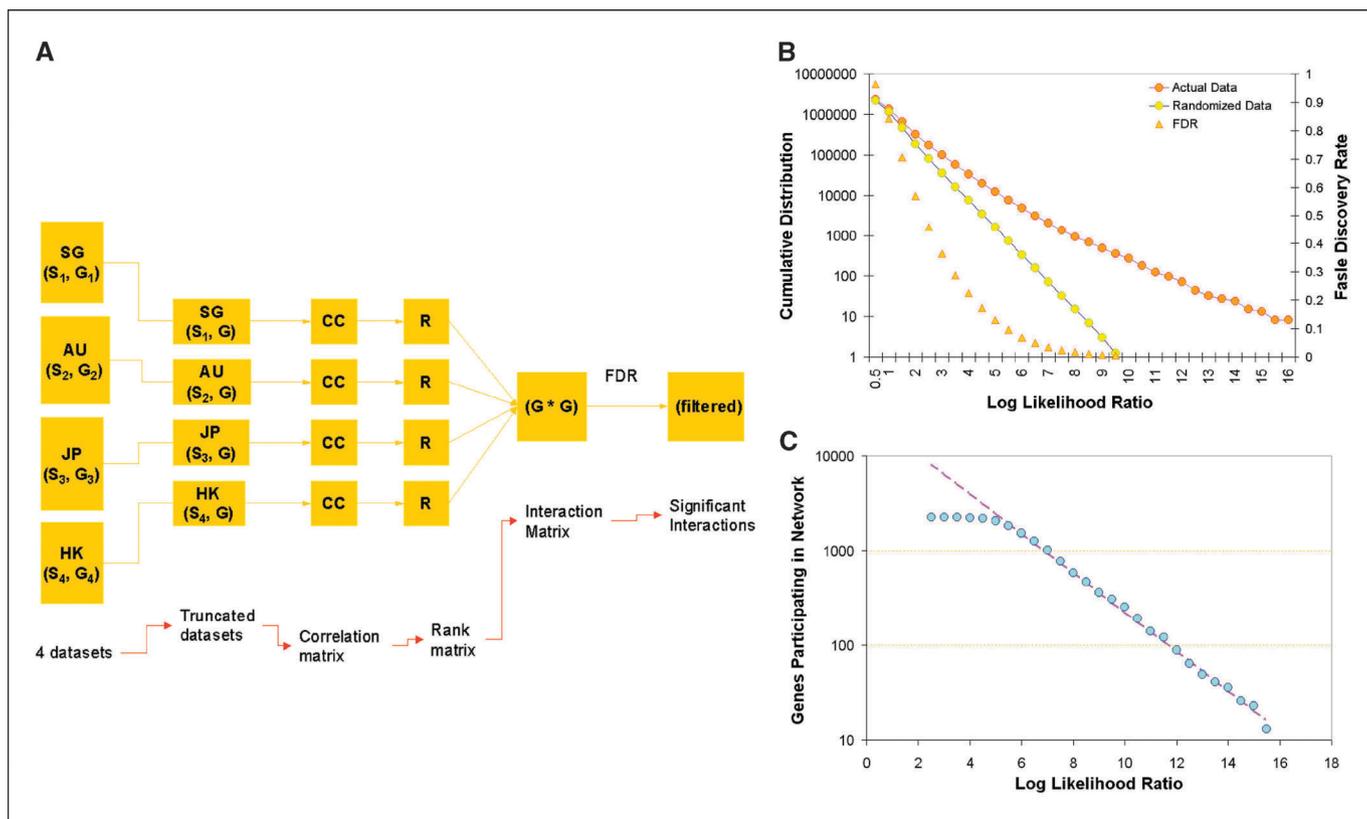


Figure 1. Identification and distribution of conserved coexpression links. **A**, general schematic for generating the coexpression meta-network (gastrome). The four data sets (SG, AU, JP, HK) were each truncated to 2,252 common Unigenes (G), and Pearson's correlations were calculated for each gene pair to construct correlation matrices (CC), which were subsequently ranked (R). A rank-statistic is used to evaluate the consistency of ranks observed for each gene pair, creating an interaction matrix of $2,252 \times 2,252$ LLR values (see Materials and Methods). On the basis of FDR values, an LLR cutoff is used to identify significant interactions. **B**, cumulative distribution of expression links in actual (orange circles) and randomly permuted networks (yellow circles) for absolute numbers of links (left axis, log scale) and FDR (right axis, linear scale), the latter being the ratio of random/actual links at a given LLR. Expression links were binned over LLR intervals of 0.5. Values for random data are the average of 50 permutation runs. **C**, number of genes participating in the network with changes in LLR.

of randomly permuted meta-data sets in which the gene names within each individual data set were internally shuffled, and determining the number of links found to be “significant” within the randomized data (see Materials and Methods).

The results of this procedure are shown in Fig. 1B. First, we observed that at all LLR values >0.5 , the numbers of significant expression links in the actual data exceeded those in randomly permuted data. The gap between actual and random data increased at higher LLRs, reducing the FDRs and leading to higher specificity (Fig. 1B, yellow triangles). For example, at LLRs 5 and 8, ~ 12.3 k and 925 links were found to be significant, with associated FDRs of $\sim 13\%$ and $\sim 1.6\%$, respectively. This result indicates that a substantial number of expression links are indeed conserved across the different data sets, which are likely to represent bona fide biological interactions. Second, we found that increasing the LLR cutoff caused the numbers of significant expression links to decrease at a greater rate than the number of individual genes (Fig. 1B, right), indicating that increases in specificity can be apparently obtained without incurring a similarly dramatic penalty in the absolute number of genes participating in the network. For example, increasing the LLR stringency from 4 to 8 caused a 36-fold drop in the numbers of expression links (~ 33.5 k to 925) but only a 3.7-fold reduction in the number of genes (~ 2.2 k to 588). This result is consistent with the hypothesis that the majority of genes are associated with relatively few highly specific expression links, a concept that is further developed later in the report. Third, to test if the significant expression links were robust or sensitive to the presence of particular sample types, we generated 10 randomly truncated meta data sets wherein each 50% of the original samples were removed, and compared the significant links in each truncated set to those found in the original meta-data set. We found that at $LLR \geq 5$, $\sim 57\%$ of the significant links in the truncated data sets were conserved with the complete data set, and that the conservation increased to $\sim 64\%$ at $LLR \geq 8$ (corresponding FDRs are $\sim 21.3\%$ and $\sim 3.2\%$). This result suggests that the sample population used in this study is reasonably broad and diverse with respect to gastric cancer physiology, as the majority of the significant expression links are fairly robust to the presence of any particular sample type. We also checked the dependence of the expression links on the specific presence of nonmalignant tissues by removing the 70 nonmalignant samples (including normal, chronic gastritis, and intestinal metaplasia) and repeating the entire analysis. Despite the removal of nonmalignant tissues, a similar result was obtained in which $\sim 56\%$ of the expression links were conserved between the “malignant-only” and the complete data set at $LLR \geq 5$, increasing to $\sim 61\%$ at $LLR \geq 8$ (corresponding FDRs $\sim 13\%$ and $\sim 1.5\%$), once again indicating their robustness. These results show that rank-order statistical approaches can be successfully used to identify conserved and robust gene expression interactions from multiple disparate data sets, despite the latter being derived from distinct patient populations and array technologies. Having identified these interactions, we assembled them into a gene coexpression meta-network of gastric cancer, termed the “gastrome.” We now analyze the gastrome in terms of its general topology, functional modules, and constituent genes.

