Corresponding Author: Do-Hyun Nam, Raul Rabadan, Jeeyun Lee

Date: June 7, 2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   The sample size of pharmacological data from 462 patient-derived cells (PDCs) were as large as that from Cancer Cell Line Encyclopedia (CCLE, n=479), which were proven to be sufficient to extract statistically significant lineage-specific and gene-drug associations.

2. Data exclusions

   Describe any data exclusions.

   Drug sensitivity data (Area Under Curve, AUC) from non-converged or non-fitted dose-response curve according to the 4 parameter logistics were excluded.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   All drug screening data were obtained from technical replicates and reliably reproduced.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Tumor specimens had been randomly collected from patients with informed consents, and there has been no intended bias of allocation into experimental design. For in vivo experiments, mice were allocated to each group based on weight to ensure that all groups were within similar weight distribution.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   The investigators and authors have been consistently blinded to the group allocation during data collection and/or analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

R version 3.2.1, python 2.7.11, Matlab 2014b, Ayasdi 7.3.0.2, Burrows-Wheeler Aligner 0.6.2, Picard 1, Genome Analysis ToolKit 2.5.2, SAVI2, SAMtools 0.1.18, Excavator 2.2, STAR 2.4.0b, Chimerascan v0.4.3, Pegasus, PRADA 1.1, GraphPad Prism 6, ssGSEA2-2.2.1, dNetFS 1.0, STRING 10.0, ELDA

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

The patient-derived cell/mouse models were restricted on availability, and can be distributed only with appropriate material transfer agreements.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Immunoblot: p-EGFR 1:1000 (Cell Signaling, #3777), EGFR 1:1000 (abcam, ab52894), p-STAT3 1:1000 (Cell Signaling, #9145), STAT3 1:1000 (Cell Signaling, #8768), pERK 1:2000 (Cell Signaling, #9101), ERK 1:2000 (Cell Signaling, #9102), pAKT 1:1000 (Cell Signaling, #4060), AKT 1:1000 (Cell Signaling, #9272), pS6K 1:1000 (abcam, ab126818), S6K 1:1000 (abcam, ab131440), Beta-actin 1:5000 (Santa Cruz, sc-81178), HDAC4 1:1000 (R&D Systems, AF6205), SIK2 1:1000 (R&D Systems, AF5737), NRG1 1:1000 (R&D Systems, AF396NA), GAPDH 1:5000 (Cell Signaling, #2118).

Immunofluorescence: EGFR 1:250 (abcam, ab52894), pERK 1:250 (Cell Signaling, #9101).

Immunohistochemistry: pEGFR 1:400 (Cell Signaling, #3777), pAKT 1:100 (Cell Signaling, #4060), pERK 1:100 (Cell Signaling, #9101).

All antibodies were validated for usage on human and the corresponding assays by the distributor.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

SW480 cell line was obtained from ATCC

b. Describe the method of cell line authentication used.

We performed short tandem repeat (STR) analysis (AmpFISTR Identifier, Applied Biosystems) and matched the resulting data to ATCC STR database.

c. Report whether the cell lines were tested for mycoplasma contamination.

We used Universal Mycoplasma Detection Kit (ATCC 30-1012K) to verify absence of mycoplasma contamination in our cell line.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Balb/c nude mice (6~8 week-old female)

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Biospecimens from 429 patients were used in this study: 229 males, 199 females, and 1 unknown. Median age at diagnosis was 56 years, ranging from 18 to 83.