nature portfolio

Corresponding author(s): Peng Zhang

Last updated by author(s): Mar 20, 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	Clinical data was recorded in CRF (case report form) in paper version, restored in Execl (version 16.73, Microsoft Excel for Mac)
Data analysis	Clinical data were analyzed by SPSS software (version 26.0, IBM Corp, Armonk, NY). Exploratory analyses were conducted using R programing (version 4.2.0). Following RNA sequencing, the raw FASTQ files were trimmed using fastp (v0.23.4) and aligned to the GRCh38 reference genome using STAR (v2.7.9a) with default settings. After obtaining BAM files, read counts were summarized using featureCounts (v2.0.6) and TPM (Transcripts Per Million) values were generated using Salmon (v0.14.1). For RNAseq data, DESeq2 was used to calculate differential gene expression between sample groups. Immune cell infiltration was assessed using MCPCounter (v1.1.0), CIBERSORT (v1.01), Quantiseq (v1.10.0), and TIMER2.

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 OncoSG cohort data used in this study is publicly accessible through the OncoSG database (https://src.gisapps.org/OncoSG/) under accession number GIS031. The original RNA sequencing data have been archived in the Genome Sequence Archive at the National Genomics Data Center, under accession number HRA008462 (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA008462).

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	We declare that this clinical trial and following exploratory analysis do not involve sex and gender considerations for study
Reporting on race, ethnicity, or other socially relevant groupings	This is a single-center trial conducted at Shanghai Pulmonary Hospital in China, involving participants who are exclusively Chinese and representing the East-Asian population. No subgroup analyses regarding race , ethnicity or other socially relevant
Population characteristics	This single-arm phase II trial was conducted at Shanghai Pulmonary Hospital, which was registered at https:// www.clinicaltrials.gov before participant enrollment (NCT04685070).
Recruitment	Treatment naive inoperable stage III NSCLC harboring EGFR mutations, older than 18 years, with an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, were enrolled in this trial. The diagnosis of NSCLC was confirmed by immunohistochemical detections by endobronchial ultrasound (EBUS) or percutaneous lung biopsy. EGFR-mutant status was diagnosed by polymerase chain reaction. The clinical staging was confirmed by chest computed tomography (CT), and brain magnetic resonance imaging (MRI), positron emission tomography/computed tomography (PET/CT). Patients were required to have normal hematological indexes, qualified hepatic, renal, and pulmonary functions so that neoadjuvant therapy followed by radical resection could be tolerated.
Ethics oversight	The study was approved by the independent ethic committee in Shanghai Pulmonary Hospital (L20-427), and conducted in accordance with the Declaration of Helsinki (as revised in 2013).

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	The primary endpoint of this research was ORR, which was taken to calculate the sample size. Meta analysis calculated that the ORR of neoadjuvant the first-generation EGFR-TKIs for locally advanced NSCLCm in previous studies (EMERGING-CTONG 1103) was 52%. The ORR of Aumolertinib for advanced NSCLCm (APOLLO) was 68.9%. Therefore, the ORR was expected to be 70% in this trial, and the neoadjuvant the first-generation EGFR-TKIs was taken as the historical control group. It was calculated with $\alpha = 0.05$ (one-tailed) and the power of test $(1-\beta) = 80\%$. The sample size will be 45 as calculated. With a drop-off rate of 20%, the total sample size will be 56.
Data exclusions	Pregnant or breast-feeding patients, patients with unstable systemic disease (interstitial lung disease, pulmonary fibrosis, cardiovascular disease and so on), patients with any anticancer therapy outsides of this trial (EGFR-TKIs, chemotherapies, immunotherapies, and so on) were ineligible for this trial.
Replication	Bulk-RNA-seq data was not applicable for replication due to inability to acquire multiple fresh specimens from the same participants. However, the transcriptional signatures identified in our dataset were consistent with prior studies.
Randomization	None (single-arm study).
Blinding	None (single-arm study).

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	This single-arm phase II trial was conducted at Shanghai Pulmonary Hospital, which was registered at https://www.clinicaltrials.gov before participant enrollment (NCT04685070).
Study protocol	Participants received 16 weeks (4 cycles) administration of Aumolertinib (Hansoh Pharmaceutical Group Co, Ltd, Shanghai, China) (Aumolertinib 110mg, oral, once-daily). Chest CT were performed at the 8th and 16th weeks after treatment or before surgical treatment. All CT scans of participants were assessed by the same radiologist and 3 experienced thoracic surgeons via Response Evaluation Criteria in Solid Tumors measurement criteria (RECIST, version 1.1). Radiologically partial response or responder were defined as at least a 30% decrease in the sum of TLD versus baseline. If participants after neoadjuvant treatment could be resected completely, surgery could be performed within two weeks after Aumolertinib discontinuation. Adjuvant Aumolertinib treatment could be administrated within eight weeks postoperatively. We strongly recommend that the participants continue long-term adjuvant therapy for at least three years or until the disease progresses. However, if participants were still inoperable after 16 weeks neoadjuvant treatment, they should be treated based on multi-disciplinary therapy (MDT).
Data collection	The clinical data was record in the Case Report Form (CRF) in paper version and was stored in Excel. Besides, we collected the primary tumor tissue by percutaneous pulmonary biopsy, bronchoscopy biopsy or endobronchial ultrasound (EBUS) biopsy before and the tumor tissue by surgery before and after drug administration respectively. The fresh tumor tissues were collected immediately after biopsy and surgical resection for RNA-seq. Tumor samples were fixed with formalin and embedded in paraffin for IHC.
Outcomes	The primary endpoint was objective response rate (ORR) accessed by the RECIST criteria. The secondary endpoints were MPR rate (proportion of patients with no more than 10% residual viable tumor cells), the complete resection rate, EFS, OS, and TRAEs.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*