A topological analysis of the gastrome reveals a hierarchical scale-free architecture with embedded modularity. Previous reports analyzing the topologies of various biological networks in simpler model organisms such as bacteria and yeast have suggested a common “scale-free” character. In a scale-free network,

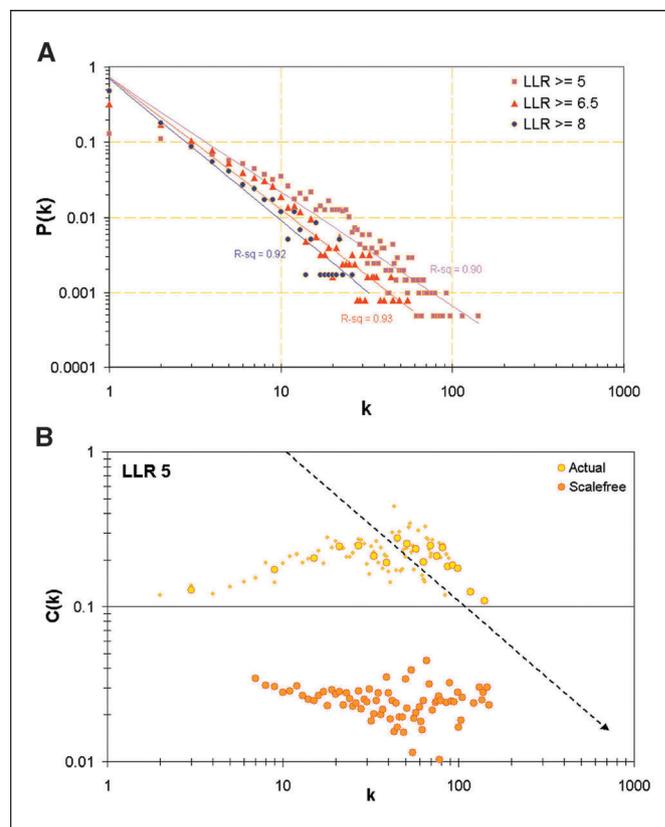


Figure 2. Topological characteristics of the gastrome. *A*, power law distribution of the gastrome, with $P(k)$ (*y*-axis, log scale) being the probability of a gene having k expression links (*x*-axis, log scale). $P(k) = n(k)/n_{sig}$, where $n(k)$ represents the number of genes with k significant expression links, and n_{sig} the total number of genes. Three different LLR cutoffs are depicted: ≥ 5 (pink squares), ≥ 6.5 (red triangles) and ≥ 8 (blue circles). *B*, evidence for a hierarchical assembly in the gastrome. The clustering coefficient “ $C(k)$ ” (*y*-axis) is plotted on a log-log graph as a function of k (*x*-axis) at $LLR_{crit} \geq 5$. In comparison to an equivalent pure scale-free network (orange circles), a decrease in $C(k)$ with increasing k is observed. Large circles, smoothed values from the original data (small circles).

most of the connections are confined to a few major nodes (hubs), which link the other, lesser-connected nodes in the network. In contrast, in random or Gaussian networks, the connections are spread in a statistically homogenous manner across the nodes, such that most nodes have similar numbers of links (see refs. 16, 20). To determine the overall network architecture of the gastrome, we computed the probability $P(k)$ of a gene having “ k ” expression links, and found that $P(k)$ was inversely related to k following a power law, a hallmark of scale-free networks (Fig. 2A). We confirmed this finding across a range of LLRs, indicating that an overall scale-free topology is a fairly robust feature of the gastrome, similar to that reported for other gene coexpression, protein interaction, and metabolic networks. However, despite this overt similarity in scale-free character, we report here the possibility that there might exist formal differences between the different network types. Specifically, whereas γ , the slope of the power law distribution, typically behaves as $\gamma \in (2, 3)$ for metabolic (20) or protein interaction networks (21), γ seems to be consistently <2 for a number of gene coexpression networks [$\gamma = 1.5$ –1.8, 1.5, and 1.1–1.8 for this study, ref. (22) and ref. (23), respectively]. In general, a γ exponent in the range $\in (1, 2)$ will result in a network not possessing a characteristic mean and variance, along with being

dominated by nodes with large degrees (24). An important consequence of this is that because the overall number of nodes is finite, nodes of large degrees will be preferentially connected to one another rather than to nodes of lesser degrees. To test this, we analyzed the cohort of highly connected genes within the gastrome, and found that genes with large numbers of expression links were indeed highly biased in connecting to other “highly connected” genes rather than genes with few connections ($P < 10^{-5}$), and that this phenomenon was not an automatic property of a generic scale-free system (Supplementary Information S3). This result suggests that gene coexpression networks, despite being scale-free, may nevertheless possess topological properties distinct from biological networks from other cellular levels. This peculiar network design may facilitate the cell’s ability to control the activity of multiple transcriptional programs in a coherent fashion. However, this is clearly speculative and more work is required to investigate the functional properties conferred by this particular systems’ organization.

