A Gene Expression Signature that Can Predict Green Tea Exposure and Chemopreventive Efficacy of Lung Cancer in Mice

Yan Lu,1 Ruisheng Yao,1 Ying Yan,1 Yian Wang,1 Yukihiko Hara,2 Ronald A. Lubet,3 and Ming You1

1Department of Surgery and The Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri; 2Food Research Laboratories, Mitsui Norin Co, Shizuoka, Japan; and 3Chemoprevention Branch, National Cancer Institute, Bethesda, Maryland

Abstract
Green tea has been shown to be a potent chemopreventive agent against lung tumorigenesis in animal models. Previously, we found that treatment of A/J mice with either green tea (0.6% in water) or a defined green tea catechin extract (polyphenon E; 2.0 g/kg in diet) inhibited lung tumor tumorigenesis. Here, we described expression profiling of lung tissues derived from these studies to determine the gene expression signature that can predict the exposure and efficacy of green tea in mice. We first profiled global gene expressions in normal lungs versus lung tumors to determine genes which might be associated with the tumorigenic process (TUM genes). Gene expression in control tumors and green tea–treated tumors (either green tea or polyphenon E) were compared to determine those TUM genes whose expression levels in green tea–treated tumors returned to levels seen in normal lungs. We established a 17-gene expression profile specific for exposure to effective doses of either green tea or polyphenon E. This gene expression signature was altered both in normal lungs and lung adenomas when mice were exposed to green tea or polyphenon E. These experiments identified patterns of gene expressions that both offer clues for green tea’s potential mechanisms of action and provide a molecular signature specific for green tea exposure. (Cancer Res 2006; 66(4): 1956-63)

Introduction
Lung cancer is the leading cause of cancer death in the U.S. (1). Exposure to tobacco is involved in 90% of lung carcinomas. Smokers’ risk of lung cancer is 20 times that of persons who have never smoked (2). Former heavy smokers retain an elevated risk for lung cancer even decades after they stopped smoking. One potential strategy to prevent lung cancer in high-risk populations is to use chemopreventive agents to regress existing intraepithelial neoplastic lesions, prevent the progression of these lesions to cancer, or inhibit the development of new lesions (3). Green tea contains flavanols or catechins including epigallocatechin gallate (EGCG; refs. 4–6). These polyphenols have various biological activities including antioxidation, modulation of enzyme systems for metabolizing chemical carcinogens, inhibition of nitrosation reactions, scavenging of activated metabolites of chemical carcinogens, and inhibition of tumor promotion (4–6). Although epidemiologic studies on the cancer-preventive effects of tea produced inconsistent results, a subset of studies suggested that tea consumption might reduce the risk of human lung, skin, breast, and gastrointestinal cancers (4–6). A prospective cohort study over 10 years in Japan showed that the consumption of 10 or more cups of green tea per day delayed the onset of cancer in both smokers and never smokers (7).

Preclinical studies have shown the inhibitory action of green tea or green tea extracts against tumorigenesis on different organ sites such as skin, lung, oral cavity, esophagus, stomach, small intestine, colon, liver, pancreas, and mammary gland (8–10). Green tea and one of its components, EGCG, has been shown to inhibit 4-(methylthio)nitrosamino)-1-(3-pyridyl)-1-butanone (NNK)–induced mouse lung tumorigenesis by 63% and 28%, respectively (11). We have recently shown the chemopreventive efficacy of green tea and polyphenon E in A/J mice (12, 13). In the green tea study, administration of green tea as the sole drinking source beginning 1 week after carcinogen administration significantly reduced tumor multiplicity in A/J mice (ref. 12; Fig. 1). Tumor multiplicity for the mice treated with a carcinogen was 5.1 and decreased to 2.4 in mice treated with green tea (P < 0.0001). In the polyphenon E study (13) treatment of mice with 2% of polyphenon E in diet caused a significant decrease in tumor multiplicity (10.8 tumors/mouse in controls to 5.9 tumors/mouse in those treated with polyphenon E; P < 0.05). The lung tissues derived from these studies were used to determine the potential mechanisms of action and gene expression signature that can predict both the efficacy and pharmacodynamics of green tea in mice.

