Supplementary Material and Methods

Cell lines culture and drug sensitivity testing to sorafenib

The cell lines used in this manuscript, except for A549, were established at the National Cancer Institute (NCI-H series) or at the Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center (HCC series). A549 was obtained from ATCC. The identity of each cell line was confirmed by DNA fingerprinting using Powerplex 1.2 (Promega). The reference fingerprints were from ATCC. More details are available elsewhere (1). All lines were cultured in RPMI 1640 supplemented with 5% heat-inactivated fetal bovine serum (Hyclone, Logan, UT) in 5% CO_2 with 100% humidity. Sorafenib stock solutions were prepared in DMSO and stored at −20°C. The drug was diluted in fresh medium before each experiment, and the final DMSO concentration was <0.1%. Growth inhibition was assessed using a modified MTT assay as described previously[2]. Briefly, 2×10^3 cells are plated in each well of 96-well flat-bottomed microtiter plates. Sorafenib was added the next day, and after 5-day incubation, a 2 mg/mL solution of MTT (50 μL) dissolved in RPMI 1640 was added to each well. The microtiter plates were incubated for 4 hours at 37°C. The visible absorbance of each well was measured using an automated plate reader (Molecular Devices, Sunnyvale, CA), and data were analyzed using the Divisa software (manuscript in preparation) to determine IC_{50} of sorafenib (2, 3).

Gene expression profiling of NSCLC cell lines

Global gene expression analysis of a panel of 68 NSCLC cell lines were performed using Illumina HumanWG-6 v3.0 expression beadchip (Illumina®, San Diego, CA). Total RNA was

extracted from snap-frozen tissues using the mirVanaTM miRNA Isolation Kit (Ambion, Austin, TX) according to manufacturer's protocol. Biotin labeled cRNA samples for hybridization were prepared by using Illumina Total Prep RNA Amplification Kit (Ambion Inc., Austin, TX). One microgram of total RNA was used for the synthesis of cDNA and followed by an amplification and biotin labeling. Each of 1.5 μg of biotinylated cRNAs was hybridized to Illumina HumanWG-6 v3.0 expression beadchip. Signals were developed by Amersham fluorolink streptavidin-Cy3 (GE Health care Bio-Sciences, Little Chalfont, UK). Gene expression data were collected by using Illumina bead Array Reader confocal scanner (BeadStation 500GXDW; Illumina Inc.). All steps from hybridization to generation of raw microarray data were processed at the University of Texas Southwestern Medical Center Microarray Facility, Dallas, Texas.

Gene expression profiling of samples prospectively collected in the BATTLE program

Core tumor biopsies were taken from each patient before treatment at either the primary lung tumor or a metastatic site. About one third of the core from each sample was used for total RNA extraction and global gene expression analysis. Gene expression profiles were available in 101/255 (40%) patients who were randomized and evaluable in the BATTLE trial including 47/105 (45%) patients treated with sorafenib. We excluded 3 of the 47 profiles generated from samples with no tumor or malignant cells detected on the H&E control section (4). Among the 44 remaining patients, 7 had a tumor with *EGFR* mutation, leaving 37 patients in whom the sorafenib sensitivity signature was tested.

RNA extracted from OCT-embedded tissue was purified using the RNeasy Mini Kit (Qiagen) including on-column DNase (Qiagen) digestion as described by the manufacturer's protocol. Quantification was done using a ND-1000 spectrophotometer (Nanodrop

Technologies). All RNAs were serially diluted in RNAse-free water to obtain a 250 pg/µL stock solution. RNA quality was ensured by analyzing separation trace of RNA using the RNA6000 PicoAssay for the Bioanalyzer 2100 (Agilent). Aliquots were prepared and stored at -80°C. The same RNA was used for all experiments as starting RNA for amplification. Each aliquot was used once.

RNA amplifications were performed using the WT-Amplification™ Pico (NuGEN) kit. For all experiments, the manufacturers' protocols were strictly followed. In contrast with other manufacturer protocols, the WT-Ovation™ Pico RNA Amplification System (NuGEN) is not based on T7 polymerase cRNA synthesis. NuGEN has designed a technique called Ribo-SPIA™, which is a three-step process that generates amplified cDNA from as little as 500 picograms of total RNA. First strand cDNA is prepared from total RNA using a unique first strand DNA/RNA chimeric primer mix. The primers have a DNA portion that hybridizes either to the 5' portion of the poly(A) sequence or randomly across the transcript. Reverse transcriptase extends the 3' DNA end of each primer generating first strand cDNA/mRNA hybrid. Second strand cDNA synthesis step generates double stranded products with RNA-DNA heteroduplex at one end. The third step is the DNA amplification, called SPIA™ amplification using a specific DNA/RNA chimeric primer, DNA polymerase and RNase H in a homogeneous isothermal assay that provides highly efficient amplification of DNA sequences. RNase H is used to degrade RNA in the DNA/RNA heteroduplex at the 5' end of the first cDNA strand. This results in the exposure of a DNA sequence that is available for binding a second SPIA™ DNA/RNA chimeric primer. DNA polymerase then initiates replication at the 3' end of the primer, displacing the existing forward strand. The RNA portion at the 5' end of the newly synthesized strand is again removed by RNase H, exposing part of the unique priming site for initiation of the next round of

