NSCLC Test 2

Generated on 2026-09-17

Leaderboard

Posterior mean with a 94% interval instead of a point score. Overlapping intervals are NOT separated — there the evidence supports no order.

Front group: ALK and MET; back group: KRAS and EGFR. The separation between the groups is robust (the intervals do not overlap); within a group the order is not decided.

Of 240 judgements, 53 are abstentions. Q12 and Q17: almost only abstentions, no contribution to the order. Q3 and Q4: at least half abstentions. Not in the overall rating: Q12. For Q3, Q12, and Q17 this suggests a run context that does not match the dossiers (run: Small Molecule, dossiers: small-molecule inhibitor). For Q4 this suggests a gap in the dossier schema, not thin dossiers or the context.

For Q1 and Q8 the judge changed its answer in 3 of 6 mirrored pairs: position, not evidence. For Q18 and Q19 the judge changed its answer in 4 of 6 mirrored pairs: position, not evidence.

The order only compares these 4 candidates with each other, only applies in the context Lung, non-small cell lung cancer, Small Molecule and 8 more and is based on the dossiers as of 09/09/2026.

Rank	Candidate	Posterior μ	94% interval	Obs.	Abst.
1	ALK (not decided)	0.913	[0.51, 1.32]	89	13
2	MET	0.418	[0.02, 0.82]	90	12
3	KRAS (not decided)	-0.538	[-0.94, -0.13]	89	13
4	EGFR	-0.793	[-1.20, -0.39]	84	18

Criteria set: Standard v2.3

Weighting: Q1 ×3, Q2 ×3, Q3 ×1, Q4 ×1, Q5 ×1, Q6 ×3, Q7 ×1, Q8 ×1, Q9 ×1, Q10 ×1, Q11 ×2, Q12 ×1 (no judgements, not in the overall ranking), Q13 ×1, Q14 ×1, Q15 ×1, Q16 ×1, Q18 ×1, Q20 ×1

BBT versus OpenSkill

Both estimators read THE SAME ledger — no judge said anything different here. Where the ranks diverge, the same material has merely been added up differently. This measures the stability of the ranking, not agreement between judges.

Rank correlation (Kendall τ-b): 1.00 over 4 shared candidates.

No candidate moves by 2 places or more — the order does not depend on the choice of estimator.

Candidate	BBT	OpenSkill	Δ places
ALK	1	1	0
MET	2	2	0
KRAS	3	3	0
EGFR	4	4	0

Appendix: Judgements

The contested and the decisive observations of this run, with the judge’s reasoning and the evidence it cited. An outcome names the candidate, not its displayed position; the order shown is stated separately. Abstention and tie are different answers.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.99

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: FIRST has direct regulatory and clinical evidence that ALK positivity, identified by approved assays, selects patients for effective ALK inhibition: the FDA label establishes the predictive biomarker and assay requirement (C1-EV_BIOMARKER_0001), PROFILE 1014 shows superior efficacy over chemotherapy in centrally confirmed ALK-positive NSCLC (C1-EV_BIOMARKER_0002; C1-EV_CLINICAL_0001), and FDA-authorized FISH, IHC, and sequencing-based companion diagnostics are documented (C1-EV_TRACTABILITY_0006). Longitudinal plasma ALK mutation kinetics provide a clinical pharmacodynamic signal consistent with target inhibition (C1-EV_BIOMARKER_0005). KRAS has evidence of clinically validated small-molecule tractability and catalogue-level statements about KRAS-G12C treatment, but the dossier does not provide comparably strong, specific evidence for a validated NSCLC patient-selection assay, clinical target-engagement/PD readout, and linked early efficacy signal. The ALK limitations—nonproductive FISH results and resistance mechanisms—reduce universality but do not negate the substantially stronger overall biomarker strategy.

Decisive evidence:

- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_BIOMARKER\_0005`
- `C1-EV\_TRACTABILITY\_0006`
- `C1-EV\_CLINICAL\_0001`

Contradicting evidence:

- `C1-EV\_BIOMARKER\_0004`
- `C1-EV\_BIOMARKER\_0006`

Missing evidence: For KRAS, a regulatory-validated NSCLC biomarker assay linked to patient selection, direct clinical pharmacodynamic or target-engagement measurements, and randomized or otherwise clearly documented early efficacy evidence in the specified NSCLC context.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET offers the more complete clinical biomarker chain. MET exon 14 skipping has prospective clinical validation for patient selection, serial METex14 ctDNA depletion is an on-treatment pharmacodynamic biomarker associated with radiographic response, and a validated pMET:total-MET assay provides a direct target-engagement readout. These biomarkers are connected to durable clinical responses and FDA-approved testing and treatment. KRAS has clinically validated tractability and mutation-based selection for KRAS G12C NSCLC, but the supplied dossier does not provide an equally explicit clinically validated target-engagement/PD assay linked to early efficacy; the cited KRAS-GTP and pERK measures are presented as biological readouts rather than established clinical assays. MET IHC and lower-level amplification have important limitations, but the METex14 strategy is specifically supported despite those caveats.

Decisive evidence:

- `C2-EV\_BIOMARKER\_0001`
- `C2-EV\_BIOMARKER\_0002`
- `C2-EV\_BIOMARKER\_0005`
- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`

- C2-EV_CLINICAL_0003
- C1-CAT-ONCO-BIO-091b
- C1-EV_TRACTABILITY_0001

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_CLINICAL_0006
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0006

Missing evidence: For KRAS, a prospectively validated clinical target-engagement and pharmacodynamic assay in NSCLC that links on-treatment KRAS pathway suppression to early efficacy, rather than primarily documenting mutation-based treatment selection.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) MET · 2) EGFR

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has direct, indication-relevant clinical validation in biomarker-defined NSCLC: two selective small-molecule MET inhibitors have regulatory approval for MET exon 14 skipping disease, with prospective molecular selection, substantial response rates, durable responses, and pharmacodynamic ctDNA evidence associated with radiographic response (C1-EV_CLINICAL_0001, C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0003, C1-EV_BIOMARKER_0001, C1-EV_BIOMARKER_0002). The dossier also documents small-molecule tractability and a second approved agent (C1-EV_TRACTABILITY_0002, C1-EV_TRACTABILITY_0004). EGFR has strong mechanistic and genetic support and a generic claim of multiple approved TKIs (C2-EV_TRACTABILITY_0001), but the supplied dossier does not provide comparable indication-specific clinical trial datasets, dose/exposure details, endpoint results, or regulatory evidence tied to the candidate in the stated context. MET's negative studies concern broad or poorly selected MET populations and do not negate the biomarker-defined METex14 validation, consistent with the guardrail (C1-EV_MECHANISM_0006, C1-EV_MECHANISM_0007, C1-EV_BIOMARKER_0004, C1-EV_CLINICAL_0006).

Decisive evidence:

- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0003
- C1-EV_BIOMARKER_0001
- C1-EV_BIOMARKER_0002
- C1-EV_TRACTABILITY_0002
- C1-EV_TRACTABILITY_0004
- C2-EV_TRACTABILITY_0001

Contradicting evidence:

- C1-EV_CLINICAL_0006
- C1-EV_BIOMARKER_0004
- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007

Missing evidence: For EGFR, indication-specific completed clinical trial and regulatory evidence with prospectively defined NSCLC population, molecule identity and quality, dose and exposure, comparator or trial design, endpoints, response durability, target-engagement or proof-of-mechanism data, and coded stoppage reasons.; For a direct modality comparison, matched evidence under the stated lung NSCLC context, including dose/exposure and endpoint-level results for EGFR-directed molecules.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has direct, indication-relevant clinical validation in NSCLC: prospective biomarker-selected phase 2 data show durable responses with capmatinib and tepotinib, matched molecular response supports target engagement, and both agents have FDA approval for MET exon 14-skipping metastatic NSCLC (C1-EV_CLINICAL_0001, C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0003, C1-EV_BIOMARKER_0001, C1-EV_BIOMARKER_0002). The evidence also includes a second selective molecule and clinically characterized toxicity, strengthening interpretation across molecules and exposure contexts (C1-EV_SAFETY_0002, C1-EV_SAFETY_0005). Negative MET trials mainly evaluated broad or poorly predictive populations and do not invalidate the biomarker-defined result (C1-EV_MECHANISM_0006, C1-EV_MECHANISM_0007, C1-EV_BIOMARKER_0004). KRAS has evidence of clinical tractability and a catalogue claim of approved G12C inhibitors in NSCLC, but the dossier does not provide comparable indication-specific trial results, endpoint data, dose/exposure details, target-engagement data, or trial-design evidence. Therefore MET has substantially stronger clinical validation for the stated NSCLC context.

Decisive evidence:

- `C1-EV\_CLINICAL\_0001`
- `C1-EV\_CLINICAL\_0002`
- `C1-EV\_CLINICAL\_0003`
- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_SAFETY\_0002`
- `C1-EV\_SAFETY\_0005`
- `C2-EV\_TRACTABILITY\_0001`
- `C2-CAT-ONCO-BIO-091b`

Contradicting evidence:

- `C1-EV\_CLINICAL\_0006`
- `C1-EV\_BIOMARKER\_0004`
- `C1-EV\_MECHANISM\_0006`
- `C1-EV\_MECHANISM\_0007`

Missing evidence: For KRAS, indication-specific NSCLC clinical validation with completed-trial design, dose and exposure, prospectively selected population, target-engagement or pharmacodynamic data, and clinical endpoints such as response rate and duration of response.; For KRAS, direct dossier evidence describing the specific approved NSCLC indication, trial population, endpoint results, and coded trial-stoppage or failure reasons.

Anomalies: C2-CAT-ONCO-BIO-018 contains a request to specify the disease.; C2-CAT-ONCO-BIO-026 contains a request for additional SNP identifiers.; C2-CAT-ONCO-BIO-037 contains a request to specify the disease.; C2-CAT-ONCO-BIO-053 contains a request to specify the disease.; C2-CAT-ONCO-BIO-083 states that the hypothesis is missing and conditionally requests an assumption.; C2-CAT-ONCO-BIO-106 contains a request to specify the disease.; C2-CAT-ONCO-BIO-111 contains a request to specify the disease and hypotheses.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has explicit regulatory validation as a predictive biomarker, FDA-authorized testing and companion-diagnostic infrastructure, biomarker-selected randomized evidence against platinum-pemetrexed chemotherapy in untreated advanced NSCLC, and longitudinal plasma mutation kinetics consistent with target engagement and pharmacodynamic monitoring. EGFR has biologically plausible and clinically relevant mutation claims, but the supplied dossier does not provide comparably specific validated-assay, prospective pharmacodynamic, or biomarker-selected efficacy evidence. ALK therefore offers the more credible clinical biomarker strategy for the stated criterion.

Decisive evidence:

- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_BIOMARKER\_0005`
- `C1-EV\_TRACTABILITY\_0005`
- `C1-EV\_TRACTABILITY\_0006`
- `C1-EV\_CLINICAL\_0001`

Contradicting evidence:

- `C1-EV\_BIOMARKER\_0004`
- `C1-EV\_BIOMARKER\_0006`
- `C2-CAT-ONCO-BIO-006`

Missing evidence: For EGFR, a directly documented FDA-authorized clinical assay for the relevant activating NSCLC mutations, prospective pharmacodynamic or target-engagement measurements, and a biomarker-selected randomized comparison demonstrating an early efficacy signal are not supplied with comparable specificity.

Anomalies: C2-CAT-ONCO-BIO-019; C2-CAT-ONCO-BIO-112; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-061; C2-CAT-ONCO-BIO-106

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) MET · 2) EGFR

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET offers the more credible clinical biomarker strategy. MET exon 14 skipping was prospectively assessed using tissue RNA or plasma cfDNA/NGS and was associated with clinical responses, while serial METex14 ctDNA depletion provided an on-treatment pharmacodynamic signal associated with radiographic response (C1-EV_BIOMARKER_0001, C1-EV_BIOMARKER_0002). A validated pMET:total-MET assay provides a direct biochemical target-engagement readout (C1-EV_BIOMARKER_0005), and FDA approvals explicitly tie METex14 testing to treatment selection (C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0003). The strategy has recognized limitations, including failure of nonspecific MET IHC selection and assay/tumor heterogeneity (C1-EV_BIOMARKER_0004, C1-EV_BIOMARKER_0006), but these limitations reinforce the need for genotype-defined selection rather than eliminate the validated strategy. The EGFR dossier contains evidence of biological and therapeutic tractability, including approved small-molecule TKIs (C2-EV_TRACTABILITY_0001), but does not supply comparable direct evidence for a specific clinical biomarker assay, target-engagement pharmacodynamic assay, serial biomarker response, or early efficacy signal in NSCLC. Therefore, MET is clearly better supported for the full biomarker-and-target-engagement criterion.

Decisive evidence:

- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_BIOMARKER\_0005`

- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0003
- C1-EV_SAFETY_0002

Contradicting evidence:

- C1-EV_BIOMARKER_0004
- C1-EV_BIOMARKER_0006
- C1-EV_TISSUE_0004
- C1-EV_CLINICAL_0006

Missing evidence: For EGFR, prospective clinical evidence linking a defined clinical assay to patient selection, target-engagement pharmacodynamics, serial biomarker response, and an early efficacy signal in NSCLC; specifically, a validated clinical EGFR biomarker/PD assay dataset and corresponding regulatory precedent.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has extensive, indication-relevant clinical validation in NSCLC: multiple randomized phase 3 trials in molecularly selected patients, including treatment-naive populations, demonstrate improved progression-free survival and intracranial control versus active comparators, with regulatory approvals and an established small-molecule modality. The dossier also supplies a clear predictive biomarker and a defined clinical dosing example for lorlatinib. EGFR has evidence of approved small-molecule tractability and a validated driver hypothesis, but the supplied dossier does not provide comparable completed randomized clinical-trial results, endpoint data, dose/exposure information, or coded trial outcomes. The ALK evidence is therefore substantially stronger despite the limitation that several ALK trials concern advanced rather than the exact contextual disease setting and despite the lack of an unadjusted overall-survival benefit in PROFILE 1014.

