Harnessing biomedical foundation models for genomic feature engineering to investigate patient drug response

Laura-Jayne Gardiner*
IBM Research Europe
The Hartree Centre, Daresbury, UK
laura-jayne.gardiner@ibm.com

Jennifer Kelly*
IBM Research Europe
The Hartree Centre, Daresbury, UK
Jennifer.Kelly@ibm.com

Stephen Checkley
STFC, The Hartree Centre
Sci-Tech Daresbury, UK
stephen.checkley@stfc.ac.uk

Karen Bingham
REPROCELL Europe Ltd
Glasgow, Scotland, UK
karen.bingham@reprocell.com

Graeme Macluskie
REPROCELL Europe Ltd
Glasgow, Scotland, UK
graeme.macluskie@reprocell.com

David Bunton
REPROCELL Europe Ltd
Glasgow, Scotland, UK
david.bunton@reprocell.com

Abstract

Utilising pre-trained biomedical foundation models (BMFMs) for inference on multi-omic data from small cohorts, represents a promising and practical route to demonstrate advantage of this technology for real-world drug discovery tasks. Here, we show this via an innovative and unique BMFM inference workflow, where BMFMs provide discernible advantage for predicting patient drug response from omics data. We utilise open-source, fine-tuned, multi-omics BMFMs for inference to enable downstream feature selection and engineering. Firstly, predicting drug-target binding affinity (BA), enabling ranking and prioritisation of gene targets and associated SNPs, and secondly, using patient SNPs to mutate reference proteins and assess their impact on prednisolone BA. BMFM-derived features were composed and used alongside non-BMFM features to predict patient-specific prednisolone response using an explainable ML approach. We demonstrate superior predictive power of BMFM-derived feature sets, and downstream explainability distinguished SNPs that were most influential for personalised drug response prediction.

1 Introduction

By modeling and capturing the interactions between multi-omics data types, biomedical foundation models (BMFMs) offer a powerful approach to increase understanding of complex biological systems and disease mechanisms. As such, they have the potential to improve and accelerate many drug discovery and development applications, enabling evaluation of drugs within the context of a more realistic biological system. However, the use of these technologies for real-world drug discovery applications is often hindered by compute availability, computational complexities and the requirement

*Authors contributed equally

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal FMs and LLMs for Life Sciences.

for interdisciplinary expertise to build and train models. Although fine-tuning a model is significantly less computationally and resource intensive, the need for GPU and specialised expertise can similarly render this option as inaccessible to researchers.

This work addresses this need by utilizing pre-trained, fine-tuned, public BMFMs for inference only, using derived inferences to engineer new BMFM-based features that we then utilised in explainable machine learning (ML) workflows to predict patient drug response from omics data. This offers an accessible and practical solution to utilizing the power of BMFMs for drug discovery tasks.

The ability to stratify patients based on their response to a specific drug is key to facilitating the recommendation of interventions or treatments that are personalised to the individual, and could help mitigate high failure rates during Phase II/III trials [1]. Here, we focus on predicting drug response for inflammatory bowel diseases (IBDs) using data derived from surgically resected diseased tissues from IBD patients. From these tissues REPROCELL Europe Ltd derived multi-omic (genomic, transcriptomic) data for 104 patient samples, and measured inflammatory cytokine TNFa level with and without prednisolone, using a proprietary lab assay, as a proxy for preclinical drug efficacy [2].

We exploit the multi-modality and multi-task BMFM called MAMMAL (Molecular Aligned Multi-Modal Architecture and Language), pre-trained on large-scale biological datasets including proteins and small-molecules, and fine-tuned and evaluated on diverse downstream tasks, including a drug-target binding affinity (BA) task that we use here [3]. Firstly, MAMMAL was used to predict drug BA for all unique human proteins with our drug of interest, prednisolone. Proteins were subsequently ranked by BA, allowing identification of priority drug target proteins. Missense patient-specific SNPs (those that result in a different amino acid being encoded in a protein) were then used to mutate the priority drug target proteins, resulting in a patient-specific set of mutated proteins. MAMMAL was used again to predict prednisolone BA for all mutated proteins, and patient-specific BAs were calculated via comparison to BA of the reference protein.