The cohort of highly connected genes was apparently skewed towards certain biological functions, in particular cellular proliferation (*PCNA*, *CKS2*, *CRYAB*, *A2M*, and *MAD2L1*) and immune response (*CD2*, *CD53*, *CCL19*, and *THBS4*). This finding made us consider if the gastrome, within its scale-free architecture, might also harbor additional subnetworks associated with distinct cellular functions and topologies. To formally address this possibility, we tested the gastrome for the presence of modularity. Modularity in a network refers to a set of nodes that are more internally connected compared to nodes from other parts of the network, which may represent functional subnetworks. We computed clustering coefficients, \hat{C}_{No} , a metric of global network modularity, for the gastrome at four different LLR cutoffs, and compared these coefficients to those found in either simulated pure scale-free or random Gaussian networks, the latter possessing a non-scale-free topology. At all LLR ranges, the clustering coefficient of the gastrome was substantially higher than either the pure scale-free or a Gaussian network, indicating that the gastrome is indeed highly modular (Table 1). Notably, the modularity of the gastrome, as reflected by \hat{C}_{No} , became more evident when the LLR cutoff was increased—this is consistent with a hierarchical model of assembly in which core modules are clearly discernible at a high LLR stringency, and relaxing this stringency

subsequently “draws in” expression links combining these modules, resulting in less cohesive structures (25). Previous work has shown that $C(k)$ is essentially independent of k for purely scale-free or modular architectures, but for hierarchical networks $C(k)$ reduces as $1/k$ (16). To test the gastrome for such a hierarchical assembly, we computed the dependence of $C(k)$ on k at different LLRs. A typical case is shown in Fig. 2B in which the $C(k)$ decreases with increasing k , thereby indicating the presence of hierarchy. Taken collectively, our topological analysis suggests that the gastrome is organized along a hierarchical scale-free architecture, with a deeply embedded substructure of distinct subnetworks and modules.

A modular analysis of the gastrome reveals both known and novel coexpression subnetworks. To identify these functional modules in the gastrome in a systematic and unbiased manner, we then implemented a novel bottom-up stepwise methodology referred to as “chain-linking” (see Materials and Methods). For this analysis, we focused on the set of expression links above LLR8, because (a) the modular structure of the gastrome is most apparent as this stringency and (b) these links have a low associated FDR ($\sim 1.6\%$), and are thus highly specific. In “chain-linking,” the expression links, comprising 925 expression links across 588 unique genes, were combined into gene-gene maximal neighbor “chains,” and subsequently pooled into “communities” with embedded tree-like structures (Fig. 3A) without restricting genes to be exclusively present in any single assembly. We then agglomerated those communities exhibiting high overlap into larger units. This latter property distinguishes the groupings defined by this approach from the conventional “expression clusters” generated by hierarchical clustering algorithms. We constructed 298 gene-gene chains each of length ≥ 3 genes, pooled them into 31 separate communities, and combined the communities exhibiting high overlap ($>50\%$) into 13 distinct units (Fig. 3B). Using random permutation assays, we also confirmed that each of these units was more internally than externally connected ($P < 0.0005$ for all units), supporting their inherent modularity, and that the integrity of each unit was maintained even after variation of the LLR cutoff (Supplementary Information S4). We henceforth refer to these units as functional modules.

We curated the modules and found that many of them could be assigned a functional or physiologic role (Fig. 3B; complete gene lists and references are provided in the Supplementary Information S5).

Table 1. Comparison of overall clustering coefficients at different LLR_{crit} cutoffs for the gastrome (\hat{C}_{No}) and equivalent pure scale-free (\hat{C}_{sf}) and random (Gaussian) networks (\hat{C}_{rnd})

LLR _{crit}	LLR ≥ 4	LLR ≥ 5	LLR ≥ 6.5	LLR ≥ 8
No.	2,216	1,801	848	303
k_{avg}	66,488/2,216 ~ 30	24,213/1,801 ~ 14	5,837/848 ~ 8	1,565/303 ~ 5
\hat{C}_{No}	0.166	0.174	0.217	0.271
Scale-free				
m	~ 15	~ 7	~ 4	~ 3
N_{sf}	2,200	1,800	850	300
\hat{C}_{sf}	0.042 (0.006)	0.022 (0.003)	0.037 (0.004)	0.068 (0.014)
Random				
$\hat{C}_{rnd} = k_{avg}/No.$	0.0135	0.0075	0.0081	0.017

NOTE: Number of nodes in the network with more than one coexpression partner (No.), and average number of coexpression partners of a vertex (k_{avg}). Values in parentheses for \hat{C}_{sf} reflect the SD after 10 simulations.