Decisive evidence:

- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0003
- C1-EV_CLINICAL_0004
- C1-EV_CLINICAL_0005
- C1-EV_BIOMARKER_0002
- C1-EV_TRACTABILITY_0002
- C1-EV_TRACTABILITY_0003

Contradicting evidence:

- C1-EV_CLINICAL_0006
- C1-EV_SAFETY_0005
- C1-EV_SAFETY_0006
- C1-EV_SAFETY_0007

Missing evidence: For EGFR, randomized indication-relevant clinical trial records with explicit comparator, dose, exposure, population selection, endpoint results, and completed-trial status are not supplied; such evidence would most affect the comparison.; For both candidates, directly comparable dose/exposure and pharmacodynamic target-engagement data from the cited pivotal trials are incompletely supplied.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has substantially stronger indication-relevant clinical validation. Multiple randomized phase 3 studies in biomarker-selected NSCLC populations demonstrated improved progression-free survival, response, symptom or intracranial outcomes versus chemotherapy or active ALK therapy, and several ALK inhibitors have regulatory approval in NSCLC (C1-EV_CLINICAL_0001, C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0003, C1-EV_CLINICAL_0004, C1-EV_CLINICAL_0005). The PROFILE 1014 evidence directly matches untreated advanced disease and chemotherapy comparison, while the other trials provide replicated clinical proof-of-concept and modality validation. The lack of an unadjusted overall-survival benefit in PROFILE 1014 is a limitation, but extensive crossover makes it weak evidence against ALK target validity (C1-EV_CLINICAL_0006). KRAS has evidence that mutant-selective small-molecule inhibition is clinically tractable and that G12C inhibitors are approved, but the supplied dossier does not provide corresponding NSCLC trial designs, endpoints, dose or exposure data, or target-engagement-linked clinical outcomes; the approval claim alone is insufficient for a comparable adjudication (C2-EV_TRACTABILITY_0001, C2-CAT-ONCO-BIO-091b). Dose and exposure details are also incompletely supplied for ALK, but this limitation does not offset the large difference in indication-specific randomized clinical evidence.

Decisive evidence:

- `C1-EV\_CLINICAL\_0001`
- `C1-EV\_CLINICAL\_0002`
- `C1-EV\_CLINICAL\_0003`
- `C1-EV\_CLINICAL\_0004`
- `C1-EV\_CLINICAL\_0005`
- `C1-EV\_BIOMARKER\_0002`
- `C2-EV\_TRACTABILITY\_0001`
- `C2-CAT-ONCO-BIO-091b`

Contradicting evidence:

- `C1-EV\_CLINICAL\_0006`

Missing evidence: For KRAS: indication-specific NSCLC trial reports with randomized or otherwise controlled design, prespecified endpoints, enrolled KRAS-mutant population, administered dose, achieved exposure, target-engagement or pharmacodynamic data, and coded reasons for discontinuation or failure.; For both candidates: directly comparable dose-exposure and target-engagement data for the molecules underlying the cited clinical outcomes, particularly in the specified newly diagnosed, chemotherapy-backbone context.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.98

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has substantially stronger indication-relevant clinical validation. The dossier contains multiple randomized phase 3 trials in biomarker-selected, previously untreated advanced ALK-positive NSCLC, including direct comparison with platinum-pemetrexed chemotherapy, superior progression-free survival and response outcomes, intracranial efficacy, durable follow-up, and regulatory approval extending into the adjuvant setting (C2-EV_CLINICAL_0001 through C2-EV_CLINICAL_0005; C2-EV_BIOMARKER_0002; C2-EV_TRACTABILITY_0005). These results use an established oral small-molecule modality and validated companion-diagnostic selection. KRAS has evidence of approved G12C inhibitors in NSCLC and clear small-molecule tractability (C1-CAT-ONCO-BIO-091b; C1-EV_TRACTABILITY_0001), but the dossier does not provide comparable

randomized trial-level efficacy, dose/exposure, target-engagement, or endpoint detail for KRAS in the stated context. The ALK evidence is therefore clearly stronger, despite acknowledged limitations including lack of unadjusted overall-survival significance in PROFILE 1014 and resistance or heterogeneous dependence in some models.

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0003
- C2-EV_CLINICAL_0004
- C2-EV_CLINICAL_0005
- C2-EV_BIOMARKER_0002
- C2-EV_TRACTABILITY_0005
- C1-CAT-ONCO-BIO-091b

Contradicting evidence:

- C2-EV_CLINICAL_0006
- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006

Missing evidence: For KRAS, randomized NSCLC trial reports with molecule-specific dose, pharmacokinetic exposure, target-engagement/pharmacodynamic data, comparator outcomes, and prespecified endpoint results would most change the comparison.; For KRAS, coded clinical trial-stoppage reasons and detailed evidence distinguishing target failure from inadequate exposure or molecule limitations are not supplied.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) EGFR · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET offers the more credible integrated clinical biomarker strategy. MET exon 14 skipping has prospective clinical selection using tissue RNA or plasma cfDNA/NGS, is linked to objective responses and regulatory-approved treatment, and has a reported on-treatment molecular response based on serial METex14 ctDNA depletion. A validated pMET:total-MET assay provides a direct target-engagement and pharmacodynamic readout. EGFR has strong clinical precedent for mutation-directed therapy, but the supplied evidence does not provide an equally explicit clinical target-engagement or pharmacodynamic assay strategy; the pEGFR item is presented as a general biological measurement and is not shown to be clinically validated. MET's limitations—heterogeneous expression, assay discordance, and failure of MET-IHC alone—are documented, but they reinforce use of the more specific METex14/genomic strategy rather than outweighing it.

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_BIOMARKER_0002
- C2-EV_BIOMARKER_0005
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0003

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_CLINICAL_0006
- C1-CAT-ONCO-BIO-099

Missing evidence: For EGFR, an explicitly validated clinical target-engagement assay, on-treatment pharmacodynamic biomarker, and biomarker-linked early efficacy dataset are not supplied; the dossier describes pEGFR as a biological readout but does not establish it as a clinical assay supporting the required decision.; For the stated newly diagnosed, chemotherapy-backbone, in-vivo PDX context, neither dossier supplies a directly matched clinical biomarker-validation dataset.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) ALK · 2) MET

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has stronger indication-relevant clinical validation. It is supported by randomized phase 3 evidence in previously untreated ALK-positive advanced NSCLC directly against platinum-pemetrexed chemotherapy, with improvements in progression-free survival, response, symptoms and quality of life, plus multiple later randomized first-line trials and durable intracranial/systemic benefit. The modality is an approved oral small-molecule inhibitor selected by an approved biomarker test. METex14 has meaningful validation through selective-inhibitor phase 2 responses and traditional approvals, including treatment-naive patients, but the dossier provides no randomized first-line comparison against chemotherapy; its strongest evidence is single-arm and biomarker-defined. Negative MET trials largely concern nonselective IHC selection or broad pathway inhibition and therefore do not invalidate selective METex14 targeting, but they reinforce the importance of population and biomarker selection. The dossiers do not provide detailed dose or exposure data sufficient for a finer molecule-quality comparison.

Decisive evidence:

- `C1-EV\_CLINICAL\_0001`
- `C1-EV\_CLINICAL\_0002`
- `C1-EV\_CLINICAL\_0003`
- `C1-EV\_CLINICAL\_0004`
- `C1-EV\_TRACTABILITY\_0002`
- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`
- `C2-EV\_CLINICAL\_0003`

Contradicting evidence:

- `C1-EV\_CLINICAL\_0006`
- `C2-EV\_CLINICAL\_0006`
- `C2-EV\_MECHANISM\_0006`
- `C2-EV\_MECHANISM\_0007`

Missing evidence: A randomized, first-line METex14-skipping NSCLC trial directly comparing a selective MET inhibitor with platinum-based chemotherapy, including prespecified endpoint results, dose/exposure adequacy, and pharmacodynamic target-engagement data in the indication-relevant population.; Clinical validation specifically in mucus-producing glandular NSCLC cells or a matching patient-derived xenograft context.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) EGFR · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has substantially stronger indication-relevant clinical validation. The dossier provides genotype-selected phase 2 evidence with durable responses in MET exon 14-skipping metastatic

NSCLC, traditional FDA approvals for capmatinib and tepotinib, prospective biomarker selection, and on-treatment METex14 ctDNA depletion associated with response (C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0002, C2-EV_CLINICAL_0003, C2-EV_BIOMARKER_0001, C2-EV_BIOMARKER_0002). It also provides a defined approved oral regimen for tepotinib (C2-EV_COMPETITIVE_0002). Negative results for nonspecific MET immunohistochemistry, low-level amplification, and broad MET-pathway inhibition do not invalidate METex14 because they concern different biomarker populations (C2-EV_CLINICAL_0006, C2-EV_BIOMARKER_0004, C2-EV_MECHANISM_0006, C2-EV_MECHANISM_0007). EGFR is described as clinically established and approved in EGFR-mutant NSCLC, but the supplied EGFR dossier does not provide comparable indication-specific trial datasets, endpoint results, dose/exposure information, or target-engagement data; its clinical claims are largely catalogue-level summaries (C1-CAT-ONCO-BIO-091b, C1-CAT-ONCO-BIO-046).

Decisive evidence:

- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`
- `C2-EV\_CLINICAL\_0003`
- `C2-EV\_BIOMARKER\_0001`
- `C2-EV\_BIOMARKER\_0002`
- `C2-EV\_COMPETITIVE\_0002`
- `C1-CAT-ONCO-BIO-091b`
- `C1-CAT-ONCO-BIO-046`

Contradicting evidence:

- `C2-EV\_CLINICAL\_0006`
- `C2-EV\_BIOMARKER\_0004`
- `C2-EV\_MECHANISM\_0006`
- `C2-EV\_MECHANISM\_0007`

Missing evidence: For EGFR, indication-specific completed-trial reports with prespecified endpoints, comparator arms, enrolled biomarker-defined populations, dose and exposure data, pharmacodynamic target engagement, and coded trial-stoppage reasons.; For both targets, direct clinical validation in the specified mucus-producing glandular-cell, newly diagnosed, chemotherapy-backbone population is not provided.

Q11 · Competitive differentiation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK offers greater potential for clinically meaningful differentiation over chemotherapy because ALK-directed therapy has direct randomized evidence in previously untreated, biomarker-selected NSCLC showing improved progression-free survival and response rate versus platinum-pemetrexed chemotherapy (C2-EV_CLINICAL_0001, C2-EV_BIOMARKER_0002). The dossier also supports clinically meaningful differentiation through intracranial efficacy and durable long-term disease control (C2-EV_TRACTABILITY_0003, C2-EV_COMPETITIVE_0002), with mechanistic differentiation against resistance mutations such as G1202R (C2-EV_COMPETITIVE_0004). Although the ALK segment is highly competitive and differentiation among current TKIs is not uniformly established (C2-EV_COMPETITIVE_0001, C2-EV_COMPETITIVE_0006), the evidence directly addresses the stated chemotherapy comparison. KRAS has strong biological and small-molecule tractability evidence, but the supplied dossier does not provide comparable direct evidence of superiority over chemotherapy in newly diagnosed NSCLC. This judgment concerns scientific and clinical differentiation evidence, not market attractiveness.

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_BIOMARKER_0002
- C2-EV_TRACTABILITY_0003
- C2-EV_COMPETITIVE_0002
- C2-EV_COMPETITIVE_0004

Contradicting evidence:

- C2-EV_CLINICAL_0006
- C2-EV_COMPETITIVE_0001
- C2-EV_COMPETITIVE_0006

Missing evidence: A direct randomized comparison of a KRAS-directed therapy with chemotherapy in newly diagnosed KRAS-altered NSCLC.; A current, disease-specific KRAS competitor pipeline and treatment-landscape analysis addressing differentiation over first-line chemotherapy.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET offers the more credible biomarker strategy. MET exon 14 skipping has prospective clinical selection using tissue RNA or plasma cfDNA/NGS, an on-treatment ctDNA depletion pharmacodynamic readout associated with radiographic response, and a validated pMET:total-MET assay for direct target engagement. Capmatinib and tepotinib approvals using FDA-approved tests provide strong regulatory precedent. The strategy is not perfect: MET IHC alone failed as a predictive biomarker and intratumoral heterogeneity can impair classification. KRAS has evidence of clinically validated small-molecule tractability, but the supplied dossier does not provide comparable direct NSCLC evidence for a clinical selection assay, target engagement/pharmacodynamics, or linked early efficacy signal; its relevant evidence is primarily generic or catalogue-level. Therefore FIRST is clearly better for this criterion.

