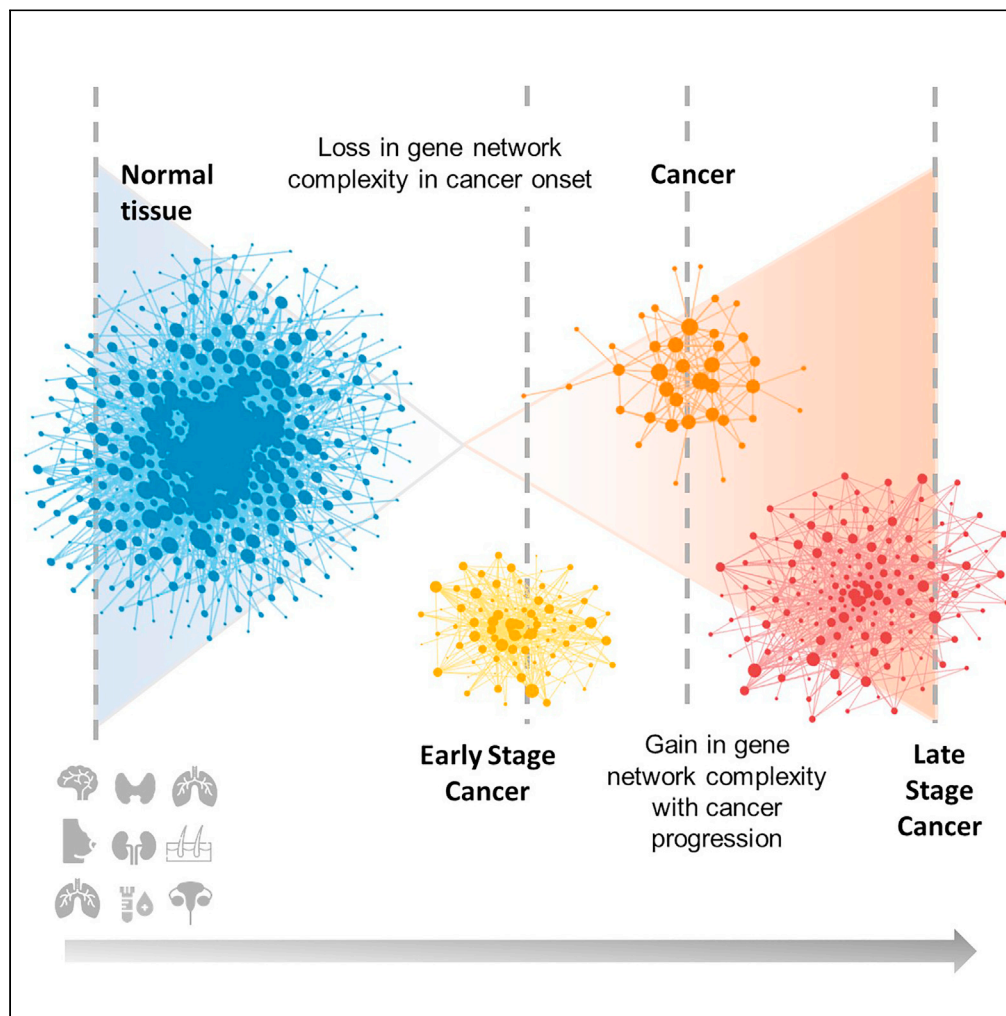


Article

Changes in gene-gene interactions associated with cancer onset and progression are largely independent of changes in gene expression

Zainab Arshad,
John F. McDonaldjohn.mcdonald@biology.
gatech.edu**Highlights**

Gene-gene network complexity is reduced in the transition from normal to cancer

Network similarity across cancer types is higher in early-stage versus late-stage cancers

Network interactions among housekeeping genes are stable through cancer development

<10% of changes in network interactions in cancer involve changes in gene expression

Arshad & McDonald, iScience
24, 103522
December 17, 2021 © 2021
The Authors.
<https://doi.org/10.1016/j.isci.2021.103522>

Article

Changes in gene-gene interactions associated with cancer onset and progression are largely independent of changes in gene expression

Zainab Arshad¹ and John F. McDonald^{1,2,*}

SUMMARY

Recent findings indicate that changes underlying cancer onset and progression are not only attributable to changes in DNA structure and expression of individual genes but to changes in interactions among these genes as well. We examined co-expression changes in gene-network structure occurring during the onset and progression of nine different cancer types. Network complexity is generally reduced in the transition from normal precursor tissues to corresponding primary tumors. Cross-tissue cancer network similarity generally increases in early-stage cancers followed by a subsequent loss in cross-tissue cancer similarity as tumors reacquire cancer-specific network complexity. Gene-gene connections remaining stable through cancer development are enriched for “housekeeping” gene functions, whereas newly acquired interactions are associated with established cancer-promoting functions. Surprisingly, >90% of changes in gene-gene network interactions in cancers are not associated with changes in the expression of network genes relative to normal precursor tissues.

INTRODUCTION

Cancer is a complex disease for multiple reasons. As with most diseases, there is an environmental and genetic component to cancer, and these two components may interact through a variety of epigenetic mechanisms (Herceg and Vaissière, 2011). Although, in some instances, the propensity for developing cancer is inherited in a Mendelian fashion (Goodrich, 2006), the disease is generally considered to be polygenic (Sugimura et al., 1992), albeit with a variety of specific genes (oncogenes and tumor suppressor genes) capable of “driving” onset and development in different cellular contexts (Bailey et al., 2018; Martínez-Jiménez et al., 2020). In more recent years, an added level of complexity has been recognized and attributed to the multiple levels of regulatory interactions that exist between genes comprising the human genome (Ashworth et al., 2011; Tutuncuoglu and Krogan, 2019). Thus, the changes that underlie cancer onset and progression may not only be credited to changes in the DNA structure and expression of individual genes but to changes in gene-gene interactions that emerging evidence (Billmann et al., 2018) indicates can contribute significantly to functions generally considered to be hallmarks of cancer (Hanahan and Weinberg, 2000, 2011).

A number of excellent studies have been carried out in recent years supporting the significance of co-expression network analysis in understanding cancer. These include analyzing changes in correlated expression patterns among genes under contrasting conditions, e.g., cancer versus normal precursor tissue (Hill and McDonald, 2015; Yu et al., 2017; Andonegui-Elguera et al., 2021) and cancers of different tissues of origin (Yang et al., 2014; Yu et al., 2017; Paci et al., 2021) and subtypes (Costa et al., 2018; Andonegui-Elguera et al., 2021). However, little is currently known concerning changes in gene-network structure associated with progression of the disease from early to late stages of development. We report here the results of a study of changes in gene-network structure occurring during the onset and progression of nine different cancer types. We consistently observe a drop in network complexity (reduction in number of network nodes and edges) in the transition from normal non-cancerous tissues to corresponding primary tumors. With the exception of thyroid cancer, cancer progression from early to later stages of development is marked by a renewed increase in complexity but involving gene-gene connections distinct from those observed in the transition from normal precursor cells. In addition, cross-cancer type (i.e., different tissues of origin) network similarity was found to increase during cancer onset, followed by a subsequent loss in similarity across different types of cancer as tumors progress.

