Reliable mRNA Expression Profiles with FFPE Tissues Using the Agilent SurePrint G3 Human Gene Expression v2 8x60K Microarray

Application Note

Authors

Tri B. Doan, Elise Chang-Wong and Alexander Wong
Agilent Technologies Inc.
Santa Clara, CA
USA

Abstract

A study has been performed to demonstrate the ability to reliably profile mRNA expression from poor quality RNA derived from FFPE tissues, using an Agilent microarray gene expression workflow that incorporated the SurePrint G3 Human Gene Expression Microarray, the SureTag Complete DNA Labeling Kit, and the GeneSpring GX Software. Normal and tumor gastric and colorectal FFPE tissues preserved on FFPE slides were used to perform the studies. In spite of the low quality of the extracted RNA (RIN 2.4 to 2.7), the resulting microarray data was highly reproducible, and generated expression data on as many as 3800 genes. Statistical analysis of the colorectal dataset revealed 10 genes and 4 lincRNAs with the highest PCA1 scores. Subsequent pathway analysis using GeneSpring GX revealed correlations with canonical pathways known to be involved in tumorigenesis. The workflow described in this study can thus produce reliable gene expression data from FFPE tissue and enable understanding of its biological context.

Introduction

The development of microarray technology has revolutionized molecular biology, enabling researchers to conduct global gene expression analysis on tens of thousands of genes, representing the entire genome of the species simultaneously. Microarray-based gene expression profiling has been used in gene discovery, biomarker identification, disease classification, gene regulation and metabolic pathways. For typical workflows on gene expression analysis using microarrays, it is recommended that RNA extracted from freshly frozen tissues have RIN scores of 7 or greater be used.

However, in many cases, formalin-fixed paraffin-embedded (FFPE) tissues are the only samples available. FFPE samples represent a valuable resource of information that can allow for elucidation of molecular events linked to disease onset, progression and response to course of treatment. Nucleic acids isolated from paraffin-embedded samples, however, are usually suboptimal in quality.

This study demonstrates that samples with RIN scores typical of FFPE-derived RNA, in tandem with a good labeling kit, microarrays with excellent resolution and a powerful software analysis suite, can still be used for expression analysis to generate biologically-relevant data. Total RNA was extracted from one or two 20 µm sections of paraffin block from tumor and paired

Introduction *(continued)*

adjacent normal tissue (gastric and colorectal). Analysis of the microarray results using GeneSpring GX 12.5 demonstrated excellent assay reproducibility between the technical replicates. Principal component analysis of array signals demonstrated clear separation between the tumor and normal samples with grouping of replicate samples. Statistical analysis of the expression level variations between tumor and normal gastric and colorectal samples revealed close to 1500 (gastric) and 3800 (colorectal) genes with a fold-change greater than 2 and a corrected p-value of 0.05. Significant differences in gene expression were observed between normal and tumor samples in many of the key tumor suppressor genes and oncogenes, which corresponded to a number of biologically-relevant pathways identified using the GeneSpring NLP Network Discovery Tool.

Materials and Methods

RNA Isolation and Qualification

FFPE blocks from tumor and paired adjacent normal tissue (gastric and colorectal) were purchased from Cureline. For each FFPE sample, two or four 20 µm sections were cut from each paraffin block using a microtome, combined, and weighed. Samples were then processed according to the Qiagen RNeasy FFPE kit specifications. FFPE tissue sections were treated with xylene to remove paraffin and ethanol to dry. The pellet was then re-suspended with buffer PKD and Proteinase K at 56˚C for 15 minutes, then 80˚C for 15 minutes for complete protein digestion, incubated on ice for 3 min, then centrifuged for 15 min at 13,500 rpm to collect the supernatant. The supernatant was then treated with DNase I at room temperature for 15 minutes and the collected total RNA was purified on RNeasy MinElute columns (Qiagen). RNA was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific) according to the manufacturer's recommendations. RNA quality was assessed on the Agilent Bioanalyzer, which determined an RNA Integrity Number (RIN).

RNA Amplification & cDNA Labeling

The TransPlex Whole Transcriptome Amplification kit (Sigma Aldrich) was used to generate amplified cDNA from 100 ng of total RNA input with a minor modification utilizing TITANIUM Taq DNA polymerase (Clontech) during the library amplification step to produce whole transcriptome amplification (WTA) product. Three technical replicates were run for each condition. Amplified cDNA was purified using the Qiagen QIAquick PCR Purification kit and quantified using the NanoDrop spectrophotometer. The Agilent SureTag Labeling kit (P/N 5190-3400) was used to enzymatically label 1.8 µg of amplified cDNA with cyanine 3-dUTP for 2 hours at 37 °C. Cyanine 3-labeled samples were cleaned up using an Amicon Ultra-0.5, Ultracel-30 Membrane, 30 kDa filter provided with the SureTag Kit. Yield and specific activity were determined using the NanoDrop spectrophotometer.

Hybridization and Washing

Each of the purified and labeled samples was combined with Agilent 10x Blocking Agent and 2x Hi-RPM Hybridization Buffer Solution (both components of the Agilent Gene Expression Hybridization Kit) and the mixtures were denatured at 95 °C for 3 minutes prior to array hybridization. The samples were hybridized to the Agilent SurePrint G3 Human Gene Expression v2 8x60K Microarray. Each slide contains eight identical arrays containing 50,599 biological probes, representing over 24,500 RefSeq genes and over 10,000 lincRNAs. Hybridization was carried out at 10 rpm at a temperature of 65 oC for 17 hours. Following hybridization, the arrays were washed according to the procedure outlined in the Agilent One-Color Microarray-based Gene Expression Analysis Manual, version 6.6, September 2012, Publication No. G4140-90040.

Microarray Scanning & Data Analysis

Scanning and image analysis were performed using the Agilent SureScan Microarray Scanner (P/N G4900DA) according to Publication No. G4140-90040. Feature Extraction Software v10.7 was used for data extraction from raw microarray images files. Data visualization and analysis were performed with GeneSpring GX.