Decisive evidence:

- C1-EV_BIOMARKER_0001
- C1-EV_BIOMARKER_0002
- C1-EV_BIOMARKER_0005
- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0003

Contradicting evidence:

- C1-EV_BIOMARKER_0004
- C1-EV_BIOMARKER_0006

Missing evidence: For KRAS: prospective NSCLC clinical evidence linking a defined genomic biomarker to patient selection, a validated clinical target-engagement or pharmacodynamic assay, and an early efficacy signal with regulatory companion-diagnostic precedent.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.97

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has substantially stronger indication-relevant clinical validation in NSCLC: genotype-selected phase 2 trials demonstrated durable responses to selective MET inhibitors, with traditional FDA approvals for MET exon 14–skipping disease and prospective molecular

pharmacodynamic evidence linking METex14 depletion to response (C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0002, C2-EV_CLINICAL_0003, C2-EV_BIOMARKER_0001, C2-EV_BIOMARKER_0002). The dossier also includes a validated pMET target-engagement assay (C2-EV_BIOMARKER_0005). MET amplification is weaker and threshold-dependent, and broad MET or IHC-selected strategies had negative trials (C2-EV_CLINICAL_0006, C2-EV_BIOMARKER_0004, C2-EV_MECHANISM_0006, C2-EV_MECHANISM_0007), but these limitations do not negate the strong METex14 evidence. KRAS has evidence of approved mutant-selective inhibitors and clinical tractability (C1-EV_TRACTABILITY_0001, C1-CAT-ONCO-BIO-091b), yet lacks a comparably detailed, indication-specific clinical validation dataset in the supplied dossier. The advantage is therefore clear, while not establishing superiority for the exact newly diagnosed chemotherapy context.

Decisive evidence:

- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`
- `C2-EV\_CLINICAL\_0003`
- `C2-EV\_BIOMARKER\_0001`
- `C2-EV\_BIOMARKER\_0002`
- `C2-EV\_BIOMARKER\_0005`
- `C2-EV\_TRACTABILITY\_0003`
- `C2-EV\_TRACTABILITY\_0004`

Contradicting evidence:

- `C2-EV\_CLINICAL\_0006`
- `C2-EV\_BIOMARKER\_0004`
- `C2-EV\_MECHANISM\_0006`
- `C2-EV\_MECHANISM\_0007`

Missing evidence: An indication-matched, head-to-head comparison of KRAS-directed therapy versus MET-directed therapy in newly diagnosed NSCLC receiving chemotherapy, including randomized design, dose, pharmacokinetic exposure, target engagement, and clinically relevant endpoints.; Direct KRAS NSCLC clinical-trial evidence with comparable detail on patient selection, response durability, target engagement, and dose/exposure; the KRAS dossier mainly supplies approval and tractability claims without a comparable trial dataset.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK offers the more credible end-to-end clinical biomarker strategy. ALK rearrangement is explicitly established as a regulatory predictive biomarker with FDA-approved testing, including FISH, IHC, and companion-diagnostic infrastructure (C2-EV_BIOMARKER_0001; C2-EV_TRACTABILITY_0006). Biomarker-selected randomized evidence demonstrates an early efficacy signal versus chemotherapy (C2-EV_BIOMARKER_0002; C2-EV_CLINICAL_0001), while longitudinal plasma ALK mutation kinetics provide clinical pharmacodynamic evidence consistent with target engagement (C2-EV_BIOMARKER_0005). Durable systemic and intracranial efficacy further supports the strategy (C2-EV_CLINICAL_0004). KRAS has strong genetic and therapeutic precedent, including approved KRAS-G12C inhibitors and mutation-based selection (C1-CAT-ONCO-BIO-091b; C1-CAT-ONCO-BIO-115), but the supplied dossier does not document an equivalently complete clinical assay, target-engagement, pharmacodynamic, and biomarker-linked early efficacy chain. ALK evidence has important limitations: ALK positivity can be nonproductive by FISH alone and resistance or variant heterogeneity can weaken dependence (C2-EV_BIOMARKER_0004; C2-EV_BIOMARKER_0006; C2-EV_MECHANISM_0006; C2-EV_MECHANISM_0007), but these caveats do not outweigh its substantially stronger clinical validation.

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_BIOMARKER_0002
- C2-EV_BIOMARKER_0005
- C2-EV_TRACTABILITY_0006
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0004
- C1-CAT-ONCO-BIO-091b
- C1-CAT-ONCO-BIO-115

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007

Missing evidence: For KRAS, a directly documented FDA-authorized clinical assay strategy, clinically measured target-engagement pharmacodynamics, and biomarker-linked early NSCLC efficacy data comparable in specificity to the ALK evidence.

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK supports the clearer prospectively testable patient-selection hypothesis. ALK rearrangement is a defined molecular subtype with validated testing, regulatory companion-diagnostic infrastructure, and randomized prospective evidence showing greater benefit from ALK inhibition than chemotherapy in ALK-selected NSCLC (C2-EV_BIOMARKER_0001, C2-EV_BIOMARKER_0002, C2-EV_CLINICAL_0001, C2-EV_TRACTABILITY_0006). KRAS is also a strong genomic stratifier, with recurrent mutation-defined NSCLC subsets and prospective sequencing evidence (C1-EV_GENETICS_0002, C1-EV_GENETICS_0003, C1-EV_GENETICS_0004), but the dossier does not provide equivalently direct prospective clinical evidence operationalizing KRAS selection in the stated context. ALK heterogeneity, nonproductive FISH results, and variant-specific sensitivity require assay and subtype refinement but do not remove the established feasibility of prospective stratification (C2-EV_BIOMARKER_0006, C2-EV_TISSUE_0004, C2-EV_MECHANISM_0007).

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_BIOMARKER_0002
- C2-EV_CLINICAL_0001
- C2-EV_TRACTABILITY_0006
- C2-EV_GENETICS_0005
- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003
- C1-EV_GENETICS_0004

Contradicting evidence:

- C2-EV_BIOMARKER_0006
- C2-EV_TISSUE_0004
- C2-EV_MECHANISM_0007
- C2-EV_MECHANISM_0006

Missing evidence: A prospective, biomarker-selected KRAS NSCLC trial using a prespecified genomic assay and defined KRAS allelic/co-mutation strata, directly comparable with the established prospective

ALK-selection evidence.

Anomalies: C1-CAT-ONCO-BIO-018; C1-CAT-ONCO-BIO-026; C1-CAT-ONCO-BIO-037; C1-CAT-ONCO-BIO-083; C1-CAT-ONCO-BIO-106; C1-CAT-ONCO-BIO-111

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK supports the clearer prospectively testable selection hypothesis. The ALK-rearranged subgroup is explicitly operationalized through FDA-authorized assays and companion-diagnostic infrastructure, and its predictive value is demonstrated in randomized biomarker-selected NSCLC trials against chemotherapy. EGFR has a biologically defined subgroup based on activating mutations and the dossier describes mutation-directed testing and treatment selection, but the supporting evidence is predominantly catalogue-level and lacks comparable direct prospective clinical-validation or regulatory diagnostic evidence. ALK heterogeneity, subclonality, and occasional nonproductive FISH results qualify implementation but do not negate the established operational selection framework.

Decisive evidence:

- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_TRACTABILITY\_0005`
- `C1-EV\_TRACTABILITY\_0006`
- `C1-EV\_CLINICAL\_0001`
- `C2-EV\_GENETICS\_0001`
- `C2-CAT-ONCO-BIO-036`
- `C2-CAT-ONCO-BIO-091b`

Contradicting evidence:

- `C1-EV\_BIOMARKER\_0006`
- `C1-EV\_TISSUE\_0004`
- `C1-EV\_TISSUE\_0005`
- `C1-EV\_MECHANISM\_0007`

Missing evidence: A prospective, biomarker-defined EGFR NSCLC trial or regulatory companion-diagnostic record directly demonstrating operational patient selection, comparable to the ALK-selected randomized trials and authorized assays.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has direct, peer-reviewed pharmacologic perturbation evidence in six MET-amplified lung-cancer PDX models showing antitumor responses, plus independent patient-derived model evidence of in-vivo sensitivity to MET inhibition. The MET CRISPR result in A549 cells weakens claims of universal MET essentiality but does not negate biomarker-restricted PDX activity. KRAS has evidence of tractability and general preclinical tool use, but the supplied dossier does not provide a comparably direct KRAS perturbation result in NSCLC PDX models demonstrating reversal of disease-relevant phenotypes. Therefore, MET has stronger reproducible PDX perturbational evidence.

Decisive evidence:

- `C1-EV\_PERTURBATION\_0002`

- C1-EV_PERTURBATION_0003

Contradicting evidence:

- C1-EV_PERTURBATION_0005

Missing evidence: A KRAS-directed perturbation study in NSCLC patient-derived xenografts reporting reversal of tumor-growth or other disease-relevant phenotypes, with orthogonal genetic or pharmacologic confirmation.

Anomalies: C2-CAT-ONCO-BIO-018; C2-CAT-ONCO-BIO-037; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-083; C2-CAT-ONCO-BIO-106; C2-CAT-ONCO-BIO-111

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK supports the clearer patient-selection hypothesis. ALK rearrangement is explicitly defined as a predictive biomarker, selected using FDA-authorized assays, and tested prospectively in randomized NSCLC trials against chemotherapy. This directly satisfies the requirement for an operationalizable genomic stratification hypothesis. KRAS has strong evidence for recurrent mutation-defined subsets and established small-molecule tractability, but the dossier provides less direct prospective clinical selection evidence for KRAS in NSCLC and emphasizes allele-, histology-, co-mutation-, and context-dependence. ALK's operational limitations—nonproductive FISH signals, heterogeneity, and histologic variability—warrant assay refinement but do not outweigh its established prospective testing infrastructure.

Decisive evidence:

- C1-EV_BIOMARKER_0001
- C1-EV_BIOMARKER_0002
- C1-EV_CLINICAL_0001
- C1-EV_TRACTABILITY_0005
- C1-EV_TRACTABILITY_0006
- C2-EV_GENETICS_0001
- C2-EV_GENETICS_0004
- C2-EV_TRACTABILITY_0001

Contradicting evidence:

- C1-EV_BIOMARKER_0006
- C1-EV_TISSUE_0004
- C1-EV_TISSUE_0005
- C2-CAT-ONCO-BIO-036
- C2-CAT-ONCO-BIO-041

Missing evidence: For KRAS, a prospective NSCLC trial or regulatory companion-diagnostic evidence showing that a prespecified KRAS allele, especially G12C, prospectively selects patients for benefit from KRAS-directed therapy would most change the answer.

Q8 · Druggability and tractability

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK is more tractable for the specified small-molecule inhibition modality and lung setting. It has multiple approved oral small-molecule inhibitors, regulatory validation, established

companion-diagnostic selection, demonstrated systemic and intracranial activity, and precedent extending into adjuvant NSCLC treatment (C1-EV_TRACTABILITY_0001 through C1-EV_TRACTABILITY_0006). KRAS also has genuine small-molecule tractability, including approved mutant-selective covalent inhibitors and a structurally defined switch-II-pocket strategy, but the strongest evidence is allele-specific to KRAS-G12C and broader KRAS-G12D/pan-KRAS tractability remains less established (C2-EV_TRACTABILITY_0001; C2-CAT-ONCO-BIO-032; C2-CAT-ONCO-BIO-115). Resistance and bypass mechanisms reduce durability for both targets but do not outweigh ALK's broader and more mature modality, delivery, biomarker, and clinical precedent.

Decisive evidence:

- `C1-EV\_TRACTABILITY\_0001`
- `C1-EV\_TRACTABILITY\_0002`
- `C1-EV\_TRACTABILITY\_0003`
- `C1-EV\_TRACTABILITY\_0004`
- `C1-EV\_TRACTABILITY\_0005`
- `C1-EV\_TRACTABILITY\_0006`
- `C2-EV\_TRACTABILITY\_0001`
- `C2-CAT-ONCO-BIO-032`
- `C2-CAT-ONCO-BIO-115`

Contradicting evidence:

- `C1-EV\_TRACTABILITY\_0007`
- `C2-CAT-ONCO-BIO-013`
- `C2-CAT-ONCO-BIO-063`

Missing evidence: Direct comparative structural pocket-quality metrics, binding affinities, and experimentally resolved structures for ALK versus the specific KRAS allele relevant to the lung mucus-producing glandular-cell context.; Direct delivery and target-engagement data in patient-derived NSCLC xenografts involving mucus-producing glandular cells for both targets.; Evidence establishing small-molecule tractability of KRAS-G12D or other non-G12C KRAS alleles at a level comparable to the established G12C precedent.

Anomalies: C2-CAT-ONCO-BIO-018; C2-CAT-ONCO-BIO-026; C2-CAT-ONCO-BIO-037; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-083; C2-CAT-ONCO-BIO-106; C2-CAT-ONCO-BIO-111

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) MET · 2) EGFR

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET supports the clearer prospectively testable patient-selection hypothesis. MET exon 14 skipping was prospectively assayed using tissue RNA or plasma cfDNA/NGS, and selective MET inhibitors produced responses in the defined subgroup; capmatinib labeling explicitly requires an FDA-approved test. The perturbational evidence also supports restriction to molecularly defined subsets rather than broad MET expression. EGFR has strong evidence that exon 19 deletions and L858R are causal drivers, but the supplied dossier does not provide comparably direct prospective trial-level evidence linking a specified genomic assay-defined EGFR subgroup to patient selection and treatment response. Its catalogue evidence is largely general or non-validated for alternative germline polymorphism-based stratification.

Decisive evidence:

- `C1-EV\_BIOMARKER\_0001`
- `C1-EV\_CLINICAL\_0003`
- `C1-EV\_TRACTABILITY\_0002`
- `C1-EV\_CLINICAL\_0001`

- C1-EV_PERTURBATION_0004

Contradicting evidence:

- C1-EV_BIOMARKER_0006
- C1-EV_CLINICAL_0006
- C2-CAT-ONCO-BIO-006

Missing evidence: For EGFR, a prospective clinical validation dossier directly demonstrating that a specified genomic assay-defined EGFR-mutant NSCLC subgroup can be operationally enrolled and predicts response to a defined small-molecule treatment.

