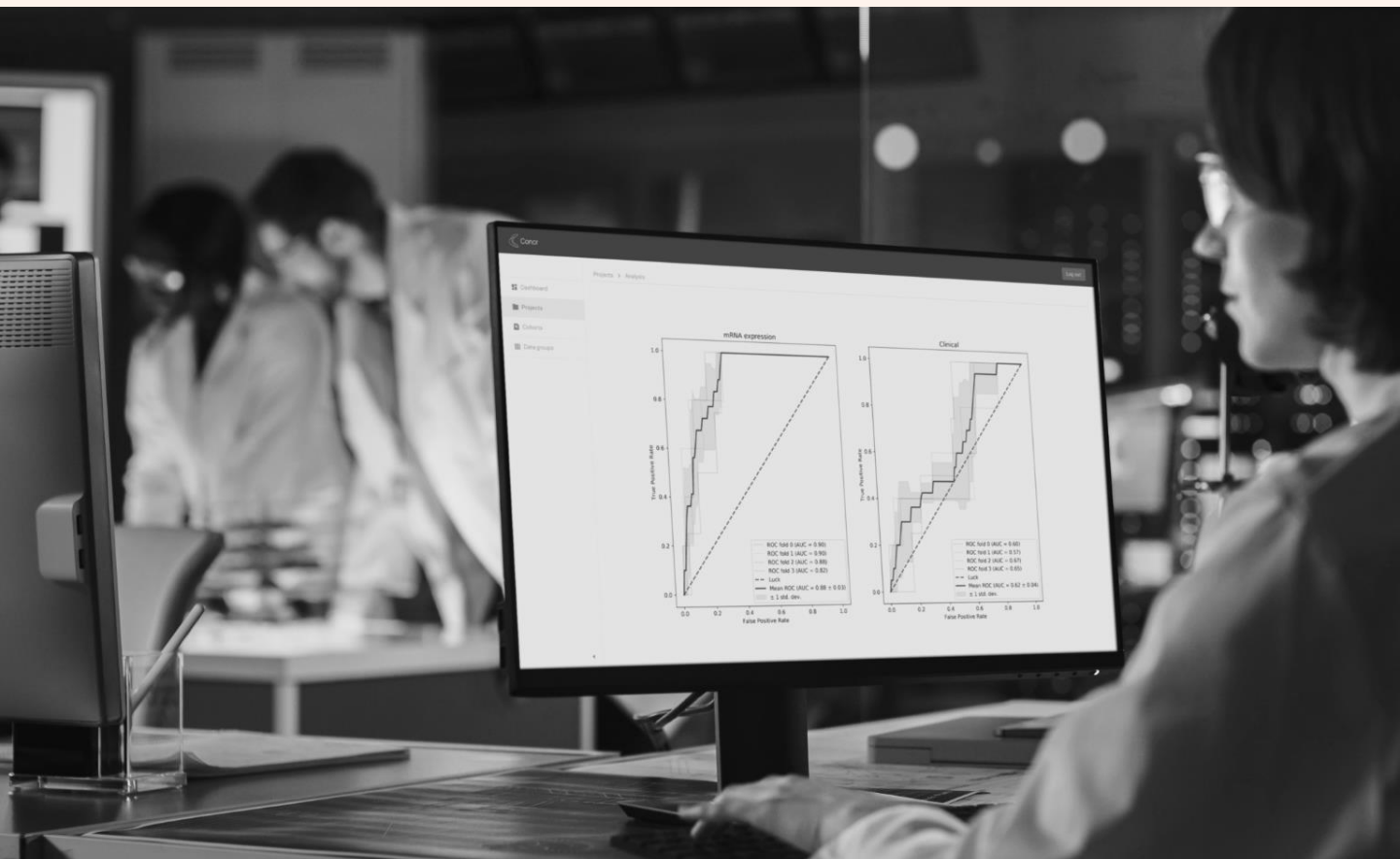


Perspective

Going Further Together

How Other Cancers Can Enable Robust Patient Outcome Predictions For All Cancers

By Matt Foster, Matthew Griffiths and Irina Babina



Summary

Cancer research often overlooks less common cancers, despite these accounting for nearly 50% of patients and 55% of cancer-related deaths. Their diversity and smaller datasets make predictions challenging, but this variability can strengthen predictive modelling.

Predictive models thrive on diverse data. While models trained on homogeneous datasets quickly reach performance ceilings, incorporating data from the wide biological variability of these lesser-studied cancers significantly improves their robustness and accuracy.

