This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Predicting Clinically Actionable LUAD Mutations from Whole-Slide Histopathology Using Prov-GigaPath Foundation Model Embeddings

Arian Baniassadi, Thomas Bezza, and Kevin Zhu(*)

Algoverse AI Research, San Francisco, CA, USA
abaniassadi77@gmail.com, thomasdb2017@gmail.com,
kevin@algoverseairesearch.org
*Corresponding author

Abstract. In lung adenocarcinoma (LUAD), identifying actionable driver mutations guides therapy, but next-generation sequencing (NGS) is costly, slow, and unevenly available. Predicting mutation status directly from routine H&E whole-slide images (WSIs) offers a rapid, low-cost complement, yet prior work has focused on common drivers and largely ignored rarer, highly actionable oncogenes. We present a data-efficient framework that pairs frozen Prov-GigaPath foundation-model embeddings with lightweight classical classifiers to predict TP53, PIK3CA, and NTRK3 status from TCGA-LUAD WSIs, and to isolate the role of macro-scale spatial context. Using the LongNet slide encoder, which aggregates tile embeddings with their spatial coordinates, we achieve pooled out-of-fold AUCs of 0.666 (TP53), 0.684 (PIK3CA), and 0.688 (NTRK3), to our knowledge, the first reported image-based benchmarks for PIK3CA and NTRK3 on TCGA-LUAD. When benchmarked against mean-pooling and multiple-instance learning (MIL) baselines, LongNet shows point-estimate gains on the rare oncogenes that are largest for NTRK3, where it has a statistically significant out-performance on both MIL baselines (DeLong $p \leq 0.011$ nominal; FDR-adjusted $p = 0.048$). The remaining rare-gene gains (PIK3CA and NTRK3 versus mean pooling) consistently favor LongNet but did not reach significance. Furthermore, LongNet's advantage in prediction scales inversely with mutation frequency, as for the common driver, TP53, spatial context adds essentially nothing (ΔAUC +0.01 over mean pooling) but yields larger gains for the rare oncogenes PIK3CA (+0.10) and NTRK3 (+0.11), where position-free baselines fall toward chance. These results suggest that foundation-model embeddings can make for a scalable molecular prescreening tool without needing end-to-end training.

Keywords: Computational pathology · Whole-slide images · Foundation models · Mutation prediction · Lung adenocarcinoma · Spatial context

1 Introduction

With an estimated 2.5 million new cases and 1.8 million deaths in 2022, lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related deaths worldwide [1]. Within lung cancer, LUAD has emerged as the predominant histologic subtype, accounting for around 46% of male lung cancer cases globally [2]. Treatment for LUAD has increasingly been influenced by the tumor's molecular characteristics. As with other cancers, therapies and immune checkpoint inhibitors are most effective against tumors harboring specific driver alterations, and survival outcomes depend on identifying these alterations at an early stage [3, 4].

We focus on three alterations with direct relevance to clinical decision-making in LUAD. TP53, the most frequently altered gene in LUAD, is an emerging prognostic biomarker associated with differential responses to immune checkpoint blockade and to EGFR tyrosine kinase inhibitors (TKIs) [5–7]. PIK3CA activates the PI3K/AKT/mTOR pathway and has been linked both to acquired EGFR-TKI resistance in EGFR-mutant LUAD and to PI3K-pathway inhibitor strategies in non-small cell lung cancer (NSCLC) [8, 9]. NTRK3 fusions, though rare in NSCLC (0.1–0.3%), offer a tissue-agnostic indication for the FDA-approved TRK inhibitors larotrectinib and entrectinib [10–12]. Detecting such driver alterations is therefore central to precision oncology in LUAD, even though a gene's actionability depends on its specific alteration.

While the current standard for mutation identification, next-generation sequencing (NGS), is accurate, it remains expensive and unevenly distributed amongst institutions [13]. Even when available, turnaround times frequently exceed the recommended 10 business days, and more than 20% of NSCLC patients begin first-line therapy before molecular results are returned [14, 15]. In contrast, hematoxylin and eosin (H&E) whole-slide images (WSIs) are produced for nearly every cancer patient and are increasingly digitized, motivating a growing body of work that predicts mutation status directly from H&E WSIs as a rapid, low-cost complement to NGS [16, 17].