¹Integrated Cancer Research Center, School of Biological Sciences, Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, 315 Ferst Drive, Atlanta, GA 30619, USA

²Lead contact

*Correspondence:

john.mcdonald@biology.gatech.edu

<https://doi.org/10.1016/j.isci.2021.103522>



Although changes in gene-network structure are often assumed to be associated with significant changes in gene expression, we find that this is frequently not the case. Indeed our results indicate that, on average, 90.5% of lost network nodes and 91.2% of acquired network nodes in cancer involve genes displaying no significant change in expression relative to their precursor normal tissues. Network features (nodes and edges) that remain stable through cancer development are enriched for “housekeeping” gene functions, whereas newly acquired networks are enriched for functions previously associated with cancer development, including regulation of inflammatory and immune response, as well as changes in extracellular matrix organization facilitating migration. Collectively, our findings indicate that changes in gene-network interactions associated with cancer onset and development may not only expand our understanding of the underlying molecular basis of the disease but may enhance our ability to identify important new candidates for targeted gene therapy.

RESULTS

Cancers display a significant loss in transcriptional network complexity relative to normal precursor tissues

Gene expression levels (RNA-seq) of tumor (early and late stages combined) and control tissue samples representing nine different cancer types (Glioblastoma multiforme, GMB; thyroid carcinoma, THCA; breast cancer, BRCA; lung adenocarcinoma, LUAD; lung squamous cell carcinoma, LUSC; skin cutaneous melanoma, SKCM; kidney renal clear cell carcinoma, KIRC; ovarian cancer, OV; acute myeloid leukemia, LAML) were downloaded from The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) database for the cancer samples and the Genotype-Tissue Expression (GTEx) (Ardlie et al., 2015) database for the control samples. A minimum of 100 tumor and 100 control samples were analyzed for each of the nine cancer types to help ensure the robustness of the derived networks (Table S1).

Co-expression networks for each dataset were constructed using pairwise Pearson’s correlations comparing the expression levels of 304.3 million pairs of genes ($17,444 \times 17,444$) across patient samples and selecting gene pairs displaying highly correlated (positive or negative, >0.85) changes in expression levels across samples. Each dataset was randomly down-sampled to a hundred observations, and edges conserved over all iterations were used to define the final networks (see [co-expression network construction](#) in STAR Methods). This resulted in 2,261 to 80,842 (Figure 1A) significantly correlated changes in expression (edges) between pairs of genes (nodes) across the nine normal tissue-of-origin samples (Table S2). The same analyses were carried out across each of the pre-cancer tissue types, resulting in 155 to 5,832 highly correlated changes in expression (edges) between pairs of genes (nodes) (Table S2). Because the vast majority of conserved network changes (96.93% of normal and 100% tumor connections) presented positive correlations between genes, positive and negative correlations were treated in an identical manner in our network analysis.

For all nine of the cancers examined, we observed a dramatic loss (average 96.7%) in network connections (edges) relative to those present in their respective normal precursor tissues (Figure 1A). On average, only 3.27% of the connections present in the precursor normal tissues were conserved in the cancers. The relative degree of loss in network connections varied across cancer types with LAML (99.93%) and OV (99.48%) displaying the most pronounced percent loss in connections and LUSC (96.2%) and THCA (83.2%) displaying the least (see Table S3).

While there was a large reduction in the number of connections in cancers that were present in precursor normal tissues (96.7%), there was a concomitant gain of new connections in the cancers (69.05%) that were not present in the normal precursor tissues. The increase of new network connections in cancer was again observed to vary across cancer types with SKCM (94.08%) and LAML (91.16%) displaying the most significant percent gain and OV (41.8%) displaying the least (see Table S3).

To quantify cross-tissue similarity among normal and among cancer networks, defined here as the proportion of network nodes common between two networks, we calculated pairwise network similarity for the nine tissues in normal and the nine tissues in disease state. Figure 1B displays heatmaps representing network similarity among sets of normal precursor tissues (left) and among cancer tissues (right). On average, 28.8% of genes (nodes) are in common among cancer networks, an overall increase in network similarity among cancer tissues relative to network similarity among their respective normal precursor tissues (average 19.1%).

Using Jaccard similarity between cancer networks for reference, we built a similarity network of cancers where each node corresponds to a distinct cancer type and each edge represents the number of genes

