



StarGazer: A Hybrid Intelligence Platform for Drug Target Prioritization and Digital Drug Repositioning Using Streamlit

OPEN ACCESS

Edited by:

Rinku Sharma,
Harvard Medical School,
United States

Reviewed by:

Rajesh Kumar Pathak,
Chung-Ang University, South Korea
Sezen Vatansever,
Icahn School of Medicine at Mount
Sinai, United States

*Correspondence:

Khader Shameer
shameer.khader@astrazeneca.com

[†]These authors contributed equally to
this work

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 01 February 2022

Accepted: 29 April 2022

Published: 31 May 2022

Citation:

Lee C, Lin J, Prokop A,
Gopalakrishnan V, Hanna RN, Papa E,
Freeman A, Patel S, Yu W, Huhn M,
Sheikh A-S, Tan K, Sellman BR,
Cohen T, Mangion J, Khan FM,
Gusev Y and Shameer K (2022)
StarGazer: A Hybrid Intelligence
Platform for Drug Target Prioritization
and Digital Drug Repositioning
Using Streamlit.
Front. Genet. 13:868015.
doi: 10.3389/fgene.2022.868015

Chiyun Lee^{1†}, Junxia Lin^{2†}, Andrzej Prokop³, Vancheswaran Gopalakrishnan⁴,
Richard N. Hanna⁵, Eliseo Papa⁶, Adrian Freeman⁷, Saleha Patel⁷, Wen Yu⁸, Monika Huhn⁹,
Abdul-Saboor Sheikh¹, Keith Tan¹⁰, Bret R. Sellman⁴, Taylor Cohen⁴, Jonathan Mangion¹,
Faisal M. Khan⁸, Yuriy Gusev² and Khader Shameer^{8*}

¹Data Science and Artificial Intelligence, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, ²Georgetown University, Washington, DC, United States, ³Biometrics, Oncology R&D, AstraZeneca, Warsaw, Poland, ⁴Discovery Microbiome, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, United States, ⁵Early Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, United States, ⁶Research Data and Analytics, R&D IT, AstraZeneca, Cambridge, United Kingdom, ⁷Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, ⁸Data Science and Artificial Intelligence, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD, United States, ⁹Biometrics and Information Sciences, BioPharmaceuticals R&D, AstraZeneca, Mölndal, Sweden, ¹⁰Neuroscience, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom

Target prioritization is essential for drug discovery and repositioning. Applying computational methods to analyze and process multi-omics data to find new drug targets is a practical approach for achieving this. Despite an increasing number of methods for generating datasets such as genomics, phenomics, and proteomics, attempts to integrate and mine such datasets remain limited in scope. Developing hybrid intelligence solutions that combine human intelligence in the scientific domain and disease biology with the ability to mine multiple databases simultaneously may help augment drug target discovery and identify novel drug-indication associations. We believe that integrating different data sources using a singular numerical scoring system in a hybrid intelligent framework could help to bridge these different omics layers and facilitate rapid drug target prioritization for studies in drug discovery, development or repositioning. Herein, we describe our prototype of the StarGazer pipeline which combines multi-source, multi-omics data with a novel target prioritization scoring system in an interactive Python-based Streamlit dashboard. StarGazer displays target prioritization scores for genes associated with 1844 phenotypic traits, and is available via <https://github.com/AstraZeneca/StarGazer>.

Keywords: multi-omics, target prioritization, drug discovery, repositioning, data integration, streamlit, stargazer, hybrid intelligence

INTRODUCTION

Drug repositioning has been rapidly gaining attention in the drug discovery domain during the past decade (Xue et al., 2018). Drug repositioning/repurposing describes the act of identifying alternative uses for a drug beyond the scope of its original indication, regardless of whether it has been FDA-approved or has failed in clinical trials (Pushpakom et al., 2019). The reasons for investing into drug repositioning are very numerous indeed.

Traditionally, a standard drug development cycle is estimated to take around 10 years and requires billions of dollars of investment, notwithstanding the still disappointingly high failure rate at clinical trials (Li et al., 2016). In light of these problems, drug repositioning holds potential to drastically reduce the time and money needed to bring a drug to the market: it has been estimated to reduce the time by half and cut costs by 5-fold when compared to developing a new drug from scratch (Shameer et al., 2015). These factors alone highlight the appealing opportunity to bring medicines to patients faster, and potentially into areas of unmet therapeutic demand. Moreover, it allows for the existing arsenal of approved drugs to be more broadly utilized, and for the opportunity to salvage some costs involved in the development of drugs that failed in clinical trials. Finally, the sheer variety in successful and promising repositioning strategies to date speaks to the potential for unearthing profound biological links between different diseases, driving paradigm shifts in our approach to modern medicine (Lee and Bhakta, 2021).

Drug target prioritization is an essential step for repositioning as it aims to highlight the potential drug targets for a particular disease. Applying computational methods to analyze and process multi-omics data is an effective approach for achieving this (Ashburn and Thor, 2004; Glicksberg et al., 2014; Shameer et al., 2018a; Pushpakom et al., 2019; Guo et al., 2021; Rapicavoli et al., 2022). Whilst there is now a vast wealth of biochemical and biomedical data in the current era of high-throughput omics technology, our ability to integrate and interpret these data has lagged behind and is presenting a great challenge in disease biology (Shameer et al., 2015). While machine learning approaches are generally used to develop tools to integrate, analyze and interpret multi-omics data, it remains a challenge that mere automation of predicting biological insights might overrepresent hypotheses that cannot be validated using function test experiments (Hodos et al., 2016; Peters et al., 2017). In such a scenario, we recommend the application of a hybrid intelligence platform that enables visual intelligence, quick search, contextual interpretations with quantitative approaches as a way to address this problem. Hybrid intelligence systems have been developed to address challenging problems in biomedicine, including remote patient diagnosis (Abu-Doleh et al., 2012; Li et al., 2014a; Akata et al., 2020; Guo et al., 2021; Weissler et al., 2021). However, such approaches are not readily available to address challenges in data integration and mining associated with drug target prioritization and drug repositioning.

Data from genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS) have been used

for drug target prioritization (Ferrero and Agarwal, 2018). Whilst GWAS aim to identify associations between genetic variants with a single phenotype, PheWAS interrogate numerous phenotypic traits at once (Denny et al., 2010). As of 06 October 2021, the EMBL-EBI GWAS catalog collates associations from 5,370 studies that, in total, identified more than 290,000 associations. The utility of this GWAS dataset can be further amplified by narrowing down the genes of interest to only those with known drug indications (Sanseau et al., 2012). Importantly, a three-step strategy for drug repositioning using PheWAS data has already been proposed (Rastegar-Mojarad et al., 2015): (Xue et al., 2018)—identify all genes with known associations with the phenotypic trait of interest using PheWAS data; (Pushpakom et al., 2019);—identify all drugs with associations with the previously identified genes using data from DrugBank; and (Li et al., 2016)—return all the drugs identified in the previous step as candidates for repositioning for the original phenotypic trait of interest. Others have gone further by incorporating a combination of data from GWASs (Khosravi et al., 2019), expression profile analysis (Lau and So, 2020), functional annotation, biological network analysis, and gene-set association (Reay and Cairns, 2021).

Taken together, these data highlight the potential of using GWAS and PheWAS data for drug target prioritization. However, the field is still young, and integrating disparate data sources remains relatively limited in scope (Gallo et al., 2021). We hypothesize that integrating multimodal data sources using a singular numerical scoring system could accelerate the discovery and prioritization of drug targets. In light of this, we present our interactive dashboard, StarGazer, which aims to address these challenges by integrating three different datatypes (i.e., disease-target association, target druggability, and target protein-protein interaction) into a novel scoring system, utilizing real-time API calls and Python-based Streamlit technology. While these types of datasets have been used for numerous repositioning studies separately (Liu et al., 2014; Khaladkar et al., 2017; Hermawan et al., 2020; Wijetunga et al., 2020; Adikusuma et al., 2021; Attique et al., 2021; Ghousaini et al., 2021; Portelli et al., 2021; Tan et al., 2021; Varghese and Majumdar, 2022; Zhao et al., 2022), StarGazer represents the first ever integration of the PheWAS catalog, Open Targets, STRING and Pharos, all of which are well-curated, well-studied, open access databases. Furthermore, computational repositioning studies focus largely on singular diseases, phenotypes or drugs, but StarGazer is equipped for flexible investigation into any of the 1,844 phenotypes and traits within the dashboard. Much of the data is up-to-date with the latest science, as it is loaded in real-time before it is analyzed in real time. StarGazer's drug target prioritization mode allows for rapid identification of potential drug targets for a disease of interest, also providing immediate analysis of various aspects surrounding drug development, such as druggability and the nature of the target-disease association. In addition to this target prioritization feature, we anticipate that StarGazer's ability to display all phenotypes associated with genes or gene variants of interest in an easily digestible manner to be of great value to exploratory or analytical workflows. Furthermore, StarGazer's other features include the support of initial