BMFM-derived feature sets were composed and used alongside non-BMFM feature sets to predict patient-specific prednisolone response using an explainable ML approach. Results revealed superior predictive power of the BMFM-derived feature sets for predicting patient-specific drug response and identification of biologically informative predictive features via explainability methods. We demonstrate the downstream application of BMFMs to gain advantage for a small patient cohort. Moreover, we use only pre-trained, fine-tuned FMs to showcase the benefits of existing public multi-omic BMFMs. In doing so, we aim to expand contexts for FM use for non-experts and for those with limited computational resource.

2 Methods

2.1 BMFM selection and inference

For all inference tasks the open-source MAMMAL model was used (https://huggingface.co/ibm-research/biomed.omics.bl.sm.ma-ted-458m). This model was fine-tuned and evaluated on eleven diverse downstream tasks, where it reached a new state of the art (SOTA) in nine tasks and was SOTA comparable in two tasks [3]. The specific task that we used was for drug-target binding affinity (BA) prediction. Binding affinities (BAs) of every unique reference protein in the human GRCh38 reference genome (and by proxy their associated genes and SNPs), with the SMILES representation of our drug of interest prednisolone, were predicted using this fine-tuned MAMMAL FM. To derive a representative set of unique human proteins in the GRCh38 reference, overlapping the requirements of MAMMAL for sequence lengths under 1250bp, we focused on those proteins over 200aa in length and less than 1240aa (37,517 of which 11,417 unique genes were represented so we selected the primary protein isoform labeled ".1" for each gene). Each of the 11,417 proteins was run in turn through the MAMMAL drug-target BA model for inference to assess its BA to our target drug prednisolone. The SMILES representation that we used for our target drug of interest prednisolone was taken from Drugbank (https://go.drugbank.com/drugs/DB00860).

Proteins were then ranked by BA (Figure 1A) and the top 5% most tightly binding proteins were selected, resulting in a feature set of 600 proteins. Missense SNPs were identified per patient from the genomics data and were used to mutate the 600 proteins (python package Bio.seq.MutableSeq), resulting in a per-patient mutated protein set. BAs of mutated proteins with prednisolone were then

predicted using the fine-tuned MAMMAL model. A final patient-specific BA was then calculated by subtracting the reference BA from the patient mutated protein BA.

2.2 Drug response prediction

To determine if using BMFM inference for feature selection and engineering improved prediction of drug response, we collated three separate feature sets. Feature set (1) consisted of raw SNP data, comprising unique SNPs per person converted to either; 2 (homozygous alternate allele), 1 (heterozygous) or 0 (homozygous reference if no SNP was recorded for that patient at that position). Feature set (2) consisted of the the raw SNPs in feature set (1) filtered to contain only those overlapping the top 600 proteins with the strongest BAs to prednisolone as derived using BMFM inference. Feature set (3) consisted of patient-specific differences in prednisolone BAs with respect to the reference protein set, for the top 600 binding proteins. Accordingly, positive patient-specific BAs indicated proteins with predicted increased prednisolone BA compared to the reference protein whilst negative values indicated the opposite.

These three feature sets were then used to train ML classifiers to predict TNFa level (a proxy for patient response to prednisolone), as a binary classification task where we predict high versus low TNFa level response groups. As all human tissue data derived in the laboratory was normalized against a patient-matched non-treated control group, lower TNFa levels (class 0), was interpreted as a better patient response group to the test drug (reduced inflammation post-treatment). For prediction of drug response using feature set (1) we also performed traditional feature selection of SNPs using ANOVA F-values for each column of features in the training data in relation to the target for prediction. For downstream ML (for the three feature sets) we used AutoXAI4Omics to compare a range of classifiers (Random Forest, AdaBoost, KNN and AutoSKlearn) with a train/test split of 80/20 and a random search for hyperparameter tuning. AutoXAI4Omics trains and tunes a model per classifier and then selects a "best" tuned model using an algorithm to balance performance metrics such as overfitting, performance on 5-fold cross validation (CV) and on held-out test data. Here we report the defined best tuned model for each feature set. AutoXAI4Omics applies the explainable AI algorithm SHAP (SHapley Additive exPlanations) to derived ML models.