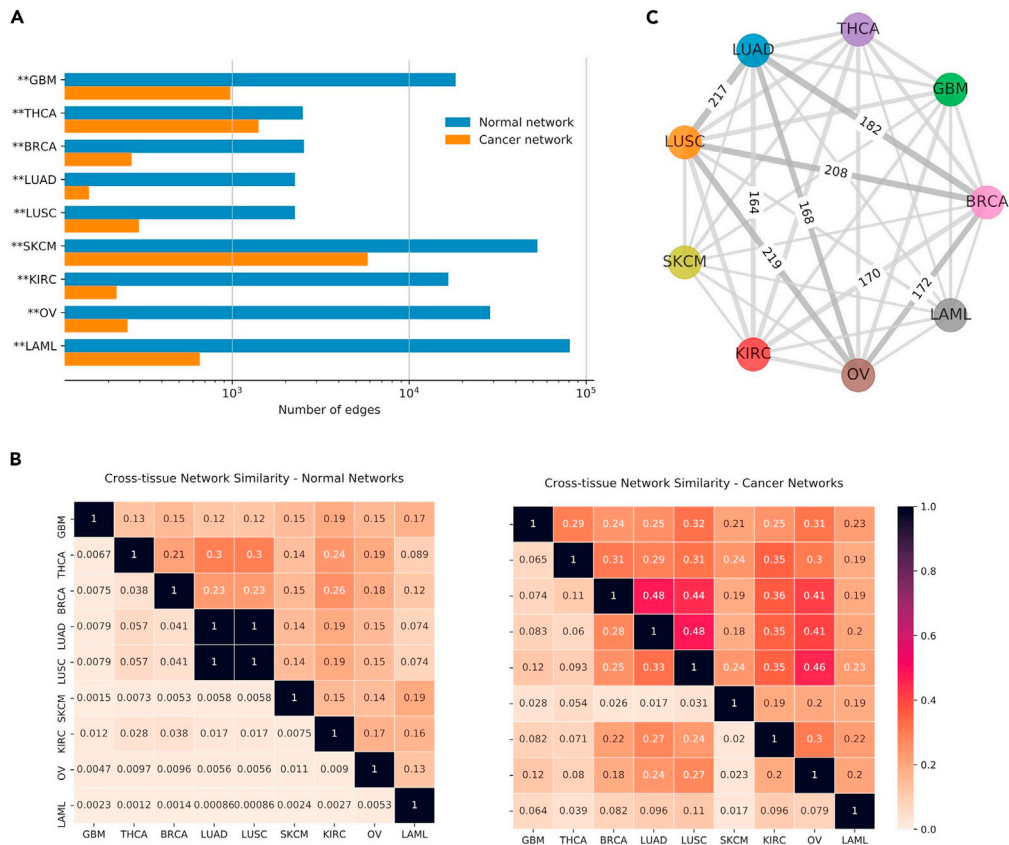


Figure 1. Network comparisons between nine cancer types

(A) Number of connections in gene co-expression networks for cancers and their respective normal precursor tissues. p Values were calculated based on t test for independent samples (GBM = 3.3×10^{-124} , THCA = 4.7×10^{-46} , BRCA = 3.3×10^{-91} , LUAD = 1.9×10^{-124} , LUSC = 3.4×10^{-98} , SKCM = 4.1×10^{-139} , KIRC = 2.0×10^{-141} , OV = 2.3×10^{-213} , LAML = 2.7×10^{-291} ; * = $p < 0.05$; ** = $p < 0.01$). Networks were constructed from conserved gene-gene correlations from 100 iterations.

(B) Cross-tissue Jaccard similarity in co-expression networks, for shared network nodes (upper triangular) and edges (lower triangular) in normal precursor networks (left) and their corresponding cancer networks (right). Numbers in the matrix are Jaccard Similarity indexes that are ratios with value between 0 and 1.

(C) Similarity network for cancer networks. Each node corresponds to a distinct cancer type, and each edge represents the number of genes shared between two cancer networks. Network pairs that share a substantial number of genes (Jaccard similarity more than 0.35) are labeled with the number of overlapping genes.

shared between two cancer networks (Figure 1C). Edges for cancer network pairs with Jaccard index greater than 0.35 (upper quantile) are labeled in the graph. Of interest, we found that four cancers (BRCA, LUAD, LUSC, and OV) displayed pairwise Jaccard similarity scores >0.4 and contributed most to the increase in network similarity in the cancers, relative to normal tissues. In addition to this, KIRC also displayed notable pairwise correlation with BRCA and LUAD, presenting 170 and 164 overlapping network nodes, respectively.

Changes in network complexity associated with cancers are largely attributable to lost or acquired network genes (nodes) displaying no significant change in expression

Comparing normal (Figure 2A) and cancer (Figure 2B) networks, connections are lost, conserved, or acquired in cancer (Figure 2C). Loss of connections (edges) in a cancer network may result from loss of one or both of the nodes (genes) in the cancer relative to normal precursor tissue or loss of a connection without loss of one or both nodes (genes) between cancer and normal precursor tissue (Figure 2D). Similarly, acquired connections in cancer may or may not be associated with the gain of acquired nodes (genes) between the tumor and normal precursor tissue (Figure 2E).

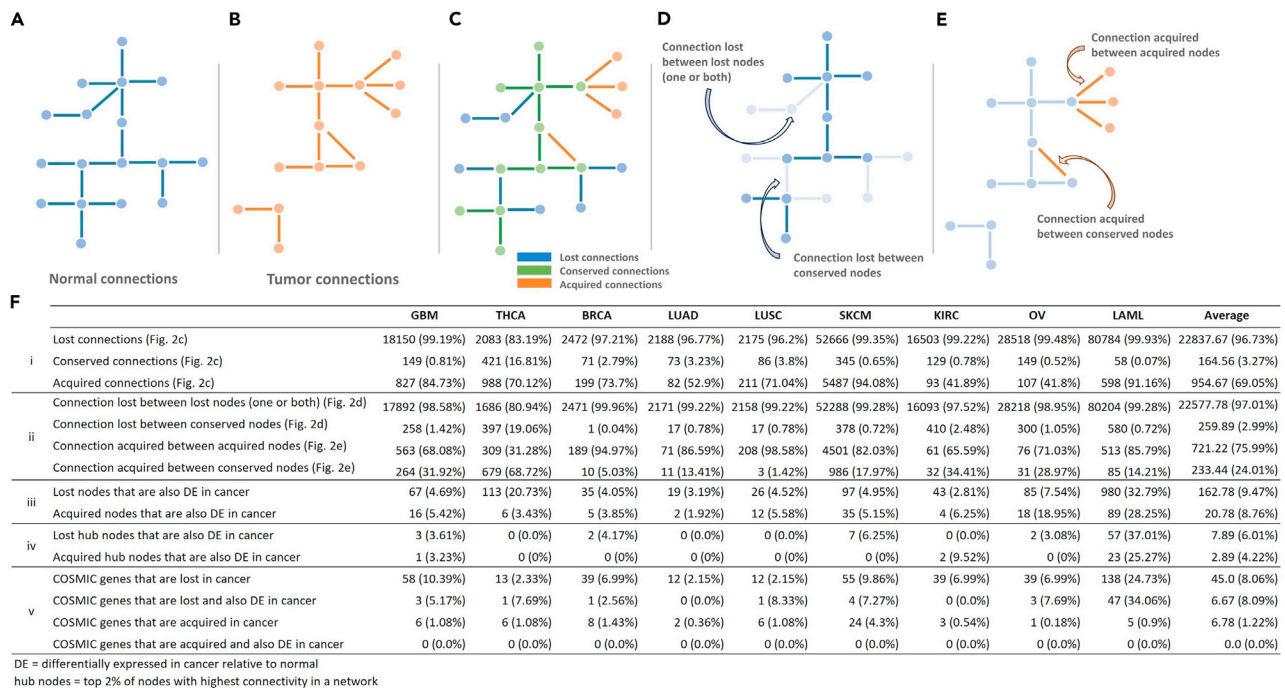


Figure 2. Changes in network structure (loss/gain of connections) in cancer

Lost connections (C) are edges present in the normal precursor tissue network (A) but not in the tumor network (B); conserved connections (C) are edges present in both the normal precursor network and the tumor network. Lost connections may arise with or without loss of associated nodes (D); acquired connections are edges NOT present in the normal precursor tissue network (A) connecting either existing nodes or newly acquired nodes (E).

