BIBS seminar

Biologically informed deep neural network for prostate cancer discovery

2024. 1. 25

Sangyeon Shin

Article

Biologically informed deep neural network for prostate cancer discovery

https://doi.org/10.1038/s41586-021-03922-4

Received: 6 November 2020

Accepted: 17 August 2021

Published online: 22 September 2021

Open access

Check for updates

Haitham A. Elmarakeby[1,2,3], Justin Hwang[4], Rand Arafeh[1,2], Jett Crowdis[1,2], Sydney Gang[1], David Liu[1,2], Saud H. AlDubayan[1,2], Keyan Salari[1,2,5], Steven Kregel[6], Camden Richter[1], Taylor E. Arnoff[1,2], Jihye Park[1,2], William C. Hahn[1,2] & Eliezer M. Van Allen[1,2]✉

The determination of molecular features that mediate clinically aggressive phenotypes in prostate cancer remains a major biological and clinical challenge[1,2]. Recent advances in interpretability of machine learning models as applied to biomedical problems may enable discovery and prediction in clinical cancer genomics[3–5]. Here we developed P-NET—a biologically informed deep learning model—to stratify patients with prostate cancer by treatment-resistance state and evaluate molecular drivers of treatment resistance for therapeutic targeting through complete model interpretability. We demonstrate that P-NET can predict cancer state using molecular data with a performance that is superior to other modelling approaches. Moreover, the biological interpretability within P-NET revealed established and novel molecularly altered candidates, such as *MDM4* and *FGFR1*, which 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.

- Published in nature, 2021
- Citation: about 170

Contents

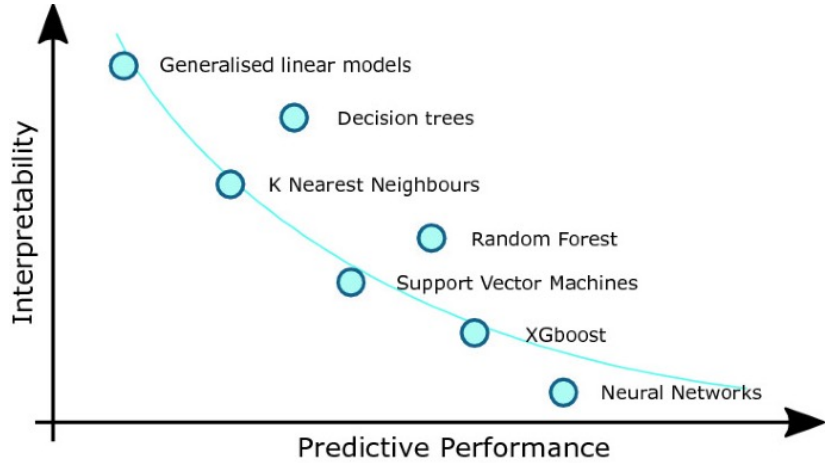
Introduction

Interpretability of predictive models

- Due to the advancement of molecular profiling technologies, **discovery of individual genes, pathways, and complexes** that promote cancer have been enabled.
- However, the **relationships** between molecular features and their biological contributions **remain uncharacterized**.
  - Disease progression, Drug resistance, and Lethal outcomes
- In translational cancer genomics, interpretability of predictive models is critical,
  - Patient care
  - Insights into the underlying biological processes
  - Functional investigation and therapeutic targeting

Limitations of existed deep learning model

- Trade-offs of accuracy and interpretability:
  - Linear regression model: high interpretability, low accuracy
  - Deep learning model: low interpretability, high accuracy
- Fully connected dense deep learning approach
  - Overfitting, computationally expensive, and less interpretable
- Sparse model can(rather than dense model),
  - Decrease storage requirements
  - Improve computational performance
- Tools that enhance the deep learning explainability:
  - LIME, DeepLIFT, DeepExplain, SHAP



Giovanni Ciatto, Michael Ignaz Schumacher, Andrea Omicini, and Davide Calvaresi. 2020. Agent-Based Explanations in AI: Towards an Abstract Framework.

Castration-resistant prostate cancer(CRPC)

- The term CRPC was initially proposed by the Prostate Cancer Working Group 2 in 2008.
- Define the state of prostate cancer in an environment with very **low serum testosterone concentration**.
  - a serum testosterone concentration maintained below 50 ng/dL or 1.7 nmol/dL.
- The paper classified:
  - **Castration-resistant metastatic v.s. Primary prostate cancers**

Morote J, Aguilar A, Planas J, Trilla E. Definition of Castrate Resistant Prostate Cancer: New Insights. Biomedicines. 2022 Mar 17;10(3):689. doi: 10.3390/biomedicines10030689. PMID: 35327491; PMCID: PMC8945091.

Preview of P-NET

- **Sparse** deep learning architecture
- Encode **biological information**
- Incorporate **explainability** algorithms
- Achieve **superior predictive performance** compared with established models
- Reveal **novel patterns of treatment resistance** in prostate cancer with translational implications

Materials and Methods

Materials

- 1,013 prostate cancers
  - Armenia, et al.(2018, *Nat Genet*)
  - CRPC(n = 333) and Primary cancer(n = 680)
    - Somatic mutation and copy number data
    - RNA sequence data for secondary analysis
- External dataset from
  - Fraser, et al.(Primary cancer, n = 277)
  - Robinson, et al.(CRPC, n = 500)
- Reactome pathway(a set of 3,007 pathway)
  - Fabregat, et al.(2018, *Nucleic Acids Res*)

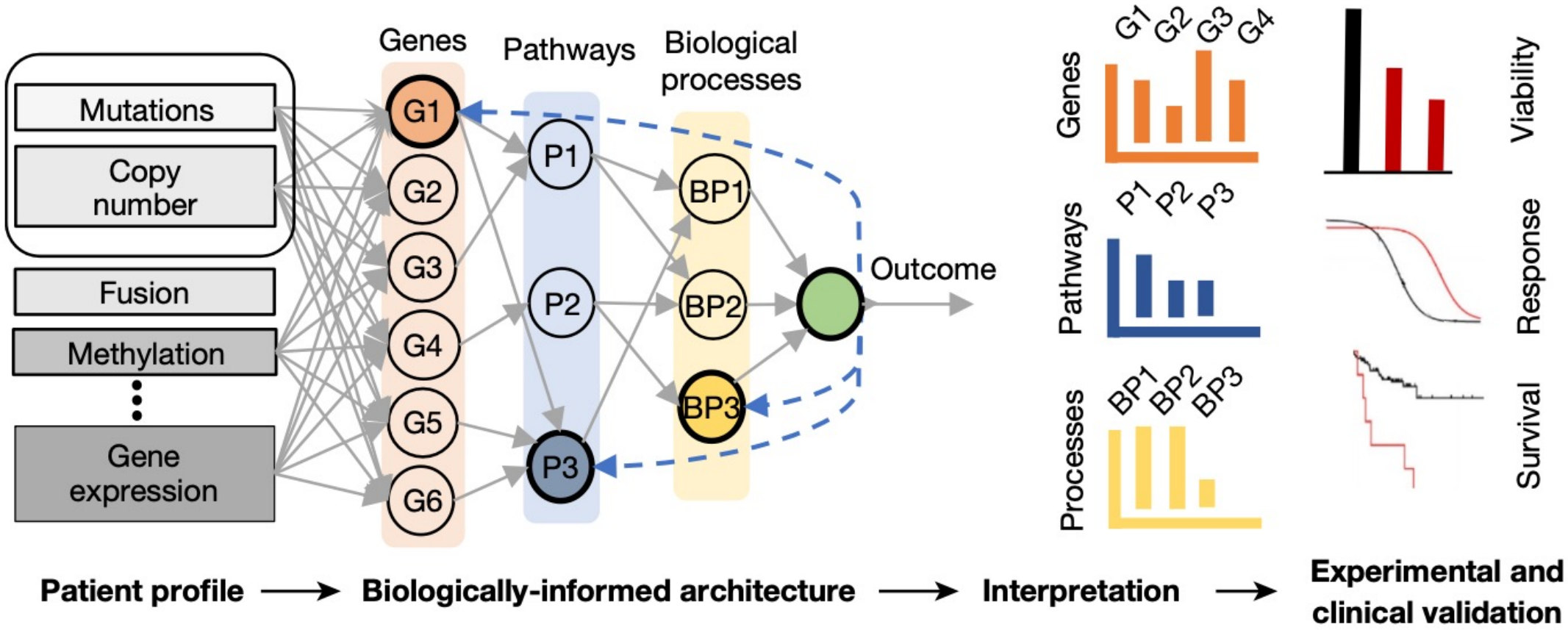
Processing of Somatic mutation and CNV

- The mutations were aggregated on the gene level with focus on non-synonymous mutations.
- Use prostate cancer whole-exome datasets,
  - excluding silent, intron, 3′ untranslated region (UTR), 5′ UTR, RNA and long intergenic non-coding RNA (lincRNA) mutations
- The copy number alterations for each gene were assigned on the basis of the called segment-level copy number as defined by GISTIC2.0
  - emphasizing high gains and deep deletions
  - excluding single-copy amplification and deletions