We address this with a data-efficient framework that pairs Prov-GigaPath tile and slide encoders with classical classifiers to predict TP53, PIK3CA, and NTRK3 status from TCGA-LUAD diagnostic H&E WSIs, using binary labels derived from cBioPortal [22, 23]. To our knowledge, this establishes the first image-based prediction benchmarks for PIK3CA and NTRK3 on TCGA-LUAD, showing that foundation-model embeddings recover morphological signal for rare, clinically actionable mutations overlooked by prior histopathology work. By comparing LongNet slide embeddings against both mean-pooled ViT-Giant and attention-based MIL baselines, we isolate the contribution of spatial-context modeling, demonstrating that tissue-level organization is predictive specifically for rarer mutations, whereas common drivers such as TP53 are recoverable from coarse, tile-intrinsic morphology alone.

2 Related Works

Prior efforts to predict LUAD mutation status from histopathology have largely focused on common driver alterations. Coudray et al. demonstrated that deep convolutional neural networks (DCNNs) could classify NSCLC subtypes and predict several driver mutations, including EGFR and STK11, directly from TCGA-LUAD WSIs, with AUROC values of 0.73–0.86 [16]. Pan-cancer studies extend this study across dozens of tumor types and hundreds of genes, but performance on rarer targets remains modest and mixed [17, 18]. Of these studies, less common but clinically relevant alterations such as PIK3CA and NTRK3 have not been specifically benchmarked on the TCGA-LUAD cohort; prior work on LUAD focused on genes mutated in at least 10% of cases, ignoring PIK3CA and NTRK3, and pan-cancer investigations that may include these target genes do not specifically mention or discuss low-frequency gene-tumor pairs [17, 18]. Currently, most existing approaches rely on convolutional tile encoders trained from scratch or initialized from natural-image pretraining, with tile features aggregated through basic multiple-instance learning (MIL) frameworks such as CLAM [19].

Alternatively, pathology foundation models offer a route beyond previous constraints. Prov-GigaPath, pretrained on more than 1.3 billion 256×256 H&E and immunohistochemistry tiles from 171,189 WSIs across 28 cancer centers, achieves state-of-the-art performance on most digital pathology benchmarks [20], and its slide encoder, adapted from the LongNet dilated-attention architecture to gigapixel pathology, aggregates tile embeddings with their spatial coordinates into slide vectors that preserve WSI context [20, 21]. Unlike MIL approaches, the LongNet encoder provides holistic spatial representation. Through freezing the base layers and extracting embeddings, the encoder enables prediction with lightweight classifiers, a method that is tailored to the needs of rare, low-frequency mutations. However, whether these embeddings can recover signals for less common, clinically actionable LUAD mutations, and whether their spatial modeling confers an advantage over both pooling and MIL attention, remains unanswered.

3 Methods

Patient clinical data and genomic mutation profiles were obtained from the TCGA-LUAD dataset via cBioPortal. Binary labels were derived from the cBioPortal mutations molecular profile of the TCGA-LUAD PanCancer Atlas study. A patient was labeled positive for a gene if any sequenced sample carried at least one somatic mutation call in that gene, and negative otherwise. Labels were defined one-versus-rest per gene, so the negative class for one gene may include patients mutated in another. Additionally, for NTRK3, the clinically actionable event is gene fusion rather than the somatic point mutations captured here. Thus, we treat the NTRK3 mutation status as a proof-of-concept target for whether foundation-model embeddings can recover signal for rare alterations.

