Technical notes provide unique applications, innovative methods, and clear protocols designed specifically for Stratagene reagents, instruments, and software.

Technical Note

The PathwayArchitect® Software can be used to efficiently and systematically study a large number of biological molecules obtained from either genomic and/or proteomic experiments and builds biological pathways and networks. This Technical Note will demonstrate how PathwayArchitect Software can predict the inter-connectivity of a large list of differentially expressed genes, predict their roles within a large breast cancer network, and predict potential drug molecules and targets for breast cancer.

Background

Breast cancer accounts for approximately one-fourth of all cancers in women and is second only to cervical cancer as the leading cause of cancer deaths in women.[1] Although Caucasian women and women from higher socio-economic background are the highest risk groups, more recently, incidence of breast cancer is rising amongst urban women in other countries as well. Higher susceptibility to breast cancer has been attributed to both genetic and socio-economic causes, including food and lifestyles. Chances of getting breast cancer increase with a woman's age, family history of breast cancer, early commencement of menstruation, late menopause, late age at the time of first full-term pregnancy, and even obesity in post menopausal women (www.breastcancer.org).

The treatment of breast cancer may require, surgery, radiotherapy, hormonal therapy, and chemotherapy.[2] The prognosis and choice of the therapy for a breast cancer patient depends on many tumor biomarkers including the presence of estrogen and progesterone receptors. Estrogen and progesterone induce cellular growth and proliferation through binding their receptors present on the surface of tumor cells. Estrogen receptor positive tumors have a lower propensity for recurrence and metastasis and respond better to hormonal therapy than receptor negative tumors. Also, chemotherapy which is effective in case of active cellular proliferation, works better in hormone receptor positive tumors in comparison to the receptor negative ones.[3,4]

Tumor suppressor genes, BRCA1 and BRCA2 have been implicated to function in DNA repair have been associated with breast cancer.[5] Increased risk to breast cancer susceptibility is associated with mutations in these genes. Recent advances in microarray technology have helped monitor the expression of thousands of additional genes within breast cancer tissues simultaneously and classify a set of signature genes with distinct expression changes that mark the onset, development and prognosis of breast cancer. Furthermore, several signal transduction cascades that turn on transcription of genes associated with cell cycle regulation, cell proliferation and apoptosis have been characterized for malfunction in breast cancer tissues compared to normal tissues. However, a complete network of the biological interactions that clearly explains the inter-connectivity amongst all the genes with altered expression in breast cancer tissues cannot be built based on the currently available tools. This would entail creating a large network that would link together various well-characterized signaling pathways operating in breast cancer, e.g. TGF-beta, MAP kinase, apoptosis, less characterized pathways such as iron metabolism, as well as genes that have not yet been classified in common signaling pathways.

Most drugs against cancer are designed to break and damage DNA of tumor cells and reduce cell division by blocking replication of DNA and the mitotic phase of the cell cycle. Thus, drugs in the market that are mainly used for chemotherapy are not specific, there-by leading to common side effects. Along with cancer cells, they also attack the rapid turnover of healthy cells like skin, hair, and bone-marrow. A single drug cure for cancer remains challenging because there is usually not a single type of cancer and also because the biochemical differences between normal and healthy cells are small. The vision for post-genomic age discovery of cancer drugs is to develop agents that will be directed towards specific molecular abnormalities of the cell and help develop personalized drug combinations targeting specific tumors. However, this requires a deeper understanding of the entire network of biochemical reactions within a cancer cell and how perturbations created by any drug in a known sub-section of the network will affect other connected molecules.

Experimental Approach

In recent times, genome-wide gene expression changes between cancer and healthy tissues can be easily detected using microarray experiments. Analysis of microarray data can provide comprehensive sets of differentially expressed genes with altered gene expression in cancerous tissues. But how these genes interact with each other and connect to form various signal transduction networks within the biological tissues, is relatively less explored.

A recent approach to solve this problem at a genome-wide scale comes from deploying a natural language processing based algorithm (NLP) to search MEDLINE publication abstracts, that are by far the best and the most extensive repository of empirical experimental data. Using this tool available within the PathwayArchitect® Software, we have analyzed microarray data from breast cancer versus healthy tissue samples and created networks of signaling pathways. Further, using statistical methods available within the software, we have predicted additional proteins and chemical molecules that fit into the breast cancer network. We found several experimentally validated proteins and small molecules with known biological importance to breast cancer in our predicted results. Thus, the PathwayArchitect software provides an invaluable tool in analysis of large data sets of biological molecules. This would expand our current understanding of signaling pathways in cancerous tissues and help predict biologically important molecules, some of which may be potential drugs or drug targets.

Data Analysis

In order to identify a set of genes with expression changes in breast cancer tissues, we re-analyzed previously published microarray data comparing breast cancer tissue samples from patients or short-term epithelial breast cancer cell cultures, with Stratagene Universal Human Reference RNA. Expression data were obtained from spotted 2-color dye cDNA microarray experimental results available under accession number GSE 4000 in Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/projects/geo/) and analyzed using ArrayAssist® Expression Software.

Following background correction and loess normalization of all samples, fold expression change for every gene was calculated in each cancer sample against the corresponding reference sample. Expression ratios from replicate cancer tissues, cell lines or normal samples, were grouped separately and one-way ANOVA tests performed between normal and the two breast cancer groups separately. The ANOVA tests yielded genes with (*p*-value) significant expression differences between the normal and cancer groups using a significance (*p*-value) cut-off of 0.001. Thus we defined our selection of up- or down- regulated genes in breast cancer samples.