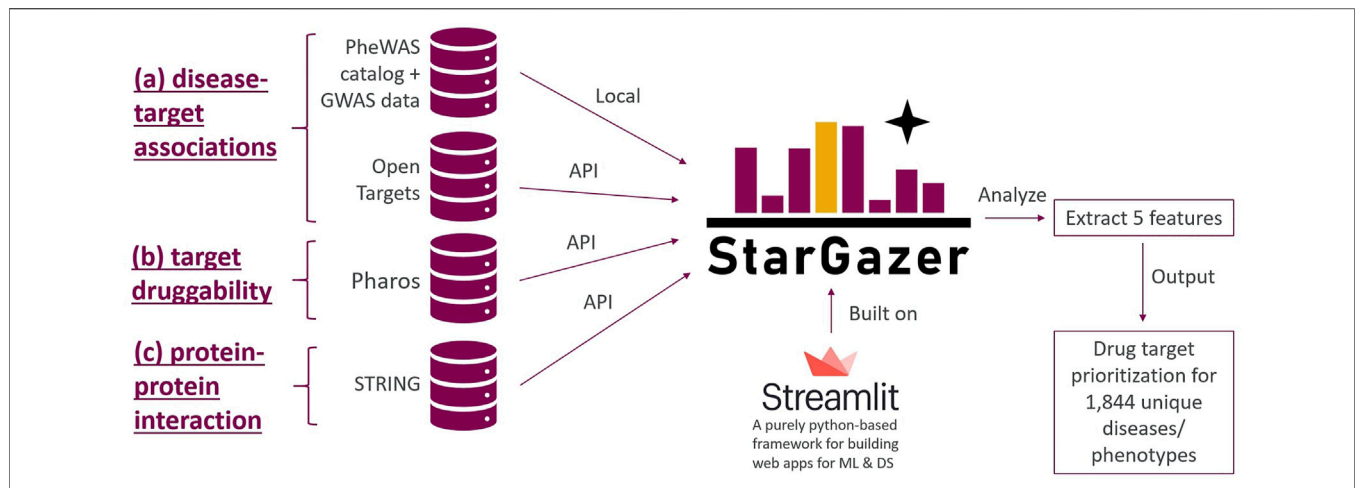


FIGURE 1 | The StarGazer drug target prioritization framework considers the following five features for each of the 1844 diseases in StarGazer’s disease list (Xue et al., 2018):—the odds ratios of association between targets and phenotypic variants of interest from GWAS and PheWAS data (Pushpakom et al., 2019);—the target-disease association scores from Open Targets (Li et al., 2016);—the druggability data of genes of interest from Pharos (Shameer et al., 2015);—the degree of nodes in protein-protein interaction networks of genes of interest from STRING; and (Lee and Bhakta, 2021)—the presence of the gene variant of interest in both PheWAS and GWAS datasets. All data, except the PheWAS and GWAS data, are loaded in real-time by API calls and therefore present the latest evidence for drug repositioning strategies. The above five features are then integrated to provide a singular numerical StarGazer score which quantifies the drug repositioning potential of a gene. StarGazer is built on the Python-based Streamlit platform, which is largely used for building sleek and modern web applications for machine-learning and data science.

discoveries by interrogating the precise contribution of evidence from each data source.

DATA

Disease-target data are acquired from the PheWAS catalog (<https://phewascatalog.org/phewas>) and OpenTargets (<https://genetics.opentargets.org/>). The latest PheWAS catalog was created in 2013 by generating odds ratios of association between 3,144 SNPs identified in GWASs and 1,358 phenotypes derived from the electronic medical records of 13,835 individuals of European ancestry, and the data is loaded locally (Denny et al., 2013). The list of phenotypic variants from the PheWAS catalog as well as from the GWASs within the PheWAS catalog were aggregated and filtered to remove duplicates, producing a list of 1844 phenotypic traits which StarGazer uses for subsequent analysis. OpenTargets version 22.02 is the latest version at the time of writing, and provides 7,980,448 target-disease association scores extracted from 21 public databases containing diverse forms of evidence, from genetic and drug associations to text mining and animal model data amongst others (Ochoa et al., 2021). Data from OpenTargets is acquired in real-time via API calls.

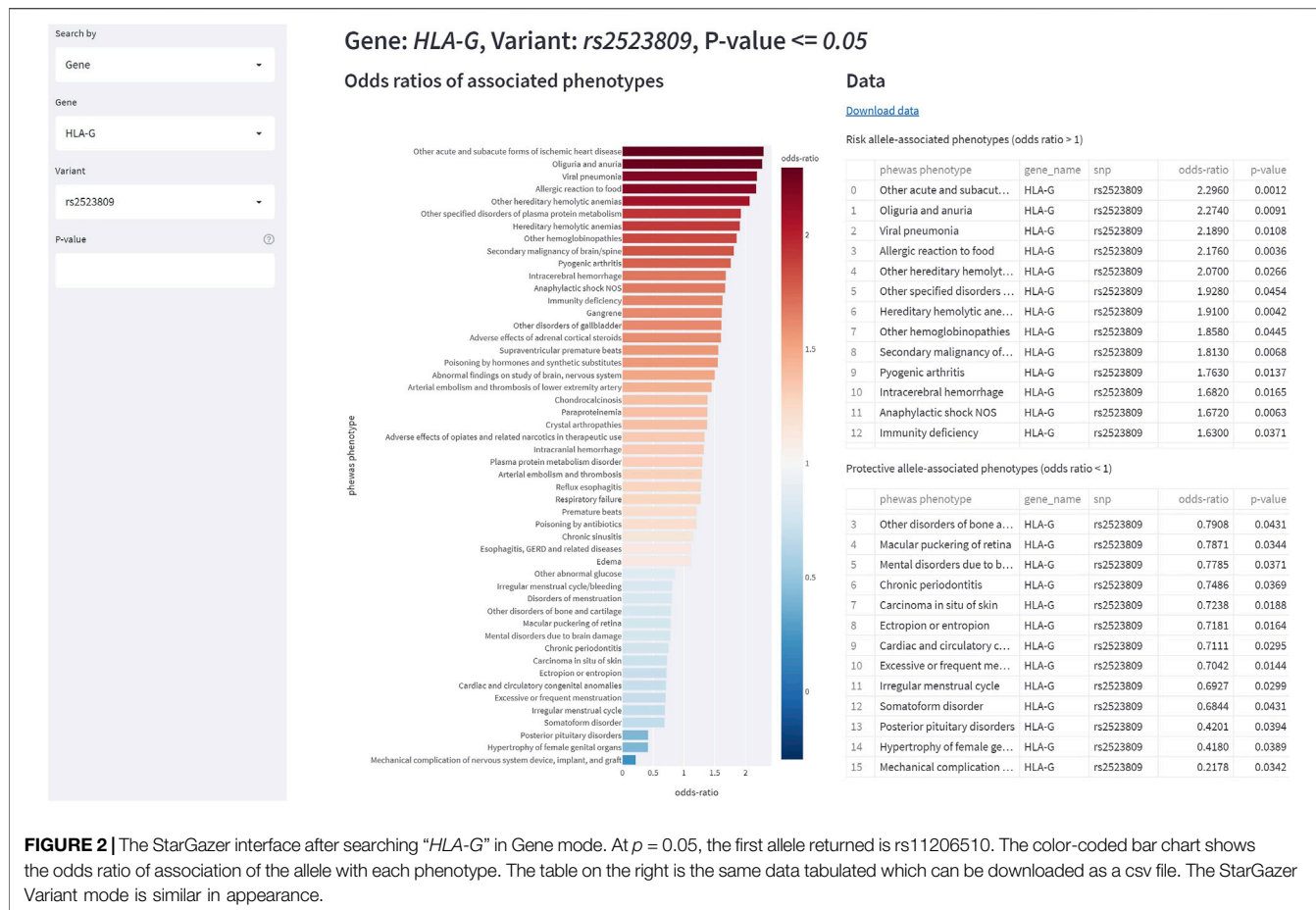
Target druggability data are acquired in real-time via API calls through Pharos (<https://pharos.nih.gov/>) to access the Target Central Resource Database (TCRD) (Sheils et al., 2021). The TCRD categorizes 20,412 targets, at the time of writing, into four groups of increasing druggability evidence: Tdark, Tbio, Tchem, and Tclin. A variety of evidence is integrated for classification, such as data from ChEMBL

(Mendez et al., 2019), Guide to Pharmacology (Armstrong et al., 2019), DrugCentral (Avram et al., 2021), and antibodypedia (Kiermer, 2008), amongst many more, as well as gene ontology and text-mining analysis. Tclin genes are already targets of approved drugs, whilst Tchem genes have drugs with evidence of sufficient activity against the gene. Tbio genes have weak evidence for druggability, and Tdark genes have an unknown level of druggability.