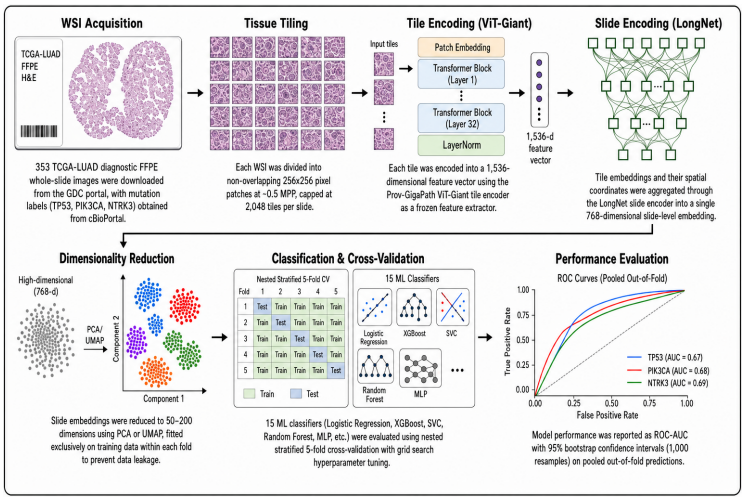

Fig. 1. An overview of our mutation-prediction pipeline.

Cohort sizes were based on data availability, yielding a TP53 cohort ($n = 353$; 214 mutant, 139 wild-type), an NTRK3 cohort ($n = 234$; 39 mutant, 195 wild-type), and a PIK3CA cohort ($n = 234$; 27 mutant, 207 wild-type). Each gene was modeled as an independent experiment with its own cross-validation, so no patient ever appears in the training split for one gene and the test split for another; within each gene, the outer folds are patient-disjoint. The PIK3CA and NTRK3 analyses use an identical 234-patient cohort that is a subset of the 353-patient TP53 cohort.

Diagnostic formalin-fixed paraffin-embedded (FFPE) whole-slide images (WSIs) were downloaded from the Genomic Data Commons portal and deduplicated to one slide per patient. Due to memory constraints, when processing untrimmed slides containing excessive background white space, 37 slides were excluded from the study. These exclusions and the 2,048-tile cap were applied uniformly and are independent of mutation status, so they are unlikely to bias the labels, though the cap may omit tissue on the largest slides. Each WSI was tiled into non-overlapping 256×256-pixel patches, capped at 2,048 tiles per slide. Tile embeddings (1,536-d) were extracted with the Prov-GigaPath ViT-Giant tile encoder and aggregated, together with their spatial coordinates, into a single 768-d slide embedding per patient using the Prov-GigaPath LongNet slide encoder (Fig. 1).

Slide embeddings were dimensionally reduced with PCA and UMAP at six target sizes (50, 64, 100, 128, 150, 200 components; 12 configurations in total), and fifteen models were trained to predict mutation status. Reducers were fit only on training data within each fold and applied out-of-sample to prevent leakage. Class imbalance was handled with balanced class weighting, and calibrated probabilities (3-fold CalibratedClassifierCV) were used for classifiers lacking probability estimates. Performance was evaluated by patient-level nested

stratified cross-validation: a 5-fold outer loop, with an inner 5-fold grid search tuning classifier hyperparameters on the reduced outer-training data before prediction on the held-out fold. Within each outer fold, the classifier and reduction were selected by inner-CV AUC on the training partition, and all results are reported as pooled out-of-fold ROC-AUC with 95% bootstrap confidence intervals (1,000 resamples).

To benchmark the LongNet slide-encoder results against established approaches, we trained two trainable attention-MIL models (AbMIL [24] and CLAM [19]) on the same 1,536-dimensional ViT-Giant tile features and matched patient cohorts, so the comparison isolates the aggregation mechanism rather than the underlying tile features. Additionally, we compare against a position-free mean-pooling baseline that shares LongNet's downstream pipeline. Per-patient out-of-fold predictions from the LongNet, MIL, and mean-pool pipelines were paired on shared patients within each gene, and AUCs were compared with DeLong's non-parametric test for correlated ROC curves. Since multiple tests were performed, we report Benjamini–Hochberg false-discovery-rate–adjusted p-values alongside nominal values.

