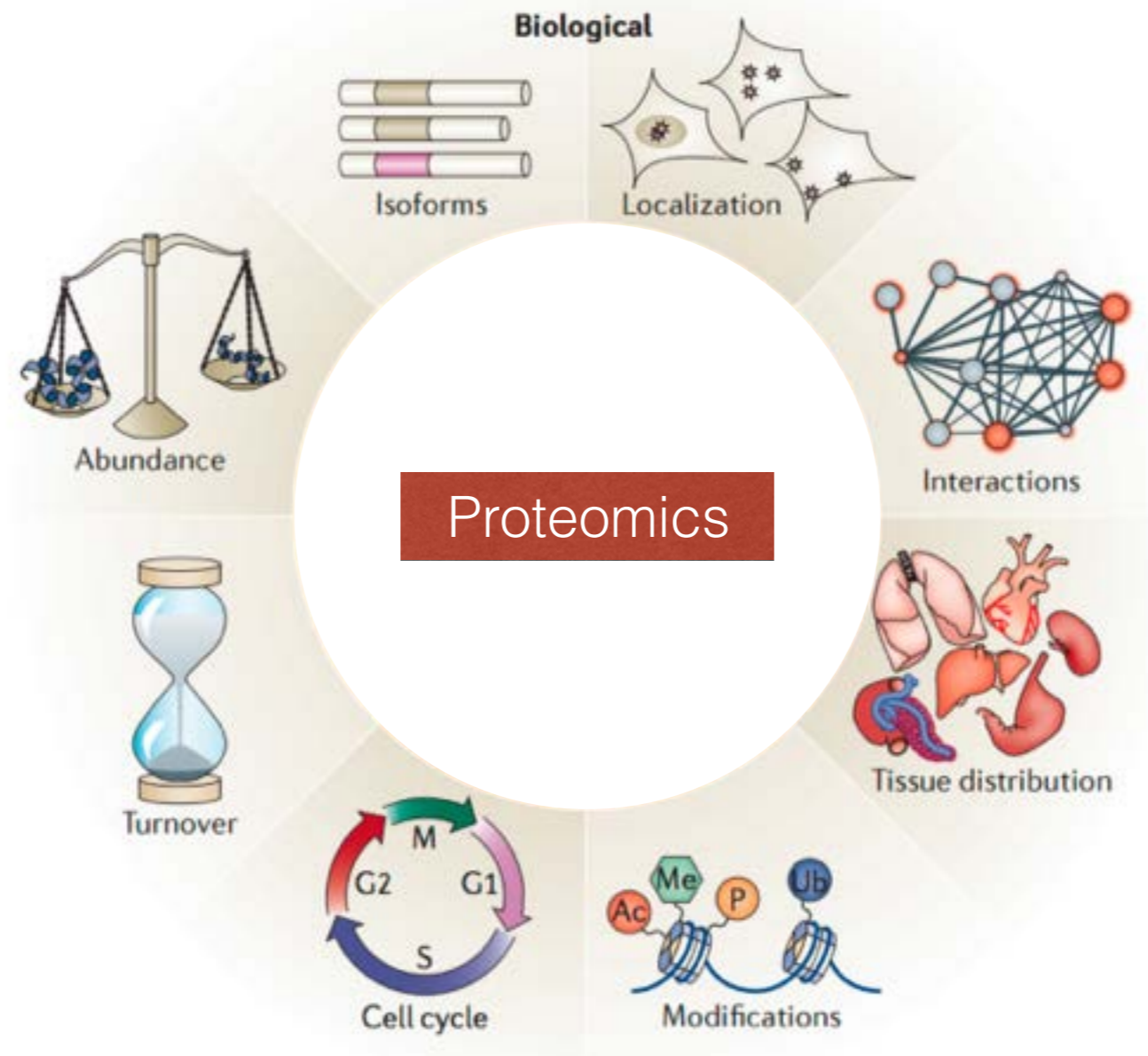


# DeepPep: deep proteome inference from peptide profiling

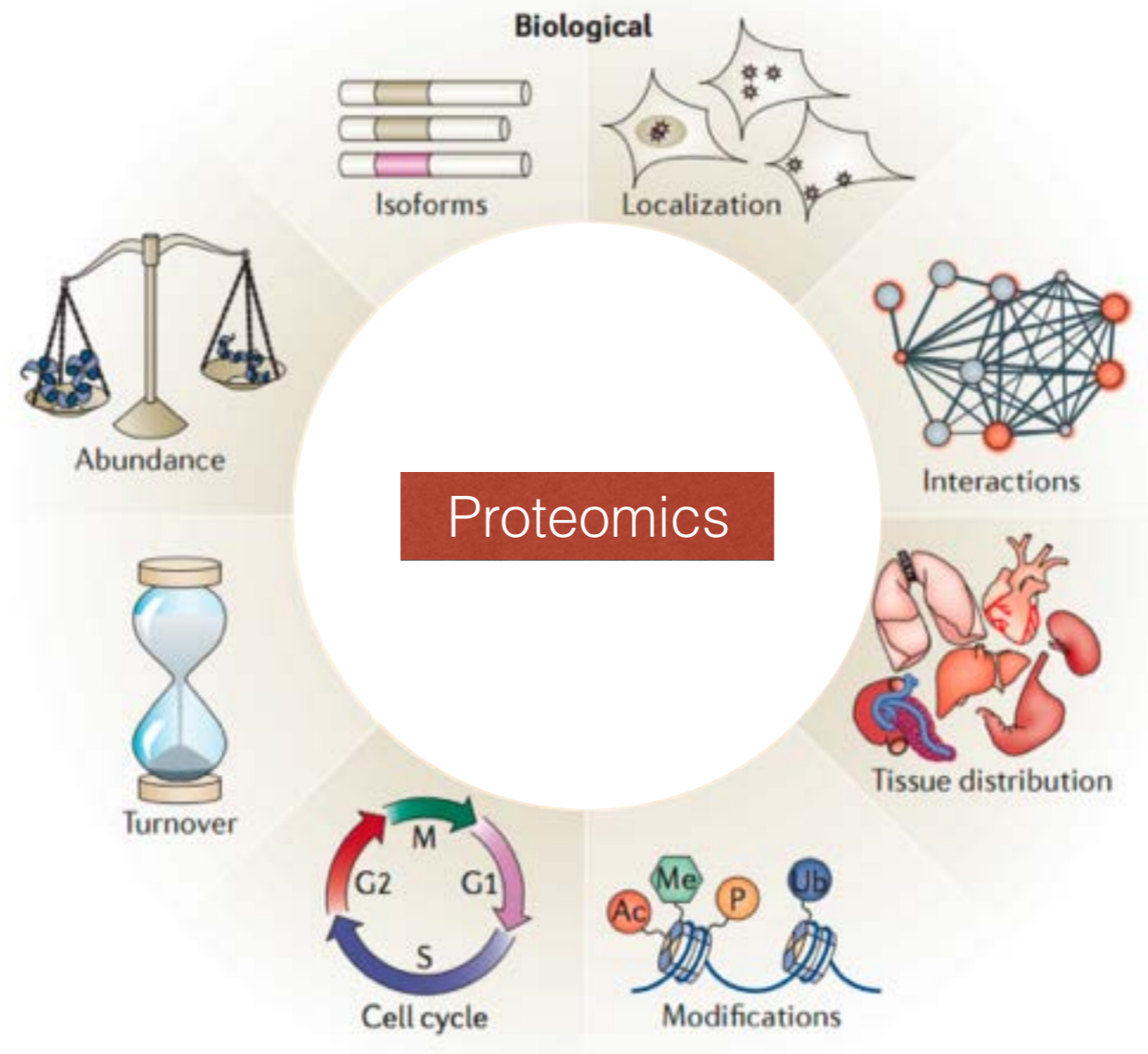
Minseung Kim

## Large-scale study of proteins, particularly their structures and functions



Larance et al., 2015

## Large-scale study of proteins, particularly their structures and functions



Larance et al., 2015

### Applications

Disease biomarker discovery

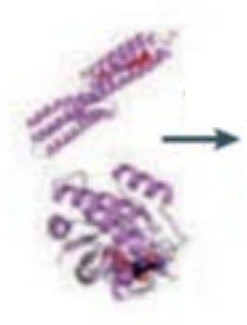
Drug target

Host-pathogen interaction

...