Although there are many suggested mechanisms of action for green tea, including the induction of apoptosis and cell cycle arrest, the exact mechanisms for the action of green tea are not yet clear. Expression profiling is a powerful technique to uncover patterns or pathways indicative of potential mechanisms of action. Additionally, we will identify gene expression profiles that may be predictive of the chemopreventive efficacy of green tea for lung cancer, as well as uncover a specific expression signature for green tea exposure. The results from our study could be used for either clinical trials or perhaps in epidemiologic studies.

Materials and Methods

Reagents
Benzo(a)pyrene [B(a)P; 99% pure] and tricaprylin were purchased from Sigma Chemical Co. (St. Louis, MO). NNK (99% pure) was from Chemsyn Science Laboratories (Lenexa, KS). Bulk green tea extract powder was obtained from the National Cancer Institute. Polyphenon E was obtained from Tokyo Food Techno Co., Ltd. (Tokyo, Japan). Chemical carcinogens were prepared immediately before use in bioassays; NNK was dissolved in warmed PBS, and B(a)P was prepared in tricaprylin.
Chemoprevention Studies with Green Tea

A/J mice were obtained from The Jackson Laboratory (Bar Harbor, ME). These mice were randomized into four groups, two each (of males and females) for green tea and polyphenon E. For the green tea groups, mice were given two i.p. injections of NNK (100 mg/kg) 1 week apart. Beginning 1 week after the final injection of carcinogen, mice in group 2 were given a solution of 0.6% green tea as their sole source of drinking fluid until the end of the experiment. Group 1 received deionized water. For polyphenon E groups, mice received a single dose of B(a)P (100 mg/kg body weight) in 0.2 mL tricaprylin by i.p. injection. One week after giving B(a)P, mice in group 4 were fed with 2% of polyphenon E in AIN-76A-purified powder diet (Dyets, Inc., Bethlehem, PA) for 20 weeks. Group 3 mice received control AIN-76A powder alone. Fluids and food were available ad libitum. The experiment was terminated 20 weeks following exposure to carcinogens by carbon dioxide asphyxiation (Fig. 1). Portions of tumor and normal tissues were quickly frozen in liquid nitrogen and then reserved at −80°C until use. The remaining lung was fixed in Tellyesniczky’s [90% ethanol (70% v/v), 5% glacial acetic acid, 5% formalin (10% v/v buffered formalin)] solution overnight, followed by 70% ethanol.

RNA Isolation and Amplification

Total RNA from normal and tumor lung tissues of tea-treated mice and the controls were isolated with Trizol (Invitrogen, Carlsbad, CA) and purified using the RNeasy Mini Kit and RNase-free DNase Set (Qiagen, Valencia, CA) according to the manufacturer’s protocols. In vitro transcription-based RNA amplification was done on each sample. cDNA for each sample was synthesized using a Superscript cDNA Synthesis Kit (Invitrogen) and a T7-oligo(dT)24 primer: 5'-GGCCAGTGAATTGTAATACGACT-3'. The cDNA was cleaned using phase-lock gel (Fisher Scientific, Pittsburgh, PA) phenol/chloroform extraction. Then, the biotin-labeled cDNA was transcribed in vitro from cDNA using a BioArray High-Yield RNA Transcript Labeling Kit (ENZO Biochem, New York, NY) and purified, again using the RNeasy Mini Kit.

Microarrays

RNA samples were further purified, labeled, and processed according to standard manufacturer’s recommendations. Singleton cRNA preparations were produced from 30 µg of total RNA from each specimen and 10 µg equivalent aliquots were hybridized to each Affymetrix oligonucleotide array (Santa Clara, CA). The labeled cRNA from NNK-induced mice was hybridized onto the Murine Genome U74Av2 Array (MG-U74Av2), which consists of >12,000 genes and expressed sequence tags on one array. The labeled cRNA from B(a)P-induced mice was hybridized onto the Mouse Genome 430A 2.0 Array (ME430Av2), which contains 22,960 genes and expressed sequence tags on one array. Arrays were then scanned and digitized. Sixteen slides from the NNK-induced model and 15 slides from the B(a)P-induced model were obtained for data analysis. In the B(a)P-induced model, one slide from a tumor tissue treated with polyphenon E had poor data quality and thus was excluded from data analysis.

