Genome Wide Expression Profiling of Paired Cancerous and Normal Breast Tissue

Kelly Li, Irene Lui, and Gary P. Schroth Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404

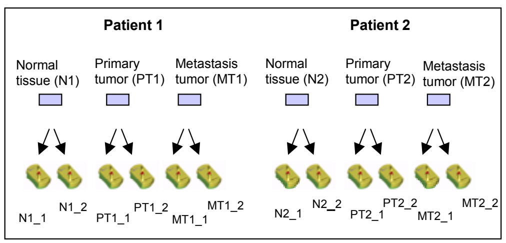
ABSTRACT

In this study we profile over 30,000 genes in paired normal and cancerous breast tissues on the Applied Biosystems Expression Array System. By comparing paired breast tissue specimens from the same patient we can identify genes that are differentially expressed in normal tissue, primary tumor, and metastasis tumor. Both biological replicates and technical replicates have been included in this study.

Using statistical data analysis tools, we have identified numerous genes that are up- or down-regulated in primary tumor and metastasis tumor compared to their normal control. Clear changes in gene expression patterns from normal to primary tumor and to metastasis tumor are observed. Gene expression level changes for genes detected by the Applied Biosystems Expression Array System have been validated using quantitative real-time PCR (TaqMan® probe-based Assays). This study demonstrates the use of microarrays for genome wide screening for gene expression in combination with quantitative real-time PCR for validating and extending results for genes of interest.

EXPERIMENTAL DESIGN

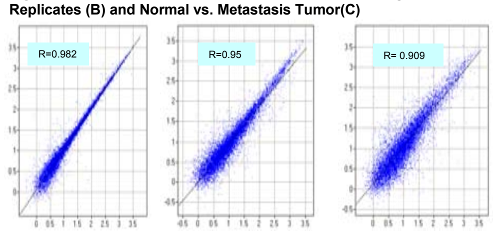
- **Two biological replicates**
- **Three disease status**
- **Two technical replicates**



Total RNA samples from 2 individual patients and 3 conditions (normal, primary and metastasis tumor) were processed into Digoxigenin (DIG) labeled cRNA using the Applied Biosystems RT-IVT Labeling Kit. Dig-labeled cRNAs were hybridized to the Applied Biosystems Human Genome Survey Microarray. Following hybridization a chemiluminescent detection assay was performed and the arrays were read on the 1700 Microarray Analyzer. Two experimental replicates were produced for each condition of both biological replicates.

RESULTS

Figure 1. Scatter Plots of Technical Replicates (A), Biological Replicates (B) and Normal vs. Metastasis Tumor(C)



A) Good reproducibility is seen in technical replicates when comparing expression from the same sample condition and from the same patient: B) Reduced correlation is seen in biological replicates by comparing expression of the same disease condition from 2 patients, suggesting individual differences of gene expression. C) Further reduced correlation is seen when comparing expression from different conditions of the same patient, indicating differential gene expression across different samples.

