nature portfolio

Corresponding author(s): Ferdinandos Skoulidis, Martin Schuler

Last updated by author(s): Apr 22, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ 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 all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using Medidata Rave electronic data capture (EDC). Baseline tissue samples (fresh/archival) were analyzed using a targeted 648 gene DNA sequencing panel (Tempus xT ; Tempus Labs, Inc., Chicago, IL, USA) and a whole-transcriptome RNA assay (Tempus RS.v2). Post-treatment samples were assessed using plasma next-generation DNA sequencing: Resolution ctDx Lung assay (23 genes; Exact Sciences Corporation, Madison, WI, USA).
Data analysis	Statistical analysis was performed using the R stats package (v4.2.3; https://rpkgs.datanovia.com/). Time-to-event analysis (PFS and OS) was performed with Cox proportional hazards regression modeling using the survminer package (v0.4.9). Linear mixed-effects model analysis used the lmerTest (v3.1) and emmeans (v1.8.9) packages. Genetic variant annotation utilized data from the Cosmic Database Cancer Gene Census classification, and OncoKB.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE295032 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE295032).
Data sharing requests relating to data in this manuscript will be considered after the publication date and 1) this product and indication (or other new use) have been granted marketing authorization in both the US and Europe, or 2) clinical development discontinues and the data will not be submitted to regulatory authorities. There is no end date for eligibility to submit a data sharing request for these data.
Qualified researchers may submit a request containing the research objectives, the Amgen product(s) and Amgen study/studies in scope, endpoints/outcomes of interest, statistical analysis plan, data requirements, publication plan, and qualifications of the researcher(s).
In general, Amgen does not grant external requests for individual patient data for the purpose of reevaluating safety and efficacy issues already addressed in the product labelling. A committee of internal advisors reviews requests. If not approved, requests may be further arbitrated by a Data Sharing Independent Review Panel. Requests that pose a potential conflict of interest or an actual or potential competitive risk may be declined at Amgen's sole discretion and without further arbitration.
Upon approval, information necessary to address the research question will be provided under the terms of a data sharing agreement. This may include anonymized individual patient data and/or available supporting documents, containing fragments of analysis code where provided in analysis specifications.
Further details are available at the following: http://www.amgen.com/datasharing.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex and gender were not considered in this analysis of molecular determinants of clinical efficacy from the CodeBreaK 100 and CodeBreaK 200 datasets.
Reporting on race, ethnicity, or other socially relevant groupings	Socially relevant groupings were not considered in this analysis of molecular determinants of clinical efficacy from the CodeBreaK 100 and CodeBreaK 200 datasets. Despite the large cohort size, the subset of biomarker-evaluable patients is more limited, reducing the statistical power of molecular analyses. We cannot rule out potential bias that may result from analyzing a subset of the enrolled population and/or instances of informative censoring that could confound interpretation of biomarker data. Additionally, most analyses reported here are post hoc in nature and were not prespecified.
Population characteristics	Population characteristics were not considered in this molecular analysis from the CodeBreaK 100 and CodeBreaK 200 datasets.
Recruitment	The multicenter, single-group, open-label, phase 1/2 CodeBreaK 100 study (NCT03600883) enrolled patients ≥18 years old with KRAS G12C–mutated locally advanced or metastatic NSCLC after progression on prior therapies. Key eligibility criteria included locally advanced or metastatic NSCLC, KRAS G12C mutation confirmed through molecular testing, an Eastern Cooperative Oncology Group (ECOG) performance score of ≤2 (phase 1) and ≤1 (phase 2), measurable disease according to RECIST v1.1, and treatment with 1 or more prior systemic therapies unless intolerant of or ineligible for available therapies known to provide clinical benefit. Exclusion criteria included previous treatment with a KRASG12C inhibitor and active brain metastases. The randomized, multicenter, open-label, phase 3 CodeBreaK 200 study (NCT04303780) compared the efficacy of sotorasib versus docetaxel in patients with KRAS G12C–mutated advanced NSCLC. Enrolled patients were ≥18 years who had disease progression after previous platinum-based chemotherapy and PD-1 or PD-L1 inhibitor. Key eligibility criteria included histologically or cytologically documented, locally advanced, and unresectable or metastatic NSCLC, KRAS G12C mutation confirmed via central laboratory testing, tumor progression after receiving at least one previous systemic therapy for advanced disease, ECOG performance status of 0 or 1, and measurable disease according to RECIST v1.1. Patients with treated, stable brain metastases were eligible. Exclusion criteria included new or progressing untreated brain lesions or symptomatic brain lesions, previously identified oncogenic driver mutation other than KRAS G12C for which an approved therapy is available, previous treatment with docetaxel (with conditions), previous treatment with a direct KRASG12C inhibitor, systemic anticancer therapy within 28 days of study day 1, and therapeutic or palliative radiation therapy within 2 weeks of treatment initiation.
Ethics oversight	Institutional review board approval before study initiation and participating country regulatory authority approval were received, all patients provided written informed consent. See the Data Supplement for a full list of study sites and Institutional Review Board/Independent Ethics Committees.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Datasets included 317 biomarker-evaluable patients with previously treated advanced KRAS G12C–mutated NSCLC who enrolled in CodeBreaK 200 and 112 biomarker-evaluable patients treated in CodeBreaK 100. This analysis was not pre-specified, and was a post hoc exploratory analyses that were not intended to be descriptive.
Data exclusions	Inclusion criteria for biomarker-eligible patients were prospectively defined. Patients who provided appropriate consent were biomarker-evaluable if they met protocol inclusion criteria and had either genomic (baseline and/or treatment) or transcriptomic (baseline) data for baseline associations, ctDNA dynamics, or acquired resistance.
Replication	Replication was not applicable to this study as this analysis was based on two different datasets for validation from Phase 2 and Phase 3 to validate key findings.
Randomization	As previously reported in CodeBreaK 200, patients were randomly assigned (1:1) to either sotorasib or docetaxel in an open-label manner using interactive response technology. The randomization number was provided to the study center by the interactive response technology system. Randomization was stratified by the number of previous lines of therapy in advanced disease (1 vs 2 vs >2), ethnicity (Asian vs non-Asian), and history of CNS metastases (present or absent).
Blinding	The study was not blinded. Since the two treatment arms have distinct adverse events, blinding would be non-informative.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*NA*
Research sample	*NA*
Sampling strategy	*NA*
Data collection	*NA*
Timing	*NA*
Data exclusions	NA
Non-participation	*NA*
Randomization	*NA*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*NA*
Research sample	*NA*
Sampling strategy	*NA*
Data collection	*NA*
Timing and spatial scale	*NA*
Data exclusions	*NA*