# What proteins are expressed?

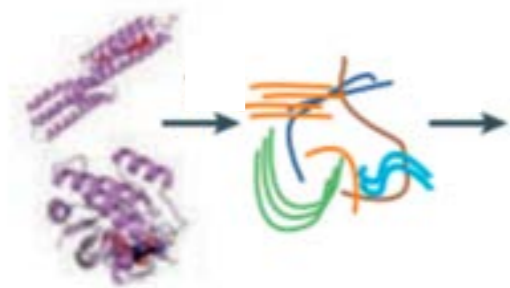
**using shotgun proteomics based on Mass Spectrometry**



Proteins

# What proteins are expressed?

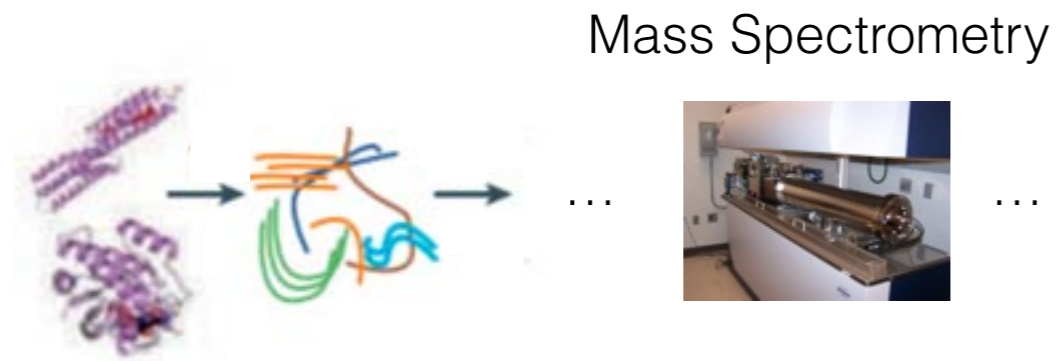
**using shotgun proteomics based on Mass Spectrometry**



Proteins Peptides

# What proteins are expressed?

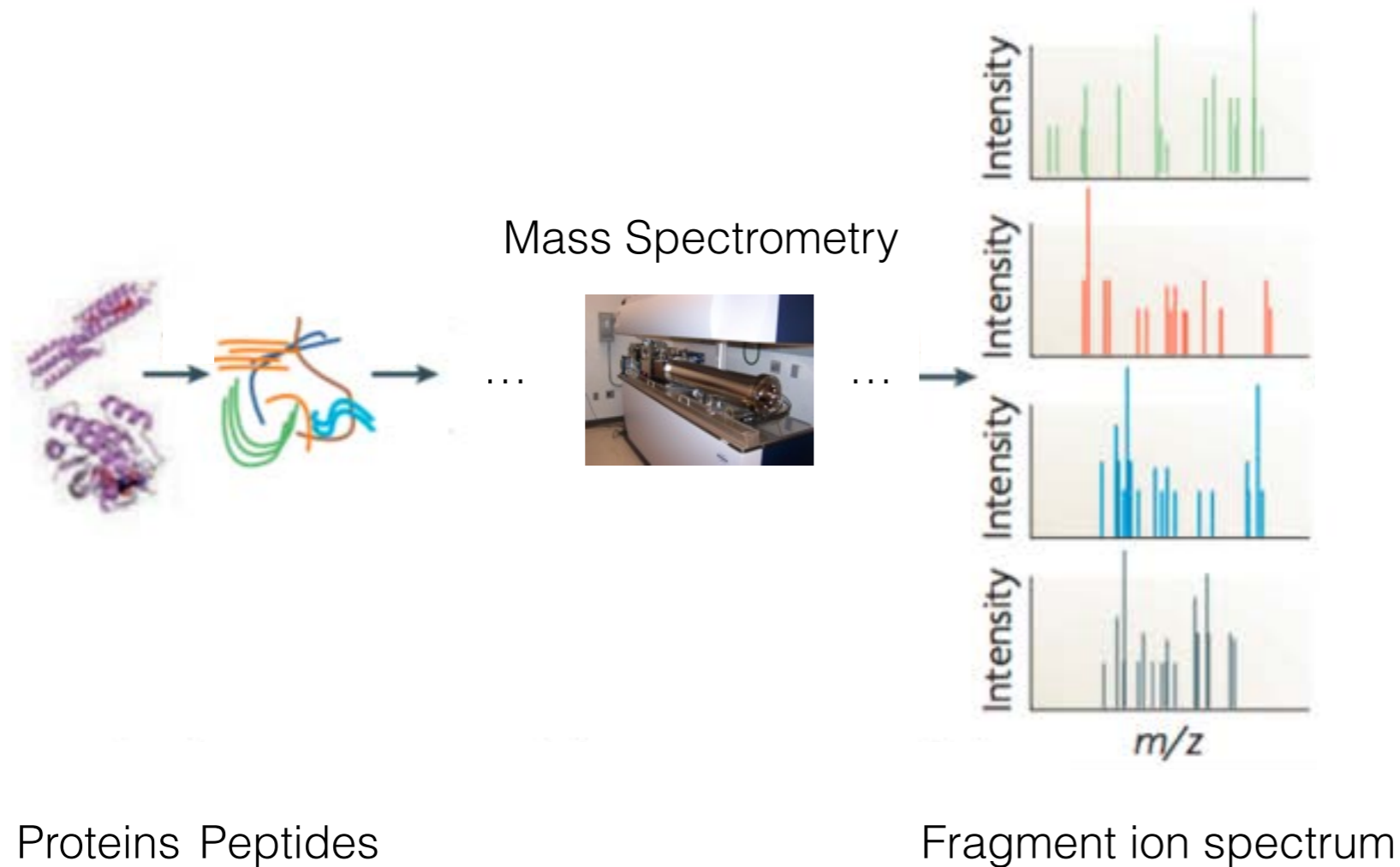
## using shotgun proteomics based on Mass Spectrometry