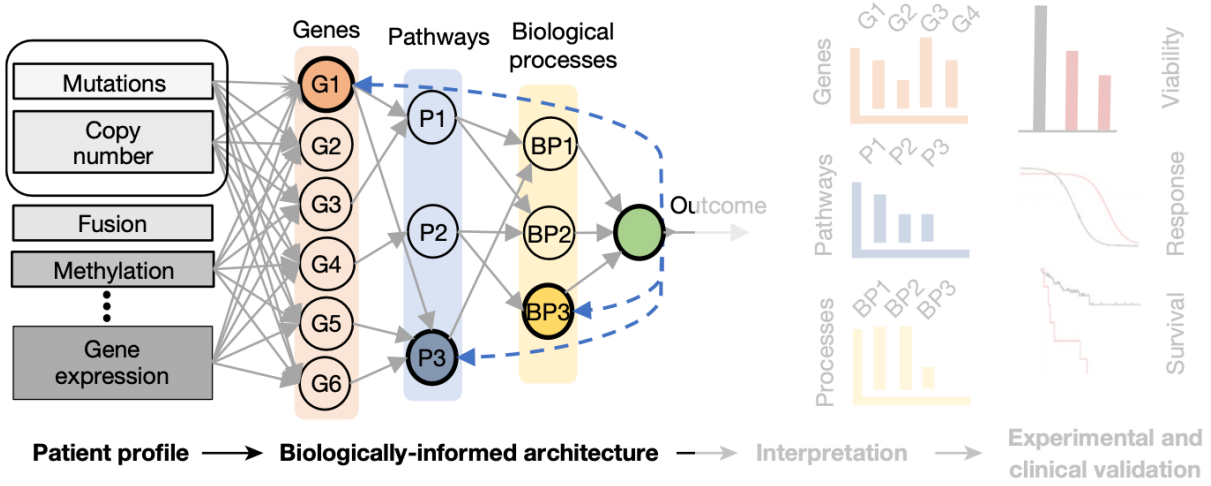
Processing of RNA sequencing data

- For secondary analyses involving RNA data, bulk whole transcriptomes from the subset of the Armenia et al. cohort.
  - n = 455 from TCGA, n = 204 from SU2C-PCF consortia
- Adapters were trimmed with cutadapt v2.2
- Reads were aligned using STAR aligner v2.7.2b
- STAR-aligned bam files were passed into RSEM to quantifications

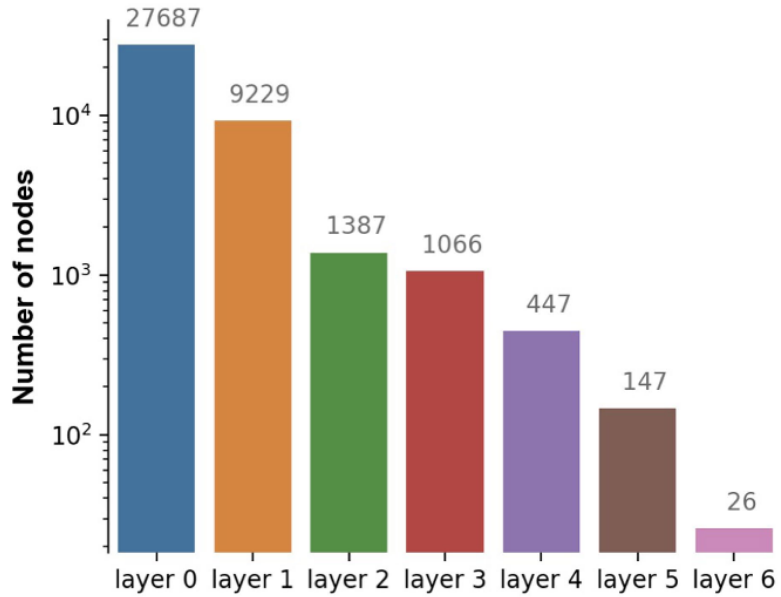
P-NET Design


Genes
Pathways
Biological processes
Mutations
Copy number
Fusion
Methylation
Gene expression
G1
G2
G3
G4
G5
G6
P1
P2
P3
BP1
BP2
BP3
Outcome
Genes
Pathways
Processes
Viability
Response
Survival
Patient profile
Biologically-informed architecture
Interpretation
Experimental and clinical validation

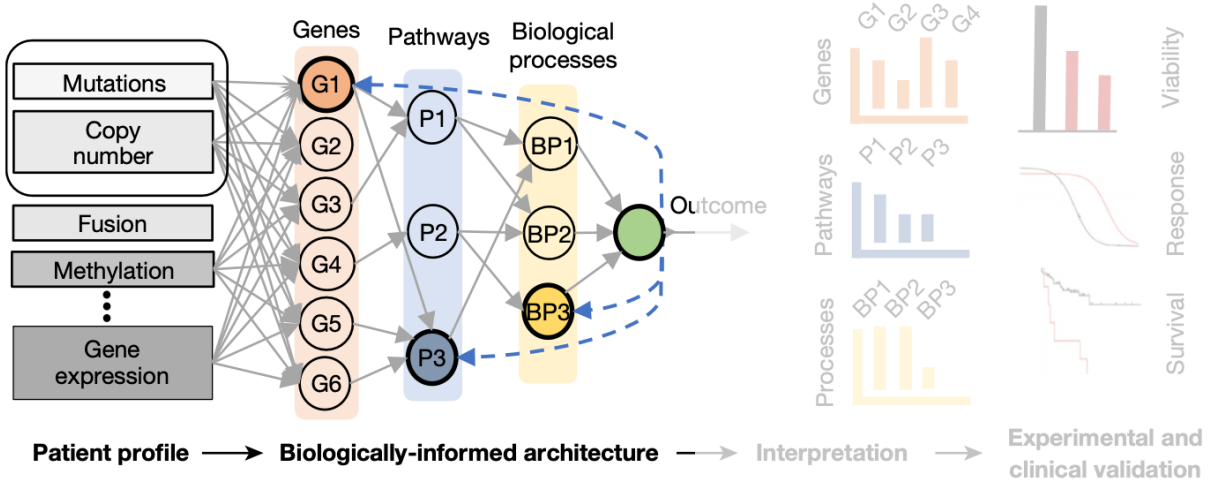
Architecture



- Built using the Reactome pathway datasets(Fabregat, et al.)
- Constraints on the nodes and edges
  - Nodes: Encode biological entity(genes, pathway)
  - Edges: Known relationship between the entities
- Layers: 5 layers of pathways, 1 layer of genes
- About 71,000 weights → Sparse
  - Dense model has 270 million weights with the first layer



