Journal of Machine Learning Research 1–15

DeepVul: A Multi-Task Transformer Model for Joint Prediction of Gene Essentiality and Drug Response

Ala Jararweh[1,2], David Arredondo[1], Oladimeji Macaulay[1], Mikaela Dicome[1,3], Luis Tafoya[1], Yue Hu[1], Kushal Virupakshappa[1], Genevieve Boland[4], Keith Flaherty[4], Avinash Sahu[1,2]

AJARARWEH@UNM.EDU

ASAHU@SALUD.UNM.EDU

[1]*Comprehensive Cancer Center, The University of New Mexico,* [2]*Department of Computer Science, The University of New Mexico,* [3]*Department of Biomedical Engineering, The University of New Mexico,* [4]*Department of Medicine, Massachusetts General Hospital Cancer Center, Harvard Medical School*

Abstract

Despite their potential, current precision oncology approaches benefit only a small fraction of patients due to their limited focus on actionable genomic alterations. To expand its applicability, we propose DeepVul, a multi-task transformer-based model designed to predict gene essentiality and drug response from cancer transcriptome data. DeepVul aligns gene expressions, gene perturbations, and drug perturbations into a latent space, enabling simultaneous and accurate prediction of cancer cell vulnerabilities to numerous genes and drugs. Benchmarking against existing precision oncology approaches revealed that DeepVul not only matches but also complements oncogene-defined precision methods. Through interpretability analyses, DeepVul identifies underlying mechanisms of treatment response and resistance, as demonstrated with BRAF vulnerability prediction. By leveraging whole-genome transcriptome data, DeepVul enhances the clinical actionability of precision oncology, aiding in the identification of optimal treatments across a broader range of cancer patients. DeepVul is publicly available at https://github.com/alaaj27/DeepVul.git.

1. Introduction

Precision oncology seeks to match patients with the most effective drugs. Current approaches benefit only 10% of cancer patients due to their narrow focus on a small number of actionable genomic alterations [14, 2, 20]. Consequently, over 90% of cancer patients lacking actionable alterations in their tumors are ineligible for these therapies. To expand the coverage of precision oncology beyond this 10%, "avatar models" have been proposed. These models involve creating cell lines, xenografts, organoids, or ex vivo systems from tumor biospecimens from a patient, which can then be tested with numerous drugs or gene knockouts to find the most optimal treatment [8, 13]. The time required to develop these experimental models impedes their clinical adoption and practical application. To address this, researchers are now mapping targetable cancer vulnerabilities in a large number of existing cell lines and mouse xenografts, aiming to translate insights gained from these experimental approaches, including drug treatments and gene essentiality screens, into patient care.

© A. Jararweh *et al.*

Gene essentiality in cancer cells refers to the degree to which the survival and proliferation of the cancer cell depends on the function of a specific gene [31, 11, 23]. It is quantified by the change in growth rate resulting from the knockout or knockdown of that gene, indicating how critical the gene is to the cell's viability. Mapping essential genes reveals specific cancer cell vulnerabilities, providing potential therapeutic targets. By identifying selective gene knockouts that can kill specific tumor cells without affecting normal cells, researchers can design drugs targeting these genes, thereby increasing the coverage of precision oncology beyond the few currently known actionable genes. Experimental approaches such as CRISPR knockout screens reveal essential genes on a genome-wide scale [11, 30]. For instance, the DepMap project has mapped essential genes in thousands of cancer cell lines. [28, 6].

Complementing experimental approaches, computational approaches have been proposed for predicting gene essentiality. This includes traditional machine learning approaches like Support Vector Machines (SVM) using features from nucleotide sequence data and protein-protein interaction (PPI) networks, which require manual feature selection and domain knowledge [9, 34, 32]. Deep learning approaches have also been adopted for essentiality prediction because of their ability to perform feature extraction and engineering without human intervention [33, 34, 12].

However, gene essentiality varies dramatically across cancer cell lines, making it highly context-specific [28]. This context-specificity raises questions about how findings in cell lines can be translated to patient tumors. The relationship between the cancer genetic landscape and its vulnerabilities is non-linear and complex, necessitating advanced modeling for accurate predictions. Current methods fail to capture this context-specificity accurately, preventing us from effectively assessing the impacts of gene knockouts on cancer cells without modeling cancer as a system.

Here, we posit that mapping gene expression to cancer vulnerability could enable the identification of optimal treatments for cancer patients from their tumor transcriptome. We propose DeepVul, a multi-task transformer-based model that aligns gene expression and essentiality into a latent space, allowing the joint prediction of drug response and gene essentiality. Our framework jointly predicts the vulnerability of cancer cells to thousands of genes and their responses to thousands of drugs simultaneously. Our interpretability framework also identifies underlying response mechanisms, as demonstrated for the BRAF oncogene. By leveraging whole-genome transcriptome data, our approach aims to expand the clinical actionability of precision oncology, identifying optimal treatments for cancer patients.

2. Related Work

There are two types of gene essentiality: universal and context-specific. Universal essential genes are vital for the survival of an organism across all contexts, external conditions, or environments, and are typically involved in critical cellular processes (Appendix C). Context-specific essential genes refer to those genes that are vital for a cell's survival in specific environmental conditions, and targeting these genes would kill specific cells without affecting others. Therefore, context-specific essential genes are of particular interest in

personalized medicine, as they may be targets for drug development, especially for diseases like cancer. In this study, we focus on context-specific essential genes.

