Biologically informed deep neural network for prostate cancer classification and discovery

Authors: Haitham A. Elmarakeby[1,2,3], Justin Hwang[4], David Liu[1,2], Saud H. AlDubayan[1,2], Keyan Salari[1], Camden Richter[1], Taylor E Arnoff[1,2], Jihye Park[1,2], William C. Hahn[1,2], Eliezer Van Allen[1,2]*.

Affiliations:
[1]Dana-Farber Cancer Institute, Boston, MA, USA.
[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[3]Al-Azhar University, Cairo, Egypt.
[4]University of Minnesota, Division of Hematology, Oncology and Transplantation, Minneapolis, MN 55414.
*Correspondence to: eliezerm_vanallen@dfci.harvard.edu

Abstract: Determination of molecular features that mediate clinically aggressive phenotypes in prostate cancer (PrCa) remains a major biological and clinical challenge. Here, we developed a biologically informed deep learning model (P-NET) to stratify PrCa patients by treatment resistance state and evaluate molecular drivers of treatment resistance for therapeutic targeting through complete model interpretability. Using a molecular cohort of 1,238 prostate cancers, we demonstrated that P-NET can predict cancer state using molecular data that is superior to other modeling approaches. Moreover, the biological interpretability within P-NET revealed established and novel molecularly altered candidates, such as *MDM4* and *FGFR1*, that were implicated in predicting advanced disease and validated *in vitro*. Broadly, biologically informed fully interpretable neural networks enable preclinical discovery and clinical prediction in prostate cancer and may have general applicability across cancer types.

Main Text:

With the advancement of molecular profiling technologies, the ability to observe millions of genomic, transcriptional, and additional features from cancer patients and their tumors has grown significantly over the past decade. Specifically in prostate cancer, the availability of rich molecular profiling data linked to clinical annotation has enabled discovery of many individual genes, pathways, and complexes that promote lethal castration resistant prostate cancer (CRPC), which has led to both biological investigations and clinical evaluations of these individual features for predictive utility (1–9). However, the relationships between these molecular features, and their combined predictive and biological contributions to disease progression, drug resistance, and lethal outcomes, remains largely uncharacterized.

When developing a predictive model, one may choose from a large range of approaches, although each comes with tradeoffs of accuracy and interpretability. In translational cancer genomics, interpretability of predictive models is critical, as properties that contribute to the predictive capabilities of the model may not only inform patient care, but also provide insights into the underlying biological processes to prompt functional investigation and therapeutic targeting. Linear models such as logistic regression tend to have high interpretability with less accurate predictive performance, whereas deep learning models often have less interpretability but higher predictive performance (10, 11). Using a typical fully connected dense deep learning approach for building predictive models may also result in overfitting unless the network is well regularized, and such models have a tendency to be computationally expensive and less interpretable (12).

Efforts to search for slimmer architecture and sparse networks given a full model demonstrated that sparse models can decrease storage requirements and improve computational performance (13–15). However, finding such a sparse model may be challenging, since the typical training-pruning-retraining cycle is usually computationally expensive, and recent studies indicate that building a sparse model *de novo* may be easier (16). Additionally, efforts to enhance the interpretability of deep learning models and the need to explain model decisions led to the development of multiple attribution methods, including LIME (17), DeepLIFT (10), DeepExplain (18), and SHAP (19), that can be used to enhance the deep learning explainability and understand how the model is processing information and making decisions.

Taken together, the advances in sparse model development and attribution methods have informed the development of deep learning models to solve biological problems using customized neural network architectures that are inspired by biological systems. For example, Ma et al, developed a visible neural network system, DCell, to model the effect of gene interaction on cell growth in yeast (20). A Pathway-Associated Sparse Deep Neural Network (PASNet) PASNet used a flattened version of pathways to predict patient prognosis in Glioblastoma multiforme (21). However, whether biologically informed neural networks can accelerate both biological discovery with translational potential, and simultaneously enable clinical predictive modeling, is largely unknown. Here, we hypothesized that a biologically informed deep learning model built upon advances in sparse deep learning architectures, encoding of biological information, and incorporation

of explainability algorithms would achieve superior predictive performance compared to established models and reveal novel patterns of treatment resistance in prostate cancer with translational implications.

Results:

We developed a deep learning predictive model that incorporates prior biologically established hierarchical knowledge in a neural network language to predict cancer state in prostate cancer patients based on their genomic profiles. A set of 3,007 curated biological pathways were used to build a pathway-aware multi-layered hierarchical network (P-NET) (Methods). In P-NET, the patient molecular profile is fed into the model and distributed over a layer of nodes representing a set of genes of interest using weighted links (Figure 1a). Later layers of the network encode a set of pathways with increasing levels of abstraction, whereby lower layers represent fine pathways while later layers represent more complex biological pathways and biological processes. The connections between different layers are constrained to follow known child-parent relationships among encoded features, genes, and pathways, and as a result the network is fully interpretable.

We trained and tested P-NET with a set of 1,013 prostate cancers (CRPC = 333, primary = 680), divided into 80% training, 10% validation, and 10% testing, to predict disease state (primary or metastatic disease). The trained P-NET outperformed typical machine learning models, including Support Vector Machine, Logistic Regression, and Decision Trees (AUC=0.93, AUPRC= 0.88, Accuracy 0.83) (Methods, Figure 1b, c, area under ROC curves are shown in Figure S3, 5-fold cross validation results are shown in Figure S4). Furthermore, we evaluated whether the sparse model had characteristics distinct from a dense fully connected deep learning model. We trained a dense model with the same number of parameters as in the P-NET model on training sets with a logarithmically increasing number of samples from 100 to 811 (80% of the total number of samples). The mean performance (determined by AUC) of the P-NET model was higher than the dense model over all the sample sizes, and this difference was statistically significant in smaller sample sizes (<=500) (e.g. mean AUC of 5-fold cross validation was significantly higher for P-NET compared to a dense network trained on 155 samples, *p-value* = 0.003) (Figure 2a, more metrics are shown in Figure S5).

We next performed external validation of the predictive aspects of the model using two additional PrCa validation cohorts, one primary (*22*) and one metastatic (*23*) (total n=225, links are included in Data S1). The trained P-NET model correctly classified 73% of the primary tumors and 76% of the castration resistant metastatic tumors, indicating that the model can generalize to unseen samples with an adequate predictive performance (Figure 2b). We hypothesized that patients with primary tumors samples incorrectly classified by P-NET as castration resistant metastatic tumors may in fact have worse clinical outcomes. Patients with high P-NET scores misclassified as resistant disease were significantly more likely to have biochemical recurrence than patients with low P-NET scores (*p* = 8*10-5; log-rank) indicating that for patients with primary prostate cancer, the P-NET score may be used to predict potential biochemical recurrence (Figure 2c).

To understand the interactions between different features, genes, pathways, and biological processes that contributed to the predictive performance, and to study the paths of impact from the input to the outcome, we visualized the whole structure of P-NET with the fully interpretable layers after training (Figure 3a and S6). Among aggregate molecular alterations, copy number variation (CNV) was more informative compared to mutations, consistent with prior reports (24). In addition, P-NET selected a hierarchy of pathways (out of 3,007 pathways on which P-NET was trained) as relevant to classification, including post-translational modification (PTM) (including Ubiquitination and SUMoylation) and transcriptional regulation by *RUNX2* and *TP53*. Ubiquitination and SUMoylation pathways contribute to the regulation of multiple tumor suppressors and oncogenes, and dysregulation of these pathways has been linked to prostate cancer initiation and progression in preclinical models (25). *RUNX2* is an osteogenic transcription factor that regulates cell proliferation and is associated with metastatic disease in prostate cancer patients (26).

To evaluate the relative importance of specific genes contributing to the model prediction, we inspected the genes layer and utilized the DeepLIFT attribution method to obtain the average ranking of genes (Methods) (10). Highly ranked genes included *AR*, *PTEN*, *RB1*, and *TP53*, which are known PrCa drivers previously associated with metastatic disease (2, 5, 7, 27). In addition, alterations in less expected genes, such as *MDM4*, *FGFR1*, *NOTCH1*, and *PDGFA*, strongly contributed to predictive performance (Figure 3b). To understand the behavior of trained P-NET, we checked the activation of each node in the network and asked whether this activation changed with the change of the input sample class (Primary vs. Metastatic) (Methods). We observed that the difference in the node activation was higher in higher layers and more concentrated in highly ranked nodes in each layer (Figure S7). For example, the activation distribution of the nodes of layer 3 was different when P-NET was given a primary sample compared to a resistant sample (Figure 3c). Thus, the interpretable architecture of P-NET can be interrogated to understand how the input information is transformed through layers and nodes, enabling further understanding of the state and importance of the involved biological entities.

Through evaluation of multiple layers in the P-NET trained model, we observed convergence in *TP53*-associated biology contributing to CRPC. Tracing the relevance of *TP53*-related pathways to the gene levels, we detected that in addition to *TP53* and *MDM2*, each established in prostate cancer disease progression (27–31), alterations in *MDM4* strongly contributed to this network convergence. *MDM4* can inhibit wild-type *TP53* expression by binding to and masking the transcriptional activation domain, though its role in prostate cancer treatment resistance is incompletely characterized (32).

We further studied the *MDM4* profile both in clinical samples and functional models. *MDM4* focal amplification was more prevalent in resistant samples compared to primary samples (X2Yates correction = 40.8251, *p-value* <0.00001) (Figure 4a). Copy number and mutation distributions of other top genes are shown in Figure S8. In a genome-wide gain-of-function preclinical screen using 17,255 open reading frames ORFs in LNCaP cells, *MDM4* overexpression was significantly associated with resistance to enzalutamide, a second-generation antiandrogen medication which is used in CRPC

patient populations (*33*) (Figure 4c). We thus utilized CRISPR-Cas9 to target *MDM4* in LNCaP prostate cancer cell lines. In comparison to a negative control, proliferation of LNCaP cells was reduced by 60% (*p-value* <0.0001) (Figure 4d) in response to *MDM4* depletion using two distinct sgRNAs (Figure 4e). This indicated that selective therapeutic targeting of *MDM4* may be viable in TP53-wildtype advanced prostate cancer patients. We then sought to study the effect of inhibiting *MDM4* in prostate cell lines with mutant and wild-type *TP53*. Prostate cells with wild-type TP53 were more sensitive to the *MDM4* inhibitor RO-5963 compared to mutant *TP53* cell lines (2μM of RO-5963 reduced the viability of LNCaP cells by nearly 50%; Figure 4f). Overall, convergence of TP53 pathway dysregulation across multiple layers of the trained P-NET model identified specific vulnerabilities involving *MDM4*, which can be therapeutically targeted with *MDM4*-selective inhibition in a genomically stratified prostate cancer patient population.

Discussion:
Broadly, P-NET leveraged a biologically informed, rather than arbitrarily overparameterized, architecture for prediction. As a result, P-NET dramatically reduced the number of parameters for learning, which led to enhanced interpretability. The sparse architecture in P-NET has better predictive performance when compared to other machine learning models, including dense networks, and may be applicable to other similar tasks. Application of P-NET to a molecular cohort of prostate cancer patients demonstrated (i) model performance that may enable prediction of clinically aggressive disease in primary prostate cancer patient populations, and (ii) convergent biological processes that contribute to a metastatic prostate cancer clinical phenotype that harbor novel therapeutic strategies in molecularly stratified populations.

Furthermore, P-NET provided a natural way for integrating multiple molecular features (e.g. mutations and copy number variations) weighted differently to reflect their importance in predicting the final outcome, which previously required different significance approaches for each feature to enable cancer gene discovery (*34, 35*). Even more, P-NET provided a framework for encoding hierarchical prior knowledge using neural network languages, and turning these hierarchies into a computational model that can be used both for prediction and biological discovery. Specifically, P-NET accurately predicted advanced prostate disease based on the patient's genomic profile and had the ability to predict potential BCR before happening. P-NET visualization enabled a multilevel view of the involved biological pathways and processes, which may guide researchers to develop hypotheses regarding the underlying biological processes involved in cancer progression and translate these discoveries into therapeutic opportunities. Specifically, P-NET rediscovered known genes such as *AR, PTEN, TP53,* and *RB1*. Moreover, P-NET nominated *MDM4* as a novel therapeutic target, which was experimentally validated and may inform further genomically directed therapeutic options for this patient population.

While P-NET provides a framework for outcome prediction and hypothesis generation, this model still requires tuning and training before being used in any given context. As with all deep learning models, the final trained model heavily depends on the hyperparameters used to train the model. In addition, P-NET encodes biological pathways

inside the network in a hardcoded way, which makes the model depend on the quality of the annotations used to build the model. Thus, the portability of this approach across different histological and clinical contexts requires further evaluation.

In conclusion, P-NET, a biologically informed deep neural network, accurately classified castration resistant metastatic versus primary prostate cancers. Visualizing the trained model generated novel hypotheses for mechanisms of metastasis in prostate cancer, and provided insights with direct potential for clinical translation in molecularly stratified prostate cancer patient populations. Biologically guided neural networks represent a novel approach to integrating cancer biology with machine-learning by building mechanistic predictive models, providing a platform for biological discovery that may be broadly applicable across cancer prediction and discovery tasks

References:

1. G. Gundem, P. Van Loo, B. Kremeyer, L. B. Alexandrov, J. M. C. Tubio, E. Papaemmanuil, D. S. Brewer, H. M. L. Kallio, G. Högnäs, M. Annala, K. Kivinummi, V. Goody, C. Latimer, S. O'Meara, K. J. Dawson, W. Isaacs, M. R. Emmert-Buck, M. Nykter, C. Foster, Z. Kote-Jarai, D. Easton, H. C. Whitaker, ICGC Prostate Group, D. E. Neal, C. S. Cooper, R. A. Eeles, T. Visakorpi, P. J. Campbell, U. McDermott, D. C. Wedge, G. S. Bova, The evolutionary history of lethal metastatic prostate cancer. *Nature*. **520**, 353–357 (2015).

2. D. Robinson, E. M. Van Allen, Y.-M. Wu, N. Schultz, R. J. Lonigro, J.-M. Mosquera, B. Montgomery, M.-E. Taplin, C. C. Pritchard, G. Attard, H. Beltran, W. Abida, R. K. Bradley, J. Vinson, X. Cao, P. Vats, L. P. Kunju, M. Hussain, F. Y. Feng, S. A. Tomlins, K. A. Cooney, D. C. Smith, C. Brennan, J. Siddiqui, R. Mehra, Y. Chen, D. E. Rathkopf, M. J. Morris, S. B. Solomon, J. C. Durack, V. E. Reuter, A. Gopalan, J. Gao, M. Loda, R. T. Lis, M. Bowden, S. P. Balk, G. Gaviola, C. Sougnez, M. Gupta, E. Y. Yu, E. A. Mostaghel, H. H. Cheng, H. Mulcahy, L. D. True, S. R. Plymate, H. Dvinge, R. Ferraldeschi, P. Flohr, S. Miranda, Z. Zafeiriou, N. Tunariu, J. Mateo, R. Perez-Lopez, F. Demichelis, B. D. Robinson, A. Sboner, M. Schiffman, D. M. Nanus, S. T. Tagawa, A. Sigaras, K. W. Eng, O. Elemento, A. Sboner, E. I. Heath, H. I. Scher, K. J. Pienta, P. Kantoff, J. S. de Bono, M. A. Rubin, P. S. Nelson, L. A. Garraway, C. L. Sawyers, A. M. Chinnaiyan, Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell*. **162**, 454 (2015).