Layers



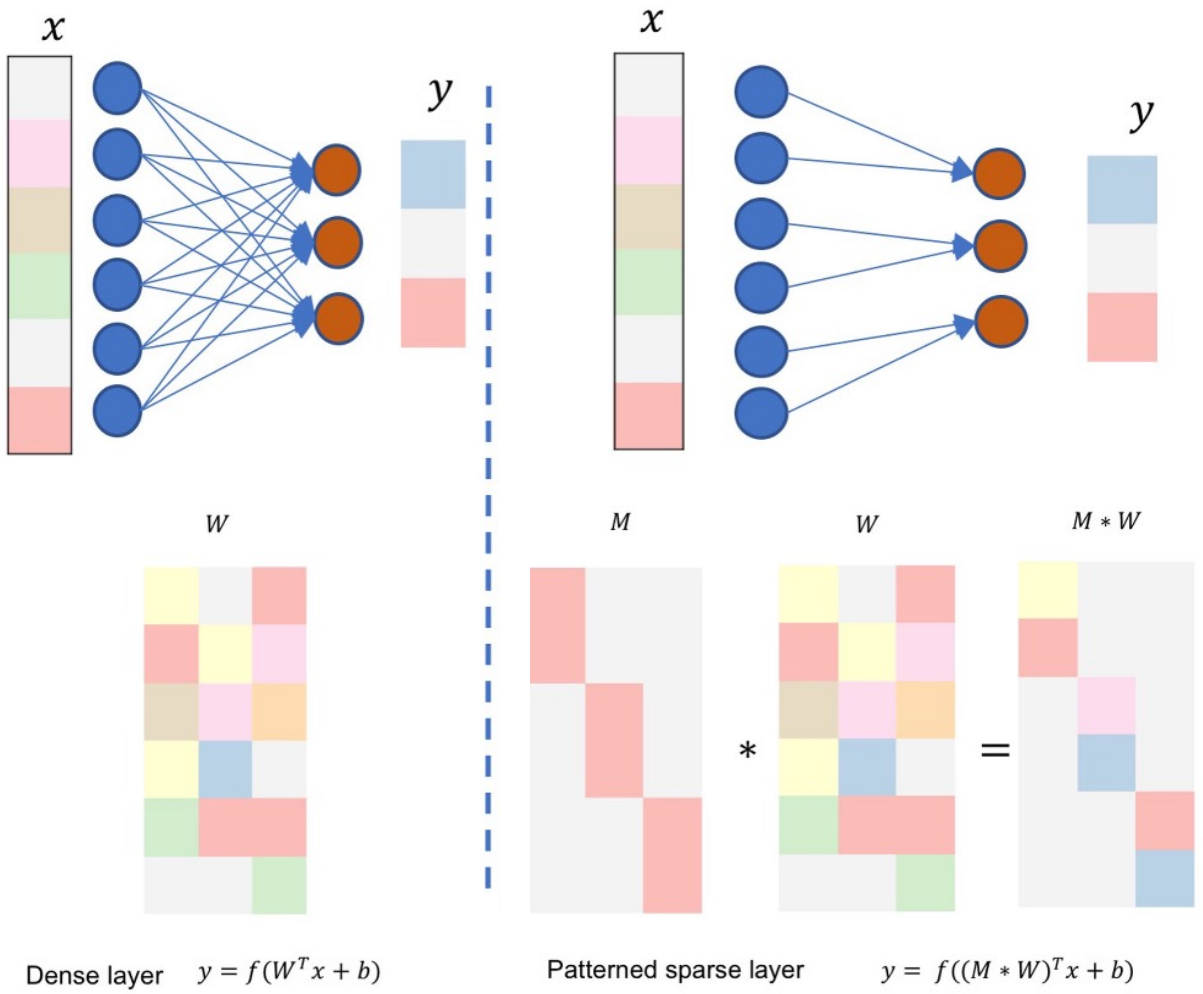
- 1 layer: a set of genes
  - Each node has three connection with input layer
    - mutations, copy number amplification, copy number deletion
- 2 – 6 layers: Hierarchy of pathways and biological processes
  - manually curated by Reactome pathway dataset, reflecting the real parent-child relationships

$$y = f[(M * W)^T x + b]$$

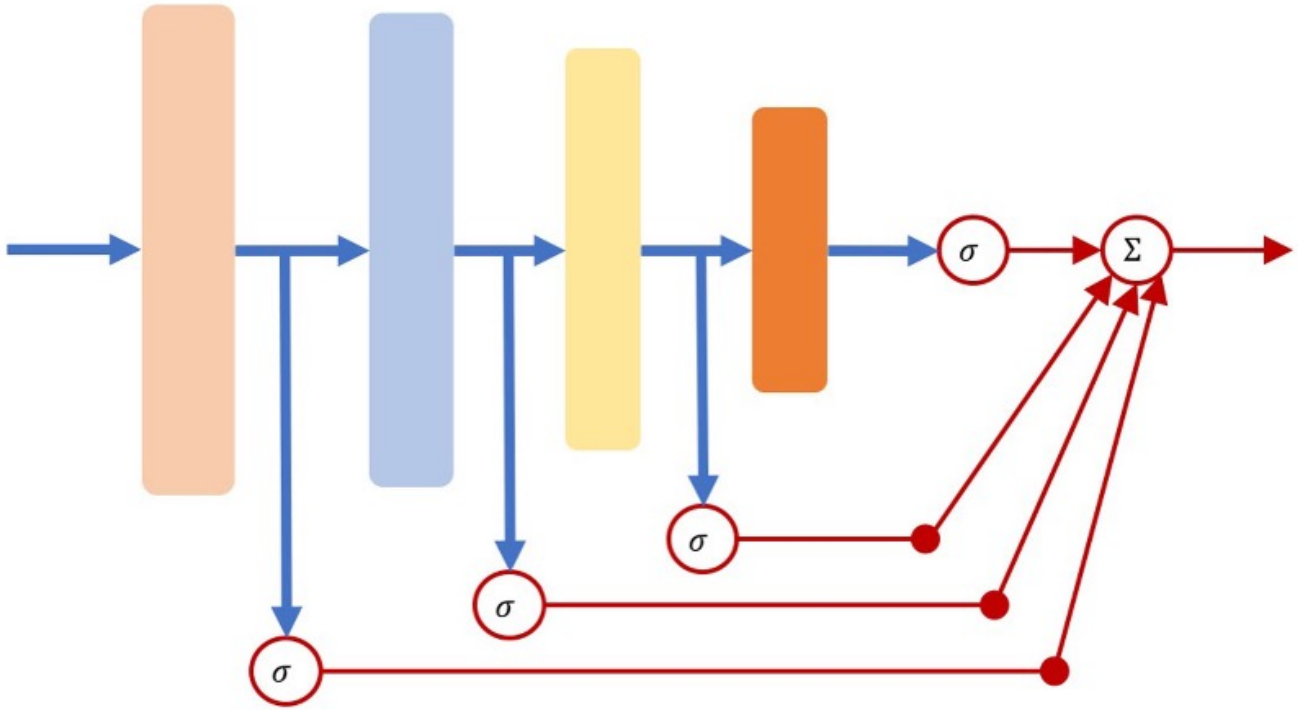
y: output of each layer
M: mask matrix
W: weights matrix
b: bias vector
$*$: Hadamard product
f: $\tanh = (e^{2x} - 1)/(e^{2x} + 1)$
σ: $1/(1 + e^{-x})$ → outcome layer