Proteins Peptides

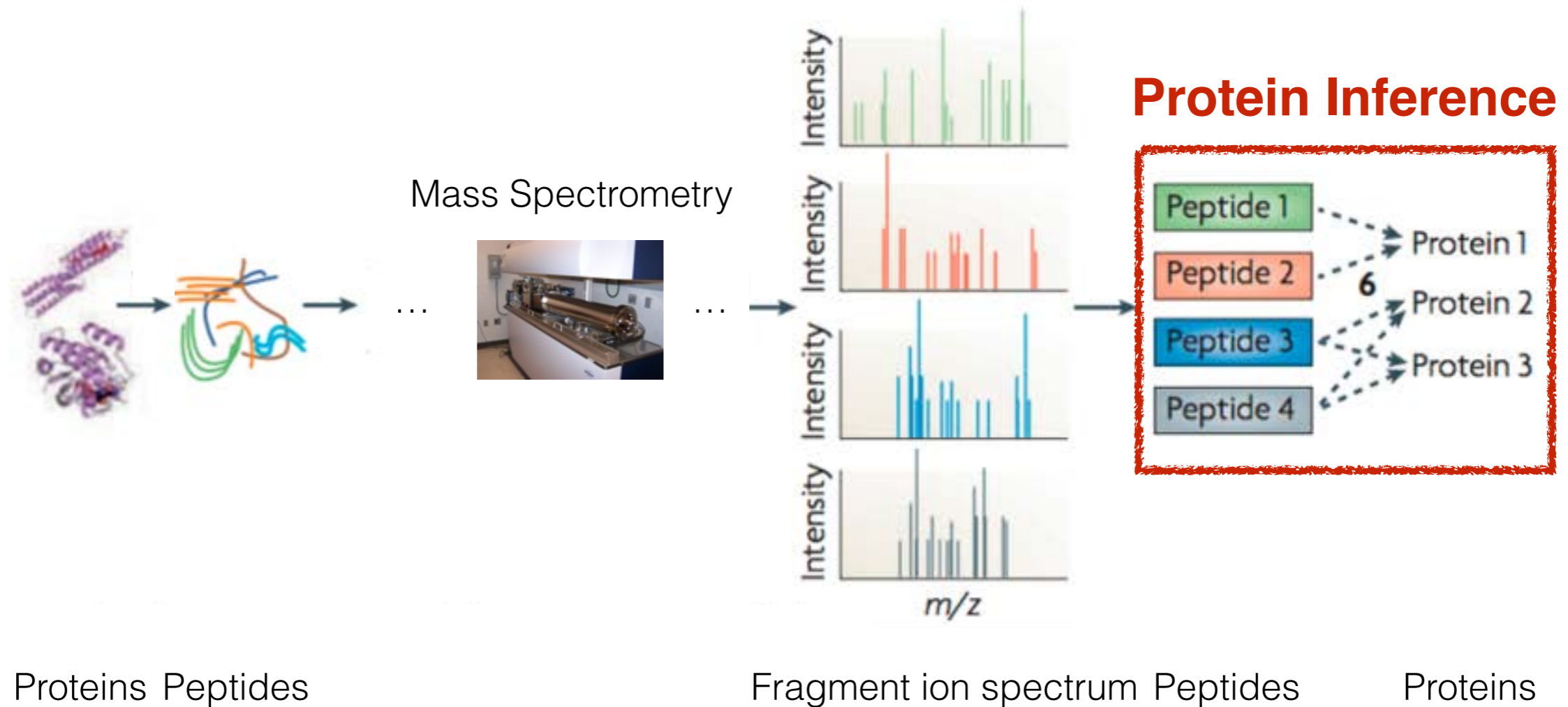
# What proteins are expressed?

## using shotgun proteomics based on Mass Spectrometry



# What proteins are expressed?

## using shotgun proteomics based on Mass Spectrometry



Proteins Peptides

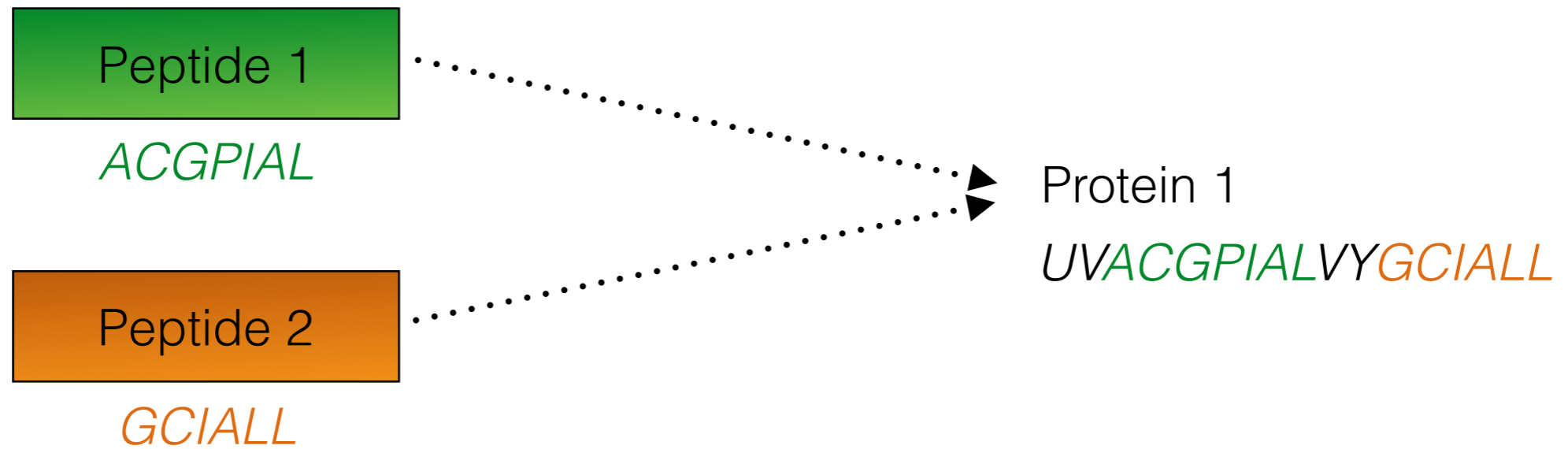
Fragment ion spectrum Peptides

Proteins

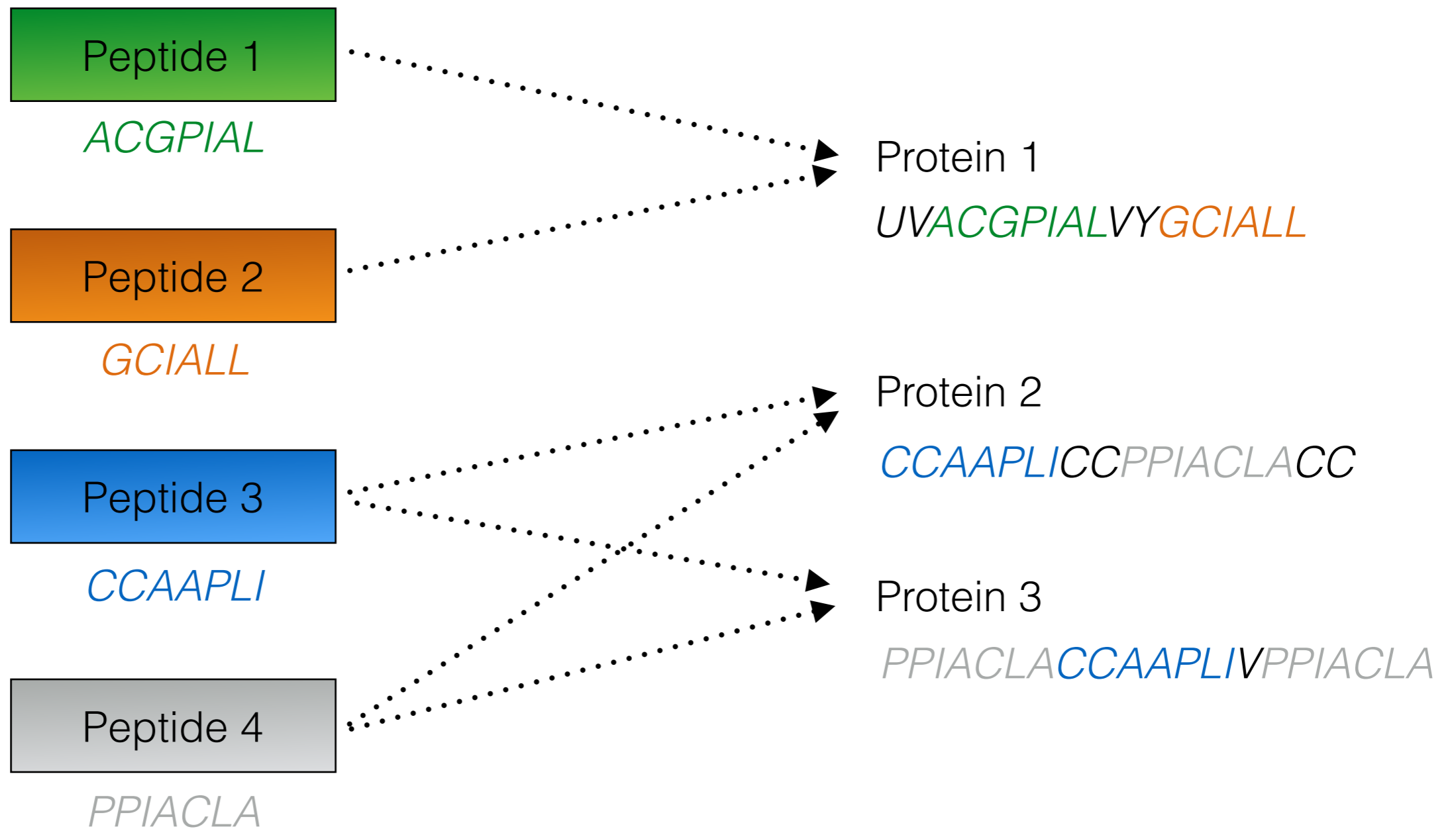
Ahrens et al., 2010



## Infer proteins that are origins of observed peptides



## Infer proteins that are origins of observed peptides



## A. Training data preparation

	ID	Peptide	Prob
Observed peptides	Pep <sub>1</sub>	TRPAY	0.8
	Pep <sub>2</sub>	PATR	0.2
	...	...	...