The raw fluorescence intensity data within CEL files were processed with Robust Multichip Average algorithm (14), as implemented with R packages from Bioconductor.4 This algorithm analyzes the microarray data in three steps: a background adjustment, a quantile normalization, and finally, a summation of the probe intensities for each probe set using a log scale linear additive model for the log transform of (background corrected, normalized) PM intensities.

RT-PCR

To evaluate the reliability of the array results, 12 genes were randomly selected from the genes differentially expressed between normal and tumor tissues in the microarray assay for further confirmation by real-time PCR. Two micrograms of total RNA per sample, collected as described above, were converted to cDNA using the SuperScript First-Strand Synthesis system for RT-PCR (Invitrogen). RT-PCR assay was done using the SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA). One microliter of cDNA was added to a 25 µL total volume reaction mixture containing water, SYBR Green PCR Master Mix, and primers. Each real-time assay was done in duplicate on Stratagene Mx3000. Data was collected and analyzed on the Stratagene Mx3000 software. GAPDH was used as an internal standard. The GAPDH value, a reflection of the number of cycles needed to reach a threshold of fluorescence was subtracted from the cycle value for the individual gene whose expression was being assessed.

Data Analyses

Matching mouse probes from two microarray systems. Because two microarray platforms were used, the probe sets should be corresponded. The batch query tool provided by Affymetrix5 was used for the correspondence between probe sets (15). There are a total of 8,904 pairs of probe sets corresponding to each other (representing the same gene) on the two microarray systems MG-U74Av2 and ME430Av2. All of the following comparisons between the two mouse models are based on these probe set equivalents.

Identifying differentially expressed genes. The following ANOVA model was used to test if a gene has significantly different transcription levels between different tissues and/or tea treatment in the two mouse models. Let yijk be the gene expression level of gene n from tissue i, treatment j and sample k. The gene expression level yijk can be expressed as,

\[ y_{ijk} = \mu_{n} + C_{i_{n}} + T_{j_{n}} + CT_{ij_{n}} + \epsilon_{ijk} \]

where i = 1 or 2, represents cancer or normal tissue; j = 1 or 2, represents tea treatment or control; k = 1 to 4, represents replicates in each group. The gene effect \( \mu_{n} \) captures the overall mean expression level of gene n across the tissues and tea treatments. The term \( C_{i_{n}} \) accounts for gene-specific tissue effects representing overall differences between two tissues. The term \( T_{j_{n}} \) accounts for tea effects that capture overall differences between tea

---

4 http://www.bioconductor.org/.
5 http://www.affymetrix.com/analysis/netaffx/batch_query.affx.
treatment and control in the samples. The term $CT_{ij}$ accounts for the interaction effect between tissue and tea treatment. The term $e_{ijkn}$ represents random errors. For each gene, we did 10,000 permutation tests and obtained empirical $P$ values for each variance component. The ANOVA and permutation tests were implemented using the R statistical package (16).

Pathway analysis. The visualization tool GenMAPP$^6$ was used to illustrate pathways containing differentially expressed genes. The differential gene expression was based on tea treatment versus nontreatment expression change by the ANOVA analysis ($P < 0.005$).