Context-Specific Gene Essentiality DeepHE, [34], a neural network, classifies essential genes in human cancer cell lines using sequence and PPI data; however, assumes that genes are universally essential across cancer cell lines. FluxGAT, [25] is a GNN-based approach for predicting gene essentiality directly from graphical representations of flux sampling data. DeepDEP, [3], a deep learning model, predicts essential genes for cancer cell proliferation using integrative genomic profiles.

Multi-Task Learning Multi-task learning refers to the joint training of a single machine learning model on multiple tasks. To improve retinal blood vessel segmentation, a novel encoder-decoder architecture leveraging mutli-task learning has been proposed, [16]. Clustering-Constrained-Attention Multiple-Instance Learning (CLAM) is another multitask framework for whole slide image (WSI) processing, which utilizes attention-based learning for predicting multiple tasks, [18]. Another multitask architecture, DeX [26], applies a transformer-based approach to model interactions between input features using the attention mechanism. Given the interactive nature of genes, this architecture is effective for genomics applications. Our proposed method, DeepVul, draws inspiration from this architecture.

3. Methodology

We propose DeepVul, a multi-task transformer-based model that connects gene expression and perturbations into a latent distribution, inspired by DeX [26]. The paper mainly discusses the significant challenge of predicting throughput in real-time communication in extreme conditions to enable proactive network management. The proposed model comprises three major components, feature selection, feature extraction via a transformer-based framework, and multi-task predictions to predict regular throughput, traffic peaks, traffic valleys, etc. We draw an analogy of predicting multiple gene essentiality for an expression profile and predicting different throughput tasks from packet-level information from network traffic. However, we expand upon their model by integrating a few modifications tailored to the essentiality task. Our framework comprises three main components(Figure 1): feature transformation, feature extraction, and fine-tuning. During feature transformation and feature extraction, the model learns a latent space of gene expression relative to gene essentiality, and perturbation data, by utilizing a multi-task loss function that reconstructs essentiality data from expression. Next, we fine-tune the latent space by training a second multi-task objective that predicts drug responses by minimizing each drug's Minimum Squared Error (MSE). Finally, we evaluate the model's performance across different predictive tasks on unseen cohorts of essentiality data.

3.1. Data Preparation

We utilize the open-source gene essentiality, gene expression, and drug response datasets from the DepMap Achilles project [5] (Appendix C). Because the experiment measured drug response data in the same cell lines, it presents rich information for finding the latent space described below.

We build a multimodal model that integrates the three modalities to identify drug treatments by matching them to the gene expression profiles of specific tumors. From the dataset, we retained cell lines with both gene expression data, and either essentiality or drug response data available. Our study focuses primarily on highly variable and actionable genes and drugs that exhibit significant variability across cells. We utilize the Variance-stabilizing transformation (VST) method to identify the top 1000 highly variable genes and drugs from each dataset [21, 10].

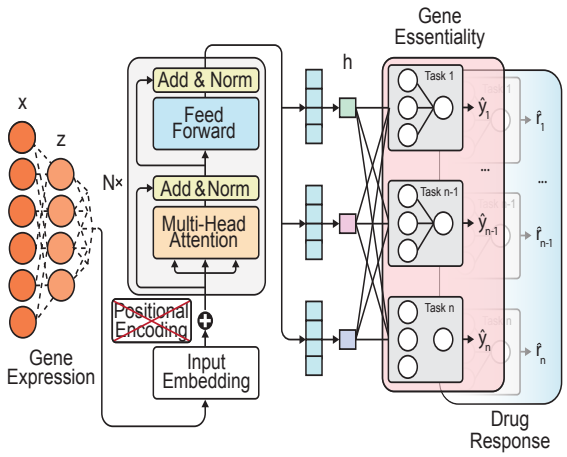


Figure 1: DeepVul architecture: a multi-task deep learning model for predicting gene essentiality and drug response from gene expression using latent space and sequential fine-tuning.

3.2. DeepVul architecture

Our goal is to map gene expression profiles of cancer cells to their essentiality and drug vulnerabilities. For this, we train an encoder to encode gene expression into a latent space and use the space to predict essentiality or drug response using decoders. The encoder is shared across all tasks, while decoders are task-specific. Below we describe the architecture and training of encoders and decoders.

Encoder: Finding a Latent Space Input: Gene expression; Output: Latent space that connects expression and essentiality. Initially, we reduce the dimensionality of the gene expression data, $\mathbf{x} \in \mathbb{R}^n$ through a series of linear transformations, resulting in $\mathbf{z} \in \mathbb{R}^d$. We then utilize a stack of transformer layers to process the transformed input, aiming to capture complex relationships among genes:

$$\text{Self-Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{1}$$

$$\mathbf{Q} = \mathbf{W}_Q\mathbf{z}, \quad \mathbf{K} = \mathbf{W}_K\mathbf{z}, \quad \mathbf{V} = \mathbf{W}_V\mathbf{z}, \tag{2}$$

where $(\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V)$ are learnable weights, and d_k is a dimension of the key vector used for scaling. The output, h, is then pooled and passed through a feedforward network:

$$\mathbf{FFN}(\mathbf{h}) = \text{ReLU}(\mathbf{W}_1\mathbf{h} + \mathbf{b}_1) \tag{3}$$

Since positional information is irrelevant in gene expression data, we eliminate positional encoding from the transformer architecture. Attention captures dependencies between genes based on their expression profiles without considering proximity.