Results and Discussion

Microarray gene expression analysis of RNA extracted from FFPE human tissues was performed by first extracting total RNA, then producing cDNA of the mRNA population, followed by cDNA labeling and cleanup. The labeled cDNA was hybridized with the Agilent SurePrint G3 Human Gene Expression v2 8x60K Microarray, which was then washed and scanned with the Agilent SureScan. The scanned images were extracted using Agilent Feature Extraction v.10.7, and analyses were performed using the Agilent GeneSpring GX Software. An overview of the workflow is shown in Figure 1.

RNA Yield and Purity

At least 4 µg of total RNA with A260/A280 ratios of 2.0 were derived from each of the FFPE samples (Table1). The assessed quality of the RNA using the 2100 Bioanalyzer was typical of FFPE-derived samples, with RIN values in the 2-3 range (Figure 2).

Concordance of Technical Replicate Samples

Excellent reproducibility was observed in the scatter plots comparing the $\log^2$ intensities of the 75th percentile-normalized green processed signals (gProcessedSignals) of all probes between technical replicate samples. Correlation values (R^2) for all the replicate comparisons averaged 0.996. (Figure 3A and 3B) All assays spanned dynamic ranges of four or more orders of magnitude.

Principal component analysis (PCA) of gastric and colon samples was performed using the Quality Control step in the gene expression analysis workflow of GeneSpring GX. When comparing all samples in the experiment, the largest separation was seen between the tissue types, gastric and colon samples. The next largest differentiator was seen within tissue types, between normal and tumor samples. Moreover, the normal and tumor samples were clearly separating along PCA1, which captures the largest amount of variation in the datasets. (Figure 4).

Hierarchical Clustering Analysis

Another quality control tool which can be utilized in GeneSpring GX is hierarchical clustering based on all samples in the experiment. The hierarchical cluster shows the largest branch between tissue types (Figure 5). The next largest branch of separation is between normal and tumor disease state within each tissue type. All the replicates are clustered together, confirming the robustness of the SureTag Labeling Kit and its compatibility with FFPE samples.

Statistically Significant mRNA Expression Differences

There are many ways to explore the sea of data that is generated through the use of microarrays, and a wealth of information can be extrapolated from these studies. The best answers are the ones that can be validated by a few different statistical approaches, and then validated again by PCR. This study used two main approaches to illustrate some of the statistical analysis capabilities of the GeneSpring Software. The Student's T-test was used for this dataset, along with PCA, to discover which genes are most responsible for the variations observed in the experiment.

In the colorectal experiment containing three replicate tumor and three replicate normal samples, probes that did not have a "Detected" signal in 100 percent of the replicate samples in each condition were filtered away. PCA was performed on this Flag Filtered gene list, looking for the top 2 major components. PCA is a mathematical method to condense the overall variations in a dataset into a couple of major trends. The dominant trend in the data is captured in the eigenvector with the biggest associated eigenvalue, PCA1. In this dataset, 61.08% of the variations in the entire dataset can be summed up, condensed, or explained by PCA1. PCA2 explains 8.99% of the remaining variations not captured in PCA1. PCA is more informative when looking at datasets with more than 2 conditions, but can still be used in this case.

Selecting the genes with the highest score for PCA1 provided the genes that most reflected the variations captured by PCA1, and thus the largest trend in the dataset. This list is shown in Table 2, sorted by PCA1 scores of 10 genes, and 4 lincRNAs with the highest PCA1 scores.

Also using the Flag Filtered list, a volcano plot analysis of the differential expression between normal and tumor tissue was performed on both colorectal and gastric samples. An asymptotic p-value computation was performed with Benjamini-Hochberg False Discovery Rate (FDR) multiple testing correction. Genes with an absolute differential fold change of at least 2-fold with a corrected p-value of at least 0.05 are colored in red (Figure 6).

A Venn diagram (Figure 7) was created to examine the genes that had the highest PCA1 scores, at least a 2-fold expression change, and that had passed the t-test, with FDR multiple testing correction. The 7 genes and 1 lincRNA passing these criteria are shown in Table 3, sorted by largest fold change. The most significant trend for these genes was down regulation (Figure 8).

Results and Discussion
(continued)

Pathway Analysis

In order to determine how these genes might be involved in carcinogenesis, a biochemical network was created from the 10 genes with the highest PCA1 scores, using the GeneSpring Natural Language Processing (NLP) Network Discovery function (Figure 9). Both probes for BMP3 were over 35-fold down-regulated, with corrected p-values of 3.33E-03, and 2.21E-03 respectively. This data corroborates the finding that the inactivation of BMP3 is an early and frequent event in colorectal cancer development[1].

There are many possible effects from the down regulation of BMP3 in colorectal cancer. The pathway generated using GeneSpring GX provided information about the possible mechanisms (Figure 7). In this pathway, BMP3 is a regulator of SMAD2 through protein modification. SMAD2 has also been linked to tumorigenesis. In addition, BMP3 is a negative regulator of BMP2. BMP2 is a positive regulator of BMP3, and both proteins are regulators of the transforming growth factor beta (TGFβ) family. It has also been found that disruption of the TGFβ pathway implicates tumorigenesis. The pathway generated In GeneSpring GX is an interactive diagram visualizing these relationships. Citation information can be accessed by clicking on any of the nodes to provide the PubMed links.

Another interesting finding is that an isoform of SLC26A3 covering the genomic coordinates chr7:107408063-107408004 was found to be both statistically differentially expressed, with a high PCA1 score, and was down-regulated by over 43 fold. This gene, with synonyms of DRA and CLD, has also been cited in many publications linked to colorectal cancer when down-regulated[2,3,4,5].

Biochemical networks generated using the NLP Network discovery tool in GeneSpring GX can help extrapolate or infer relationships between genes that in this case could be involved in tumorigenesis. An example inference that can be drawn from this network is that BMP3 affects the process of DNA-dependent transcription[6]. Clicking on the node between BMP3, IL1β, and SLC26A3 provides a reference that indicates that IL-1βhas been shown to reduce DRA (SLC26A3) mRNA expression in vitro by inhibiting gene transcription[7]. The loss of transport function in the surface epithelium of the colon by attenuation of transporter gene expression, perhaps inhibited at the level of gene transcription by pro-inflammatory cytokines, may play a role in the pathogenesis of diarrhea in colitis[8] and possibly colorectal cancer[8,9]. Unfortunately, these two relationships are extracted from the same citation. Thus, this finding is very preliminary, but it is a starting point for further research. Users should always click on the relationship nodes and examine the PubMed abstract to confirm any relationships found by NLP.

