A systems biology-informed machine learning framework to predict drug response using clinically available NGS data

Maayan Baron, Andrey Chursov, Brandon Funkhouser, Jacob Kaffey, Sunil Kumar, George Komatsoulis, Felicia Kuperwaser, Jeff Sherman, Dillon Tracy, and Emily Vucic. Zephyr AI, McLean, VA, USA

Background

Tremendous progress towards understanding the molecular underpinnings of cancer has ushered in an era of precision medicine, where actionable biomarkers inform rational design of novel drugs, repurposing of existing medications or new combination therapies, agnostic of tumor type. However, less than 20% of cancer patients match with therapies using existing predictive biomarkers. There is therefore an acute need for new methods to identify the expanded population of patients who could benefit from existing drugs and improve patient selection and design of clinical trials evaluating new or repurposed drugs.

Zephyr's ML algorithms identify pharmacological sensitivities and genetic vulnerabilities

We developed systems biology-informed AI methods trained on vast multi-omics and multi-modal datasets from DepMap, TCGA and other publicly available sources, to uncover latent vulnerabilities existing in tumors (Figure 1). These vulnerabilities map to pharmacological sensitivities and can be assigned to individual patients using clinically available molecular data. Foundational to the interpretability and novel biological insights generated by our models is the use of structured biological networks[3,4], which are a commonly used framework in systems biology and increasingly used to build interpretable ML models for computational biology.

Figure 1 | Zephyr AI computational framework. Structural embeddings of small molecule drugs of interest along with mutation and copy number data from commercial NGS cancer panels and a subset of clinical data are used to generate drug sensitivity predictions for each drug on a per patient basis (Left). Biological interpretability is achieved through assignment of a Vulnerability Network (VN), which represents latent genetic vulnerabilities operative in tumors that correspond to pharmacological sensitivities (Right). VNs further expose potential targets and provide biological context for existing targets.

Zephyr's algorithms generate interpretable drug response predictions from Real-World Data

To evaluate our model, we focused on predicting sensitivities to Tagrisso, a third-generation tyrosine kinase inhibitor (TKI), which we assessed in non-small cell lung cancer (NSCLC) cell lines, patient derived xenografts (PDXs) and a clinicogenomics cohort we assembled from two distinct RWD cohorts (n=334 patients treated with TKI of interest).

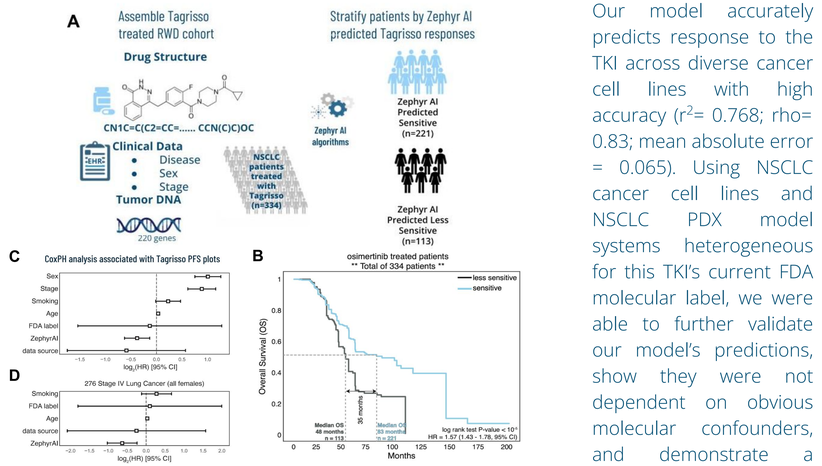


Figure 2 | NSCLC patients Zephyr predicted to be sensitive to Tagrisso have longer progression free survival. A) Data obtained in real-world settings were used to assign drug sensitivity predictions to individual patients in a RWD cohort of NSCLC patients treated with Tagrisso. B) Kaplan-Meier plot depicting corresponding OS measurements of patients our algorithms predicted to be sensitive vs. insensitive. NSCLC patients our algorithms predicted to be sensitive experienced a 35 month increased median OS than the predicted less sensitive group; Log-rank test p-value<10^{-5}; HR= 1.57 (1.43-1.78, 95% CI). C) CoxPH analysis indicates that in addition to ZephyrAI prediction, sex and stage might be confounding factors. D) Subsetting the population to stage IV females still produces significance based on Zephyr AI prediction.