	Protein	Sequence
All proteins	P <sub>A</sub>	PTRPAYR...
	P <sub>B</sub>	PATRPAYA..A
	...	...

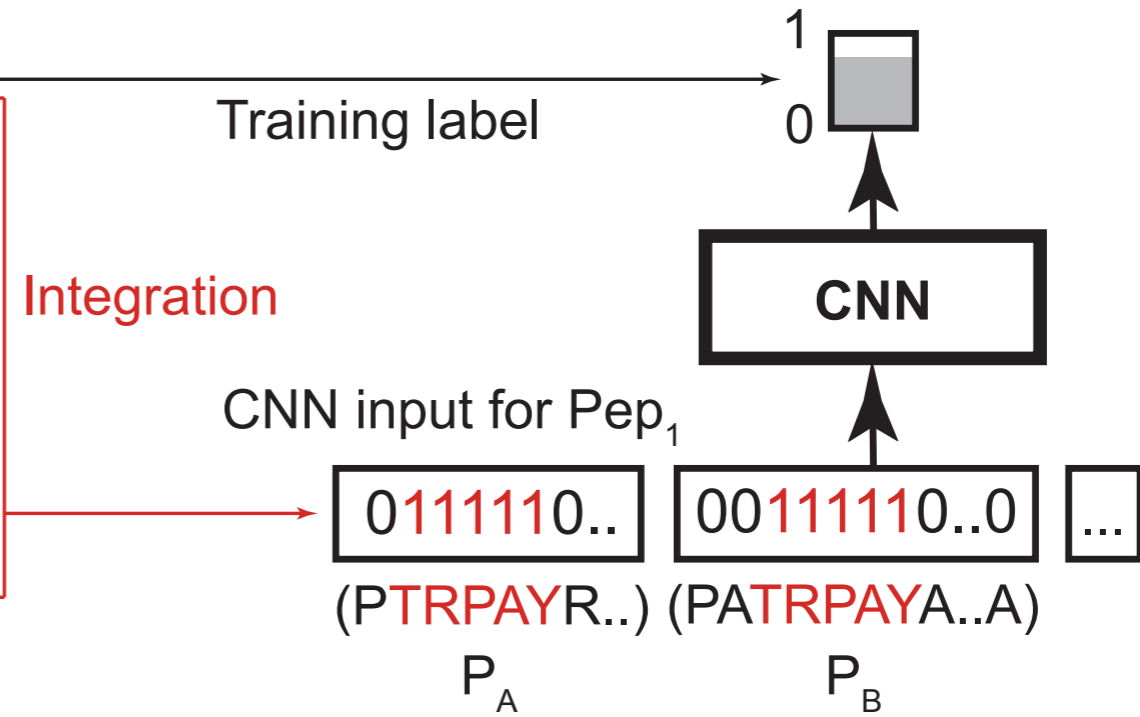
# DeepPep - method overview

## A. Training data preparation

	ID	Peptide	Prob
Observed peptides	Pep <sub>1</sub>	TRPAY	0.8
	Pep <sub>2</sub>	PATR	0.2
	...	...	...

	Protein	Sequence
All proteins	P <sub>A</sub>	PTRPAYR...
	P <sub>B</sub>	PATRPAYA..A
	...	...

## B. Train model to predict peptide probability given protein sequences

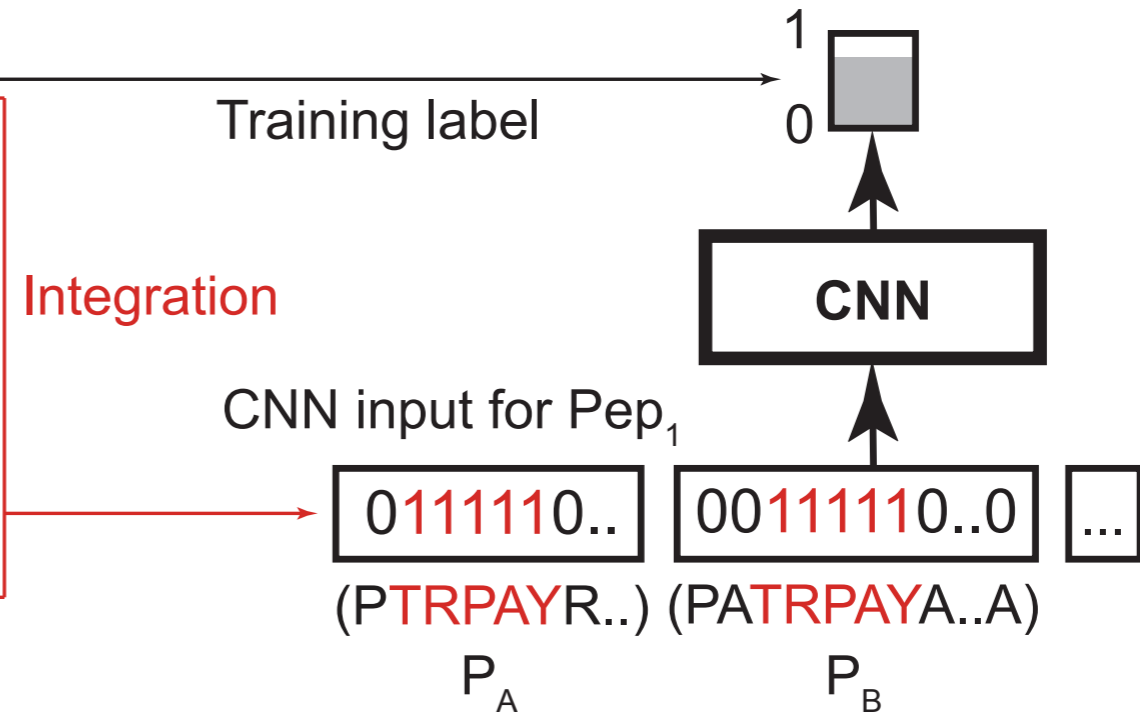


## A. Training data preparation

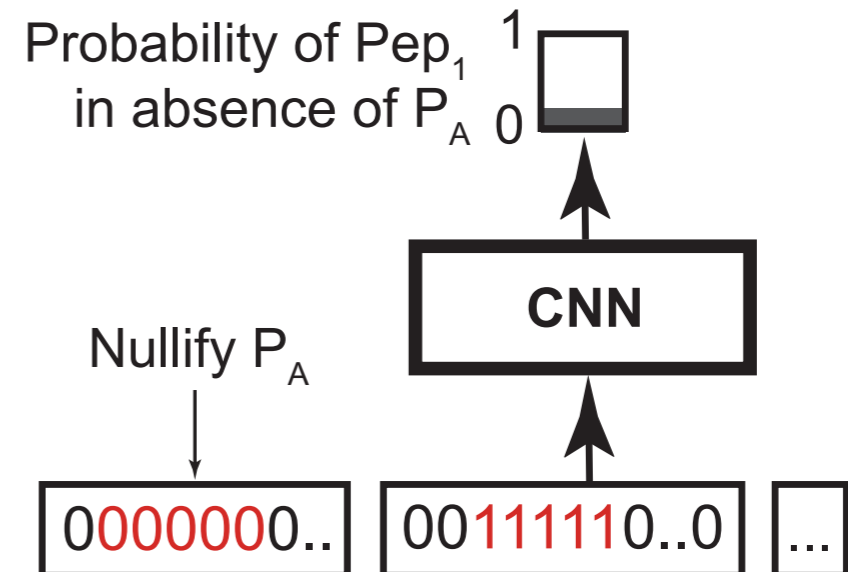
	ID	Peptide	Prob
Observed peptides	Pep <sub>1</sub>	TRPAY	0.8
	Pep <sub>2</sub>	PATR	0.2
	...	...	...