Protein-protein interaction data are acquired in real-time via API calls from the STRING database (<https://string-db.org/>). STRING version 11.5 contains data of 20,052,394,042 protein-protein interactions from 14,094 organisms, of which only human genes and orthologous genes were used in StarGazer (Szkłarczyk et al., 2021), which were analyzed using the Python package, pyvis. Gene ontology enrichment analysis is also performed by STRING.

METHODS

StarGazer was built using Streamlit (<https://streamlit.io/>), a relatively new Python-based tool for developing web applications for machine learning and data science. It enables data scientists to build web applications purely from Python scripts quickly and seamlessly. The Streamlit dashboard allows for local files to be loaded, as well as data to be requested from databases via real-time API calls. The StarGazer drug target prioritization framework considers the following five features for each disease (Figure 1): (Xue et al., 2018)—the odds ratios of association between targets and phenotypic variants of interest from GWAS and PheWAS data; (Pushpakom et al., 2019);—the target-disease association scores from Open Targets; (Li et al.,



2016);—the druggability data of genes of interest from Pharos; (Shameer et al., 2015);—the degree of nodes in protein-protein interaction networks of genes of interest from STRING; and (Lee and Bhakta, 2021)—the presence of the gene variant of interest in both PheWAS and GWAS datasets. Each gene was analyzed with respect to each of these five features, and five scores were computed corresponding to each of the above features. These five scores were then normalized to ensure equal maximum contribution, before summing the five normalized scores to obtain an overall score (i.e., the StarGazer score) which has a maximum score of 1. The targets were then ranked in descending order to facilitate target prioritization.

Processing of Disease-Target Data

Analysis of the PheWAS and GWAS odds ratios involved identifying risk associations where the odds ratio ≥ 1 (i.e., more associated with the occurrence of the phenotype), and protective associations where the odds ratio < 1 (i.e., more associated with the non-occurrence of the disease). In the risk allele-based target prioritization, odds ratios were taken as they were. However, in protective-allele-based target prioritization, odds ratios were subtracted by 1, as the lower ratio implies higher magnitude of association. An average value was taken for odds ratios from multiple studies of the same gene, before normalizing to generate the feature score. Another feature score was generated

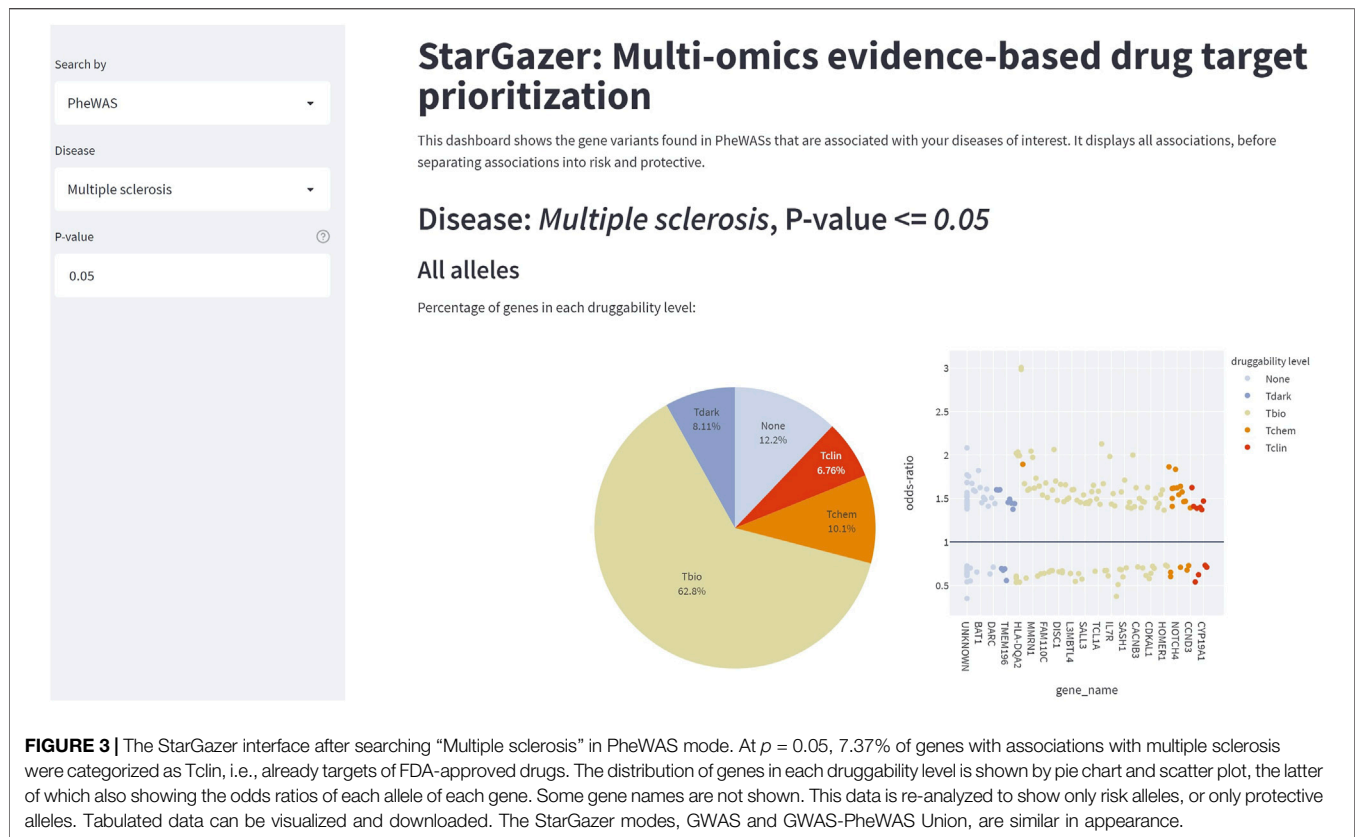
by determining if the gene target was present in both the PheWAS and GWAS datasets, assigning a score of 1 for the PheWAS-GWAS intersection score, which is otherwise 0. Finally, the target-disease association feature scores from OpenTargets were values between 0 and 1, calculated in a similar manner as the PheWAS catalogue analysis.

Processing of Target Druggability Data

For analysis of the druggability data from Pharos, the number of distinct druggability levels that a target has was counted, with the exception of Tdark, e.g., a target with Tbio, Tclin, and Tdark labels is scored 2 (1 + 1 + 0). These scores were then normalized against the highest druggability feature score of each gene.

Processing of Protein-Protein Interaction Data

The degree of the node in the protein-protein interaction networks from STRING is the number of proteins directly connected to the target node via functional associations, which include experimentally confirmed interactions, predicted interactions and text mining data. Node degrees were computed for each gene in a network and calculated as a ratio of the highest node degree in that network, as a gene with higher interactivity within a STRING network is more likely to be



biologically underpinning the molecular pathway that contributes to a phenotype. The calculation of node degrees scores this way also reduces effects of false positive interactions.