3. R. Aggarwal, J. Huang, J. J. Alumkal, L. Zhang, F. Y. Feng, G. V. Thomas, A. S. Weinstein, V. Friedl, C. Zhang, O. N. Witte, P. Lloyd, M. Gleave, C. P. Evans, J. Youngren, T. M. Beer, M. Rettig, C. K. Wong, L. True, A. Foye, D. Playdle, C. J. Ryan, P. Lara, K. N. Chi, V. Uzunangelov, A. Sokolov, Y. Newton, H. Beltran, F. Demichelis, M. A. Rubin, J. M. Stuart, E. J. Small, Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J. Clin. Oncol.* **36**, 2492–2503 (2018).

4. J. Armenia, S. A. M. Wankowicz, D. Liu, J. Gao, R. Kundra, E. Reznik, W. K. Chatila, D. Chakravarty, G. C. Han, I. Coleman, B. Montgomery, C. Pritchard, C. Morrissey, C. E. Barbieri, H. Beltran, A. Sboner, Z. Zafeiriou, S. Miranda, C. M. Bielski, A. V. Penson, C. Tolonen, F. W. Huang, D. Robinson, Y. M. Wu, R. Lonigro, L. A. Garraway, F. Demichelis, P. W. Kantoff, M.-E. Taplin, W. Abida, B. S. Taylor, H. I. Scher, P. S. Nelson, J. S. de Bono, M. A. Rubin, C. L. Sawyers, A. M. Chinnaiyan, PCF/SU2C International Prostate Cancer Dream Team, N. Schultz, E. M. Van Allen, PCF/SU2C International Prostate Cancer Dream Team, The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).

5. D. A. Quigley, H. X. Dang, S. G. Zhao, P. Lloyd, R. Aggarwal, J. J. Alumkal, A. Foye, V. Kothari, M. D. Perry, A. M. Bailey, D. Playdle, T. J. Barnard, L. Zhang, J. Zhang, J. F. Youngren, M. P. Cieslik, A. Parolia, T. M. Beer, G. Thomas, K. N. Chi, M. Gleave, N. A. Lack, A. Zoubeidi, R. E. Reiter, M. B. Rettig, O. Witte, C. J. Ryan,

L. Fong, W. Kim, T. Friedlander, J. Chou, H. Li, R. Das, H. Li, R. Moussavi-Baygi, H. Goodarzi, L. A. Gilbert, P. N. Lara Jr, C. P. Evans, T. C. Goldstein, J. M. Stuart, S. A. Tomlins, D. E. Spratt, R. K. Cheetham, D. T. Cheng, K. Farh, J. S. Gehring, J. Hakenberg, A. Liao, P. G. Febbo, J. Shon, B. Sickler, S. Batzoglou, K. E. Knudsen, H. H. He, J. Huang, A. W. Wyatt, S. M. Dehm, A. Ashworth, A. M. Chinnaiyan, C. A. Maher, E. J. Small, F. Y. Feng, Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell*. **175**, 889 (2018).

6. D. N. Rodrigues, P. Rescigno, D. Liu, W. Yuan, S. Carreira, M. B. Lambros, G. Seed, J. Mateo, R. Riisnaes, S. Mullane, C. Margolis, D. Miao, S. Miranda, D. Dolling, M. Clarke, C. Bertan, M. Crespo, G. Boysen, A. Ferreira, A. Sharp, I. Figueiredo, D. Keliher, S. Aldubayan, K. P. Burke, S. Sumanasuriya, M. S. Fontes, D. Bianchini, Z. Zafeiriou, L. S. T. Mendes, K. Mouw, M. T. Schweizer, C. C. Pritchard, S. Salipante, M.-E. Taplin, H. Beltran, M. A. Rubin, M. Cieslik, D. Robinson, E. Heath, N. Schultz, J. Armenia, W. Abida, H. Scher, C. Lord, A. D'Andrea, C. L. Sawyers, A. M. Chinnaiyan, A. Alimonti, P. S. Nelson, C. G. Drake, E. M. Van Allen, J. S. de Bono, Immunogenomic analyses associate immunological alterations with mismatch repair defects in prostate cancer. *J. Clin. Invest.* **128**, 5185 (2018).

7. W. Abida, J. Cyrta, G. Heller, D. Prandi, J. Armenia, I. Coleman, M. Cieslik, M. Benelli, D. Robinson, E. M. Van Allen, A. Sboner, T. Fedrizzi, J. M. Mosquera, B. D. Robinson, N. De Sarkar, L. P. Kunju, S. Tomlins, Y. M. Wu, D. Nava Rodrigues, M. Loda, A. Gopalan, V. E. Reuter, C. C. Pritchard, J. Mateo, D. Bianchini, S. Miranda, S. Carreira, P. Rescigno, J. Filipenko, J. Vinson, R. B. Montgomery, H. Beltran, E. I. Heath, H. I. Scher, P. W. Kantoff, M.-E. Taplin, N. Schultz, J. S. deBono, F. Demichelis, P. S. Nelson, M. A. Rubin, A. M. Chinnaiyan, C. L. Sawyers, Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11428–11436 (2019).

8. W. S. Chen, R. Aggarwal, L. Zhang, S. G. Zhao, G. V. Thomas, T. M. Beer, D. A. Quigley, A. Foye, D. Playdle, J. Huang, P. Lloyd, E. Lu, D. Sun, X. Guan, M. Rettig, M. Gleave, C. P. Evans, J. Youngren, L. True, P. Lara, V. Kothari, Z. Xia, K. N. Chi, R. E. Reiter, C. A. Maher, F. Y. Feng, E. J. Small, J. J. Alumkal, West Coast Prostate Cancer Dream Team, Genomic Drivers of Poor Prognosis and Enzalutamide Resistance in Metastatic Castration-resistant Prostate Cancer. *Eur. Urol.* **76**, 562–571 (2019).

9. S. G. Zhao, W. S. Chen, H. Li, A. Foye, M. Zhang, M. Sjöström, R. Aggarwal, D. Playdle, A. Liao, J. J. Alumkal, R. Das, J. Chou, J. T. Hua, T. J. Barnard, A. M. Bailey, E. D. Chow, M. D. Perry, H. X. Dang, R. Yang, R. Moussavi-Baygi, L. Zhang, M. Alshalalfa, S. Laura Chang, K. E. Houlahan, Y.-J. Shiah, T. M. Beer, G. Thomas, K. N. Chi, M. Gleave, A. Zoubeidi, R. E. Reiter, M. B. Rettig, O. Witte, M. Yvonne Kim, L. Fong, D. E. Spratt, T. M. Morgan, R. Bose, F. W. Huang, H. Li, L. Chesner, T. Shenoy, H. Goodarzi, I. A. Asangani, S. Sandhu, J. M. Lang, N. P. Mahajan, P. N. Lara, C. P. Evans, P. Febbo, S. Batzoglou, K. E. Knudsen, H. H. He, J. Huang, W. Zwart, J. F. Costello, J. Luo, S. A. Tomlins, A. W. Wyatt, S. M.

Dehm, A. Ashworth, L. A. Gilbert, P. C. Boutros, K. Farh, A. M. Chinnaiyan, C. A. Maher, E. J. Small, D. A. Quigley, F. Y. Feng, The DNA methylation landscape of advanced prostate cancer. *Nat. Genet.* **52**, 778–789 (2020).

10. A. Shrikumar, P. Greenside, A. Kundaje, Learning Important Features Through Propagating Activation Differences. *arXiv [cs.CV]* (2017), (available at http://arxiv.org/abs/1704.02685).

11. W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22071–22080 (2019).

12. Q. Xu, M. Zhang, Z. Gu, G. Pan, Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs. *Neurocomputing.* **328**, 69–74 (2019).

13. S. Han, J. Pool, J. Tran, W. Dally, in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett, Eds. (Curran Associates, Inc., 2015), pp. 1135–1143.

14. H. Li, A. Kadav, I. Durdanovic, H. Samet, H. P. Graf, Pruning Filters for Efficient ConvNets. *arXiv [cs.CV]* (2016), (available at http://arxiv.org/abs/1608.08710).

15. J. Frankle, M. Carbin, The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. *arXiv [cs.LG]* (2018), (available at http://arxiv.org/abs/1803.03635).

16. T. Dettmers, L. Zettlemoyer, Sparse Networks from Scratch: Faster Training without Losing Performance. *arXiv [cs.LG]* (2019), (available at http://arxiv.org/abs/1907.04840).

17. M. T. Ribeiro, S. Singh, C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), pp. 1135–1144.

18. M. Ancona, E. Ceolini, C. Öztireli, M. Gross, Towards better understanding of gradient-based attribution methods for Deep Neural Networks. *arXiv [cs.LG]* (2017), (available at http://arxiv.org/abs/1711.06104).

19. S. M. Lundberg, S.-I. Lee, in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds. (Curran Associates, Inc., 2017), pp. 4765–4774.

20. J. Ma, M. K. Yu, S. Fong, K. Ono, E. Sage, B. Demchak, R. Sharan, T. Ideker, Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* (2018), doi:10.1038/nmeth.4627.

21. J. Hao, Y. Kim, T.-K. Kim, M. Kang, PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC*

Bioinformatics. **19**, 510 (2018).

22. M. Fraser, V. Y. Sabelnykova, T. N. Yamaguchi, L. E. Heisler, J. Livingstone, V. Huang, Y.-J. Shiah, F. Yousif, X. Lin, A. P. Masella, N. S. Fox, M. Xie, S. D. Prokopec, A. Berlin, E. Lalonde, M. Ahmed, D. Trudel, X. Luo, T. A. Beck, A. Meng, J. Zhang, A. D'Costa, R. E. Denroche, H. Kong, S. M. G. Espiritu, M. L. K. Chua, A. Wong, T. Chong, M. Sam, J. Johns, L. Timms, N. B. Buchner, M. Orain, V. Picard, H. Hovington, A. Murison, K. Kron, N. J. Harding, C. P'ng, K. E. Houlahan, K. C. Chu, B. Lo, F. Nguyen, C. H. Li, R. X. Sun, R. de Borja, C. I. Cooper, J. F. Hopkins, S. K. Govind, C. Fung, D. Waggott, J. Green, S. Haider, M. A. Chan-Seng-Yue, E. Jung, Z. Wang, A. Bergeron, A. Dal Pra, L. Lacombe, C. C. Collins, C. Sahinalp, M. Lupien, N. E. Fleshner, H. H. He, Y. Fradet, B. Tetu, T. van der Kwast, J. D. McPherson, R. G. Bristow, P. C. Boutros, Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*. **541**, 359–364 (2017).

23. D. R. Robinson, Y.-M. Wu, R. J. Lonigro, P. Vats, E. Cobain, J. Everett, X. Cao, E. Rabban, C. Kumar-Sinha, V. Raymond, S. Schuetze, A. Alva, J. Siddiqui, R. Chugh, F. Worden, M. M. Zalupski, J. Innis, R. J. Mody, S. A. Tomlins, D. Lucas, L. H. Baker, N. Ramnath, A. F. Schott, D. F. Hayes, J. Vijai, K. Offit, E. M. Stoffel, J. S. Roberts, D. C. Smith, L. P. Kunju, M. Talpaz, M. Cieślik, A. M. Chinnaiyan, Integrative clinical genomics of metastatic cancer. *Nature*. **548**, 297–303 (2017).

24. H. Hieronymus, N. Schultz, A. Gopalan, B. S. Carver, M. T. Chang, Y. Xiao, A. Heguy, K. Huberman, M. Bernstein, M. Assel, R. Murali, A. Vickers, P. T. Scardino, C. Sander, V. Reuter, B. S. Taylor, C. L. Sawyers, Copy number alteration burden predicts prostate cancer relapse. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11139–11144 (2014).

25. Z. Chen, W. Lu, Roles of ubiquitination and SUMOylation on prostate cancer: mechanisms and clinical implications. *Int. J. Mol. Sci.* **16**, 4560–4580 (2015).

26. C. Ge, G. Zhao, Y. Li, H. Li, X. Zhao, G. Pannone, P. Bufo, A. Santoro, F. Sanguedolce, S. Tortorella, M. Mattoni, S. Papagerakis, E. T. Keller, R. T. Franceschi, Role of Runx2 phosphorylation in prostate cancer and association with metastatic disease. *Oncogene*. **35**, 366–376 (2016).

27. A. J. Levine, p53: 800 million years of evolution and 40 years of discovery. *Nat. Rev. Cancer*. **20**, 471–480 (2020).

28. N. M. Navone, P. Troncoso, L. L. Pisters, T. L. Goodrow, J. L. Palmer, W. W. Nichols, A. C. von Eschenbach, C. J. Conti, p53 protein accumulation and gene mutation in the progression of human prostate carcinoma. *J. Natl. Cancer Inst.* **85**, 1657–1669 (1993).

29. K. R. Leite, M. F. Franco, M. Srougi, L. J. Nesrallah, A. Nesrallah, R. G. Bevilacqua, E. Darini, C. M. Carvalho, M. I. Meirelles, I. Santana, L. H. Camara-Lopes, Abnormal expression of MDM2 in prostate carcinoma. *Mod. Pathol.* **14**, 428–436 (2001).

30. T. Schlomm, L. Iwers, P. Kirstein, B. Jessen, J. Köllermann, S. Minner, A. Passow-Drolet, M. Mirlacher, K. Milde-Langosch, M. Graefen, A. Haese, T. Steuber, R. Simon, H. Huland, G. Sauter, A. Erbersdobler, Clinical significance of p53 alterations in surgically treated prostate cancers. *Mod. Pathol.* **21**, 1371–1378 (2008).

31. F. Y. Feng, Y. Zhang, V. Kothari, J. R. Evans, W. C. Jackson, W. Chen, S. B. Johnson, C. Luczak, S. Wang, D. A. Hamstra, MDM2 Inhibition Sensitizes Prostate Cancer Cells to Androgen Ablation and Radiotherapy in a p53-Dependent Manner. *Neoplasia*. **18**, 213–222 (2016).

32. D. Pei, Y. Zhang, J. Zheng, Regulation of p53: a collaboration between Mdm2 and Mdmx. *Oncotarget*. **3**, 228–235 (2012).

33. J. H. Hwang, J.-H. Seo, M. L. Beshiri, S. Wankowicz, D. Liu, A. Cheung, J. Li, X. Qiu, A. L. Hong, G. Botta, L. Golumb, C. Richter, J. So, G. J. Sandoval, A. O. Giacomelli, S. H. Ly, C. Han, C. Dai, H. Pakula, A. Sheahan, F. Piccioni, O. Gjoerup, M. Loda, A. G. Sowalsky, L. Ellis, H. Long, D. E. Root, K. Kelly, E. M. Van Allen, M. L. Freedman, A. D. Choudhury, W. C. Hahn, CREB5 Promotes Resistance to Androgen-Receptor Antagonists and Androgen Deprivation in Prostate Cancer. *Cell Rep.* **29**, 2355–2370.e6 (2019).

34. C. H. Mermel, S. E. Schumacher, B. Hill, M. L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).

35. M. S. Lawrence, P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, M. Meyerson, S. B. Gabriel, E. S. Lander, G. Getz, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. **505**, 495–501 (2014).

36. A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).

37. D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]* (2014), (available at http://arxiv.org/abs/1412.6980).

38. E. G. Bluemn, I. M. Coleman, J. M. Lucas, R. T. Coleman, S. Hernandez-Lopez, R. Tharakan, D. Bianchi-Frias, R. F. Dumpit, A. Kaipainen, A. N. Corella, Y. C. Yang, M. D. Nyquist, E. Mostaghel, A. C. Hsieh, X. Zhang, E. Corey, L. G. Brown, H. M. Nguyen, K. Pienta, M. Ittmann, M. Schweizer, L. D. True, D. Wise, P. S. Rennie, R. L. Vessella, C. Morrissey, P. S. Nelson, Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell*. **32**, 474–489.e6 (2017).