Q4 · Human natural experiments

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.96

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK is better supported by human pharmacological natural experiments: randomized clinical studies in ALK-positive NSCLC show superior efficacy versus chemotherapy and durable benefit with ALK inhibition (C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0002, C2-EV_CLINICAL_0004). Human regulatory safety evidence describes a generally manageable profile while also documenting important toxicities (C2-EV_SAFETY_0005), so tolerability is supported but not unqualified. KRAS has human tumor-genomic driver evidence and germline RASopathy evidence, but the supplied germline variants are activating and associated with developmental disease, not human loss-of-function or protective phenotypes supporting inhibition; the required genetic-direction match is therefore absent (C1-CAT-ONCO-BIO-005, C1-CAT-ONCO-BIO-016, C1-CAT-ONCO-BIO-056).

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0004
- C2-EV_SAFETY_0005

Contradicting evidence:

- C2-EV_CLINICAL_0006
- C2-EV_SAFETY_0006
- C2-EV_SAFETY_0007
- C1-CAT-ONCO-BIO-005
- C1-CAT-ONCO-BIO-016
- C1-CAT-ONCO-BIO-056

Missing evidence: For KRAS, a human loss-of-function or protective-allele study, or a long-term human KRAS-inhibitor exposure cohort directly reporting both NSCLC efficacy and tolerability, is missing. The supplied KRAS germline disease evidence reflects activating variants and therefore does not match therapeutic inhibition.

Q11 · Competitive differentiation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.95

Presented as: 1) MET · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK offers greater potential for clinically meaningful differentiation over chemotherapy because untreated, biomarker-selected ALK-positive patients showed superior progression-free survival and response rate with crizotinib versus platinum-pemetrexed chemotherapy in randomized phase 3 evidence (C2-EV_BIOMARKER_0002; C2-EV_CLINICAL_0001). The ALK landscape also contains clinically

demonstrated differentiation through superior CNS and long-term disease control with next-generation inhibitors, although the class is highly competitive (C2-EV_COMPETITIVE_0001; C2-EV_COMPETITIVE_0002). For MET, the dossier supports activity of selective inhibitors in treatment-naive METex14 disease, but explicitly states that chemotherapy remains a background option and provides no evidence of a comparative efficacy advantage over chemotherapy (C1-EV_COMPETITIVE_0003). MET is additionally a crowded competitive segment with multiple approved or clinically active agents and a non-TKI competitor (C1-EV_COMPETITIVE_0005; C1-EV_COMPETITIVE_0006). This conclusion addresses differentiation over chemotherapy, not overall scientific merit or market attractiveness.

Decisive evidence:

- C2-EV_BIOMARKER_0002
- C2-EV_CLINICAL_0001
- C2-EV_COMPETITIVE_0001
- C1-EV_COMPETITIVE_0003

Contradicting evidence:

- C2-EV_CLINICAL_0006
- C2-EV_COMPETITIVE_0006
- C1-EV_COMPETITIVE_0005
- C1-EV_COMPETITIVE_0006

Missing evidence: A randomized, direct comparison of a selective MET inhibitor versus chemotherapy in newly diagnosed METex14-positive NSCLC would most change the answer for MET.

Q6 · Translational model validity

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.95

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has the stronger translational efficacy evidence. Its dossier includes selective perturbation in EML4-ALK-positive NSCLC cells, evidence in a patient-derived NSCLC model, genetically engineered lung tumor models, and randomized clinical efficacy in previously untreated ALK-positive NSCLC against platinum-pemetrexed chemotherapy. These models and the clinical setting are directly relevant to NSCLC and newly diagnosed disease. KRAS has strong human genetic and mechanistic support, but the supplied evidence does not provide comparably detailed efficacy results in NSCLC PDX, organoid, primary-cell, or genetically engineered lung models; the cited efficacy example is a MiaPaCa-2 xenograft, which is pancreatic rather than NSCLC. ALK dependence is not uniform across variants or models, but these limitations do not offset the substantially more direct and clinically anchored translational efficacy evidence.

Decisive evidence:

- C2-EV_PERTURBATION_0001
- C2-EV_MECHANISM_0005
- C2-EV_TISSUE_0003
- C2-EV_CLINICAL_0001
- C2-EV_BIOMARKER_0002

Contradicting evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_PERTURBATION_0006

Missing evidence: Direct KRAS efficacy results in NSCLC patient-derived xenografts or genetically engineered lung models, specifically using newly diagnosed or treatment-naive disease settings.; Direct

comparative efficacy data for KRAS and ALK interventions in patient-derived or genetically engineered NSCLC models.

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.95

Presented as: 1) EGFR · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: SECOND (MET) has the more robust meta-evidence package. Its dossier identifies study types and provenance, distinguishes independent genomic, tissue, perturbational, clinical, and biomarker evidence, explicitly flags shared cohorts, and includes multiple contradictory or negative findings, including failed phase III trials and biomarker discordance (C2-EV_MECHANISM_0006, C2-EV_MECHANISM_0007, C2-EV_BIOMARKER_0004, C2-EV_TISSUE_0004, C2-EV_TISSUE_0006). The package is therefore auditable for source dependence and selective reporting, even though several efficacy items share GEOMETRY, VISION, or LUMINOSITY datasets. EGFR has strong biological claims but is dominated by catalogue-style entries with limited provenance metadata, no explicit cohort-overlap or replication map, and little directly documented negative-result availability; the single safety contradiction (C1-EV_SAFETY_0001) does not provide comparable meta-evidence. This decision is based on evidence structure and transparency, not on the larger number of MET findings.

Decisive evidence:

- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0003
- C2-EV_GENETICS_0004
- C2-EV_PERTURBATION_0001
- C2-EV_PERTURBATION_0002
- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0006
- C2-EV_BIOMARKER_0004

Contradicting evidence:

- C2-EV_GENETICS_0005
- C2-EV_GENETICS_0007
- C2-EV_PERTURBATION_0004
- C2-EV_PERTURBATION_0005
- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0006
- C2-EV_BIOMARKER_0004
- C2-EV_CLINICAL_0006

Missing evidence: For EGFR, an explicit provenance and source-family map identifying the underlying studies, cohorts, sample overlap, independent replications, and systematically available negative results is missing.; For EGFR, independent replication and negative-result evidence directly addressing the stated lung, mucus-producing glandular-cell, PDX, and in-vivo proof-of-concept context is missing.; For both candidates, a complete publication-bias assessment or prospectively defined evidence search is missing.

Q2 · Disease biology and mechanism coherence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.95

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has the more coherent inhibition mechanism for NSCLC: a recurrent human EML4-ALK fusion is shown to transform cells and is confirmed in human lung tumors, with defined constitutive kinase signaling and downstream MAPK dependence (C1-EV_GENETICS_0002, C1-EV_GENETICS_0003, C1-EV_MECHANISM_0003, C1-EV_MECHANISM_0004). Direct ALK knockdown selectively impaired viability and clonogenic growth in ALK-rearranged NSCLC models versus ALK-negative controls, providing experimentally causal support for the direction of modulation (C1-EV_PERTURBATION_0001). The ALK mechanism is not uniform across all tumors or fusion variants, as shown by resistance and heterogeneity evidence (C1-EV_MECHANISM_0006, C1-EV_MECHANISM_0007, C1-EV_PERTURBATION_0004, C1-EV_PERTURBATION_0005, C1-EV_PERTURBATION_0006), but these qualify the scope rather than eliminate the causal model. KRAS has strong human-genetic and pathway plausibility in NSCLC (C2-EV_GENETICS_0002, C2-EV_GENETICS_0003, C2-EV_MECHANISM_0001), but the dossier supplies substantially less direct, high-quality NSCLC-specific perturbational evidence establishing that inhibiting KRAS itself causally alters tumor pathophysiology. The decision therefore rests on mechanistic and perturbational coherence, not on druggability or clinical tractability.

Decisive evidence:

- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003
- C1-EV_MECHANISM_0003
- C1-EV_MECHANISM_0004
- C1-EV_PERTURBATION_0001
- C1-EV_TISSUE_0002
- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0003
- C2-EV_MECHANISM_0001

Contradicting evidence:

- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007
- C1-EV_TISSUE_0004
- C1-EV_PERTURBATION_0004
- C1-EV_PERTURBATION_0005
- C1-EV_PERTURBATION_0006
- C2-CAT-ONCO-BIO-052
- C2-CAT-ONCO-BIO-054
- C2-CAT-ONCO-BIO-070

Missing evidence: A direct, well-controlled KRAS inhibition or allele-specific genetic perturbation study in KRAS-mutant NSCLC models, ideally including rescue or pathway-level causal validation and comparison with KRAS-wild-type controls.; Human NSCLC pathology or spatial evidence specifically localizing the relevant KRAS-mutant causal state to the disease-relevant tumor cell population.

Anomalies: C2-CAT-ONCO-BIO-018; C2-CAT-ONCO-BIO-037; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-083; C2-CAT-ONCO-BIO-106; C2-CAT-ONCO-BIO-111

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.95

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has substantially stronger indication-relevant clinical validation. Its dossier includes regulatory biomarker selection, multiple randomized phase 3 trials in ALK-positive NSCLC,

chemotherapy-controlled evidence, superiority on progression-free survival and response endpoints, intracranial efficacy, long-term follow-up, and an adjuvant approval. The evidence also directly supports the small-molecule modality and clinical deployment through approved dosing and companion diagnostics. EGFR is supported by approved EGFR-directed drugs and broad claims of clinical validation in EGFR-mutated NSCLC, but the supplied dossier provides far less trial-level detail on dose, exposure, design, population, and endpoints. ALK's limitations—crossover-confounded overall survival in PROFILE 1014 and resistance or nonproductive-rearrangement issues—do not overturn the much stronger aggregate clinical validation, because they qualify durability or biomarker implementation rather than negate demonstrated target engagement and efficacy.

Decisive evidence:

- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`
- `C2-EV\_CLINICAL\_0003`
- `C2-EV\_CLINICAL\_0004`
- `C2-EV\_CLINICAL\_0005`
- `C2-EV\_BIOMARKER\_0001`
- `C2-EV\_BIOMARKER\_0002`
- `C2-EV\_TRACTABILITY\_0002`
- `C2-EV\_TRACTABILITY\_0005`
- `C1-CAT-ONCO-BIO-091b`
- `C1-CAT-ONCO-BIO-088`

Contradicting evidence:

- `C2-EV\_CLINICAL\_0006`
- `C2-EV\_BIOMARKER\_0004`
- `C2-EV\_BIOMARKER\_0006`

Missing evidence: For EGFR, indication-matched randomized clinical-trial data with explicit molecule, dose, achieved exposure, comparator, biomarker-defined population, and endpoint results comparable in detail to the ALK dossier.; For both candidates, direct clinical validation specifically in newly diagnosed NSCLC with the stated mucus-producing glandular-cell context and chemotherapy backbone.

Q20 · Integrated therapeutic hypothesis

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK represents the stronger integrated target–disease–modality hypothesis because the dossier connects a causal NSCLC alteration to a defined oncogenic mechanism, selective perturbational dependence, validated biomarker selection, established small-molecule tractability, and multiple randomized clinical demonstrations of benefit over chemotherapy, including durable systemic and intracranial activity in previously untreated disease. The hypothesis is not uniformly applicable: ALK dependence varies by model and fusion variant, resistance is common, toxicity is clinically relevant, and the segment is crowded. EGFR also has a coherent causal and mechanistic rationale and established small-molecule tractability, but its dossier is dominated by repetitive catalogue summaries and lacks comparable study-level randomized clinical evidence for the stated newly diagnosed context or a demonstrated differentiated medicine opportunity. The available EGFR evidence therefore supports plausibility but not an overall hypothesis as strongly as the ALK evidence.

Decisive evidence:

- `C1-EV\_GENETICS\_0001`
- `C1-EV\_MECHANISM\_0002`

- C1-EV_PERTURBATION_0001
- C1-EV_TRACTABILITY_0002
- C1-EV_BIOMARKER_0002
- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0004

Contradicting evidence:

- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007
- C1-EV_COMPETITIVE_0001
- C1-EV_COMPETITIVE_0005
- C2-EV_SAFETY_0001
- C2-CAT-ONCO-BIO-004
- C2-CAT-ONCO-BIO-013

Missing evidence: Direct randomized clinical validation of EGFR inhibition versus chemotherapy in newly diagnosed EGFR-mutant NSCLC, with evidence of differentiated benefit against current EGFR-directed standards.; Direct head-to-head evidence demonstrating that either target provides differentiated value in the specified mucus-producing glandular-cell, patient-derived xenograft, in-vivo proof-of-concept context.; For ALK, evidence specifically in the stated mucus-producing glandular-cell PDX model and a differentiation strategy that overcomes the crowded ALK-TKI segment.

Anomalies: C2-CAT-ONCO-BIO-019 contains an unresolved {Disease} placeholder and requests specification of the disease.; C2-CAT-ONCO-BIO-042, C2-CAT-ONCO-BIO-053, C2-CAT-ONCO-BIO-061, C2-CAT-ONCO-BIO-083, C2-CAT-ONCO-BIO-086, C2-CAT-ONCO-BIO-106, and C2-CAT-ONCO-BIO-112 contain unresolved placeholders or requests for additional inputs.

Q9 · Biomarker and target engagement

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK offers the more credible end-to-end clinical biomarker strategy. ALK rearrangement is explicitly established as a regulatory predictive biomarker requiring an FDA-approved test, has randomized evidence of superior efficacy versus platinum-pemetrexed chemotherapy in biomarker-selected patients, has an established companion-diagnostic infrastructure, and has longitudinal plasma mutation kinetics consistent with pharmacodynamic target engagement (C2-EV_BIOMARKER_0001, C2-EV_BIOMARKER_0002, C2-EV_TRACTABILITY_0006, C2-EV_BIOMARKER_0005, C2-EV_CLINICAL_0001). Direct ALK knockdown selectively affecting ALK-positive models further supports mechanistic engagement and early efficacy interpretation (C2-EV_PERTURBATION_0001). EGFR has strong clinical target precedent and catalogue evidence for mutation testing and pEGFR as a biological readout, but the supplied dossier provides less direct, dossier-level evidence connecting a validated clinical assay, pharmacodynamic measurement, and early efficacy signal than the ALK evidence. ALK's strategy is not perfect: FISH may detect nonproductive rearrangements and ALK positivity does not guarantee durable dependence (C2-EV_BIOMARKER_0004, C2-EV_BIOMARKER_0006), but these limitations do not outweigh its explicit regulatory, diagnostic, pharmacodynamic, and randomized efficacy chain.