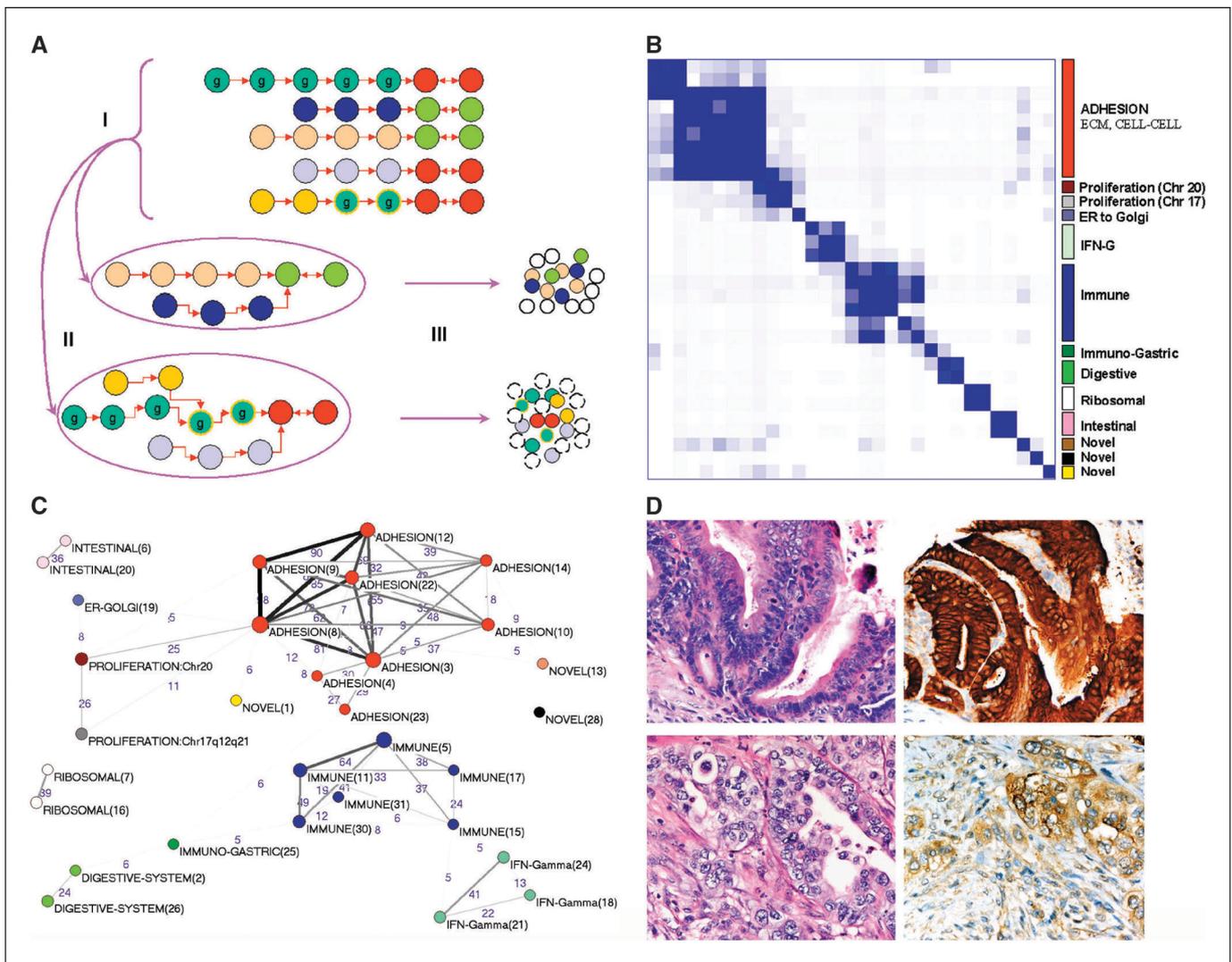


Figure 3. Functional modules in the gastrome. **A**, general schematic of the “chain-linking” methodology for organizing expression links into communities and subsequent modules. Each gene starts a chain that extends by connecting the highest linked neighbors (*unidirectional red arrows*, each circle represents a distinct gene). The linking process continues until a gene is maximally linked with the preceding gene (*bidirectional red arrows*), whereupon the chain is terminated (step 1). Multiple chains ending in the same chain terminator pair are combined (step 2), and genes within the assembly whose links exceed an LLR threshold (*white circles*) are combined to form a community. **B**, identification of modules from expression communities. A blue-white heat map depicts the similarity of each community to all other communities, with the depth of color reflecting the significance of overlap. Each community achieves the lowest probability, i.e., perfect overlap, with itself (*dark blue diagonal*). Each module is enriched for specific biological functions, shown on the right-hand side. **C**, higher-order relationships between communities and modules. *Circles*, community of various genes (list provided in Supplementary Material), and different colors represent distinct modules. Similarities between any two communities were computed as in **(B)**, and lines connecting various communities and modules correspond to cases in which the significance of overlap $\leq 10^{-5}$. *Thicker lines*, overlaps of higher significance (lower *P* values). **D**, *Villin1* expression in gastric adenocarcinomas. Two tumors are shown (*top and bottom*) both exhibiting glandular architecture corresponding with Lauren’s intestinal subtype: H&E staining (*left*); *Villin1* immunostaining (*right*). Intense *Villin1* expression is clearly noted in the tumor cells of the top tumor but not in the bottom tumor.

Validating our basic methodology, some of these modules were previously described in the original reports as coexpression clusters; including those associated with (a) proliferation (*PCNA*, *TOP2A*, *CDC2*, *CKS2*, and *MCM3*), (b) IFN activity (*ISGF3G*, *IFITM2*, *IFIT1*, *TRIM22*, and *SP110*), and (c) ribosomal function (*RPS28*, *RPL35*, *RPL31*, *RPL19*, and *RPS10*). However, in addition to these known modules, other modules were novel and not obviously discernible from the original single data sets (A. Aggarwal, data not shown), demonstrating the utility of the meta-approach in improving and refining the inherent data substructure of such genome-wide expression data. For example, module 1 is a novel module containing a highly significant number of nuclear and DNA-binding genes associated with the cell cycle (e.g., *GSTPI*,

CKS1B, *CKS2*, *PPP5C*, *P* values are $<10^{-5}$, $= 0.0045$, $<10^{-5}$ for nuclear localization, DNA-binding, and cell cycle activity, respectively, by the hypergeometric distribution), whereas module 28 contains several genes related to epidermal development (*LAMA3*, *EMPI*, *ANXA1*, *ITGB4*, and *SERPINB5*; see Supplementary Information S6 for a more in-depth analysis of these novel clusters). We believe that these novel modules are likely to be functionally meaningful, as their constituent member genes are both biologically coherent and robustly coexpressed and at high levels of specificity across all four independent microarray data sets. These examples show that the meta-analytic approach can indeed uncover novel biological modules that can be subsequently targeted for further experimentation.