Table 1. Partial list of genes shown in Figure 2B

Gene	GeneName	MT2/N2	ANOVA P	Panther_family	Panther_MolecularFunction
TGFBI	transforming growth factor, beta-induced, 68kDa	0.06	0.0032	FASCICLIN DOMAIN CONTAINING PROTEIN	Cell adhesion molecule -> Other cell adhesion molecule
MUM-1	melanoma ubiquitous mutated protein	0.13	0.0030		
BCAR3	breast cancer anti-estrogen resistance 3	0.14	0.0017	CHAT - RELATED	Molecular function unclassified
MBIP	MUK-binding inhibitory protein	0.18	0.0000	MBIP PROTEIN	Select regulatory molecule -> Kinase modulator -> Kinase inhibitor
LIMK1	LIM domain kinase 1	0.20	0.0063	LIM KINASE-RELATED	Kinase -> Protein kinase -> Non-receptor serine/threonine protein ki
LUM	lumican	0.22	0.0002	LEUCINE-RICH REPEAT PROTEIN	Extracellular matrix -> Other extracellular matrix
CXCR6	chemokine (C-X-C motif) receptor 6	0.25	0.0003	G-PROTEIN COUPLED RECEPTOR-RELATED	Receptor -> G-protein coupled receptor
NPHP1	nephronophthisis 1 (juvenile)	0.31	0.0006	NEPHROCYSTIN	Cell junction protein
FN1	fibronectin 1	0.34	0.0084	TITIN-RELATED TENASCIN/FIBRONECTIN	Extracellular matrix -> Extracellular matrix linker protein
OXTR	oxytocin receptor	0.34	0.0092	G-PROTEIN COUPLED RECEPTOR	Receptor -> G-protein coupled receptor
MARK3	MAP/microtubule affinity-regulating kinase 3	0.35	0.0099	SERINE/THREONINE PROTEIN KINASE	Kinase -> Protein kinase -> Non-receptor serine/threonine protein ki
CXCL9	chemokine (C-X-C motif) ligand 9	0.36	0.0006	SMALL INDUCIBLE CYTOKINE-RELATED	Signaling molecule -> Cytokine -> Other cytokine
		0.39	0.0001	IMMUNOGLOBULIN KAPPA CHAIN VARIABLE REGION	Defense/immunity protein -> Immunoglobulin receptor family membe
CCL11	chemokine (C-C motif) ligand 11	0.40	0.0001	SMALL INDUCIBLE CYTOKINE A	Signaling molecule -> Chemokine
		0.45	0.0000	IMMUNOGLOBULIN KAPPA VARIABLE REGION	Molecular function unclassified
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CD	0.45	0.0069	CYCLIN-DEPENDENT KINASE 4 INHIBITOR-RELATED	Select regulatory molecule -> Kinase modulator -> Kinase inhibitor
CXCL10	chemokine (C-X-C motif) ligand 10	0.49	0.0019	SMALL INDUCIBLE CYTOKINE-RELATED	Signaling molecule -> Cytokine -> Other cytokine
ROR2	receptor tyrosine kinase-like orphan receptor 2	0.54	0.0025	TYROSINE PROTEIN KINASE	Kinase -> Protein kinase -> Tyrosine protein kinase receptor Recept
FKSG42	FKSG42	0.58	0.0030	GTPASE ACTIVATING PROTEIN HOMEOBOX PROTEINS	Transcription factor -> Homeobox transcription factor Select regulat
CCR1	chemokine (C-C motif) receptor 1	0.71	0.0084	G PROTEIN-COUPLED RECEPTOR	Receptor -> G-protein coupled receptor
		0.96	0.0000	IMMUNOGLOBULIN KAPPA LIGHT CHAIN VARIABLE REGION	Defense/immunity protein -> Immunoglobulin
		2.58	0.0078	ADP-RIBOSYLATION FACTOR AND RAB-RELATED	Select regulatory molecule -> G-protein -> Small GTPase
ANAPC4	anaphase-promoting complex subunit 4	4.07	0.0044	ANAPHASE-PROMOTING COMPLEX SUBUNIT 4	Ligase -> Other ligase
ITGB4	integrin, beta 4	5.21	0.0066	IG DOMAIN SUPERFAMILY INTEGRIN BETA CHAIN-RELATED TITIN-REL	Phosphatase -> Protein phosphatase Cell adhesion molecule Molecu
MAPRE1	microtubule-associated protein, RP/EB family, member 1	5.43	0.0062		
RAB33B	RAB33B, member RAS oncogene family	6.12	0.0075	SMALL GTPASE	Select regulatory molecule -> G-protein -> Small GTPase
NPY1R	neuropeptide Y receptor Y1	10.64	0.0095	G PROTEIN-COUPLED RECEPTOR	Receptor -> G-protein coupled receptor
		16.65	0.0039		
PPP2R5E	protein phosphatase 2, regulatory subunit B (B56), epsilon isofor	17.19	0.0036	PHOSPHATASE 2A REGULATORY SUBUNIT B-RELATED	Phosphatase -> Protein phosphatase
CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gam	19.19	0.0014	PROTEIN KINASE SERINE/THREONINE PROTEIN KINASE	Kinase -> Protein kinase -> Non-receptor serine/threonine protein ki
GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	20.18	0.0029	GABA-B RECEPTOR	Receptor -> G-protein coupled receptor

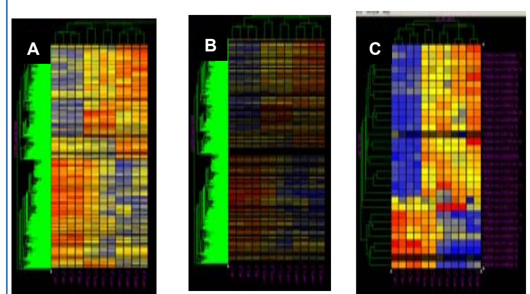
Using the Panther classification system, we identify 200 genes that are involved in various signal transduction pathways. These genes are among the 2,508 genes which have p-values <0.01.