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_BIOMARKER_0002
- C2-EV_BIOMARKER_0005
- C2-EV_TRACTABILITY_0006
- C2-EV_CLINICAL_0001
- C2-EV_PERTURBATION_0001

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_TISSUE_0004
- C2-EV_MECHANISM_0006

Missing evidence: A direct, head-to-head comparison of validated EGFR and ALK biomarker and pharmacodynamic strategies in the specified newly diagnosed NSCLC, chemotherapy-backbone, in-vivo PDX context.; For EGFR, a dossier-level regulatory companion-diagnostic record plus prospective clinical pharmacodynamic and early-efficacy data directly linking the selected EGFR genomic alteration to target engagement.

Q6 · Translational model validity

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has direct efficacy evidence in patient-derived NSCLC cells and in vivo validation, including a patient-derived ALK-rearranged model, plus randomized efficacy evidence in previously untreated advanced ALK-positive NSCLC. KRAS has strong genetic and mechanistic support, but the dossier does not provide comparable direct efficacy results in clinically relevant newly diagnosed NSCLC patient-derived or genetic models; its cited xenograft efficacy example is not NSCLC. ALK model heterogeneity and variable dependency temper, but do not overturn, the stronger translational efficacy evidence.

Decisive evidence:

- C1-EV_PERTURBATION_0001
- C1-EV_PERTURBATION_0002
- C1-EV_SAFETY_0002
- C1-EV_CLINICAL_0001

Contradicting evidence:

- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007

Missing evidence: Direct KRAS efficacy results in newly diagnosed NSCLC-relevant patient-derived organoids, primary patient cells, NSCLC PDXs, or genetically engineered lung models, with comparator and treatment-response data.

Anomalies: C2-CAT-ONCO-BIO-018; C2-CAT-ONCO-BIO-037; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-083; C2-CAT-ONCO-BIO-111

Q11 · Competitive differentiation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK offers stronger evidence of clinically meaningful differentiation over chemotherapy: in untreated advanced ALK-positive NSCLC, crizotinib improved progression-free survival and response rate versus platinum-pemetrexed chemotherapy (C2-EV_BIOMARKER_0002; C2-EV_CLINICAL_0001). Subsequent ALK inhibitors provide evidence of differentiated long-term and intracranial control (C2-EV_COMPETITIVE_0002) and broad activity against resistance mutations (C2-EV_COMPETITIVE_0004), while the documented fourth-generation pipeline shows an active differentiation landscape (C2-EV_COMPETITIVE_0005). The ALK case is not uniformly differentiated:

there is substantial competition without direct head-to-head superiority among preferred agents (C2-EV_COMPETITIVE_0001), limited differentiation in one head-to-head comparison (C2-EV_COMPETITIVE_0006), and resistance or non-ALK-dependent progression can occur (C2-EV_BIOMARKER_0004). EGFR has strong mechanistic and tractability evidence, but the supplied dossier does not provide comparable direct chemotherapy-comparator or competitor-landscape evidence for the stated newly diagnosed setting. This verdict addresses scientific and clinical differentiation evidence, not market attractiveness.

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_BIOMARKER_0002
- C2-EV_COMPETITIVE_0002
- C2-EV_COMPETITIVE_0004
- C2-EV_COMPETITIVE_0005

Contradicting evidence:

- C2-EV_COMPETITIVE_0001
- C2-EV_COMPETITIVE_0006
- C2-EV_BIOMARKER_0004

Missing evidence: For EGFR: a randomized, newly diagnosed NSCLC comparison of an EGFR-directed small molecule against chemotherapy, with clinically meaningful endpoints such as progression-free survival, overall survival, or intracranial control.; For EGFR: a documented current competitor pipeline and direct evidence of differentiation versus existing EGFR-directed treatments.

Q10 · Clinical validation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) MET · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: Both targets have indication-relevant clinical validation through biomarker-selected responses, approved small-molecule inhibitors, and evidence of on-target activity. MET has strong clinical evidence in MET exon 14-skipping NSCLC, including durable responses with capmatinib and traditional approval of capmatinib and tepotinib (C1-EV_CLINICAL_0001, C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0003). However, the principal MET evidence is phase 2 and largely non-randomized, and validation is highly dependent on the specific alteration; lower-level MET amplification and nonspecific MET IHC were not reliably predictive (C1-EV_CLINICAL_0006, C1-EV_BIOMARKER_0004). ALK has stronger indication-relevant validation because it includes randomized phase 3 evidence against platinum-pemetrexed chemotherapy in untreated advanced ALK-positive NSCLC, with improved progression-free survival, response rate, symptoms, and quality of life (C2-EV_CLINICAL_0001, C2-EV_BIOMARKER_0002), followed by randomized phase 3 confirmation with alectinib, brigatinib, and lorlatinib showing systemic and intracranial benefit (C2-EV_CLINICAL_0002, C2-EV_CLINICAL_0003, C2-EV_CLINICAL_0004). Regulatory approval of ensartinib based on randomized phase 3 evidence further supports reproducibility across molecules (C2-EV_CLINICAL_0005). The lack of unadjusted overall-survival significance in PROFILE 1014 is a limitation, but substantial crossover makes this less informative about target validity (C2-EV_CLINICAL_0006). The ALK evidence therefore has the stronger trial design, comparator, endpoint breadth, treatment-naive population relevance, and cross-molecule reproducibility for the stated NSCLC inhibition context.

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0003
- C2-EV_CLINICAL_0004

- C2-EV_CLINICAL_0005
- C2-EV_BIOMARKER_0002
- C2-EV_TRACTABILITY_0005
- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0003

Contradicting evidence:

- C2-EV_CLINICAL_0006
- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C1-EV_CLINICAL_0006
- C1-EV_BIOMARKER_0004
- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007

Missing evidence: Direct randomized comparative clinical evidence for MET inhibition versus chemotherapy or another active control in newly diagnosed, biomarker-selected NSCLC, with exposure-confirmed target engagement and outcomes specifically reported for mucus-producing glandular tumor cells.; Direct comparative pharmacokinetic/exposure and dose-response data between the MET and ALK molecules in the stated indication and model context.; Clinical validation specifically in the stated mucus-producing glandular-cell population and in a patient-derived xenograft-to-clinical translational bridge.

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has the clearest operationalized patient-selection hypothesis: MET exon 14 skipping was prospectively assayed using tissue RNA or plasma cfDNA/NGS, linked to response in a defined NSCLC population, and incorporated into FDA-approved testing and indications (C1-EV_BIOMARKER_0001, C1-EV_CLINICAL_0003, C1-EV_TRACTABILITY_0002). KRAS has strong genomic evidence for recurrent, prospectively identifiable molecular subsets and established mutant-selective small-molecule tractability (C2-EV_GENETICS_0003, C2-EV_TRACTABILITY_0001), but the supplied dossier does not provide comparably direct evidence of a prospectively specified, biomarker-selected NSCLC treatment study. MET therefore supports the clearer prospectively testable selection hypothesis.

Decisive evidence:

- C1-EV_BIOMARKER_0001
- C1-EV_CLINICAL_0003
- C1-EV_TRACTABILITY_0002
- C1-EV_CLINICAL_0001
- C2-EV_GENETICS_0003
- C2-EV_TRACTABILITY_0001

Contradicting evidence:

- C1-EV_BIOMARKER_0006
- C1-EV_CLINICAL_0006

Missing evidence: For KRAS: a prospective NSCLC treatment-selection study using a prespecified genomic assay and reporting outcomes for a defined KRAS molecular subgroup, particularly KRAS G12C, rather than evidence limited to prospective prevalence profiling or general clinical tractability.

Anomalies: C2-CAT-ONCO-BIO-018 contains a request to specify the disease placeholder.; C2-CAT-ONCO-BIO-026 contains a request for additional SNP identifiers before answering.;

C2-CAT-ONCO-BIO-037 contains a request to specify the disease placeholder.; C2-CAT-ONCO-BIO-053 contains a request to specify the disease placeholder.; C2-CAT-ONCO-BIO-083 requests clarification of the proposed hypothesis.; C2-CAT-ONCO-BIO-106 contains a request to specify the disease placeholder.; C2-CAT-ONCO-BIO-111 contains a request to specify the disease placeholder.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.94

Presented as: 1) MET · 2) EGFR

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has explicit pharmacologic perturbation evidence in six MET-amplified lung-cancer PDX models showing antitumor responses, plus independent patient-derived EGFR-mutant/MET-amplified models responding to MET inhibition in vivo (C1-EV_PERTURBATION_0002, C1-EV_PERTURBATION_0003). The evidence is biomarker-restricted rather than universal, but that is consistent with the defined genomic-biomarker context. The EGFR dossier contains broad catalogue assertions about xenograft and PDX studies, but no comparably specific, model-level PDX result demonstrating reproducible disease-phenotype reversal. MET therefore has stronger reproducible perturbational evidence for this criterion.

Decisive evidence:

- `C1-EV\_PERTURBATION\_0002`
- `C1-EV\_PERTURBATION\_0003`

Contradicting evidence:

- `C1-EV\_PERTURBATION\_0004`
- `C1-EV\_PERTURBATION\_0005`

Missing evidence: A model-specific, reproducible EGFR perturbation study using patient-derived xenografts, with documented tumor regression or reversal of disease-relevant phenotypes and preferably orthogonal genetic and pharmacologic perturbations.

Q11 · Competitive differentiation

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) ALK · 2) MET

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK offers stronger evidence for clinically meaningful differentiation over chemotherapy in newly diagnosed disease: randomized phase 3 evidence showed superior progression-free survival and response rate versus platinum-pemetrexed chemotherapy in biomarker-selected patients. Its differentiation is also supported by durable systemic and intracranial control, although the ALK segment is crowded and incremental differentiation among preferred ALK inhibitors is not established. MET has clinically meaningful treatment-naive response data in METex14-positive disease and a stated opportunity relative to non-targeted treatment, but the dossier does not provide a direct randomized comparison against chemotherapy. MET differentiation is further limited by multiple approved or clinically active MET inhibitors and competition from a non-TKI modality. This conclusion concerns evidence for scientific/clinical differentiation over chemotherapy, not market attractiveness.

Decisive evidence:

- `C1-EV\_CLINICAL\_0001`
- `C1-EV\_BIOMARKER\_0002`
- `C1-EV\_COMPETITIVE\_0001`
- `C1-EV\_COMPETITIVE\_0002`
- `C2-EV\_COMPETITIVE\_0003`

- C2-EV_CLINICAL_0001

Contradicting evidence:

- C1-EV_COMPETITIVE_0005
- C1-EV_COMPETITIVE_0006
- C2-EV_COMPETITIVE_0005
- C2-EV_COMPETITIVE_0006
- C2-EV_CLINICAL_0006

Missing evidence: A randomized first-line trial directly comparing a MET-directed therapy in METex14-positive newly diagnosed NSCLC with chemotherapy, with clinically meaningful efficacy and intracranial outcomes

Q8 · Druggability and tractability

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: Both candidates have evidence supporting small-molecule tractability in lung cancer. KRAS has a clinically validated covalent small-molecule modality, but the strongest evidence is allele-specific and concentrated on KRAS-G12C, with additional resistance and bypass mechanisms documented. ALK has broader and more mature tractability evidence: multiple approved oral small-molecule TKIs, regulatory companion-diagnostic infrastructure, randomized clinical efficacy in NSCLC, and evidence of systemic and intracranial activity. These data directly support binding-site accessibility, modality feasibility, delivery to lung cancer, and therapeutic precedent. ALK therefore has the stronger overall small-molecule tractability dossier, while acknowledging that exact evidence in mucus-producing glandular cells is missing.

Decisive evidence:

- C2-EV_TRACTABILITY_0001
- C2-EV_TRACTABILITY_0002
- C2-EV_TRACTABILITY_0003
- C2-EV_TRACTABILITY_0004
- C2-EV_TRACTABILITY_0006
- C2-EV_CLINICAL_0001
- C1-EV_TRACTABILITY_0001
- C1-CAT-ONCO-BIO-032
- C1-CAT-ONCO-BIO-115

Contradicting evidence:

- C2-EV_TRACTABILITY_0007
- C2-EV_BIOMARKER_0004
- C1-CAT-ONCO-BIO-013
- C1-CAT-ONCO-BIO-063

Missing evidence: Direct target-expression, binding-site engagement, intracellular exposure, and pharmacodynamic delivery data for ALK and KRAS specifically in mucus-producing glandular cells of lung PDX tumors.; Comparative lung tissue and cell-type distribution data for the candidate small molecules in the specified mucus-producing glandular-cell context.; For KRAS, tractability evidence for the full stated KRAS scope, particularly G12D, rather than predominantly clinically validated G12C.

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK has a more auditable evidence structure: it includes distinct source families such as an original discovery study, TCGA profiling, a large real-world cohort, multicenter tissue pathology, functional perturbation studies, and regulatory or clinical sources. It also explicitly preserves contradictory findings, including variable model dependence and bypass resistance. These features support assessment of replication and publication-bias vulnerability, even though some ALK items share the original discovery study or resistance experiments. EGFR has relevant evidence, but its dossier is overwhelmingly composed of catalogue-answer summaries with sparse provenance and no documented sample-overlap, replication, or negative-result framework. Thus its apparent breadth cannot be treated as independent robustness under this meta-evidence criterion.

