## LipiDetective shows that

# deep learning models are able to identify

# lipid fragmentation patterns

Logo created with BioRender.com



Presenter Vivian Würf

#### 1. Identify lipids independent of

- $\rightarrow$  collision energy
- → platform
- → reference spectra

#### Motivation

- 2. Integrate into database matching pipeline
  - $\rightarrow$  to improve estimation of false discovery rates (FDR)
  - → for more and better identifications at lower FDR
  - $\rightarrow$  increasing the quality and reliability of data analysis

#### Introduction

- Hypothesis: artificial neural networks can learn characteristic lipid fragmentation patterns in mass spectra
- Aim: test the ability of neural networks to identify lipid species from MS2 spectra with pilot model LipiDetective
- **Benefit:** enables analysis of more complex datasets in less time and could assist in the estimation of false discovery rates

### Methods

- Training on **reference dataset** containing 49 different phospholipid standards → 6942 MS2 spectra total
- 13 different **collision energies** from 20 to 50 eV in 2.5 eV steps
- Three **preprocessing steps**: resampling, baseline correction and normalization
- Two methods of **cross validation**: via lipid species and via collision energy
- Feedforward and convolutional neural network implemented with PyTorch framework





- Prediction output: headgroup and side chain masses
- Hyperparameter tuning and inclusion of dropout layers to optimize network performance
- Common definition of accuracy not well suited in this case
- → Implementation of new measure called closeness

### Results

#### Feedforward Network

- **Overfitting** after the 30th epoch  $\rightarrow$  might be good time point for early stopping
- Network shows ~70 % accuracy for predicting lipids that it encountered before at different collision energy (cross validation via collision energy)
- Model struggles with identifying lipids it has never seen before (cross validation via lipid species)
- Higher resolution needed to distinguish between fatty acids one desaturation step away from each other
- Performance might be improved by **expanding** the dataset to include more fragmentation spectra of different lipids

18:0/22:6 | 18:1/16:0 | 18:1/18:0 | 20:0/20:0 | 22:0/22:0



#### Learning Process





#### Network Performance





18:0/18:0

#### **Convolutional Network**

- Less overfitting in convolutional network compared to feedforward network
- Similar accuracy as feedforward network, but performance still needs to be improved via hyperparameter tuning

### Conclusion

- Pilot model shows that using neural networks for identifying lipid species from fragmentation spectra is generally feasible
- Further optimization of the network parameters is crucial for enhancing predictive performance
- Increasing the amount of training data is key