Essentiality Decoder: Learning Multi-gene Essentiality Predictors. The decoder comprises independent pipelines (MLP layers) for each gene, allowing tunable layers specific to each gene while maintaining the same latent space (Fig 1). Each gene's pipeline operates independently, unaffected by the pipelines of other genes. The encoder and essentiality decoder are learned jointly leveraging a multi-task objective that fine-tunes the latent space and predicts the essentiality of every gene at a time as:

$$\mathcal{L}_{\text{multi-gene}} = \sum_{i=1}^{G} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \ell(y_{ij} - \hat{y}_{ij})^2 \right), \tag{4}$$

where G is the number of genes in the essentiality dataset, N_i is the number of samples for the i-th gene, and $\hat{y}_{ij} = \text{MLP}_i(Z)$ is the prediction for the i-th gene by feeding the latent space to a multi-layer perception.

Drug Decoder: Learning Multi-drug Response Predictor through Fine-tuning. The drug decoder also comprises an independent pipeline for each drug, similar to the essentiality decoder. The latent space learned by predicting gene essentiality from expression profiles is then fine-tuned to predict drug responses using multi-task objective layers as:

$$\mathcal{L}_{\text{multi-drug}} = \sum_{i=1}^{D} \left(\frac{1}{M_i} \sum_{j=1}^{M_i} \ell(r_{ij} - \hat{r}_{ij})^2 \right), \tag{5}$$

where D is the number of genes in the essentiality dataset, M_i is the number of samples for the i-th drug, and $\hat{r}_{ij} = \text{MLP}_i(Z)$ is the prediction for the i-th drug by feeding the latent space to a multi-layer perception. The drug response pipeline leverages the weights of the latent space to initially create the latent space for the multi-task drug layers. Subsequently, the weights of the latent space can be either fine-tuned or frozen while the multi-task layers are always trainable.

3.3. SHAP Analysis

SHapley **A**dditive Ex**P**lanations (SHAP) is a method of assigning contributions of inputs to a black box model's output based on game theory [19]. For a given input, the contribution of each feature sums to the difference between the model output and the model expected value. We calculate the SHAP values for all 19,193 gene expression inputs as they contribute to the model output of gene essentiality. We then use a correlation analysis [15] in which the Pearson correlation between the SHAP contribution and the expression value for a given gene is used to rank genes with the highest and lowest correlations.

4. Results

4.1. Predictions of Essential Genes in Cancer Cell Lines

We benchmarked DeepVul against multiple baselines (refer to Appendix E and Appendix D for baselines and training details). Evaluating the flattened test data by ignoring cell lines and genes demonstrates that DeepVul shows a high correlation between predicted and actual essentiality across genes and cell lines (Figure 2a). The baseline Autoencoder-NN model also performs similarly in terms of correlation (Figure 2b). However, when we calculated the correlation for each gene separately, DeepVul outperforms the baselines including Autoencoder-NN (Figure 2c). This suggests that when the essentiality prediction task was assessed one gene at a time, predicting how the essentiality of a given gene varies across cell lines, the task becomes challenging. All models including DeepVul showed zero or often negative correlations for many genes. Nonetheless, DeepVul predicted essentiality for 46% of genes with a correlation greater than 0.3.

The performance gap between DeepVul and the baselines increased when we analyzed important genes known to be actionable (Figure 2d). Calculating the correlation between predicted and actual essentiality for each cell line separately, DeepVul and Autoencoder-NN showed similar performance (Figure 2e). DeepVul matched or outperformed the baseline models, particularly in predicting the essentiality of a single gene and how it varies across cell lines.

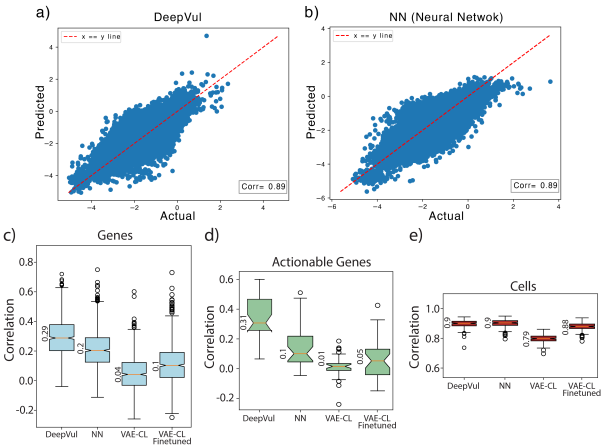


Figure 2: Evaluation of DeepVul compared to various baseline methods in predicting gene essentiality. a,b) Scatter plot shows the predicted vs. actual values of DeepVul and the neural network baseline. c,d,e) Bar charts illustrate a detailed evaluation of our model to the baselines on various prediction categories such as genes(c), actionable genes (d), and cells(e).

Generalizability to New Cohorts Model generalizability to new cohorts is a primary challenge widespread across biomedical applications of AI. We evaluated DeepVul's ability to generalize predictions to an independent cohort without additional training. Using data from the Sanger Institute, we assessed DeepVul's performance, which was initially trained on DepMap essentiality data from the Broad Institute. DeepVul's performance on the Sanger Institute data showed a similar correlation (Figure 3a,b), indicating that DeepVul's predictions could generalize to new datasets without requiring further training. In contrast, the baseline model failed to generalize effectively (Figure A.8).

