SUPPLEMENTARY METHODS

Robust Gene Expression Signature from Formalin-Fixed Paraffin-Embedded Samples Predicts Prognosis of Non-Small-Cell Lung Cancer Patients Yang Xie et al.

RNA extraction. The extraction of RNA from tissue samples was done by a proprietary procedure of Response Genetics, Inc. (United States Patent Application 20090092979) designed to optimize the yield of higher molecular weight RNA fragments from FFPE specimens. In brief, tissue samples to be extracted are incubated for 16 hr at 50 C in a solution of 20 mM Tris, pH 8.0, 3.6 mM EDTA, 1% SDS containing 1 µg/mL of proteinase K. After mixing this solution with phenol/chloroform/isoamyl alcohol (PCI) and glycogen, the RNA is precipitated by addition of isopropanol, centrifuged down and washed with ethanol. A second precipitation is then carried out to further purify the RNA from contaminating DNA. The RNA pellet is mixed with 0.5% sarcosine - guanidine isothiocyanate (GITC), 20 mM DTT solution, 50 µL of 5 mM Tris and 50 µL of 2M sodium acetate followed by PCI extraction and precipitation of the RNA with glycogen and isopropanol. After washing it with ethanol, the RNA pellet is dissolved in 5 mM Tris, pH 7.5. The concentration of nucleic acids in the preparation (the yield) is measured by reading the absorbance at 260 nm in a Nanodrop 2000 (Thermo Scientific) and the purity of the RNA is estimated by the 260/280 absorbance ratio. A quality control procedure is then done (see below) to ascertain the amount of “higher quality” (250 bp and greater) RNA and the amount of DNA contamination.

RNA quality control. RNA isolated from FFPE tissue is generally quite extensively fragmented. We found that a certain level of 300 base-length or greater RNA fragments in the RNA preparation was critical to the success of microarray analysis of FFPE tissues. Therefore, each RNA isolation was checked to determine if its content of the higher molecular weight RNA reached the required level. The isolated RNA is converted to cDNA as previously described (1) except that oligo dT primers are used. Thus, only mRNA fragments containing a 3'-oligo A tail will be extended and converted to cDNA, thereby providing a starting point from which to measure fragment length. PCR amplification of β-actin mRNA was used to represent the total population of mRNA. Primers were chosen to amplify approximately 85 base-length segments of the β -actin gene representing locations approximately 300 and 400 bases from the 3'-end of the mRNA. The PCR primer sequences were as follows: for the segment 206-293 bases from 3' end, GCCACCCCACTTCTCTCTAAGG, ATAATTTACACGAAAGCAATGCTATCAC, 6FAM-ATGGCCCAGTCCTCTCCCAAGTCCA; for the segment 322-407 from 3' end, CATCCCCCAAAGTTCACAATGT,CAATGCATCTCATATTTGGAATGACT, 6FAM-CAACAATGTGCAATCAAAGTCCTCGGC. PCR was carried out as previously described (1). Based on our experience, we judged the RNA preparation to be suitable for the microarray if the β-actin Ct for both amplicons was 31 cycles or less. To assess DNA contamination, PCR was performed on the RNA solution without first performing reverse transcription (*i.e.*, a "no-RT" control). Thus, any signal here would be due to DNA contamination. We judged a preparation to be suitable for the microarray if the β-actin Ct in the no-RT control was at least 3 cycles less than that of the cDNA preparation.

Hybridization of RNAs to oligonucleotide arrays. Total RNA was processed for analysis on the Affymetrix U133 plus 2.0 arrays according to Affymetrix protocols for first- and second-strand synthesis, biotin labeling and fragmentation. Genechip®Two-Cycle cDNA Synthesis kit from Affymetrix was used according to the manufacturer's protocol for amplification of signal. Samples were hybridized to the arrays overnight at 45°C, rotating at 60 r.p.m., and were washed and scanned as per the Affymetrix protocol.

Gene set enrichment and pathway analysis. The fold changes of expression values between RGS group 1 and group 2 was used to rank the genes and then applied to Gene set enrichment analysis (GSEA) software to test which gene sets are enriched in RGS group 1 and group 2. GSEA and Ingenuity Pathway Analysis were applied to the 1400-gene list to test which gene sets and pathways are enriched in the 1400 robust gene set.

Sweave report. In order to allow others to reproduce any or all parts of our statistical analyses, we used Sweave to generate our program and report. Sweave is a literate programming R package for reproducible research; it embeds R code inside LaTeX and replaces the code with the result of running the code. The complete Sweave report is included in the supplemental material.

Reference

1. Schneider S, Uchida K, Salonga D, Yochim JM, Danenberg KD, Danenberg PV. Quantitative determination of p16 gene expression by RT-PCR. Methods Mol Biol 2004;281: 91-103.