cDNA synthesis. The process of SPIA™ DNA/RNA primer binding, DNA replication, strand displacement and RNA cleavage is repeated, resulting in rapid accumulation of cDNA with a sequence complementary to the original mRNA. WT-Ovation™ Pico products (NuGEN) are labeled using the FL-Ovation™ cDNA Biotin Module V2 (NuGEN). Each labeled cRNA targets are synthesized according to manufacturer's protocols. The quantity and quality of the amplified cRNA or cDNA were assessed by a ND-1000 spectrophotometer (Nanodrop Technologies), and Agilent Bioanalyzer (Agilent Technologies), respectively.

All steps from hybridization to generation of raw microarray data were processed at the University of Texas MD Anderson Cancer Center Microarray and Affymetrix Facility. Hybridization mixtures were prepared according to Affymetrix procedures to accommodate 5 μg of cDNA targets from NuGEN amplification. Human Gene 1.ST platform from Affymetrix were hybridized, revealed and washed according to the Affymetrix protocol. Gene chips were scanned using a 7 G scanner (Affymetrix) and images (DAT files) were converted to CEL files using GCOS software (Affymetrix).

Development of the sorafenib sensitivity signature

Data analysis was performed using R packages in Bioconductor (http://www.bioconductor.org) (Supplementary Methods_Training the sorafenib sensitivity signature and Supplementary Methods_Testing the sorafenib sensitivity signature). Affymetrix raw data of microarrays were processed using quantile normalization, robust multi-array average (RMA) algorithm[5] and $\log_2$ transformed. Illumina BeadArray data were processed using the method of model-based background correction for BeadArrays (MBCB) developed elsewhere[6]. Only probes with gene symbol were considered.

The sorafenib signature was trained *in vitro* using gene expression profiling of 68 NSCLC wilt-type *EGFR* cell lines with available IC_{50} of sorafenib. Spearman correlation of IC_{50} with each individual probe expression level was computed. To address the multiple testing problems, histogram of p-values was plotted and false discovery rates (FDR) of genes were calculated according to BUM model[7]. The top 50 probes were selected (*ad hoc* choice) and used for a two-way hierarchical clustering of the cell lines used the Pearson's *correlation* distance between genes, Euclid distance between samples, and *Ward's* linkage method. To summarize the effect of the selected 50 genes, a Principal Component Analysis (PCA) was computed with the first two components. First principal component (PC) was correlated with IC_{50} of sorafenib using the Spearman correlation.

To test the signature, the first PC was computed in BATTLE samples after identifying the Human Gene 1.ST platform probesets corresponding to the 50 genes selected in cell lines using the Illumina HumanWG-6 v3.0 expression beadchip. Kaplan-Meier curves were used to compare progression-free survival in patients with high versus low first PC based on the median of the first PC. Log-rank test was used to test the difference in survival between the 2 groups. A univariate cox proportional hazards model was computed. All tests were two-sided. All statistical tests were two-sided, and p values of 0.05 or less were considered to be statistically significant.

Gene Set Enrichment Analysis (GSEA)

Functional analyses were performed using GSEA software v2.0.4 (5). Three required data inputs were generated: (1) Spearman correlation of IC_{50} with each individual probe included in the Human Gene 1.ST platform (2) a mapping file for identification of Human Gene 1.ST, and (3) a catalog of functional gene sets from Molecular Signature Database (MSigDB web site v3.82

released Oct 7, 2011, www.broad.mit.edu/gsea/msigdb/msigdb_index.html). Using the Spearman correlation of IC_{50} with each individual probe to rank them, a GSEA was performed. After collapsing the probesets into gene symbols, 19874 genes were considered. A total of 2509 gene sets were included in the analysis. Default parameters were used throughout (inclusion gene set size was set between 15 and 500 and the phenotype was permutated 1,000 times). Gene sets positively enriched are associated with resistance and gene sets negatively enriched are associated with sensitivity to sorafenib.

References

1. Gazdar AF, Girard L, Lockwood WW, Lam WL, Minna JD. Lung cancer cell lines as tools for biomedical discovery and research. J Natl Cancer Inst. 2010;102:1310-21.
2. Coldren CD, Helfrich BA, Witta SE, Sugita M, Lapadat R, Zeng C, et al. Baseline gene expression predicts sensitivity to gefitinib in non-small cell lung cancer cell lines. Mol Cancer Res. 2006;4:521-8.
3. Gandhi J, Zhang J, Xie Y, Soh J, Shigematsu H, Zhang W, et al. Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. PLoS One. 2009;4:e4576.
4. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin Cancer Res. 2013;19:279-90.
5. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-50.