We next evaluate DeepVul's transfer learning capability by assessing its ability to predict responses to multiple drugs. We test two transfer learning paradigms: one with a frozen expression encoder (referred to as freeze-shared) and another with a tunable expression encoder fine-tuned for drug response (referred to as tuned-shared). Surprisingly, the model with the frozen encoder outperforms the tunable encoder in predicting drug responses (Figure 4).

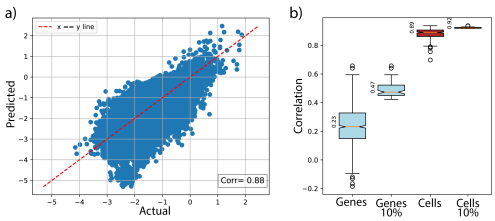


Figure 3: Evaluation of DeepVul predictions on Sanger cohort. a) Scatter plot shows the predicted vs actual gene essentiality values on the Sanger dataset. b) The boxplot illustrates the predictive performance across gene and cell lines.

4.2. Predicting Drug Response via Transfer learning

This outcome may be likely due to drugs targeting a smaller set of pathways compared to gene essentiality, which is measured genome-wide. As a result, gene essentiality might capture cancer at a systems level more comprehensively than drug response. Additionally, the limited training data available for drug response may explain why fine-tuning the parameters does

not enhance performance for drug prediction. This finding aligns with transfer learning paradigms in the imaging field, where pre-trained models like ResNet and VGG show no performance gain when fine-tuned on small datasets [4]. Fine-tuning is recommended only when large training datasets are available.

4.3. Benchmarking Against Current Precision Approaches

We benchmarked DeepVul against current actionable-mutation-based precision oncology approaches, which recommend drugs based on the presence of single actionable mutations in tumors. For instance, EGFR tyrosine kinase inhibitors (EGFR-TKIs) are recommended as the first-line treatment for patients with EGFR-mutant non-small cell lung cancer [7]. Similarly, BRAF inhibitors are the first-line treatment for patients with BRAF-mutant melanoma, particularly those with the BRAF V600E mutation [27].

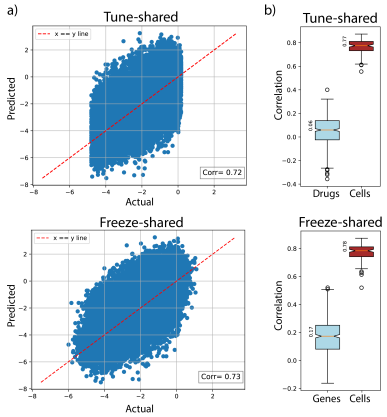


Figure 4: Comparison of drug response prediction prediction methods: Fine-tuning or Freezing the latent shared space. (a) The correlation between the actual and predicted drug response values when the latent shared space is either fine-tuned (left) or frozen (right). (b) The Drug and cell-wise correlation of the two training paradigms.

DeepVul's predictions were compared to these established precision oncology approaches. Specifically, DeepVul predicted higher essentiality of the BRAF gene in BRAF-mutant cell lines than in those without BRAF mutations (Figure 5a). The DepMap project defines an essentiality threshold of -1 to indicate sensitivity [28]. We used this threshold but for predicted essentiality to identify cell lines sensitive to BRAF targeting and compared the experimental essentiality of our predicted sensitive cell lines with BRAF-mutant cell lines. The experimental essentiality of our predicted sensitive cell lines was slightly lower than that of the BRAF-mutant cell lines (Figure 5b), although the difference was not significant due to the small number of cell lines with predicted essentiality of less than -1 in our test set. This suggests that DeepVul can identify cell lines sensitive to BRAF targeting similar to existing precision oncology.

We extended this analysis to other 32 actionable mutations, such as PIK3CA and EGFR [29]. Similar to the BRAF analysis, DeepVul predicted higher essentiality for mutant cell lines, indicating its ability to capture current oncogenic-based populations (Figure 5c). For each actionable mutation, we predicted sensitive cell lines based on the threshold of predicted essentiality of -1. DeepVul-predicted sensitive cell lines showed significantly lower experimental essentiality than those selected based on oncogenic-defined populations (Figure 5d). This suggests that DeepVul can identify target populations for current targeted therapies, complementing existing precision oncology approaches.

4.4. SHAP and LISA interpretability identify resistance mechanism

SHAP values provide the relative contribution of a given input feature. Gene expressions are often highly correlated due to co-regulation by transcription factors. To decouple these

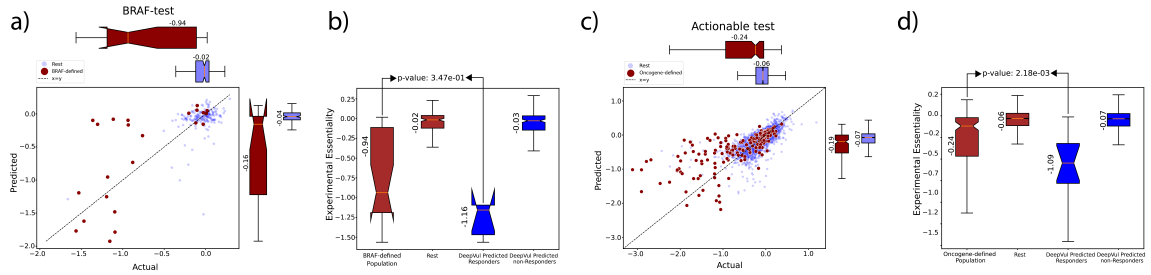


Figure 5: Comparative analysis between DeepVul and other clinical decisions regarding actionable genes. a) Predicted vs. actual essentiality values for BRAF-defined (red) and other cell lines (blue). Boxplots on the side summarize the essentiality difference between BRAF-defined vs other. b) Experimental essentiality of cell lines predicted as responders and non-responders by DeepVul and the corresponding oncogene-based clinical decisions. Wilcox P-value for the difference. c) Similar to (a), but for all actionable genes instead of just BRAF. d) Similar to (b), but for all actionable genes instead of just BRAF.