(F i-v) Total number (percentage) of connections lost, conserved, or acquired in the tumor networks; (F ii) lost connections between conserved or lost nodes, and acquired connections between conserved or acquired nodes; (F iii) nodes lost or acquired and differentially expressed between normal and cancer tissues; (F iv) lost or acquired hub nodes (top 2% of nodes with highest connectivity in a network) that are differentially expressed; (F v) COSMIC (cancer driver genes) nodes lost or acquired, and differentially expressed between normal and cancer tissues.

We found that most changes in gene-gene (node to node) interactions observed in cancer networks resulted from the loss or gain of a node (Figure 2F i). On average, 97.01% of all lost connections had one or both of their nodes lost in cancer, whereas acquired cancer nodes accounted for 75.99% of all gained edges (Figure 2F ii, and lines 5 and 6 Table S4).

To determine if lost or acquired network nodes (genes) exhibit any significant expression changes between normal and cancer, differential gene expression levels were examined for all nine tissue types. Gene expression profiling identified 790 (GBM) to 5,919 (LAML) significantly differentially expressed genes (mRNAs) between normal and cancer samples (see Table S5). On average, only 9.47% of lost network nodes (Figure 2F iii) and 8.76% of acquired cancer nodes (Figure 2F iii) displayed a significant change in expression between cancer and normal (see also Table S6). LAML displayed the highest number of network nodes that were differentially expressed, including 32.79% lost nodes (10.5% up expressed and 22.3% down expressed) and 28.25% acquired nodes (8.25% up expressed and 20% down expressed) (f iii).

We defined the top 2% of nodes with the highest connectivity in our cancer networks as hub genes. Consistent with what we found for all genes lost or gained in cancer, only a small minority of network hub genes lost (6.01%) or acquired (4.22%) in cancer were found to display significant changes in expression in the cancers (Figure 2F iv; see also Table S7).

To determine if these trends remained consistent for those genes previously identified as drivers of cancer onset and progression, we focused on genes in the COSMIC (Catalogue of Somatic Mutations in Cancer) (Sondka et al., 2018) dataset. The COSMIC cancer gene set consists of cancer driver genes, including 315 oncogenes and 315 tumor suppressor genes. On average, only 8.09% of lost network COSMIC genes (nodes) and 0% of acquired network COSMIC genes (nodes) (Figure 2F v) displayed a significant change in expression between cancer and normal (see also Table S8).

Late-stage cancers display a net gain in network complexity relative to early-stage cancers

To explore the possibility that changes in network structure may also vary as cancers progress from early to late stages, we divided the TCGA cancer samples into early-stage (stage I and stage II) and late-stage (stage III and stage IV) samples. Datasets with less than 100 samples were removed from this analysis (SCKM and OV). GBM and LAML were also dropped owing to missing stage information. For the remaining five datasets, co-expression networks were constructed as employed above. As shown in [Figure 3A](#), four of the five cancer types tested displayed on average a 268.7% (BRCA = 69.47%, LUAD = 348.41%, LUSC = 896.97%, KIRC = 60.23%) increase in the number of gene-gene associations in late-stage cancers relative to early stage. THCA was again the outlier, displaying, in contrast, a decrease in the number of connections in late- versus early-stage samples. Cross-tissue network comparison revealed an average of 6.1% loss in Jaccard similarity coefficient for late-stage cancers relative to early-stage cancers ([Figure 3B](#)). Average Jaccard similarity for overlapping network nodes increased by 9.6 from 0.315 in normal tissue (194 overlapping nodes) to 0.401 in early-stage cancer (412 overlapping nodes) across the five tissues. Sixty-three of these cross-tissue conserved network nodes are shared between normal and early-stage networks.

Genes lost, conserved, or acquired in cancer networks are enriched for distinct biological functions

To determine if those genes (nodes) that were lost, conserved, or acquired in the cancer networks were differentially enriched for biological functions, we submitted the sets of genes lost, conserved, or acquired in cancer for pathway enrichment analysis to GSEAPy ([Subramanian et al., 2005](#)). This resulted in a total of 304, 96, and 142 hits for lost, conserved, or acquired network nodes, respectively (see [Table S9](#)). The most common significant GO hits (adjusted p value <0.05) across the nine cancers are presented in [Figure 4A](#). mRNA pre-processing, transcription, and translation were the biological processes most significantly enriched for lost nodes. Conserved nodes in cancer networks were predominantly enriched for metabolic and immune system processes, whereas acquired cancer nodes were enriched for extracellular matrix organization and immune-related processes.

GO analysis for early-stage and late-stage cancers revealed 77 pathways for lost nodes, 78 pathways for conserved nodes, and 113 pathways associated with network nodes acquired in late-stage cancer, a majority of which are unique to individual cancers. Similar to the normal versus cancer comparison, consensus hits for nodes lost in late stage showed enrichment for mRNA pre-processing, transcription, and translation ([Figure 4B](#)). Early-stage network nodes conserved in late stage were predominantly enriched for metabolic functions and immune response processes including multiple hits for antigen processing and presentation via major histocompatibility complex (MHC) molecules. Acquired nodes in BRCA displayed enrichment for 76 pathways (of the total 113 for all cancers) that include negative regulation of G2/M transition of mitotic cell cycle, regulation of stem cell differentiation, and regulation of signaling pathways for NIK/NF- κ B and Wnt. Genes shared across all five early-stage networks (412 in total) presented significant enrichment for eight biological processes associated with regulation (metabolic) and development. In contrast, network nodes common in their corresponding five precursor normal tissues did not show significant enrichment for any biological process (see [Table S10](#)).

DISCUSSION

Cancer cells typically display a loss in regulatory control over a variety of cellular functions relative to their normal precursor cells ([Hanahan and Weinberg, 2000](#)). On the molecular level, loss in regulatory control has traditionally been monitored by changes in gene expression ([Zhang et al., 1997](#); [Cadenas et al., 2014](#)). However, because genes operate within the context of complex and highly interactive molecular networks, a fuller appreciation of the regulatory changes that underlie cancer onset and progression must also include consideration of alterations in gene-gene network interactions. Toward this end, we constructed gene co-expression networks for nine human tissues and their derivative cancers. We employed an unsupervised approach of network construction to uncover global changes in network structure associated with cancer onset and progression. We found that cancer is generally associated with an overall reduction in network complexity with 96.73% of gene-gene connections (edges) present in normal tissues being lost in their corresponding cancer tissues. However, we also found that many new gene-gene connections are acquired in cancer accounting for, on average, 69.05% of network connections present in these transformed cells. Our results provide quantitative evidence for the dynamic