Our model accurately predicts response to the TKI across diverse cancer cell lines with high accuracy (r^2= 0.768; rho= 0.83; mean absolute error = 0.065). Using NSCLC cancer cell lines and NSCLC PDX model systems heterogeneous for this TKI's current FDA molecular label, we were able to further validate our model's predictions, show they were not dependent on obvious molecular confounders, and demonstrate a potential label expansion opportunity for this drug beyond the current narrow molecular label.

Zephyr AI sensitivity predictions validated in preclinical NSCLC PDX models in both FDA-on and FDA-off label contexts

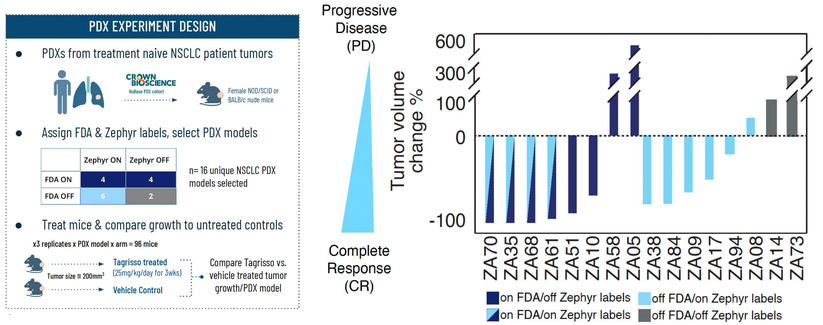
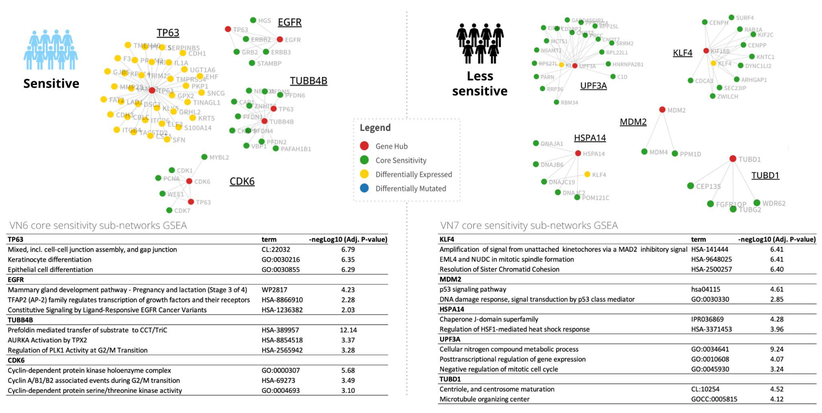


Figure 3 | Preclinical validation of Zephyr AI predicted response to a TKI and demonstration of a label expansion opportunity within NSCLC. The drug's current FDA molecular label was used to assign PDXs to FDA ON or FDA OFF label groups based on the presence or absence of DNA alterations indicated for this drug. PDX tumors were subcutaneously implanted into female NOD/SCID or BALB/c nude mice, in triplicate for a total of 96 mice (n=48 treated; 48 vehicle-treated controls). After implanted tumors reached a size of 200mm^3, mice were randomized into treatment and control groups. Treated mice were dosed with 25mg/kg/day of the TKI for 3 weeks. Changes in tumor volume were calculated as previously described[2].