	Protein	Sequence
All proteins	P <sub>A</sub>	PTRPAYR...
	P <sub>B</sub>	PATRPAYA..A
	...	...

## B. Train model to predict peptide probability given protein sequences



## C. Predict probability of peptide i in absence of protein j for all i and all j

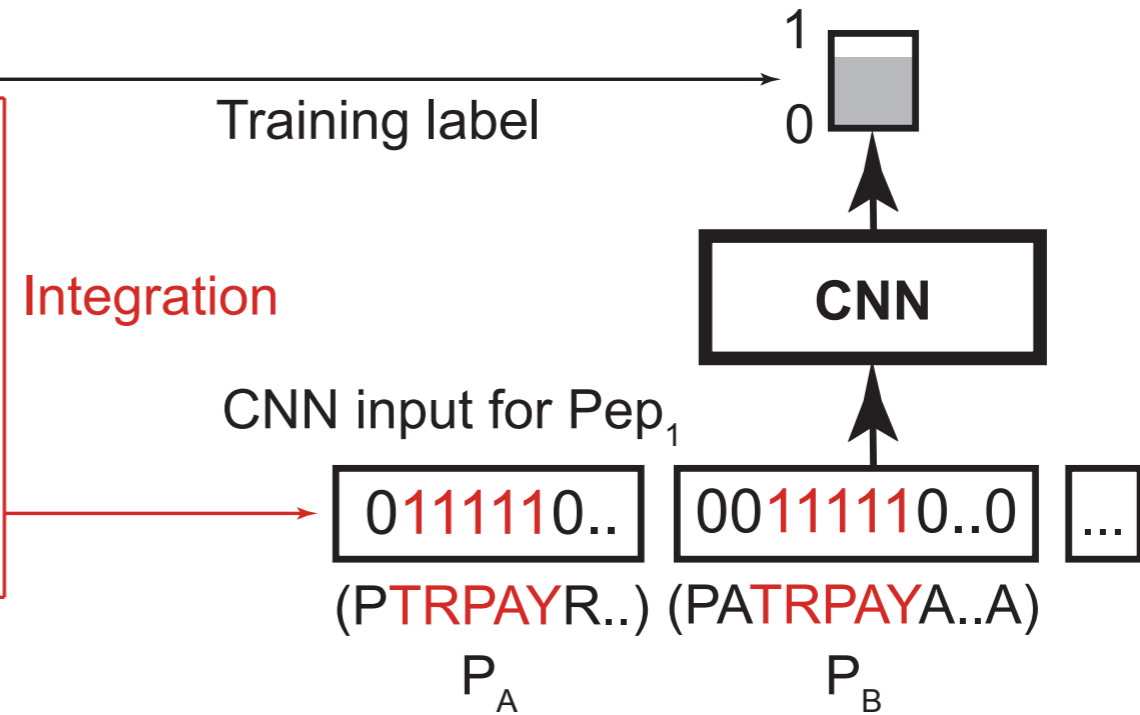


## A. Training data preparation

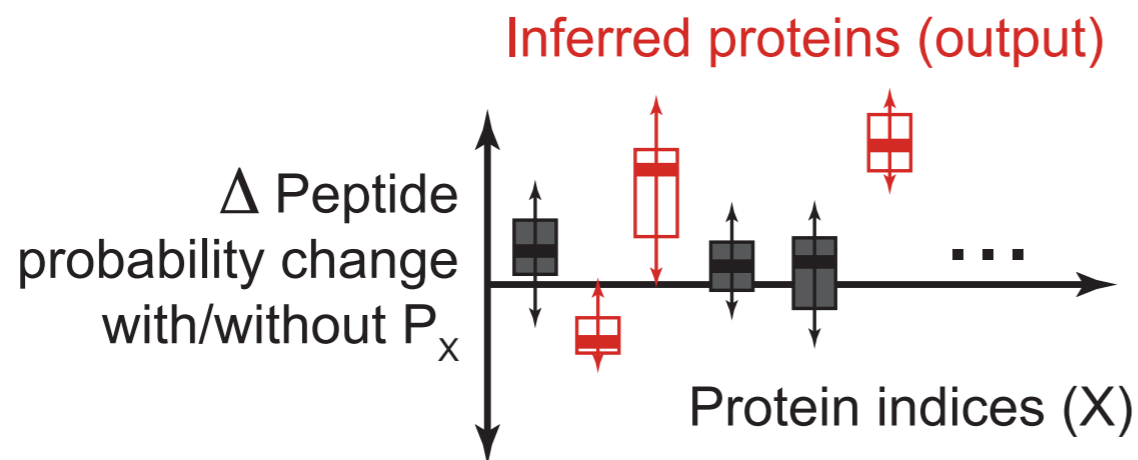
	ID	Peptide	Prob
Observed peptides	Pep <sub>1</sub>	TRPAY	0.8
	Pep <sub>2</sub>	PATR	0.2
	...	...	...

	Protein	Sequence
All proteins	P <sub>A</sub>	PTRPAYR...
	P <sub>B</sub>	PATRPAYA..A
	...	...

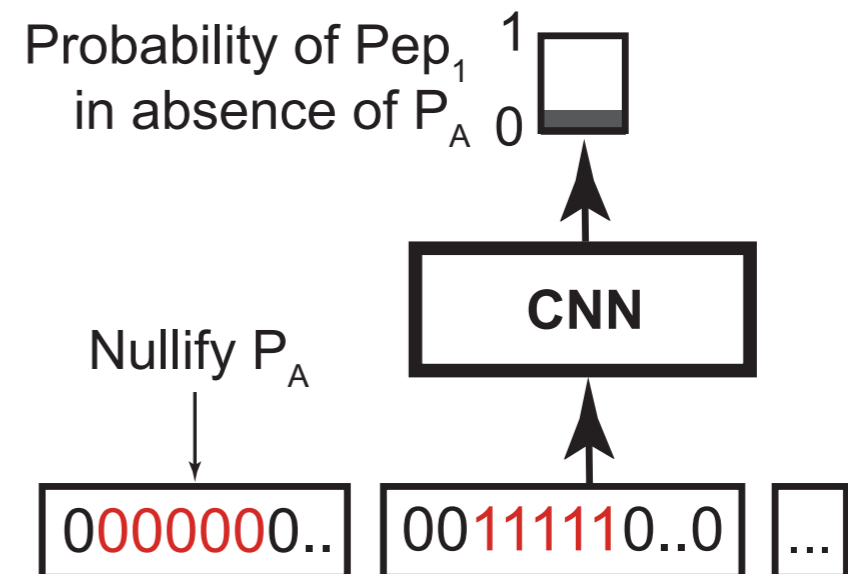
## B. Train model to predict peptide probability given protein sequences