Canonical Pathway Analysis

Canonical pathways, such as Wikipathways and Biocyc pathways can also be employed to provide insight into the microarray data. Starting with a larger gene list, genes that are at least 2 fold down-regulated, passing the Standard t-test, with Benjamini Hochberg FDR multiple testing, and with a corrected p-value of 0.05, were selected to create a list of 1797 genes total. A "Single Experiment Analysis" was performed on this list of 1797 genes to find relationships in the canonical pathways.

The canonical TGFβ signaling pathway was found to be enriched, with a p-value of 6.48E-03 (Figure 10). Of the 118 entities involved in the pathway derived from analysis of the experimental data, 11 were also found in this list (highlighted in yellow), confirming the results derived from the pathway analysis.

The experimental data can be mined further, using the many statistical and pathway analysis capabilities of GeneSpring Software to elucidate the pathways and possible mechanisms involved in colorectal cancer development using FFPE tissue.

Conclusion

Formalin-fixed paraffin-embedded (FFPE) archival tissue samples are a valuable source of material for global gene expression profiling studies of human diseases. They offer a unique challenge but are of significant value, because in many cases they are frequently the only source of tissues available. The data presented in this study demonstrate that poor quality RNA derived from FFPE samples can produce reliable mRNA profiles using the Agilent microarray gene expression workflow described in this study, which utilized a high-performance labeling kit and microarrays with excellent resolution. Using GeneSpring GX Software provides powerful, accessible statistical tools for fast visualization and analysis of gene expression data, including pathway analysis. Designed specifically for the needs of biologists, GeneSpring GX offers an interactive desktop computing environment that promotes investigation and enables understanding of microarray data within a biological context.

References

1. Loh, K., Chia, J. A., Greco, S., Cozzi, S.-J., Buttenshaw, R. L., Bond, C. E., Simms, L. A., Pike, T., Young, J. P., Jass, J. R., Spring, K. J., Leggett, B. A. and Whitehall, V. L. J. (2008). Bone morphogenic protein 3 inactivation is an early and frequent event in colorectal cancer development. Genes Chromosom. Cancer, 47: 449–460.

2. Schweinfest CW, Henderson KW, Suster S et al. (1993). "Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas". Proceedings of the National Academy of Sciences of the United States of America. 90 (9): 4166–70.

3. Taguchi T, Testa JR, Papas TS, Schweinfest C (1994). "Localization of a candidate colon tumor-suppressor gene (DRA) to 7q22-q31.1 by fluorescence in situ hybridization". Genomics 20 (1): 146–7.

4. Höglund P, Haila S, Gustavson KH et al. (1998). "Clustering of private mutations in the congenital chloride diarrhea/down-regulated in adenoma gene". Human Mutation. 11 (4): 321–7.

5. Antalis TM, Reeder JA, Gotley DC et al. (1998). "Down-regulation of the down-regulated in adenoma (DRA) gene correlates with colon tumor progression". Clinical Cancer Research. 4 (8): 1857–63.

6. Grgurevic L, Macek B, Healy DR, Brault AL, Erjavec I, Cipcic A, Grgurevic I, Rogic D, Galesic K, Brkljacic J, Stern-Padovan R, Paralkar VM, Vukicevic, "Circulating bone morphogenetic protein 1-3 isoform increases renal fibrosis". Journal of the American Society of Nephrology SJ Am Soc Nephrol. 2011 Apr;22(4):681-92. Doi: 10.1681/ASN.2010070722. Epub 2011 Mar 17. .

7. Jiang W, Furth EE, Wen X, Katz JP, Sellon RK, Silberg DG, Antalis TM, Schweinfest CW, Wu GD., "Intestinal inflammation reduces expression of DRA, a transporter responsible for congenital chloride diarrhea". The American Journal of Physiology. 1998 Dec;275(6 Pt 1):G1445-53.

8. Dupaul-Chicoine J, Yeretssian G, Doiron K, Bergstrom K. S. B, McIntire C.R, LeBlanc P. M, Meunier C, Turbide C, Gros P, Beauchemin N, Vallence B, Saleh M, "Control of Intestinal Homeostasis, Colitis, and Colitis-Associated Colorectal Cancer by the Inflammatory Caspases". Immunity Vol 32, Issue 3, 26 March 2010, Pages 367-378

9. Sohn K, Shah S.A, Reid R, Choi M, Carrier J, Comiskey M, Terhorst C, Kim Y. "Molecular Genetics of Ulcerative Colitis-associated Colon Cancer in the Interleukin 2- and β2-Microglobulin-deficient Mouse". Cancer Research September 15, 2001; 61: 6912-6917

Tables

Sample Type	Conc. (ng/µl)	A260/280	A260/230	Yield (µg)
FFPE Normal Gastric	278	2.01	2.02	13.90
FFPE Tumor Gastric	95	2.03	1.91	4.75
FFPE Normal Colorectal	130	1.98	1.91	6.50
FFPE Tumor Colorectal	127	1.96	1.87	6.35

Table 1. FFPE Sample Yield and Purity

Probe Name	Component 1 (61.08%)	Component 2 (8.99%)	Component 3 (8.98%)	Gene Symbol
A_23_P18362	13.998987	-0.036361217	-3.3001862	SLITRK3
A_33_P3379816	13.882183	0.20662713	0.51133084	SMYD1
A_33_P3316621	13.8042145	0.47750616	-0.70329964	BMP3
A_33_P3258561	13.587834	-0.9991226	-0.3361261	SLC26A3
A_23_P48229	13.557139	-1.584934	-0.5683942	KCNA1
A_24_P453497	13.047358	0.20624113	-0.122980595	RBM20
A_23_P81058	12.86558	-1.0063794	-0.8440573	BMP3
A_23_P127760	12.663666	-0.2693293	-0.60042703	MS4A12
A_33_P3316953	12.6256275	-0.5278287	0.4249158	RBFOX3
A_21_P0008597†	12.506699	-4.023875	2.3335567	XLOC_011181
A_21_P0009451†	12.4837265	-1.9801863	0.6664485	XLOC_012622
A_21_P0000045	12.313205	0.37963986	-0.3057946	MS4A12
A_21_P0008591†	12.23699	-2.893068	2.3092976	XLOC_011178
A_33_P3307187	12.173254	-1.4096905	-0.002263904	NPAS3
A_33_P3337049	12.070871	-0.41763687	0.003252029	GLRA2
A_21_P0008868†	12.013392	-0.4367206	0.42453027	XLOC_011603

