nature research

Corresponding author(s): Jinghui Zhang

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but 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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   We analyzed all genome and transcriptome data collected from diagnosis tumor samples available under the accession number cited in the Online Methods section, except two cases: one is a duplicated neuroblastoma case, and the other is an osteosarcoma case which was collected from a patient age > 40 yrs.

2. Data exclusions

   Describe any data exclusions.

   Besides the two cases excluded from the analysis as described above, WES data from another 23 osteosarcoma samples were included only for determining driver mutation prevalence but not other analyses (see section "Whole exome data analysis" of the Online Methods for details).

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   The western blot for KRAS novel isoforms shown in Extended Data Figure 6g and 6h were done in duplicates, with similar observations.

4. Randomization

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

   Not relevant. Patient samples were grouped by disease types in this analysis.

5. Blinding

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

   The investigators were not blinded to group allocation during the analysis. Patient samples were grouped by disease types. Detailed comparison was carried out among different type of tumors.

Note: all in vivo studies must report how sample size was determined and 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 in 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

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.

Public available softwares: bwa (v0.7.12) for WES data alignment; Bambino for SNV/Indel analysis from WES; GRIN, MuSiC (v0.1), MutSigCV (v1.4) for driver mutation analysis; HTSeq for expression analysis from RNA-seq; R packages (v3.3.2) for statistical analysis; ESTIMATE (v1.0.13) and CIBERSORT for immune infiltration estimation; HotNet2 for pathway analysis; Circos (v0.69) for circular genome visualization; ProteinPaint portal for genomic data visualization.
In house tools: STRONGARM for RNA-seq data alignment; CICERO for fusion detection from RNA-seq; Medal Ceremony for mutation pathogenicity analysis; ad hoc perl scripts were used in filtering genomic variants from CGI dataset.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

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 third party.

No unique materials were used.

9. Antibodies

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

Monoclonal antibody against human KRAS-N terminus (catalogue # H00003845-M02 Novusbio Littleton, CO), anti-beta-actin (catalogue # 4967, Cell Signaling Tech Danvers, MA), and anti-flag antibody (catalogue # TA50011, Origene Rockville, MD) were used for western blot.

10. Eukaryotic cell lines

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

293T cells purchased from ATCC (catalogue CRL-3216).

b. Describe the method of cell line authentication used.

The 293T cell line was purchased from ATCC (CRL-3216) and authenticated by ATCC.

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

We tested the mycoplasma contamination on 293T cells using MycoAlert kit (Lonza) and the result is negative.

d. If any of the cell lines used 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.

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

No animals were used.

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.

Patients were recruited through collaborations with the Children's Oncology Group (COG) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) project. Diagnosis for these patients include B-ALL (N=689, age information missing for 6 cases, median age: 5.4 yr), T-ALL (N=267, age information missing for 3 cases, median age: 9.3 yr), AML (N=210, median age: 9.4 yr), NBL (N=316, median age: 3.1 yr), WT (N=128, median age: 4.2 yr) and OS (N=89, median age: 14.4 yr). Gender (N=8, female N=3, male N=5) and race (N=6, Hispanic N=1, Caucasion N=4, Asian N=1) information were included for the patients with UV signature in Extended Data Figure 2. Other clinical information was not used for this study.