4 Results

The mean-pooled ViT-Giant baseline, which discards tile positions, performed comparably to LongNet for the common mutation TP53, achieving an AUC of 0.655 (0.598–0.715) versus LongNet's 0.666 (0.607–0.721), a difference (ΔAUC +0.011) that was not statistically significant ($p = 0.711$). For the rarer oncogenes, however, LongNet consistently outperformed mean pooling. In the case of NTRK3, the mean-pooled baseline dropped to near chance at 0.584 (0.475–0.688), well below LongNet's 0.688 (0.584–0.779) (ΔAUC +0.114). Similarly, for PIK3CA, mean pooling achieved only 0.588 (0.457–0.706) compared with LongNet's 0.684 (0.560–0.795) (ΔAUC +0.100). In both cases, the mean-pooled baseline performed only marginally above chance, whereas LongNet retained clear discriminative ability. Although these rare-gene differences did not reach statistical significance ($p = 0.137$ for NTRK3 and $p = 0.307$ for PIK3CA), the magnitude and consistency of the gains indicate that the spatial context preserved by LongNet is particularly valuable for the subtler, spatially distributed phenotypes of rare mutations (Fig. 2; Table 1).

Attention MIL baselines for TP53 performed at or slightly better than LongNet, with AbMIL achieving an AUC of 0.711 (0.655–0.770) and CLAM achieving 0.714 (0.659–0.773), outperforming LongNet's AUC of 0.666 (0.607–0.721). However, this was not a statistically significant result ($p = 0.101$ and $p = 0.097$, respectively). For the rarer oncogenes, LongNet consistently outperformed the MIL baselines. In the case of NTRK3, both attention MILs dropped to near chance at 0.481 (0.379–0.585) and 0.512 (0.414–0.606) for AbMIL and CLAM, respectively. These results were statistically significantly worse than LongNet's AUC of 0.688 (0.584–0.779): nominal DeLong $p = 0.011$ (AbMIL) and 0.011 (CLAM), both of which remain significant after Benjamini–Hochberg correction

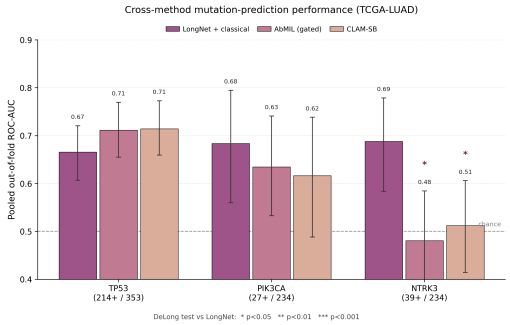

Fig. 2. Cross-method mutation-prediction performance on TCGA-LUAD. Bars show pooled out-of-fold ROC-AUC for LongNet and two attention-MIL baselines (AbMIL, CLAM), all trained on identical 1,536-d ViT-Giant tile features. Error bars are 95% bootstrap confidence intervals (1,000 resamples). The dashed line represents chance (AUC = 0.5). Asterisks denote DeLong-test significance versus LongNet (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Per-gene mutant/total counts are given on the x-axis.

(adjusted $p = 0.048$). In the case of PIK3CA, the attention MIL baselines showed AUCs of 0.635 (0.533–0.741) and 0.616 (0.488–0.738) for AbMIL and CLAM, respectively, both lower than LongNet's 0.684 (0.560–0.795), though not statistically significant ($p = 0.567$ and $p = 0.485$). Beyond ROC-AUC, on the rare oncogenes LongNet achieved the highest PR-AUC (PIK3CA 0.26, NTRK3 0.33; against prevalence baselines of 0.12 and 0.17) and the best calibration (expected calibration error 0.009 and 0.047 vs. 0.10–0.20 for the MIL models), though sensitivity at 95% specificity remained limited for all methods (LongNet 0.11 and 0.13).

5 Discussion & Limitations

The improvement in prediction via Prov-GigaPath's LongNet slide encoder, especially as it pertains to the rarer mutations, NTRK3 and PIK3CA, shows how spatial relationships between tissue regions do carry predictive signals for mutation status, and LongNet does indeed capture these relationships. This is further supported by the fact that the largest gains for LongNet over both baselines fall on the two rare oncogenes, rather than on the common driver TP53, as seen in Fig. 3. Since mutations do not manifest in single cells but rather in tissue-level patterns, a tumor not only changes what an individual cell looks like but also its spatial organization, which LongNet can capture. However, simple mean pooling results in the loss of crucial slide-level spatial context and destroys spatial relationships due to its averaging of tile vectors into one.