Methods

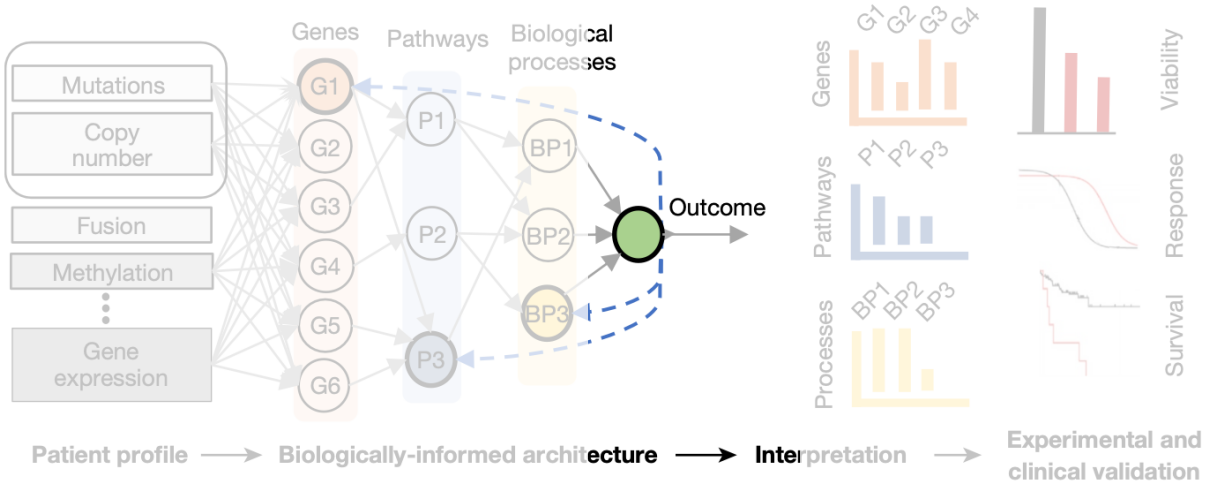
$$y = f[(M * W)^T x + b]$$



$$y = f[(M * W)^T x + b]$$



Optimization Process

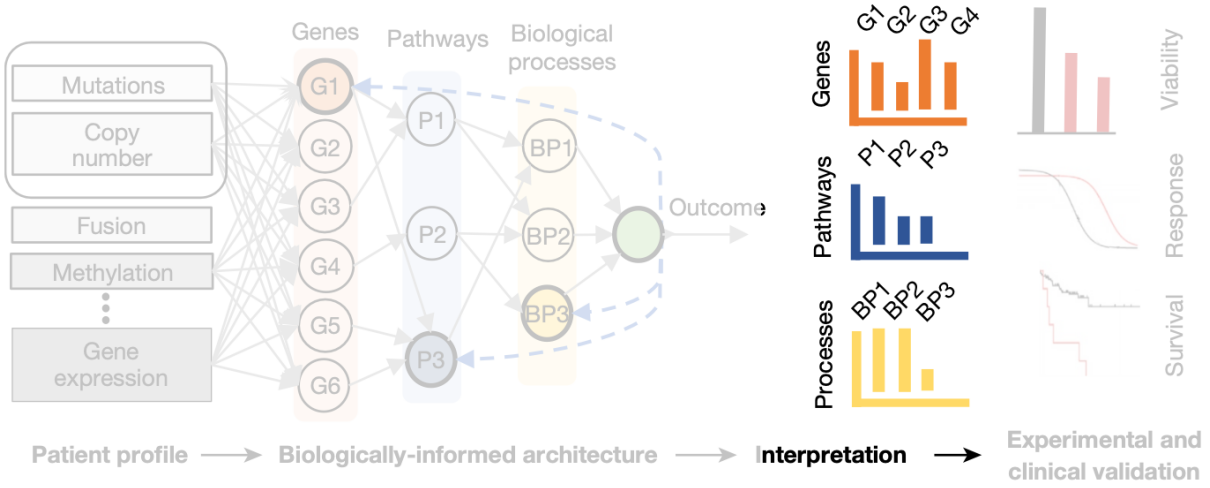


- Learning rate: 0.001
- Epochs: 50
- Binary cross-entropy loss functions

$$H = -\frac{1}{N}\sum y_i \log\left(p(y_i)\right) + (1 - y_i)\log(1 - p(y_i))$$

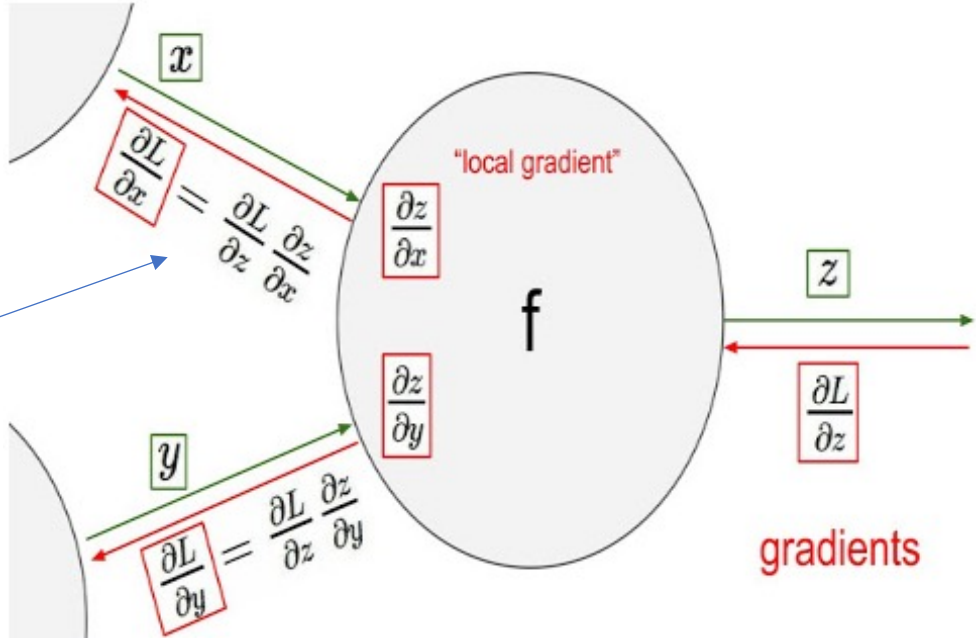
- To reduce loss, Adam optimizer was used.

DeepLIFT



- Backpropagation-based attribution approach
- Shrikumar, et al.(2017, *international conference on machine learning*)
  - # of citations: 3813
- Calculate importance score for each node in each layer

Chain Rule



DeepLIFT

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta t} = \Delta t$$

Δt : Target neuron t with difference of output from reference (e.g. Primary cancer sample v.s. CRPC sample)

$x_1, x_2, \ldots, x_n$: some neurons in some intermediate layer

$C_{\Delta x_i \Delta t}$: Contribution score

Multipliers

$$m_{\Delta x \Delta t} = \frac{C_{\Delta x \Delta t}}{\Delta x}$$

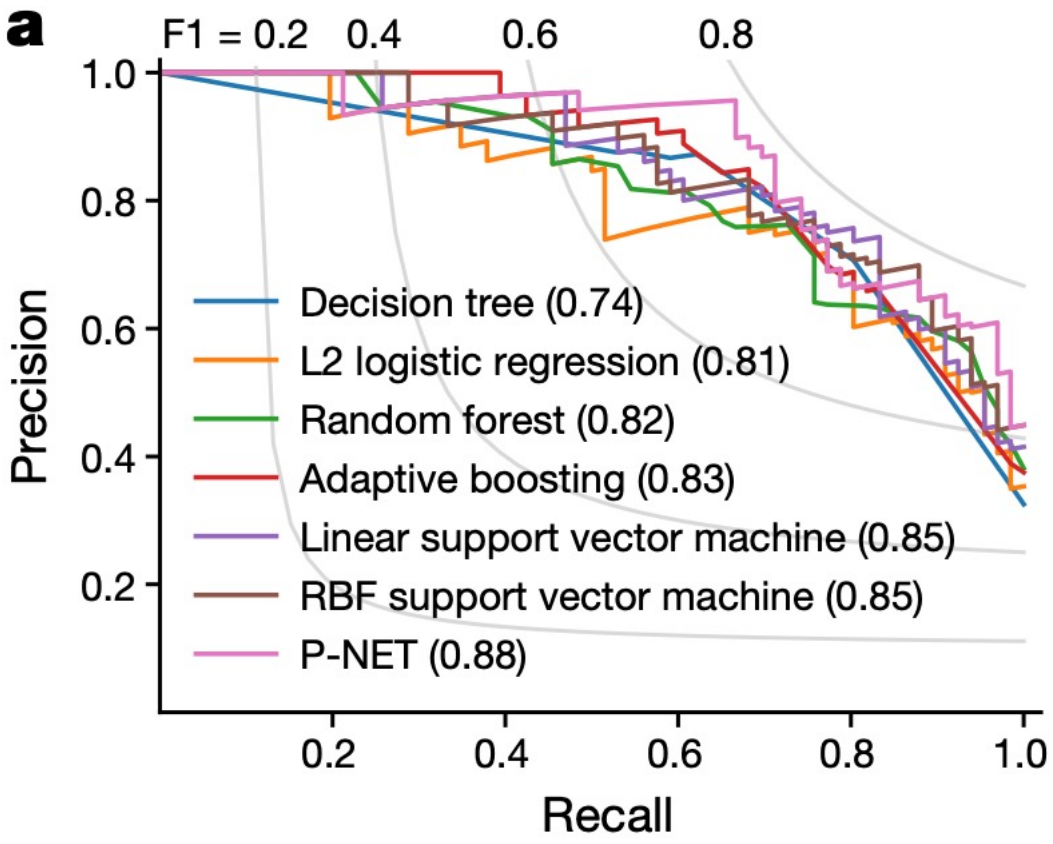
Chain Rule for multipliers

$$m_{\Delta x_i \Delta t} = \sum_{j} m_{\Delta x_i \Delta y_j} m_{\Delta y_j \Delta t}$$