co-expression patterns, we identified genes whose SHAP values were correlated with their expression levels and identified their transcription factor regulators using LISA [22]. We applied this procedure to the essentiality of BRAF (Figure A.7).

For instance, in the case of positive correlation, a cell with higher expression levels of APOL1 is likely to depend on BRAF1 as an essential gene. The transcription factor regulators of these genes are listed in Table B.1. Among them, STAT3 is the third most likely transcription factor regulator. Inhibition of STAT3 is known to enhance sensitivity to BRAF inhibitors [35], while activation of this pathway is associated with BRAF inhibitor resistance [17]. This analysis suggests that the interpretability framework of DeepVul could identify treatment response and resistance mechanisms.

5. Conclusion

We introduced DeepVul, a multi-task transformer-based model designed to predict gene essentiality and drug response from gene expression profiles. Our approach aligns gene expressions, gene perturbations, and drug perturbations into a latent space, enabling simultaneous predictions of cancer cell vulnerabilities to numerous genes and drugs. For gene essentiality prediction, DeepVul matched or outperformed the baseline models and generalized to new datasets without requiring further training. Using BRAF as a use case, we demonstrated that DeepVul can identify cell lines sensitive to targeting hundreds of genes with similar accuracy to existing precision oncology approaches currently available for only a few actionable genes. Furthermore, its interpretability framework could identify treatment response and resistance mechanisms. Thus, DeepVul can aid in identifying target populations for current targeted therapies and complement existing precision oncology approaches.

References

[1] Christoph Bock, Paul Datlinger, Florence Chardon, Matthew A Coelho, Matthew B Dong, Keith A Lawson, Tian Lu, Laetitia Maroc, Thomas M Norman, Bicna Song, et al. High-content crispr screening. *Nature Reviews Methods Primers*, 2(1):1–23, 2022.

[2] T. H. Carr, R. McEwen, B. Dougherty, J. H. Johnson, J. R. Dry, Z. Lai, and et al. Defining actionable mutations for oncology therapeutic development. *Nature Reviews Cancer*, 16:319–329, 2016. doi: 10.1038/nrc.2016.35.

[3] Yi-Chin Chiu, Shijie Zheng, Liangjian Wang, Bradley S Iskra, Manjusha K Rao, Peter J Houghton, Yanzhi Huang, and Yidong Chen. Predicting and characterizing a cancer dependency map of tumors with deep learning. *Science Advances*, 7(34):eabh1275, 2021. doi: 10.1126/sciadv.abh1275.

[4] Ana Davila, Jacinto Colan, and Yasuhisa Hasegawa. Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image and Vision Computing*, 146:105012, June 2024. ISSN 0262-8856. doi: 10.1016/j.imavis.2024.105012. URL http://dx.doi.org/10.1016/j.imavis.2024.105012.

[5] Joshua M Dempster, Isabella Boyle, Francisca Vazquez, David E Root, Jesse S Boehm, William C Hahn, Aviad Tsherniak, and James M McFarland. Chronos: a cell population dynamics model of crispr experiments that improves inference of gene fitness effects. *Genome biology*, 22:1–23, 2021.

[6] Broad DepMap. Depmap 20q2 public. https://depmap.org/portal/, 2020. Broad Institute of MIT and Harvard.

[7] Kai Fu, Fachao Xie, Fang Wang, and Liwu Fu. Therapeutic strategies for egfr-mutated non-small cell lung cancer patients with osimertinib resistance. *Journal of Hematology & Oncology*, 15(1):173, 2022. ISSN 1756-8722. doi: 10.1186/s13045-022-01391-4. URL https://doi.org/10.1186/s13045-022-01391-4.

[8] Hu Gao, Joshua M Korn, Stefano Ferretti, John E Monahan, Yidong Wang, Meeta Singh, Cheng Zhang, Carl Schnell, Guanghui Yang, Ying Zhang, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine*, 21(11):1318–1325, 2015. doi: 10.1038/nm.3954. URL https://www.nature.com/articles/nm.3954.

[9] Feng-Biao Guo, Chuan Dong, Hong-Li Hua, Shuo Liu, Hao Luo, Hong-Wan Zhang, Yan-Ting Jin, and Kai-Yue Zhang. Accurate prediction of human essential genes using only nucleotide composition and association information. *Bioinformatics*, 33(12):1758–1764, 2017.

[10] Yuhan Hao, Tim Stuart, Madeline H Kowalski, Saket Choudhary, Paul Hoffman, Austin Hartman, Avi Srivastava, Gesmira Molla, Shaista Madad, Carlos Fernandez-Granda, and Rahul Satija. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 2023. doi: 10.1038/s41587-023-01767-y. URL https://doi.org/10.1038/s41587-023-01767-y.