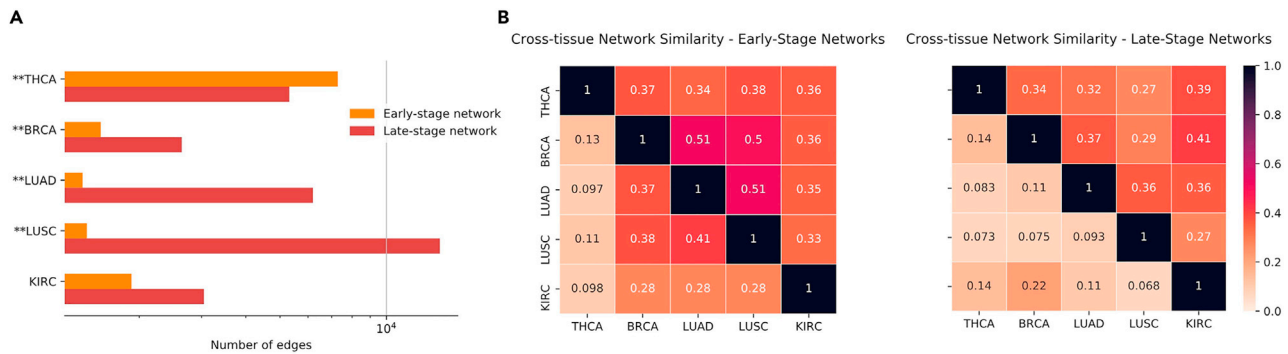


Figure 3. Increase in co-expressed genes in late-stage cancer

(A) For each tissue, the number of connections in networks representing early-stage (stages I and II) and late-stage (stages III and IV) samples. p Values were calculated based on t test for independent samples (THCA = 4.8×10^{-39} , BRCA = 8.4×10^{-14} , LUAD = 1.5×10^{-70} , LUSC = 6.0×10^{-9} , KIRC = 9.5×10^{-1} ; * = $p < 0.05$; ** = $p < 0.01$).

(B) Cross-tissue Jaccard similarity in co-expression networks, for shared networks nodes (upper triangular) and edges (lower triangular) in early-stage networks (left) and late-stage networks (right) across pairs of tissues.

rewiring of normal biological pathways in cancer (Chow, 2010; Billmann et al., 2018; Hjaltelin et al., 2019), as well as the acquisition of novel regulatory controls (Gupta and Qin, 2003; Balkwill and Coussens, 2004) previously associated with cancer. Moreover, in examining cross-tissue network similarity between precursor normal tissues and their respective cancers, we noted a 9.7% increase in shared network genes (nodes) (7.6% increase in shared connections [edges]) across the nine types of cancer in our study. These findings are consistent with the hypothesis that cancer cells revert to more undifferentiated states relative to their normal precursor cells (Carvalho, 2020) as evidenced, e.g., by elevated levels of progenitor (stem cell) markers and lower expression of differentiation markers (Friedmann-Morvinski et al., 2012). This hypothesis is further supported by our downstream analysis of early-stage networks, which presented twice as many network nodes conserved across early-stage cancers as their precursor normal tissues, with an overall 9.7% increase in cross-tissue network similarity.

Four cancers (BRCA, LUAD, LUSC, and OV) were found to contribute most to the increase in network similarity in cancer, with pairwise Jaccard indexes >0.4 each. These results provide confirmatory evidence for previously reported genetic correlations between breast, lung, and ovarian cancers (Jiang et al., 2019), along with previously unreported common genetic features of KIRC with BRCA, LUAD, and LUSC, as indicated by significant overlap of network nodes (Jaccard indexes >0.35).

Large sample sizes of five TCGA datasets allowed us to further explore network properties of early-stage and late-stage cancers. Although our results indicate reduced gene-gene interaction networks in cancers relative to their precursor normal tissues, they also reveal an increase in network complexity as cancers progress from early stage to late stage (268.7% increase). The significance of this increase in network complexity in later-staged cancers remains to be determined but may reflect a gain in regulatory control of cellular functions (e.g., metastasis, angiogenesis) characteristic of late-stage cancers. This possibility is consistent with our gene ontology analysis (Figure 4 and see below). Collectively, our results suggest that network complexity may complement current efforts to utilize stage-specific differences in gene expression in cancer staging (Sun et al., 2017; Yu et al., 2020). Furthermore, we found that late-stage cancers displayed 6.1% loss in shared network nodes relative to early stage, suggesting an increase in tissue-specific differentiation with cancer progression to late stage. The observation that increased tissue specificity is associated with cancer progression may be significant in view of previously reported divergence in mutational landscape of cancer evolution, from an initial set of common mutations in driver genes to rare mutations in a broad spectrum of genes at late stages of cancer development (Gerstung et al., 2020).

In contrast to the majority of cancers analyzed in our study, thyroid cancers were found to display a further drop in number of network connections in later stages of cancer progression. Although the reason for this dichotomy is currently unknown, the fact that patients with thyroid cancer in the TCGA database are less