Decisive evidence:

- `C1-EV\_GENETICS\_0002`
- `C1-EV\_GENETICS\_0003`
- `C1-EV\_GENETICS\_0005`
- `C1-EV\_TISSUE\_0002`
- `C1-EV\_MECHANISM\_0006`
- `C1-EV\_MECHANISM\_0007`
- `C1-EV\_PERTURBATION\_0005`
- `C2-EV\_GENETICS\_0001`
- `C2-EV\_MECHANISM\_0001`
- `C2-EV\_TRACTABILITY\_0001`
- `C2-EV\_SAFETY\_0001`

Contradicting evidence:

- `C1-EV\_MECHANISM\_0006`
- `C1-EV\_MECHANISM\_0007`
- `C1-EV\_PERTURBATION\_0004`
- `C1-EV\_PERTURBATION\_0005`
- `C2-EV\_SAFETY\_0001`

Missing evidence: For EGFR: independently identified primary-study references with provenance, cohort/sample-overlap flags, explicit replication status, and available negative or null results; the current dossier is dominated by catalogue summaries.; For both candidates: a systematic source-family and sample-overlap audit across all listed studies, especially for repeated cohorts and summaries of the same experiments.; For EGFR: directly documented contradictory findings or failed critical experiments in the specified NSCLC context, rather than only general statements of context dependence.

Q2 · Disease biology and mechanism coherence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has a coherent causal chain from recurrent human NSCLC driver alterations to a defined molecular mechanism: METex14 removes the CBL-binding region, impairs receptor down-regulation, sustains MET signaling, and activates downstream MAPK pathways (C1-EV_GENETICS_0002, C1-EV_MECHANISM_0002, C1-EV_MECHANISM_0003). This is reinforced by selective genetic dependency and pharmacologic inhibition in MET-amplified NSCLC models, including patient-derived xenografts (C1-EV_PERTURBATION_0001, C1-EV_PERTURBATION_0002). KRAS has strong human-genetic plausibility and a coherent RAS-MAPK model (C2-EV_GENETICS_0002, C2-EV_GENETICS_0003, C2-EV_MECHANISM_0001), but the supplied structured evidence contains substantially less direct, primary, NSCLC-specific perturbational validation of inhibition altering disease biology. MET evidence is not universal: dependency is biomarker-restricted and heterogeneous, which limits generalization but does not overturn the stronger mechanistic support for the defined

MET-altered subset. This judgment concerns biological coherence and experimental support, not druggability.

Decisive evidence:

- C1-EV_MECHANISM_0002
- C1-EV_MECHANISM_0003
- C1-EV_PERTURBATION_0001
- C1-EV_PERTURBATION_0002
- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003
- C2-EV_MECHANISM_0001
- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0003

Contradicting evidence:

- C1-EV_PERTURBATION_0004
- C1-EV_PERTURBATION_0005
- C1-EV_GENETICS_0005
- C1-EV_TISSUE_0006

Missing evidence: For KRAS: primary, disease-specific mechanistic perturbation studies in NSCLC using genetic knockdown, knockout, or allele correction, ideally in patient-derived models and with pathway readouts.; For both targets: direct pathology and perturbation evidence specifically in the stated mucus-producing glandular lung-cell context.

Anomalies: C2-CAT-ONCO-BIO-018; C2-CAT-ONCO-BIO-037; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-083

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has the more robust meta-evidence package. Its dossier explicitly identifies multiple shared cohorts and prevents treating dependent reports as independent, while also documenting contradictory findings and negative clinical, biomarker, tissue, and perturbation results (C2-EV_MECHANISM_0006, C2-EV_MECHANISM_0007, C2-EV_BIOMARKER_0004, C2-EV_PERTURBATION_0004, C2-EV_PERTURBATION_0005, C2-EV_PERTURBATION_0006). It also spans distinct source families including genomic cohorts, mechanistic studies, randomized trials, phase 2 trials, regulatory records, and patient-derived models. The KRAS dossier contains supporting human-genomic and mechanistic claims but does not supply comparable provenance metadata, explicit sample-overlap analysis, replication status, or systematic negative-result availability. This is a metadata robustness advantage for MET, not a judgment based on the larger number of MET items.

Decisive evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_BIOMARKER_0004
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0006
- C2-EV_PERTURBATION_0004
- C2-EV_PERTURBATION_0005
- C2-EV_PERTURBATION_0006
- C2-EV_GENETICS_0005
- C2-EV_GENETICS_0007

Contradicting evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_BIOMARKER_0004
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0005
- C2-EV_TISSUE_0006
- C2-EV_PERTURBATION_0004
- C2-EV_PERTURBATION_0005
- C2-EV_PERTURBATION_0006
- C2-EV_GENETICS_0005
- C2-EV_GENETICS_0007

Missing evidence: For KRAS, explicit provenance metadata, source-family structure, cohort-overlap flags, replication status, and systematically reported negative-result availability comparable to the MET dossier; For both candidates, a complete publication-level audit identifying all shared cohorts, unpublished or null studies, and independent replication datasets

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has the more robust meta-evidence package. Its dossier identifies distinct source families and study designs, including a primary discovery study, independent tumor-genome profiling, large prevalence cohorts, tissue cohorts, perturbation studies, randomized clinical trials, regulatory sources, and explicitly reported contradictory or limiting findings. The contradictions are not hidden: ALK dependence is shown to vary across models, fusion variants, resistance mechanisms, and diagnostic assays (C2-EV_MECHANISM_0006, C2-EV_MECHANISM_0007, C2-EV_BIOMARKER_0004, C2-EV_BIOMARKER_0006). EGFR contains strong biological claims but is dominated by numerous catalogue-style entries with limited provenance, replication-status, cohort-overlap, or negative-result metadata. Under this criterion, ALK therefore has superior source structure and contradiction transparency, despite some documented dependence among individual ALK items.

Decisive evidence:

- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0003
- C2-EV_GENETICS_0005
- C2-EV_MECHANISM_0003
- C2-EV_PERTURBATION_0001
- C2-EV_PERTURBATION_0002
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0006

Contradicting evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_TISSUE_0004
- C2-EV_PERTURBATION_0005
- C2-EV_PERTURBATION_0006
- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006

Missing evidence: For EGFR, item-level source provenance, independent replication mapping,

sample-overlap flags, and systematically identified negative or null studies would most change the comparison.; For both candidates, a formal publication-bias assessment and complete study-level cohort-overlap table are not supplied.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) EGFR · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has specific, high-trust perturbational evidence across orthogonal modalities: MET knockdown selectively impaired growth and survival in MET-amplified NSCLC models, pharmacologic MET inhibition produced antitumor responses in six MET-amplified PDX models, and a separate patient-derived model series showed responses to MET inhibition alone in EGFR-mutant, MET-amplified disease. The evidence is biomarker-restricted rather than universal, and CRISPR knockout in A549 plus nonselective tivantinib activity provide countervailing findings, but these do not directly negate the PDX evidence. EGFR has catalogue-level statements that tools were used in PDX studies, but the dossier does not provide comparably specific, reproducible EGFR PDX perturbation results or orthogonal intervention data. Therefore MET has stronger reproducible evidence for reversing disease-relevant phenotypes in PDX.

Decisive evidence:

- `C2-EV\_PERTURBATION\_0001`
- `C2-EV\_PERTURBATION\_0002`
- `C2-EV\_PERTURBATION\_0003`

Contradicting evidence:

- `C2-EV\_PERTURBATION\_0005`
- `C2-EV\_PERTURBATION\_0006`
- `C2-EV\_PERTURBATION\_0004`

Missing evidence: A directly described EGFR perturbation study in patient-derived xenografts showing disease-relevant phenotype reversal, ideally with orthogonal genetic and pharmacologic perturbations and biomarker-matched models.; For MET, direct confirmation that the reported patient-derived xenograft models represent the specified mucus-producing glandular-cell context.

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET supports the clearer operational patient-selection hypothesis: MET exon 14 skipping was prospectively detected using tissue RNA or plasma cfDNA/NGS and was associated with responses to selective MET inhibitors in clinical studies, with an FDA-approved test specifying the eligible population (C2-EV_BIOMARKER_0001, C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0003). Serial METex14 ctDNA depletion also provides an operational pharmacodynamic measure (C2-EV_BIOMARKER_0002). KRAS has strong evidence for recurrent genotype-defined NSCLC subsets and prospective genomic profiling (C1-EV_GENETICS_0002, C1-EV_GENETICS_0003), but the dossier does not provide comparably direct prospective interventional evidence tying a prespecified KRAS subgroup-selection strategy to treatment outcomes. MET's hypothesis is not universal across all MET assays: MET IHC alone failed as a selector and amplification thresholds are less reliable (C2-EV_BIOMARKER_0004, C2-EV_TISSUE_0004, C2-EV_BIOMARKER_0006), but the METex14 genomic subtype remains clearly actionable.

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0003
- C2-EV_BIOMARKER_0002

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_TISSUE_0004
- C2-EV_BIOMARKER_0006
- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003

Missing evidence: A prospective interventional NSCLC trial directly demonstrating treatment benefit in a KRAS-defined subgroup, with a prespecified genomic selection assay and outcome comparison, would most narrow the difference.

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.93

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK supports the clearer prospectively testable selection hypothesis. ALK rearrangement is defined by an actionable molecular alteration, has FDA-authorized companion-diagnostic infrastructure, and was used to select patients in randomized clinical trials comparing ALK inhibition with chemotherapy. The genetic driver and its operational assay are therefore directly linked to prospective enrollment. EGFR has strong evidence for mutation-defined NSCLC subgroups and clinical targetability, but the supplied dossier is less direct about a single prospectively operationalized selection rule and relies more heavily on heterogeneous catalogue evidence. The ALK limitations—subclonality, nonproductive FISH results, and resistance—complicate interpretation but do not negate the feasibility of prospective ALK-positive selection.

Decisive evidence:

- C2-EV_BIOMARKER_0001
- C2-EV_TRACTABILITY_0006
- C2-EV_CLINICAL_0001
- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0005

Contradicting evidence:

- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_TISSUE_0004

Missing evidence: A comparably direct prospective, biomarker-selected clinical validation package for EGFR mutation-defined NSCLC, including an explicitly specified assay and trial eligibility rule, would most challenge the preference for ALK.

Q4 · Human natural experiments

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.92

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has direct human pharmacological evidence matching the requested inhibitory direction: genotype-selected MET inhibition produced substantial and durable responses in MET exon

14-skipping NSCLC, including in treatment-naive patients, with a second inhibitor confirming activity (C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0002, C2-EV_SAFETY_0002, C2-EV_SAFETY_0005). Toxicities are substantial but documented and monitorable, providing evidence relevant to tolerability (C2-EV_SAFETY_0004, C2-EV_SAFETY_0006, C2-EV_SAFETY_0007). KRAS has human somatic driver and germline RASopathy evidence, but no supplied human loss-of-function, protective-allele, or NSCLC pharmacological exposure evidence demonstrating both efficacy and tolerability of KRAS inhibition. The activating genetic direction for MET does not itself match inhibition, so the decision rests on the human pharmacological evidence rather than the genetic alterations.

Decisive evidence:

- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_SAFETY_0002
- C2-EV_SAFETY_0005

Contradicting evidence:

- C2-EV_SAFETY_0004
- C2-EV_SAFETY_0006
- C2-EV_SAFETY_0007
- C2-EV_CLINICAL_0006

Missing evidence: For KRAS: human pharmacological exposure data directly demonstrating efficacy and tolerability of KRAS inhibition in NSCLC, or a human loss-of-function/protective-allele phenotype whose direction matches inhibition.; For MET: long-term exposure cohorts with mature tolerability data specifically separated from efficacy evidence.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.92

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has the stronger reproducible PDX perturbational evidence. Pharmacologic MET inhibition produced antitumor responses across six MET-amplified lung-cancer PDX models, and a separate set of six patient-derived EGFR-mutant/MET-amplified models showed MET-governed signaling and sensitivity to MET inhibition in vivo (C2-EV_PERTURBATION_0002, C2-EV_PERTURBATION_0003). Genetic MET knockdown independently showed selective growth inhibition, arrest, and apoptosis in MET-amplified NSCLC cell lines (C2-EV_PERTURBATION_0001), supporting biomarker-restricted on-target dependence. The evidence is not universal: effects were largely confined to molecularly defined subsets (C2-EV_PERTURBATION_0004), and CRISPR knockout in A549 cells did not impair intrinsic proliferation although it affected dissemination-related phenotypes (C2-EV_PERTURBATION_0005). KRAS has supplied statements that tool compounds were used in PDX models, but the PDX studies and results are not comparably explicit or reproducibly characterized in the dossier (C1-CAT-ONCO-BIO-010, C1-CAT-ONCO-BIO-117b).

Decisive evidence:

- C2-EV_PERTURBATION_0001
- C2-EV_PERTURBATION_0002
- C2-EV_PERTURBATION_0003

Contradicting evidence:

- C2-EV_PERTURBATION_0004
- C2-EV_PERTURBATION_0005
- C2-EV_PERTURBATION_0006

Missing evidence: For KRAS, explicit, independently replicated PDX perturbation studies with defined

KRAS-altered NSCLC models, intervention details, and disease-relevant response measurements; the supplied KRAS PDX evidence is generic or insufficiently specified.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) MET · 2) ALK

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has explicit pharmacologic perturbation evidence in six MET-amplified lung-cancer patient-derived xenograft models, with MET-inhibition-associated signaling suppression and antitumor responses, plus concordant activity across additional patient-derived models in vivo (C1-EV_PERTURBATION_0002, C1-EV_PERTURBATION_0003). The ALK dossier contains direct knockdown evidence in cell models and CRISPR resistance-screen evidence, but does not provide comparably direct ALK modulation producing disease-relevant responses in multiple patient-derived xenografts. MET evidence is therefore stronger for this specific PDX perturbational criterion, despite limited orthogonal genetic validation in PDX.