By shifting focus to 'biological units' – networks of traits associated with therapeutic response – predictive models can move beyond tissue-specific classifications. This enables the identification of universal patterns that apply across cancer types, improving outcomes even for rare cancers with limited representation in training data. We demonstrate this approach in this report using our predictive analytics platform FarrSight®.

Leveraging biological information about "other cancers" unlocks untapped potential to advance predictive oncology. By harnessing their diversity, we can develop more adaptable and inclusive models, ensuring better treatment strategies and outcomes for all cancer patients, particularly where they are needed most.

Cancer research has traditionally focused on a few common types like lung, breast, prostate, and colorectal cancers. Due to their high frequency, these cancers dominate clinical trials and data collection efforts, driving the development of targeted therapies and precision treatments. The result is a research and funding landscape where resources and efforts revolve around these few types. This focus, however, creates a critical imbalance, leaving behind a large and diverse group frequently referred to collectively as ‘other’ cancers. Despite their name, these cancers account for nearly 50% of cancer patients and 55% of cancer-related deaths in the United States¹. This disparity highlights the untapped potential within these overlooked cancers - a potential that, if harnessed correctly, could transform oncology care. In this feature, we explore this topic using our own predictive engine, FarrSight^{®2}.

The economic divide: common vs. rare cancer

The economic focus on common cancers is clear when considering the global cancer drug market. In 2024, the oncology market was valued at approximately \$223 billion, with an expected compound annual growth rate (CAGR) of 12.8% through 2028³. Yet over 70% of this market is dominated by treatments for common cancers like lung and breast cancer. In contrast, drugs targeting rare or less common cancers collectively represent a much smaller slice of the market, receiving only 3% of the global funding for cancer research⁴. Concerningly, spending on novel cancer therapeutic development or their use in clinics does not, across all cancer types, correlate with improved patient outcomes⁵. This economic reality underscores the need for a shift in research focus - one that can drive better outcomes and more equitable investment in these under-served areas.

¹‘Cancer Facts & Figures 2024’, American Cancer Society (ACS), Atlanta, Georgia, 2024.

²Griffiths et al., ‘Computational prediction of therapeutic response and cancer outcomes’, medRxiv, 2024

³‘Global Oncology Trends 2024: Outlook to 2028’, IQVIA Institute, 2024

⁴McIntosh et al., ‘Global funding for cancer research between 2016 and 2020: a content analysis of public and philanthropic investments’, *The Lancet Oncology*, vol. 24, no. 6, pp. 636–645, 2023.

⁵Chow, Bradley & Gross, ‘Comparison of Cancer-Related Spending and Mortality Rates in the US vs 21 High-Income Countries’, *JAMA Health Forum*, vol. 3, no. 5, p. e221229, 2022

Units of biology: leveraging the hidden strength of data variability

Predictive models in oncology, like other fields, benefit immensely from data diversity. Models built on homogenous datasets – such as those limited to a specific cancer or data type – often encounter a ceiling in accuracy due to a lack of variability. This is where the data from 'other' cancers can make a difference. Though these cancers are less common individually, collectively, they provide a wealth of diverse biological data that can train more versatile models. So, we put it to the test using our own Bayesian proprietary model of human cancer biology, **FarrSight®**.