Discriminant analysis. $k$-nearest neighbors ($k$-NN) algorithm was used to select classifier. In the $k$-NN algorithm, a series of competitive models were built with a wide range of features (1-200 genes). Then, the predicting error rate of each of these models was estimated by using “leave-one-out” cross-validation approach. Finally, the best models were chosen based on their predicting error rates. Fisher’s test was used to evaluate the significance of these model predictors. The obtained classifier was then used for predicting tea status of the samples in two mouse models. GeneCluster 2.0 (17) was used to perform the discriminant analysis. We also attempted to find a common classifier to predict the tea status without reference to regular green tea or tea extraction. Because of the different microarray platforms, we only selected genes that were present in both microarray platforms. The gene expression data were integrated after standardizing the relative expression levels for both data sets. The gene expression levels for each gene were standardized separately to a mean $\mu$ and SD of 0 ± 1 in each data set. This standardization diminished the difference not only from microarray platforms but also from carcinogens. Again, the $k$-NN algorithm described above was applied to choose a classifier for integrated expression data. In order to cross-validate the classifier chosen by $k$-NN algorithm, four other clustering methods were used to predict the tea status of samples, which were implemented in Gene Expression Pattern Analysis Suite V1.1.$^7$

Results

Gene expression changes associated with tumorigenesis which were reversed by green tea. Gene expression analyses using Affymetrix microarrays showed a significant difference between mouse lung tumors and normal lungs. Specifically, 2,738 genes were differentially expressed when comparing normal lung tissue and B(a)P-induced lung tumors with 1,329 up-regulated and 1,409 down-regulated genes ($P < 0.005$). Similarly, 3,712 genes were differentially expressed when comparing normal

$^6$ http://www.genmapp.org/


Figure 2. Gene expressions altered by tea treatments ($P < 0.005$). A, genes underexpressed in tumors and restored at the normal expression levels after tea treatments.
lungs and NNK-induced tumors (1,721 up-regulated and 1,991
down-regulated genes; \( P < 0.005 \)). Comparing the tissue effects of
these genes in both chemically induced mouse lung tumor
models, we found that the expression levels of 2,036 significant
genes were consistently altered in both models, including 946
up-regulated and 1,090 down-regulated genes. This result
indicates that the gene expression changes, and presumably,
the mechanism of lung carcinogenesis for both B(\( \alpha \))P and NNK
are largely similar. Most of the common altered genes in both
models are related to several important pathways of lung
carcinogenesis (Supplementary Table S1). Interestingly, a subset
of these differentially expressed genes in tumors was reversed
towards levels found in normal lungs when animals were treated
with either green tea or polyphenon E. Figure 2A and B show 88
such genes in the NNK model, 25 of them are most likely
associated with the tumorigenic process from their functions,
including 19 underexpressed, and 6 overexpressed genes in
tumors. Detailed information for these genes is listed in
Supplementary Table S2. Dual specificity phosphatase 1 (Dusp1),
DNA topoisomerase I (Top1), Fyn proto-oncogene (Fyn),
regulator of G protein signaling 2 (Rgs2), G1 to S phase
transition 1 (Gept1), heat shock protein 8 (Hspa8), E26 avian
leukemia oncogene 2, 3' domain (Ets2), cyclin-dependent kinase
inhibitor 2C (Cdkn2c), and v-abl Abelson murine leukemia
oncogene 1 (Ab11) are related to cell cycle. In B(\( \alpha \))P mouse
model, we found that seven genes met the criteria, including
three overexpressed genes (Rbbp6, Tnfrsf9, and Kif20a) and four
underexpressed genes (Ccbl2, Drg1, Sf3a2, and Ryk) in tumors.
Because these alterations in expression seemed to reverse the
levels observed during tumorigenesis, they are candidates to be
involved in the mechanism by which green tea or polyphenon E
blocks tumorigenesis.

GenMAPP is a novel tool for visualizing expression data in the
context of biological pathways (18). We imported our data set
into the program and illustrated the pathways modulated by
treatment with green tea. We found that genes in cell cycle
pathways were significantly altered by tea treatment in both mouse
models. Figure 3A represents the cell cycle pathway following
treatment with green tea. Following treatment with green tea,
genes in S phase and G2 phase were down-regulated with tea
treatment; genes in G1 phase were up-regulated. In addition, green
tea seems to affect the inflammatory response pathway (Fig. 3B).