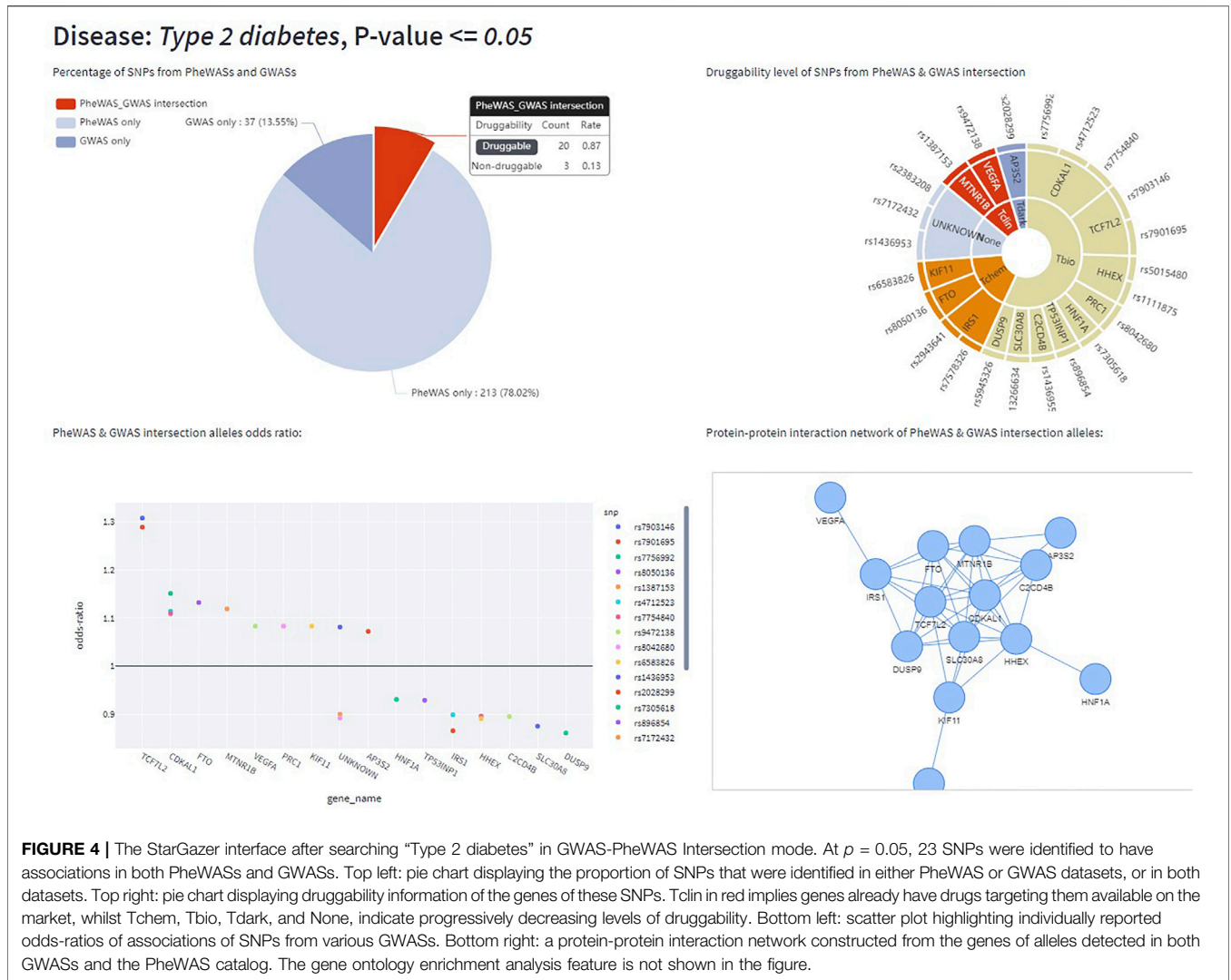
RESULTS

The StarGazer dashboard (<https://github.com/AstraZeneca/StarGazer>) offers eight modes of data exploration for drug target prioritization using the data analyzed as described in Methods. The modern yet simple interface allows for rapid navigation without the need for specialist training or programming experience. StarGazer allows users to search by genes or gene variants which displays all associated phenotypic variants ranked by odds ratio graphically, as well as in tabular format (Figure 2). Red bars indicate an odds ratio of greater than 1 (i.e., risk association), whilst blue bars indicate less than 1 (i.e., protective association). Users can also search by the PheWAS, GWAS, and GWAS-PheWAS Union modes of exploration, which returns odds ratios of all variants of genes associated with the phenotype of interest from the respective datasets, as well as their corresponding druggability levels (Figure 3). When searching in the GWAS-PheWAS Intersection mode, only variants with associations identified in both GWAS and PheWAS datasets are shown (Figure 4). For these variants, the dashboard also provides association odds ratios, druggability data, protein-protein interaction networks and gene ontology enrichment

analysis for the disease of interest (Figure 5). Finally, when users search by disease target prioritization, the overall StarGazer score is shown for each gene with association with the disease of interest (Figure 6). Contextual information on any of these genes can be found immediately using the build-in NCBI search tool. For each of these exploration modes, users can also modify the p -value to only display associations of desired statistical significance assigned by the origin data source.

Use Case: StarGazer for Understanding Complex Diseases

In the following case study, we posed as someone who was simply curious about the possible mechanistic causes of insomnia, and consequently adopted a more exploratory workflow. As insomnia is a complex and relatively understudied disorder, we set the p -value to a less stringent 0.05 to prevent issues in, for example, study sample size or sensitivity from masking any potentially true associations. This returned a list of 106 genes with associations with insomnia, 62 of which had at least one risk-associated allele, and 46 had at least one protection-associated allele (Table 1). After searching on NCBI, there were three genes found to have significant relevance to insomnia. *DISC1* encodes a scaffold protein which is involved in brain development, and its mutations have been implicated in schizophrenia and other psychiatric disorders (Dahoun et al., 2017); *MAOA* encodes a mitochondrial oxidative deaminase targeting amines such as

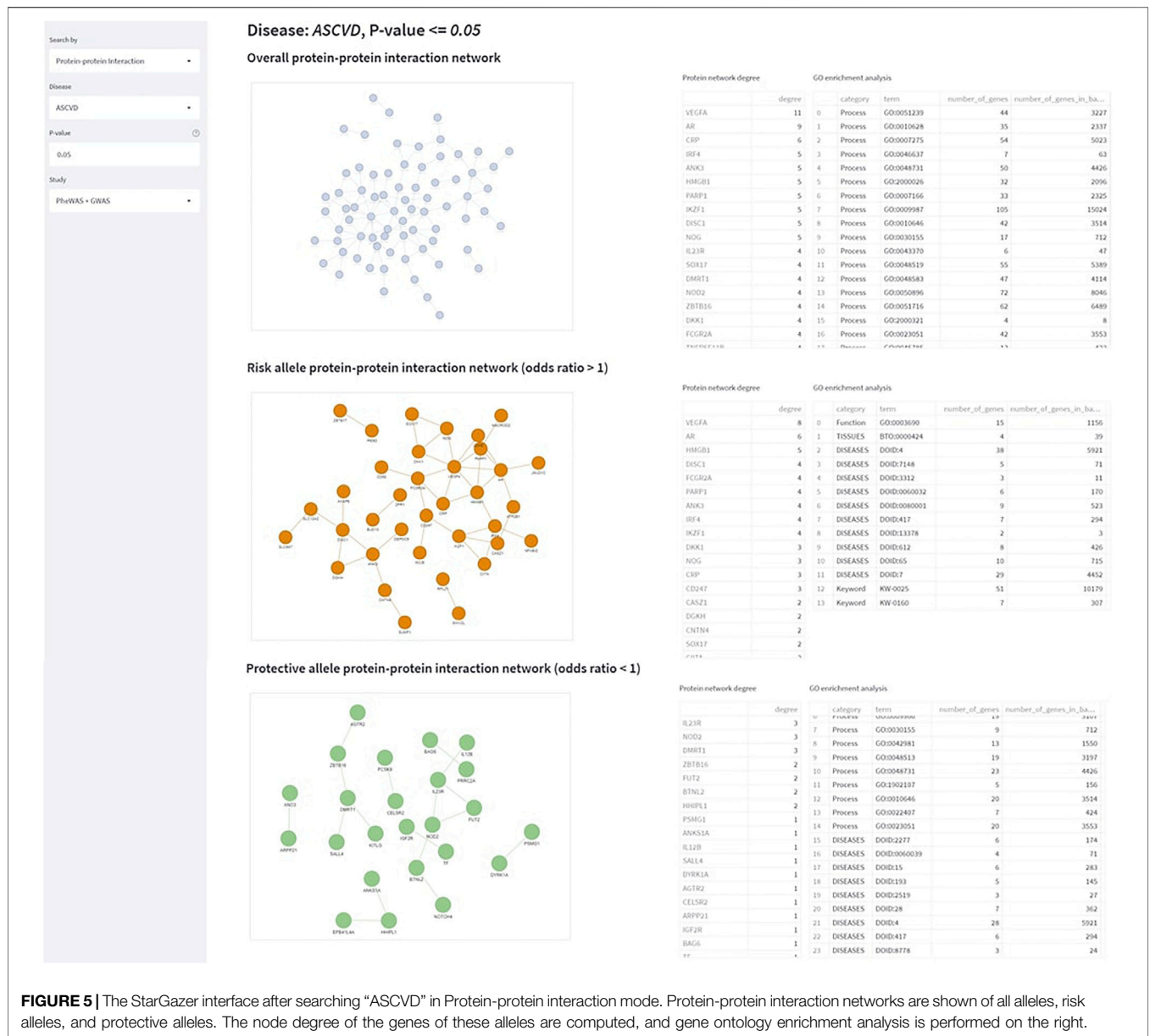


dopamine, norepinephrine, and serotonin, and mutations in the gene can result in Brunner syndrome, a psychiatric and sleep disorder (Brunner et al., 2007); *MEIS1* is a HOX gene thought to have a pleiotropic effect on chronic insomnia disorder, and have possible association with restless leg syndrome (Sarayloo et al., 2019). We also found genes with a variety of functions and unclear links with insomnia. Tumor suppressor genes, *CMTM7* (Li et al., 2014b), *NKAPL* (Okuda et al., 2015) and *ATM* (encoding ATM checkpoint kinase) (Shiloh and Ziv, 2013) may allude to aberrant DNA damage responses contributing to insomnia, and indeed, there are several reports of links between DNA damage and sleep in the literature (Carroll et al., 2016; Zada et al., 2021). HLA isoforms indicate a potential immunity-related cause of insomnia (Choo, 2007). *In vitro* mutants in vesicular trafficking protein, dynamin-1, have impaired ability to recycle neurotransmitter at synapses (Chung et al., 2010), providing a more obvious potential link with insomnia. Finally, genes with noticeably pleiotropic effect were also found to have a high