Reproducibility	NA
Randomization	NA
Blinding	NA

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	NA
Location	NA
Access & import/export	NA
Disturbance	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	NA
Validation	NA

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	NA
Authentication	NA
Mycoplasma contamination	NA
Commonly misidentified lines (See ICLAC register)	NA

Palaeontology and Archaeology

Specimen provenance	NA
Specimen deposition	NA

Dating methods	NA

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	NA
Wild animals	NA
Reporting on sex	NA
Field-collected samples	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Phase 1/2 CB100 study (NCT03600883); Phase 3 CB200 study (NCT04303780)
Study protocol	CodeBreaK 100 (https://www.nejm.org/doi/full/10.1056/NEJMoa2103695); CodeBreaK 200 (https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(23)00221-0)
Data collection	The study was conducted across 148 academic and community centers in 22 countries: Australia, Belgium, Brazil, Canada, Denmark, Germany, Great Britain, Greece, Finland, France, Hungary, Italy, Japan, Netherlands, Poland, Portugal, Republic of Korea, Russia, Spain, Sweden, Switzerland, USA.
Outcomes	For CodeBreaK 100, the primary end point was objective response (complete or partial response) as assessed by blinded, independent, central radiologic review. Tumor response was assessed by independent central review according to RECIST, version 1.1, with the use of contrast-enhanced computed tomography or magnetic resonance imaging. Key secondary end points were duration of response, disease control (defined as complete response, partial response, or stable disease, according to RECIST, version 1.1; minimum time interval for the determination of stable disease, 5 weeks), time to response, progression-free survival, overall survival, and safety. Adverse events were graded with the use of the Common Terminology Criteria for Adverse Events, version 5.0. In accordance with the protocol, response-related end points were evaluated in patients who had received at least one dose of sotorasib and had at least one measurable lesion at baseline as assessed by independent central review according to RECIST, version 1.1. For CodeBreaK 200, the primary endpoint was progression-free survival by blinded independent central review per RECIST (version 1.1). Key secondary endpoints included overall survival, overall response rate, and patient-reported outcomes (change from baseline to week 12 for dyspnoea, cough, chest pain, global health status, and physical functioning). Other secondary endpoints were duration of response, disease control rate, time to response, safety, patient-reported outcomes (time to deterioration), and pharmacokinetics of sotorasib and its major metabolites. Progression-free survival 2 was an exploratory endpoint. Progression-free survival was defined as the time from randomisation until disease progression or death from any cause, whichever occurred first. Overall survival was defined as the time from randomisation to death from any cause. Overall response rate was defined as the sum of the confirmed complete response rate and confirmed partial response rate by BICR. Duration of response was measured from the first complete or partial response until progressive disease or death, whichever occurred first. Progression-free survival 2 was defined as the time from randomisation to disease progression on subsequent treatment, as assessed by an investigator, or death from any cause, whichever occurred first. Patient-reported outcomes were assessed using European Organization for Research and Treatment of Cancer Quality-of-life Questionnaire Core 30 (EORTC QLQ-C30) and Quality-of-life Questionnaire Core 13 (EORTC QLQ-LC13). We recorded adverse events from the first study dose through up to 30 days after the end of treatment and graded them according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0).

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	*NA*
Novel plant genotypes	*NA*
Authentication	*NA*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*NA*
Files in database submission	*NA*
Genome browser session (e.g. UCSC)	*NA*

Methodology

Replicates	*NA*

Sequencing depth	NA
Antibodies	NA
Peak calling parameters	NA
Data quality	NA
Software	NA

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	NA
Instrument	NA
Software	NA
Cell population abundance	NA
Gating strategy	NA

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	NA
Design specifications	NA
Behavioral performance measures	NA

Acquisition

Imaging type(s)	NA
Field strength	NA
Sequence & imaging parameters	NA
Area of acquisition	NA

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	NA
Normalization	NA
Normalization template	NA

Noise and artifact removal NA

Volume censoring NA

Statistical modeling & inference

Model type and settings NA

Effect(s) tested NA

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference NA

(See Eklund et al. 2016)

Correction NA

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity NA

Graph analysis NA

Multivariate modeling and predictive analysis NA