†lincRNA probe

Table 2. Flag Filtered List of the Entities (or probe names) That Most Reflected the Variations Captured by PCA1

Probe Name	Corrected p-value	p-value	Absolute Fold Change	Direction of Regulation	Component 1 (61.08%)	Gene Symbol
A_33_P3379816	2.81E-03	7.12E-05	48.5175	2.81E-03	13.882183	SMYD1
A_33_P3316621	2.21E-03	4.11E-05	46.742603	2.21E-03	13.8042145	BMP3
A_33_P3258561	3.40E-03	1.10E-04	43.76897	3.40E-03	13.587834	SLC26A3
A_24_P453497	1.60E-03	2.00E-05	37.852314	1.60E-03	13.047358	RBM20
A_23_P81058	3.33E-03	1.03E-04	35.29111	3.33E-03	12.86558	BMP3
A_23_P127760	1.12E-03	7.00E-06	33.83211	1.12E-03	12.663666	MS4A12
A_33_P3316953	9.19E-04	3.47E-06	33.805286	9.19E-04	12.6256275	RBFOX3
A_21_P0000045	8.82E-04	2.41E-06	31.022337	8.82E-04	12.313205	MS4A12
A_33_P3337049	9.25E-03	6.85E-04	29.138407	9.25E-03	12.070871	GLRA2
A_21_P0008868†	1.20E-03	8.40E-06	28.560474	1.20E-03	12.013392	XLOC_011603

†lincRNA probe

Table 3. List of the Genes With a Statistically Significant Fold Change of at Least Two

Figures

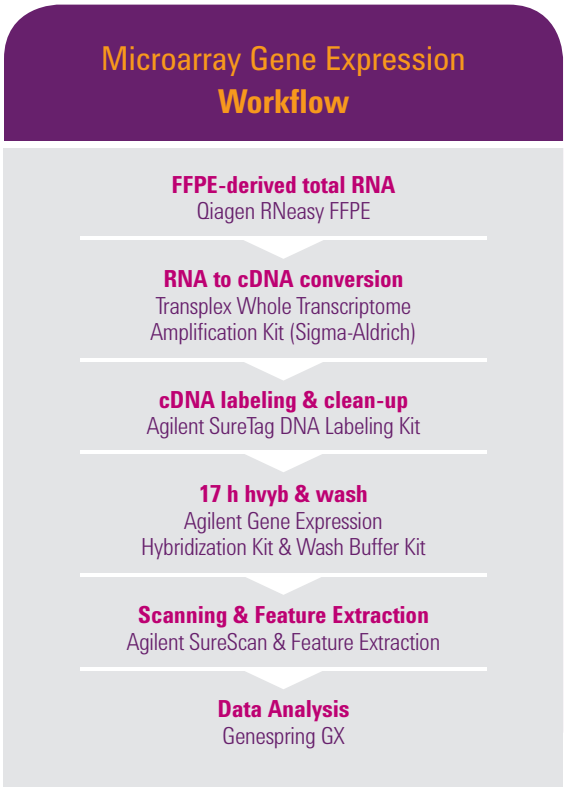

Figure 1. Workflow for processing FFPE samples using various kits in conjunction with various steps of the Agilent gene expression microarray workflow.

Figures
(continued)