Identification of a green tea gene expression signature.
Using \( k \)-NN algorithm (\( k = 3 \); Fisher’s exact test, \( P = 1.5 \times 10^{-5} \)),
we successfully selected 14 and 91 gene expression sets that are most closely associated with green tea and polyphenon E exposure, respectively, to readily discriminate lung tissues (normal lungs and lung tumors) exposed to polyphenon E or green tea from those of vehicle control mice. When we did hierarchical cluster analyses using the 14-gene and 91-gene classifiers, the tissues with green tea exposure clustered together and there are no classification errors in either treatment. The results were displayed by TreeView (Fig. 4A and B). Furthermore, we identified a classifier composed of 17 genes which are most closely associated with tea status from all of the 31 samples (Fisher’s exact test, \( P = 3.3 \times 10^{-9} \)) using the integrated expression data. This 17-gene classifier could correctly discriminate from all samples (normal lung or lung adenomas) of mice exposed to both forms of green tea (green tea and polyphenon E). To further evaluate the performance of the 17-gene classifier obtained by the \( k \)-NN algorithm, we comparatively applied four other clustering methods to integrated expression data: (a) \( k \)-mean with Pearson correlation coefficient, (b) a self-organizing map with Gaussian function neighborhood, (c) support vector machines with linear kernel, and (d) hierarchical clustering. All four methods made no misclassification errors. Figure 4C presents the results of hierarchical clustering.

**Confirmation of gene expression data using semiquantitative RT-PCR.** To confirm the significance of the genes identified to be differentially expressed using microarrays, we did RT-PCR analysis. Of the selected genes, nine of them (Lamr1, Sftpb, Clu, P4ha1, U46068, Hba-a1, Car4, Aldh2, and Pon1) were identified as
having significant tissue effects in both mouse models by ANOVA analysis, whereas Ngef, Cyr61, and Scgb1a1 were identified in either mouse model \( (P < 0.005) \). All of them were confirmed by RT-PCR. Figure 5 is the comparison of the fold changes produced by DNA microarrays with the relative expression ratio obtained from RT-PCR.

**Discussion**

In this report, we did gene expression analyses on both normal lungs and tumors treated with either green tea or polyphenon E to examine mechanistic questions regarding the efficacy of green tea or polyphenon E. We showed that a subset of the differentially expressed genes in tumors treated with green tea were reversed towards levels found in normal lungs, suggesting that they might be involved in mediating the chemopreventive effect of green tea or polyphenon E. We also identified a green tea gene expression signature that can readily discriminate green tea–treated lung tissues from those without exposure to green tea in mouse models.

As can be seen in Fig. 1, treatment with either polyphenon E (2.0% in diet) or green tea (0.6% in water) administered beginning 1 week following the last dose of carcinogen, decreased lung tumor multiplicity by 45% and 55%, respectively. Presuming consumption of 5 mL of water and a similar amount of feed, a 25 g animal would consume roughly 1.4 g/kg of green tea and roughly 5.0 g/kg of tea polyphenols. Given the fact that polyphenols represent roughly 15% to 20% of the dry weight of

---

**Figure 4.** Hierarchical clustering of mice treated with either green tea or polyphenon E. A, 15-gene classifier from animals treated with green tea. B, 91-gene classifier from mice treated with polyphenon E. C, 17-gene classifier from the integrated gene expression data for all 31 samples in animals treated with either green tea or polyphenon E.
We first profiled global gene expressions in normal lungs versus lung tumors to determine genes which might be associated with the tumorigenic process. The obvious rationale to such an approach is that these genes are mechanistically likely to be involved in the mechanism by which green tea or polyphenon E blocks tumorigenesis. As can be seen in Fig. 2, there are a substantial number of genes which met this criteria in the NNK model. However, the genes which were modulated by green tea versus polyphenon E had limited overlap. When we took the results into GenMAPP, we found that although both agents affected cell cycle related genes, they seemed to preferentially affect different specific genes. In addition to affecting cell cycle related genes, green tea but not polyphenon E seemed to affect genes related to the inflammatory pathway. At this time, the identified genes and pathways must be considered candidate genes requiring further specific examination. EGCG, the major polyphenol from green tea, can inhibit DNA methyltransferase (DNMT) activity and reactivate methylation-silenced genes in cancer cells (20). The average level of mRNA for DNMT2 was significantly lower in hepatocellular carcinomas, colorectal cancers, and stomach cancers than in non–cancerous tissue (21, 22). In our results, DNMT2 was also found underexpressed in lung tumors and green tea was able to reverse the gene expression towards levels in normal lungs.