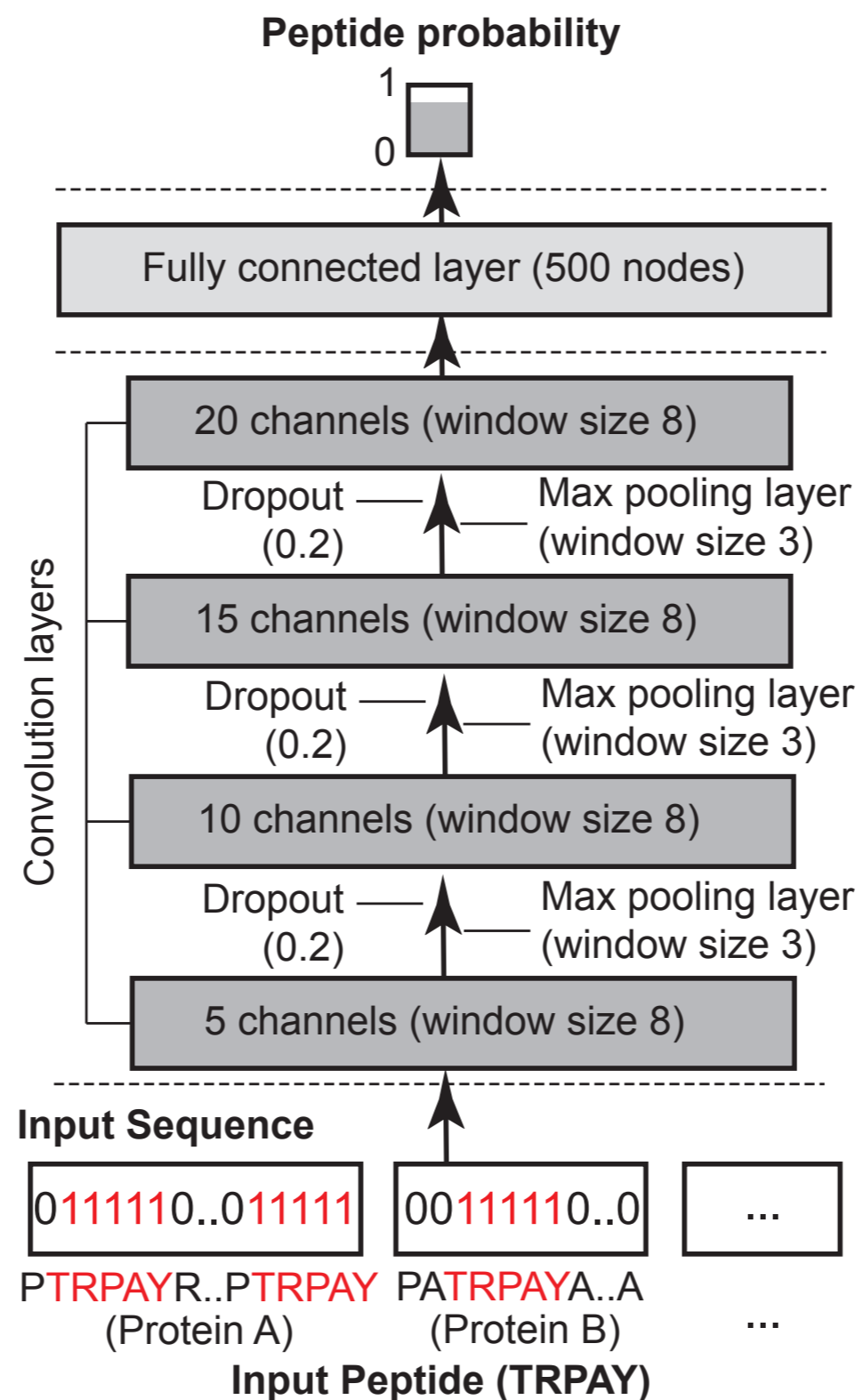
## D. Infer protein set with the most pronounced effect