We selected 914 genes that were either up- or down- regulated in both patient cancer tissue samples and primary cancer cell line cultures compared to a normal tissue sample. These genes were selected for network building using PathwayArchitect v.1.1 software. At the core of the PathwayArchitect software is a large database of proteins, small molecules, enzymes, complexes, processes and functions, and 1.2 million described biological interactions amongst them derived from scientific literature. A query list of up- or down- regulated genes in breast cancer samples were searched against the database and all possible known interactions between these entities were found. In the software, each interaction is given a quality "score" (Max, High, Medium, Low etc.) based on NLP derived criteria that depend on the number of references found associated with the interaction as well as the syntax and semantics of those reference sentences (see Technical Note "Natural Language Processing in PathwayArchitect software" for more details at www.stratagene.com/SoftwareSolutions). The software allowed us to set filters to select only interactions of a certain score, as per our requirements.

Exploring Interactions Between Well Characterized Pathways

We imported a list consisting of 914 differentially expressed genes, along with all property columns, into the PathwayArchitect software using the "Import Data table" function, under Data Import in the Workflow. All 914 entities found matched IDs in the PathwayArchitect database. We selected all the 914 entities and chose "Direct Interactions Network" under the Interactions Network in the Workflow. The "Direct Interactions Network" function only displays interactions of the maximum quality score. We found that among 914 genes, a total of 336 genes participated in 972 "direct" interactions. In this network, we did not include interactions of lower quality scores at this point. The list of genes with altered transcript levels included several known cell-proliferation and DNA repair related genes such as *BRCA1*, *TGF-beta*, *MYC*, *EGF*, *ERBB*, *CDK*, and many others.

PathwayArchitect software can further assist the analysis of a network of genes by overlaying numerical data (in this case, the microarray expression data) onto the network. Using the "Data Overlay" function under "Numerical Data Analysis" in the Workflow, we have colored every node by "fold change in gene expression" between cancer and normal samples. Up-regulated genes are colored in shades of red and down-regulated genes are in shades of green, as indicated by the scale. Thus, any data column from the microarray experiment can be overlaid onto the selected nodes, thereby showing not only biological interactions that have been described in scientific literature, but also the numerical results from a microarray study. In Figure 1, we have used the navigator function to show a small subsection of the entire breast cancer Direct Interactions Network, which includes several well characterized cancer related genes.

So far, in breast cancer, the biological roles of many of these well characterized genes or a subset of well characterized genes have been studied, but in isolation. However, a comprehensive network connecting all genes/proteins involved in breast cancer has not been possible without a tool that can simultaneously process interaction information for hundreds of known genes. Using PathwayArchitect software, we have simultaneously explored the available interaction information among all the genes that showed differential expression in cancer tissue. We have seen that the large network includes several well characterized genes which interact with one-another either directly, or through other entities having lesser known roles in breast cancer. This has increased the size of the network associated with breast cancer by bringing in new candidates and finding new interactions between known players.

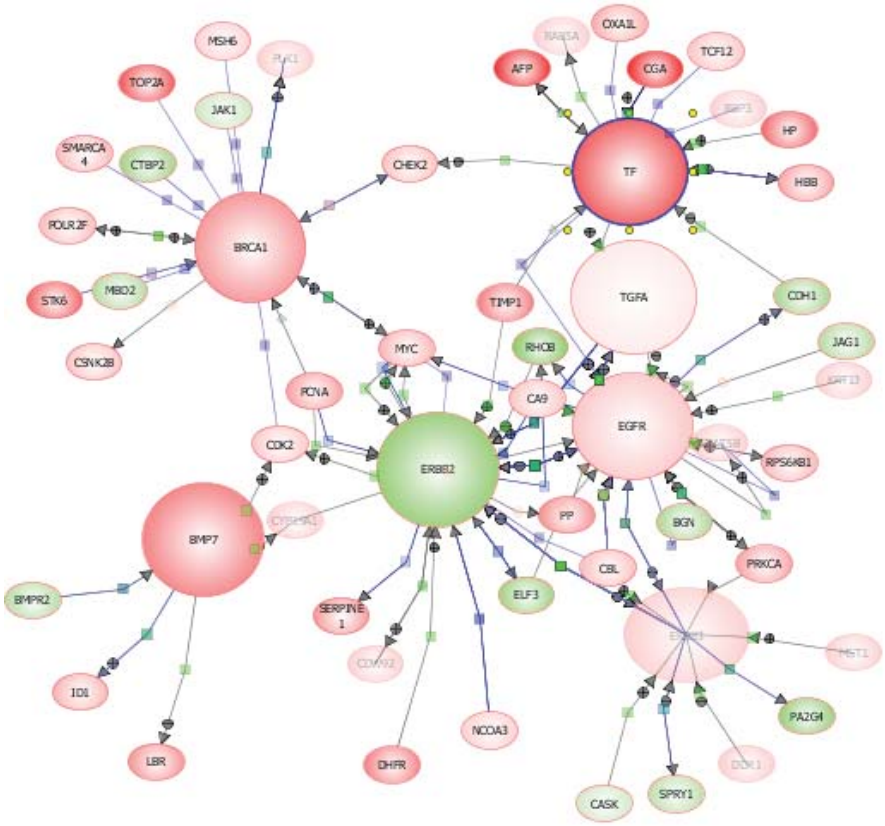

Figure 1
A Sub-Section of the Direct Interactions Network for Breast Cancer

For detailed analysis in this study, we have concentrated on a handful of genes. We chose the DNA repair protein BRCA1, a classical marker for breast cancer as a starting point of our study.[6] Using the "Navigator" function, we focused only on BRCA1 and its neighbors, in a separate viewer. In the direct interaction network, as shown in Figure 2, we found BRCA1 connected to other proteins and pathways through 14 other entities including CDK2, JAK1 and MYC, whose roles have been well elucidated with relevance to breast cancer.[7]