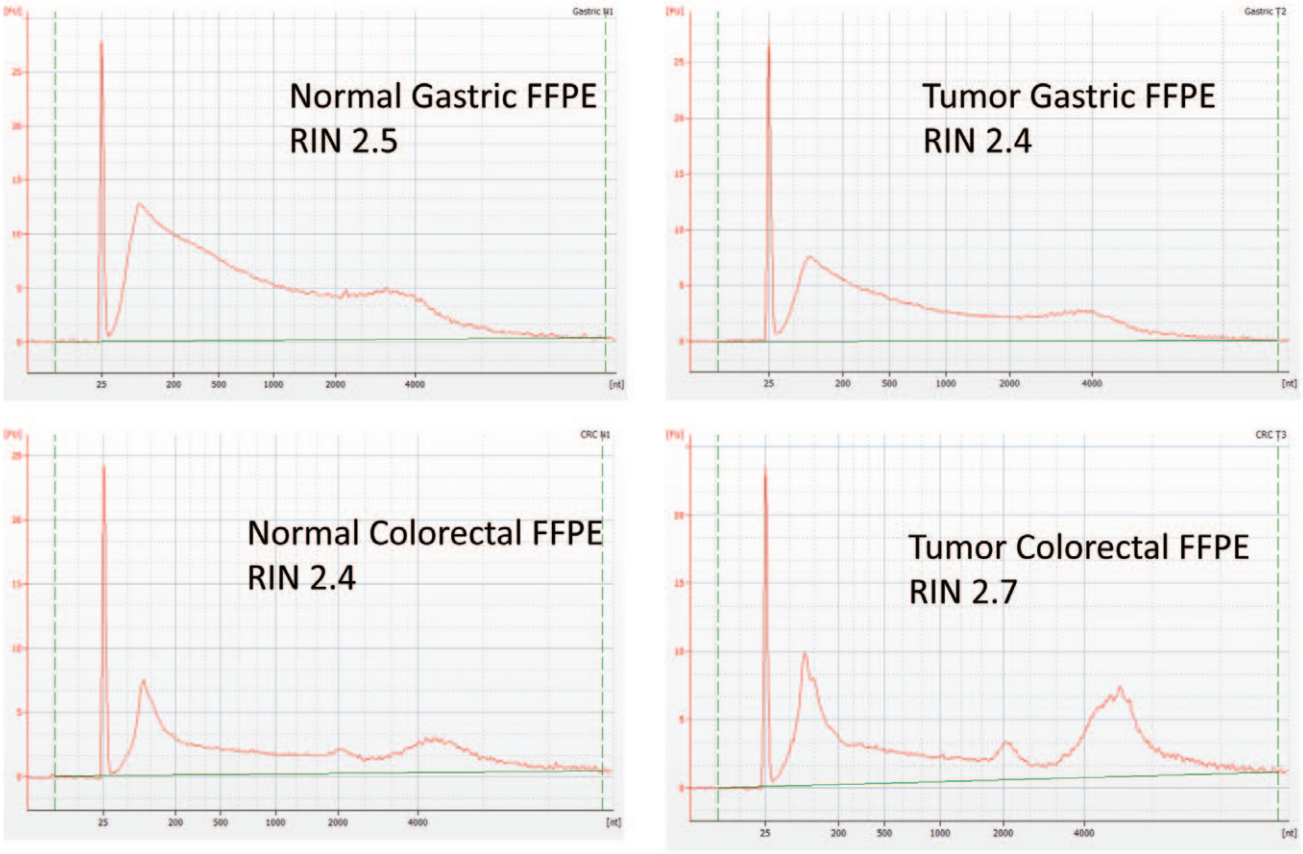

Figure 2. RNA quality was analyzed using the Agilent 2100 Bioanalyzer – Eukaryote Total RNA Nano Assay. The low RNA Integrity Number (RIN, around 2.5) was typical for FFPE extractions.

Gastric Normal and Tumor Samples

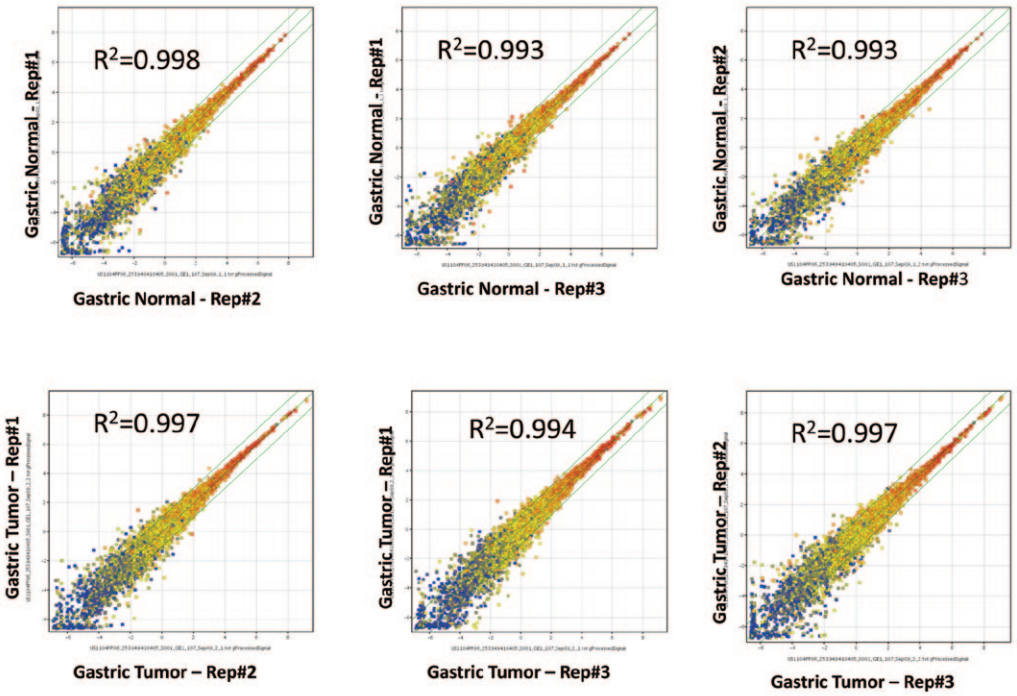

Figure 3A. Technical replicate scatter plot demonstrates excellent reproducibility.

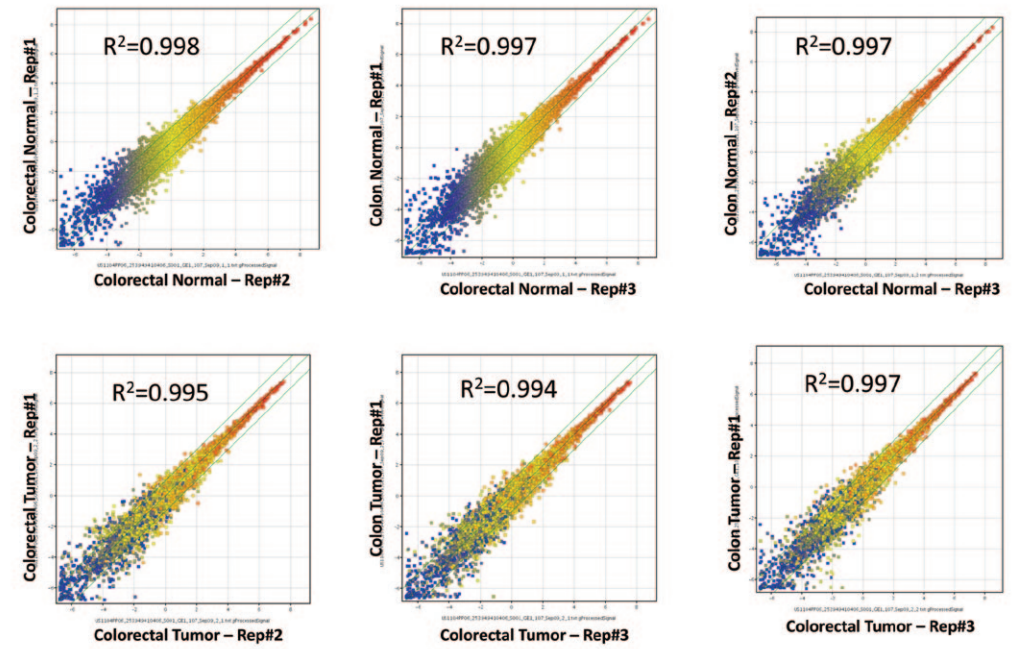

Figure 3B shows the technical replicates of the FFPE-gastric normal and tumor samples, with R^2 values for the replicate FFPE samples above 0.99. Figure 3B shows the technical replicates of the FFPE-colorectal normal and tumor samples, with R^2 values for the replicate FFPE samples also above 0.99.