39. G. C. Han, J. Hwang, S. A. M. Wankowicz, Z. Zhang, D. Liu, C. Cibulskis, G. C.

Gaviola, V. Ghazikhanian, R. R. McKay, G. J. Bubley, S. L. Carter, S. P. Balk, W. C. Hahn, M.-E. Taplin, E. M. Van Allen, Genomic Resistance Patterns to Second-Generation Androgen Blockade in Paired Tumor Biopsies of Metastatic Castration-Resistant Prostate Cancer. *JCO Precis Oncol*. **1** (2017), doi:10.1200/PO.17.00140.

40. B. Graves, T. Thompson, M. Xia, C. Janson, C. Lukacs, D. Deo, P. Di Lello, D. Fry, C. Garvie, K.-S. Huang, L. Gao, C. Tovar, A. Lovey, J. Wanner, L. T. Vassilev, Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11788–11793 (2012).

41. T. P. Howard, T. E. Arnoff, M. R. Song, A. O. Giacomelli, X. Wang, A. L. Hong, N. V. Dharia, S. Wang, F. Vazquez, M.-T. Pham, A. M. Morgan, F. Wachter, G. H. Bird, G. Kugener, E. M. Oberlick, M. G. Rees, H. L. Tiv, J. H. Hwang, K. H. Walsh, A. Cook, J. M. Krill-Burger, A. Tsherniak, P. C. Gokhale, P. J. Park, K. Stegmaier, L. D. Walensky, W. C. Hahn, C. W. M. Roberts, MDM2 and MDM4 Are Therapeutic Vulnerabilities in Malignant Rhabdoid Tumors. *Cancer Res.* **79**, 2404–2414 (2019).

Fig. 1. Interpretable biologically informed deep learning model for molecular discovery in metastatic prostate cancer. a) P-NET; neural network architecture that encodes different biological entities into a neural network language with customized connections between consecutive layers (i.e. features from patient profile, genes, pathways, biological processes, and outcome). The trained P-NET provides a relative ranking of nodes in each layer which can be used to generate biological hypotheses regarding the relevance of different biological entities to the outcome of interest. Candidate genes are experimentally and clinically validated to understand the function and the mechanism of action of these genes. b) The Precision-Recall curve of multiple predictive models including Random forest, Support Vector Machine, Decision Trees, and adaptive Boosting as trained and tested on a cohort of prostate cancer patients. P-NET achieves better area under precision recall curve (AUPRC). c) The confusion matrix of P-NET showing the correct classification percentage of samples in the testing set. The model is biased toward predicting more primary samples given that the training set is biased toward having more primary samples (70% primary and 30% metastatic).

13
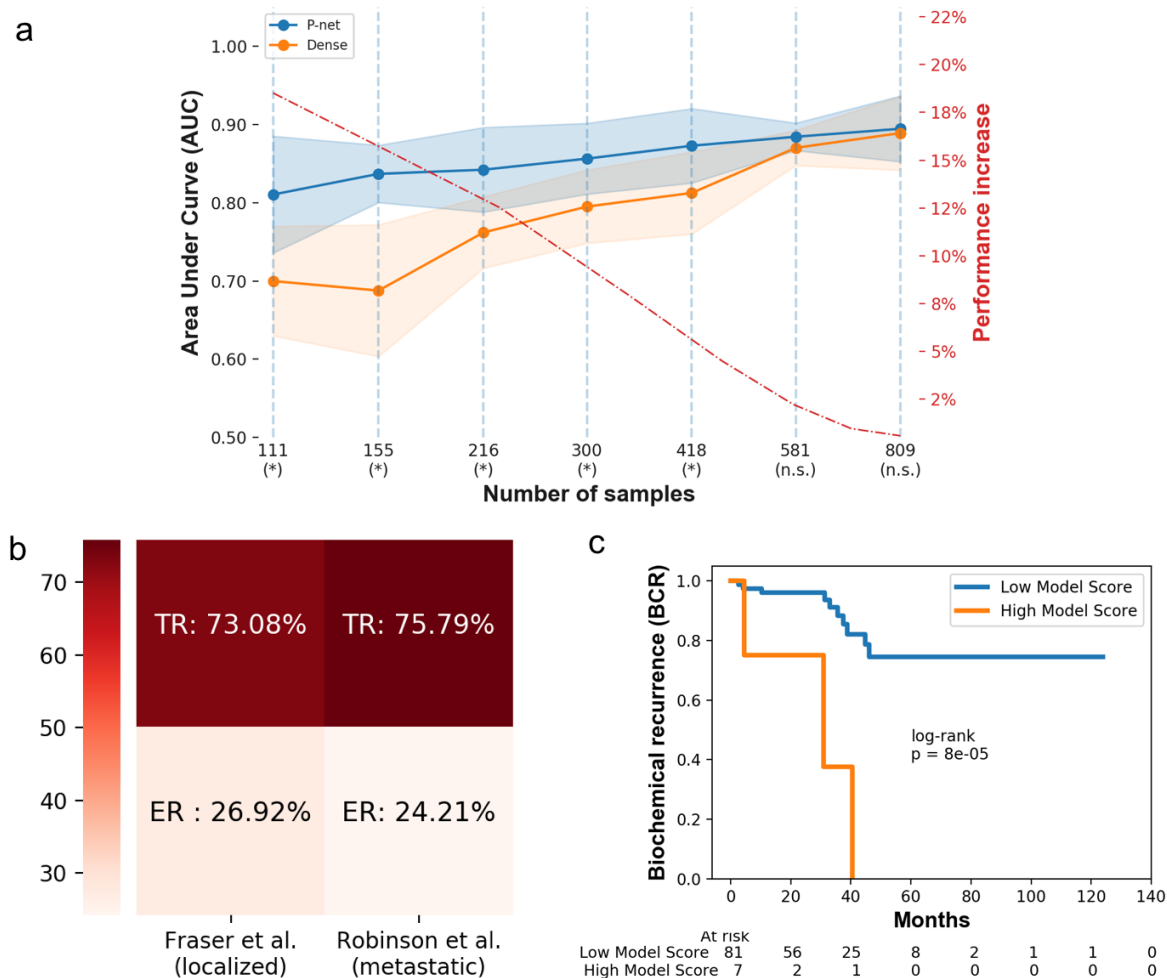
Fig. 2. Prediction performance and interpretation of P-NET. a) Prediction performance of P-NET when compared to a fully connected network with the same number of parameters trained using an increasing number of training samples. P-NET achieves better performance (measured as the average AUC over 5 cross validation splits) with smaller numbers of samples compared to a dense fully connected network. The ratio of performance increase is polynomially increasing with the decrease of the number of training samples as shown with the dashed red line. The difference in performance between the two models is statistically significant in all sample sizes less than or equal to 500. Sample sizes marked by (*) indicate statistically significant differences while those marked by (n.s.) are not. b) External validation of P-NET using two independent cohorts (22, 23). The P-NET model achieves 73% and 75.79% true prediction rate (TR) respectively showing that the P-NET can generalize to classify unseen samples. c) Wrong predictions of P-NET were inspected for clinical insights. Patients with high P-NET scores (wrongly classified by P-NET to be resistant samples) have more tendency to have biochemical recurrence (BCR) compared to patients with lower P-NET scores who tend to have progression free survival (PFS). This shows that the P-NET model may be useful in stratifying patients in the clinic and predicting potential BCR before it happens (raw data are included in Data S4).

14
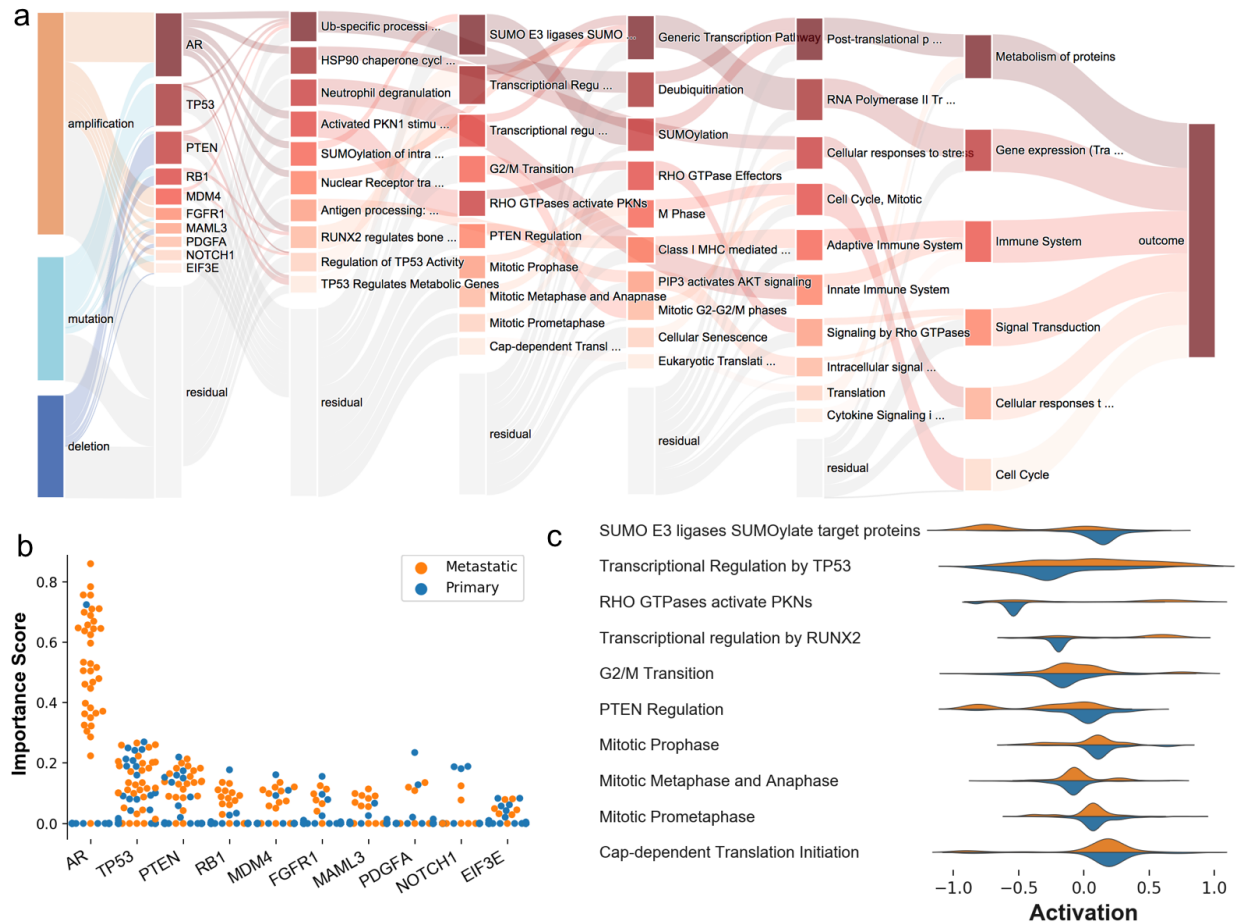
Fig. 3. Inspecting and interpreting P-NET. a) Visualization of inner layers of P-NET shows the estimated relative importance of different nodes in each layer. Nodes on the far left represent feature types (mutation and copy number amplification and deletion). The second layer nodes represent genes and layers representing higher level biological entities shown in layers to the right. The final layer represents the model outcome. Nodes with darker color are shown to be more important while nodes with transparent color represent the residual importance of non-shown nodes in each layer. The contribution of datatype to the importance of each gene is depicted using the Sankey diagram. The importance of *AR* gene is driven mainly by gene amplification, the importance of *TP53* is driven by mutation, and the importance of *PTEN* gene is driven by deletion. b) Top genes are ranked based on the average importance of each gene. The distribution of sample-level importance calculated for the testing set is shown in the Swarm diagram. In general, metastatic samples tend to have more influence on the importance of different genes. c) Inspecting the activation output of each node shows how the outcome of different nodes changes with changing the class of input samples (primary-blue vs. metastatic-orange). The activation distribution of top nodes in layer 3 shows that the "Transcription Regulation by *TP53*" pathways is ranked second among pathways in the same layer. Metastatic samples tend to push the activation of this pathway toward positive values while primary samples tend to have negative activation of this pathway.
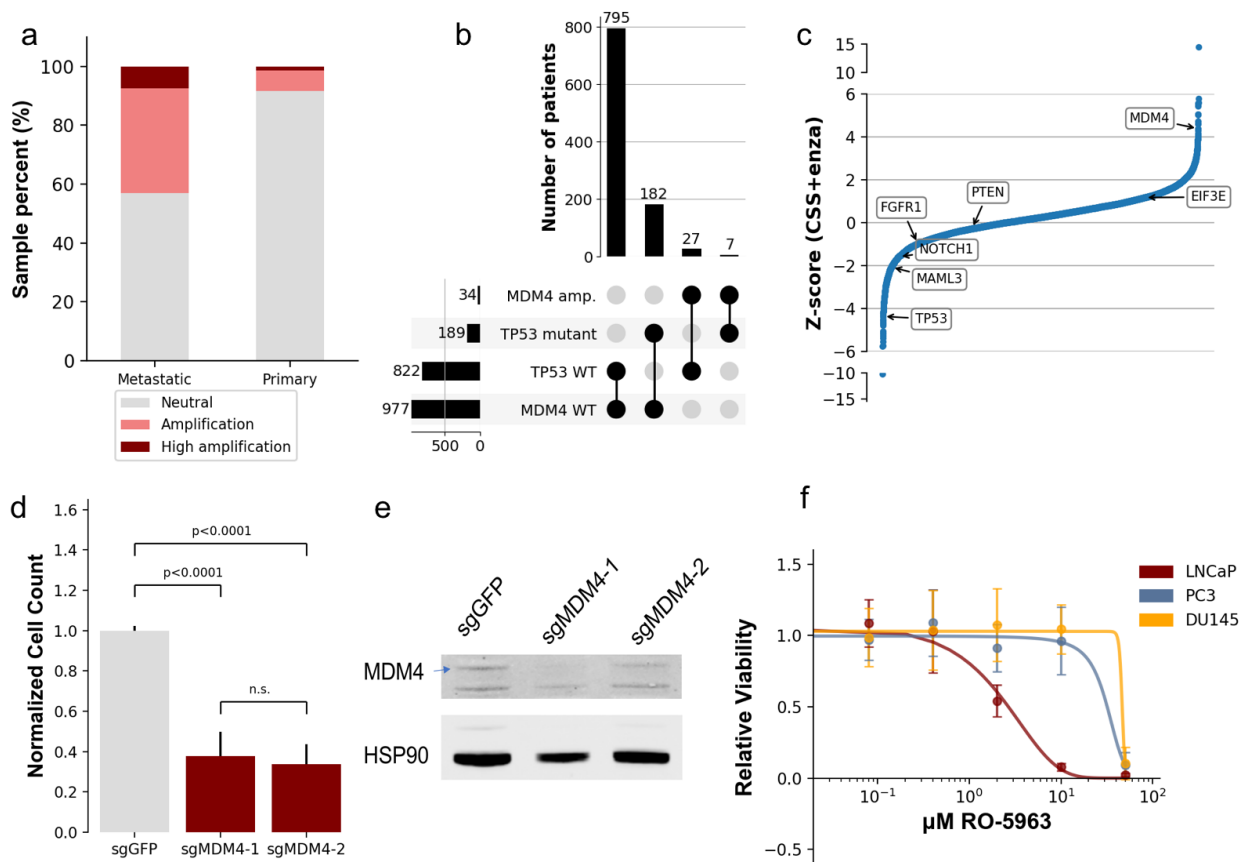
Fig. 4. Functional validation. a) Distribution of *MDM4* alterations across 1,013 PrCa samples showing more prevalence of *MDM4* amplification in resistant samples compared to primary samples. b) Joint distribution of TP53 mutations and *MDM4* amplifications across 1,013 PrCa samples. c) Analysis of enzalutamide resistant genes in LNCaP cells based on a genome-scale screen including 17,255 ORFs. The relative enzalutamide resistance of each ORF (x-axis) is plotted as a Z-score (y-axis) (raw data are included in Data S2). A positive Z-score indicates that the gene promotes resistance. *MDM4* and other hits are highlighted on the graph, with *MDM4* scoring as the strongest hit. D) Relative viability of LNCaP cells after transduction of CRISPR-Cas9 and sgRNAs targeting *MDM4* (2 guides, red) or control GFP (black). Data represents the mean $\pm$ SD of seven replicates (raw data are included in Data S3). e) Immunoblot confirming MDM4 gene deletion in LNCaP cells. HSP90 is a loading control. f) Sensitivity of LNCaP, PC3, and DU145 cells to RO-5963. Relative viability is shown at each indicated dosage of RO-5963. Data represents the mean $\pm$ SD of six replicates (raw data are included in Data S3)

16

Methods

We introduce P-NET, an artificial neural network with biologically informed, parsimonious architecture that accurately predicts metastasis in PrCa patients based on their genomic profiles. P-NET is a feedforward neural network with constraints on the nodes and edges. In P-NET, each node encodes some biological entity (e.g. genes and pathways) and each edge represents a known relationship between the corresponding entities. The constraints on the nodes allow for better understanding of the state of different biological components. The constraints on the edges allow us to use a large number of nodes without increasing the number of edges, which leads to a smaller number of parameters compared to fully connected networks with the same number of nodes, and hence potentially less computations. The architecture was built using the Reactome pathway data sets (*36*). The whole Reactome dataset was downloaded and processed to form a layered network of five layers of pathways, one layer of genes, and one layer for features. This sparse model had slightly over 71,000 weights with the number of nodes per layer distributed as shown in Fig S1.B. A dense network with the same number of nodes would have more than 270 million weights with the first layer containing more than 94% of the weights. A hybrid model which contains a sparse layer followed by dense layers still contains over 14 million weights. The number of dense weights is calculated as $w_l = n_l * (n_{l-1} + 1)$ where w_l is the number of weights per layer l and n_l is the number of nodes of the same layer.