Functional modules have highly distinct subtopologies consistent with their different biological functions. We then compared the individual connectivities of the different modules. In principle, the topologies of modules could belong to one of two contrasting scenarios. For example, a module might connect to other modules via several genes and expression links—in this case, parts of the former module would thus be integrated with and potentially regulatable by other modules. Alternatively, a module could be isolated, meaning that other modules would share very few to none of the former module's constituent genes and expression links—in this case, the formers' activity might be fairly autonomous and independent from other modules in the global network. To quantify the relative integration or isolation of the different modules, we defined a simple metric termed the "isolation index," a ratio of external (intermodule) to internal (intramodule) links, and computed this index for all modules (Table 2). We observed a considerable range in the isolation indexes of the different modules—for example, some modules like the ribosomal module were highly isolated (isolation index of 0) whereas other modules in comparison were relatively integrated (module 9, index 2.1). To ensure that these findings were not simply due to the spurious presence of a few key "outlier" genes, we generated permuted data sets in which 10% to 20% of the genes within each module were randomly removed and repeated the analysis, comparing the module rankings obtained from the modified data to the original complete data set. Although there were slight variations in the rankings of adjacent modules (i.e., modules in which highly similar indexes might sometimes switch places), the overall finding that certain modules seem to be integrated whereas others highly isolated was robust, supporting the idea that different functional modules can exhibit highly distinct topological structures (Supplementary Information S7). There is biological consistency to this finding, as shown in Fig. 3C, in which the interrelatedness of the modules is depicted. For example, modules associated with cellular proliferation were negatively linked and integrated with modules associated with cellular adhesion and

extracellular matrix production, which may reflect the requirement of a rapidly proliferating tumor to dissociate the extracellular matrix potentially inhibiting its growth. In contrast, the externally isolated but highly internally connected nature of the ribosomal module may provide a mechanism for tightly balancing the relative transcript levels of each ribosomal subunit, thereby ensuring that their protein products are produced in the appropriate stoichiometric balance for correct physical association and formation of a functional ribosomal complex. As we discuss later, such examples of "modular insularity" may emerge as a general design principle used by cells to (a) "shield" particular functional modules from excessive external regulation, and (b) to ensure that genes within that specific module will act as a concerted unit.

Intestinal metaplasia refers to the process where gastric epithelium begins to exhibit phenotypic traits reminiscent of intestinal epithelia. Frequently found in association with gastric cancer, intestinal metaplasia is regarded by some investigators as a potential preneoplastic cursor to gastric carcinogenesis (26). We observed that one module in the gastrome contained several coexpressed genes implicated in intestinal function and differentiation (*ANXA13*, *CDXI*, *GUCY2C*, and *VILL1*), suggesting that it might be related to the intestinal metaplasia process. Strikingly, this module exhibited an extremely high degree of topological isolation, being second only to the ribosomal module. The high insularity of this "intestinal module" raises the intriguing possibility that its specific topological properties may contribute to the frequent occurrence of intestinal metaplasia in gastric cancer. To confirm that the insularity of this module did not result from the inclusion of nonmalignant intestinal metaplasia tissues in the sample population, we repeated our analysis after deliberately removing all the noncancerous intestinal metaplasia samples and obtained similar results (Supplementary Information S8). As further evidence that this coexpression module is indeed present in tumor cells, we then examined the histologic expression of *VILL1*, a hallmark gene of this module, and found that it was clearly observed within the malignant carcinoma cells of certain tumors (Fig. 3D). Interestingly, there was

Table 2. Isolation indexes of functional modules (links out/links in) at LLR \geq 8

	Function	Unique genes in each functional unit	Number of genes that are shared with other functional units	Links out/links in
1	adhesion	119	25; 21%	0.27 (178/662)
2	immune reaction	78	9; 11.5%	0.21 (70/328)
3	proliferation Chr:20	24	17; 70.8%	1.71 (116/68)
4	ribosomal	23	0; 0%	0 (0/96)
5	proliferation Chr:17q12q21	19	11; 57.9%	1.23 (81/66)
6	IFN- γ	33	4; 12.1%	0.29 (34/116)
7	intestinal	20	2; 10%	0.07 (4/54)
8	digestive system	23	2; 8.7%	0.18 (9/50)
9	novel (13)	13	8; 61.5%	2.11 (59/28)
10	immunogastric	12	5; 41.7%	0.88 (21/24)
11	endoplasmic reticulum to Golgi transport	12	4; 33%	1.23 (32/26)
12	novel (28)	10	1; 10%	0.18 (5/28)
13	novel (1)	10	3; 30%	1.72 (31/18)
		344 unique		

NOTE: The number of unique genes in each functional module and genes that are shared between modules are also shown. Numbers in parentheses depict the absolute number of links used to calculate the isolation index.

little correlation between a tumor expressing this marker and belonging to the classic “intestinal” histologic subtype as defined by Lauren (Fig. 3D, bottom; also see Supplementary Information S9). Because the inherent autonomy of this intestinal module may prevent it from being strongly regulated by other modules or cellular pathways, its topological isolation may thus contribute to the process of intestinal metaplasia by rendering the intestinal differentiation process, once initiated, to be essentially irreversible. Such autonomous modules may also play an important role in normal development and tissue differentiation, as evidenced by a separate module containing genes associated with normal gastric physiology (*PGC*, *MSMB*, and *ALDH3A1*) also exhibiting a high isolation index. However, in cancer, these results raise the possibility that the inadvertent expression of such “self-sustaining” modules may play a significant role in the pathologic expression of cancer-specific phenotypic traits.

A gene neighborhood analysis of the gastrome reveals novel interactions between phospholipase *PLA2G2A* and the *EphB2* receptor. To gain insights into the component pathways affected in gastric cancer, we then performed an analysis of “gene neighborhoods,” referring to the set of closest coexpression partners associated with any particular gene. As an example of such “neighborhood analysis,” junction plakoglobin/ γ -catenin (*JUP*), which resides on chromosome 17q12q21, possesses a coexpression neighborhood of several other 17q12q21 genes within the gastrome (*TOP2A*, *GRB7*, *TRAP100*, and *PSMB3*), suggesting that a 17q12q21 coexpression amplicon is likely to be present in a certain proportion of tumors from all four patient populations. In this report, we focused on phospholipase II group 2A (*PLA2G2A*), a secreted phospholipase whose expression was recently found to be of prognostic significance in gastric cancer (15), but for which little is known concerning its actual molecular function in gastric tumors. Using neighborhood analysis, we found that the coexpression neighborhood of *PLA2G2A* in the gastrome contained the cell surface receptor *EphB2*, and β -catenin, a central member of the Wnt signaling pathway (see Supplementary Information S10 for *PLA2G2A* neighbors). To validate the association between *PLA2G2A* and *EphB2*, we analyzed a series of gastric tumor tissue microarrays and found that the histologic expression of *PLA2G2A* mRNA and *EphB2* protein was indeed significantly correlated ($P = 0.017$, χ^2 test; Fig. 4A; Supplementary Information S11). We also asked if the association between *PLA2G2A* and *EphB2* could have been discerned by conventional single-data set analysis rather than the meta-analytic approach by reexamining the former, and found that in the individual data sets, *EphB2* expression was only modestly correlated to *PLA2G2A* (Pearson’s correlation values of 0.4521, 0.3746, 0.4908, and 0.2129) with thousands of other gene pairs exhibiting higher correlation scores (Supplementary Information S10). In the absence of knowing that this moderate correlation is conserved across multiple data sets, it is quite unlikely that such an interaction would have been identified as significant in any of the original analyses. Indeed, it was stated in one of the original reports that “...the pattern of variation in *PLA2G2A* expression among (the) gastric cancer samples was not closely related to that of any other genes” (15).

