Background The development of immune checkpoint inhibitor (ICI) therapy has fundamentally changed the landscape of cancer treatment. While ICIs have exhibited remarkable efficacy across diverse cancer types, the majority of cancer patients do not respond to these therapies.1 Tools to better identify patients who would benefit from ICI therapy are urgently needed to facilitate personalized care. Models for ICI response that incorporate tumor microenvironment (TME) features in addition to molecular data have demonstrated improved predictive power of patient response to therapy.2 3 These features reflect the coordinated activity of multiple cell types and therefore, are best captured by mRNA expression. Transcriptomic profiles are not however readily assayed in clinical settings. Extracting TME features from molecular data already collected in clinical settings provides an opportunity to bridge the gap between predictive models that rely on these features and their translation into clinical practice.

Methods We developed an ML model to reconstruct tumor gene expression profiles using genetic information from clinically available commercial NGS panels and embeddings4 generated by a language model (figure 1). This model was trained on publicly available data including ~8000 tumors representing 32 cancer types5 and validated in additional heterogeneous cohorts.

Results Gene expression reconstruction using this model was highly correlated with true expression (mean correlation per sample = 0.88, [0.8818 - 0.8858, 95% CI, N=847]). We applied these data to the prediction of a set of TME signatures, previously associated with response to ICI therapy7 and which describe TME composition and phenotype (mean correlation per sample = 0.81, [0.8008, 0.8170, 95% CI, N=756]). We demonstrate how reconstructed TME signatures were predictive of survival and provide interpretable biological insight into differences in patient outcomes across these cohorts.

Conclusions Our flexible analytic framework for reconstructing gene expression profiles from clinicoanalytics data allows for integration of additional features and enables prediction of cancer type- and subtype-specific features across diverse patient cohorts. This approach also has the potential to expand the number of patients who may otherwise have been overlooked for these therapies, ultimately providing more precise and effective individualized treatment options with the potential for improved outcomes for more cancer patients.

Acknowledgements The authors would like to acknowledge members of the Zephyr AI science, engineering and data teams.

REFERENCES

5. The Cancer Genome Atlas Program (TCGA) [https://www.cancer.gov/tcga]

http://dx.doi.org/10.1136/jitc-2023-SITC2023.1296