StarGazer score. One such example is estrogen receptor (*ESR1*), important for gestation in women but is in addition expressed in many non-reproductive tissues in both sexes, as it has roles more broadly in growth and metabolism (Barros and Gustafsson, 2011). Not only is estrogen receptor linked with breast cancer but also with osteoporosis (Gennari et al., 2007), and thus makes for a peculiar hit on the StarGazer analysis. Although additional investigations are required to ascertain the link between these genes and phenotypes, it is exciting to hypothesize about the underlying molecular mechanisms. This is especially the case for insomnia, a disorder of sleep which is a biological process we still have a relatively poor understanding of.

DISCUSSION

StarGazer is a novel application built for rapid investigation of drug repositioning strategies. It combines multi-source, multi-



omics data with a novel target prioritization scoring system in an interactive Python-based Streamlit dashboard. StarGazer analyzes and integrates disease-target associations, druggability data, and protein-protein interaction data before extracting five features from the data to create an overall StarGazer score for every potential target associated with StarGazer's curated list of 1844 phenotypic variants.

StarGazer is adapted to facilitate exploration of the human biology landscape from a birds-eye view, allowing rapid digestion of information from PheWASs/GWASs, which otherwise contains many tens of thousands of complex multivariate datapoints. Streamlit, as a user interface package adapted for complex data visualization and user interactivity, was considered to be a well-suited technology for such a task. Indeed, the importance of the flexibility in visualization methods, and live

data retrieval and analysis is becoming increasingly clear, with their applications ever-expanding (Badgeley et al., 2016; Moosavinasab et al., 2016).

We demonstrate the utility in integrating several omics datasets and returning easy-to-interpret analysis metrics in an interactive dashboard. One can easily imagine the power of such a strategy as we incorporate state-of-the-art, machine learning-based, multi-omics integration techniques, as well as a wider variety of high quality data. In an era where the speed at which we can generate data is accelerating at a higher rate than we can analyze it, we anticipate that integrative scores and visualization tools will grow increasingly essential in biology, and that we must begin to break away from the more rigid, single-use analysis framework that forms the modern paradigm for analyzing not just GWAS and PheWAS data (Diogo et al., 2018; Ferrero and

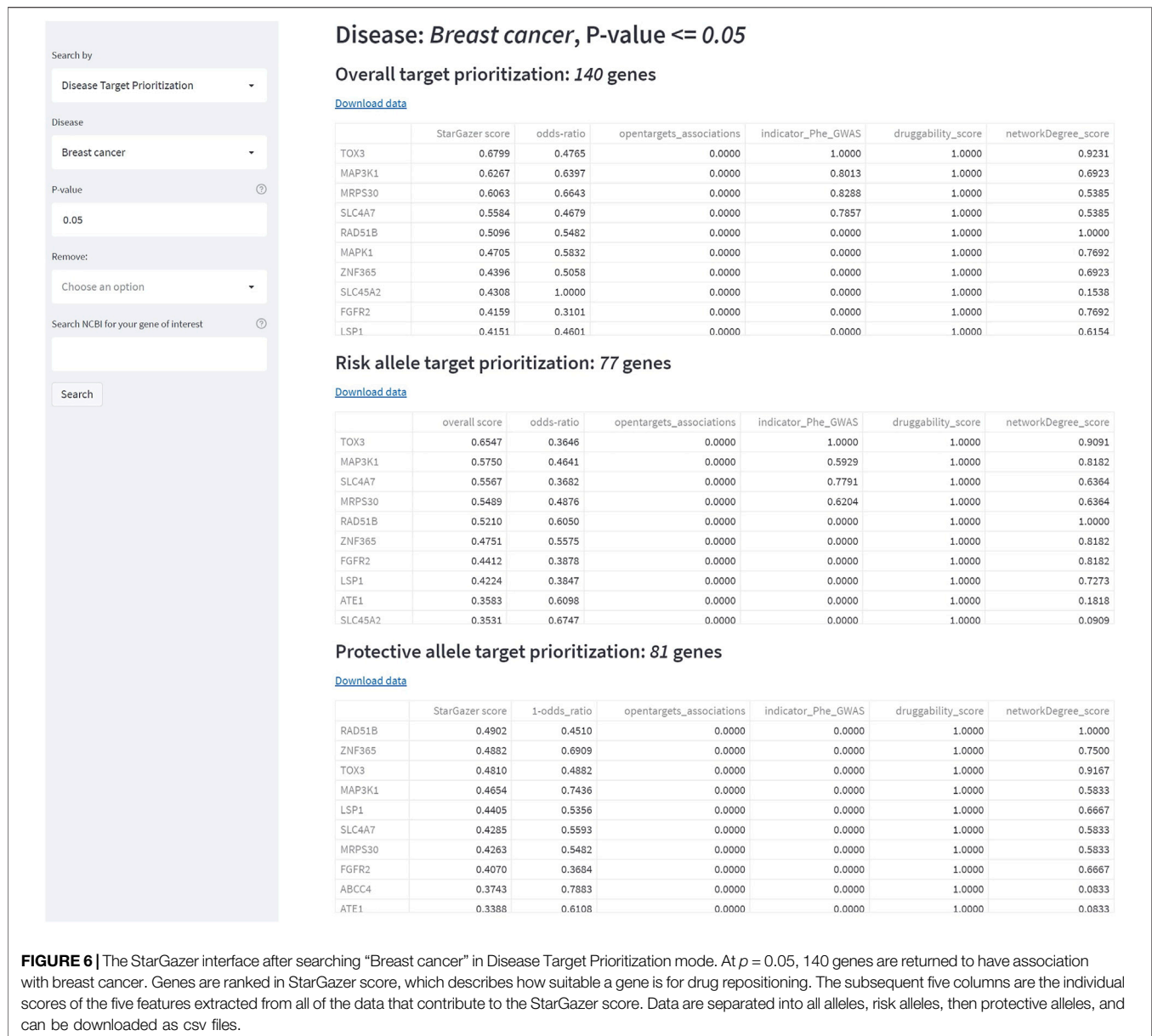


FIGURE 6 | The StarGazer interface after searching “Breast cancer” in Disease Target Prioritization mode. At $p = 0.05$, 140 genes are returned to have association with breast cancer. Genes are ranked in StarGazer score, which describes how suitable a gene is for drug repositioning. The subsequent five columns are the individual scores of the five features extracted from all of the data that contribute to the StarGazer score. Data are separated into all alleles, risk alleles, then protective alleles, and can be downloaded as csv files.

Agarwal, 2018; Robinson et al., 2018; Lau and So, 2020), but multi-omics data in general (Subramanian et al., 2020).

StarGazer has been built with the goal of pushing multi-omics integration towards upward scalability by providing users with immediate access to contextual information on genes of potential interest by automatically performing several steps of follow-up analysis on all genes - this saves a considerable amount of time from performing speculative follow-up analysis. These follow-up analysis steps are completed in bulk through the processing of the single-omic layers, which removes the need for users to analyze every gene separately for various properties and then later compare the results to make sense of the evidence. Not only does integrating single-omic layers increase the speed of exploratory data analysis, but it also provides additional value from combining multiple pieces of evidence as opposed to

focusing on individual single high-confidence pieces of information, especially when the different types of data are likely to have an intimate biological relationship, e.g., combining a gene’s DNA, RNA and protein information together is likely to be more valuable than analyzing them independently as they are functionally coupled. This approach may be our best strategy for uncovering complex and profound relationships and hence, the phrase “the whole is greater than the sum of its parts” holds particularly true in the context of multi-omics data analysis. A more integrated strategy may also be more useful in helping us understand the genetic basis of complex diseases driven by genes and gene variants with pleiotropic functions or effects. Applying the latest ideas on pleiotropy in biological systems to future work may allow us to obtain a more complete understanding of genome-phenome relationships and

TABLE 1 | Top 30 hits from Disease Target Prioritization mode analysis of “Insomnia” using StarGazer.