* Order of layers: x → y → t(node)

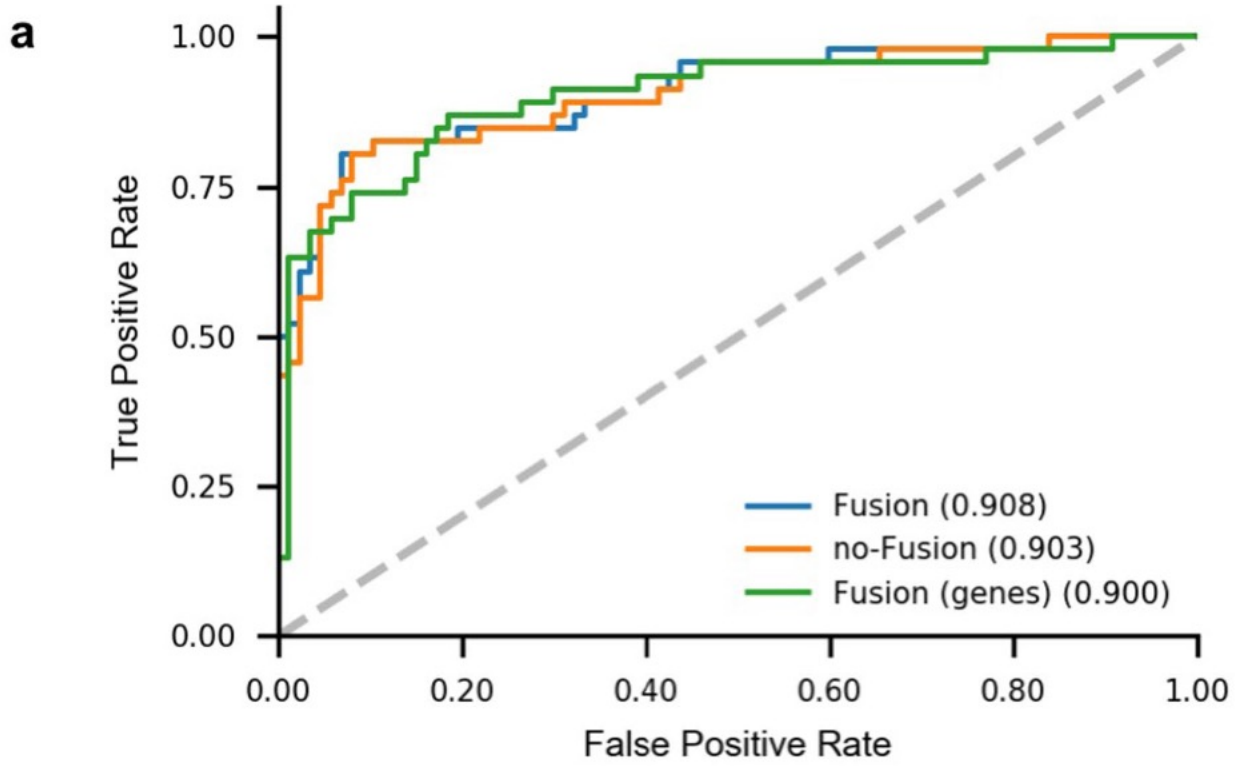
Results

Fig 2. Prediction performance of P-NET



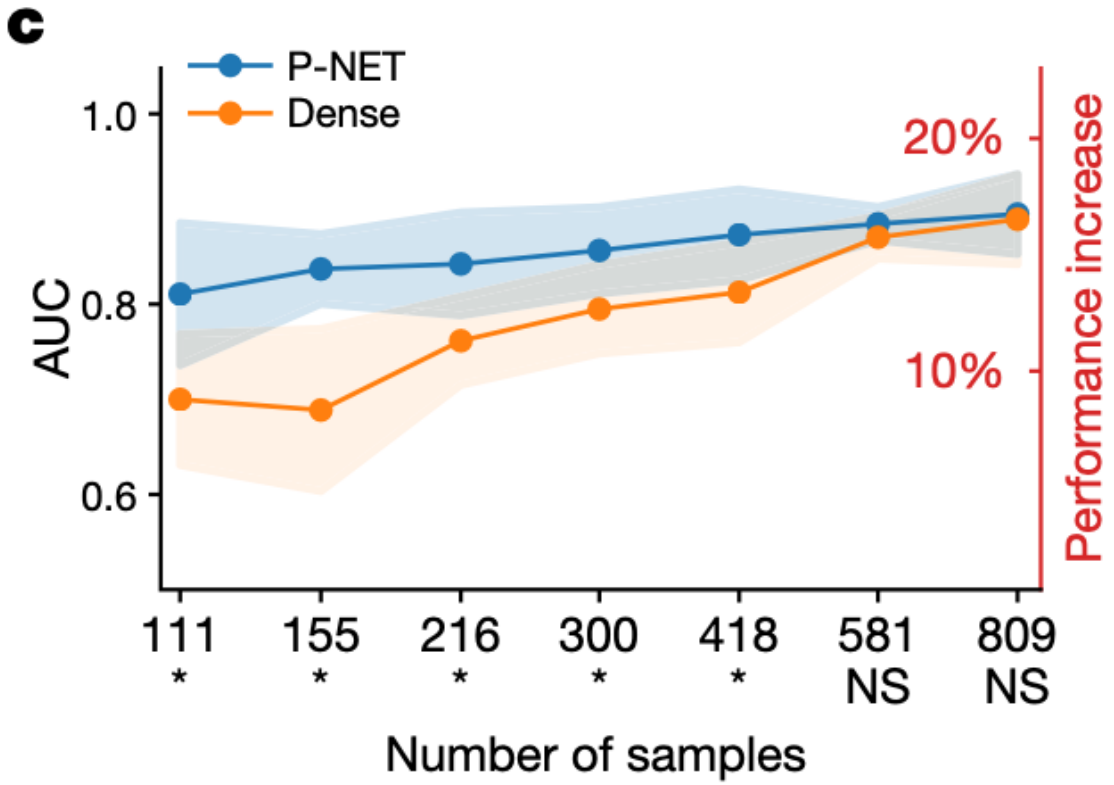
- Compare the performance of P-NET with six methods
- Input data is divided into 80% training, 10% validation and 10% testing
- The P-NET outperformed at AUC-ROC, AUPRC