Figures
(continued)

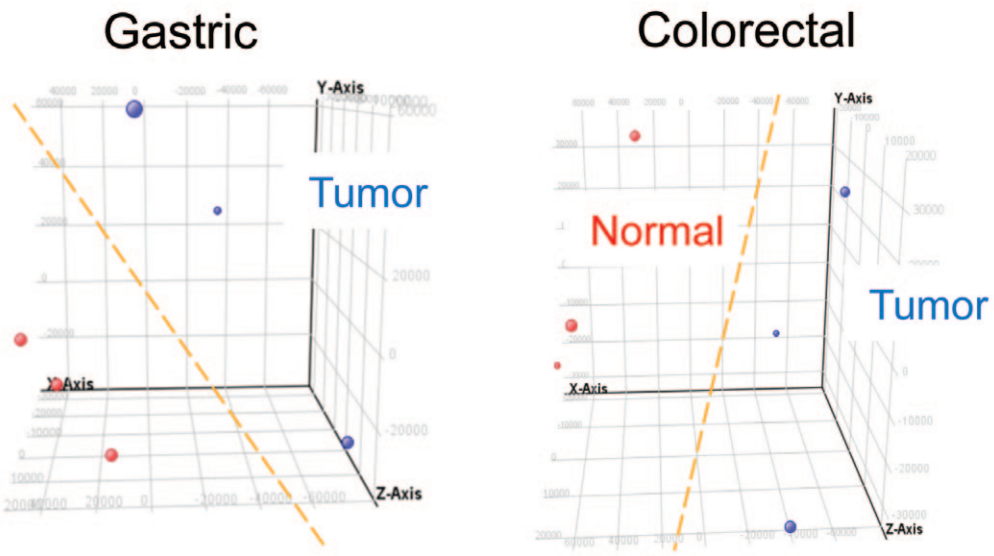

Figure 4. Principal Component Analysis (PCA) on the samples shows a clear separation along PCA1 (x-axis) representing the disease state of each tissue type.

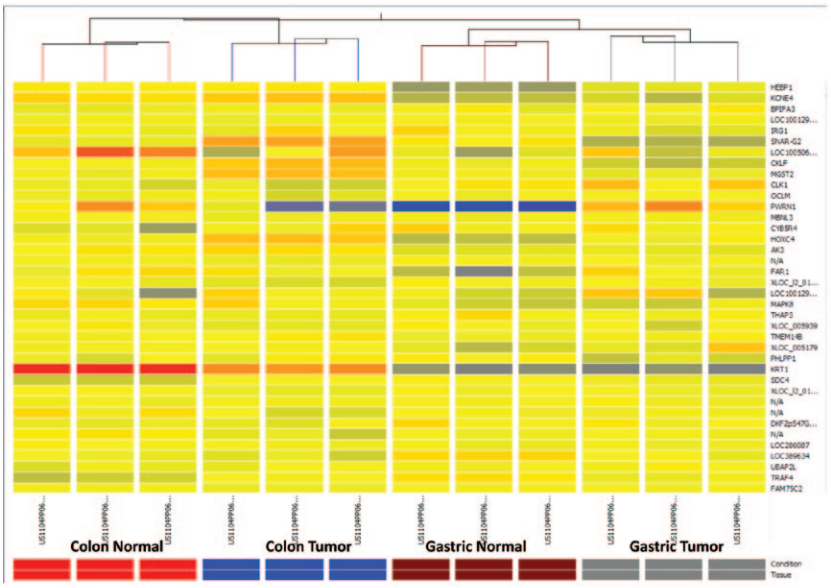

Figure 5. Hierarchical clustering analysis of the gene expression data. Without filtering away any probes so that there are no biases, the replicate samples clustered together, demonstrating that Agilent microarrays used with the SureTag Labeling Kit can detect distinct expression differences between conditions.

Figures
(continued)

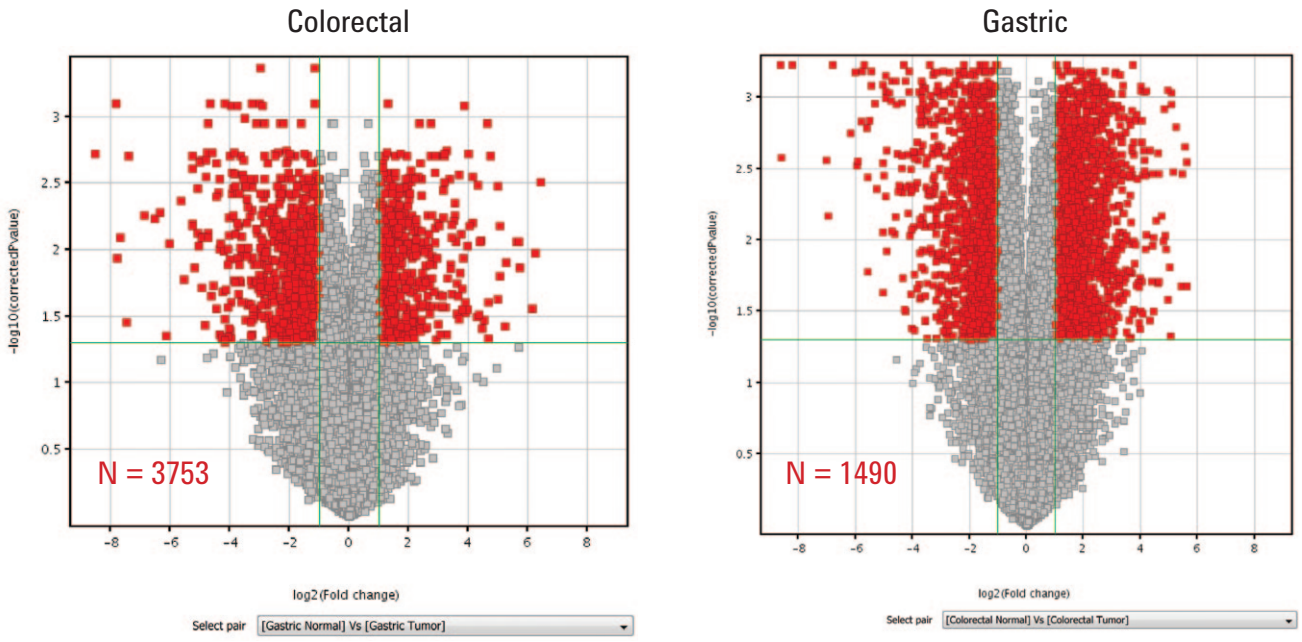

Figure 6. Volcano plot analysis of the differential expression between Normal versus Tumor FFPE samples. Genes with an absolute differential expression fold change of at least 2-fold with a corrected p-value of at least 0.05 are colored red.