The meaning of the nodes, layers, and connection of P-NET is encoded through a carefully engineered architecture and a set of restrictions on the connections of the network. The input layer is meant to represent features that can be measured and fed into the network. The second layer represents a set of genes of interest. The higher layers represent a hierarchy of pathways and biological processes that are manually curated. The first layer of P-NET is connected to the next layer via a set of one-to-one connections where each node in the next layer is connected to exactly three nodes of the input layer representing mutations, copy number amplification, and copy number deletions. This scheme results in a much smaller number of weights in the first layer compared to a fully connected network and the special pattern of the connection matrix results in more efficient training. The second layer is restricted to have connections reflecting the gene-pathway relationships as curated by the Reactome pathway dataset. The connections are encoded by a mask matrix *M* that is multiplied by the weights matrix *W* to zero-out all the connections that don't exist in the Reactome pathway dataset. For the next layers, a similar scheme is devised to control the connection between consecutive layers to reflect the real parent-child relationships that exist in the Reactome dataset. The output of each layer is calculated as $y = f[(M * W) x + b]$ where *f* is the activation function, *M* is the mask matrix, *W* is the weights matrix, *x* is the input matrix, and *b* is the bias vector (see Fig. S1.A). The activation of each node is kept into the range of [-1,1] by applying the tanh function $f = tanh = (e^{2x} - 1)/(e^{2x} + 1)$ to the weighted inputs of the node. The activation of the outcome layers is calculated by the sigmoid function $\sigma = 1/(1 + e^{-x})$.

To allow each layer to be useful by itself, we added a predictive layer after each hidden layer. P-NET has a smaller number of nodes per layer in the later layers compared to the first layers Fig. S1.B. Since it is more challenging to fit the data using a smaller number of weights in the later layers, we used a higher loss weight for later layer outcomes during the optimization process. The final prediction of the network was calculated by taking the average of all the layer outcomes, Fig S1.C. The learning rate was initialized to be 0.001 and actively reduced after every 50 epochs to allow for smooth convergence. Since we have an unbalanced dataset, we weighted the classes differently to reduce the network bias toward one class based on the bias in the training set. The model was trained using Adam optimizer (*37*) to reduce the binary cross-entropy loss functions $H = -\frac{1}{N} \Sigma \, y_i \, . \, log \, (p(y_i)) \, + \, (1 - y_i \,) . \, log \, (1 - p(y_i))$, where y_i is the label for sample i, $p(y_i)$ is the probability that sample i has a metastatic cancer as calculated using the sigmoid function σ, and N is the total number of samples. We checked different gradient based attribution methods to rank the features in all the layers, and we chose to use the DeepLIFT scheme as implemented in the DeepExplain library (*10*).

To check the utility of the developed model, we trained P-NET to predict cancer state (primary/metastatic) of prostate patients based on their genomic profiles. We used whole-exome sequencing of 1,013 patients along with the corresponding somatic mutations and copy number alterations (*4*). The mutations were aggregated on the gene level, excluding Silent, Intron, 3'UTR, 5'UTR, RNA, and lincRNA mutations (Fig. S2). The copy number alterations for each gene was assigned based on the called segment level copy number emphasizing focal amplifications and deletions and excluding single copy amplification and deletions. The prediction performance was measured using the average area under the ROC curve (AUC), the area under precision-recall curve (AUPRC), and F1 score. The corresponding measures were reported for the testing split and also for the cross-validation setup. The input data was divided into a testing set (10%) and a development set (90%). The development set was further divided into a validation set that has the same size as the testing set and the remaining samples are reserved for training. For the cross-validation experiments, the development data set was divided into 5 folds stratified by the label classes to account for the bias in the dataset. The implementation of the proposed system along with the reproducible results are available on Github (https://github.com/marakeby/pnet_prostate_paper).

Analysis of a genome-scale ORF screen
A genome-scale ORF screen was previously performed in LNCaP cells (33) . In brief, cells were infected with a pooled ORF library, subject to puromycin selection to isolate cells containing the respective ORFs, and then seeded in low androgen media (CSS) with enzalutamide. The relative effect of each ORF on cell proliferation was determined after 25 days in culture and is represented as Z-scores. We postulated that amplified genes identified by P-NET regulate oncogenic functions in metastatic castration resistant

prostate cancer. To validate this hypothesis, we analyzed this previously published genome-scale ORF screen performed in LNCaP cells which identified genes that, when overexpressed, promoted resistance to the AR inhibitor, enzalutamide (Figure 4C) (*33*). LNCaP cells are dependent on AR and treatment with enzalutamide attenuates cell proliferation. Based on this analysis, *MDM4* scored as a robust enzalutamide resistant gene relative to other hits, including cell cycle regulators (*CDK4, CDK6*) or those with roles in FGF signaling (*FGFR2, FGFR3, FGF6*); these are two pathways implicated in driving resistance to anti-androgen therapies in clinical prostate cancers (*38*, *39*).

Sensitivity to RO-5963
LNCaP, DU145, and PC3 cells were seeded in 12-well plates at 100k, 20k, or 20k, respectively. After 24 hours, cells were treated with increasing concentrations of RO-5963 between 80nM and 50µM. Media containing the inhibitor was refreshed after 3 days. Relative cell viability was determined using a Vi-Cell after 6 days of treatment, and cell counts were used to calculate IC_{50} values.

MDM4 gene depletion experiments
Blasticidin-resistant Cas9 positive LNCaP cells were cultured in 150µg/mL blasticidin (Thermo Fisher Scientific, NC9016621) for 72 hours to enrich cells with optimal Cas9 activity. 2 million cells were seeded in parallel 10cm plates and infected with lentiviruses expressing puromycin-resistant sgRNAs targeting *MDM4* or GFP control 24 hours later. Cells were then subject to puromycin selection for 4 days, at which point one plate was harvested for immunoblotting and the other was counted using a Vi-Cell and seeded for a proliferation assay. 7 days later, cells were counted again with a Vi-Cell to assess viability, representing a total of 12 days. The target sequence against GFP was CACCGGCCACAAGTTCAGCGTGTCG (sgGFP). The target sequences against *MDM4* were AGATGTTGAACACTGAGCAG (sgMDM4-1) and CTCTCCTGGACAAATCAATC (sgMDM4-2).

Immunoblotting
Cell pellets were lysed in RIPA buffer (MilliporeSigma, 20-188) containing Protease/Phosphatase Inhibitor Cocktail (Cell Signaling Technology, 5872S). Protein concentrations were calculated using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, PI23225), and protein was then denatured in NuPAGE LDS sample buffer (Thermo Fisher Scientific, NP0007) with 5% β-Mercaptoethanol. 13µg of each protein sample was electrophoresed using NuPAGE 4-12% Bis-Tris Protein gels (Thermo Fisher Scientific) and run with NuPAGE MOPS SDS Running Buffer (Thermo Fisher Scientific, NP0001). Proteins were transferred to nitrocellulose membranes using an iBlot apparatus (Thermo Fisher Scientific). Membranes were blocked in Odyssey Blocking Buffer (LI-COR Biosciences, 927-70010) for one hour at room temperature, and membranes were then cut and incubated in primary antibodies diluted 1:1000 in Odyssey Blocking Buffer at 4°C overnight. The following morning, membranes were washed with Phosphate-Buffer Saline, 0.1% Tween (PBST) and incubated with fluorescent anti-rabbit secondary antibodies (Thermo Fisher Scientific, NC9401842) for one hour at room temperature. Membranes were again washed with PBST and then imaged using an Odyssey Imaging

19

System (LI-COR Biosciences). Primary antibodies used include MDM4 (Thermo Fisher Scientific, A300287A) and HSP90 (Cell Signaling Technology, 4874S).
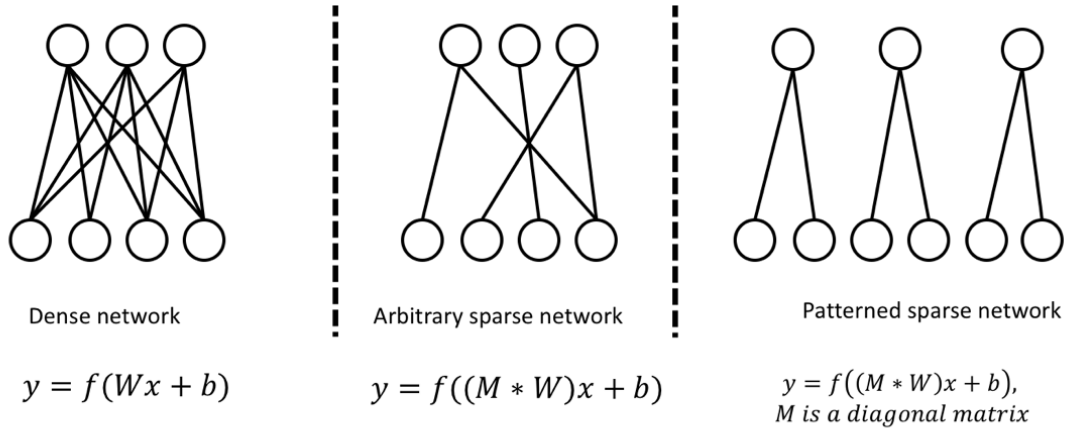
Chemical inhibition of MDM4 reduces prostate cancer cell viability

Given the proposed role that *MDM4* plays in driving enzalutamide resistance in prostate cancer cells, we sought to determine the response of prostate cancer cells to chemical inhibition of *MDM4*. We evaluated RO-5963, a small molecule *MDM2/4* dual inhibitor with the greatest selectivity towards MDM4 in its class (*40*). This drug has previously demonstrated robust efficacy against *MDM4* dependent cancer cell lines (*41*). We evaluated the effects of increasing concentrations of RO-5963 on prostate cancer cell proliferation.

Gene depletion of *MDM4* reduces prostate cancer cell viability

To determine how prostate cancer cells would respond to precision tools that target *MDM4* at the gene level, we utilized CRISPR-Cas9 and two sgRNAs targeting distinct sequences of *MDM4* in LNCaP cells. In comparison to a negative control sgRNA (GFP), viability of LNCaP cells was reduced by about 60% (Figure 4D) in response to *MDM4* depletion (Figure 4F) after 12 days in culture. Altogether, we concluded that *MDM4* regulates enzalutamide resistance, and that targeting *MDM4* through either chemical or genetic approaches significantly attenuated the viability of prostate cancer cell lines. Our observations indicate that antagonizing *MDM4* in metastatic castration resistant prostate cancers that harbor wild-type *p53* is an attractive precision strategy.