VN6 core sensitivity sub-networks GSEA

TP63	term	-negLog10 (Adj. P-value)
Mixed, incl. cell-cell junction assembly, and gap junction	CL:22032	6.79
Keratinocyte differentiation	GO:0030216	6.35
Epithelial cell differentiation	GO:0030855	6.29
EGFR		
Mammary gland development pathway - Pregnancy and lactation (Stage 3 of 4)	WP2817	4.23
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors	HSA-8866910	2.28
Constitutive Signaling by Ligand-Responsive EGFR Cancer Variants	HSA-1236382	2.03
TUBB4B		
Prefoldin mediated transfer of substrate to CCT/TriC	HSA-389957	12.14
AURKA Activation by TPX2	HSA-8854518	3.37
Regulation of PLK1 Activity at G2/M Transition	HSA-2565942	3.28
CDK6		
Cyclin-dependent protein kinase holoenzyme complex	GO:0000307	5.68
Cyclin A/B1/B2 associated events during G2/M transition	HSA-69273	3.49
Cyclin-dependent protein serine/threonine kinase activity	GO:0004693	3.10

VN7 core sensitivity sub-networks GSEA

KLF4	term	-negLog10 (Adj. P-value)
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	HSA-141444	6.41
EML4 and NUDC in mitotic spindle formation	HSA-9648025	6.41
Resolution of Sister Chromatid Cohesion	HSA-2500257	6.40
MDM2		
p53 signaling pathway	hsa04115	4.61
DNA damage response, signal transduction by p53 class mediator	GO:0030330	2.85
HSPA14		
Chaperone J-domain superfamily	IPR036869	4.28
Regulation of HSF1-mediated heat shock response	HSA-3371453	3.96
UPF3A		
Cellular nitrogen compound metabolic process	GO:0034641	9.24
Posttranscriptional regulation of gene expression	GO:0010608	4.07
Negative regulation of mitotic cell cycle	GO:0045930	3.24
TUBD1		
Centriole, and centrosome maturation	CL:10254	4.52
Microtubule organizing center	GOCC:0005815	4.12

Figure 4 |Interpretability of Zephyr AI Tagrisso predictions via Vulnerability Networks. Tagrisso-sensitive tumors harbor genetic sensitivity to EGFR perturbation, but are not defined by EGFR mutations, TP63 and CDK6 genetic dependence. Combination of CDK4/6 inhibitor and Tagrisso was demonstrated to overcome Tagrisso acquired resistance[6]. Tagrisso-insensitive tumors are defined in part by a genetic sensitivity to Krüppel-like factor 4 (KLF4). KLF4 has been demonstrated to promote c-Met amplification-mediated TKI resistance in NSCLC[7] and efficacy of TKIs in TNBC by repressing EGFR[8].

Conclusion

We have developed an ML model that uses clinically available NGS data to accurately predict increased sensitivity to Tagrisso, a third-generation TKI, validated in NSCLC cell lines, PDX models and a RWD patient cohort.

References

1. https://www.aacr.org/professionals/research/aacr-project-genie/aacr-project-genie-data/
2. Gao, H. et al., Nat. Med. 21, 1318–1325 (2015).
3. Kim, C. Y. et al Nucleic Acids Res. 50, D632–D639 (2022).
4. Huang, J. K. et al., Cell Systems vol. 6 484–495.e5 (2018).
5. Yi, M. et al., Cell. Mol. Life Sci. 77, 4325–4346 (2020).
6. Qin, Q. et al., Thorac. Cancer 11, 2389–2397 (2020).
7. Feng et al., Cancer Sci. 2018 Jun; 109(6): 1775–1786.
8. Roberts et al., Breast Cancer Res. 2020; 22: 66.

Acknowledgements

The authors would like to acknowledge the following:

- Zephyr AI™ Scientific Board members and advisors: Dr. Justin Stebbing, Dr. Tom Maniatis, Dr. Richard L. Schilsky, Dr. Alan Barge, Dr. Siddhartha Mukherjee, Dr. Fanziska Michor and Dr. George Miller
- Zephyr AI engineering and data teams
- Candy Zhu
- Aster Insights® (previously M2GEN®)
- Chelsea Bailey and Crown Bio®

Contact: maayan@zephyrai.bio

Abstract #3136
Poster #334