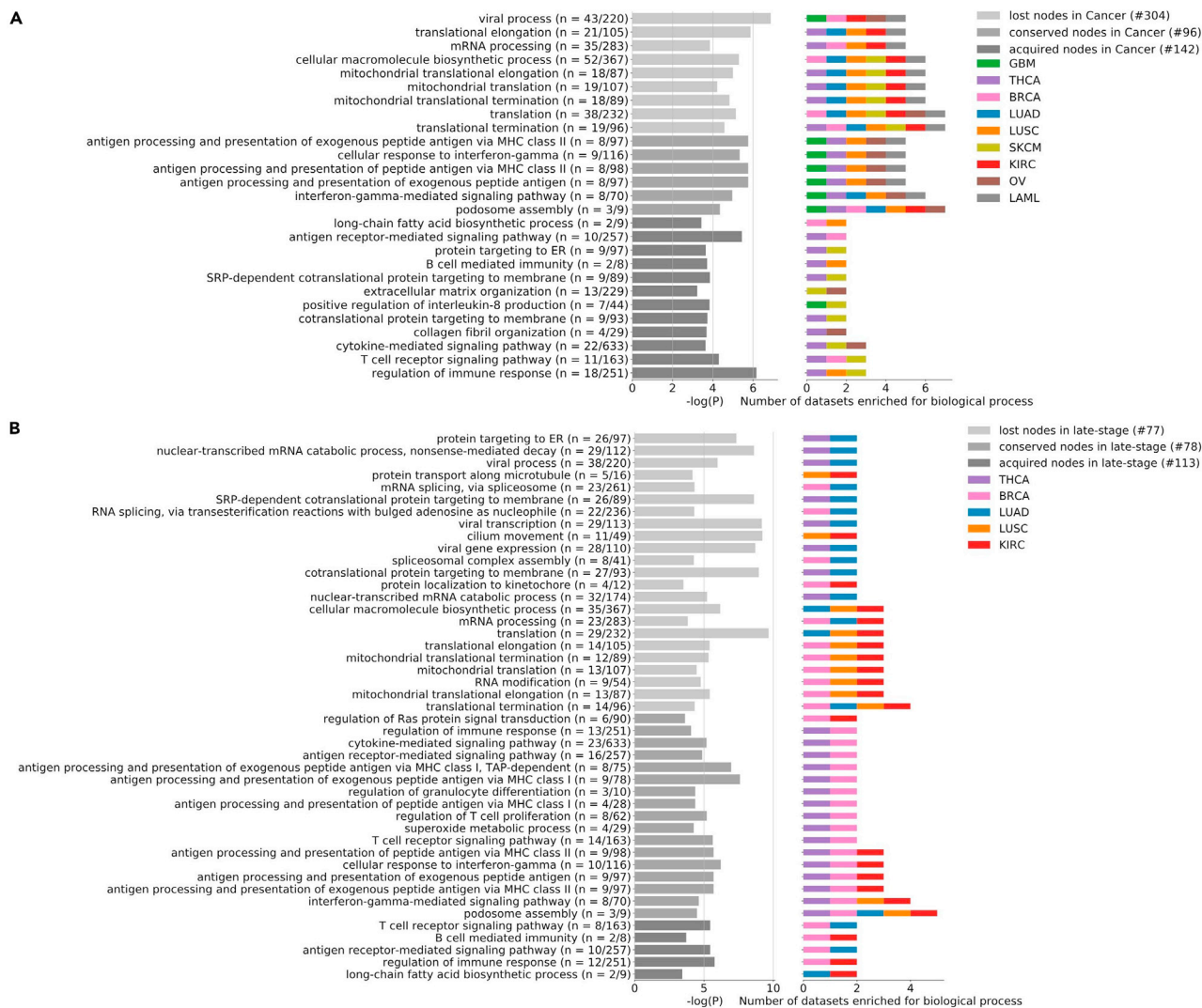


Figure 4. Pathway enrichment in network nodes lost, conserved, or acquired

(A) In cancers compared with their respective normal precursor tissues

(B) In late-stage cancers compared with their respective early-stage cancers. p Values for consensus GO hits (x axis) found to be enriched for lost, conserved, or acquired network nodes for cancers of different tissues of origin. (n = the number of overlapping genes [between network nodes and the annotated pathway] over total number of genes associated with the pathway).

likely to have received neo-adjuvant therapy than patients with other types of cancer (see [Figure S1](#)) suggests that therapeutic treatment *per se* may contribute to changes in gene-network interactions. Further studies will be required to rigorously test this hypothesis.

As noted above, changes in the levels of gene expression are currently widely employed to monitor putative regulatory changes associated with cancer onset and progression. Interesting, we found that genes displaying significant changes in gene-gene connections in the transition from normal to cancer cells were often not differentially expressed. DEGs were also found to be largely depleted in network hub nodes. In addition, we found nearly equal numbers of cancer driver genes (COSMIC database) to be associated with changes in gene-gene connections (52) and significant changes in gene expression (57); however, the two groups of genes were largely non-overlapping (only six in common). This observation highlights the combined power of gene-network analysis and differential expression analysis to uncover unique sets of disease-associated genes. Overall, our results suggest that changes in network structure in cancer onset/development represent a clinically significant dynamic that may go undetected in standard gene

expression analysis. Genes not differentially expressed in cancer but playing a critical role in newly acquired gene networks may constitute novel candidates for targeted gene therapy.

To explore the functional significance of changes in gene-network structure in cancers, we conducted detailed gene ontology (GO) analyses for those genes (nodes) lost, conserved, or acquired in cancer. Our results indicate that genes lost in cancer are enriched for functions associated with mRNA pre-processing and translational regulation. These findings are consistent with the view that cancer reflects an overall loss of regulatory control (Hartwell and Kastan, 1994; Wang et al., 2019). In contrast, genes acquired in cancer networks are enriched for functions associated with inflammation and extracellular matrix organization, two established hallmarks of cancer (Hanahan and Weinberg, 2000, 2011). Those genes conserved in normal and cancer networks are enriched for functions associated with metabolic activity, immune response, and other basic housekeeping cell functions.

Collectively, our findings reinforce the view that changes in gene-network interactions are playing a significant role in cancer onset and progression. We believe such system-level approaches to cancer may not only help unify and complement our understanding of the significance of the vast array of individual structural and regulatory variants associated with the disease but may also facilitate identification of potential new targets for chemotherapy.

Limitations of the study

There are limitations that need to be noted. For example, co-expression networks are sensitive to sample sizes with small datasets potentially giving rise to false positives. To minimize this possibility, we limited our networks to include conserved edges from multiple iterations of down-sampled equally (>100) sized datasets. In addition, it is important to keep in mind that, although correlated changes in the expression of genes associated with the onset and development of cancer are reliable system-level indicators of changes in network structure, this does not imply that each correlation is necessarily indicative of a direct (gene-to-gene) regulatory change on the molecular level. Indeed, it is likely that a number of the significantly correlated changes computationally identified between genes (represented as edges between nodes in the network) are the result of indirect regulatory interactions on the molecular level.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Data pre-processing
 - Co-expression network construction
 - Network similarity measures
 - Differential expression analysis
 - Gene ontology enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103522>.

ACKNOWLEDGMENTS

This research was supported by the Mark Light Integrated Cancer Research Center Student Fellowship, the Deborah Nash Endowment Fund, and the Ovarian Cancer Institute (Atlanta). The results shown here are based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

AUTHOR CONTRIBUTIONS

Z.A. and J.F.M. conceived the project; Z.A. conducted the computational analyses; Z.A. and J.F.M. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 19, 2021