A

Dense network

$$y = f(Wx + b)$$

Arbitrary sparse network

$$y = f((M * W)x + b)$$

Patterned sparse network

$$y = f((M * W)x + b),$$
$$M \text{ is a diagonal matrix}$$

B

C
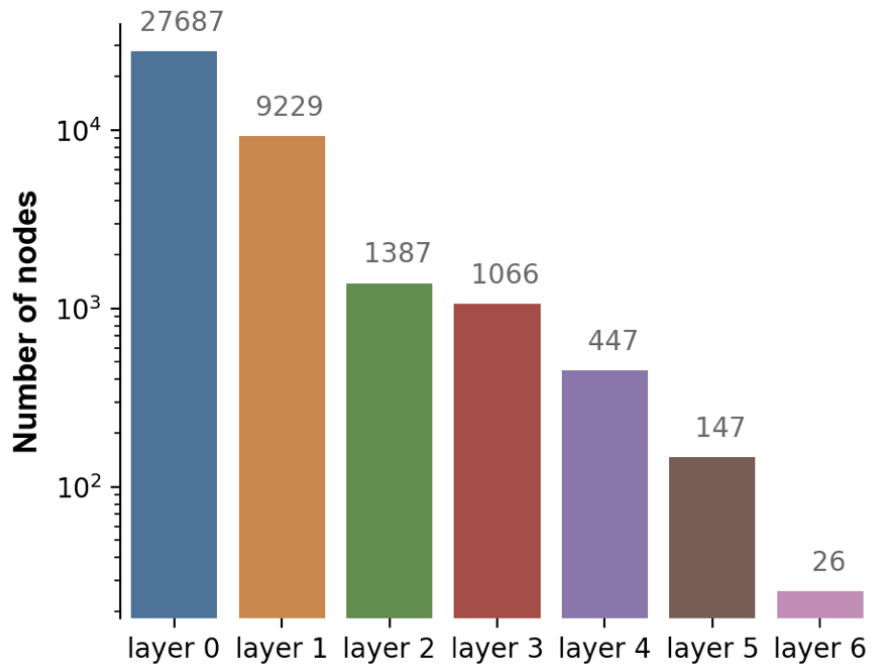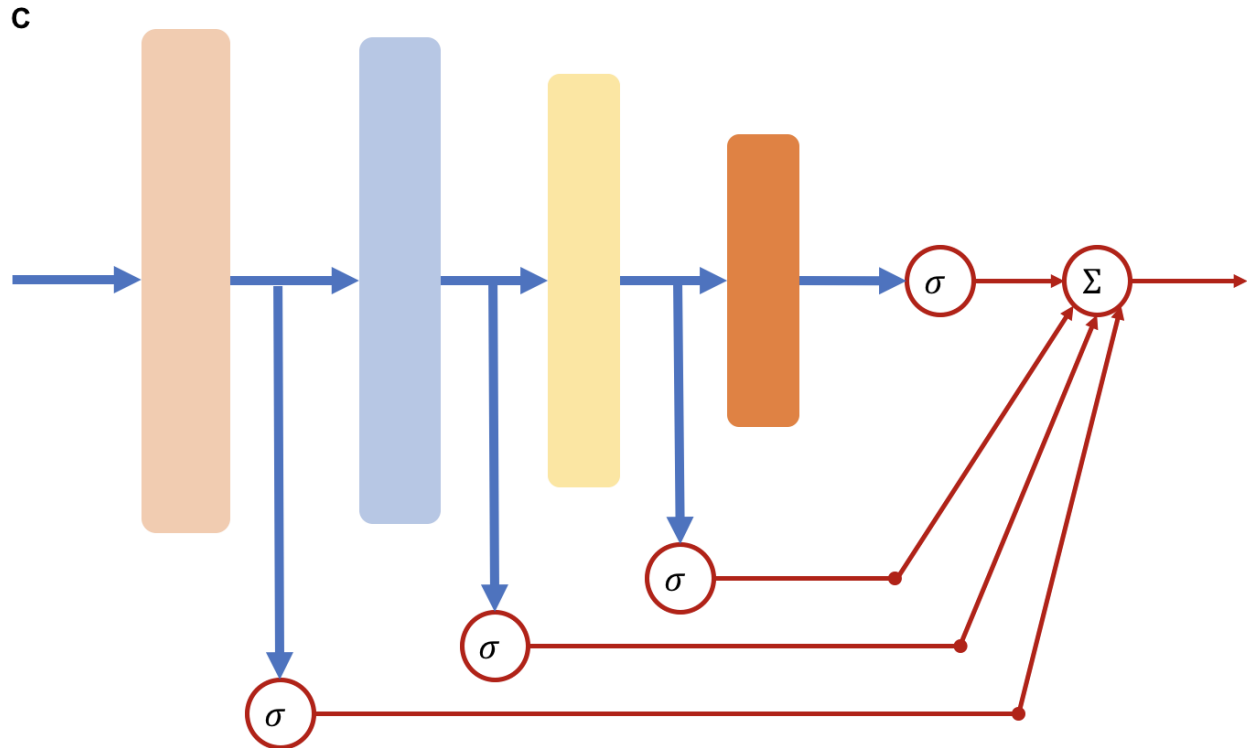
Fig. S1. A) The difference between dense and sparse layers. Sparse layers can be arbitrary sparse or patterned sparse. Arbitrary sparse layers are flexible to encode any connection scheme. Patterned sparse layers can make computations more efficient. B) The number of nodes in each layer of the developed P-NET showing a decreasing number of nodes for higher layers. C) A predictive node is connected to each hidden layer in P-NET, and the final prediction is calculated by taking the average of all the predictive elements in the network
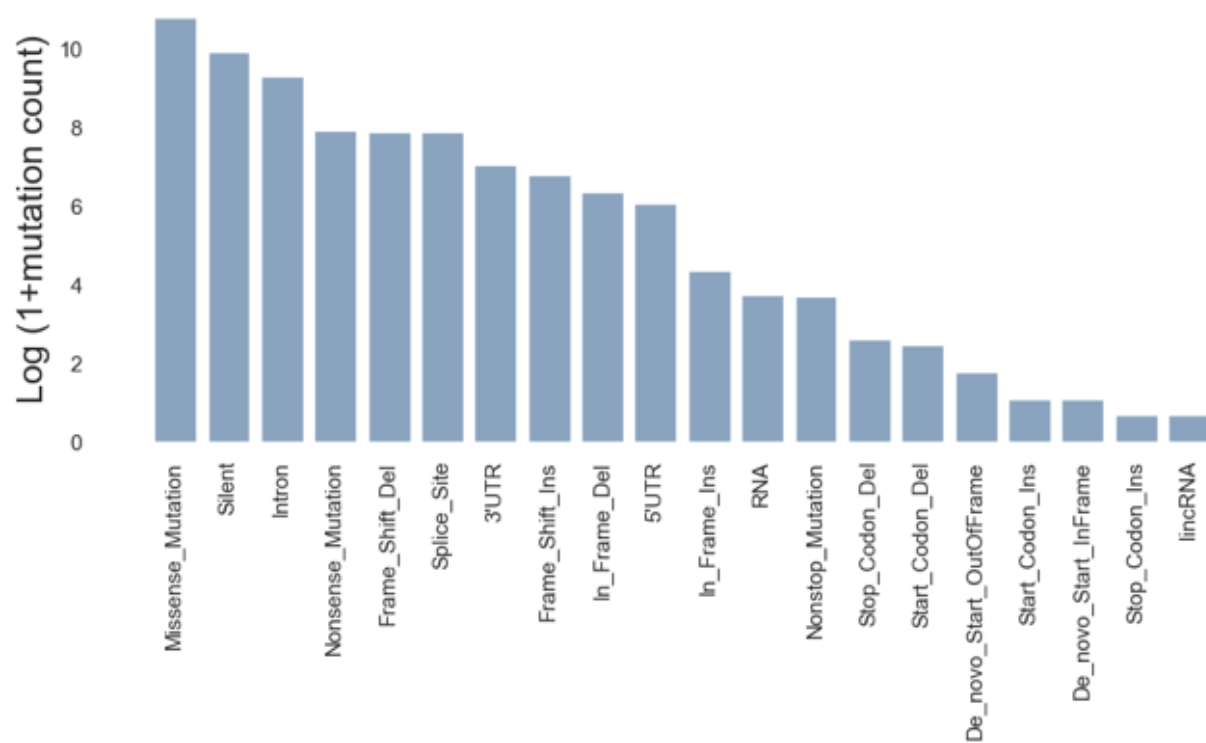
Fig. S2.

The distribution of mutation types in the P100 dataset. Silent, Intron, 3'UTR, 5'UTR, RNA, and lincRNA mutations are excluded from the modeling.
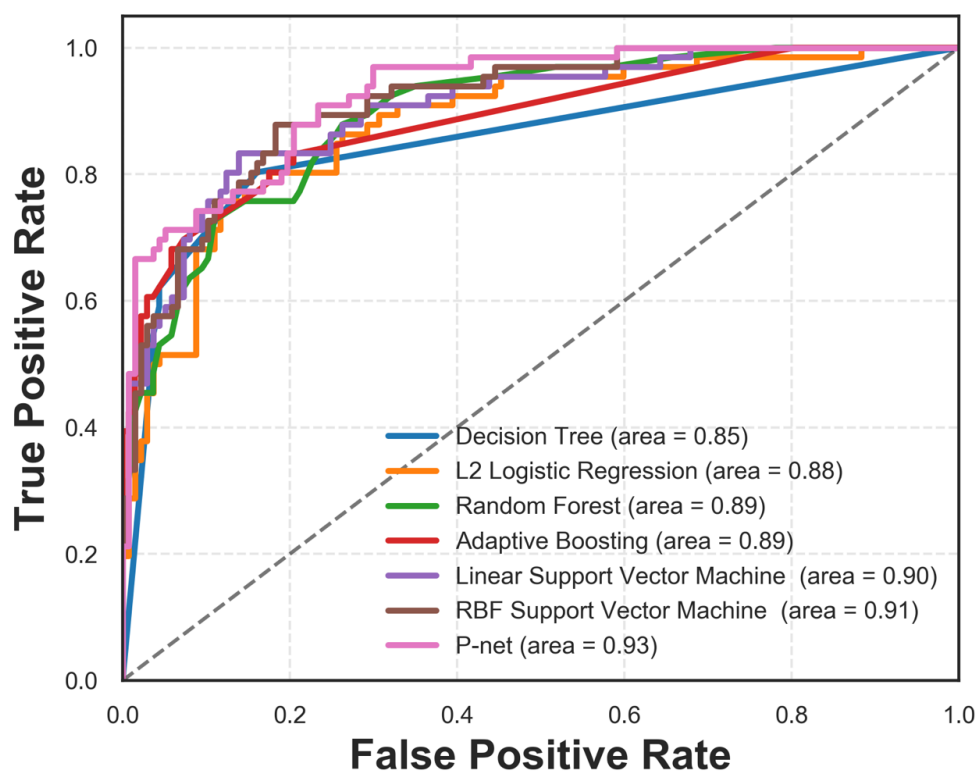
Fig. S3. The area under the ROC curve of different models when trained and tested on the P1000 dataset. P-NET outperforms other models leading to higher AUC.
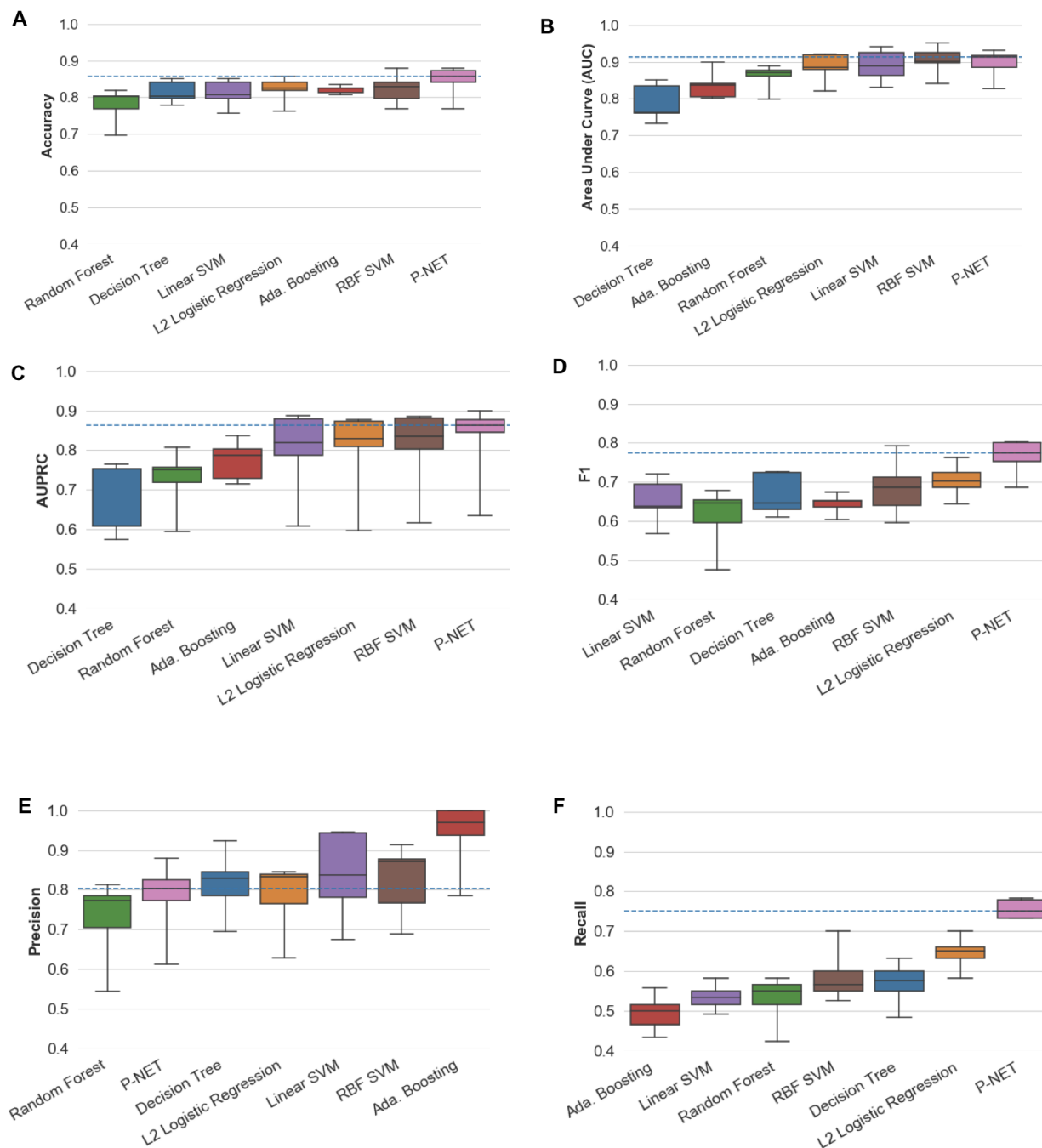
Fig. S4. Cross-validation experiment using 5 folds over the development split of the data. P-NET outperforms other models on average using all the metrics (A: Accuracy, B: Area under curve, C: AUPRC, D: F1, and F: recall) except Precision (E).
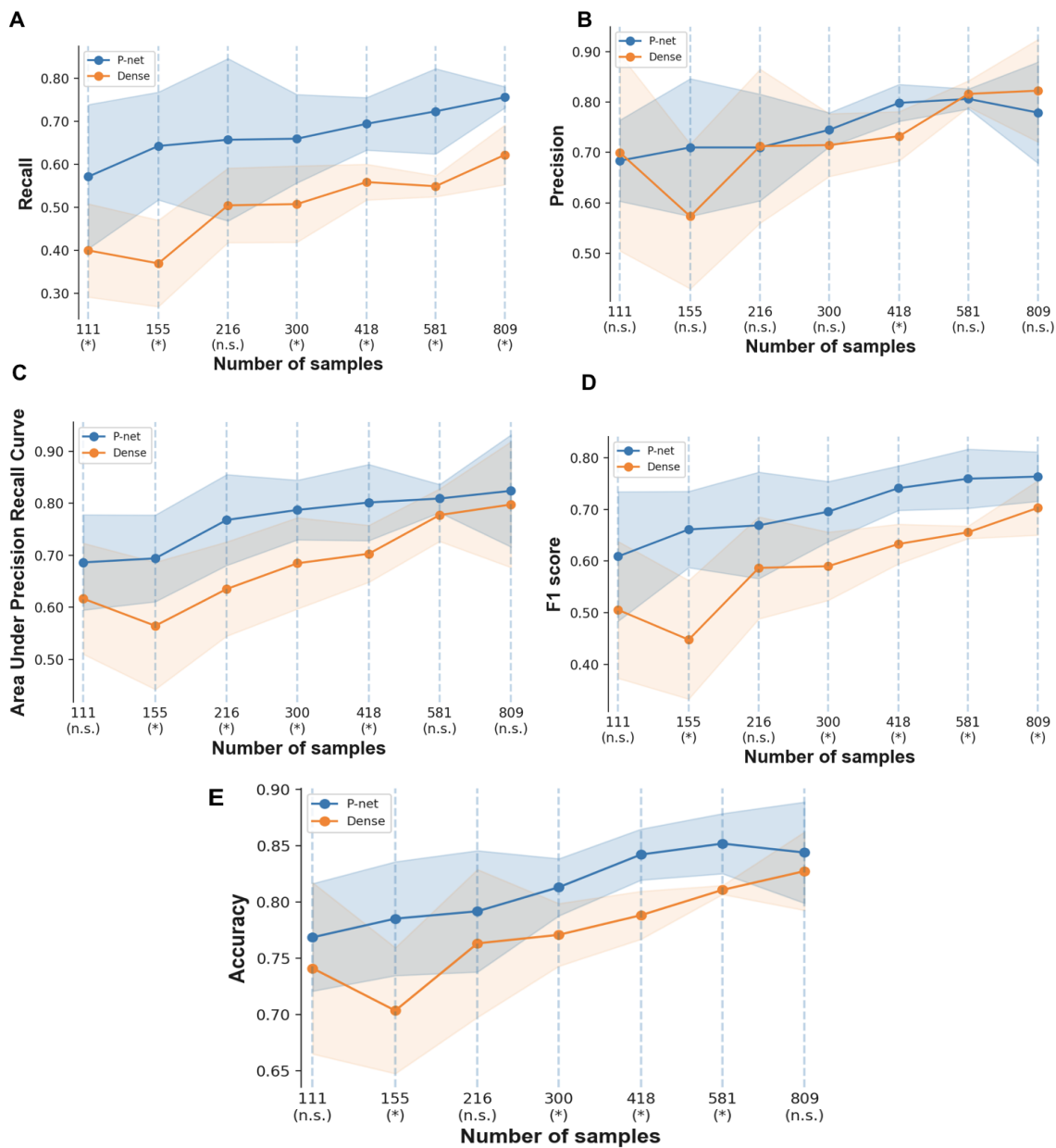
Fig. S5. Comparing the performance of P-NET to a dense network with the number of parameters using different sizes of training sets (A: Recall, B: Precision, C: AUPRC, D:

F1, E: Accuracy). Sample sizes marked by (*) indicate statistically significant differences while those marked by (n.s.) are not.