Decisive evidence:

- `C1-EV\_PERTURBATION\_0002`
- `C1-EV\_PERTURBATION\_0003`

Contradicting evidence:

- `C1-EV\_PERTURBATION\_0005`
- `C2-EV\_PERTURBATION\_0002`
- `C2-EV\_PERTURBATION\_0005`

Missing evidence: Direct ALK knockdown or knockout in multiple ALK-positive patient-derived xenografts with tumor-regression or disease-control readouts, ideally with rescue or an orthogonal perturbation.; Orthogonal genetic MET perturbation validated in MET-dependent patient-derived xenografts.

Q20 · Integrated therapeutic hypothesis

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET provides the more coherent target–disease–modality hypothesis for this context: recurrent NSCLC driver biology, a defined mechanism of constitutive activation, selective perturbational and PDX evidence, validated pharmacodynamic biomarkers, oral small-molecule tractability, and direct clinical responses in treatment-naive METex14 NSCLC. KRAS has strong driver genetics and established small-molecule tractability, but the supplied KRAS evidence is dominated by broad catalogue material and lacks comparable direct clinical-validation, newly diagnosed, or NSCLC PDX evidence for the proposed setting. MET is not an unqualified choice: activity is alteration-dependent, resistance and toxicity are documented, and the class is competitively crowded. Those limitations reduce differentiation but do not offset the substantially stronger integrated evidence base.

Decisive evidence:

- `C1-EV\_GENETICS\_0002`
- `C1-EV\_MECHANISM\_0002`
- `C1-EV\_PERTURBATION\_0002`
- `C1-EV\_SAFETY\_0002`
- `C1-EV\_TRACTABILITY\_0002`
- `C1-EV\_BIOMARKER\_0001`

- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002
- C2-EV_GENETICS_0002
- C2-EV_MECHANISM_0001
- C2-EV_TRACTABILITY_0001

Contradicting evidence:

- C1-EV_GENETICS_0005
- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007
- C1-EV_PERTURBATION_0004
- C1-EV_SAFETY_0004
- C1-EV_CLINICAL_0006
- C1-EV_COMPETITIVE_0005
- C1-EV_COMPETITIVE_0006
- C2-CAT-ONCO-BIO-004
- C2-CAT-ONCO-BIO-052
- C2-CAT-ONCO-BIO-054
- C2-CAT-ONCO-BIO-063

Missing evidence: The Q1–Q19 scores and criterion-weighted composite required by the Q20 guardrail were not supplied; Direct head-to-head evidence in newly diagnosed, mucus-producing glandular-cell NSCLC PDX models comparing MET inhibition with KRAS inhibition on a chemotherapy backbone; Clinical validation of a KRAS small-molecule strategy specifically in newly diagnosed NSCLC, beyond the supplied general tractability evidence; Evidence that either target produces differentiated efficacy over existing targeted competitors in the newly diagnosed setting

Anomalies: C2-CAT-ONCO-BIO-018 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-037 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-053 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-083 states that the proposed hypothesis is missing and asks the reader to assume one; C2-CAT-ONCO-BIO-106 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-111 contains a request to specify the disease placeholder

Q2 · Disease biology and mechanism coherence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) MET · 2) EGFR

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has a more experimentally triangulated mechanism: recurrent human METex14 driver alterations are linked to impaired receptor down-regulation and sustained MAPK signaling, and genetic and pharmacologic perturbation selectively affects MET-dependent NSCLC models, including six MET-amplified patient-derived xenografts (C1-EV_GENETICS_0002; C1-EV_GENETICS_0003; C1-EV_MECHANISM_0002; C1-EV_PERTURBATION_0001; C1-EV_PERTURBATION_0002; C1-EV_PERTURBATION_0003). EGFR also has strong mechanistic plausibility and causal activating-mutation evidence, but much of the supplied support is catalogue-level synthesis rather than direct primary perturbation evidence, and the dossier explicitly describes context-dependent dependence (C2-EV_GENETICS_0001; C2-EV_MECHANISM_0001; C2-CAT-ONCO-BIO-031; C2-CAT-ONCO-BIO-070). MET's limitations indicate biomarker-restricted rather than incoherent biology, so they do not outweigh its stronger direct experimental support for inhibition altering NSCLC pathophysiology.

Decisive evidence:

- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003

- C1-EV_MECHANISM_0002
- C1-EV_PERTURBATION_0001
- C1-EV_PERTURBATION_0002
- C1-EV_PERTURBATION_0003
- C2-EV_GENETICS_0001
- C2-EV_MECHANISM_0001

Contradicting evidence:

- C1-EV_GENETICS_0005
- C1-EV_PERTURBATION_0004
- C1-EV_PERTURBATION_0005
- C2-CAT-ONCO-BIO-031
- C2-CAT-ONCO-BIO-070

Missing evidence: Directly comparable, primary mechanistic perturbation studies of MET versus EGFR inhibition in the specified mucus-producing glandular NSCLC cell context, including matched patient-derived xenograft models.

Anomalies: C2-CAT-ONCO-BIO-019; C2-CAT-ONCO-BIO-042; C2-CAT-ONCO-BIO-053; C2-CAT-ONCO-BIO-061; C2-CAT-ONCO-BIO-083; C2-CAT-ONCO-BIO-086; C2-CAT-ONCO-BIO-112

Q7 · Safety and essentiality

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has the more favourable anticipated on-target safety profile. The dossier supports relatively restricted adult expression and viable, fertile homozygous ALK-knockout mice, although with reproductive and neuroendocrine phenotypes (C2-EV_SAFETY_0001, C2-EV_SAFETY_0004). In contrast, EGFR is broadly important in normal epithelia, including lung, skin and gastrointestinal tissues; constitutive knockout causes severe epithelial, developmental and respiratory phenotypes (C1-CAT-ONCO-BIO-098, C1-CAT-ONCO-BIO-102, C1-CAT-ONCO-BIO-114). Direct on-target EGFR inhibition is also associated with dermatologic and gastrointestinal toxicity (C1-EV_SAFETY_0001). ALK inhibitors have important clinical toxicities, including hepatotoxicity, pneumonitis, cardiac, visual, metabolic and neurologic effects (C2-EV_SAFETY_0005, C2-EV_SAFETY_0006, C2-EV_SAFETY_0007), but these are less compelling evidence of broad target essentiality and may include molecule- or modality-specific components. Therefore, despite residual ALK safety concerns, ALK is favoured for anticipated on-target safety.

Decisive evidence:

- C2-EV_SAFETY_0001
- C2-EV_SAFETY_0004
- C1-CAT-ONCO-BIO-102
- C1-CAT-ONCO-BIO-098
- C1-CAT-ONCO-BIO-114
- C1-EV_SAFETY_0001

Contradicting evidence:

- C2-EV_SAFETY_0005
- C2-EV_SAFETY_0006
- C2-EV_SAFETY_0007

Missing evidence: Direct comparative human loss-of-function and PheWAS data for EGFR versus ALK.; Quantitative, matched tissue-expression breadth and pan-essentiality scores for both targets, including the specified lung mucus-producing glandular cells.; Direct human genetic evidence defining the adult normal-tissue consequences of ALK loss.

Q14 · Patient stratification

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) EGFR · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET supports the clearer prospectively testable selection hypothesis. MET exon 14 skipping is a recurrent genomic driver with diverse splice-site alterations, and patients were prospectively selected using tissue RNA or plasma cfDNA/NGS in clinical studies, with regulatory labeling requiring an FDA-approved test. This directly demonstrates an operational genomic biomarker-to-treatment hypothesis. EGFR has strong evidence for a molecularly defined NSCLC subgroup and clinical feasibility of mutation-specific targeting, but the supplied evidence is less explicit about prospective trial selection and operational validation. MET amplification and MET protein expression are weaker or heterogeneous selection strategies, but these limitations do not outweigh the unusually direct prospective evidence for METex14.

Decisive evidence:

- `C2-EV\_BIOMARKER\_0001`
- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0003`
- `C2-EV\_GENETICS\_0003`
- `C2-EV\_GENETICS\_0004`
- `C1-CAT-ONCO-BIO-091b`
- `C1-CAT-ONCO-BIO-036`

Contradicting evidence:

- `C2-EV\_CLINICAL\_0006`
- `C2-EV\_TISSUE\_0004`
- `C2-EV\_BIOMARKER\_0004`

Missing evidence: For EGFR, prospective biomarker-selected clinical-trial evidence explicitly demonstrating operational testing of Exon-19 deletion or L858R selection in the stated treatment context would most affect the comparison.; For MET amplification, prospective evidence defining a reproducible gene-copy threshold and assay workflow is needed; the dossier supports METex14 more clearly than amplification.

Q6 · Translational model validity

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.91

Presented as: 1) EGFR · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK has the stronger translational-model evidence. Its dossier supplies specific perturbational efficacy in EML4-ALK-positive NSCLC models including a patient-derived NSCLC cell line, genetically engineered mouse models producing lung tumors, and human cancer-cell validation, with additional randomized evidence in previously untreated ALK-positive NSCLC supporting clinical relevance. EGFR has strong biological and target-validation evidence, including described transgenic mouse and PDX applications, but the supplied dossier does not provide comparably specific, independently documented efficacy results from human organoids, primary patient cells, humanised models, or ex vivo tissue. The ALK evidence is therefore more directly tied to mechanistically and clinically relevant NSCLC models, despite documented heterogeneity of ALK dependence and fusion-variant sensitivity.

Decisive evidence:

- `C2-EV\_PERTURBATION\_0001`

- C2-EV_TISSUE_0003
- C2-EV_MECHANISM_0005
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002

Contradicting evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_TISSUE_0004
- C2-EV_PERTURBATION_0006

Missing evidence: Directly comparable EGFR and ALK efficacy studies in matched patient-derived models or human organoids from newly diagnosed NSCLC, including the specified mucus-producing glandular cell context.; For EGFR, a specific peer-reviewed efficacy result in a patient-derived, primary-human, humanised, or ex vivo NSCLC model is not supplied; the dossier mainly provides general model descriptions and catalogue summaries.

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) MET · 2) KRAS

Outcome: Winner: MET (1) · Margin: clear

Reasoning: MET has a substantially better documented meta-evidence structure: its dossier identifies source types and provenance, records explicit cohort and trial dependence, includes independent human-genomic cohorts, replicated clinical validation with different selective inhibitors, and multiple negative or contradictory findings such as METLung, ATTENTION, biomarker discordance, and limited activity in low-level amplification. KRAS has evidence of biological and clinical tractability, but its dossier is dominated by catalogue-style answers and lacks explicit source-family mapping, sample-overlap assessment, replication status, and systematic negative-result availability. This is not treated as evidence that KRAS is biologically weaker; it means its robustness cannot be established from the supplied meta-evidence. The documented dependence and contradiction handling for MET therefore make FIRST more robust on this criterion.

Decisive evidence:

- C1-EV_GENETICS_0002
- C1-EV_GENETICS_0003
- C1-EV_GENETICS_0004
- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007
- C1-EV_BIOMARKER_0004
- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002

Contradicting evidence:

- C1-EV_GENETICS_0005
- C1-EV_GENETICS_0007
- C1-EV_TISSUE_0004
- C1-EV_PERTURBATION_0005
- C2-EV_SAFETY_0001

Missing evidence: For KRAS: explicit provenance metadata, source-family structure, sample-overlap flags, replication status, and availability of negative results across the cited catalogue answers and clinical/preclinical evidence; For KRAS: independent replicated studies and documented contradictory or null findings for the specific NSCLC context, mucus-producing glandular cells, and patient-derived xenograft setting

Anomalies: C2-CAT-ONCO-BIO-018 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-026 contains a request to provide specific KRAS SNP identifiers; C2-CAT-ONCO-BIO-037 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-053 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-083 contains a request to provide the proposed hypothesis; C2-CAT-ONCO-BIO-106 contains a request to specify the disease placeholder; C2-CAT-ONCO-BIO-111 contains a request to specify the disease placeholder

Q20 · Integrated therapeutic hypothesis

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) ALK · 2) KRAS

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: ALK represents the stronger integrated target–disease–modality hypothesis because the dossier directly connects ALK rearrangement to NSCLC biology, provides a validated predictive biomarker and established oral small-molecule modality, and includes multiple randomized first-line clinical studies showing superiority or durable benefit over chemotherapy or crizotinib (C1-EV_BIOMARKER_0001, C1-EV_BIOMARKER_0002, C1-EV_CLINICAL_0001, C1-EV_CLINICAL_0002, C1-EV_CLINICAL_0004). Its main limitation is differentiation in a crowded ALK-TKI segment, together with resistance and toxicity concerns. KRAS has greater mutation prevalence and clinically validated small-molecule tractability, particularly for G12C (C2-EV_GENETICS_0002, C2-EV_GENETICS_0004, C2-EV_TRACTABILITY_0001), but the supplied evidence is less directly relevant to newly diagnosed patients, chemotherapy comparison, the specified PDX setting, and a differentiated G12C/G12D medicine. Much of the KRAS support is catalogue-level, generic, or explicitly context-dependent. Therefore the direct clinical coherence of ALK outweighs KRAS's broader but less mature and less first-line-specific hypothesis.