As mentioned above, the preclinical studies showing the efficacy of tea and tea extracts as well as the more limited epidemiologic data have encouraged the use of clinical trials employing tea or tea extracts (9). Therefore, another major objective of this study was to identify potential pharmacodynamic markers which might be useful for clinical trials with this class of agents. Although there are genes that were modulated only in tumors or in histologically normal tissue, we queried the array data to define genes which were modulated in both tumors and histologically normal tissues. We initially defined potential pharmacodynamic markers for individual treatment agents, e.g., green tea or polyphenon E. Therefore, in Fig. 4A and B, we have defined genes for either green tea or polyphenon E that can differentiate tissues exposed to effective levels of an agent versus control. As can be seen, we can readily differentiate histologically normal or tumor tissue treated with green tea from control tumors or control normal lungs. Similarly, we found a group of genes that differentiate histologically normal or tumor tissue treated with polyphenon E from control tumors or control normal lungs. These genes might be useful in examining a clinical trial employing either of these specific agents. The fact that these gene candidates can be used in either histologically normal lung tissues or lesions would seem to be a plus because it allows one to use samples from either set of tissues. It will be of some importance to determine whether these samples are relevant for testing bronchial washings as well because they may be the most readily accessible tissue from a phase II study in lung (23, 24). Finally we mined the data to determine whether we could determine a more limited number of genes that would differentiate tea exposed (green tea or polyphenon E) normal tissues and lesions from control normal tissues and lesions.

The advantage of these potential pharmacodynamic markers is that they might be applicable to a wide variety of tea compounds and they might prove applicable to epidemiologic studies assessing tea consumption where the specific makeup of the tea, caffeinated...
or decaffeinated, may not be known. Thus, such a generalized set of markers might have some general applicability.

One should be aware of certain aspects of these pharmacodynamic markers. Firstly, because the gene changes are not necessarily involved in the mechanism of action of these agents, they are not necessarily efficacy end points. Thus, they only directly associated with the pharmacodynamic/physiologic end point (gene modulation) and that at doses which achieved these changes in an animal that you achieved efficacy. Therefore, if you fail to observe these changes in a clinical trial, you may feel that it is unlikely that the dose employed will be effective. Secondly, multiple genes should be examined simultaneously because some genes which are modulated in animal models by a given agent may not be similarly modulated in a human. However, with various genes, this is less likely to be a problem. Finally, these genes become potential pharmacodynamic markers for the agent in tissues other than the tissue examined. Thus, genes modulated in the lung by tea might be similarly modulated by tea in other tissues, which remains to be determined in these other tissues.

In summary, we found multiple gene changes associated with lung tumorigenesis. Interestingly, both green tea and tea polyphenols could reverse certain of these gene changes associated with tumorigenesis. However, the specific genes altered are significantly different for green tea and polyphenol E, implying potentially different mechanisms of action. Finally, we defined potential pharmacodynamics for both green tea alone and polyphenol E alone and additionally defined genes which were modulated by both treatments. We feel that these genes are potentially useful for ongoing and proposed clinical trials using tea or tea extracts in humans.

Acknowledgments

Received 9/7/2005; revised 11/28/2005; accepted 12/5/2005.

Grant support: NIH grants (P01 CA94964 and N01-CN-43308).

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

References

A Gene Expression Signature that Can Predict Green Tea Exposure and Chemopreventive Efficacy of Lung Cancer in Mice

Yan Lu, Ruisheng Yao, Ying Yan, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/4/1956

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/02/16/66.4.1956.DC1

Cited articles
This article cites 23 articles, 6 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/4/1956.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/66/4/1956.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

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

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/66/4/1956.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.