Revised: November 7, 2021

Accepted: November 23, 2021

Published: December 17, 2021

REFERENCES

- Andonegui-Elguera, S.D., Zamora-Fuentes, J.M., Espinal-Enríquez, J., and Hernández-Lemus, E. (2021). Loss of long distance co-expression in lung cancer. *Front. Genet.* 12, 625741. <https://doi.org/10.3389/fgene.2021.625741>.
- Ardlie, K.G., Deluca, D.S., Segre, A.V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M., et al. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. <https://doi.org/10.1126/science.1262110>.
- Ashworth, A., Lord, C.J., and Reis-Filho, J.S. (2011). Genetic interactions in cancer progression and treatment. *Cell* 145, 30–38. <https://doi.org/10.1016/j.cell.2011.03.020>.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* 173, 371–385. <https://doi.org/10.1016/j.cell.2018.02.060>.
- Balkwill, F., and Coussens, L.M. (2004). Cancer: an inflammatory link. *Nature* 431, 405–406. <https://doi.org/10.1038/431405a>.
- Billmann, M., Chaudhary, V., ElMaghraby, M.F., Fischer, B., and Boutos, M. (2018). Widespread rewiring of genetic networks upon cancer signaling pathway activation. *Cell Syst.* 6, 52–64.e4. <https://doi.org/10.1016/j.cels.2017.10.015>.
- Cadenas, C., van de Sandt, L., Edlund, K., Lohr, M., Hellwig, B., Marchan, R., Schmidt, M., Rahmenführer, J., Oster, H., and Hengstler, J.G. (2014). Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell Cycle* 13, 3282–3291. <https://doi.org/10.4161/15384101.2014.954454>.
- Carvalho, J. (2020). Cell reversal from a differentiated to a stem-like state at cancer initiation. *Front. Oncol.* 10, 541. <https://doi.org/10.3389/fonc.2020.00541>.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., , Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.
- Chow, A.Y. (2010). Cell cycle control, oncogenes, tumor suppressors. *Nat. Educ.* 3, 7.
- Costa, R.L., Boroni, M., and Soares, M.A. (2018). Distinct co-expression networks using multi-omic data reveal novel interventional targets in HPV-positive and negative head-and-neck squamous cell cancer. *Sci. Rep.* 8, 1–13. <https://doi.org/10.1038/s41598-018-33498-5>.
- Friedmann-Morvinski, D., Bushong, E.A., Ke, E., Soda, Y., Marumoto, T., Singer, O., Ellisman, M.H., and Verma, I.M. (2012). Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice. *Science* 338, 1080–1084. <https://doi.org/10.1126/science.1226929>.
- Gerstung, M., Jolly, C., Leschchiner, I., Dentro, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J., Rubanove, Y., Anur, P., Yu, K., et al. (2020). The evolutionary history of 2,658 cancers. *Nature* 578, 122–128. <https://doi.org/10.1038/s41586-019-1907-7>.
- Goodrich, D.W. (2006). The retinoblastoma tumor-suppressor gene, the exception that proves the rule. *Oncogene* 25, 5233–5243. <https://doi.org/10.1038/sj.onc.1209616>.
- Gupta, M.K., and Qin, R.Y. (2003). Mechanism and its regulation of tumor-induced angiogenesis. *World J. Gastroenterol.* 9, 1144–1155. <https://doi.org/10.3748/wjg.v9.i6.1144>.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
- Hartwell, L.H., and Kastan, M.B. (1994). Cell cycle control and cancer. *Science* 266, 1821–1828. <https://doi.org/10.1126/science.7997877>.
- Herceg, Z., and Vaissière, T. (2011). Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 6, 804–819. <https://doi.org/10.4161/epi.6.7.16262>.
- Hill, C.G., and McDonald, J.F. (2015). Evidence and potential clinical significance of changes in gene network interactions in ovarian cancer. *J. Biomed. Eng. Inform.* 2, 1. <https://doi.org/10.5430/jbei.v2n1p1>.
- Hjaltelin, J.X., Izarzugaza, J.M.G., Jensen, L.J., Russo, F., Westergaard, D., and Brunak, S. (2019). Identification of hyper-rewired genomic stress non-oncogene addition genes across 15 cancer types. *NPJ Syst. Biol. Appl.* 5, 1–10. <https://doi.org/10.1038/s41540-019-0104-5>.
- Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Lesueur, C., Kuchenbaecker, K.B., Dennis, J., et al. (2019). Shared heritability and functional enrichment across six solid cancers. *Nat. Commun.* 10, 1–23. <https://doi.org/10.1038/s41467-018-08054-4>.
- Martínez-Jiménez, F., Muinos, F., Sentis, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., et al. (2020). A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* 20, 555–572. <https://doi.org/10.1038/s41568-020-0290-x>.
- Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X., Noushmehr, H., Colaprico, A., and Papaleo, E. (2019). New functionalities in the TCGAAbiolinks package for the study and integration of cancer data from GDC and GTEx. *Plos Comput. Biol.* 15, e1006701. <https://doi.org/10.1371/journal.pcbi.1006701>.
- Paci, P., Fisco, G., Conte, F., Wang, R.-S., Farina, L., and Loscalzo, J. (2021). Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. *NPJ Syst. Biol. Appl.* 7, 1–11. <https://doi.org/10.1038/s41540-020-00168-0>.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705. <https://doi.org/10.1038/s41568-018-0060-1>.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
- Sugimura, T., Terada, R., Yokota, J., Hirohashi, S., and Wakabayashi, K. (1992). Multiple genetic

alterations in human carcinogenesis. *Environ. Health Perspect.* 98, 5–12. <https://doi.org/10.1289/EHP.92985>.

Sun, Q., Zhao, H., Zhang, C., Hu, T., Wu, J., Lin, Z., Luo, D., Wang, C., Meng, L., Xi, L., et al. (2017). Gene co-expression network reveals shared modules predictive of stage and grade in serous ovarian cancers. *Oncotarget* 8, 42983–42996. <https://doi.org/10.18632/oncotarget.17785>.

Tutuncuoglu, B., and Krogan, N.J. (2019). Mapping genetic interactions in cancer: a road to rational combination therapies. *Gen. Med.* 11, 1–12. <https://doi.org/10.1186/s13073-019-0680-4>.

Wang, L.H., Wu, C.F., Rajasekaran, N., and Shin, Y.K. (2019). Loss of tumor suppressor gene function in human cancer: an overview. *Cell. Physiol. Biochem.* 51, 2647–2693. <https://doi.org/10.1159/000495956>.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Gen.* 45, 1113–1120. <https://doi.org/10.1038/ng.2764>.

Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat. Commun.* 5, 3231. <https://doi.org/10.1038/ncomms4231>.

Yu, W., Zhao, S., Wang, Y., Zhao, B.N., Zhao, W., and Zhou, X. (2017). Identification of cancer prognosis-associated functional modules using differential co-expression networks. *Oncotarget* 8, 112928–112941. <https://doi.org/10.18632/oncotarget.22878>.