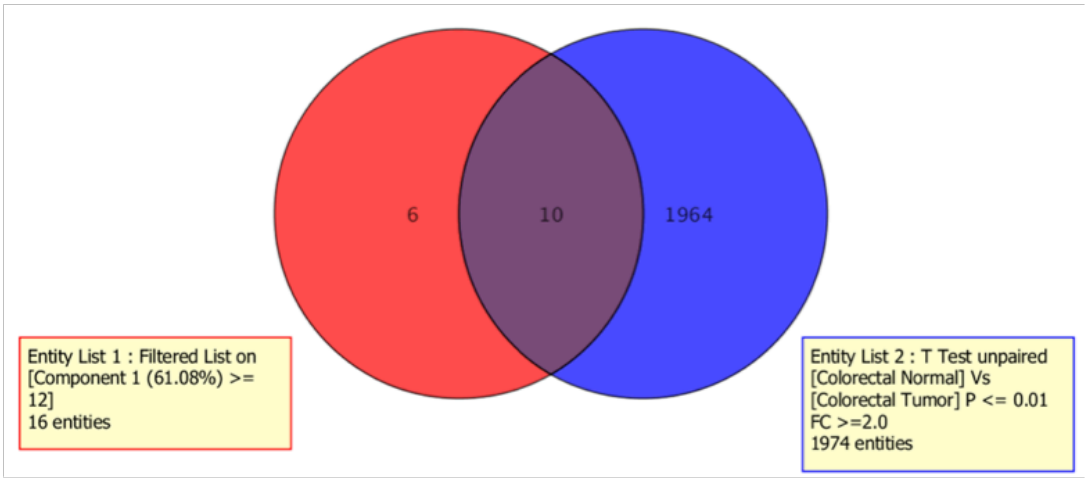

Figure 7. Venn Diagram. Entity List 1 includes the 16 entities/probe names representing (10 genes and 4 lincRNAs) with highest PCA1 scores. Entity List 2 includes the entities (or probe names) found to be significantly differentially expressed by the t-test unpaired, with Benjamini Hochberg FDT with corrected p-value of 0.01 and 2 folds changed. The intersection of the two lists provides 10 probe names representing 7 genes and 1 lincRNA, listed in Table 3.

Figures
(continued)

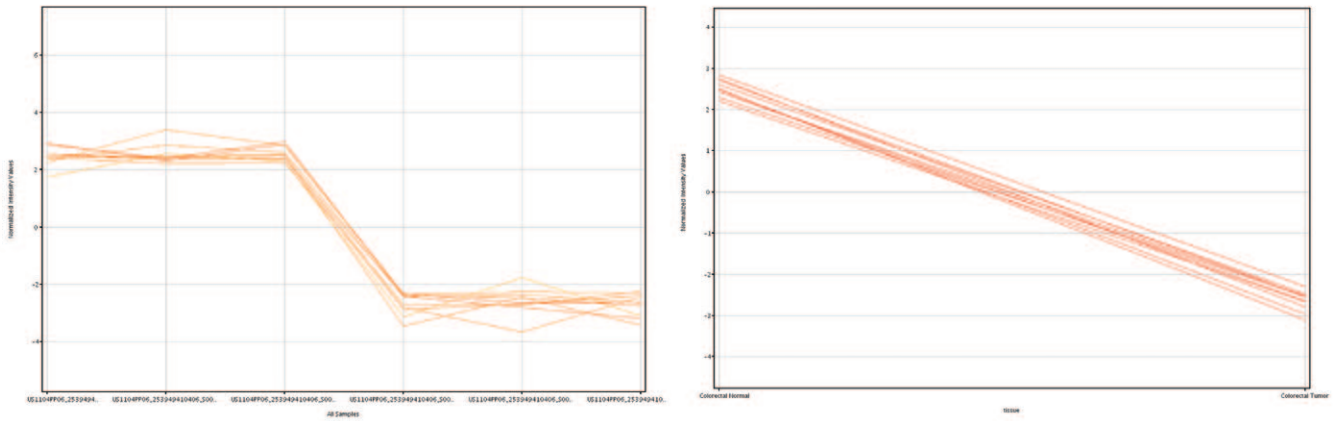

Figure 8. Profile plots of the genes with the highest PCA1 scores, illustrating the trend toward down-regulation.