EphB2 is a known target gene of the Wnt/ β -catenin signaling pathway (27). Thus, the significant coexpression interaction between *EphB2* and *PLA2G2A* expression made us consider that the Wnt signaling may also regulate *PLA2G2A* in human gastric cancer (notably, β -catenin is also present in the gene neighborhood of *PLA2G2A*). To explore the relationship between Wnt signaling

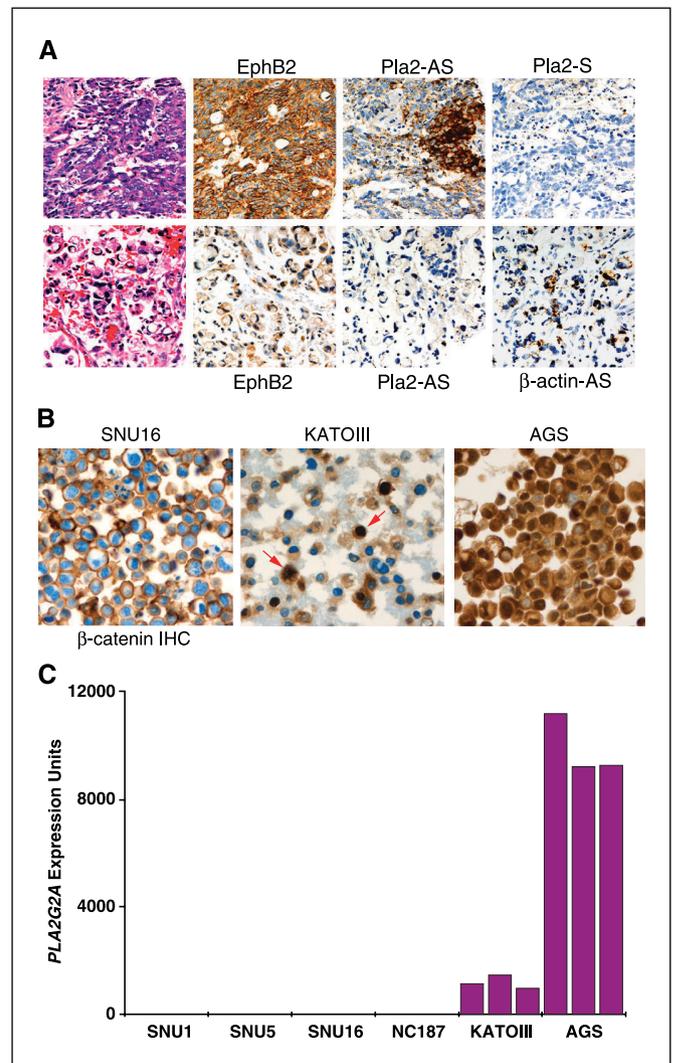


Figure 4. Expression interactions between *EphB2*, *PLA2G2A*, and β -catenin. **A**, a tissue microarray showing two gastric adenocarcinomas (top and bottom rows) examined by H&E staining (first column), *EphB2* expression (second column, deep brown membranous staining), and *PLA2G2A* expression (in situ hybridization, brown cytoplasmic granular staining). AS, antisense riboprobe; S, sense riboprobe. The top tumor coexpresses *EphB2* and *PLA2G2A*, whereas the bottom tumor is negative for both *EphB2* and *PLA2G2A* expression. A control hybridization for β -actin confirms the RNA integrity of the bottom tumor. **B**, immunohistochemical staining of β -catenin in gastric cancer cell lines. Aberrant nuclear expression of β -catenin is observed in cell lines KATOIII (red arrows) and AGS (right), but not in SNU16 (left). Similar nonnuclear staining was observed for cell lines SNU1, SNU5, and NCIN87 (SYL, data not shown). **C**, expression levels of *PLA2G2A* in gastric cancer cell lines. The graph depicts the expression level of *PLA2G2A* across the six cell lines (three replicates each) as determined by Affymetrix U133A arrays. Only KATOIII and AGS are seen to express *PLA2G2A*. Note that the level of *PLA2G2A* expression in KATOIII and AGS correlates with the extent of nuclear localized β -catenin in (B).

and *PLA2G2A*, we assayed a series of gastric cancer cell lines for both *PLA2G2A* expression and the presence of nuclear-localized β -catenin, the latter indicating high Wnt signaling activity. Out of six cell lines, two lines (KATOIII and AGS) exhibited nuclear staining of β -catenin, and remarkably, these were also the only two cell lines to express *PLA2G2A* (Fig. 4B and C). This result further supports an association between Wnt signaling and the expression of *PLA2G2A*. Finally, to ask if manipulation of the Wnt pathway could influence *PLA2G2A* expression, we then reanalyzed two public gene expression data sets—the first from *APC*^{-/-} mice in

which Wnt signaling is constitutively activated (28), and the second from a colon cell line in which Wnt signaling is inhibited by a dominant-negative *TCF4* transgene (ref. 29; dnTCF4). We found that in the former, *PLA2G2A* was transcriptionally up-regulated in *APC*($-/-$) intestinal epithelium compared with wild-type tissue (>7-fold), whereas in the latter, *PLA2G2A* was modestly suppressed (1.5- to 2-fold) in cells expressing dnTCF4 compared with cells in which the transgene was not induced (Supplementary Information S12). Taken collectively, these results are thus consistent with a model for *PLA2G2A* being a target of the Wnt signaling pathway. More work is required to address the functional aspects of *PLA2G2A* on Wnt pathway activity (see below).