A

B

C

D

Layer4

Generic Transcription Pathway
Deubiquitination
SUMOylation
RHO GTPase Effectors
M Phase
Class I MHC mediated antigen processing & presentation
PIP3 activates AKT signaling
Mitotic G2-G2/M phases
Cellular Senescence
Eukaryotic Translation Initiation
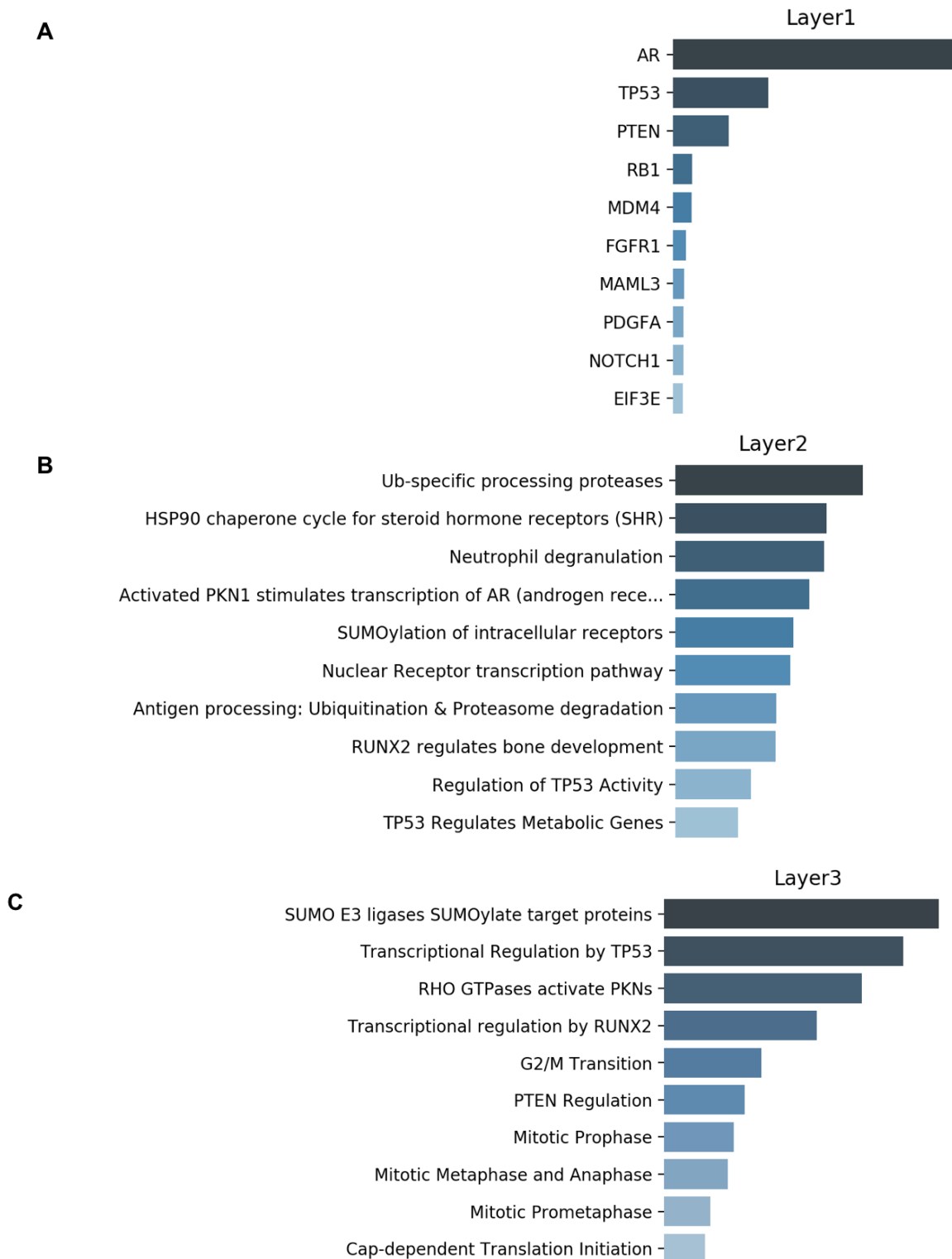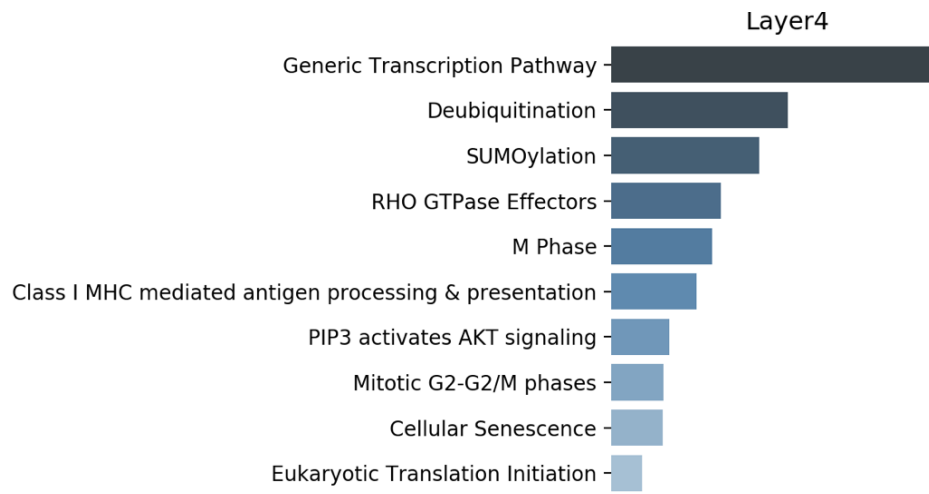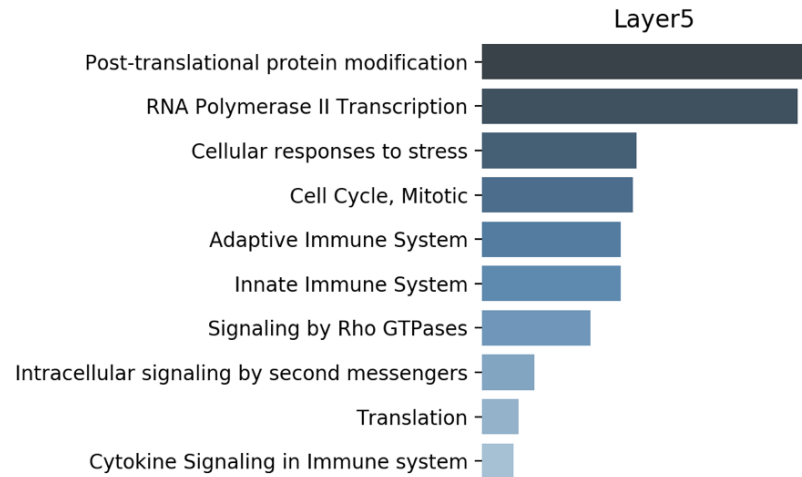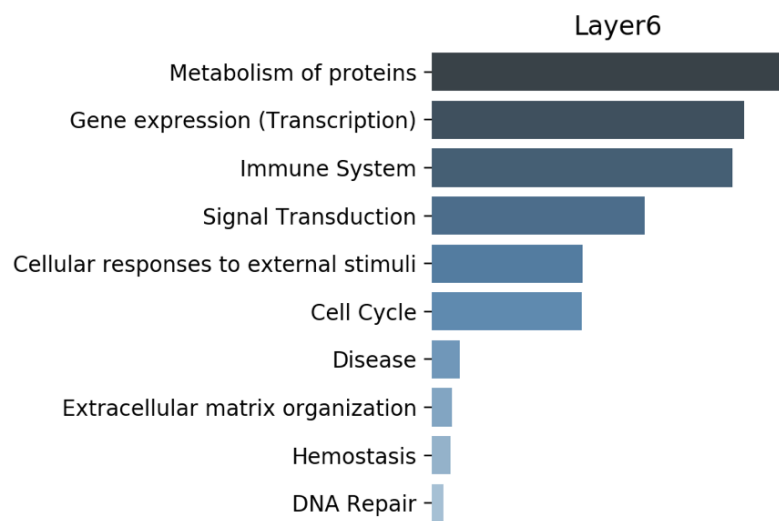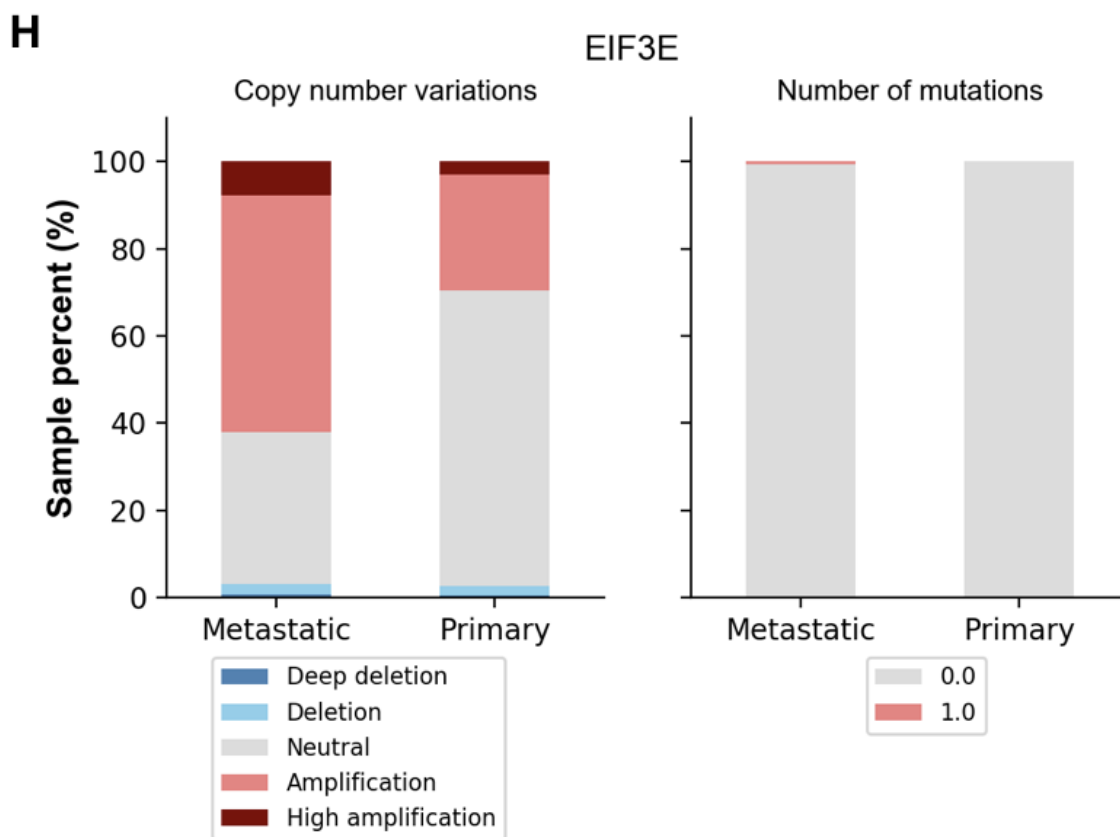