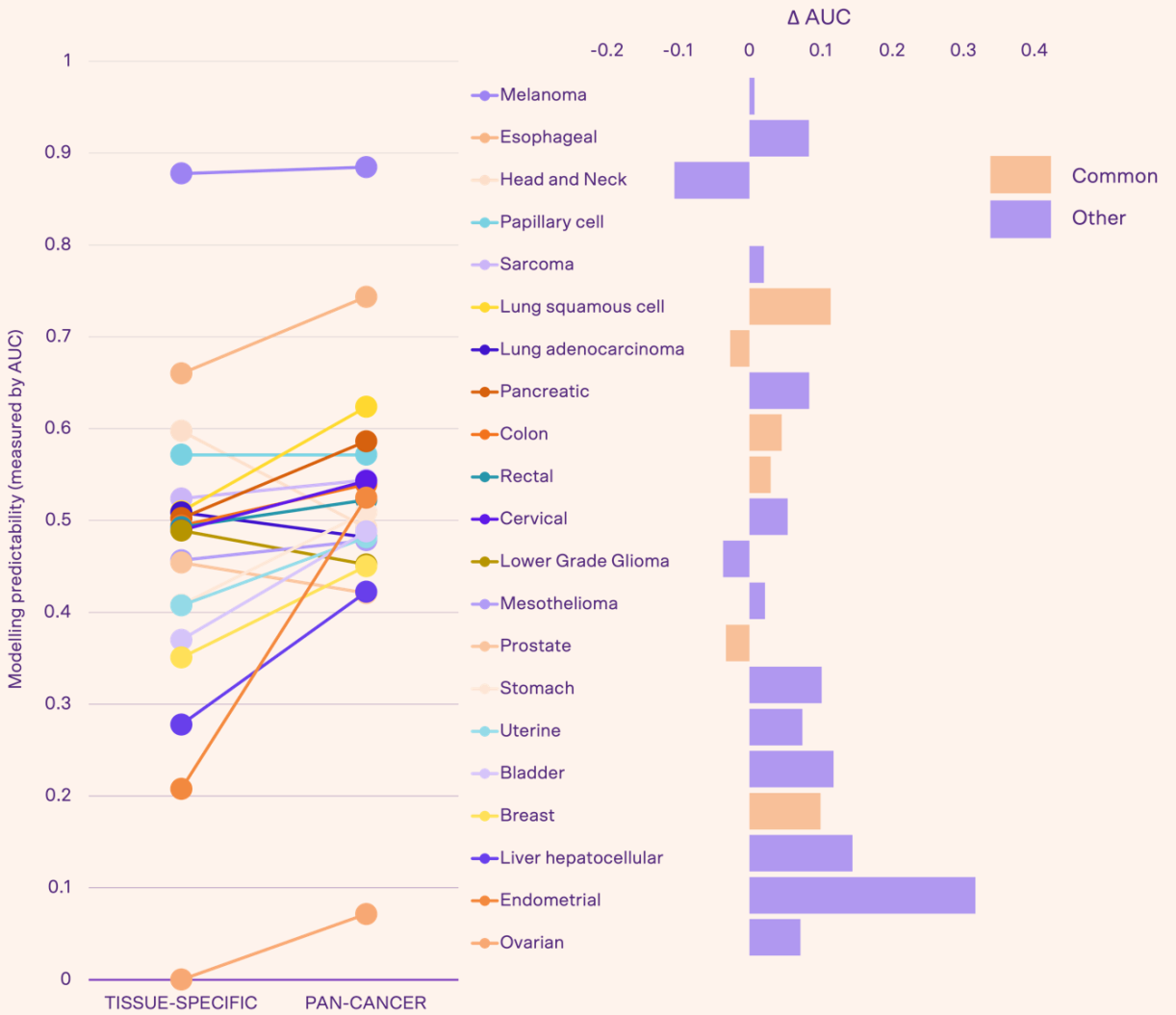
Models built on homogenous datasets often encounter a ceiling in accuracy due to a lack of variability.

We found that when **FarrSight®** was trained using data from a broader set of cancers, including rare ones, the accuracy of treatment outcome predictions increased by up to 153% compared to models trained solely on common cancer datasets (Exhibit 1). Therefore, by integrating data from these varied cancer types, predictive models gain access to a wider range of biological behaviours and generalise across cancer types even for cancers that are not directly represented in the training data. This adaptability is crucial in oncology, where biological mechanisms often overlap across different cancers.

To overcome the limitations of traditional, tissue-specific cancer classifications, **FarrSight®** identifies 'biological units' - clusters of traits and multi-modal molecular signatures that are shared across different cancers. These units go beyond the conventional approach of classifying cancers by their tissue of origin (e.g. breast or lung) and instead focus on the underlying mechanisms driving the cancer. By identifying these shared features, we can create models that predict outcomes for rare cancer types even when those specific types were not part of the original training set. This makes the data from 'other' cancers a valuable resource for enriching predictive models, turning what was once considered a limitation – high variability and small cohort sizes – into a powerful advantage.

Exhibit 1

FarrSight® performance, measured by Area Under the Curve (AUC), when trained on data from tissue-specific cancer, and then using the pan-cancer approach, using data across all cancers, including “others”.