Gene Name	StarGazer Score	Odds-Ratio	OpenTargets Associations	Indicator Phe/GWAS	Druggability Score	Network Degree Score
HLA-DRB1	0.456	0.725	0.000	0.000	1.000	0.556
ESR1	0.433	0.655	0.009	0.000	0.500	1.000
GRIN2B	0.407	0.756	0.000	0.000	0.500	0.778
MEIS1	0.395	0.251	1.000	0.000	0.500	0.222
MAOA	0.344	0.888	0.000	0.000	0.500	0.333
DNM1	0.337	0.630	0.000	0.000	0.500	0.556
HLA-DQB1	0.320	0.653	0.000	0.000	0.500	0.444
BMP4	0.307	0.591	0.000	0.000	0.500	0.444
ATM	0.293	0.188	0.000	0.000	0.500	0.778
CMTM7	0.288	0.941	0.000	0.000	0.500	0.000
NKAPL	0.288	0.938	0.000	0.000	0.500	0.000
GRIA1	0.286	0.263	0.000	0.000	0.500	0.667
TOMM40	0.280	0.677	0.000	0.000	0.500	0.222
NR5A2	0.278	0.668	0.000	0.000	0.500	0.222
HDAC9	0.276	0.768	0.000	0.000	0.500	0.111
MS4A6A	0.271	0.631	0.000	0.000	0.500	0.222
DISC1	0.267	0.166	0.000	0.000	0.500	0.667
ST6GAL1	0.265	0.716	0.000	0.000	0.500	0.111
SLC22A3	0.264	0.600	0.000	0.000	0.500	0.222
EFNA5	0.264	0.600	0.000	0.000	0.500	0.222
NRGN	0.263	0.706	0.000	0.000	0.500	0.111
DRD2	0.263	0.000	0.150	0.000	0.500	0.667
RNASET2	0.263	0.591	0.000	0.000	0.500	0.222
FGFR2	0.263	0.257	0.000	0.000	0.500	0.556
UBE2L3	0.262	0.701	0.000	0.000	0.500	0.111
YDJC	0.260	0.690	0.000	0.000	0.500	0.111
CDC42BPB	0.260	0.690	0.000	0.000	0.500	0.111
LAMP3	0.258	0.791	0.000	0.000	0.500	0.000
ARG1	0.254	0.660	0.000	0.000	0.500	0.111
CCND3	0.251	0.643	0.000	0.000	0.500	0.111

thus drive novel discoveries previously inaccessible in the biomedical field (Shameer et al., 2021).

Limitations

This should, of course, highlight to the reader the current co-dependence between broader exploratory analytical approaches, such as StarGazer, with those that possess stronger statistical power, aimed at target confirmation at the cost of breadth and fewer omics layers, and of course, experimental confirmation. Moving forwards, we should hope that the field develops more sophisticated strategies for these types of analysis. All in all, we anticipate StarGazer to be potentially useful in providing insights into many types of biological pathways, as long as the molecular perturbations that are linked with disease lie close to the genetic level. Whilst it is easy to imagine StarGazer’s utility for studying diseases caused by variants of proteins or nucleic acids due to their more direct connection to genome-level information, studying metabolic disorders of carbohydrates and lipids would be possible but more difficult.

We wish to highlight that, although the barrier to entry for multi-omics data analysis is low, there seems yet a limitless space for improvement in the field at the time of writing. In the future, we aim to incorporate gene ontology terms enrichment analysis, gene semantic similarity, and gene expression data into our target prioritization framework, and improve on the implementation of protein-protein interaction networks (Shameer et al., 2016; Peters

et al., 2017). Whilst the current version of StarGazer extracts several features for target-disease associations, the assessment of target druggability uses only one dataset to generate one feature. Although the knowledge-based classification of the genome that Pharos provides is very high quality data, it is less indicative of future potential developments as it reflects only the current status of the druggability landscape of human biology. Therefore, more predictive datasets, such as computational docking predictions using structural data from molecular techniques or even AI-based computational prediction, may provide more robust insight into the future (Baek et al., 2021; Jumper et al., 2021).

StarGazer’s use of API calls allows for the majority of its data to be updated automatically with the latest relevant studies, aside from the PheWAS catalog which was performed in 2013—it would be invaluable if a similar study was repeated to include the GWAS data which was generated during the decade that has elapsed since the original effort. Furthermore, a variety of machine learning strategies have been applied to multi-omics data analysis and show great promise in assisting precision medicine and repositioning (Shameer et al., 2018b; Nicora et al., 2020; Reel et al., 2021), and is therefore an area we are interested in developing for StarGazer. Another avenue for future development is to improve on the standardization of clinical terms between the different datasets, which is a problem not unique to StarGazer but found ubiquitously in healthcare-related work (Wears, 2015; Beck et al., 2019). This problem manifested itself as data from OpenTargets

being underrepresented in the overall StarGazer score. We hypothesize that using a combination of standardized codes for clinical terms, such as ICD-9/-10 (<https://www.cdc.gov/nchs/icd/icd9.htm>, <https://www.cdc.gov/nchs/icd/icd10.htm>), and EFO (<https://www.ebi.ac.uk/efo/faq.html>), would help with this problem, as well as further curate our list of 1844 phenotypic variants. Currently, the code for installation can be found on GitHub (<https://github.com/AstraZeneca/StarGazer>).

CONCLUSION

We have created StarGazer (<https://github.com/AstraZeneca/StarGazer>), an interactive dashboard that facilitates rapid investigation of potential novel drug targets and repositioning strategies. It integrates three different types of data (disease-target data, target druggability data, and protein-protein interaction data) from four different knowledgebases (the PheWAS catalog, OpenTargets, Pharos, and STRING) to extract five features that are then processed to return a singular normalized “StarGazer” score. All genes with associations with any of the 1844 phenotypic variants in the StarGazer disease list are then ranked in suitability for drug repositioning strategies for the disease of interest.

We demonstrate the utility in integrating several omics datasets to return easy-to-interpret analysis metrics in an interactive dashboard. One can easily imagine the power of such a strategy as we incorporate machine learning techniques as well as a wider variety of high quality data. It is anticipated that such integrative analysis strategies will become commonplace as biomedical data science grows to explore more multi-disciplinary and multi-omic datasets. Integrative scores and visualization

REFERENCES

- Abu-Doleh, A. A., Al-Jarrah, O. M., and Alkhateeb, A. (2012). Protein Contact Map Prediction Using Multi-Stage Hybrid Intelligence Inference Systems. *J. Biomed. Inf.* 45 (1), 173–183. doi:10.1016/j.jbi.2011.10.008
- Adikusuma, W., Irham, L. M., Chou, W.-H., Wong, H. S.-C., Mugiyanto, E., Ting, J., et al. (2021). Drug Repurposing for Atopic Dermatitis by Integration of Gene Networking and Genomic Information. *Front. Immunol.* 12, 724277. doi:10.3389/fimmu.2021.724277
- Akata, Z., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., et al. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* 53, 18–28. doi:10.1109/mc.2020.2996587
- Armstrong, J. F., Faccenda, E., Harding, S. D., Pawson, A. J., Southan, C., Sharman, J. L., et al. (2019). The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: Extending Immunopharmacology Content and Introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* 48, D1006–D1021. doi:10.1093/nar/gkz951
- Ashburn, T. T., and Thor, K. B. (2004). Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discov.* 3 (8), 673–683. doi:10.1038/nrd1468
- Attique, Z., Ali, A., Hamza, M., al-Ghanim, K. A., Mehmood, A., Khan, S., et al. (2021). In-silico Network-Based Analysis of Drugs Used against COVID-19: Human Well-Being Study. *Saudi J. Biol. Sci.* 28 (3), 2029–2039. doi:10.1016/j.sjbs.2021.01.006
- Avram, S., Bologna, C. G., Holmes, J., Bocci, G., Wilson, T. B., Nguyen, D.-T., et al. (2021). DrugCentral 2021 Supports Drug Discovery and Repositioning. *Nucleic Acids Res.* 49 (D1), D1160–D1169. doi:10.1093/nar/gkaa997

tools for high dimensional data will become essential as we navigate science in this era where we are generating data at a such an enormous pace, thus we have positioned StarGazer to push multi-omics integration towards upward scalability.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/AstraZeneca/StarGazer>.