For TP53, this has little to no effect, as TP53 mutations cause widespread, cell-intrinsic morphological changes visible in individual tiles regardless of spatial context. However, for rarer oncogenes such as PIK3CA and NTRK3, which produce subtler, more spatially distributed changes such as altered growth patterns,

Table 1. Comparison of the LongNet pipeline with attention-MIL and mean-pooled ViT-Giant baselines on TCGA-LUAD. ΔAUC and p are from DeLong's test versus LongNet (positive ΔAUC favors LongNet); significant p in **bold**. PR-AUC should be read against the prevalence chance baseline (TP53 0.61, PIK3CA 0.12, NTRK3 0.17); Sens@95 is the sensitivity at 95% specificity; ECE is the expected calibration error (10 bins, lower is better).

Gene	Method	AUC (95% CI)	ΔAUC	p	PR-AUC	Sens@95 / ECE
TP53	LongNet	0.666 (0.607–0.721)	–	–	0.742	0.140 / 0.098
	AbMIL	0.711 (0.655–0.770)	−0.051	0.101	0.790	0.192 / 0.248
	CLAM	0.714 (0.659–0.773)	−0.054	0.097	0.788	0.145 / 0.252
	Mean-pool ViT-Giant	0.655 (0.598–0.715)	+0.011	0.711	0.736	0.104 / 0.111
PIK3CA	LongNet	0.684 (0.560–0.795)	–	–	0.260	0.111 / 0.009
	AbMIL	0.635 (0.533–0.741)	+0.049	0.566	0.202	0.148 / 0.106
	CLAM	0.616 (0.488–0.738)	+0.067	0.485	0.192	0.074 / 0.123
	Mean-pool ViT-Giant	0.588 (0.457–0.706)	+0.100	0.307	0.172	0.077 / 0.189
NTRK3	LongNet	0.688 (0.584–0.779)	–	–	0.333	0.128 / 0.047
	AbMIL	0.481 (0.379–0.585)	+0.207	**0.011**	0.168	0.051 / 0.186
	CLAM	0.512 (0.414–0.606)	+0.176	**0.011**	0.185	0.051 / 0.181
	Mean-pool ViT-Giant	0.584 (0.475–0.688)	+0.114	0.137	0.205	0.027 / 0.203

Nominal DeLong p-values shown; NTRK3 LongNet-vs-MIL comparisons remain significant after Benjamini–Hochberg correction (adjusted $p = 0.048$).

the spatial relationships between tiles carry critical signals that mean pooling destroys. Attention-MIL models, such as AbMIL and CLAM, suffer from a similar drawback as they are permutation-invariant and therefore do not use positional information. LongNet, on the other hand, aggregates tile embeddings together with their spatial positions, which is what allows it to encode tissue-level organization (Fig. 4).

Finally, we would like to note that we attempted to benchmark our results with TransMIL, but were unable to train it reliably due to TransMIL's substantially larger data requirements relative to our limited cohort sizes and small number of positive cases. Additionally, with so few positive cases, the corresponding AUC estimates carry wide confidence intervals, and several of LongNet's gains over the baselines, though large and consistent, did not reach statistical significance. Furthermore, another key limitation of our study was that since NTRK3 fusions are too rare in TCGA-LUAD to model directly, we had to utilize NTRK3 mutation status as a proxy for rare alterations. Thus, we provide the opportunity for future work to build directly on our approach. By assembling larger, multi-institutional cohorts, future works can look to test if models can predict these fusion events rather than relying on mutation status as a proxy.

6 Conclusion

Utilizing the Prov-GigaPath pathology foundation model paired with classical machine learning classifiers, we not only demonstrate the potential for automated mutation prediction in lung adenocarcinoma (LUAD) without the need for end-to-end deep learning training or extensive computational resources, but also