Fig 2. Prediction performance of P-NET



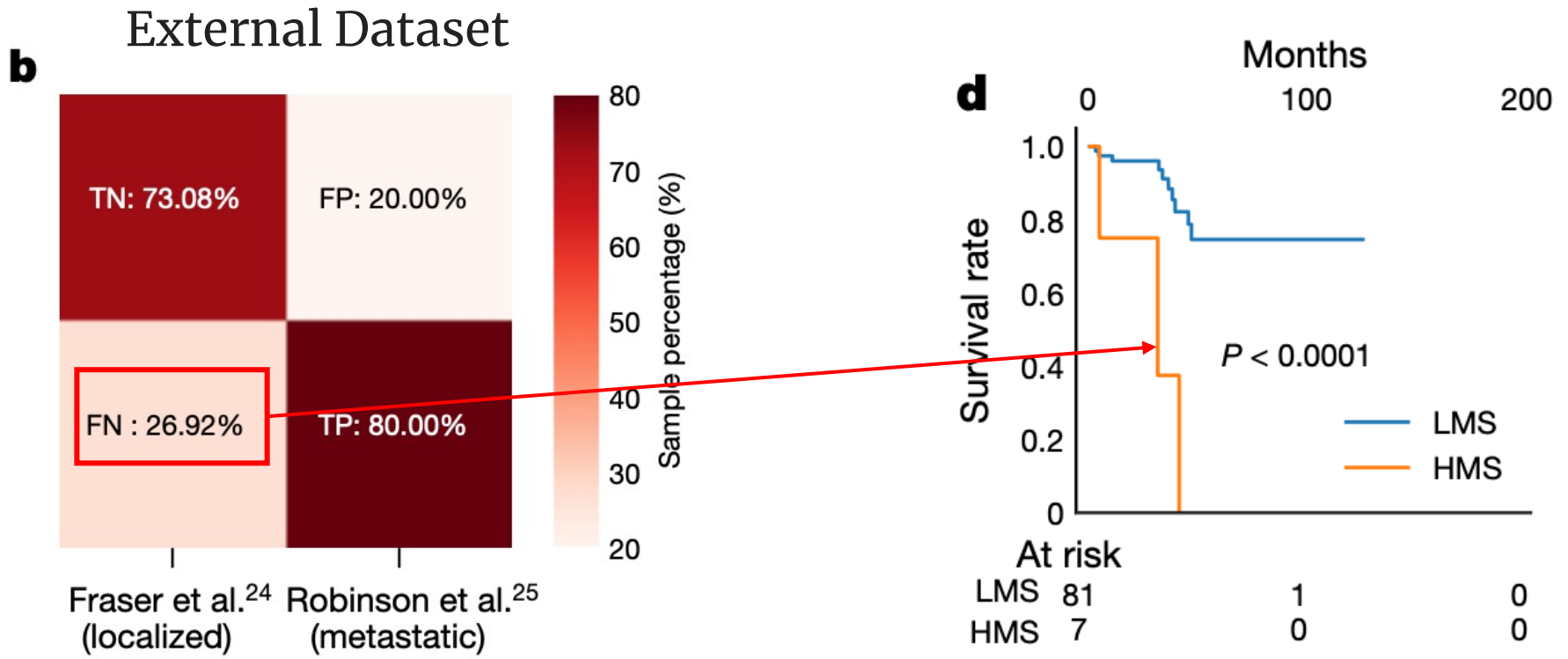
- Compare the performance between three models
  - Fusion: include binary variable to indicate whether a sample has fusion or not(ETS fusion and oncogene fusion)
  - Fusion(genes): binary variables for each gene
- Not impact the performance

Fig 2. Prediction performance of P-NET



- Average AUC over five cross-validation splits
- Sparse model has higher performance than dense model
  - Statistically significant in Sample sizes up to 500(t-test, $p<0.05$)

Fig 2. Prediction performance of P-NET



- Adequate predictive performance with **unseen samples**
- Patients with **High P-NET scores** misclassified have **biochemical recurrence**