[11] Thomas Hart, Anthony HY Tong, Kah Wai Chan, Jos Van Leeuwen, Anantharaman Seetharaman, Michael Aregger, Mihir Chandrashekhar, Nicole Hustedt, Saurabh Seth, Adam Noonan, Aisha Habsid, Olga Sizova, Lyudmila Nedyalkova, Rachel Climie, Luke Tworzyanski, Kimberly Lawson, Marco Alessandro Sartori, Sepideh Alibeh, David Tieu, Shaffat Masud, Pauliina Mero, Alexander Weiss, Kevin R Brown, Matej Usaj, Markus Billmann, Muhammad Rahman, Michael Constanzo, Chad L Myers, Brenda J Andrews, Charles Boone, Daniel Durocher, and Jason Moffat. Evaluation and design of genome-wide crispr/spcas9 knockout screens. *G3: Genes, Genomes, Genetics*, 7(8): 2719–2727, 2017. doi: 10.1534/g3.117.041277. URL https://www.g3journal.org/content/7/8/2719.

[12] Md Abid Hasan and Stefano Lonardi. Deeplyessential: a deep neural network for predicting essential genes in microbes. *BMC bioinformatics*, 21:1–19, 2020.

[13] Manuel Hidalgo, Frederic Amant, Andrew V Biankin, Eva Budinska, Aideen T Byrne, Carlos Caldas, Paul A Clarke, Steven de Jong, Jos Jonkers, Gunhild M Mælandsmo, et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discovery*, 4(9):998–1013, 2014. doi: 10.1158/2159-8290.CD-14-0001. URL https://cancerdiscovery.aacrjournals.org/content/4/9/998.

[14] Richard Hodson. Precision oncology. *Nature Publishing Group UK*, 585:S1, 2020. doi: 10.1038/d41586-020-02673-y.

[15] Joseph D. Janizek, Ayse B. Dincer, Safiye Celik, Hugh Chen, William Chen, Kamila Naxerova, and Su-In Lee. Uncovering expression signatures of synergistic drug responses via ensembles of explainable machine-learning models. *Nature Biomedical Engineering*, 7(6):811–829, June 2023. ISSN 2157-846X. doi: 10.1038/s41551-023-01034-0.

[16] Ge Jin, Xu Chen, and Long Ying. Deep multi-task learning for an autoencoder-regularized semantic segmentation of fundus retina images. *Mathematics*, 10(24):4798, 2022.

[17] J. Limberg, C.E. Egan, K.D. Gray, M. Singh, Z. Loewenstein, Y. Yang, M.C. Riascos, H. Al Asadi, P. Safe, S. El Eshaky, H. Liang, T.M. Ullmann, W. Wang, W. Li, T. Zhang, J. Xiang, D. Stefanova, M.M. Jin, R. Zarnegar, T.J. Fahey, and I.M. Min. Activation of the jak/stat pathway leads to braf inhibitor resistance in brafv600e positive thyroid carcinoma. *Mol Cancer Res*, 21(5):397–410, 2023. doi: 10.1158/1541-7786.MCR-21-0832.

[18] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

[19] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[20] J. Marquart, E. Y. Chen, and V. Prasad. Estimation of the percentage of us patients with cancer who benefit from genome-driven oncology. *JAMA Oncology*, 4:1093–1098, 2018. doi: 10.1001/jamaoncol.2018.1660.

[21] Alex McBratney. Everitt, B.S., 2002. The Cambridge Dictionary of Statistics. 2nd edition. Cambridge University Press, Cambridge, UK. 410 pp. ISBN 0 521 81099. AU$75. *Geoderma*, 121(1-2):155–156, July 2004. doi: 10.1016/j.geoderma.2003.11.001.

[22] Qian Qin, Jingyu Fan, Rongbin Zheng, Changxin Wan, Shenglin Mei, Qiu Wu, Hanfei Sun, Myles Brown, Jing Zhang, Clifford A. Meyer, and X. Shirley Liu. Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and chip-seq data. *Genome Biology*, 21(1):32, 2020. ISSN 1474-760X. doi: 10.1186/s13059-020-1934-6. URL https://doi.org/10.1186/s13059-020-1934-6.

[23] Jonathan Rosenski, Sigal Shifman, and Tomer Kaplan. Predicting gene knockout effects from expression data. *BMC Medical Genomics*, 16(1):26, 2023. doi: 10.1186/s12920-023-01446-6. URL https://doi.org/10.1186/s12920-023-01446-6.

[24] João Schapke, Anderson Tavares, and Mariana Recamonde-Mendoza. Epgat: gene essentiality prediction with graph attention networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1615–1626, 2021.

[25] Kieren Sharma, Lucia Marucci, and Zahraa S Abdallah. Fluxgat: Integrating flux sampling with graph neural networks for unbiased gene essentiality classification. *arXiv preprint arXiv:2403.18666*, 2024.

[26] Tailai Song, Paolo Garza, Michela Meo, and Maurizio M. Munafò. Dex: Deep learning-based throughput prediction for real-time communications with emphasis on traffic extremes. *Computer Networks*, 249:110507, 2024. ISSN 1389-1286. doi: https://doi.org/10.1016/j.comnet.2024.110507. URL https://www.sciencedirect.com/science/article/pii/S1389128624003396.