Similarly, the role of TGF-beta signaling pathway has been studied in great detail in various cancer tissues including breast cancer. But how do BRCA1 and TGF-beta signaling, two critical components of breast cancer network, connect to each other? We found that our list of differentially expressed genes in cancer includes BMP7, a known ligand of the TGF-beta signaling pathway. BMP7 binds to BMP receptor and triggers on TGF-beta signaling pathway. Similarly, using the "Navigator", we explored BMP7 and its neighbors, and investigated if there are common neighbors between BRCA1 and BMP7. As shown in Figure 3, in this network, both BRCA1 and BMP7 share a common neighbor, CDK2. From the pathway exploration, we see that BMP7 inhibits cyclin dependent kinase CDK2 which is a cell cycle regulator. Thus inhibiting cell cycle regulation probably triggers unchecked cell division in the cancer cells, leading to their proliferation. CDK2 negatively regulates BRCA1 by binding and disintegrating its ubiquitination complex with BARD1, which stabilizes nucleophosmin (NPM) protein association with the centrosome. Although the mechanism of sporadic breast cancer formation is not well understood, it is possible that CDK2-BRCA1-NPM pathway coordinately regulate centrosome duplication in the S-phase of cell cycle.[8]

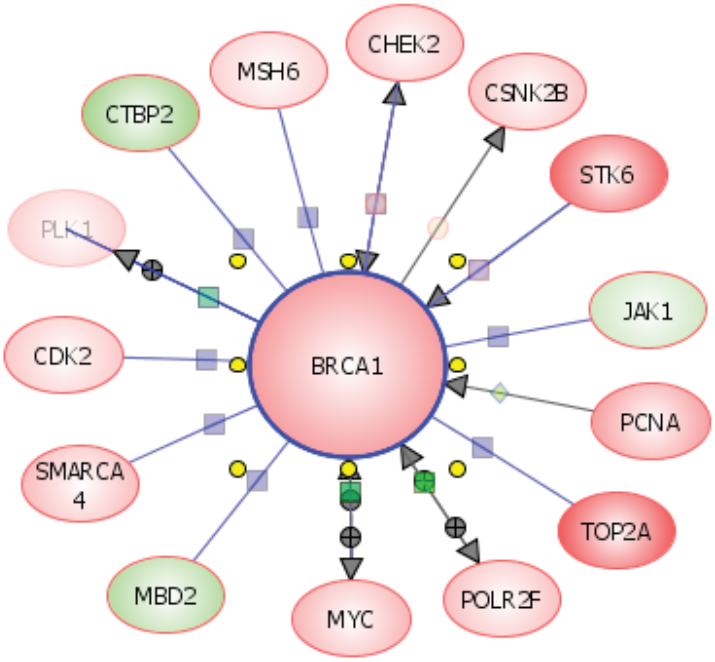

Figure 2
Neighbors of BRCA1 in the Direct Interactions Network

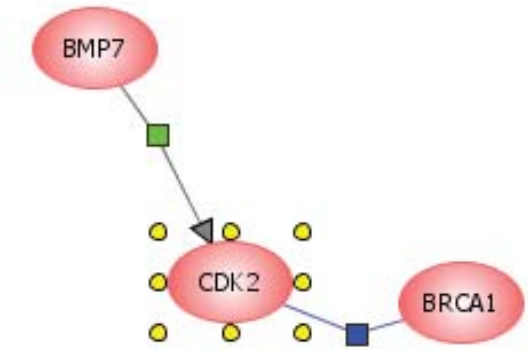

Figure 3
Connecting BRCA1 with TGF-beta Signaling Pathway

While CDK2 may suggest a possible connection between the induced TGF-beta signaling and the cell cycle regulation as well as DNA repair involving BRCA1, there may be other candidate molecules that connect BMP7 and BRCA1. We selected BRCA1 and BMP7 and used the "Shortest Path Network" function from the Interactions Network in the PathwayArchitect workflow pane, to find the shortest connections between the two genes.