## C. Predict probability of peptide i in absence of protein j for all i and all j



# DeepPep - model architecture



**1)** Many thanks to ...

Deep-learning environment setup : ***Galen Arnold***

GPU environment setup : ***JaeHyuk Kwack***

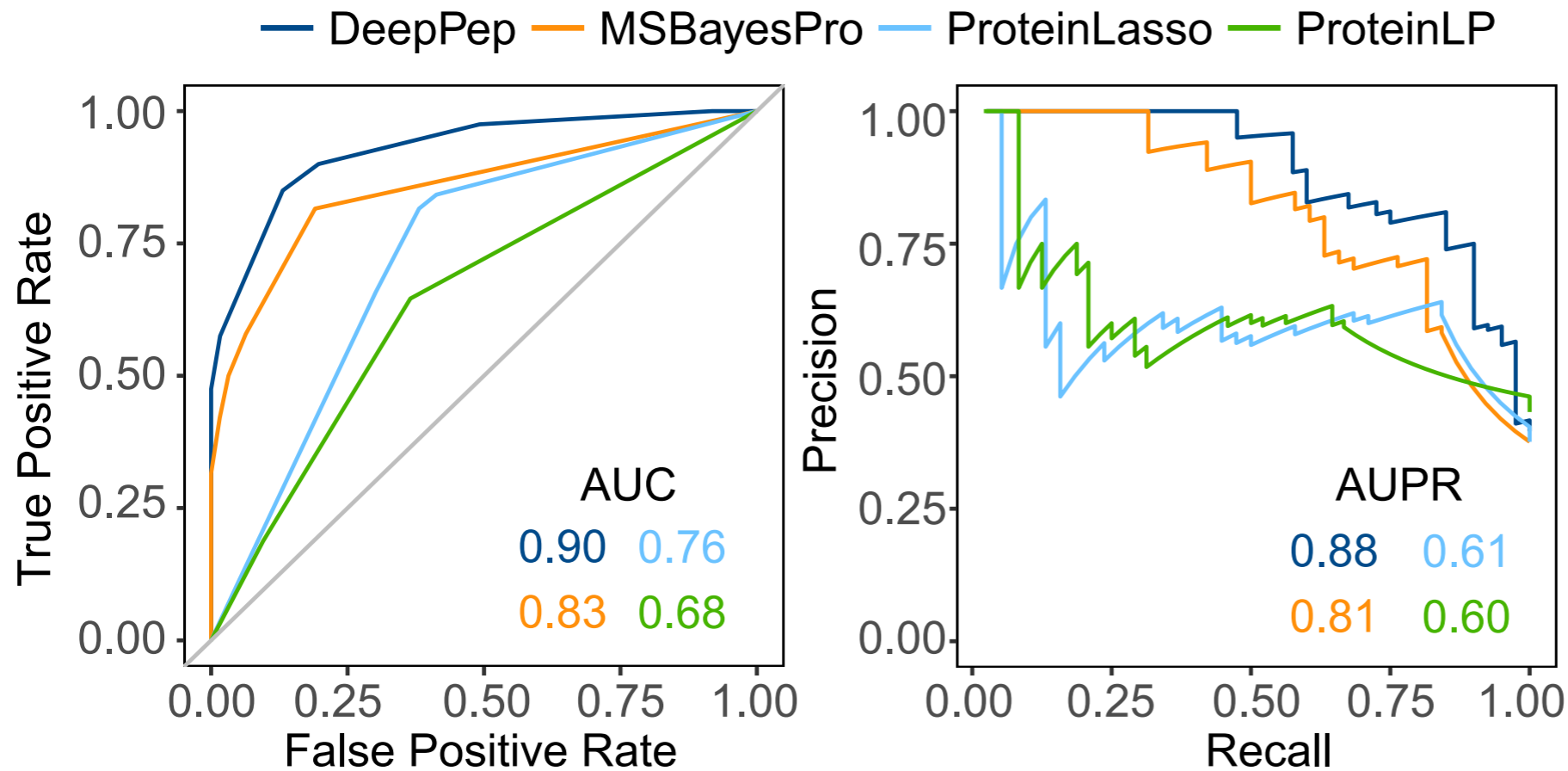
Others : ***Brett Bode, Tom Cortese***

**2)** BW GPU (XK) computes DeepPep **48** times faster than BW CPU (XE)

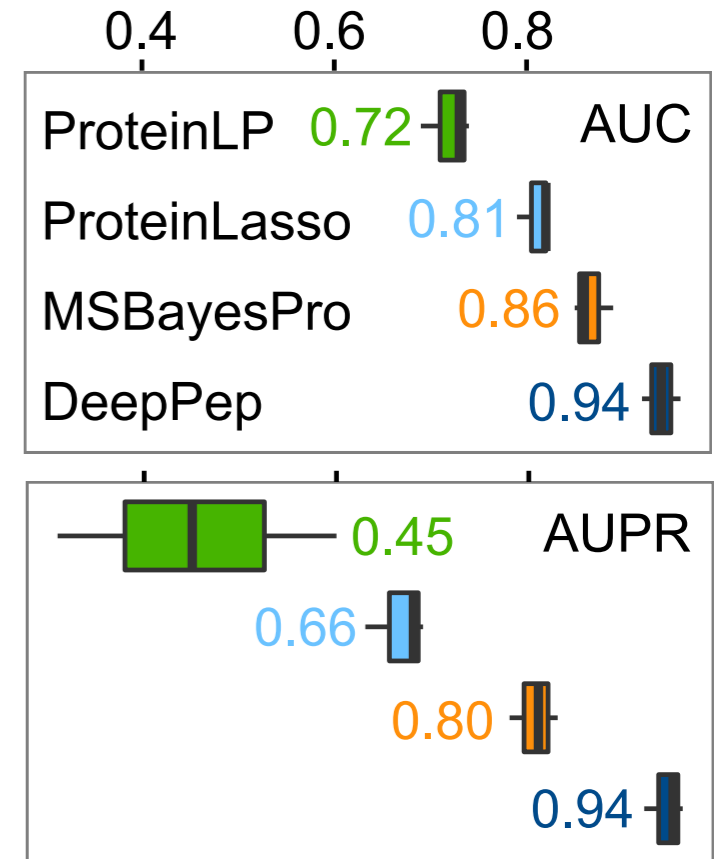


DeepPep outperforms three existing methods across 31 different datasets

## Sigma49 Dataset

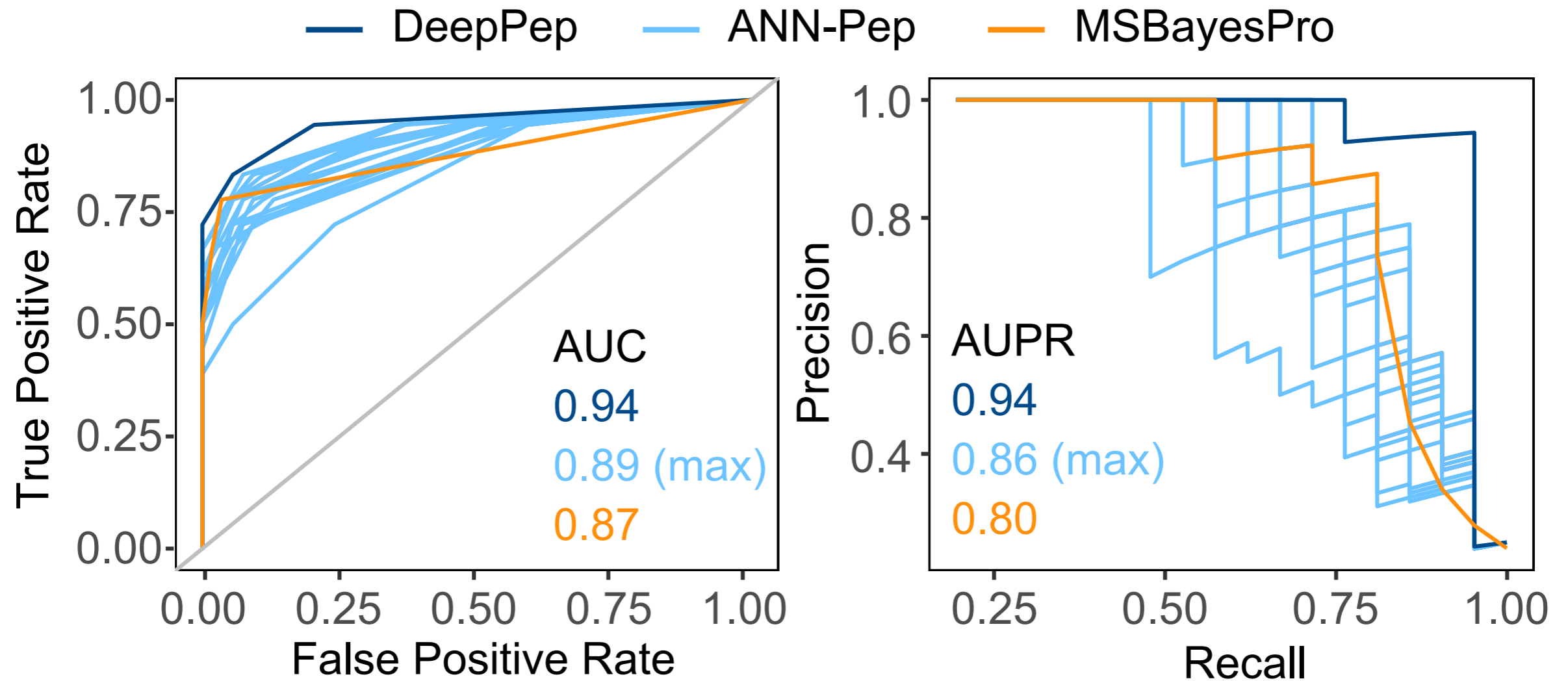


## ISB 30 Datasets

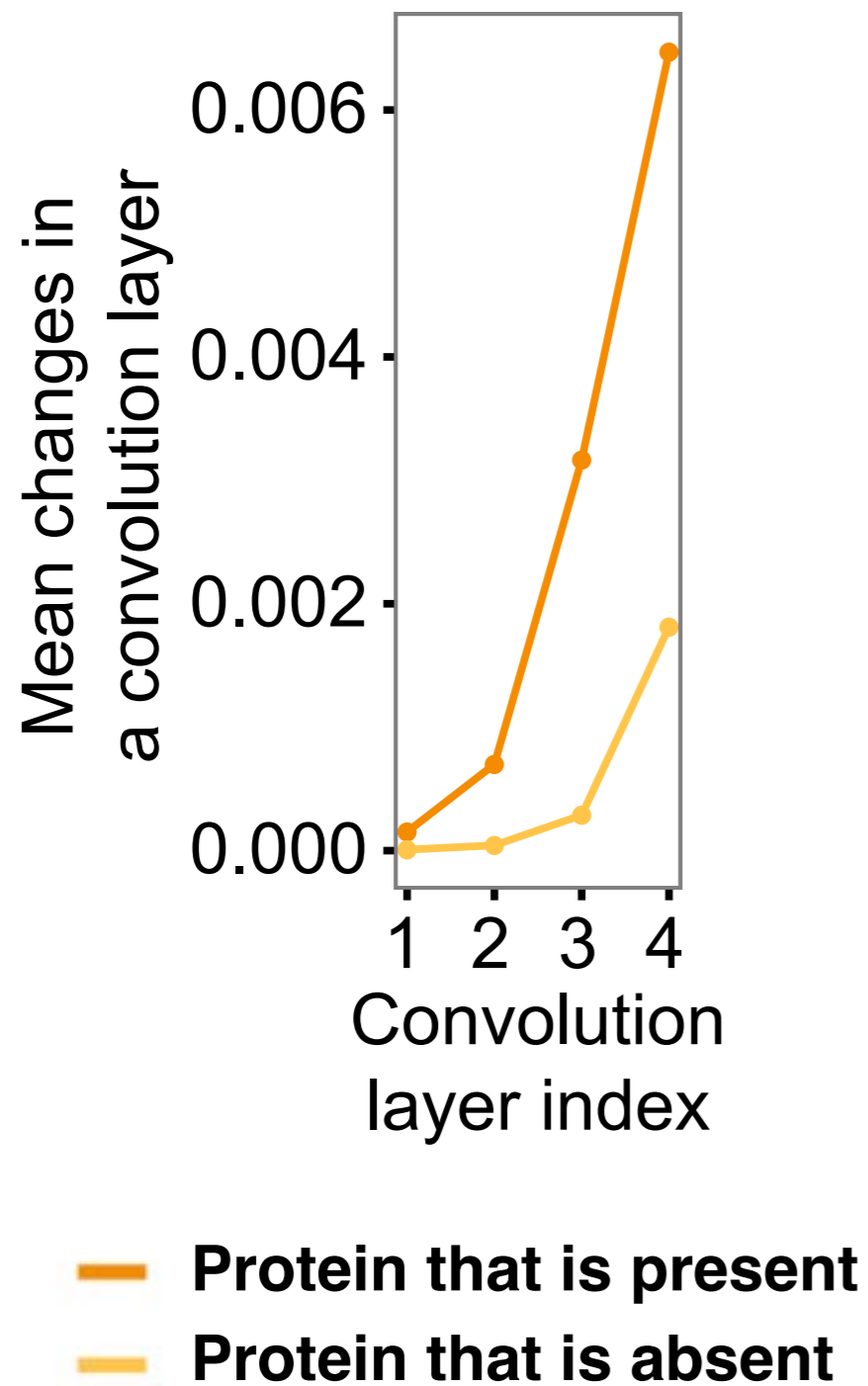


# RESULTS - method comparison

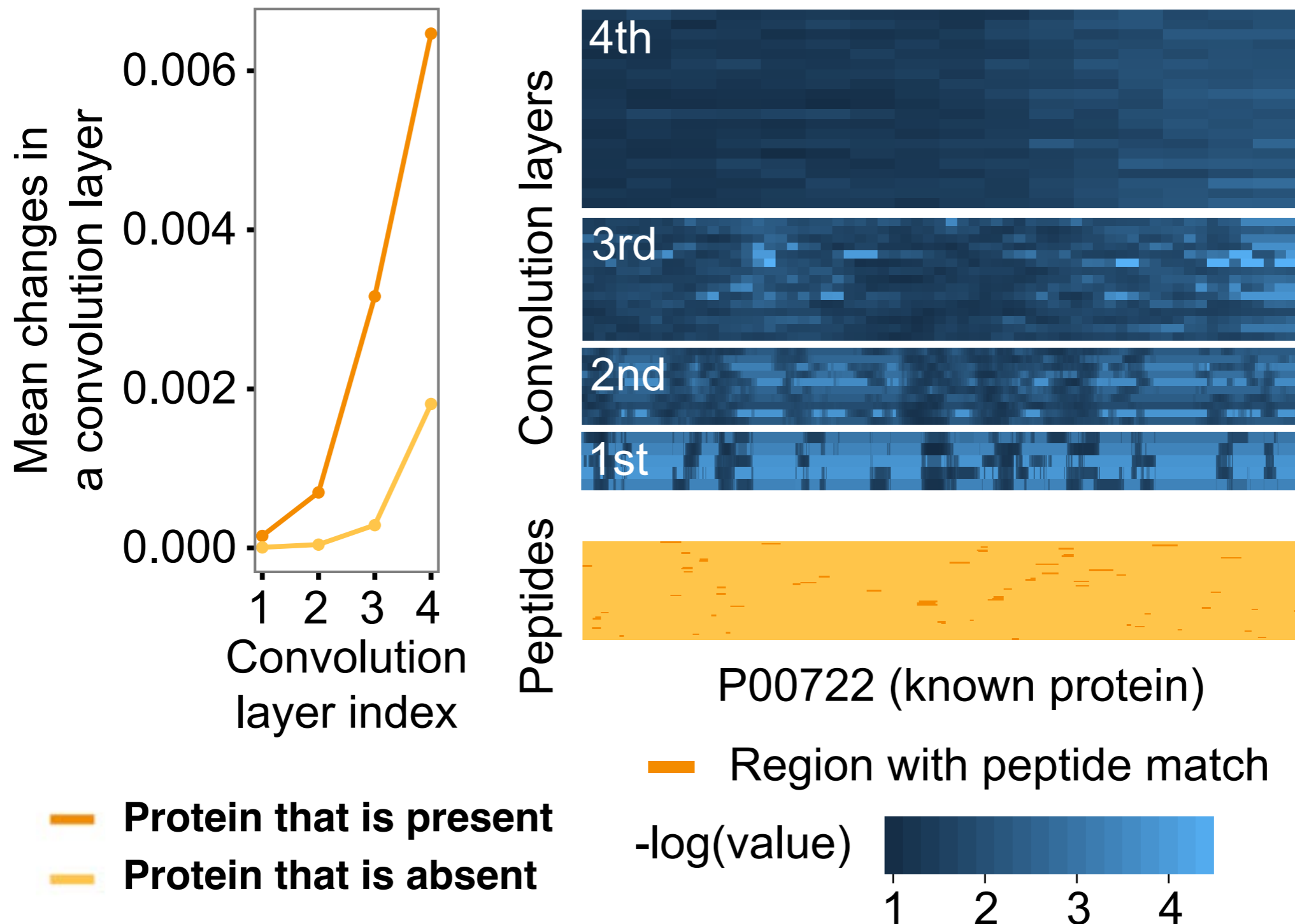
*DeepPep* outperforms plain *Artificial Neural Networks* demonstrating importance of convolutional effects in protein inference.



*Sensitivity is enhanced as we go into deeper layers*

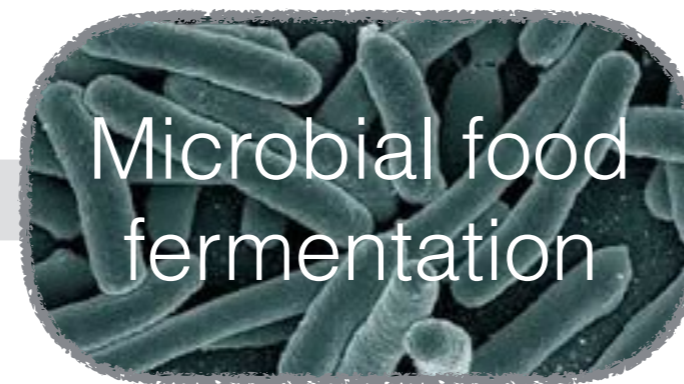


*Sensitivity is enhanced as we go into deeper layers*