[27] Chen Sun, Sofia España, Nina Richarz, Carme Solé-Blanch, Aram Boada, Anna Martinez-Cardús, Alan Chu, Zongwen Liu, and Jose Luis Manzano. Targeted therapy or immunotherapy in braf-mutated metastatic melanoma: a spanish center's decade of experience. *Frontiers in Oncology*, 14, 2024. ISSN 2234-943X. doi: 10.3389/fonc.2024.1322116. URL https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1322116.

[28] Aviad Tsherniak, Francisca Vazquez, Patrick G Montgomery, Barbara A Weir, Gregory Kryukov, Glenn S Cowley, Stephanie Gill, William F Harrington, Stephanie Pantel, Jenna M Krill-Burger, Robin M Meyers, Laetitia Ali, Andrew Goodale, Youngsoo Lee, Gentry Jiang, Jenny Hsiao, William FJ Gerath, Sean Howell, Emily Merkel, Mahmoud Ghandi, Levi A Garraway, David E Root, Todd R Golub, Jesse S Boehm, and William C Hahn. Defining a cancer dependency map. *Cell*, 170(3):564–576, 2017. doi: 10.1016/j.cell.2017.06.010.

[29] Michael R. Waarts, Aaron J. Stonestrom, Young C. Park, and Ross L. Levine. Targeting mutations in cancer. *Journal of Clinical Investigation*, 132(8):e154943, 2022. doi: 10.1172/JCI154943. URL https://doi.org/10.1172/JCI154943.

[30] Tim Wang, John J Wei, David M Sabatini, and Eric S Lander. Genetic screens in human cells using the crispr-cas9 system. *Science*, 343(6166):80–84, 2014. doi: 10.1126/science.1246981. URL https://www.science.org/doi/10.1126/science.1246981.

[31] Tim Wang, Kivanc Birsoy, Nathan W Hughes, Kristopher M Krupczak, Rebecca J Postel, Julissa Jessurun, Pallavi Hingorani, and David M Sabatini. Identification and characterization of essential genes in the human genome. *Science*, 350(6264):1096–1101, 2015. doi: 10.1126/science.aac7041. URL https://www.science.org/doi/10.1126/science.aac7041.

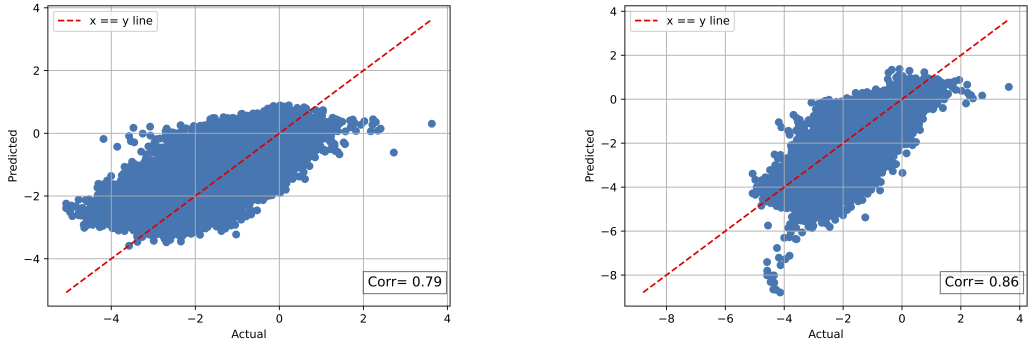
[32] Weiwei Xu, Yuting Dong, Jing Guan, Peng Liu, and Anjie Li. Identifying essential proteins from protein–protein interaction networks based on influence maximization. *BMC Bioinformatics*, 23(Suppl 8):339, 2022. doi: 10.1186/s12859-022-04874-w. URL https://doi.org/10.1186/s12859-022-04874-w.

[33] Min Zeng, Min Li, Fang-Xiang Wu, Yaohang Li, and Yi Pan. Deepep: a deep learning framework for identifying essential proteins. *BMC bioinformatics*, 20:1–10, 2019.

[34] Xue Zhang, Weilin Xiao, and Wen Xiao. Deephe: Accurately predicting human essential genes based on deep learning. *PLOS Computational Biology*, 16(9):e1008229, 2020. doi: 10.1371/journal.pcbi.1008229.

[35] Kun Zhao, Qi Zhao, Xinzhi Dai, Xue Wen, Xing Luo, Yi Duan, Zhihui Yang, and Qiong Dai. Alantolactone enhances the sensitivity of melanoma to mapk pathway inhibitors by targeting inhibition of stat3 activation and down-regulating stem cell markers. *Cancer Cell International*, 24(1):191, 2024. doi: 10.1186/s12935-024-03371-9. URL https://doi.org/10.1186/s12935-024-03371-9.

Appendix A. Figures



(a) VAE-CL (b) VAE-CL fine-tuned

Figure A.6: The prediction and actual essentiality values of the baseline VAE-CL before and after fine-tuning.