Fig 3. Inspecting and interpreting P-NET

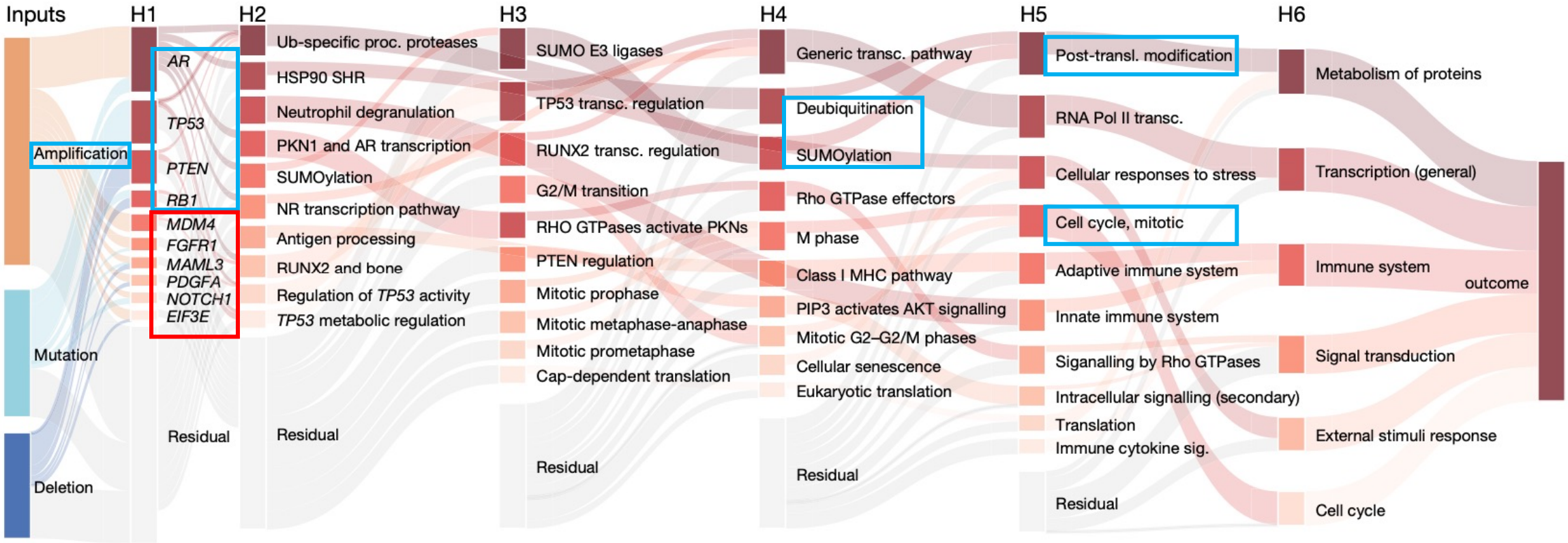


Fig 3. Importance score of each entities

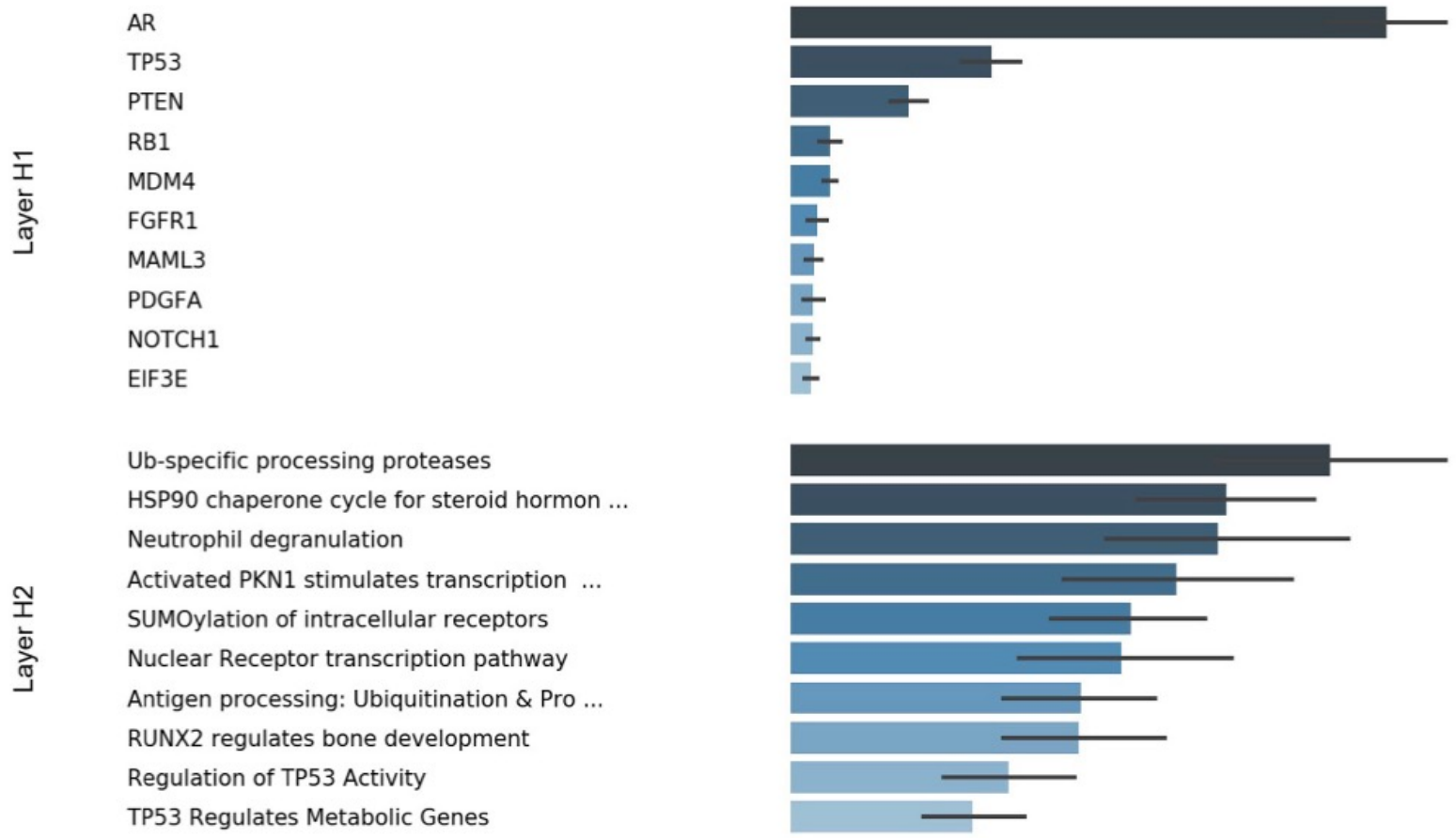
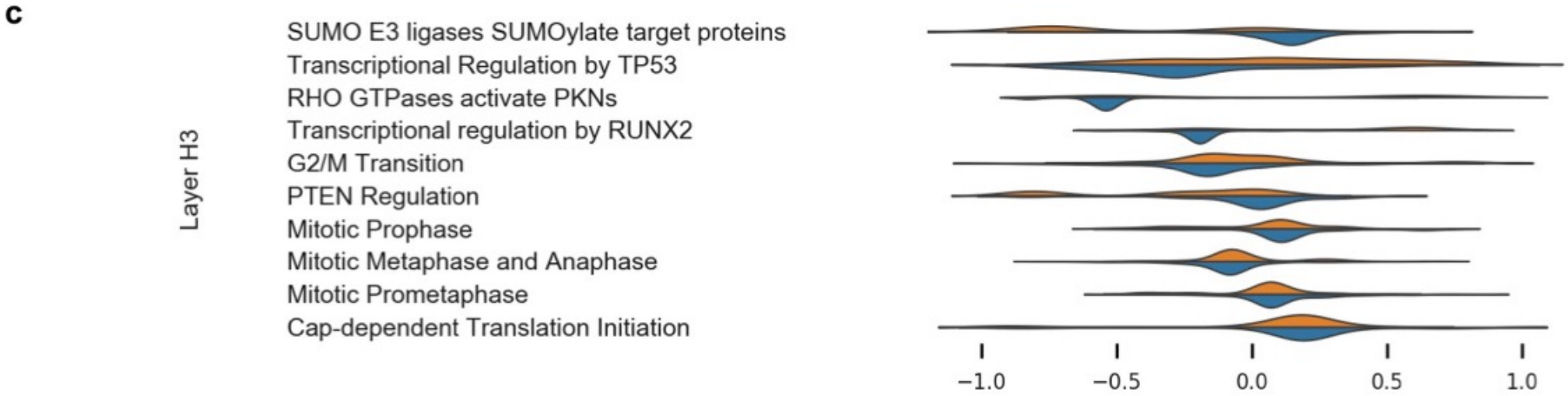
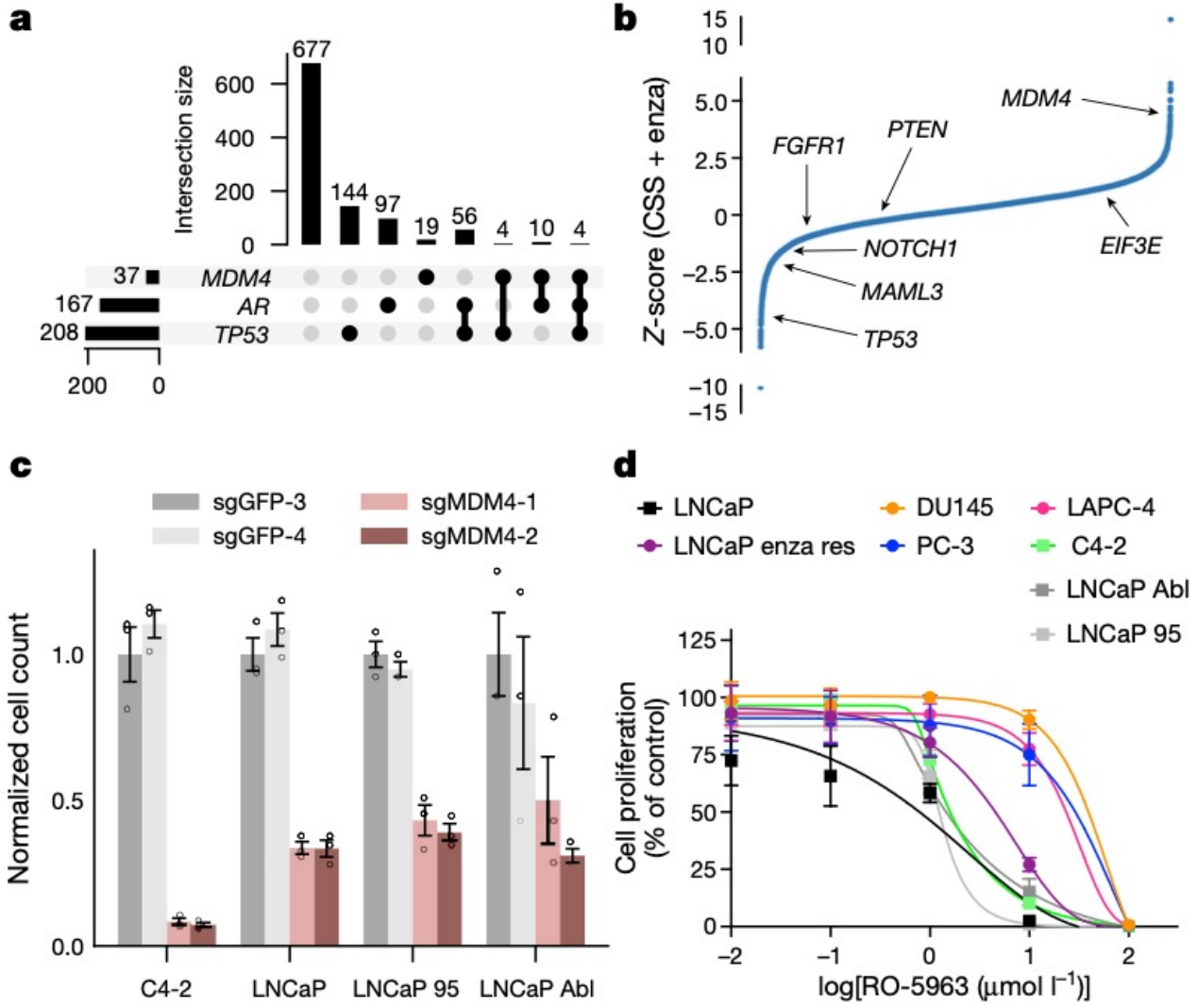


Fig 3. Activation distribution



- Activation: outcome of a certain node given its inputs
- Observed difference in the node activation between Primary and Metastatic
- Higher node activation in

Figure 4.


a
Intersection size
677
600
400
200
0
144
97
19
56
4
10
4
37
167
208
MDM4
AR
TP53
200
0
b
Z-score (CSS + enza)
15
10
5.0
2.5
0
−2.5
−5.0
−10
−15
MDM4
PTEN
FGFR1
NOTCH1
EIF3E
MAML3
TP53
c
sgGFP-3
sgGFP-4
sgMDM4-1
sgMDM4-2
Normalized cell count
1.0
0.5
0.0
C4-2
LNCaP
LNCaP 95
LNCaP Abl
d
LNCaP
DU145
LAPC-4
LNCaP enza res
PC-3
C4-2
LNCaP Abl
LNCaP 95
Cell proliferation (% of control)
125
100
75
50
25
0
−2
−1
0
1
2
log[RO-5963 (μmol l−1)]

Discussion

Discussion

- P-NET leveraged a biologically informed, rather than arbitrarily overparameterized architecture for prediction.
- P-NET reduced the number of parameters which led to enhanced interpretability.
- The sparse architecture in P-NET has better predictive performance when compared to other machine learning models.
- Application of P-NET to a molecular cohort of patients with prostate cancer demonstrated,
  - model performance that may enable prediction of clinically aggressive disease in populations of patients with primary prostate cancer
  - convergent biological processes that contribute to a metastatic prostate cancer that harbor novel therapeutic strategies in molecularly stratified populations.