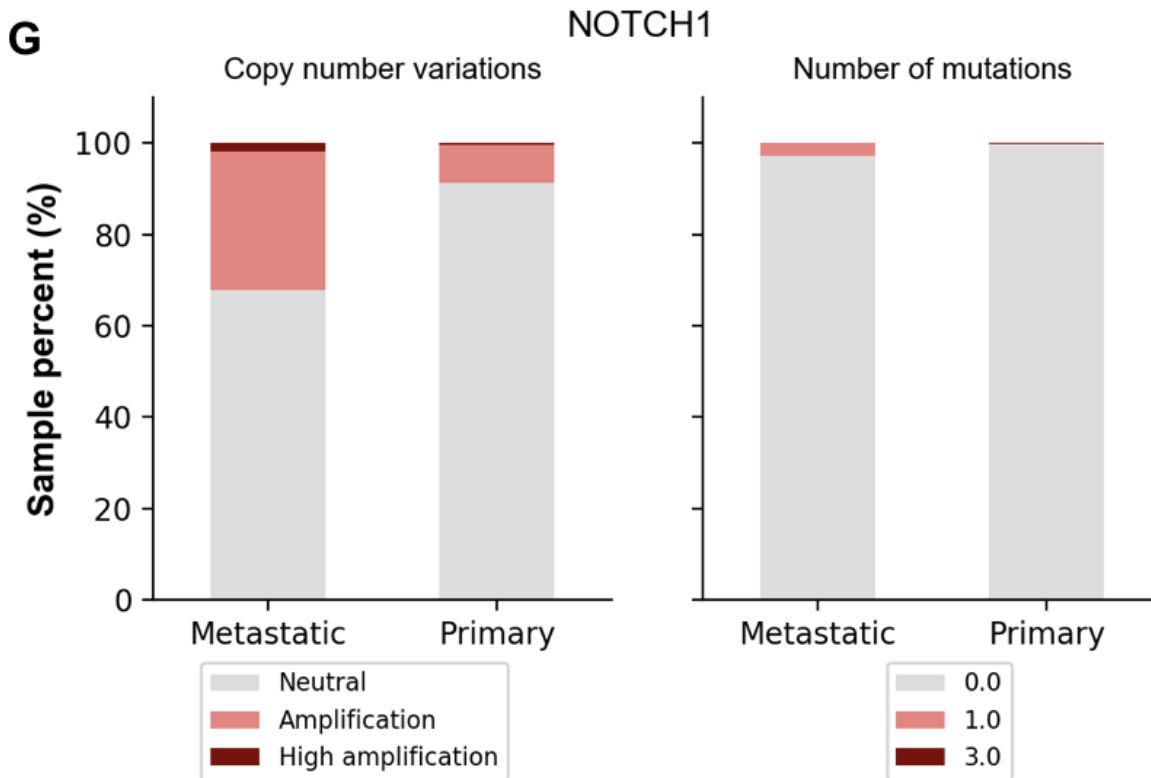**Fig. S6.** Relative ranking of nodes in each layer
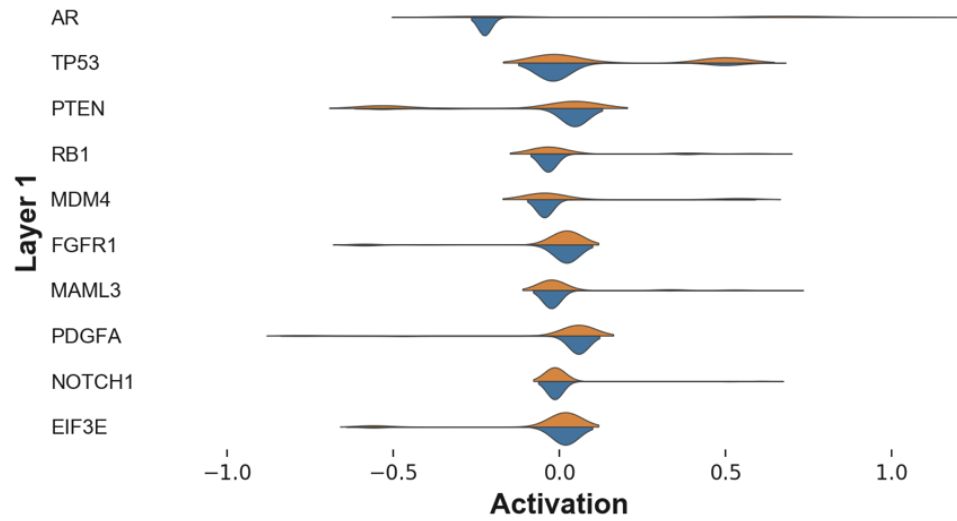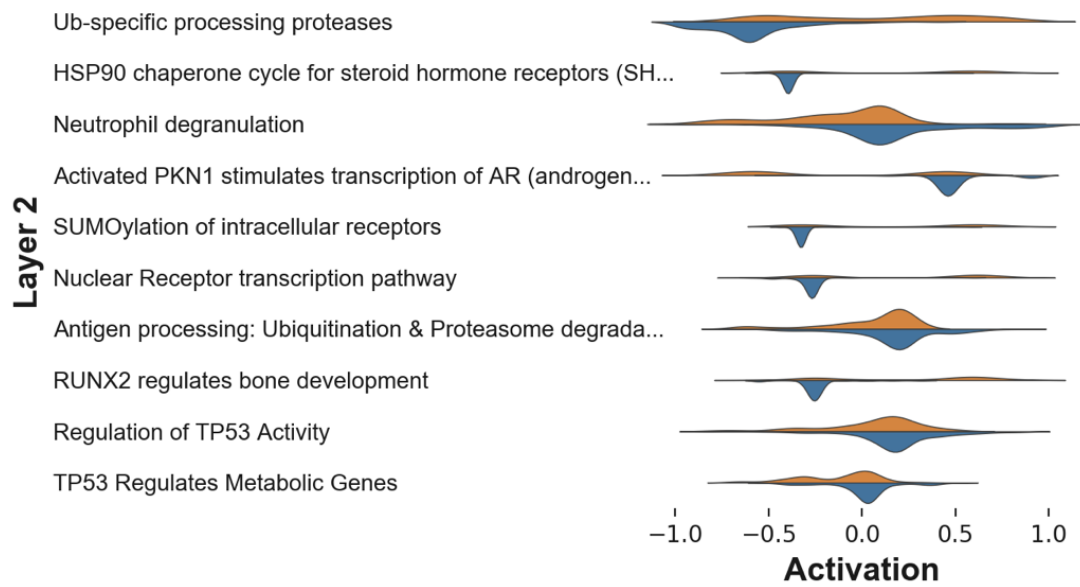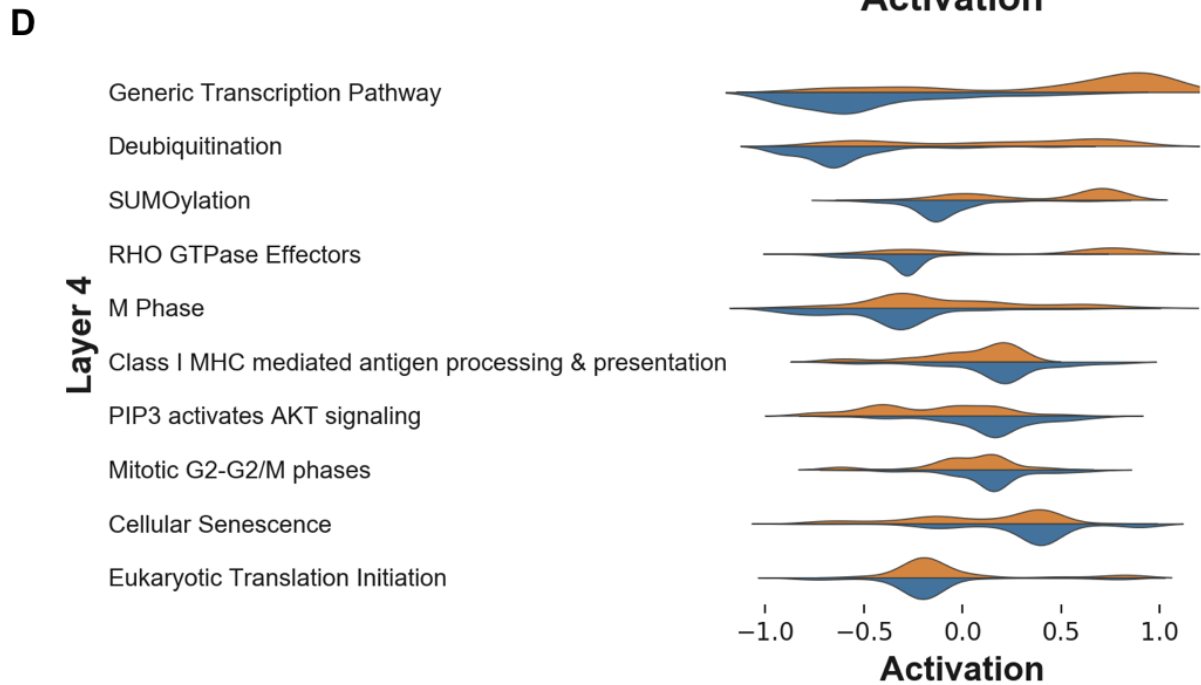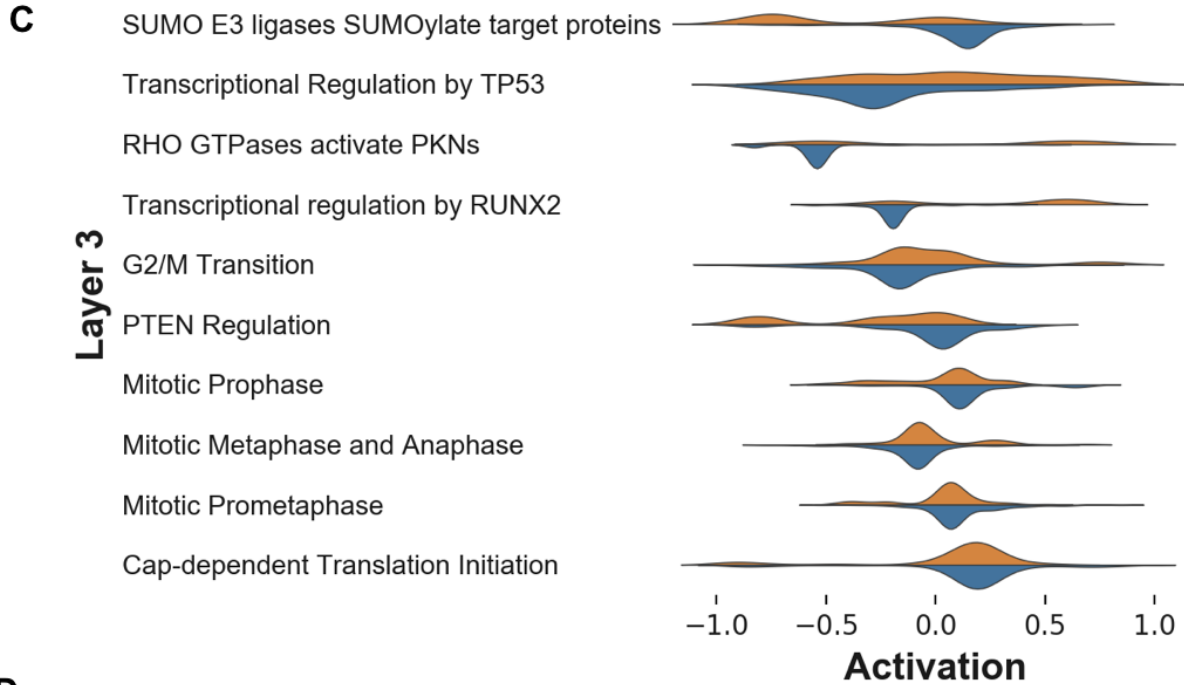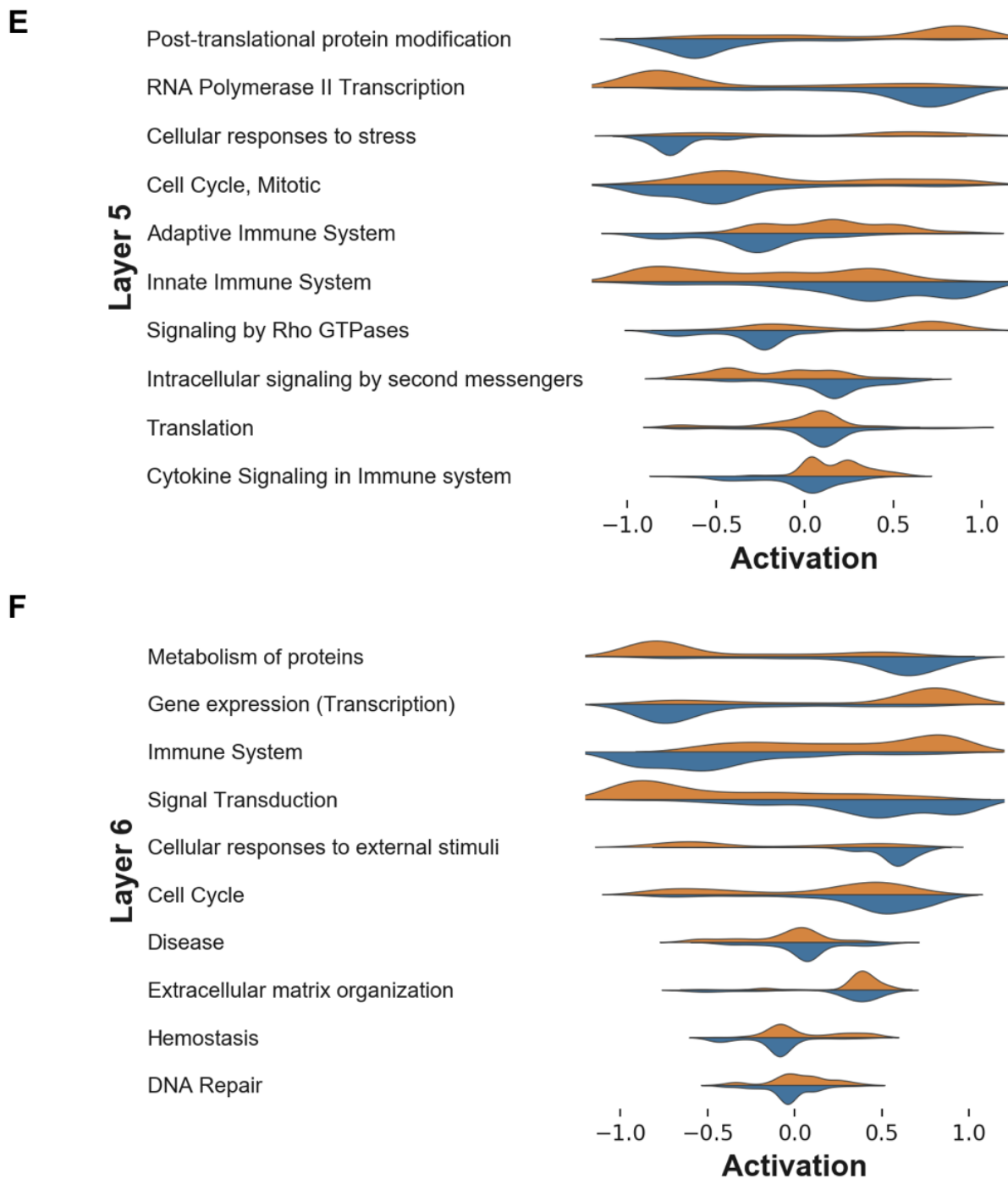
Fig. S7. Activation of top ranked nodes in each layer showing better discrimination between sample classes (Primary- blue vs. Metastatic-orange) in higher layers