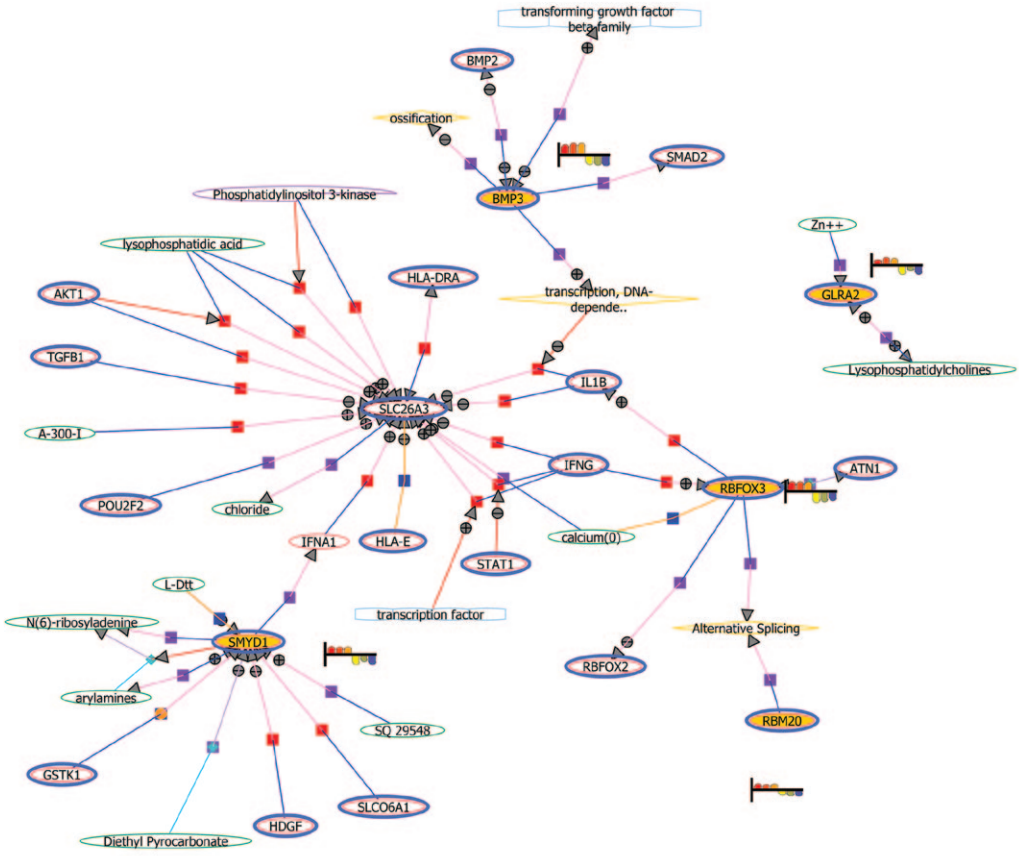

Figure 9. NLP biochemical network generated in GeneSpring of the most significantly differentially expressed genes, found by both t-test and PCA analysis.

Figures
(continued)

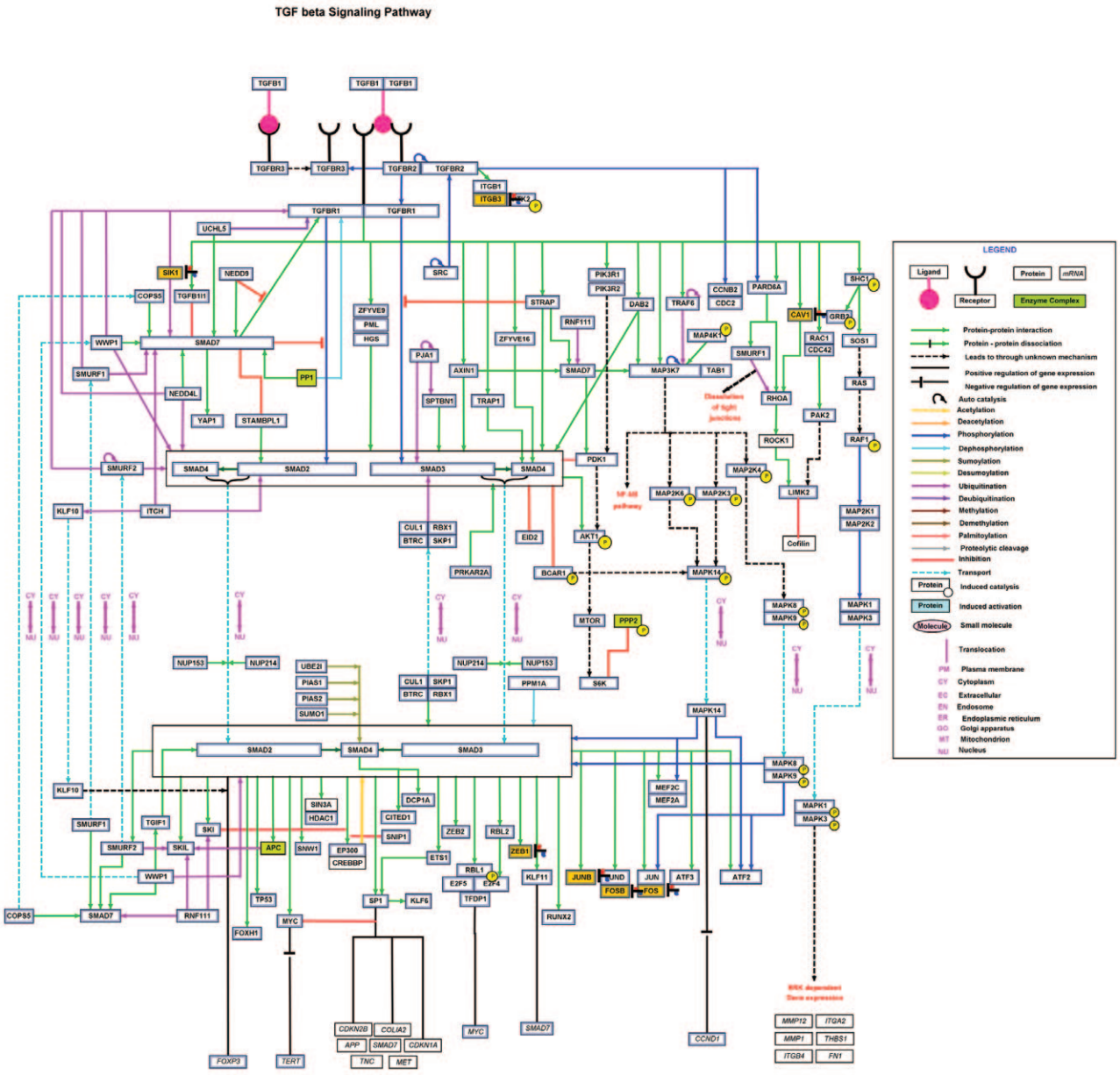

Figure 9. NLP biochemical network generated in GeneSpring of the most significantly differentially expressed genes, found by both t-test and PCA analysis.

For more information

Learn more or buy online
www.agilent.com/genomics/GEarraysFFPE

Find an Agilent customer center in your country
www.agilent.com/genomics/contactus

U.S. and Canada
1-800-227-9770
Agilent_inquiries@agilent.com

Europe
Info_agilent@agilent.com

Asia Pacific
Inquiry_lsca@agilent.com

This item is intended for Research Use Only, not for use in diagnostic procedures.

Information, descriptions, and specifications in this publication are subject to change without notice.

Published in the U.S.A., June 2013
Publication Number 5991-2385EN

Agilent Technologies