3 Results and Discussion

The top 600 proteins as identified by our BMFM-inference guided feature selection contained the two known main targets of prednisolone plus a range of additional biologically coherent targets. Prediction of patient-specific drug response using our three feature sets revealed that the BMFM inference derived BAs (feature set 3, test F1 0.91) considerably outperformed prediction using both other feature sets on held out test data (Table 1); SNP data with traditional feature selection (feature set 1, test F1 0.81) and SNP data filtered by BMFM-led feature selection i.e., most highly bound proteins (feature set 2, test F1 0.82). Furthermore, feature set 3 showed reduced variation on CV (Figure 1B). This demonstrates that the predicted change in prednisolone BA between the patient SNP-mutated proteins and corresponding reference proteins has a promising ability to accurately classify whether a patient will exhibit a high or low response to prednisolone treatment.

Table 1: Performance of ML classification of patients as class 0 - better, and class 1 - worse, responders to prednisolone for feature set (1) raw SNP data, feature set (2) raw SNPs filtered for only those overlapping with the top 600 proteins with the strongest BAs to prednisolone, and feature set (3) patient-specific differences in prednisolone BAs with respect to the reference protein set. Train F1 and Test F1(C0/C1/WM) denote the F1 scores on the training data and on the held-out test data for class 0/class 1/weighted mean respectively. CV F1 (Med/SD) denote the median and standard deviation of F1 scores after 5-fold CV during training.

Feature Set	Train F1	Test F1(C0)	Test F1 (C1)	Test F1 (WM)	CV F1 (Med)	CV F1 (SD)
1	1.000	0.75	0.86	0.81	0.79	0.19
2	1.000	0.80	0.83	0.82	0.80	0.18
3	0.975	0.88	0.92	0.91	0.79	0.07

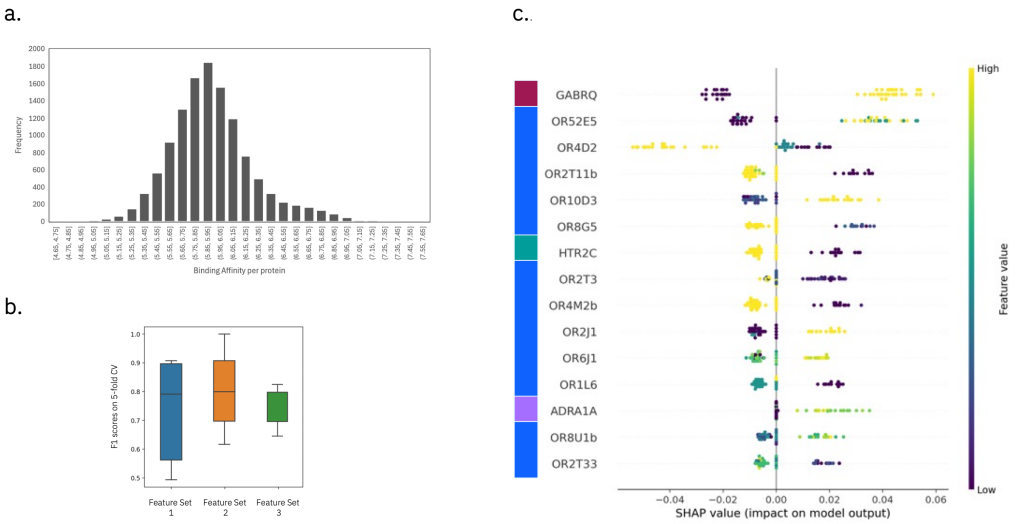

Figure 1: Performance and explainability. (A) Histogram to show derived BAs with prednisolone for all reference proteins after MAMMAL inference. (B) Box plots show F1 scores of hyper-tuned ML models during 5-fold CV for the feature sets (C) Global SHAP dot plot, explainability output relating to the best performing model (Random Forest) to predict class 0. The y-axis ranks the relative importance of the proteins on the model from high to low impact, from top to bottom, and plotted against their SHAP impact values (x-axis). Rows are BAs per protein, dots are patient samples, and dot colour represents the feature value (difference in BA of the patient compared to the reference value) of that sample for the corresponding feature. Colour bar groups proteins by superfamily.