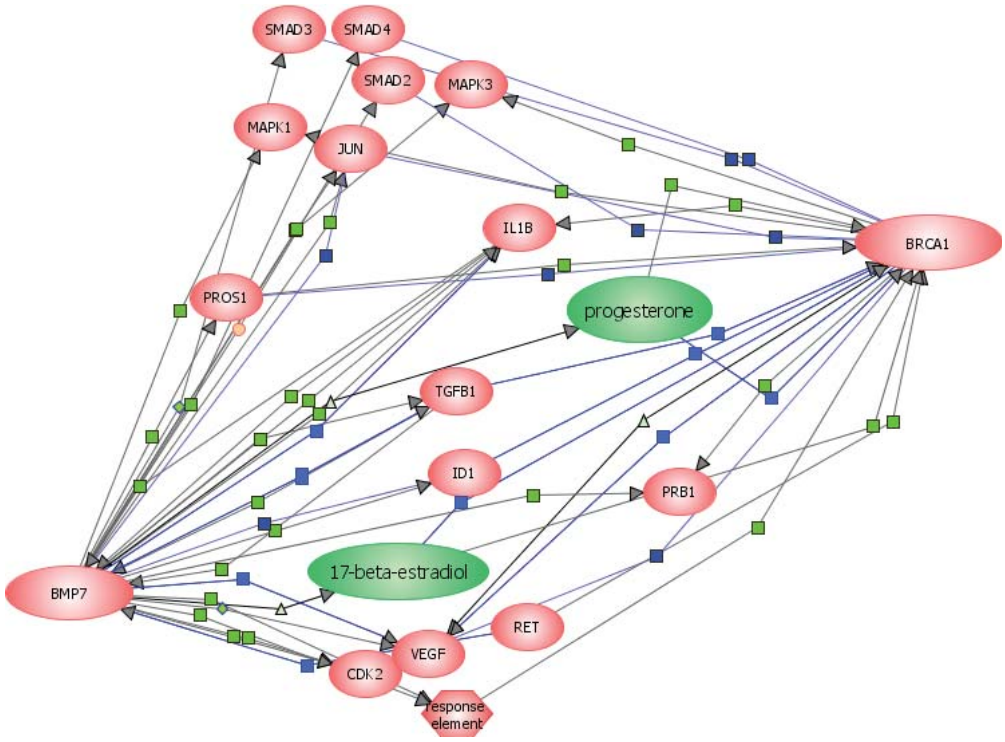

Figure 4
Shortest Network Connecting BMP7 and BRCA1

We found that the shortest connection network, as shown in Figure 4, includes various other components of the TGF-beta signaling pathway, including inducing ligand TGFB1, transcription factors SMAD 2, 3 and 4 and MAP3K. We also notice that the shortest network between BMP7 and BRCA1 includes the female hormone progesterone, whose decreasing level in women with menopause is believed to increase their susceptibility to breast cancer. In addition, the network includes 17-beta estradiol which is a form of estrogen in the body. From these results, one may speculate that the BMPs probably enhance estrogen secretion and suppress progesterone levels which together regulate the level of BRCA1 expression in the tumor. Thus we can better hypothesize the connection between these two sub-networks within a larger breast cancer network by studying the interactions, visualizing the entire network data simultaneously, and then exploring selected interactions in greater detail.

Unraveling Interactions with Lesser Known Pathways

Examining what gene ontology (GO) terms are over represented in a dataset allows the user to identify previously lesser known functions and processes within a biological network. PathwayArchitect software allows the user to examine what GO terms are enriched (having significant *p*-values) in a dataset and exactly what genes are associated with each GO term of interest. We launched the GO Browser function under Enrichment Analysis in the Workflow, and selected all the genes that are either up- or down- regulated in breast cancer samples to see what GO terms are enriched in the dataset. The GO terms that are over represented are highlighted in blue as seen in Figure 5 below.