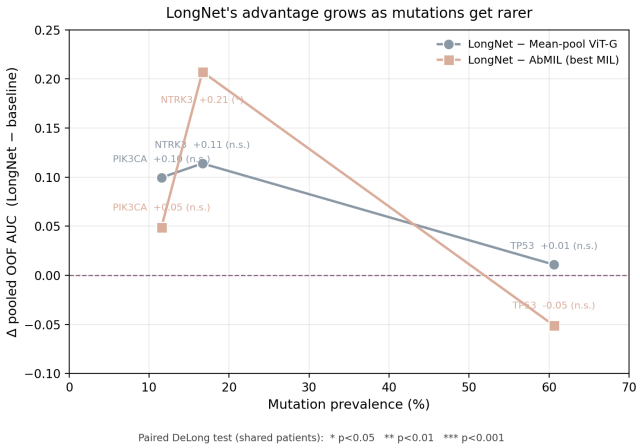

Fig. 3. LongNet's advantage over the baselines is concentrated among the rarer genes. Each point shows the difference in pooled out-of-fold ROC-AUC between LongNet and a baseline (mean-pooled ViT-Giant, blue; best MIL model, AbMIL) versus per-gene mutant prevalence. With only three genes, this is illustrative rather than a quantitative trend, and the ordering is not strictly monotone, but the largest gains clearly fall on the two rare oncogenes.

the ability to scale toward more complex genomic signals. Our study achieved AUCs of 0.666–0.688 by using the LongNet slide encoder, and when benchmarked against established attention-based multiple-instance-learning (MIL) models (AbMIL and CLAM) and mean-pooled ViT-Giant baselines, our methods show consistent point-estimate gains that reach statistical significance against MIL baselines for NTRK3, demonstrating that WSI context preservation and a lightweight screening framework can be particularly useful when working with rarer gene mutations that have posed challenges for previous studies.

Prospects of Application: Taken together, these results establish a data-efficient, low-cost pre-screening tool in digital pathology workflows, which can flag lung adenocarcinoma patients likely to harbor actionable mutations (e.g., NTRK3, PIK3CA) directly from routine H&E slides. As a result, our methods could speed up access to targeted therapies or extend genomic insight to settings lacking routine sequencing.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Ervik, M., Soerjomataram, I., Jemal, A., Torre, L.A.: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **74**(3), 229–263 (2024). https://doi.org/10.3322/caac.21834

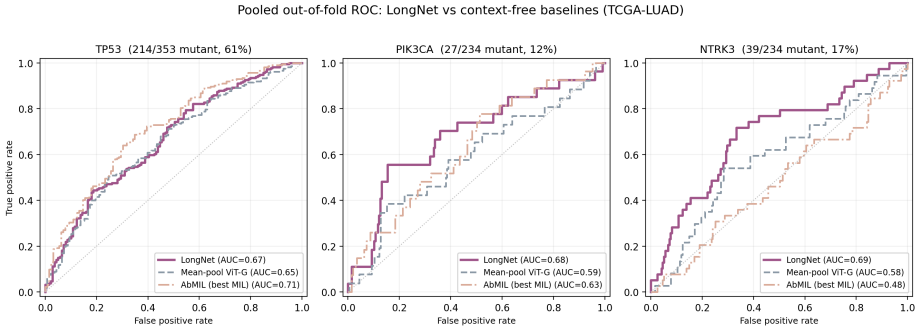

Fig. 4. Pooled out-of-fold ROC curves for LongNet versus position-free baselines on TCGA-LUAD. Each panel shows one gene, with curves for LongNet (solid), mean-pooled ViT-Giant (dashed), and the best attention-MIL model, AbMIL (dash-dot). The three methods overlap for the common driver TP53, but for the rare oncogenes PIK3CA and NTRK3, the baselines sag toward or below the diagonal, while LongNet retains clear discrimination.