Table 2. Identification of genes that are absent in normal tissue, but are expressed in breast cancers.

Gene Symbol	Gene Name	ANOVA P Value	PANTHER Molecular Function
CXCL11	chemokine (C-X-C motif) ligand 11	4.98E-11	Signaling molecule -> Cytokine -> Other cytokine
C1orf29	chromosome 1 open reading frame 29	1.77E-05	Molecular function unclassified
BLAME	B lymphocyte activator macrophage expressed	0.0030	Defense/immunity protein -> Immunoglobulin receptor family member Receptor
MMP11	matrix metalloproteinase 11 (stromelysin 3)	0.0194	Extracellular matrix -> Other extracellular matrix Protease -> Metalloprotease
CXCL9	chemokine (C-X-C motif) ligand 9	0.0006	Signaling molecule -> Cytokine -> Other cytokine
CKTSF1B1	cysteine knot superfamily 1, BMP antagonist 1	0.0030	Signaling molecule
INDO	indoleamine-pyrrole 2,3 dioxygenase	0.0000	Oxidoreductase -> Oxygenase
RAI3	retinoic acid induced 3	0.0001	Receptor -> G-protein coupled receptor
CCR5	chemokine (C-C motif) receptor 5	0.0003	Receptor -> G-protein coupled receptor
LTB	lymphotoxin beta (TNF superfamily, member 3)	0.0751	Signaling molecule -> Cytokine -> Other cytokine
HOP	homeodomain-only protein	0.0046	

This is a partial list of genes that are not detected in normal breast tissue samples, but are significantly increased in cancerous tissues across all 4 samples for the same condition. This approach can be used to identify potential markers that indicate changes in disease.

Figure 2. Clustering of Microarray Data



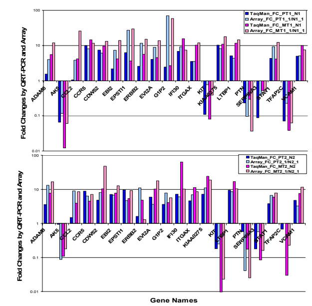
Two-dimensional hierarchical clustering analysis was performed using GeneSpring software (Silicon Genetics, Redwood City, CA). Red and blue color represents high and low expression respectively and yellow represents unchanged expression. A) A list of 2,508 genes generated from ANOVA test P<0.01 for all 12 arrays is used. First, all experimental and biological replicates are clustered together. All the arrays for normal and metastasis tumor were clustered into their own groups; however, primary tumor from the 2 patients showed differences in some gene expression. The primary tumor expression profile from patient 1 looks closer to metastasis, while that of patient 2 is more similar to normal breast tissue. The 2,508 genes are clustered into 2 main clusters; the top cluster indicates increasing expression and the bottom cluster shows decreasing expression, of genes in tumors relative to normal tissue. B) Among the 2,058 gene list, 200 genes are involved in signal transduction pathways as classified by the PANTHER system. The gene expression profiles for these genes are shown in this graph. C) Clustering analysis of 30 genes used in TaqMan® validation studies (See Figure 3).

Table 3. Panther classification suggests potential function for differentially expressed genes that are otherwise "unknown".