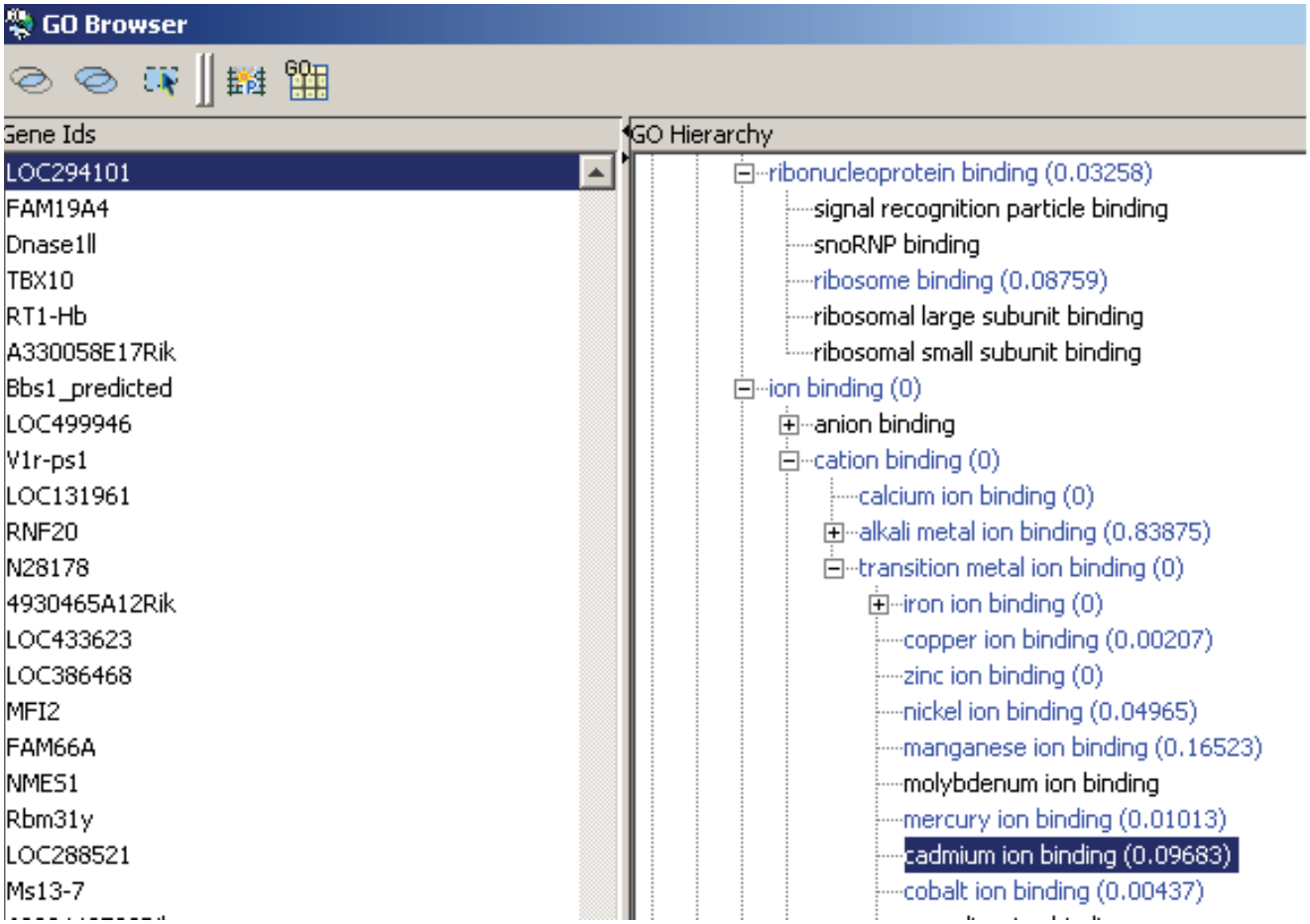

Figure 5
GO Browser Showing Over Represented GO Terms

Amongst several interesting GO terms, we noticed that GO terms corresponding to "iron ion binding" and "iron ion homeostasis" had very low *p*-values. Iron is a well known micro-nutrient for growing breast cancer cells and transferrin receptors for iron are enriched in breast cancer cells.[9] However, regulation of iron metabolism and its putative connections within the network of breast cancer related proteins have been less well-studied. Hence, we selected these terms and chose the function "Show genes with this term", from the GO Browser. The selected genes corresponding to these terms were highlighted on all open spreadsheets and pathway views.

In our microarray data, 21 putative iron ion binding genes and 4 iron ion homeostasis genes are present. For the purpose of this study, we concentrated on the classical iron binding protein Transferrin (TF) which had significantly increased transcript levels in breast cancer tissues and cell lines compared to normal samples. By examining the Direct Interactions Network we find that TF is connected to 13 other proteins in the network, as shown in Figure 6. Thus we could speculate on several alternate connections through which TF can fit into the large network of breast cancer related proteins.

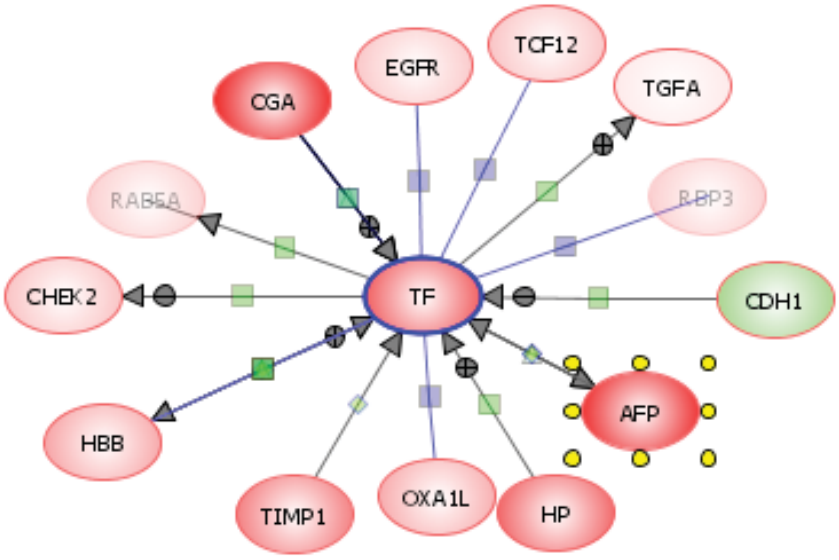

Figure 6
Neighbors of Transferrin in the Direct Interactions Network

One of the possible connections includes the epidermal growth factor (EGF) mediated signaling pathway which activates the MAPK signaling. Since the MAPK signaling pathway has well established functions in activating transcription factors that regulate a variety of growth, differentiation and cell death steps, its importance in cancer tissue has been widely studied. TF activates EGF receptor protein both by direct binding and by indirect activation of its ligand TGF-alpha (TGFA), and may thus trigger a putative mitogenic response. These explorations may be speculative and further biological testing required defining the exact role of TF in breast cancer. Nevertheless, this analysis helps narrow down the tentative roles that TF may play in breast cancer and guide biological validation experiments which are resource intensive, laborious and time-consuming.

Assuming that TF may positively regulate TGFA which in turn phosphorylates EGF receptor protein and activates the MAP kinase pathway, we chose to study the common targets of these two proteins. We selected TF and TGFA proteins and launched the "Network Targets" function under the Interactions Network, in the workflow (As before, launching the function from the workflow displays only the interactions of maximum quality). Amongst 7 small molecules that are potential targets for both TF and TGFA (Figure 7), we found cisplatin, a known anti-cancer drug which is known to bind TF. We may then speculate that one of the possible ways by which cisplatin targets the EGFR activated mitogenic cell proliferation is by binding TF and hence indirectly affecting the levels of iron available to proliferating cancer cells. One more interesting candidate among the TF and TGFA targets is the female hormone progesterone. Progesterone may block both EGF mediated cell proliferation and iron availability to the cancer cells.

