Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A retrospective benchmark of an open-source oncology drug ranker on 2,505 MSK CHORD patients

Aashish Kharel

aashishkharel7@gmail.com

Independent Researcher https://orcid.org/0009-0001-1785-8228

Research Article

Keywords: precision oncology, molecular tumor board, clinical decision support, targeted therapy, real-world data, MSK CHORD, OncoKB, CIViC, open source, benchmark

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11584932/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: The author is the founder and sole owner of OncoPurpose LLC, which develops OpenOncology, the open-source software evaluated in this preprint. The software is released under the MIT licence. No external funding was received.

Abstract

Precision oncology drug selection currently depends on a molecular tumor board that most hospitals in the world do not have. OpenOncology is an open-source pipeline that takes a tumor's variant file and returns a ranked, cited shortlist of drugs. We measured its agreement with the real prescribing decisions of a large academic cancer centre, using the published MSK CHORD cohort (Memorial Sloan Kettering, Nature 2024; 24,950 patients with IMPACT panel data). Of 5,190 patients who received at least one targeted therapy, 2,505 had a tumor alteration whose FDA label names that drug. On this benchmarkable subset, OpenOncology issued a prediction for 2,466 patients (98.4%). Its top three contained a drug of the correct molecular class for 2,466 of 2,466 scored patients (100.0%) and the exact drug the patient actually received for 2,094 of 2,466 (84.9%). Counting the 39 patients with no prediction as misses, the top-three class rate is 98.4% and the exact rate is 83.6%. The top-one drug matched the drug actually received in 70.6% of scored patients. A head-to-head comparison against an OncoKB-only baseline shows the full pipeline covers 315 additional patients (+ 12.6 pp), lifts top-three exact agreement by + 9.9 pp and top-three class agreement by + 12.6 pp. The scorer, the fully filtered cohort, both ranker runs, and the fusion extract are published in the project repository.

1. Background

A molecular tumor board interprets a tumor's variant file and recommends a therapy. It is a room of specialists: a pathologist, a geneticist, an oncologist, a bioinformatician, and curated clinical evidence databases. The Lancet Oncology Commission estimates seven to eight million cancer patients a year worldwide need molecular testing, and low- and middle-income countries host only 2% of biomarker-driven trials [1]. Over half of US cancer patients are treated in community settings without a tumor board [2]. Open evidence databases such as OncoKB [3] and CIViC [4] now cover much of the ground that a tumor board would consult, but a hospital still needs people to apply them case by case.

OpenOncology is an open-source pipeline, released under the MIT licence, that joins variant calling, pathogenicity scoring, OncoKB and CIViC lookups, and drug ranking into one Docker-deployable workspace. The ranked drug list is a shortlist for an oncologist to act on, not a prescription. The code and this benchmark are at github.com/immortal71/openoncology.

An earlier internal benchmark reported 100% agreement with an oncologist on 20 cases. We later found that the test labels and the model's evidence shared a source, so the test was circular. We retracted the number, published a plain-language explanation in docs/ONCOLOGIST_CONCORDANCE_PLAIN_LANGUAGE.md, and added a circularity check to the repository. This preprint is the replacement benchmark: a retrospective concordance study on an independent cohort, with the scorer and ground-truth map both public.

2. Methods

2.1 Cohort

We used the MSK CHORD cohort [5], published in *Nature* 2024, which contains 24,950 patients profiled on the MSK IMPACT panel, with structured treatment timelines, somatic mutations, discrete copy-number alterations, and (via the cBioPortal structural-variant endpoint) gene fusions. We restricted to patients who received at least one targeted therapy drug, giving 5,190 patients, 56 actionable genes, and 423 patients with at least one fusion in an actionable gene (389 inside the targeted-therapy cohort). The cohort file as of 2026-10-27 is at validation_results/msk_chord_targeted_cohort.json.

2.2 Ground truth

Each targeted therapy a patient received is paired against a map (scripts/mskchord/targeted_drug_targets.py) that lists, per drug, the gene-and-alteration patterns whose FDA label indicates that drug. A patient is **benchmarkable** when at least one of their received drugs is paired with an alteration in their own tumor that the label indicates. The map is deliberately conservative: for example, HER2-directed drugs require an ERBB2 amplification; CDK4/6 inhibitors require a CDK4 or CDK6 alteration (receptor-driven HR+/HER2- selection falls outside the mutation-matching scope of this benchmark and is reported as a miss). Fusion-only drugs (alectinib, crizotinib, lorlatinib, selpercatinib, pralsetinib, larotrectinib, entrectinib, pemigatinib) are matched against the per-patient fusion list.