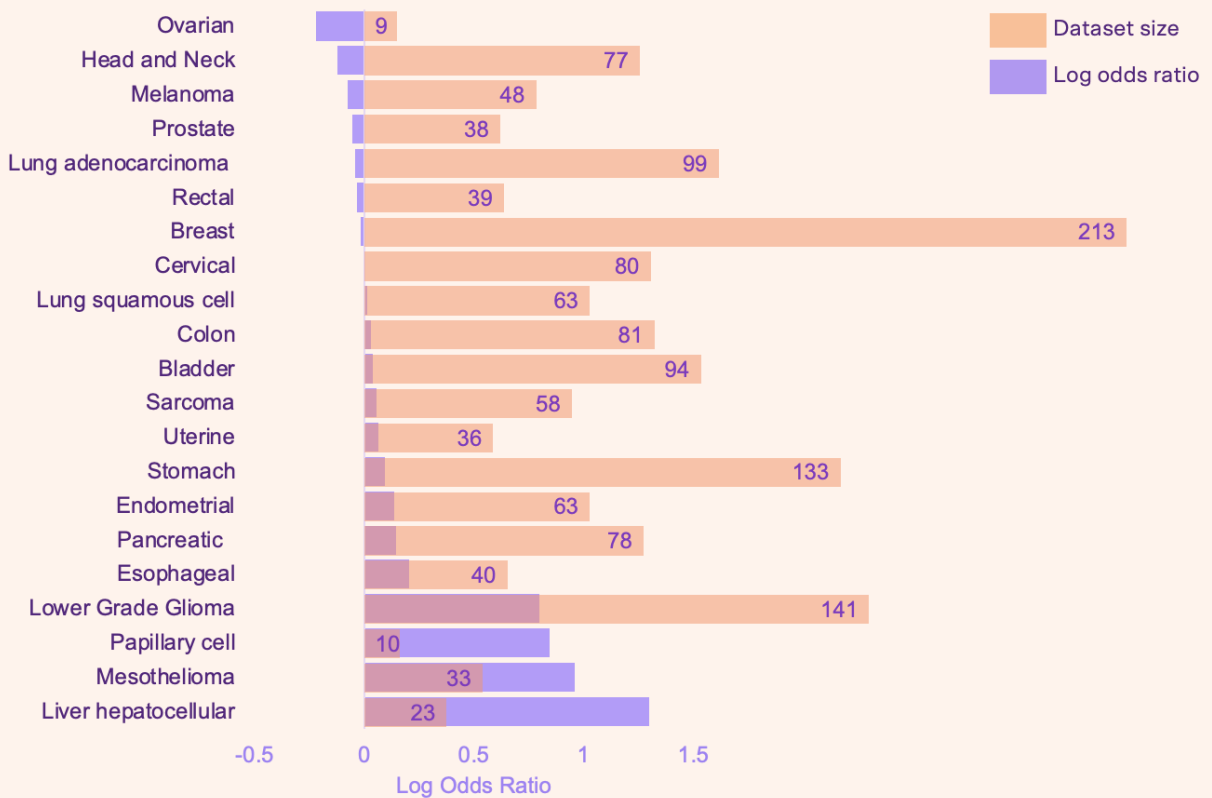
Less can be more with data

A common assumption in predictive modelling is that larger datasets yield better results. However, our analysis of pan-cancer models suggests that diversity outperforms volume, concordant with previous research⁶.

Our analysis of patient outcomes using FarrSight® trained on various cancer types revealed no significant correlation between the size of the training dataset and the improvement in outcome predictions (Exhibit 2). Instead, datasets representing a broader range of biological behaviours, like those from 'other' cancers, showed greater predictive power.

Exhibit 2

Comparison between cohort size of the training data for each cancer type (orange bar) is not correlated to the improvements in patient outcomes (purple bar).



⁶Brown et al., 'Language Models are Few-Shot Learners', NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020

Economic and clinical imperatives for change

The economic imbalance between treatments for common and rare cancers highlights the need for inclusive predictive models in oncology. While drugs targeting common cancers continue to dominate the market with revenues exceeding \$10 billion annually, significant unmet needs remain for patients with rarer cancers. Expanding predictive modelling to include diverse data from these lesser-studied cancers can drive new drug development, more tailored clinical trials, and a shift toward truly personalised medicine.

Overall, incorporating data from "other" cancers into predictive models could help reduce the overall societal burden of cancer, with the potential to improve survival rates and lower treatment costs, especially since new therapies can be prohibitively expensive. By leveraging the variability in rare cancers, models become more robust, offering better predictions across all types. This approach is a scientific imperative but a strategic advantage that can drive equitable investment, improved patient outcomes, and the vision of making personalised oncology a reality for all.

Matt Foster and Matthew Griffiths are co-founders on Concr Ltd and led the development of FarrSight®; Irina Babina serves as the CEO of Concr Ltd. The authors wish to acknowledge the rest of the Concr team for their contribution to the development of FarrSight® and the methodology behind it.



Solving data
to enable precision therapy
for all

Contact: info@concr.co

