BIOGNOSYS
NEXT GENERATION PROTEOMICS

Biognosys AG
Wagistrasse 21
8952 Schlieren (Zurich)
Switzerland

t: +41 (0) 44 738 20 40
f: +41 (0) 44 738 20 49
info@biognosys.com
www.biognosyscom

Library Based Search Improves Protein Identification from Isobarically Labeled Single Cells

Tikira Temu, Erwin Schoof, Tejas Gandhi, Benjamin Furtwängler, Bo T. Porse, Oliver M. Bernhardt, Lukas Reiter

Introduction

Large-scale single-cell analyses are of fundamental importance to capture biological heterogeneity within complex cell systems, and to understand diseases. Till recently, this type of analysis was limited to RNA-based technology. But since proteins are the cellular workhorses, we require a workflow that is capable of consistently identifying 1000s of proteins per cell across thousands of individual cells in a timely manner. This is challenging as, due to the low sample amounts, we cannot deploy the usual tricks of boosting identification such as fractionation. Here we investigate how using spectral libraries in a single cell proteomic analysis using isobaric labeling has the advantage of reducing the search space and therefore improving performance both in speed and sensitivity (Lam et. Al. 2007).

Methods

Two different human cell systems were used. One is a diluted bulk prepared as explained in Schoof et. al., 2021, and the other is a diluted pool of samples treated with a variety of drug inhibitors. The samples were labeled with TMTpro and acquired on an Eclipse or an Exploris 480 with FAIMS Pro cycling between CVs of -50V and -70V. Identification is performed with 1% FDR at PSM, peptide, and protein levels. SpectroMine (Biognosys) uses normalized retention time (iRT) (Bruderer et al. 2016) available from the libraries as well as deep learning (DL) scores to improve the identification. Here we propose to first create a library of single-cell data, which is then used to search unboosted single-cell runs.

Preliminary Data

We created a library with SpectroMine out of 36 human single-cell injections, a mixture of unboosted and boosted unfractionated single-cell samples. The library has a size of 15,723 precursors and 2,240 protein groups. We chose a subset of 6 unboosted single-cell runs to apply separately to the library, and compared those results to searching the runs with a human uniprot FASTA file (20’367 proteins). Half of the samples were measured with a 2h gradient, and the second half with a 1h gradient. Boosting samples by using a carrier channel can lead to interference, which is why using unboosted samples has an advantage (Cheung et.al, 2020). This also allows the usage of all TMTpro channels for single-cell samples, i.e. up to 18 samples per run. All analyses were done with default search settings as provided by SpectroMine’s preconfigured TMTpro and spectral library analysis pipeline.

Biognosys AG
Wagistrasse 21
8952 Schlieren (Zurich)
Switzerland

t: +41 (0) 44 738 20 40
f: +41 (0) 44 738 20 49
info@biognosys.com
www.biognosyscom

When searching the runs with the FASTA file, SpectroMine identified on average 4,028 precursors and 1,216 protein groups using 1% FDR on all three levels (PSM, peptide, protein) for the unboosted 2h gradient samples. Applying the same search settings to the 1h gradient samples resulted in an average precursor identification of 2,930, and an average protein identification of 1,015. Using a spectral library gives a boost in precursor and protein identification. For the 2h gradient samples we identify on average 12% more precursors and 13% more proteins. Applying the spectral library to the 1h gradient samples resulted in an average increase of 13% more precursors and 14% more proteins. On average 90% of the identified proteins have all 12 assigned TMTpro channels identified (without applying imputation).

Additionally, the library search was on average more than 50 % faster than using a protein database.

Novelty aspect

Library searching of TMTpro-labeled single-cell data, while making use of normalized retention time (iRT) and deep learning scores.