hCG	ANOVA P Value	PANTHER Family	PANTHER Molecular Function/Biological Process
hCG2036584	0.0014	CMF/LEK/CENP CELL DIVISION-RELATED KINESIN-RELATED	Nucleic acid binding -> Centromere DNA-binding protein Molecular function unclassified
hCG1747942.2	0.0006	DEAD BOX ATP-DEPENDENT RNA HELICASE	Cell cycle -> Mitosis -> Chromosome segregation Signal transduction -> Cell communication
hCG1998322	0.0025	EUKARYOTIC INITIATION FACTOR 4B	Nucleic acid binding -> Translation factor -> Translation initiation factor
hCG1787521.1	0.0027	HOMEOBOX PROTEINS	Transcription factor -> Homeobox transcription factor
hCG2014613	0.0026	HYPOXIA INDUCED GENE 1-RELATED	Cell structure and motility
hCG2000418.1	0.0047	HYPOXIA INDUCED GENE 1-RELATED	Biological process unclassified
hCG19910.3	0.0036	HYPOXIA INDUCED GENE 1-RELATED	Cell cycle -> Mitosis -> Chromosome segregation Signal transduction -> Cell communication
hCG1995119	0.0006	INTEGRIN BETA 1 BINDING PROTEIN (MELUSIN) RELATED	Signaling molecule -> Other signaling molecule
hCG1731660.2	0.0061	INTERFERON INDUCIBLE TRANSMEMBRANE PROTEIN	Interferon-mediated immunity Nucleoside, nucleotide and nucleic acid metabolism -> mRNA
hCG2040233	0.0001	INTERFERON INDUCIBLE TRANSMEMBRANE PROTEIN	Nucleoside, nucleotide and nucleic acid metabolism
hCG40171.2	0.0081	INTERFERON REGULATORY FACTOR	Transcription factor -> Other transcription factor
hCG2036709	0.0000	KRAB BOX ZINC FINGER TRANSCRIPTION FACTORS	Transcription factor -> Zinc finger transcription factor -> KRAB box transcription factor
hCG41814.2	0.0004	KRAB-RELATED C2H2-TYPE ZINC-FINGERS	Transcription factor -> Zinc finger transcription factor -> KRAB box transcription factor
hCG18392.3	0.0066	MYB	Transcription factor -> Other transcription factor
hCG40815.3	0.0018	SERINE/THREONINE PROTEIN KINASE	Kinase -> Protein kinase -> Non-receptor serine/threonine protein kinase
hCG1641561.3	0.0026	THYMIDYLATE KINASE	Kinase -> Nucleotide kinase
hCG1995887.1	0.0055	TUBULIN	Cytoskeletal protein -> Microtubule family cytoskeletal protein -> Tubulin

This is a partial list of "unknown" genes that are significantly differentially expressed (p-values of <0.001), based on expression data from Applied Biosystems microarrays. PANTHER classification is able to assign molecular function and biological processes to many of these "unknown" genes. This approach can be very useful for approach can be used to identify potential markers that indicate changes in disease.

Figure 3. Confirmation of differential expression detected by Applied Biosystems Expression Array System using quantitative real-time PCR (TaqMan® probe-based Assays)



Differential gene expression was determined by ANOVA testing across all 3 conditions (normal, primary and metastasis tumor) and for all 4 replicates (both technical and biological). To confirm the expression detected by the microarrays, we conducted quantitative real-time PCR analysis comparing fold change between tumor (primary or metastasis) and control (normal). Selected assays were ordered from the Applied Biosystems TaqMan® Gene Expression Assays. For microarrays, gene expression fold change was calculated as a ratio of tumor sample to matched normal tissue after median global normalization. For TaqMan® data, fold change was calculated using relative quantitation, or DDCt method, where DDCt = DCt tumor-DCt normal (DCt is GAPDH normalized Ct). Fold changes seen by TaqMan® assays and by the Expression Array System are similar. Panel A and B respectively represent data from patient 1 and patient 2 .

CONCLUSIONS

Using the Applied Biosystems Expression Array System, we profiled over 30,000 genes in paired normal and cancerous breast tissue. We have identified statistically significant gene expression changes when comparing normal and cancerous breast tissues. We also have demonstrated the utility of our array system in the identification of potential biomarkers and exploration of novel gene functions using PANTHER.

Selected TaqMan® probe-based assays confirm detected gene expression changes observed on the microarray. Genome wide expression screening using the Expression Array System can be easily linked with quantitative real time PCR for validating results for genes of interest. These TaqMan® assays can also be used to extend the array observations across many new samples.

TRADEMARKS/LICENSING

For Research Use Only. Not for use in diagnostic procedures.

Applied Biosystems is a registered trademark and AB (Design) and Applera are trademarks of Applera Corporation or its subsidiaries in the US and/or certain other countries.

All other trademarks are the sole property of their respective owners.

© 2004 Applied Biosystems. All rights reserved.