Discussion

Gastric adenocarcinoma is a leading cause of global cancer mortality, surpassed only by lung cancer (10). The improvement of gastric cancer diagnosis and treatment will undoubtedly require developing molecular classification schemes that are both robust and clinically meaningful, and several reports have recently described gene expression profiles of gastric tumors for this purpose (11–14). However, as these studies were done using different technology platforms, clinical protocols, and patient populations, a critical challenge is to validate the robustness and generality of such initial discoveries in a larger series of patients and tissue samples. As this scenario pertains to essentially all cancers for which gene expression analysis has been reported, it is likely that some of the general approaches and methodologies we have developed in this report will be applicable to many other cancer types as well.

In order to identify gene coexpression relationships that were robustly conserved across the multiple data sets, we employed a rank-based statistical methodology conceptually similar to a previous study identifying evolutionarily conserved patterns of gene coexpression (22). This approach is distinct from a number of previously reported meta-analytic reports of cancer transcriptomes, as it does not rely on performing intergroup comparisons to define differential expression signatures from multiple data sets (7), and is focused on a single cancer type as opposed to all cancers (9). In our study, the success of the meta-analytic approach in identifying hitherto undetected subtle but significant gene coexpression relationships is likely due to both having larger sample sizes for additional statistical power and also to the use of multiple independent data sets to identify conserved associations. The former (increased sample size) increases the sensitivity of detection, whereas the latter (conserved behavior) increases specificity. For example, the expression of *EphB2* is only moderately correlated to *PLA2G2A* in the individual data sets (each containing different numbers of samples), but the preservation of this correlation across the four sets allows it to be identified above other interactions that display “strong” interactions in one data set but not in others (Supplementary Information S13). The network-driven approach also possesses a number of advantages compared with the conventional hierarchical clustering algorithms commonly used in such studies. In hierarchical clustering, the distance or similarity-based metrics used to cluster groups of genes (and samples) are usually based on preprocessed gene sets selected using arbitrary fold change or variance cutoffs. Such an approach offers no correction for potential false-positive relationships. In contrast, the network analyses used in this study combines data from multiple independent studies to perform a rigorous assessment of the FDR. Furthermore, as genes could have multiple

cellular functions, another advantage of the network approach is the ability to assign a gene to multiple clusters, unlike hierarchical clustering, in which genes are assigned to single individual clusters.

We note that our study possesses a number of limitations. First, by focusing on conserved gene behavior, our study necessarily identifies common aspects of gastric cancer that are preserved across the different patient populations and array studies. As such, we are unable to infer if there are population-specific differences in the molecular features of gastric cancer between the different countries. Second, our study is limited to the number of genes commonly found across the data sets (2,252), and consequently, does not provide information regarding other potentially important genes. Nevertheless, it should be possible to use the consensus framework defined in this study as a core molecular scaffold to which additional genes can be added as more information becomes available. Third, the gene expression data used in this analysis provides only a static snapshot of the collective gene expression interactions occurring within a complex tissue sample, and is thus likely to underrepresent both the dynamic nature of these interactions, and whether they arise in either single or multiple cell types within the tumor. Fourth, the network presented in our report is undoubtedly simplified as it employs binary relationships between genes and does not consider the “strength” of the relation. Nevertheless, it is apparent that even a “simplified” version of the real network can be useful in identifying novel biological associations for hypothesis generation and subsequent experimental testing, which provides a strong biological motivation for this form of research.

We analyzed the coexpression relationships within the gastrome at the levels of network topology and functional modules. Such higher-level analyses have proved useful in elucidating various systems-level properties of biological networks, such as robustness to perturbations (30, 31), module evolution (7), and oscillatory behavior (32). It is not unreasonable that a similar systems-level description of cancer transcriptomes might also provide insights into the complex biological properties of tumors, including independence from local growth control, derangement of cellular architecture, and the adoption of novel tissue phenotypes (e.g., metaplasia). We found that the global architecture of the gastrome was consistent with a hierarchical scale-free topology containing several deeply embedded modules, and that different modules seemed to possess distinct subtopologies—certain modules were relatively integrated with others, whereas other modules were relatively isolated. Although we cannot definitely rule out that some of our topological inferences may be a consequence of our specific methodology, we believe that they are unlikely to be complete artifacts, for the following reasons: (a) we focused on gene-gene relationships that are strongly conserved multiple independent data sets; such relationships are unlikely to be technical errors, (b) the subnetworks constructed in this study display strong biological coherence, as reflected by their individual constituent genes exhibiting similar cellular functions (e.g., intestinal and gastric differentiation, etc.), and (c) using different network construction methods, other groups have also reported similar topological findings, albeit in simpler organisms such as bacteria and yeast (22, 23). However, to our knowledge, this is the first report demonstrating the presence of topological differences between different functional modules. We propose that such differential “insularity” may reflect a general design principle used by cells to delegate specific functional roles to groups of genes, and to ensure their proper functioning by inhibiting or facilitating

intermodule cross-talk (33). Regarding intestinal metaplasia, we suggest that once the intestinal module is initiated, the highly internally connected nature of the module may cause the remainder of the module members to be similarly activated to manifest the complete intestinal differentiation phenotype. In addition, the “insularity” of this module may prevent this differentiation process from being regulated by other modules. The potential role of such systems-level features in contributing to the cancer phenotype needs to be further investigated.