Decisive evidence:

- C1-EV_CLINICAL_0001
- C1-EV_CLINICAL_0002
- C1-EV_CLINICAL_0004
- C1-EV_BIOMARKER_0001
- C1-EV_BIOMARKER_0002
- C1-EV_TRACTABILITY_0002
- C1-EV_TRACTABILITY_0006
- C1-EV_PERTURBATION_0001
- C2-EV_GENETICS_0002
- C2-EV_GENETICS_0004
- C2-EV_TRACTABILITY_0001

Contradicting evidence:

- C1-EV_COMPETITIVE_0001
- C1-EV_COMPETITIVE_0005
- C1-EV_SAFETY_0005
- C1-EV_SAFETY_0006
- C1-EV_BIOMARKER_0004
- C2-EV_SAFETY_0001
- C2-CAT-ONCO-BIO-004
- C2-CAT-ONCO-BIO-013
- C2-CAT-ONCO-BIO-052
- C2-CAT-ONCO-BIO-061

Missing evidence: Q1-Q19 criterion scores and the criterion-weighted composite required by the Q20 guardrail; A randomized first-line NSCLC study directly testing a KRAS-mutant inhibitor against chemotherapy in newly diagnosed patients; Direct in vivo PDX efficacy evidence for either target in the

specified mucus-producing glandular-cell context

Anomalies: C2-CAT-ONCO-BIO-018 contains a placeholder for {Disease} and requests specification of the disease; C2-CAT-ONCO-BIO-037 contains a placeholder for {Disease} and requests specification of the disease; C2-CAT-ONCO-BIO-053 contains a placeholder for {Disease} and requests specification of the disease; C2-CAT-ONCO-BIO-083 states that the proposed KRAS hypothesis is missing and supplies a conditional substitute; C2-CAT-ONCO-BIO-106 contains a placeholder for {Disease} and requests specification of the disease; C2-CAT-ONCO-BIO-111 contains a placeholder for {Disease} and requests specification of the disease

Q6 · Translational model validity

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) KRAS · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has direct efficacy evidence in patient-derived NSCLC xenograft models, including six MET-amplified lung-cancer PDX models and additional patient-derived EGFR-mutant/MET-amplified models showing in vivo sensitivity to MET inhibition (C2-EV_PERTURBATION_0002, C2-EV_PERTURBATION_0003). Its clinical relevance is further supported by durable responses to selective MET inhibitors in treatment-naive MET exon 14–skipping NSCLC, including capmatinib and tepotinib datasets (C2-EV_CLINICAL_0001, C2-EV_CLINICAL_0002). KRAS has strong human-genetic and clinical tractability evidence, but the dossier does not provide comparably specific, high-quality KRAS efficacy results from admissible translational models in NSCLC or newly diagnosed disease. The MET evidence is biomarker- and alteration-dependent, and the negative or weaker findings for unselected MET models do not negate the stronger evidence in the defined MET-dependent subsets.

Decisive evidence:

- `C2-EV\_PERTURBATION\_0002`
- `C2-EV\_PERTURBATION\_0003`
- `C2-EV\_CLINICAL\_0001`
- `C2-EV\_CLINICAL\_0002`

Contradicting evidence:

- `C2-EV\_PERTURBATION\_0004`
- `C2-EV\_PERTURBATION\_0005`
- `C2-EV\_CLINICAL\_0006`

Missing evidence: Direct KRAS efficacy data from NSCLC patient-derived xenografts, organoids, primary patient cells, genetically engineered or humanized in vivo models, specifically in treatment-naive/newly diagnosed settings.; Direct MET-versus-KRAS comparative efficacy data using the same validated translational model and newly diagnosed treatment context.

Q20 · Integrated therapeutic hypothesis

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) MET · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: ALK represents the stronger integrated target–disease–modality hypothesis. Its oncogenic fusion mechanism is precisely defined, perturbational evidence supports selective dependence in ALK-rearranged NSCLC, and randomized evidence directly demonstrates superiority of ALK inhibition over platinum-pemetrexed chemotherapy in untreated disease (C2-EV_BIOMARKER_0002, C2-EV_CLINICAL_0001). Multiple randomized trials further support strong systemic and intracranial activity, including durable first-line benefit with lorlatinib (C2-EV_CLINICAL_0002, C2-EV_CLINICAL_0004,

C2-EV_TRACTABILITY_0003, C2-EV_TRACTABILITY_0004). MET is also a well-validated, tractable driver with treatment-naive responses and PDX support (C1-EV_SAFETY_0002, C1-EV_CLINICAL_0001, C1-EV_PERTURBATION_0002), but its evidence is more biomarker- and alteration-dependent, especially for amplification (C1-EV_TRACTABILITY_0007), and its clinical differentiation is weakened by substantial class competition and the absence of demonstrated efficacy advantage among MET inhibitors (C1-EV_COMPETITIVE_0005, C1-EV_COMPETITIVE_0006). ALK also faces meaningful competition and resistance, but the direct chemotherapy comparison and stronger intracranial and long-term first-line evidence provide the more coherent differentiated-medicine hypothesis.

Decisive evidence:

- C2-EV_BIOMARKER_0002
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0004
- C2-EV_TRACTABILITY_0003
- C2-EV_TRACTABILITY_0004
- C2-EV_PERTURBATION_0001
- C1-EV_PERTURBATION_0002
- C1-EV_SAFETY_0002
- C1-EV_CLINICAL_0001
- C1-EV_TRACTABILITY_0007

Contradicting evidence:

- C2-EV_COMPETITIVE_0001
- C2-EV_COMPETITIVE_0005
- C2-EV_SAFETY_0005
- C2-EV_BIOMARKER_0004
- C1-EV_COMPETITIVE_0005
- C1-EV_COMPETITIVE_0006
- C1-EV_CLINICAL_0006
- C1-EV_PERTURBATION_0004
- C1-EV_SAFETY_0004

Missing evidence: The Q1-Q19 criterion scores and criterion-weighted composite/posterior are not supplied, so the required Q20 coherence check against those prior judgments cannot be performed.; Direct evidence in the specified mucus-producing glandular-cell context and in the stated newly diagnosed NSCLC patient-derived xenograft context is limited or not clearly identified for either candidate.; Comparative evidence showing differentiated efficacy of a MET- or ALK-directed medicine against current standard therapy specifically in the supplied newly diagnosed PDX context would most affect the comparison.

Q2 · Disease biology and mechanism coherence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) ALK · 2) EGFR

Outcome: Winner: ALK (1) · Margin: clear

Reasoning: FIRST has the more coherent and experimentally supported inhibition mechanism. ALK rearrangements are supported as causal human NSCLC alterations, with transforming activity, defined downstream MAPK signaling, tumor-cell localization, and selective ALK knockdown effects in ALK-positive but not ALK-negative NSCLC models (C1-EV_GENETICS_0001, C1-EV_GENETICS_0002, C1-EV_MECHANISM_0003, C1-EV_MECHANISM_0004, C1-EV_PERTURBATION_0001, C1-EV_TISSUE_0001, C1-EV_TISSUE_0002). The mechanism is not uniform: resistance and variant-specific dependence weaken universal claims, but they do not negate the defined causal model (C1-EV_MECHANISM_0006, C1-EV_MECHANISM_0007). SECOND also has evidence that activating EGFR mutations are causal and produce ligand-independent kinase signaling into MAPK and PI3K pathways (C2-EV_GENETICS_0001,

C2-EV_MECHANISM_0001), but much of the supporting material is catalogue-level synthesis rather than independent direct perturbation evidence. EGFR bypass and compensatory signaling further qualify the inhibition mechanism (C2-CAT-ONCO-BIO-004, C2-CAT-ONCO-BIO-013). The judgment concerns mechanistic coherence, not druggability; clinical tractability evidence was not used as a basis for the verdict.

Decisive evidence:

- C1-EV_GENETICS_0001
- C1-EV_GENETICS_0002
- C1-EV_MECHANISM_0003
- C1-EV_MECHANISM_0004
- C1-EV_PERTURBATION_0001
- C1-EV_TISSUE_0001
- C1-EV_TISSUE_0002
- C2-EV_GENETICS_0001
- C2-EV_MECHANISM_0001

Contradicting evidence:

- C1-EV_MECHANISM_0006
- C1-EV_MECHANISM_0007
- C1-EV_TISSUE_0004
- C1-EV_PERTURBATION_0005
- C2-CAT-ONCO-BIO-004
- C2-CAT-ONCO-BIO-013

Missing evidence: Independent, direct EGFR loss-of-function or pharmacologic inhibition studies in NSCLC models specifically representing the stated mucus-producing glandular-cell context, with matched human pathology or PDX evidence.

Q5 · Perturbational evidence

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) ALK · 2) MET

Outcome: Winner: MET (2) · Margin: clear

Reasoning: MET has stronger reproducible PDX perturbational evidence: selective MET inhibition produced antitumor responses across six MET-amplified patient-derived xenograft models, and an independent set of six EGFR-mutant, MET-amplified patient-derived models responded to MET inhibition alone in vitro and in vivo (C2-EV_PERTURBATION_0002, C2-EV_PERTURBATION_0003). Genetic MET knockdown supports on-target dependence, although it uses shRNA and is therefore weaker alone (C2-EV_PERTURBATION_0001). ALK has evidence of selective siRNA effects and a limited patient-derived model mention, but the dossier does not provide comparably detailed, reproducible PDX tumor-response evidence; its CRISPR evidence primarily concerns resistance mechanisms rather than reversal of disease-relevant phenotypes (C1-EV_PERTURBATION_0001, C1-EV_PERTURBATION_0002).

Decisive evidence:

- C2-EV_PERTURBATION_0002
- C2-EV_PERTURBATION_0003
- C2-EV_PERTURBATION_0001

Contradicting evidence:

- C2-EV_PERTURBATION_0004
- C2-EV_PERTURBATION_0005
- C1-EV_PERTURBATION_0004
- C1-EV_PERTURBATION_0005

Missing evidence: A directly comparable ALK PDX study with quantified tumor-response results and

orthogonal ALK perturbations, rather than the limited knockdown evidence described in C1-EV_PERTURBATION_0001.

Q6 · Translational model validity

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.90

Presented as: 1) KRAS · 2) EGFR

Outcome: Winner: EGFR (2) · Margin: clear

Reasoning: EGFR has direct evidence in mechanistically relevant EGFR-mutant lung adenocarcinoma GEMMs, including response to EGFR inhibition, and the dossier also identifies NSCLC xenograft, PDX, and other translational models. KRAS has general statements that NSCLC PDX and CDX studies exist, but the specifically described efficacy result is from a subcutaneous MiaPaCa-2 human xenograft rather than an NSCLC model. Thus EGFR provides stronger model-validity evidence for NSCLC; this judgment is based on model relevance rather than reported effect size.

Decisive evidence:

- `C2-CAT-ONCO-BIO-117`
- `C2-CAT-ONCO-BIO-117b`
- `C2-CAT-ONCO-BIO-010`

Contradicting evidence:

- `C1-CAT-ONCO-BIO-117b`
- `C1-CAT-ONCO-BIO-010`

Missing evidence: A directly comparable KRAS NSCLC efficacy dataset in validated lung GEMM, orthotopic, humanised, or patient-derived models, with explicit relevance to newly diagnosed disease and chemotherapy combination.

Q16 · Evidence robustness (meta-evidence)

decisive — decided, with evidence

Judge: gpt-5.6-luna (openai) · **Self-reported confidence:** 0.88

Presented as: 1) KRAS · 2) ALK

Outcome: Winner: ALK (2) · Margin: clear

Reasoning: SECOND is more robust on the stated meta-evidence criterion. The ALK dossier identifies multiple source families, including a discovery study, TCGA profiling, a large real-world prevalence cohort, tissue pathology cohorts, genome-wide perturbation screens, randomized clinical trials and regulatory sources. It also preserves adverse or contradictory findings: incomplete ALK dependence, variant-specific sensitivity, subclonality, nonproductive FISH-positive cases, bypass resistance and lack of unadjusted overall-survival benefit. This makes source dependence and failure modes more visible. KRAS has some independent human-genomics support from TCGA, prospective MSK-IMPACT and a large retrospective cohort, but much of the dossier consists of repeated catalogue claims with limited provenance, no explicit negative-result or replication-status framework, and few criterion-specific contradiction or publication-bias indicators. The conclusion is therefore based on metadata quality and explicit heterogeneity, not on the larger quantity of ALK evidence.

Decisive evidence:

- `C2-EV\_GENETICS\_0002`
- `C2-EV\_GENETICS\_0003`
- `C2-EV\_GENETICS\_0005`
- `C2-EV\_MECHANISM\_0006`
- `C2-EV\_MECHANISM\_0007`
- `C2-EV\_PERTURBATION\_0002`
- `C2-EV\_PERTURBATION\_0005`

- C2-EV_PERTURBATION_0006
- C2-EV_CLINICAL_0001
- C2-EV_CLINICAL_0002
- C2-EV_CLINICAL_0004
- C2-EV_CLINICAL_0006

Contradicting evidence:

- C2-EV_MECHANISM_0006
- C2-EV_MECHANISM_0007
- C2-EV_TISSUE_0004
- C2-EV_TISSUE_0005
- C2-EV_BIOMARKER_0004
- C2-EV_BIOMARKER_0006
- C2-EV_CLINICAL_0006
- C1-CAT-ONCO-BIO-012
- C1-CAT-ONCO-BIO-020

Missing evidence: An explicit sample-overlap and provenance matrix for all studies in both dossiers, especially whether the KRAS TCGA, MSK-IMPACT and retrospective NSCLC cohorts are independent.; A systematic assessment of unpublished negative results, selective reporting and publication bias for both targets.; Direct replication-status metadata for the key KRAS perturbation and in vivo PDX experiments, comparable to the explicitly heterogeneous and contradictory ALK evidence.

A further 121 judgements met the selection but were not printed due to the length limit.