Similarly, a search for small molecules that regulate TF and TGFA signaling shows several interesting candidate molecules (Figure 8). We selected TF and TGFA and chose the "Network Regulators" function under Interactions Network, in the Workflow. The regulators of this sub-network include 17-beta estradiol, Ret-A, and phorbol esters. 17-beta estradiol is used for hormone replacement therapy in older women and may increase the risk of breast cancer while Ret-A and phorbol esters are known cancer inducers. In addition, the regulators also include vitamin D and calcium, which when included in diet, are known to lower breast cancer risk. We may speculate that anti-breast cancer activity of vitamin D and calcium is mediated through iron availability and EGF signaling. Anti-clotting agent lovinox that reduces risk to cancer and anti-cancer drug suramin that prevents angiogenesis in tumors, are also found amongst the regulators of this branch of the breast cancer network.

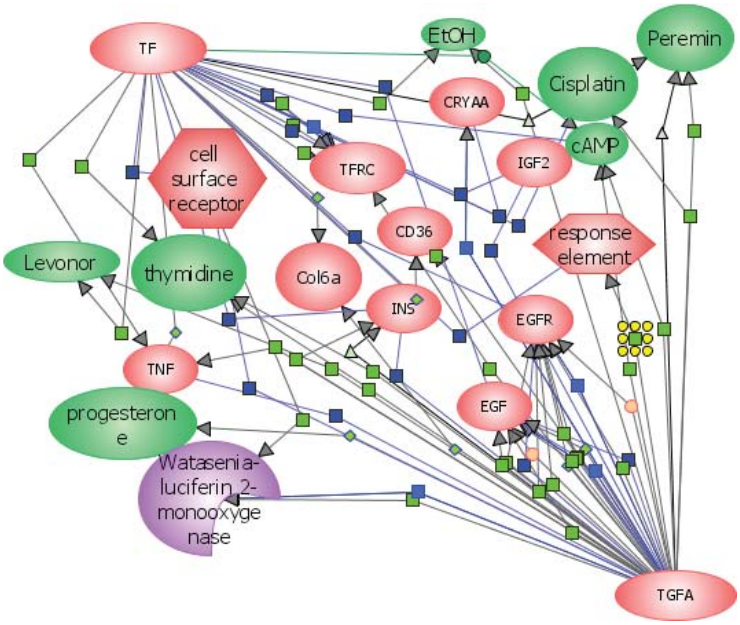

Figure 7
Network Targets of TF and EGFR-Mediated MAPK Signaling Pathway

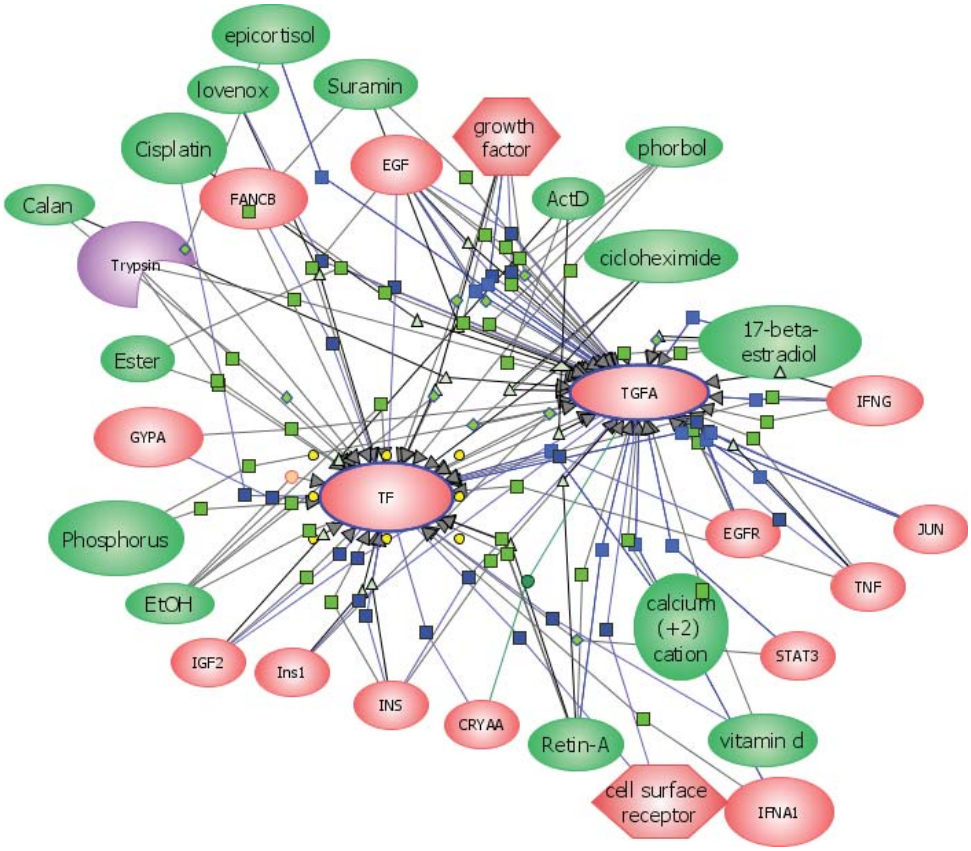

Figure 8
Network Regulators of TF and EGFR-Mediated MAPK Signaling Pathway

Amongst the enriched GO terms displayed in PathwayArchitect software, we also noticed that the term "zinc ion binding" has very low *p*-value, implying an over-representation of these genes in our breast cancer dataset. We found 45 genes were present in the breast cancer data representing this GO term. A study in 1990 showed that zinc is accumulated in breast cancer tissues and depressed in the plasma of breast cancer patients.[10] Later, it was shown that modulation of estrogen receptor zinc finger function suppresses breast cancer growth in vitro and in nude mice.[11] Apart from these sporadic studies and information, little is known about how many genes in total may be responsible for the accumulation of zinc in breast cancer tissues. The 45 zinc ion binding proteins enriched in the breast cancer dataset need further in depth study for their roles in the breast cancer tissue and whether or not they can be developed as other putative targets, whose zinc binding ability can modulate tumor growth.

