151P When LLMs Disagree: A Cross-Model Strategy for Scalable Biomarker Extraction from Pathology Reports

Daria B. Kokh, PhD [1]; Irina Surovtsova, PhD [1]; Philipp Morakis, MD[2]

[1] Clinical State Registry Baden-Württemberg GmbH, Baden-Württemberg Cancer Registry (BWCR), Stuttgart, Germany
[2] Quality Conferences Office at the Clinical State Registry Baden-Württemberg GmbH, Baden-Württemberg, Cancer Registry (BWCR), Stuttgart, Germany

ESMO AI & DIGITAL ONCOLOGY Annual Congress
BERLIN GERMANY 16-18 NOVEMBER 2026

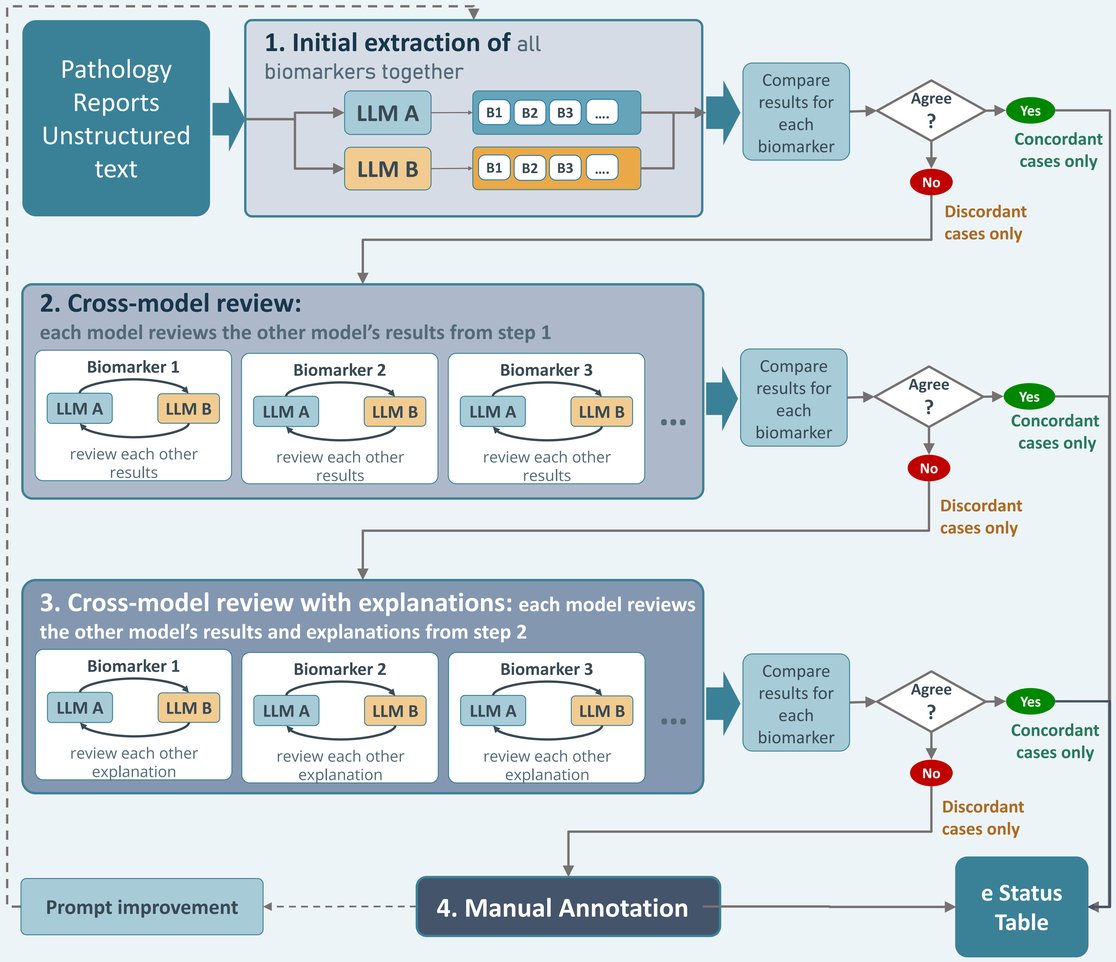
Background

Biomarker data documented in pathology reports are essential for real-world studies of cancer treatment effectiveness and survival outcomes. Large-scale studies require accurate extraction of these data into a structured format. Although large language models (LLMs) enable scalable annotation, extraction errors and the need for extensive manual validation remain major barriers. We therefore adapted, extended, and evaluated a disagreement-driven cross-model verification strategy to improve the reliability of biomarker extraction while limiting manual review to unresolved cases.

Data Source, Target Biomarkers, and LLMs

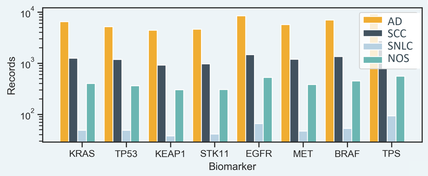
Data source: Baden-Württemberg Cancer Registry (BWCR)
Data: 44 910 pathology reports from 20 484 patients diagnosed with lung cancer between 2009 and 2025 (ICD-10: C34.*)
Models: Two locally deployed LLMs, Mistral-Small and Qwen3:14B
Extracted Information (3 categories):
(1) **Gene mutation status (mut/ WT):** KRAS; TP53; STK11; KEAP1; MET; BRAF, EGFR
(2) **Mutation type:** KRAS G12 variants;
(3) **Score:** PD-L1 tumor proportion score (TPS).