2. Luo, G., Zhang, Y., Rumgay, H., et al.: Estimated worldwide variation and trends in incidence of lung cancer by histological subtype in 2022 and over time: a population-based study. The Lancet Respiratory Medicine (Published online 2025). https://doi.org/10.1016/S2213-2600(24)00428-4
3. Cancer Genome Atlas Research Network: Comprehensive molecular profiling of lung adenocarcinoma. Nature **511**(7511), 543–550 (2014). https://doi.org/10.1038/nature13385
4. Ettinger, D.S., Wood, D.E., Aisner, D.L., et al.: Non-Small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network **20**(5), 497–530 (2022). https://doi.org/10.6004/jnccn.2022.0025
5. Gibbons, D.L., Byers, L.A., Kurie, J.M.: Smoking, p53 mutation, and lung cancer. Molecular Cancer Research **12**(1), 3–13 (2014). https://doi.org/10.1158/1541-7786.MCR-13-0539
6. Dong, Z.Y., Zhong, W.Z., Zhang, X.C., et al.: Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. Clinical Cancer Research **23**(12), 3012–3024 (2017). https://doi.org/10.1158/1078-0432.CCR-16-2554
7. Canale, M., Petracci, E., Delmonte, A., et al.: Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. Clinical Cancer Research **23**(9), 2195–2202 (2017). https://doi.org/10.1158/1078-0432.CCR-16-1781
8. Fang, Y., Wang, Y., Zeng, D., et al.: PIK3CA mutation as an acquired resistance driver to EGFR-TKIs in non-small cell lung cancer: clinical challenges and opportunities. Pharmacological Research **201**, 107087 (2024). https://doi.org/10.1016/j.phrs.2024.107087
9. Tan, A.C.: Targeting the PI3K/Akt/mTOR pathway in non-small cell lung cancer (NSCLC). Thoracic Cancer **11**(3), 511–518 (2020). https://doi.org/10.1111/1759-7714.13328

10. Cocco, E., Scaltriti, M., Drilon, A.: NTRK fusion-positive cancers and TRK inhibitor therapy. Nature Reviews Clinical Oncology **15**(12), 731–747 (2018). https://doi.org/10.1038/s41571-018-0113-0
11. Drilon, A., Laetsch, T.W., Kummar, S., et al.: Efficacy of larotrectinib in TRK fusion–positive cancers in adults and children. New England Journal of Medicine **378**(8), 731–739 (2018). https://doi.org/10.1056/NEJMoa1714448
12. Doebele, R.C., Drilon, A., Paz-Ares, L., et al.: Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials. The Lancet Oncology **21**(2), 271–282 (2020). https://doi.org/10.1016/S1470-2045(19)30691-6
13. Pennell, N.A., Mutebi, A., Zhou, Z.Y., et al.: Economic impact of next-generation sequencing versus single-gene testing to detect genomic alterations in metastatic non–small-cell lung cancer using a decision analytic model. JCO Precision Oncology **3**, 1–9 (2019). https://doi.org/10.1200/PO.18.00356
14. Lim, C., Tsao, M.S., Le, L.W., et al.: Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. Annals of Oncology **26**(7), 1415–1421 (2015). https://doi.org/10.1093/annonc/mdv208
15. Sadik, H., Pritchard, D., Keeling, D.M., et al.: Impact of clinical practice gaps on the implementation of personalized medicine in advanced non–small-cell lung cancer. JCO Precision Oncology **6**, e2200246 (2022). https://doi.org/10.1200/PO.22.00246
16. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., et al.: Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Medicine **24**(10), 1559–1567 (2018). https://doi.org/10.1038/s41591-018-0177-5
17. Fu, Y., Jung, A.W., Torne, R.V., et al.: Pan-cancer computational histopathology reveals mutations, tumor composition, and prognosis. Nature Cancer **1**(8), 800–810 (2020). https://doi.org/10.1038/s43018-020-0085-8
18. Kather, J.N., Heij, L.R., Grabsch, H.I., et al.: Pan-cancer image-based detection of clinically actionable genetic alterations. Nature Cancer **1**(8), 789–799 (2020). https://doi.org/10.1038/s43018-020-0087-6
19. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021). https://doi.org/10.1038/s41551-020-00682-w
20. Xu, H., Usuyama, N., Bagga, J., et al.: A whole-slide foundation model for digital pathology from real-world data. Nature **630**(8015), 181–188 (2024). https://doi.org/10.1038/s41586-024-07441-w
21. Ding, J., Ma, S., Dong, L., et al.: LongNet: scaling transformers to 1,000,000,000 tokens. arXiv preprint arXiv:2307.02486 (2023)
22. Cerami, E., Gao, J., Dogrusoz, U., et al.: The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery **2**(5), 401–404 (2012). https://doi.org/10.1158/2159-8290.CD-12-0095
23. Gao, J., Aksoy, B.A., Dogrusoz, U., et al.: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling **6**(269), pl1 (2013). https://doi.org/10.1126/scisignal.2004088
24. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136. PMLR (2018)