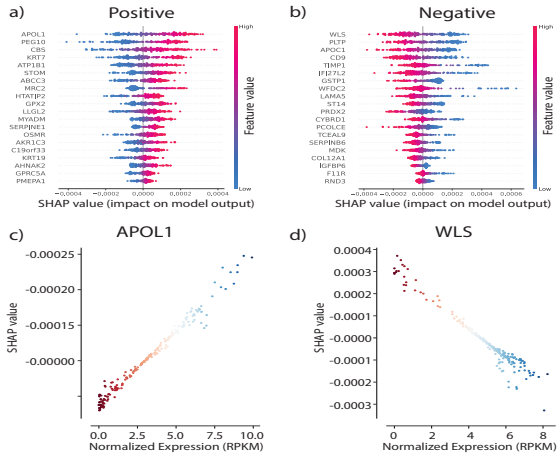


Figure A.7: SHAP-expression correlation analysis for BRAF1. **a)** Shap summary plot showing SHAP value (x-coordinate) and Normalized Expression value (color) for the top 20 genes (y-axis) with the highest positive Pearson correlation between SHAP and Expression values, where each point represents one sample in the test set. **b)** Summary plot for highest negative correlations. **c)** Scatter plot of SHAP value (y-axis) vs. Normalized Expression value (x-axis) for the gene with the highest positive correlation (APOL1) shown in a. **d)** Scatter plot for the highest negative correlation (WLS) shown in b.

Appendix B. Tables

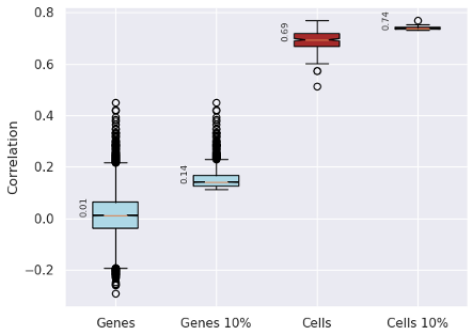


Figure A.8: The performance of the VAE-CL fine-tuned in predicting Sanger cohort. The figure reports the Spearman correlation coefficient of the predicted and actual essentiality values across genes and cell lines.

TF	**Rank**	**Sample p-value**				
		1st	**2nd**	**3rd**	**4th**	**5th**
HMG20A	1	7.09×10^{-13}	2.44×10^{-1}	N/A	N/A	N/A
FOS	2	2.31×10^{-10}	5.97×10^{-10}	7.15×10^{-10}	4.68×10^{-9}	7.23×10^{-8}
STAT3	3	1.13×10^{-9}	1.24×10^{-8}	2.23×10^{-6}	1.03×10^{-5}	1.26×10^{-5}
JUN	4	1.23×10^{-9}	3.76×10^{-9}	8.14×10^{-8}	1.21×10^{-7}	1.02×10^{-5}
YAP1	5	2.70×10^{-9}	4.45×10^{-8}	1.76×10^{-5}	2.48×10^{-3}	4.17×10^{-3}
TEAD1	6	4.47×10^{-9}	6.95×10^{-8}	4.93×10^{-5}	2.45×10^{-4}	1.26×10^{-2}
TP53	7	6.42×10^{-9}	4.39×10^{-6}	6.39×10^{-5}	1.82×10^{-4}	6.17×10^{-4}
RELA	8	2.30×10^{-8}	6.36×10^{-8}	2.86×10^{-4}	3.05×10^{-4}	3.31×10^{-4}
TEAD4	9	3.58×10^{-8}	1.22×10^{-5}	1.86×10^{-5}	2.35×10^{-5}	4.12×10^{-5}
SMAD4	10	4.74×10^{-8}	9.00×10^{-3}	1.53×10^{-1}	2.34×10^{-1}	5.69×10^{-1}

Table B.1: LISA-Inferred Transcription Factor Regulators of Important Genes Identified Using SHAP Interpretability

Appendix C. Related Work

Universal Gene Essentiality DeeplyEssential, a multi-layer perceptron (MLP) based network, predicts universal essential genes in microbes using features from DNA and protein sequences [12]. EPGAT, a Graph Neural Network (GNN) based approach predicts universal essential genes from protein-protein interaction networks and multiomics data [24]. EPGAT was benchmarked for four organisms, including humans.

Experimental Approach Gene essentiality can be measured using CRISPR-Cas9 in pooled or arrayed screens. Pooled screens deploy many gRNAs simultaneously, enabling high-throughput, cost-effective screening of thousands of genes. Arrayed screens use fewer gRNAs per gene, offering higher resolution and specificity [1]. The DepMap Achilles project aims to find essential genes in cancer cell lines via genome-scale, pooled CRISPR knockout screening, by knocking out genes in various cancer cells and measuring the effect on cell

proliferation and survival [5]. DepMap also measures response to different therapeutic compounds in various cancer cell lines [28].

Appendix D. Training Details

Both models were trained using a comprehensive set of hyperparameters to optimize the model performance. The model utilizes the Adam optimizer of the learning rate set to $5e^{-4}$, a batch size of 40, ReLU activation, 20 epochs for each model, and a dropout rate of 0.2. For the transformer layers, we use 2 transformer encoder layers with 4 heads, a hidden state of 1000 for the linear transformation layer, and a 2048 dimension for the feedforward layer. All experiments were performed on an NVIDIA A100 PCIe GPU with 80GB of memory with an estimated time of 1.2 seconds per epoch.

Appendix E. Baseline Models

In order to evaluate the performance of our proposed DeepVul framework, we compare it against several baseline models. The chosen baselines include variations of Variational Autoencoders with Contrastive Learning (VAE-CL) and a standard neural network. The VAE-CL baseline integrates a variational autoencoder with a contrastive learning objective to enhance the latent space representation. The VAE-CL Fine-Tuned baseline includes an additional phase of fine-tuning specifically tailored to the throughput prediction task of gene essentiality. A simple feedforward neural network (NN) baseline was designed to directly map input features to throughput predictions without any explicit representation learning phase.