Of 5,190 patients, 2,505 were benchmarkable, 2,632 had no qualifying alteration for any received drug (chiefly breast-cancer patients on receptor-driven therapies), and 53 were marked unevaluable.

2.3 Ranker

For each benchmarkable patient, the gene-and-variant that justifies their received drug is fed to services.oncokb_evidence.get_all_drugs_for_variant_live, which first queries OncoKB's committed static table (built from the OncoKB oncology drug database and ordered by FDA level), then falls back to a nightly release of the CIViC accepted clinical evidence table (AcceptedClinicalEvidenceSummaries.tsv) filtered to level A/B/C sensitivity/response evidence for the actionable gene set. Fusion variants are resolved against both tables. The CIViC fallback is restricted to the drug family that OncoKB lists for the gene, so cross-pathway and chemotherapy entries CIViC also carries for actionable genes do not crowd the top three. Within an OncoKB level the ranker preserves the dump's own clinical order (osimertinib before erlotinib for EGFR; trastuzumab before pertuzumab for ERBB2; trametinib before vemurafenib for BRAF), which lines up with how these patients are prescribed in practice. The top three are returned as parent drug names (salt forms stripped).

2.4 Scoring

For each benchmarkable patient, we classify the ranker's output as (i) exact top-3 hit if any drug the patient received is in the top three; (ii) same-class top-3 hit if any received drug shares a molecular target class with any top-three drug (classes defined by FDA label indication: EGFR TKIs, BRAF/MEK inhibitors, HER2-directed agents, ALK/ROS1 TKIs, PARP inhibitors, CDK4/6 inhibitors, PI3K/AKT, mTOR,

KIT/PDGFRA TKIs, KRAS G12C, RET fusion inhibitors, MET inhibitors, NTRK fusion inhibitors, IDH1/2, hedgehog, BTK, BCR-ABL); (iii) miss otherwise. Patients the ranker could not score are reported separately. The scorer is scripts/mskchord/score_mskchord.py.

2.5 Baseline

To isolate the lift from the ranker over a hospital that performs an OncoKB lookup unaided, we re-ran the pipeline with the CIViC fallback disabled, keeping everything else identical. The baseline is reported alongside the full pipeline.

3. Results

3.1 Headline numbers

On 2,466 of 2,505 benchmarkable patients where the ranker issued a prediction (98.4% coverage):

Metric	n	Rate
Top-1 exact: first drug matched the drug received	1,740	70.6%
Top-1 same-class	2,454	99.5%
Top-3 exact: drug received was in top three	2,094	84.9%
Top-3 same-class	2,466	100.0%
True misses (no class match in top three)	0	0.0%

As a lower bound counting the 39 benchmarkable patients with no prediction as misses, top-3 exact is 83.6% and top-3 class is 98.4%.

3.2 Per cancer type

Cancer type	n predicted	Top-3 exact	Top-3 class
Non-Small Cell Lung Cancer	1,710	91.8%	100.0%
Breast Cancer	507	64.1%	100.0%
Colorectal Cancer	136	67.6%	100.0%
Prostate Cancer	69	100.0%	100.0%
Pancreatic Cancer	44	88.6%	100.0%

Breast cancer top-3 exact looks low beside its 100% class rate. The reason is specific: 180 of the 507 scored breast-cancer patients are ERBB2-amplified patients whose label-matched drug was a later-line HER2 agent (lapatinib, neratinib, tucatinib, or ado-trastuzumab emtansine), and these account for 170 of the 182 breast exact misses. OncoKB ranks trastuzumab, pertuzumab and trastuzumab deruxtecan first

for ERBB2 amplification, so our top three fits a first-line decision but not the drug given after progression. The ranker is not currently line-of-therapy aware. Colorectal shows a related time effect: 28 of the 44 colorectal exact misses are BRAF-mutant patients who received dabrafenib with trametinib, vemurafenib, or cobimetinib with vemurafenib, while the ranker returns encorafenib with binimetinib and cetuximab, the regimen on the current label. A further 5 are PDGFRA patients who received regorafenib, which is not selected by a PDGFRA alteration.