Yu, X., Cao, S., Zhou, Y., Yu, Z., and Xu, Y. (2020). Co-expression based cancer staging and application. *Sci. Rep.* 10, 1–10. <https://doi.org/10.1038/s41598-020-67476-7>.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1997). Gene expression profiles in normal and cancer cells. *Science* 276, 1268–1272. <https://doi.org/10.1126/science.276.5316.1268>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA GTEx data	Mounir et al., 2019. https://doi.org/10.1371/journal.pcbi.1006701 .	https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html
COSMIC	Sondka et al., 2018. https://doi.org/10.1038/s41568-018-0060-1 .	https://cancer.sanger.ac.uk/cosmic
Software and algorithms		
Python	Python Software Foundation	https://www.python.org/
TCGAbiolinks	Mounir et al., 2019. https://doi.org/10.1371/journal.pcbi.1006701 .	https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html
GSEApY	Subramanian et al., 2005. https://doi.org/10.1073/pnas.0506580102 .	https://github.com/zqfang/gseapy

RESOURCE AVAILABILITY

Lead contact

Request for further information should be directed to and will be fulfilled by the lead contact, John F. McDonald (john.mcdonald@biology.gatech.edu)

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data: This paper analyzes existing, publicly available data. Details of these datasets are listed in the [key resources table](#).

Code: This paper does not report original code.

Additional information requests: Any additional information required to reanalyse the data used in this study is available from the lead contact upon request.

METHOD DETAILS

Data pre-processing

Nine cancer types, including carcinoma, melanoma, and leukemia, were selected from The Cancer Genome Atlas (TCGA) (Cerami et al., 2012; Weinstein et al., 2013) Glioblastoma multiforme-GMB, thyroid carcinoma-THCA, breast-BRCA, lung adenocarcinoma-LUAD, lung squamous cell carcinoma-LUSC, skin cutaneous melanoma-SKCM, kidney renal clear cell carcinoma-KIRC, ovarian-OV, acute myeloid leukemia-LAML. Healthy tissue data for each tissue type were obtained from the Genotype-Tissue Expression (GTEx) (Ardlie et al., 2015) database. Batch-corrected data from the two projects was downloaded together with Recounts2 project (Mounir et al., 2019) The RNA-seq counts values were normalized and processed with TCGAbiolinks package (Mounir et al., 2019). A summary of sample sizes for each dataset can be found in [Table S1](#).

For stage specific cancer analysis, we required that each cancer type, with stage information available, had at least one hundred samples for early-stage and late-stage each, which left us with only five cancer types: BRCA, THCA, LUAD, LUSC and KIRC.

To investigate trends in co-expressed gene pairs for cancer driver genes, 315 oncogenes and 315 tumor suppressor genes were obtained from Cancer Gene Census, COSMIC (Sondka et al., 2018). Out of these, 72 genes are labeled both oncogene and tumor suppressor gene in COSMIC.

Co-expression network construction

Pearson correlations between gene pairs were calculated using `pandas.DataFrame.corr` function. Based on our previous work where we evaluated different values of correlation thresholds (0.70–0.99), it was found that networks of random signals could appear to be connected for values of $r < 0.85$ (Hill and McDonald, 2015). Thus, to minimize false positives, the absolute value of r was limited to values > 0.85 . Baseline relationships (correlations) between pairs of genes were established by correlating the expression of 304.3 million pairs of genes ($17,444 \times 17,444$) across normal samples for each tissue, and selecting pairs satisfying $r > 0.85$, p value < 0.05 . Because on average 96.93% of baseline relationships (correlations) observed in GTEx normal tissue datasets and 100% correlations in TCGA cancer datasets, all relationships (correlations) were treated as unsigned in all downstream analyses.

The number of edges was found to be highly dependent on the number of samples. Therefore, to ensure fair comparison between tissues with variable sample sizes, each dataset was randomly down-sampled to a 100 and consensus interactions from a hundred iterations were used to define the final networks. For downstream analysis, network nodes were divided into: lost nodes representing genes displaying significant correlations across normal samples only, conserved nodes with connections in both normal and cancer samples and acquired nodes that develop gene-gene correlations in cancer samples.

Network similarity measures

To evaluate network conservation among precursor normal samples and among cancer samples, we calculated a Jaccard Index using `sklearn.metrics.jaccard_score`, for network nodes and edges shared between two networks as shown in Figure 1B. The analysis was repeated for different stages of cancer to obtain results in Figure 3B.

Differential expression analysis

DEGs are identified using `TCGAanalyze_DEA` function from TCGAAbiolinks that utilizes the edgeR package (Robinson et al., 2009). The thresholds of p value < 0.05 and $|\log_2 \text{FC (fold change)}| \geq 1$ were set to define DEGs (Figure S2). DEGs were further divided into up-regulated DEGs with $\log_2 \text{FC} \geq 1$ and down-regulated DEGs with $\log_2 \text{FC} \leq -1$.

Gene ontology enrichment analysis

Lists of genes lost, conserved or acquired for each cancer were submitted to genome ontology (GO) enrichment analysis by using the Gene Set Enrichment Analysis in Python (GSEAPy version 0.9.12; <https://github.com/zqfang/gseapy>) (Subramanian et al., 2005). Biological process terms having adjusted $P < .05$ were considered significant. GO hits for all nine cancers were consolidated to determine the mostly highly conserved biological processes across the nine cancers, for lost, conserved or acquired network nodes. GO parent terms with most enriched descendants for each node set were also mentioned in Figure 4A (see Table S9). The analysis was repeated for different stages of cancer in Figure 4B.

QUANTIFICATION AND STATISTICAL ANALYSIS

A summary of sample sizes for each dataset used in this work can be found in Table S1. For gene co-expression analysis, Pearson coefficient (r) > 0.85 and the corresponding p values < 0.05 were used. Table S2 lists the number of significantly correlated changes in expression between pairs of genes (network edges) in normal and cancer networks. p values for the loss in number of interactions between normal and cancer samples, per tissue, were calculated based on T test for independent samples and are provided within the Figure 1A legend. Jaccard index for network similarity was calculated with `sklearn.metrics.jaccard_score`, for network nodes and edges shared across tissues in normal and cancer samples, shown in Figure 1B. For differential expression of genes analysis, we used the `TCGAanalyze_DEA` function from TCGAAbiolinks that utilizes the edgeR package. Thresholds of p value < 0.05 and $|\log_2 \text{FC (fold change)}| \geq 1$ were set to define DEGs, that have been provided in Table S5 for each tissue type. p values for the difference in number of interactions between early-stage and late-stage samples, per tissue, were calculated based on T test for independent samples and are provided within the Figure 3A legend. Significant GO hits for networks nodes were identified using GSEAPy, p values < 0.05 . Hits enriched in more than one cancer are plotted in Figures 4A and 4B and the complete lists of results are included in Table S9. Statistical test results (*i.e.*, p values) are provided for all comparisons in figures, and p values < 0.05 were considered significant.