compared lower layers and in top ranked nodes compared to lower ranked ones.
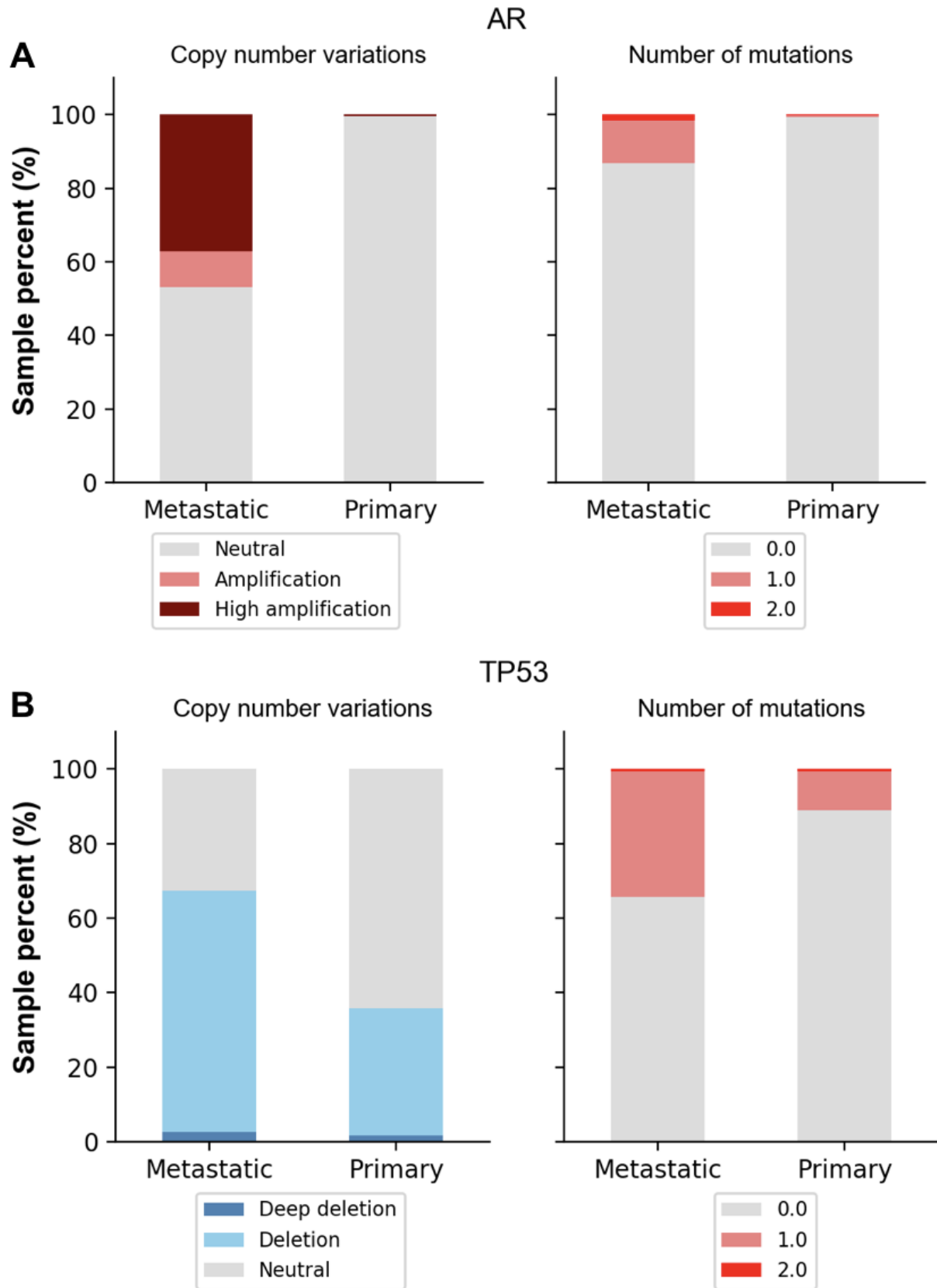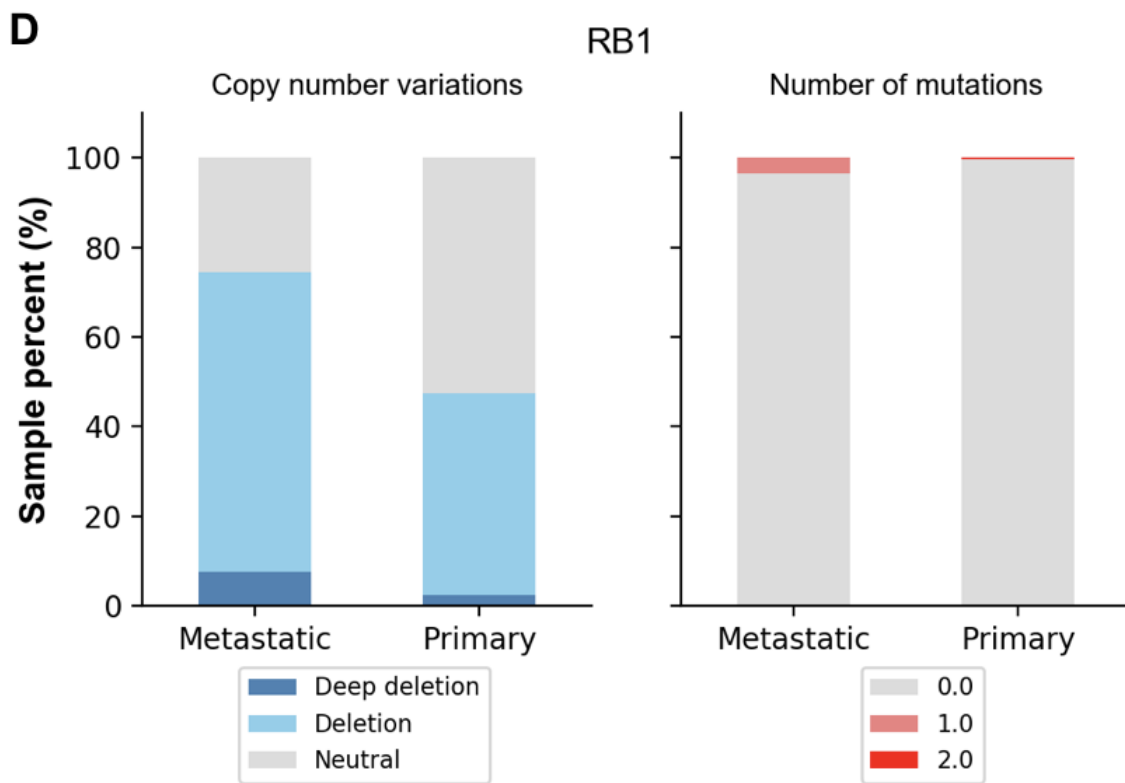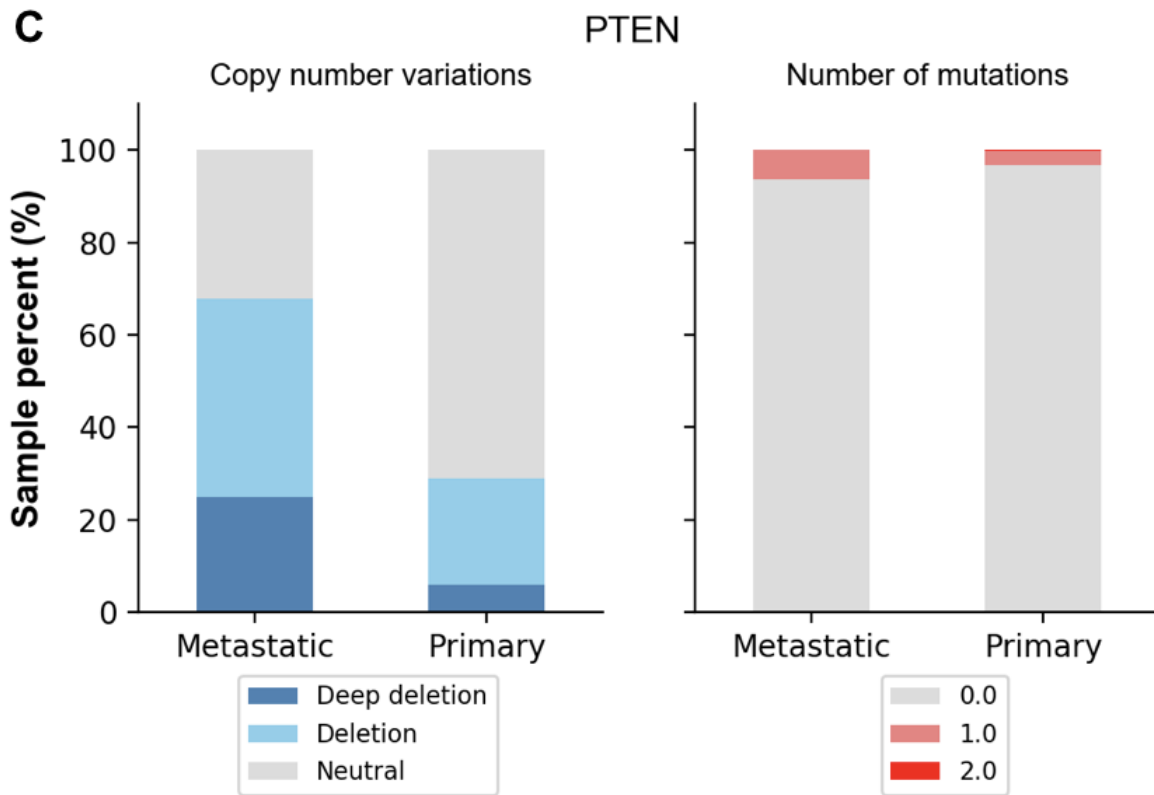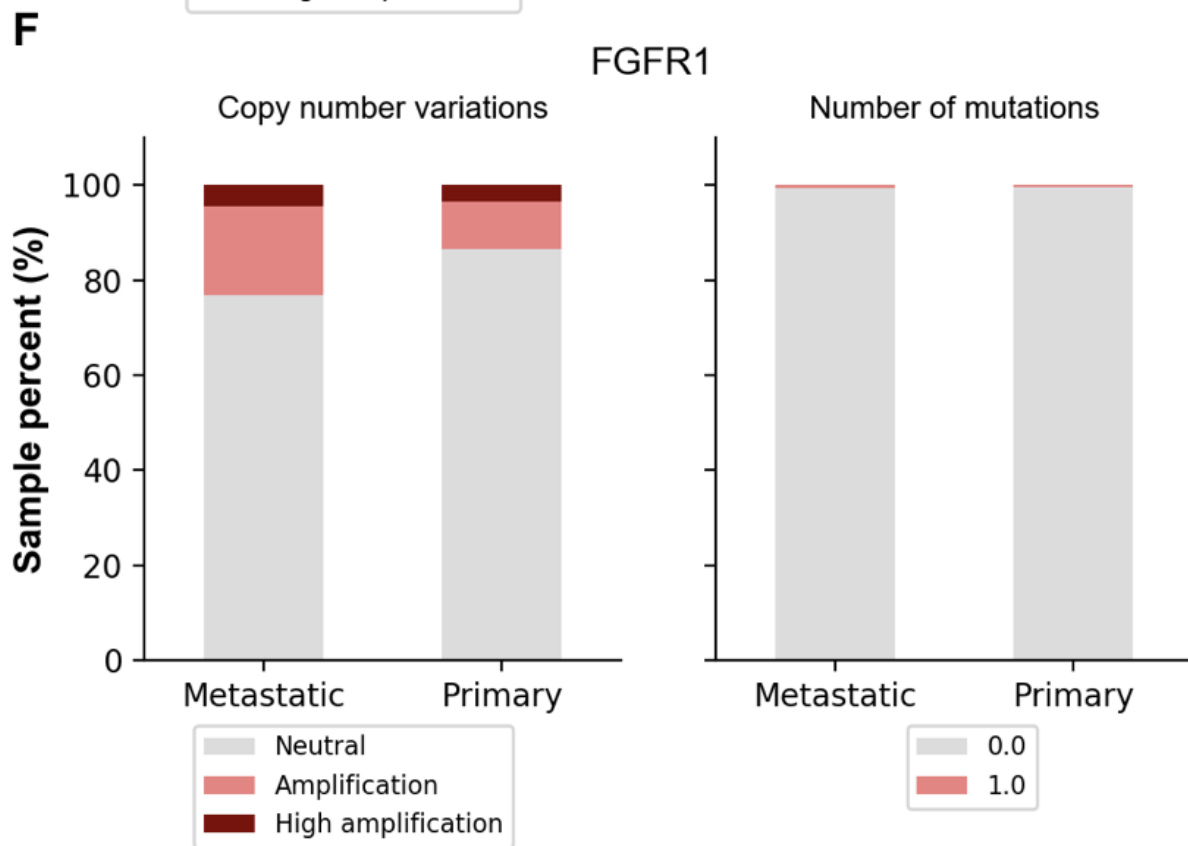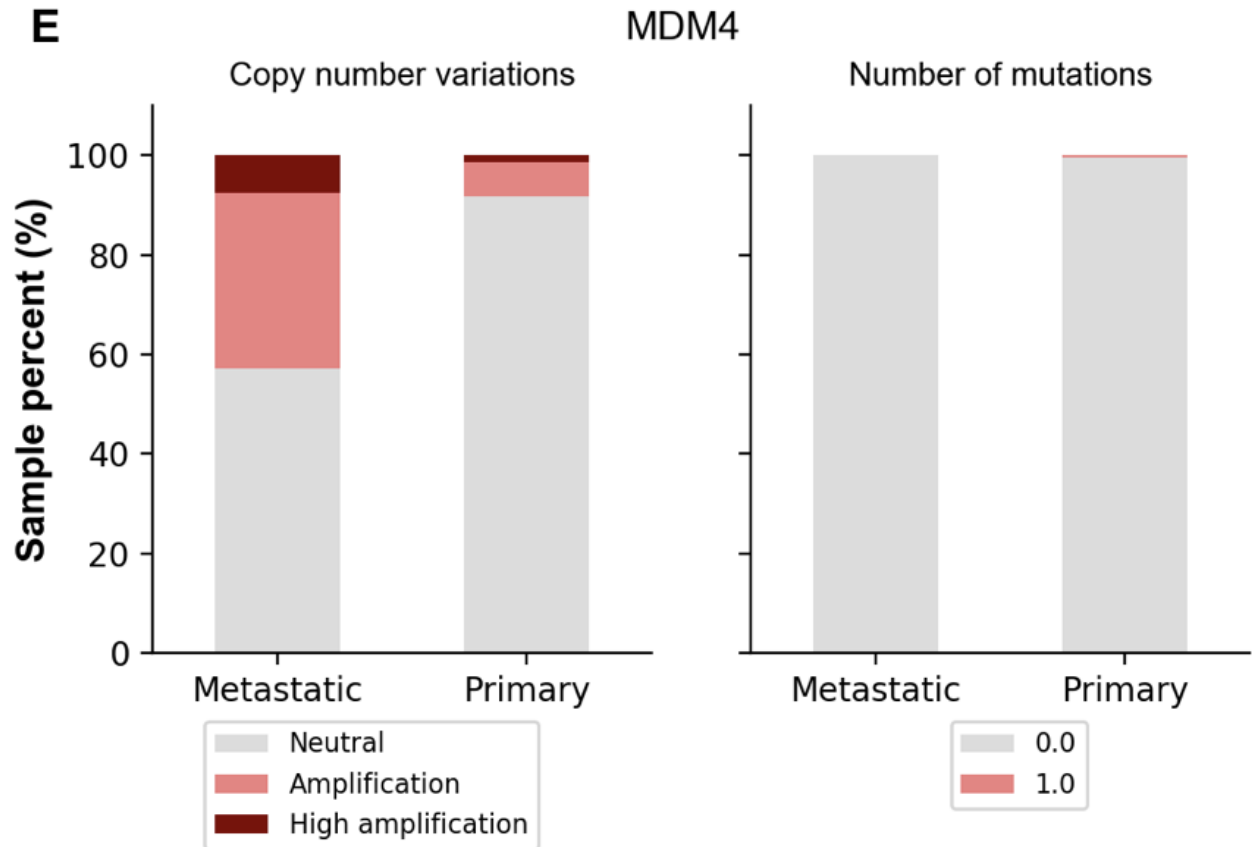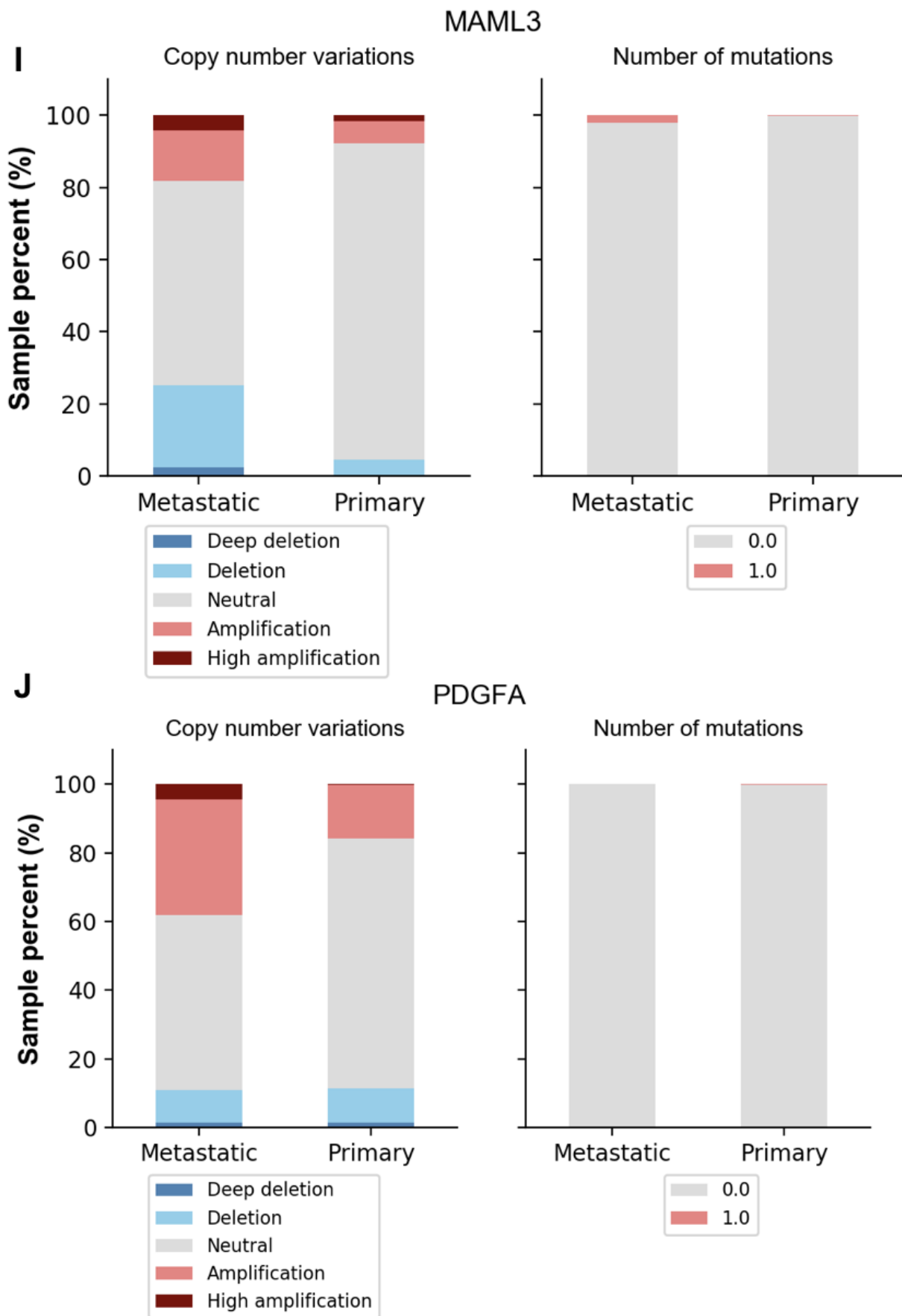
Fig. S8. The distribution of mutations and copy number variants of top ranked genes stratified by the class of the samples (Primary vs. Metastatic).