Exploring Putative Roles of Less Characterized Genes

Two proteins whose roles remain relatively less elucidated in the context of breast cancer are fatty acid desaturases 1 and 2 (FADS1 and 2). FADS1 and FADS2 mRNA are enriched in breast cancer tissues. We selected FADS 1 and 2 and launched the "Expand Network" function from Interactions Network in the Workflow. Our expanded search yielded 3 proteins - SULT1E1, SREBF1, and NOVA1. SULT1EI is a protein belonging to the sulfotransferase protein family (SULT1E1) which transfers a sulfur moiety to and from estrogen and thus may be the checkpoint of determining how much estrogen sulfate is available for hydrolysis to estradiol, which fuels breast cancer cell proliferation. The second protein SREBF1 is a steroid responsive element binding transcription factor. Paraneoplastic antigen NOVA1 antibodies are found in the sera of breast cancer patients. Thus we may speculate that FADS may play a role in breast cancer by affecting either synthesis or biological effect of estrogen.

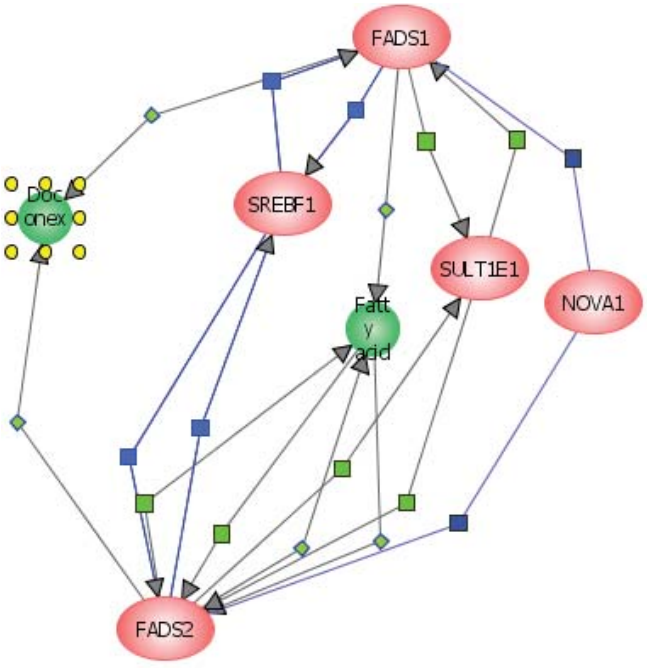

Figure 9
Identifying Less Characterized Genes in Breast Cancer

Searching for Potential Drugs Molecules and Target Proteins

The ultimate goal of any cancer study is to find potential inducers of cancer or define drug molecules and targets for drug design. We used the Relevance Interactions Network function in PathwayArchitect software to find proteins and small molecules with high "statistical relevance" to a given network that would increase the confidence score for the breast cancer network (see Technical Note "Relevance Interaction Networks in PathwayArchitect software" for more details at www.stratagene.com/SoftwareSolutions). To do this, we selected a second list of known breast cancer markers. The Lasso feature in PathwayArchitect software allows the user to select nodes and interactions across all open views and spreadsheets. There are 26 common entities between the known breast cancer marker list and the microarray experiment derived list of 336 entities that are part of the Direct Interactions Network. We selected these 26 genes and "Small molecules" that are relevant to the network, from under Relevant Interactions Network, in the Workflow.

As shown in Figure 10, amongst them, we found estrogen hormone which is one of the main inducers of breast cancer, as mentioned above. Indeed, breast cancers are classified by their sensitivity to estrogen molecules and many potential drugs are designed to target estrogen response of tumors. We also found tamoxifen, a known breast cancer drug as a statistically relevant small molecule that computationally increases the confidence score of the breast cancer network, based on relevance calculation in PathwayArchitect software.[12] Another interesting small molecule relevant to the breast cancer network is fulvestrant, a recent breast cancer drug from Astra Zeneca.[13] These findings provide us with confidence that the computational search for relevant drug molecules in breast cancer may guide us towards biologically relevant molecules.