AUTHOR CONTRIBUTIONS

CL and JL developed the project with critical inputs from VG, AF, MH, KT, BS, TC, JM, and YG. MH led app hosting, and AP and RNH conducted validation analysis. KS led and oversaw the project. The original manuscript was written by CL and JL, with additions and edits provided by VG, EP, AF, SP, WY, A-SS, KT, BS, TC, JM, FMK, YG, and KS. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to acknowledge Anshul Kanakia for insightful discussions, and Stefano Borini for assistance with GitHub.

- Badgeley, M. A., Shameer, K., Glicksberg, B. S., Tomlinson, M. S., Levin, M. A., McCormick, P. J., et al. (2016). EHDViz: Clinical Dashboard Development Using Open-Source Technologies. *BMJ Open* 6 (3), e010579. doi:10.1136/bmjopen-2015-010579
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* 373 (6557), 871–876. doi:10.1126/science.abj8754
- Barros, R. P. A., and Gustafsson, J.-Å. (2011). Estrogen Receptors and the Metabolic Network. *Cell Metab.* 14 (3), 289–299. doi:10.1016/j.cmet.2011.08.005
- Beck, T., Shorter, T., and Brookes, A. J. (2019). GWAS Central: a Comprehensive Resource for the Discovery and Comparison of Genotype and Phenotype Data from Genome-wide Association Studies. *Nucleic Acids Res.* 48, D933–D940. doi:10.1093/nar/gkz895
- Brunner, H. G. (2007). “MAOA Deficiency and Abnormal Behaviour: Perspectives on an Association,” in *Novartis Foundation Symposia [Internet]*. Editors G. R. Bock and J. A. Goode (Chichester, UK: John Wiley & Sons), 155–167. Available at: <https://onlinelibrary.wiley.com/doi/10.1002/9780470514825.ch9>. doi:10.1002/9780470514825.ch9
- Carroll, J. E., Cole, S. W., Seeman, T. E., Breen, E. C., Witarama, T., Arevalo, J. M. G., et al. (2016). Partial Sleep Deprivation Activates the DNA Damage Response (DDR) and the Senescence-Associated Secretory Phenotype (SASP) in Aged Adult Humans. *Brain, Behav. Immun.* 51, 223–229. doi:10.1016/j.bbi.2015.08.024
- Choo, S. Y. (2007). The HLA System: Genetics, Immunology, Clinical Testing, and Clinical Implications. *Yonsei Med. J.* 48 (1), 11. doi:10.3349/ymj.2007.48.1.11
- Chung, C., Barylko, B., Leitz, J., Liu, X., and Kavalali, E. T. (2010). Acute Dynamin Inhibition Dissects Synaptic Vesicle Recycling Pathways that Drive

- Spontaneous and Evoked Neurotransmission. *J. Neurosci.* 30 (4), 1363–1376. doi:10.1523/jneurosci.3427-09.2010
- Dahoun, T., Trossbach, S. V., Brandon, N. J., Korth, C., and Howes, O. D. (2017). The Impact of Disrupted-In-Schizophrenia 1 (DISC1) on the Dopaminergic System: a Systematic Review. *Transl. Psychiatry* 7 (1), e1015. doi:10.1510.1038/tp.2016.282
- Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., et al. (2013). Systematic Comparison of Phenome-wide Association Study of Electronic Medical Record Data and Genome-wide Association Study Data. *Nat. Biotechnol.* 31 (12), 1102–1111. doi:10.1038/nbt.2749
- Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., et al. (2010). PheWAS: Demonstrating the Feasibility of a Phenome-wide Scan to Discover Gene-Disease Associations. *Bioinformatics* 26 (9), 1205–1210. doi:10.1093/bioinformatics/btq126
- Diogo, D., Tian, C., Franklin, C. S., Alanne-Kinnunen, M., March, M., Spencer, C. C. A., et al. (2018). Phenome-wide Association Studies across Large Population Cohorts Support Drug Target Validation. *Nat. Commun.* 9 (1), 4285. doi:10.1038/s41467-018-06540-3
- Ferrero, E., and Agarwal, P. (2018). Connecting Genetics and Gene Expression Data for Target Prioritisation and Drug Repositioning. *BioData Min.* 11 (1), 7. doi:10.1186/s13040-018-0171-y
- Gallo, K., Goede, A., Eckert, A., Moahamed, B., Preissner, R., and Gohlke, B.-O. (2021). PROMISCUOUS 2.0: a Resource for Drug-Repositioning. *Nucleic Acids Res.* 49 (D1), D1373–D1380. doi:10.1093/nar/gkaa1061
- Gennari, L., Merlotti, D., Valleggi, F., Martini, G., and Nuti, R. (2007). Selective Estrogen Receptor Modulators for Postmenopausal Osteoporosis. *Drugs & Aging* 24 (5), 361–379. doi:10.2165/00002512-200724050-00002
- Ghousaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., et al. (2021). Open Targets Genetics: Systematic Identification of Trait-Associated Genes Using Large-Scale Genetics and Functional Genomics. *Nucleic Acids Res.* 49 (D1), D1311–D1320. doi:10.1093/nar/gkaa840
- Glicksberg, B. S., Li, L., Cheng, W.-Y., Shameer, K., Hakenberg, J., Castellanos, R., et al. (2014). “An Integrative Pipeline for Multi-Modal Discovery of Disease Relationships,” in *Biocomputing 2015 [Internet]* (Hawaii, USA: World Scientific), 407–418. Available from: http://www.worldscientific.com/doi/abs/10.1142/9789814644730_0039.
- Guo, Z., Shen, Y., Wan, S., Shang, W., and Yu, K. (2021). Hybrid Intelligence-Driven Medical Image Recognition for Remote Patient Diagnosis in Internet of Medical Things. *IEEE J. Biomed. Health Inf.* doi:10.1109/jbhi.2021.3139541
- Hermawan, A., Putri, H., and Utomo, R. Y. (2020). Functional Network Analysis Reveals Potential Repurposing of β -blocker Atenolol for Pancreatic Cancer Therapy. *DARU J. Pharm. Sci.* 28 (2), 685–699. doi:10.1007/s40199-020-00375-4
- Hodos, R. A., Kidd, B. A., Shameer, K., Readhead, B. P., and Dudley, J. T. (2016). In Silico methods for Drug Repurposing and Pharmacology. *WIREs Mech. Dis.* 8 (3), 186–210. doi:10.1002/wsbm.1337
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Khaladkar, M., Koscielny, G., Hasan, S., Agarwal, P., Dunham, I., Rajpal, D., et al. (2017). Uncovering Novel Repositioning Opportunities Using the Open Targets Platform. *Drug Discov. Today* 22 (12), 1800–1807. doi:10.1016/j.drudis.2017.09.007
- Khosravi, A., Jayaram, B., Goliaei, B., and Masoudi-Nejad, A. (2019). Active Repurposing of Drug Candidates for Melanoma Based on GWAS, PheWAS and a Wide Range of Omics Data. *Mol. Med.* 25 (1), 30. doi:10.1186/s10020-019-0098-x
- Kiermer, V. (2008). Antibodypedia. *Nat. Methods* 5 (10), 860. doi:10.1038/nmeth1008-860
- Lau, A., and So, H.-C. (2020). Turning Genome-wide Association Study Findings into Opportunities for Drug Repositioning. *Comput. Struct. Biotechnol. J.* 18, 1639–1650. doi:10.1016/j.csbj.2020.06.015
- Lee, C., and Bhakta, S. (2021). The Prospect of Repurposing Immunomodulatory Drugs for Adjunctive Chemotherapy against Tuberculosis: A Critical Review. *Antibiotics* 10 (1), 91. doi:10.3390/antibiotics10010091
- Li, H., Li, J., Su, Y., Fan, Y., Guo, X., Li, L., et al. (2014). A Novel 3p22.3 Gene CMTM7 Represses Oncogenic EGFR Signaling and Inhibits Cancer Cell Growth. *Oncogene* 33 (24), 3109–3118. doi:10.1038/onc.2013.282
- Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., and Lu, Z. (2016). A Survey of Current Trends in Computational Drug Repositioning. *Brief. Bioinform* 17 (1), 2–12. doi:10.1093/bib/bbv020
- Li, S., Kang, L., and Zhao, X. M. (2014). A Survey on Evolutionary Algorithm Based Hybrid Intelligence in Bioinformatics. *Biomed. Res. Int.* 2014, 362738. doi:10.1155/2014/362738
- Liu, Z., Borlak, J., and Tong, W. (2014). Deciphering miRNA Transcription Factor Feed-Forward Loops to Identify Drug Repurposing Candidates for Cystic Fibrosis. *Genome Med.* 6 (12), 94. doi:10.1186/s13073-014-0094-2
- Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., et al. (2019). ChEMBL: towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* 47 (D1), D930–D940. doi:10.1093/nar/gky1075
- Moosavinasab, S., Patterson, J., Strouse, R., Rastegar-Mojarad, M., Regan, K., Payne, P. R. O., et al. (2016). ‘RE:fine Drugs’: an Interactive Dashboard to Access Drug Repurposing Opportunities. *Database* 2016, baw083. doi:10.1093/database/baw083
- Nicora, G., Vitali, F., Dagliati, A., Geifman, N., and Bellazzi, R. (2020). Integrated Multi-Omics Analyses in Oncology: A Review of Machine Learning Methods and Tools. *Front. Oncol.* 10, 1030. doi:10.3389/fonc.2020.01030
- Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Urriarte, A., Malangone, C., et al. (2021). Open Targets Platform: Supporting Systematic Drug-Target Identification and Prioritisation. *Nucleic Acids Res.* 49 (D1), D1302–D1310. doi:10.1093/nar/gkaa1027
- Okuda, H., Kiuchi, H., Takao, T., Miyagawa, Y., Tsujimura, A., Nonomura, N., et al. (2015). A Novel Transcriptional Factor Nkapl Is a Germ Cell-specific Suppressor of Notch Signaling and Is Indispensable for Spermatogenesis. *PLoS ONE* 10 (4), e0124293. doi:10.1371/journal.pone.0124293
- Peters, L. A., Perrigoue, J., Mortha, A., Luga, A., Song, W.-m., Neiman, E. M., et al. (2017). A Functional Genomics Predictive Network Model Identifies Regulators of Inflammatory Bowel Disease. *Nat. Genet.* 49 (10), 1437–1449. doi:10.1038/ng.3947
- Portelli, M. A., Rakkar, K., Hu, S., Guo, Y., and Adcock, I. M. (2021). Translational Analysis of Moderate to Severe Asthma GWAS Signals into Candidate Causal Genes and Their Functional, Tissue-dependent and Disease-Related Associations. *Front. Allergy* 2, 738741. doi:10.3389/falgy.2021.738741
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168
- Rapicavoli, R. V., Alaimo, S., Ferro, A., and Pulvirenti, A. (2022). “Computational Methods for Drug Repurposing,” in *Computational Methods for Precision Oncology [Internet]*. Editor A. Laganà (Cham: Springer International Publishing), 119–141. Available at: https://link.springer.com/10.1007/978-3-030-91836-1_7. doi:10.1007/978-3-030-91836-1_7
- Rastegar-Mojarad, M., Ye, Z., Kolesar, J. M., Hebring, S. J., and Lin, S. M. (2015). Opportunities for Drug Repositioning from Phenome-wide Association Studies. *Nat. Biotechnol.* 33 (4), 342–345. doi:10.1038/nbt.3183
- Reay, W. R., and Cairns, M. J. (2021). Advancing the Use of Genome-wide Association Studies for Drug Repurposing. *Nat. Rev. Genet.* 22 (10), 658–671. doi:10.1038/s41576-021-00387-z
- Reel, P. S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using Machine Learning Approaches for Multi-Omics Data Analysis: A Review. *Biotechnol. Adv.* 49, 107739. doi:10.1016/j.biotechadv.2021.107739
- Robinson, J. R., Denny, J. C., Roden, D. M., and Van Driest, S. L. (2018). Genome-wide and Phenome-wide Approaches to Understand Variable Drug Actions in Electronic Health Records. *Clin. Transl. Sci.* 11 (2), 112–122. doi:10.1111/cts.12522
- Sanseau, P., Agarwal, P., Barnes, M. R., Pastinen, T., Richards, J. B., Cardon, L. R., et al. (2012). Use of Genome-wide Association Studies for Drug Repositioning. *Nat. Biotechnol.* 30 (4), 317–320. doi:10.1038/nbt.2151
- Sarayloo, F., Dion, P. A., and Rouleau, G. A. (2019). MEIS1 and Restless Legs Syndrome: A Comprehensive Review. *Front. Neurol.* 10, 935. doi:10.3389/fneur.2019.00935
- Shameer, K., Glicksberg, B. S., Badgeley, M. A., Johnson, K. W., and Dudley, J. T. (2021). Pleiotropic Variability Score: A Genome Interpretation Metric to Quantify Phenomic Associations of Genomic Variants. *bioRxiv*. doi:10.1101/2021.07.18.452819
- Shameer, K., Glicksberg, B. S., Hodos, R., Johnson, K. W., Badgeley, M. A., and Readhead, B. (2018). Systematic Analyses of Drugs and Disease Indications in RepurposeDB