For the BMFM-predicted BA model (feature set 3), we computed SHAP explainability to understand for which proteins the change in BAs due to SNP mutations was most influential for predicting drug response. Deriving this information is our main reason for generation of the model i.e., to assist biological investigations. This analysis revealed for that for prediction of class 0 (good drug response), a higher BA of protein GABRQ (gamma-aminobutyric acid receptor subunit theta) had a positive influence in predicting a good drug response, while a lower BA had a negative influence in predicting a good drug response (Figure 1C). Examination of this protein revealed the presence of one SNP (chrX:152652814, A/T) in 19 patients, which resulted in a change in protein sequence (Ile478Phe) and decreased the protein-drug BA (pKd) compared to the reference, whilst all other patients exhibiting zero SNPs had no change in BA (0). In consideration of this, the higher BAs for GABRQ represent the unmutated protein whilst lower BAs represent reduced prednisolone BA due to the effect of the SNP. (Figure 1C). This SNP, and its resulting lower prednisolone BA, negatively impacts prediction of a good drug response, suggesting a possible link to poor response to prednisolone.

GABRQ is a subunit of GABA-A receptors, ionotrophic channels expressed in both the enteric and central nervous systems, where they bind primary antagonist GABA (gamma-aminobutyric acid) [4, 5]. Corticosteroids like prednisolone are not direct ligands of GABA-A receptors [6], however several studies have documented that certain corticosteroids exhibit indirect regulatory effects on GABA-A, likely by influencing the gut microbiota-brain axis [7–9] or influencing production of other GABA receptor modulators [10, 11]. In the gut GABA receptor activation has also been shown to inhibit pro-inflammatory cytokine secretion and increase anti-inflammatory cytokine production [12, 13], highlighting it's role in regulating intestinal inflammation.

SHAP analysis also revealed a number of olfactory receptors (including OR52E5, OR4D2, OR2T11, OR10D3 and OR8G5, (Figure 1C)) were influential in predicting drug response. For the most predictive olfactory receptor OR52E5, a higher prednisolone BA was an important predictor of good drug response. For protein OR52E5, 38 patients had >=1 of 4 SNPs (chr11:5901063,5901270,5901402,5901477) that changed protein sequence (Phe96Cys, Ile165Thr, Asp209Gly, Pro234Leu) and resulted in decreased prednisolone BA. Patients with increased prednisolone BA were those with no SNPs and therefore expressing the unmutated OR52E5 protein.

This suggests that mutations in this protein may reduce a patients response to prednisolone treatement. Olfactory regulators have been associated with regulation of intestinal inflammation and cell differentiation, and IBD patients often have decreased olfactory function [14].

4 Conclusions

Our work utilised BMFM inference for feature selection and to engineer features that allowed accurate prediction of patient drug response, and suggested novel associations between specific SNPs and personalised drug response to prednisolone. In ML our cohort is considered small and findings are therefore tentative, though sample number was relatively high for an ex vivo study of human fresh tissues. Further work is required to confirm and validate these associations, however this demonstrates the potential of using BMFM inference to enhance personalised medicine discoveries. To the best of our knowledge, comparison to related works suggests that this is a novel approach for BMFM-enabled feature generation for drug response prediction. Furthermore, using a protein model for a primarily genomics analysis, we show the possibility to extend BMFM application beyond its intended trained modalities.