3.3 Head-to-head: ranker versus OncoKB alone

Metric (as % of 2,505 benchmarkable)	OncoKB-only	OpenOncology	Lift
Coverage (patients predicted)	85.9%	98.4%	+ 12.6 pp
Top-1 exact	63.4%	69.5%	+ 6.1 pp
Top-1 same-class	85.4%	98.0%	+ 12.5 pp
Top-3 exact	73.7%	83.6%	+ 9.9 pp
Top-3 same-class	85.9%	98.4%	+ 12.6 pp

The CIViC fallback, target-class filtering, structural-variant and copy-number alteration handling, and preserved OncoKB clinical order collectively add 315 benchmarkable patients to the “scored” set and lift every headline rate.

4. Discussion

4.1 What this benchmark does and does not measure

The ground truth is “a drug was prescribed, and its FDA label indicates it for an alteration this tumor carries.” This is label-level concordance. It is not a measurement of clinical outcomes, nor of agreement with an oncologist on cases where the label alone is ambiguous. It is also partially circular at the knowledge-base level: MSK’s tumor board consults OncoKB in its own practice, and the ranker queries OncoKB. A genuinely independent test requires a cohort treated at centres that do not consult OncoKB, or a prospective study against a tumor board blinded to the ranker’s output.

4.2 Limits

Receptor-driven selection (CDK4/6 inhibitors for HR+/HER2- breast cancer, hormone-pathway drugs in prostate cancer) is correctly reported as a miss by the ground-truth map, because the selecting signal is not a mutation. These should not be read as ranker errors. The MSK CHORD cohort is a US-based academic centre; variant distributions in South Asian or community hospital cohorts will differ, and generalisation is unverified. 39 benchmarkable patients still received no prediction, almost all with rare TSC1/TSC2 point mutations not in either knowledge base; those are precisely the cases where a tumor board earns its value. The ranker is not line-of-therapy aware, which is the chief reason the top-three

exact rate for ERBB2-amplified breast cancer is lower than its class rate; a progression patient receives the 4th-ranked HER2 drug, not the first.

4.3 Next steps

A prospective pilot of the ranker is planned beside a live tumor board at a Nepali cancer centre: a shadow-mode read against new cases for 90 days with public concordance reporting, followed by a 100-case comparison against past tumor-board decisions. An independent oncologist review of 20 cases from the current benchmark (10 hits, 5 misses, 5 edge cases) will accompany this preprint when it is complete. Multi-institutional expansion requires additional data agreements and is outside the scope of this preprint.

Declarations

Acknowledgements

MSK CHORD is a project of Memorial Sloan Kettering Cancer Center; the data was made public via cBioPortal under the terms stated at the project page. CIViC is maintained by the McDonnell Genome Institute at Washington University. OncoKB is maintained by Memorial Sloan Kettering. OpenOncology is independent work; no funding or institutional affiliation is being claimed. Dr. Jeremy Warner gave feedback on cancer indication and treatment-line checks in an earlier iteration of OpenOncology, and I am grateful for his time. The content of this preprint is my own and should not be read as an endorsement.

Data and code availability

Item	Path	Open
Cohort file	validation_results/msk_chord_targeted_cohort.json	yes
Fusion file	validation_results/msk_chord_fusions.json	yes
Ground-truth map	scripts/mskchord/targeted_drug_targets.py	yes
Ranker runner	scripts/mskchord/run_openoncology_on_mskchord.py	yes
Scorer	scripts/mskchord/score_mskchord.py	yes
Ranker results	validation_results/msk_chord_openoncology_run.json	yes
Baseline results	validation_results/msk_chord_oncokb_only_run.json	yes
CIViC evidence	validation_results/civic_actionable_evidence.json	yes
MSK CHORD cohort source	cBioPortal msk_chord_2024	public
CIViC source	civicdb.org/releases/main	public, CC0
OncoKB source	oncokb.org (static dump)	restricted-use licence

All computation runs on a single commodity machine (< 1 minute end to end for the ranker; <2 seconds for the scorer).

References

1. Lancet Oncology Commission on precision oncology (Sept 2026) for low- and middle-income countries, 2026. via Omnicuris
2. McKesson community oncology trends report (2026) (via HMP Global Learning Network)
3. Chakravarty D, Gao J, Phillips S et al (2017) OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol
4. Griffith M, Spies NC, Krysiak K et al (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet 49:170–174
5. Jee J, Fong C, Pichotta K et al (2024) Automated real-world data integration improves cancer outcome prediction. Nature 636(8043):728–736. 10.1038/s41586-024-08167-5 (MSK CHORD)