Mechanistic Autoencoders for Patient-Specific Phosphoproteomic Models

This manuscript ([permalink](https://meyer-lab.github.io/mechanismEncoder/v/f851f7fa2f9e3b90c404c0b2352f9d9ff51e2931/)) was automatically generated from [meyer-lab/mechanismEncoder@f851f7f](https://github.com/meyer-lab/mechanismEncoder/tree/f851f7fa2f9e3b90c404c0b2352f9d9ff51e2931) on April 6, 2021.

## Authors

* **Fabian Fröhlich** ORCID [0000-0002-5360-4292](https://orcid.org/0000-0002-5360-4292) · Github [FFroehlich](https://github.com/FFroehlich) · twitter [fabfrohlich](https://twitter.com/fabfrohlich) Department of Systems Biology, Harvard Medical School
* **Sara JC Gosline** ORCID [0000-0002-6534-4774](https://orcid.org/0000-0002-6534-4774) · Github [sgosline](https://github.com/sgosline) · twitter [sargoshoe](https://twitter.com/sargoshoe) Pacific Northwest National Laboratories
* **Jackson L. Chin** · Github [JacksonLChin](https://github.com/JacksonLChin) Department of Bioengineering, University of California, Los Angeles
* **Emek Demir** ORCID [0000-0002-3663-7113](https://orcid.org/0000-0002-3663-7113) · Github [emekdemir](https://github.com/emekdemir) Department of Molecular and Medical Genetics, Oregon Health & Sciences Univerity
* **Aaron S. Meyer** ORCID [0000-0003-4513-1840](https://orcid.org/0000-0003-4513-1840) · Github [aarmey](https://github.com/aarmey) · twitter [aarmey](https://twitter.com/aarmey) Department of Bioengineering, University of California, Los Angeles; Department of Bioinformatics, University of California, Los Angeles; Jonsson Comprehensive Cancer Center, University of California, Los Angeles; Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles

## Abstract

Proteomic data provides measurements that are uniquely close to the mechanism of action for many cancer therapies. As such, it can provide an unmatched perspective into the mechanism of drug action and resistance. At the same time, extracting the source of patient-to-patient differences in proteomic measurements and understanding its relevance for drug sensitivity is extremely challenging. Correlative analyses are most common but are difficult to mechanistically interpret.

## Introduction

Proteomic data provides measurements that are uniquely close to the mechanism of action for many cancer therapies. As such, it can provide an unmatched perspective into the mechanism of drug action and resistance.[1](#ref-nOENiyam),[2](#ref-dzPjnBBH) At the same time, extracting the source of patient-to-patient differences in proteomic measurements and understanding its relevance for drug sensitivity is extremely challenging. Correlative analyses are most common but are difficult to mechanistically interpret.

Mechanistic models are uniquely powerful for identifying the drivers of differences within measurements, integrating our prior knowledge, and interpreting data. However, a key question that limits their use for patient data is how to handle patient-to-patient differences. Constructing multiple patient-specific models is infeasible due to the limited data for each patient. Alternatively, universal models that use patient invariant and patient-specific parameters to integrate data across multiple individuals have been proposed.[3](#ref-JX6lNYob) However, how to estimate these patient-specific parameters is challenging as genetic and microenvironmental context influences signaling pathways in complex, non-linear, and often poorly understood ways.

At its core, the challenge of integrating mechanistic models with patient-derived measurements is an issue of how to account for patient-to-patient variation. Mechanistic dynamical models have been widely applied to data of all types but are used where the sources of variation among measurements can be explicitly identified and modeled. By contrast, variation among individuals can arise through both factors that can easily be identified, like changes in the abundance of the species being modeling, and endless other molecular and physiological factors that cannot be usefully enumerated in a mechanistic approach. Still, the structure of mechanistic models provides important constraints on the behavior of molecular pathways and interpretability that is missing from purely data-driven statistical methods.

To address this issue, we propose a model structure that is based on a variational autoencoder. Autoencoders are neural networks that embed data into low dimensional latent feature space by feeding the data through encoding and decoding layers.[4](#ref-764aXjfI) The extracted latent features then provide a reduced representation of patient-patient similarity. We integrate mechanistic information by partly replacing the decoder layers in the network with a coarse-grained mechanistic model, where the encoded, latent representation of the data defines the patient-specific parameters of the universal ordinary differential equation (ODE) model. We apply this to AML patient samples, where proteomic and phosphoproteomic measurements with high tumor purity can be collected. This model structure enables mechanistic interpretation of these data; more robust latent space representations of patient relationships; and integration of prior knowledge, other data sources such as *in vitro* experiments or other data types, and clinical measurements. Mechanistic autoencoders, therefore, offer a general solution to building mechanistic models in the presence of unexplained sample variation, such as from clinical samples.

## Results

### MAEs mechanistically account for patient-to-patient variation

Figure 1: **XXX.** A) XXX. B) XXX.

Figure 1: **XXX.** A) XXX. B) XXX.

* Schematic of autoencoder structure
* Cartoon description of other encoder structures
* Part of this would be to speak to the generality of the approach

### Figure 2

Figure 2: **XXX.** A) XXX. B) XXX.

Figure 2: **XXX.** A) XXX. B) XXX.

Initial plot of proteomic data (clustergram?) - see #23 Data-driven selection of network nodes from OHSU

### Figure 3

Figure 3: **XXX.** A) XXX. B) XXX.

Figure 3: **XXX.** A) XXX. B) XXX.

Training against actual data Description of fit model

### Figure 4

Figure 4: **XXX.** A) XXX. B) XXX.

Figure 4: **XXX.** A) XXX. B) XXX.

Cell line perturbation Description of that data

### Figure 5

Figure 5: **XXX.** A) XXX. B) XXX.

Figure 5: **XXX.** A) XXX. B) XXX.

Model/validation comparison

## Discussion

* Emphasize generality of approach
* Future possibilities
  + True validation in patient-derived samples
  + Pan-cancer modeling
  + Infer structure ala perturbation biology / neural ODEs?
* Cover other forms of mechanistic / data-driven integration?

FROM PROPOSAL

This project has the potential to enable routine use of mechanistic models to analyze clinical proteomics measurements. As such, one can easily envision applying a similar technique across many different cancer types as well as other diseases. It is hard to overstate the potential impact, as this can convert these measurements to (1) exacting predictions of which components to target in individual patients and (2) provide a mechanism-grounded view of patient-to-patient variation.

## Methods

### Data collection

Sara should likely fill this in.

### Basic autoencoder implementation

This would be Jackson’s work.

### Mechanistic model implementation and integration

Fabian knows this.

### Pathway Commons analysis

## Acknowledgements

This work was supported by an administrative supplement to NIH U01-CA215709 to A.S.M. The authors declare no competing financial interests.

## Author contributions statement

1. J.L.C. implemented and analyzed the standard autoencoder as a baseline. All authors wrote the paper.

## References

1. McDermott, J. E. *et al.* Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. *Cell Reports Medicine* **1**, 100004 (2020).

2. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964–983.e31 (2019).

3. Fröhlich, F. *et al.* Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Systems* **7**, 567–579.e6 (2018).

4. Hinton, G. E. Reducing the Dimensionality of Data with Neural Networks. *Science* **313**, 504–507 (2006).