References

1. Fountzilas, E., Tsimberidou, A. M., Vo, H. H. & Kurzrock, R. Clinical trial design in the era of precision medicine. *Genome Medicine* **14,** 101. ISSN: 1756-994X. https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-022-01102-1 (2025) (Aug. 31, 2022).
2. Inam Illahi, M. *et al.* Serum Tumor Necrosis Factor-Alpha as a Competent Biomarker for Evaluation of Disease Activity in Early Rheumatoid Arthritis. *Cureus.* ISSN: 2168-8184. https://www.cureus.com/articles/59950-serum-tumor-necrosis-factor-alpha-as-a-competent-biomarker-for-evaluation-of-disease-activity-in-early-rheumatoid-arthritis (2025) (May 29, 2021).
3. Shoshan, Y. *et al. MAMMAL – Molecular Aligned Multi-Modal Architecture and Language* Version Number: 3. 2024. https://arxiv.org/abs/2410.22367 (2025).
4. Sinkkonen, S. T., Hanna, M. C., Kirkness, E. F. & Korpi, E. R. $GABA_A$ Receptor and Subunits Display Unusual Structural Variation between Species and Are Enriched in the Rat Locus Ceruleus. *The Journal of Neuroscience* **20,** 3588–3595. ISSN: 0270-6474, 1529-2401. https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.20-10-03588.2000 (2025) (May 15, 2000).
5. Auteri, M., Zizzo, M. G. & Serio, R. GABA and GABA receptors in the gastrointestinal tract: from motility to inflammation. *Pharmacological Research* **93,** 11–21. ISSN: 10436618. https://linkinghub.elsevier.com/retrieve/pii/S104366181400187X (2025) (Mar. 2015).
6. Phulera, S. *et al.* Cryo-EM structure of the benzodiazepine-sensitive 112S tri-heteromeric GABAA receptor in complex with GABA. *eLife* **7,** e39383. ISSN: 2050-084X. https://elifesciences.org/articles/39383 (2025) (July 25, 2018).
7. So, S. Y. & Savidge, T. C. Gut feelings: the microbiota-gut-brain axis on steroids. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **322,** G1–G20. ISSN: 0193-1857, 1522-1547. https://journals.physiology.org/doi/10.1152/ajpgi.00294.2021 (2025) (Jan. 1, 2022).
8. Tetel, M. J., De Vries, G. J., Melcangi, R. C., Panzica, G. & O'Mahony, S. M. Steroids, stress and the gut microbiome-brain axis. *Journal of Neuroendocrinology* **30,** e12548. ISSN: 0953-8194, 1365-2826. https://onlinelibrary.wiley.com/doi/10.1111/jne.12548 (2025) (Feb. 2018).
9. McCurry, M. D. *et al.* Gut bacteria convert glucocorticoids into progestins in the presence of hydrogen gas. *Cell* **187,** 2952–2968.e13. ISSN: 00928674. https://linkinghub.elsevier.com/retrieve/pii/S0092867424005142 (2025) (June 2024).
10. De Lima, A. M. D. L. *et al.* Effect of prednisolone in a kindling model of epileptic seizures in rats on cytokine and intestinal microbiota diversity. *Epilepsy & Behavior* **155,** 109800. ISSN: 15255050. https://linkinghub.elsevier.com/retrieve/pii/S1525505024001811 (2025) (June 2024).
11. Di, S., Maxson, M. M., Franco, A. & Tasker, J. G. Glucocorticoids Regulate Glutamate and GABA Synapse-Specific Retrograde Transmission via Divergent Nongenomic Signaling Pathways. *The Journal of Neuroscience* **29,** 393–401. ISSN: 0270-6474, 1529-2401. https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.4546-08.2009 (2025) (Jan. 14, 2009).
12. Deng, Z. *et al.* Activation of GABA receptor attenuates intestinal inflammation by modulating enteric glial cells function through inhibiting NF-B pathway. *Life Sciences* **329,** 121984. ISSN: 00243205. https://linkinghub.elsevier.com/retrieve/pii/S0024320523006197 (2025) (Sept. 2023).
13. Ma, X. *et al.* Activation of GABAA Receptors in Colon Epithelium Exacerbates Acute Colitis. *Frontiers in Immunology* **9,** 987. ISSN: 1664-3224. http://journal.frontiersin.org/article/10.3389/fimmu.2018.00987/full (2025) (May 7, 2018).

14. Sollai, G. *et al.* Olfactory Function in Patients with Inflammatory Bowel Disease (IBD) Is Associated with Their Body Mass Index and Polymorphism in the Odor Binding-Protein (OBPIIa) Gene. *Nutrients* **13,** 703. ISSN: 2072-6643. https://www.mdpi.com/2072-6643/13/2/703 (2025) (Feb. 22, 2021).