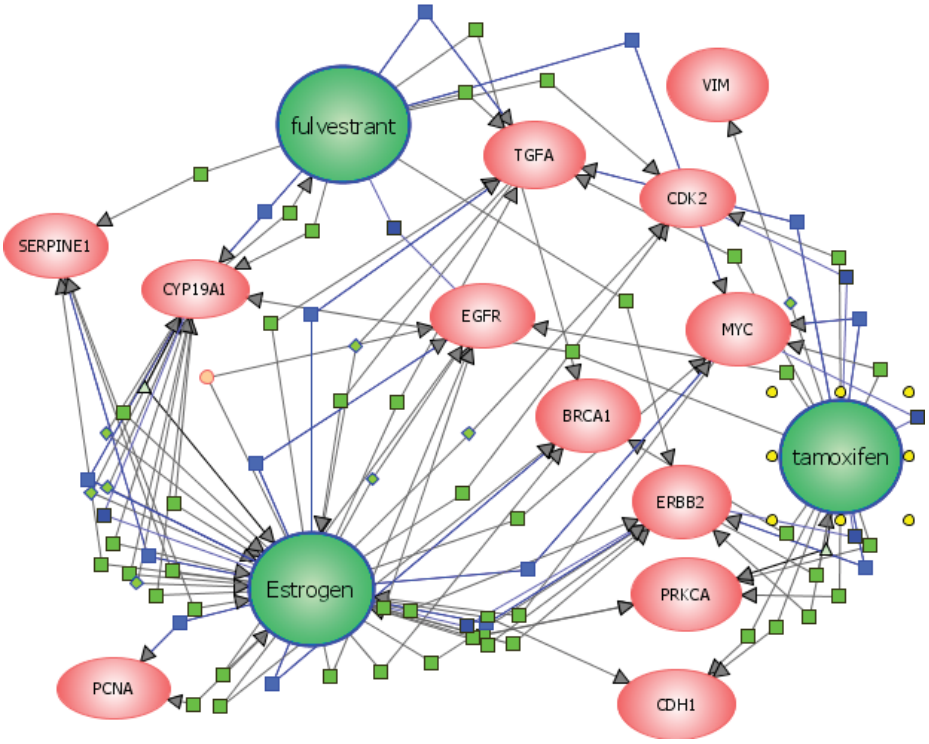

Figure 10
Identifying Less Characterized Genes in Breast Cancer

Searching for Proteins as Potential Drug Targets

In the constructed relevance network for small molecules, we noticed that the estrogen hormone molecule has the highest number of interactions with an aromatase protein molecule CYP19A1. CYP19A1 belongs to cytochrome P450 super-family of enzymes and catalyzes the last steps of estrogen biosynthesis from androgen. CYP19A1 has been recognized both as a marker and a drug target for breast cancer. Aromatase inhibitors, letrozole, anastrozole and exemestane are popular drugs that block estrogen synthesis in post-menopausal women. All three known aromatase inhibitor drugs were found to directly interact with CYP19A1. However, we also identified over 100 small molecules that interact with CYP19A1 which are either drugs or could be potential pharmacological targets particularly against breast cancer in post menopausal women. These include many other known aromatase inhibitors like rogletimide, testololactone, finrozole, YM-511, atamestane, vorozole and many others being used to treat breast cancer. The list also includes lesser studied molecules speculated to be potential drug candidates like TAN-931, MFT-279, Lg-101305, Anabasine, etc. We have included small molecules which are connected to very few other molecules in the database outside CYP19A1 (those with a connectivity less than 20), suggesting that these may target aromatase inhibitors, specific for breast cancer. More importantly, CYP19A1 may be a potential site of attack for aromatase inhibitors thus blocking estrogen production and its availability to proliferate cancer cells. These small molecules being drug candidates and CYP19A1 being a target of aromatase inhibitors could be further explored in biological experiments.

We have biased our study towards exploring interactions with relatively well studied genes and molecules that play a role in breast cancer. While this study has increased our understanding of how known proteins interact with each other and expand our current knowledge about pathway interactions in the breast cancer tissues, we have not limited the scope of this study by any means. There are several less well characterized and novel proteins and small molecules in the networks which could provide invaluable information regarding breast cancer and should be studied in more detail.

Finding molecules that are known inducers of breast cancer and drug molecules that are specific for breast cancer validate the computational algorithms in the PathwayArchitect software and show that the tool can correctly mine biologically relevant information from within our current knowledge of cancer literature. As mentioned before our method speculates potential interactions and roles based on existing knowledge in the literature. Nevertheless, it provides additional insight into possible biological functions and refining subsequent biological experiments.

Summary

We have demonstrated throughout this Technical Note that PathwayArchitect software can provide researchers with an extensive database of biological interactions, coupled with powerful computational data mining and visualization tools that can tremendously broaden the scope of knowledgebase for important diseases. Insights gained into disease mechanisms through this type of analysis can greatly accelerate drug discovery by both proposing primary and secondary drug targets and by helping scientists decide which lead compounds to first screen against the selected target.

REFERENCES

1. Chopra R. (2001) *Journal of Clinical Oncology* 19: 106-111.
2. Miller W.R., et. al. (1994) *British Medical Journal* 309: 1573-1576.
3. Lippman M.E. and Allegra J.C. (1980) *Cancer* 46: 2829-2834.
4. Ravdin P.M., et. al. (1992) *Journal of Clinical Oncology* 10: 1284-1291.
5. Shih H.A., et. al. (2002) *Journal of Clinical Oncology* 20: 994-999.
6. Ghimenti C., et. al. (2002) *Genes Chromosomes Cancer* 33: 235-242.
7. Lange, C. A., et. al. (1998) *J Biol Chem* 273;47: 31308-31316.
8. Hayami R., et. al. (2005) *Cancer Res* 65(1): 6-10.
9. Elliot R.L., et . al. (1993) *Ann N Y Acad Sci.* 30;698: 159-66.
10. Holtkamp W., et. al. (1990) *Onkologie* 13;3: 207-9.
11. Wang, L.H., et al. (2004) *Nature Med.* 10(1); 40-47.
12. Pradine, et. al. (2005) *J.Comput.Biol.* 12;2.
13. Johnston, et. al. (2002) *Curr Opin Investig Drugs* 3;2: 305-12.

LEGAL

For more information on this Technical Note
or to contact our Technical Service department:

Stratagene US and Canada
Technical Service: 800-894-1304 x2

Stratagene Europe
Technical Service: 00800-7400-7400

Stratagene Japan K.K.
Technical Service: 3-5821-8076

Visit our web page at:
www.stratagene.com/contacts/TechServices

Email our Technical Service department at:
techservices@stratagene.com

www.stratagene.com