- Reveal Pharmacological, Biological and Epidemiological Factors Influencing Drug Repositioning. *Briefings Bioinforma.* 19 (4), 656–678. doi:10.1093/bib/bbw136
- Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., and Sengupta, P. P. (2018). Machine Learning in Cardiovascular Medicine: Are We There yet? *Heart* 104 (14), 1156–1164. doi:10.1136/heartjnl-2017-311198
- Shameer, K., Readhead, B., and T. Dudley, J. (2015). Computational and Experimental Advances in Drug Repositioning for Accelerated Therapeutic Stratification. *Ctmc* 15 (1), 5–20. doi:10.2174/1568026615666150112103510
- Shameer, K., Tripathi, L. P., Kalari, K. R., Dudley, J. T., and Sowdhamini, R. (2016). Interpreting Functional Effects of Coding Variants: Challenges in Proteome-Scale Prediction, Annotation and Assessment. *Brief. Bioinform* 17 (5), 841–862. doi:10.1093/bib/bbv084
- Sheils, T. K., Mathias, S. L., Kelleher, K. J., Siramshetty, V. B., Nguyen, D.-T., Bologa, C. G., et al. (2021). TCRD and Pharos 2021: Mining the Human Proteome for Disease Biology. *Nucleic Acids Res.* 49 (D1), D1334–D1346. doi:10.1093/nar/gkaa993
- Shiloh, Y., and Ziv, Y. (2013). The ATM Protein Kinase: Regulating the Cellular Response to Genotoxic Stress, and More. *Nat. Rev. Mol. Cell Biol.* 14 (4), 197–210. doi:10.1038/nrm3546
- Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and its Application. *Bioinform Biol. Insights* 14, 117793221989905. doi:10.1177/1177932219899051
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- Tan, X., Gong, L., Li, X., Zhang, X., Sun, J., Luo, X., et al. (2021). Promethazine Inhibits Proliferation and Promotes Apoptosis in Colorectal Cancer Cells by Suppressing the PI3K/AKT Pathway. *Biomed. Pharmacother.* 143, 112174. doi:10.1016/j.biopha.2021.112174
- Varghese, R., and Majumdar, A. (2022). A New Prospect for the Treatment of Nephrotic Syndrome Based on Network Pharmacology Analysis. *Curr. Res. Physiology* 5, 36–47. doi:10.1016/j.crphys.2021.12.004
- Wears, R. L. (2015). Standardisation and its Discontents. *Cogn. Tech. Work* 17 (1), 89–94. doi:10.1007/s10111-014-0299-6
- Weissler, E. H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., et al. (2021). The Role of Machine Learning in Clinical Research: Transforming the Future of Evidence Generation. *Trials* 22 (1), 537. doi:10.1186/s13063-021-05489-x
- Wijetunga, I., McVeigh, L. E., Charalambous, A., Antanaviciute, A., Carr, I. M., Nair, A., et al. (2020). Translating Biomarkers of Cholangiocarcinoma for Theranosis: A Systematic Review. *Cancers* 12 (10), 2817. doi:10.3390/cancers12102817
- Xue, H., Li, J., Xie, H., and Wang, Y. (2018). Review of Drug Repositioning Approaches and Resources. *Int. J. Biol. Sci.* 14 (10), 1232–1244. doi:10.7150/ijbs.24612
- Zada, D., Sela, Y., Matosevich, N., Monsonogo, A., Lerer-Goldshtein, T., Nir, Y., et al. (2021). Parp1 Promotes Sleep, Which Enhances DNA Repair in Neurons. *Mol. Cell* 81 (24), 4979–4993. e7. doi:10.1016/j.molcel.2021.10.026
- Zhao, K., Shi, Y., and So, H.-C. (2022). Prediction of Drug Targets for Specific Diseases Leveraging Gene Perturbation Data: A Machine Learning Approach. *Pharmaceutics* 14 (2), 234. doi:10.3390/pharmaceutics14020234

Conflict of Interest: Authors CL, AP, VG, RNH, EP, AF, SP, WY, MH, A-SS, KT, BS, TC, JM, FMK, and KS are or were employed by AstraZeneca.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lee, Lin, Prokop, Gopalakrishnan, Hanna, Papa, Freeman, Patel, Yu, Huhn, Sheikh, Tan, Sellman, Cohen, Mangion, Khan, Gusev and Shameer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.