# Microbial food fermentation in our life

## Natural products



## Fermented foods



# Microbial food fermentation in our life

## Natural products



Let's reveal science behind



## Fermented foods



# Microbial food fermentation in our life

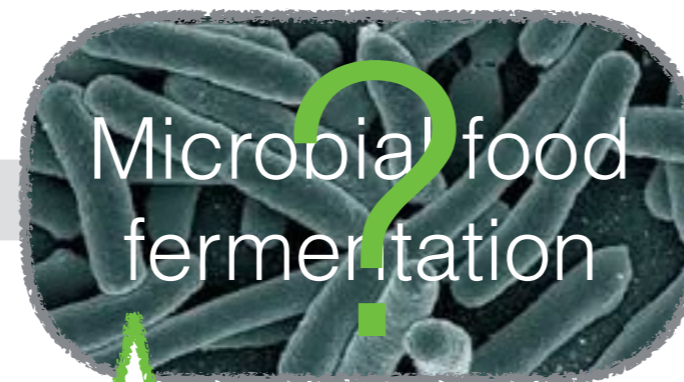
## Natural products



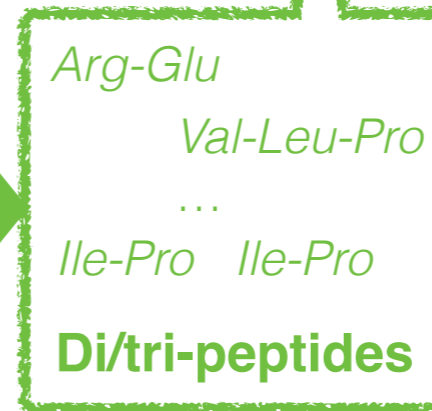
## Fermented foods



Let's reveal science behind



Microbial food fermentation



Protein degradation

# Microbial food fermentation in our life

Natural products



Let's reveal science behind



Fermented foods



? ?  
?  
**Proteins**

*Arg-Glu*  
*Val-Leu-Pro*  
...  
*Ile-Pro Ile-Pro*  
**Di/tri-peptides**



Protein inference



1) Extend DeepPep for real large datasets

which can be facilitated by **distributed learning from multiple nodes**

2) Apply the extended method to real dataset **for biological discovery**

# Acknowledgements

Advisor : Ilias Tagkopoulos (UC Davis)

Blue Waters Support Team