Biomarker Extraction Workflow Using Two LLMs

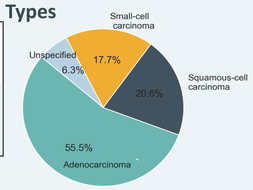


Results

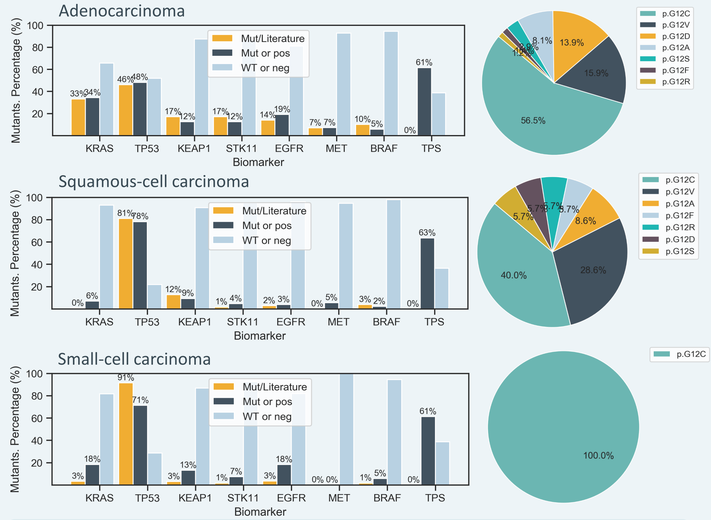
Biomarker recorded per histological type



Distribution of Histological Types



Distribution of biomarker status per histology

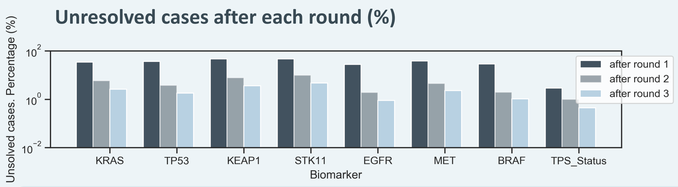


Biomarker	Biomarker status processed in each round				Validation (on 100-300 randomly selected records)
	1st round	2nd round	3rd round	Manual	Accuracy among extracted values
KRAS	16 047	5 549	947	417	94.9%
TP53	13 696	5 036	520	244	97.9%
STK11	12 407	5 956	1 258	600	96.0%
KEAP1	12 735	5 869	977	453	93.0%
MET	14 515	5 572	650	335	97.0%
BRAF	16 001	4 655	312	156	98.0%
EGFR	17 972	4 955	347	159	99.0%
KRAS (G12)	2 946	195	64	0	100%
PD-L1 (TPS)	19 095	553	190	84	99.3%

TIMING



Unresolved cases after each round (%)



Limitations

- Prompt design is critical and should be refined based on an analysis of extraction errors
- Extraction was most accurate for numerical values and mutation types.
- Wild-type status is the least accurate as it is not always directly specified in reports
- The workflow did not distinguish between comprehensive mutation testing and testing limited to specific variants.

Further Analysis

- The workflow does not currently classify variants as pathogenic or non-pathogenic. Where an exact variant is extracted, its pathogenicity can be assessed during post-processing

Conclusions

- The workflow supports scalable biomarker extraction, with manual review required for 1%–4% of records entering round 1.
- Extraction accuracy exceeds 93% for biomarkers with reported validation results.
- This approach effectively supports the structured use of pathology report data in cancer registries.

Krebsregister Baden-Württemberg
www.krebsregister-bw.de

Dr. Philipp Morakis
Head of the Quality Conferences Office
at the Clinical State Registry Baden-Württemberg GmbH,
Baden-Württemberg Cancer Registry (BWCR)

Birkenwaldstr. 149
70191 Stuttgart, Germany
Phone: +49 711 137909-101
Email: morakis@qualiko-bw.de

Disclosure Statement:

The authors report no conflicts of interest relevant to this work.

References

(biomarker status distribution):
1. Sci Transl Med. 2013, 5(209) DOI: 10.1126/scitranslmed.3006802
2. Nature. 2014; 511(7511) DOI: 10.1038/nature13385
3. Nature. 2012; 489(7417), 519 DOI: 10.1038/nature11404
4. J. of Clinical Onc. 2015, 33(15) DOI: 10.1200/jco.2015.33.15_suppl.11078
5. J. of Thoracic Dis. , 2020, 12 (7), 3776–3784 DOI: 10.21037/jtd.2019.08.61
6. BMC Res Notes 2015 8(688) DOI: 10.1186/s13104-015-1675-x

2026 ESMO AI & DIGITAL ONCOLOGY Annual Congress
BERLIN GERMANY 16-18 NOVEMBER 2026