We also studied the gastrome at the level of individual genes and pathways. Specifically, we analyzed the gene *PLA2G2A*, a secreted phospholipase identified in a previous study as a molecular prognostic marker for gastric cancer (15). Although previous work has implicated *PLA2G2A* activity in a wide variety of cellular functions, including prostaglandin biosynthesis, inflammatory responses, and antibacterial activity (34), the exact nature of *PLA2G2A*'s role in gastric cancer remains unclear. To identify potential cellular pathways affected by *PLA2G2A* in gastric cancer, we explored the coexpression neighborhood of *PLA2G2A* in the gastrome and found that it contained both β -catenin, a major component of the Wnt signaling pathway, and the *EphB2* receptor, a target of Wnt/ β -catenin signaling. These results, coupled with our additional supporting data that the Wnt pathway may regulate *PLA2G2A* expression, raises the intriguing hypothesis that one of *PLA2G2A*'s roles in gastric cancer may be to modulate the activity of the Wnt pathway. As our tumor expression profiles only provide a static snapshot of these interactions, it is not possible to definitely establish from our data if *PLA2G2A* might function as a positive or negative regulator of Wnt signaling. We note, however,

that (a) *PLA2G2A* was previously identified as a genetic modulator of colon cancer in the *APC^{min}* mouse model (35), and that (b) that high levels of *PLA2* have been shown to suppress colon cancer formation (36). This potential connection between *PLA2G2A* and the Wnt pathway definitely deserves further study.

In conclusion, we have described in this report a consensus molecular framework for gastric cancer, and shown how an analysis of this framework can deliver novel topological and functional insights. More broadly, our results show how meta-analytic approaches, by capitalizing on greater sample numbers and focusing on conserved gene behavior, could aid in identifying subtle but potentially important biological relationships relevant to tumor biology. With the rapidly increasing availability of such data sets in the public domain, it is likely that such meta-analytic approaches will play a valuable role in elucidating both general and tissue-specific molecular circuits in the oncogenome.

Acknowledgments

Received 6/27/2005; revised 9/14/2005; accepted 10/25/2005.

Grant support: Biomedical Research Council grant 01/01/31/19/209 and a National Medical Research Council core grant (P. Tan), grants-in-aid for scientific research (S) 16101006 from the Ministry of Education, Culture, Sports, Science and Technology (H. Aburatani), and The Research Grants Council of the Hong Kong Special Administrative Region (HKU7440/03M) (S.Y. Leung).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

H. Aburatani, D. Bowtell, S.Y. Leung, and P. Tan are founding members of the Gastric Cancer Genomics Consortium. We thank Hyun Cheol Chung and Sun Young Rha of Yonsei Cancer Centre, South Korea for organizing the Gastric Cancer Conference where the Gastric Cancer Genomics Consortium was first conceived.

References

- Jazaeri AA, Yee CJ, Sotiriou C, Brantley KR, Boyd J, Liu ET. Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. *J Natl Cancer Inst* 2002;94:990–1000.
- Kun Y, Lee CH, Tan PH, et al. A molecular signature of the Nottingham prognostic index in breast cancer. *Cancer Res* 2004;64:2962–8.
- van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
- Rosenwald A, Wright TR, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346:1937–47.
- Wong YF, Selvanayagam ZE, Wei N, et al. Expression genomics of cervical cancer: molecular classification and prediction of radiotherapy response by DNA microarray. *Clin Cancer Res* 2003;9:5486–92.
- Stegmaier K, Ross KN, Colavito SA, O'Malley S, Stockwell BR, Golub TR. Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat Genet* 2004;36:257–63.
- Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 2004;101:9309–14.
- Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004;36:1090–8.
- Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. *Genome Res* 2004;14:1085–94.
- World cancer report. World Health Organization; 2003.
- Boussioutas A, Li H, Liu J, et al. Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. *Cancer Res* 2003;63:2569–77.
- Chen X, Leung SY, Yuen ST, et al. Variation in gene expression patterns in human gastric cancers. *Mol Biol Cell* 2003;14:3208–15.
- Hippo Y, Taniguchi H, Tsutsumi S, et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res* 2002;62:233–40.
- Tay ST, Leong SH, Kun Y, et al. A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes. *Cancer Res* 2003;63:3309–16.
- Leung SY, Chen X, Chu KM, et al. Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc Natl Acad Sci U S A* 2002;99:16203–8.
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science* 2002;297:1551–5.
- Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999;286:509–12.
- Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4:844–7.
- Leung SY, Yuen ST, Chu KM, et al. Expression profiling identifies chemokine (C-C motif) ligand 18 as an independent prognostic indicator in gastric cancer. *Gastroenterology* 2004;127:457–69.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. *Nature* 2000;407:651–4.
- Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science* 2002;296:910–3.
- Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science* 2003;302:249–55.
- Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol* 2004;2:E9.
- Newman MEJ. Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 2005;46:323–51.
- Ravasz A, Barabasi AL. Hierarchical organization in complex networks. *Phys Rev E, Stat Nonlin Soft Matter Phys* 2003;67:026112.
- Yuasa Y. Control of gut differentiation and intestinal-type gastric carcinogenesis. *Nat Rev Cancer* 2003;3:592–600.
- Battle E, Henderson JT, Beghtel H, et al. β -Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 2002;111:251–63.
- Sansom OJ, Reed KR, Hayes AJ, et al. Loss of *Apc in vivo* immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* 2004;18:1385–90.
- van de Wetering M, Sancho E, Verweij C, et al. The β -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 2002;111:240–50.
- Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature* 2001;411:41–2.
- Kolisnychenko V, Plunkett G, Herring CD, et al. Engineering a reduced *Escherichia coli* genome. *Genome Res* 2002;12:640–7.
- Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000;403:335–8.
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999;402 Suppl:C47–52.
- Kudo I, Murakami M. Phospholipase A2 enzymes. Prostaglandins Other Lipid Mediat 2002;68–9:3–58.
- MacPhee M, Chepenik KP, Liddell RA, Nelson, KK, Siracusa LD, Buchberg AM. The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of *Apc(Min)*-induced intestinal neoplasia. *Cell* 1995;81:957–66.
- Cormier RT, Hong KH, Halberg RB, et al. Secretory phospholipase Pla2g2a confers resistance to intestinal tumorigenesis. *Nat Genet* 1997;17:88–91.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association
for Cancer Research

Topological and Functional Discovery in a Gene Coexpression Meta-Network of Gastric Cancer

Amit Aggarwal, Dong Li Guo, Yujin Hoshida, et al.

Cancer Res 2006;66:232-241.

Updated version Access the most recent version of this article at:
<http://cancerres.aacrjournals.org/content/66/1/232>

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2006/01/06/66.1.232.DC1>

Cited articles This article cites 35 articles, 16 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/66/1/232.full#ref-list-1>

Citing articles This article has been cited by 10 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/66/1/232.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.