Conclusion

- P-NET, a deep neural network informed by biology, successfully distinguished between primary and advanced prostate cancers.
- It offered new ideas about how prostate cancer spreads and useful insights for treating different patient groups.
- This approach combines cancer biology with machine learning, creating models that predict and help in understanding cancer, potentially useful in various cancer research areas.

Thanks!

DeepLIFT

$$
\begin{aligned}
\sum_i C_{\Delta x_i \Delta t} &= \sum_i \Delta x_i m_{\Delta x_i \Delta t} \text{ (By definition of } m_{\Delta x_i \Delta t}) \\
&= \sum_i \Delta x_i \sum_j m_{\Delta x_i \Delta y_j} m_{\Delta y_j \Delta t} \text{ (By the chain rule)} \\
&= \sum_i \Delta x_i \sum_j \frac{C_{\Delta x_i \Delta y_j}}{\Delta x_i} m_{\Delta y_j \Delta t} \text{ (By definition of } m_{\Delta x_i \Delta y_j}) \\
&= \sum_i \sum_j C_{\Delta x_i \Delta y_j} m_{\Delta y_j \Delta t} \\
&= \sum_j \sum_i C_{\Delta x_i \Delta y_j} m_{\Delta y_j \Delta t} \text{ (Flipping the order of summation)} \\
&= \sum_j \Delta y_j m_{\Delta y_j \Delta t} \text{ (By summation-to-delta of } C_{\Delta x_i \Delta y_j}) \\
&= \sum_j \Delta y_j \frac{C_{\Delta y_j \Delta t}}{\Delta y_j} \text{ (By definition of } m_{\Delta y_j \Delta t}) \\
&= \sum_j C_{\Delta y_j \Delta t} = \Delta t \text{ (By summation-to-delta of } C_{\Delta y_j \Delta t})
\end{aligned}
$$

DeepLIFT

- Separating Positive and Negative Contributions

$$\Delta y = \Delta y^+ + \Delta y^-$$

$$C_{\Delta y \Delta t} = C_{\Delta y^+ \Delta t} + C_{\Delta y^- \Delta t}$$

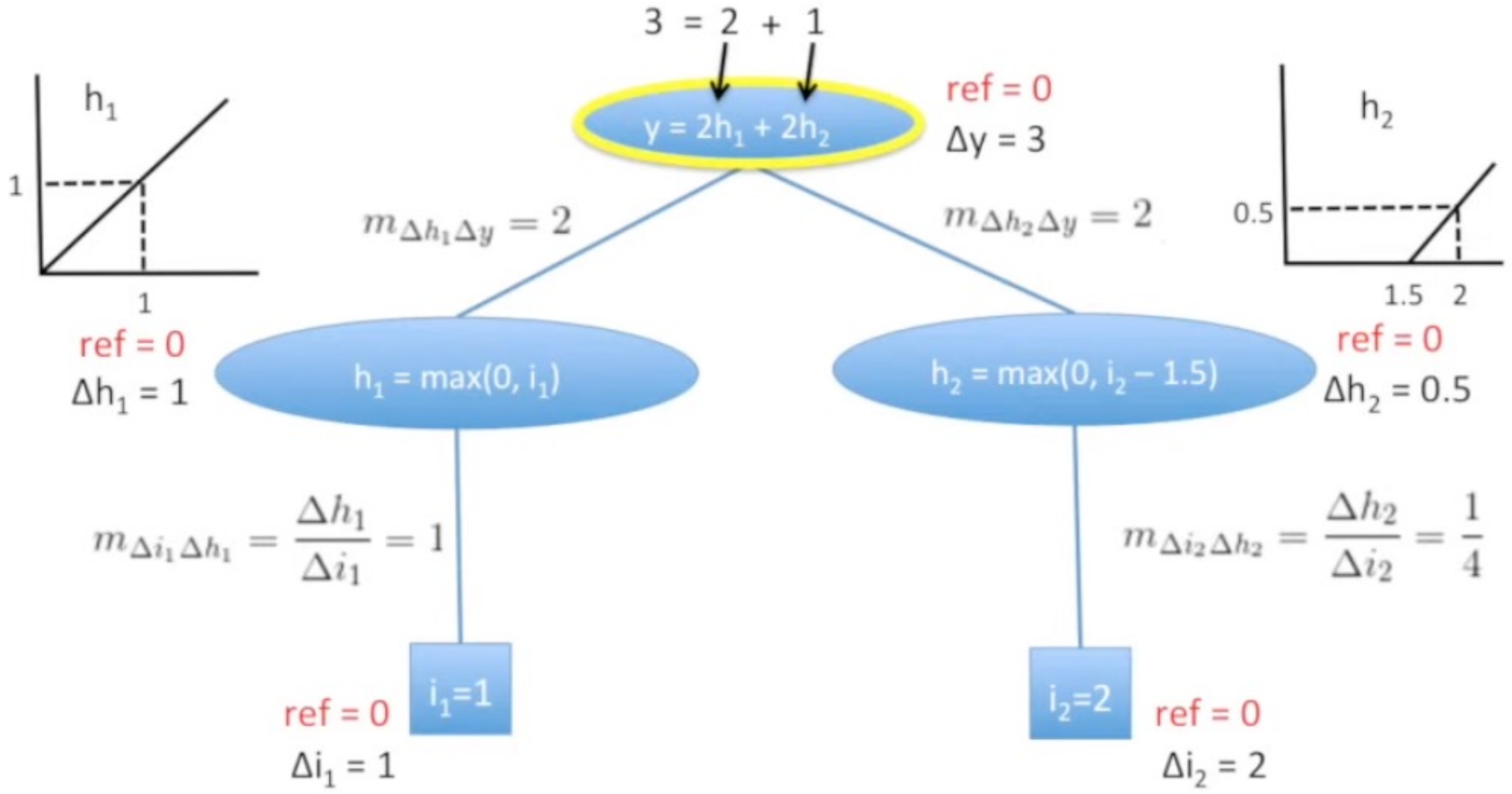
- The rescale rule

$$\Delta y^+ = \frac{\Delta y}{\Delta x} \Delta x^+ = C_{\Delta x^+ \Delta y^+}$$

$$\Delta y^- = \frac{\Delta y}{\Delta x} \Delta x^- = C_{\Delta x^- \Delta y^-}$$

$$m_{\Delta x^+ \Delta y^+} = m_{\Delta x^- \Delta y^-} = m_{\Delta x \Delta y} = \frac{\Delta y}{\Delta x}$$

DeepLIFT


3 = 2 + 1
y = 2h1 + 2h2
ref = 0
Δy = 3
h1
1
1
ref = 0
Δh1 = 1
h2
0.5
1.5 2
ref = 0
Δh2 = 0.5
m_Δh1Δy = 2
m_Δh2Δy = 2
h1 = max(0, i1)
h2 = max(0, i2 – 1.5)
m_Δi1Δh1 = Δh1/Δi1 = 1
m_Δi2Δh2 = Δh2/Δi2 = 1/4
i1=1
i2=2
ref = 0
Δi1 = 1
ref = 0
Δi2 = 2
m_Δi1Δy = m_Δi1Δh1 m_Δh1Δy = 2
C_Δi1Δy = m_Δi1Δy Δ_i1 = 2
m_Δi2Δy = m_Δi2Δh2 m_Δh2Δy = 0.5
C_Δi2Δy = m_Δi2Δy Δ_i2 = 1

P-NET training and evaluation

- Training
  - Input data is divided into 80% training, 10% validation and 10% testing
  - For the cross-validation experiments: Five folds cross-validation
- Evaluation
  - The change in the ROC-AUC between P-NET and other models is tested using DeLong test.
    - DeLong test: non-parametric approach to compare two AUC values
  - For the cross-validation experiments: using a t-test of the means
  - For the survival analysis, a nonparametric log-rank test is used