Pattern Recognition plus Elute to Wash Ratios Identify Proteins of Interest in Proximity Labeling Experiments

Jordan B. Burton [1,2,5], Nicholas Carruthers [3], Zhanjun Hou [4], Larry H. Matherly [4], & Paul M. Stemmer [2]

1.Wayne State University, Department of Chemistry, Detroit, MI

2. Wayne State University, Institute of Environmental Health Sciences, Detroit, MI

3. University of Michigan, Bioinformatics Core Facility, Ann Arbor, MI

4. Department of Oncology, Wayne State University and Karmanos Cancer Institute, Detroit, MI

5. Buck Institute for Research on Aging, Schilling Lab, Novato, CA

PROJECT GOAL

Interpretating affinity selection experiments is complicated by high background leading to false positives. Particularly with streptavidin-based selection of biotinylated targets, even highly stringent washes with detergents leave high background proteomes. Using localization of organelle proteins by isotope tagging (LOPIT) maps helps in the interpretation of streptavidin affinity selection data through pattern recognition. A proximity labeling experiment with the Bir-A biotin ligase fused to the folate transporter was completed. Proteins of interest as potential interacting partners to the folate transporter were isolated using streptavidin beads with the final wash collected and evaluated along with the elute fractions. Mapping the subcellular localization of identified proteins and establishing a cut off for the elute to wash abundance ratio (E:W) simplifies biological interpretation.

BioID Proximity Labeling

Proximity labeling experiments create fusion proteins between the promiscuous BirA Biotin Ligase and a protein of interest. The goal is to label lysine residues on interacting proteins with biotin.

- Bir-A is expressed as a fusion with a Folate Transporter.
- Located in the plasma membrane with cytosolic C-terminus.

Biotinylated proteins that are candidate interaction partners are isolated using streptavidin beads.

- The high number of background proteins contributes to difficult interpretation of proximity labeling experiments.

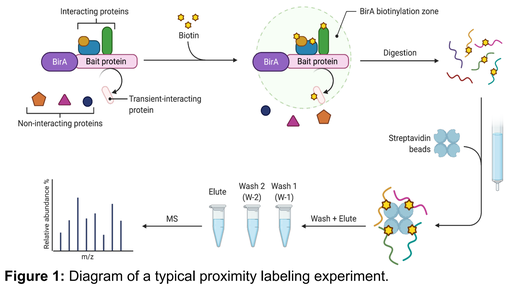

Figure 1: Diagram of a typical proximity labeling experiment.

Mass Spectrometry / Database Search

Orbitrap Fusion Tribrid (Thermo Fisher Scientific)

- 50 min data dependent acquisition (DDA) analysis
- 5 samples for each group x 3 fractions x 3 Groups = 45 samples

Proteome Discoverer 2.4

- Match-Between Runs
- Sequest Database Search Algorithm
  - Human Uniprot Protein Database
  - 20 ppm Precursor Ion Tolerance
  - 0.5 Da Fragment Ion Tolerance
- Precursor Ion Quantitation

LOPIT Map

Well documented and publicly available through Human Protein Atlas and R packages.

Coordinate-directed representation of data containing 5020 proteins.

Mass spectrometry-based localization supported by immunofluorescent microscopy.

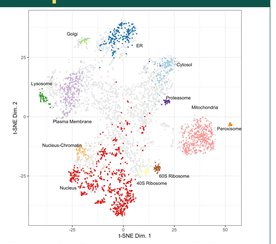

Figure 2: Distribution of LOPIT localized proteins for u2os cells from Thul, Akeson et al. 2017. *Reproduced with permission from Science.*

Database Search Results

The control group (GR) of the BioID experiment consisted of the free BirA Biotin Ligase in the cell.

The test group of the BioID experiment consisted of the BirA Biotin Ligase fused to proton-coupled folate transporter (PCFT).

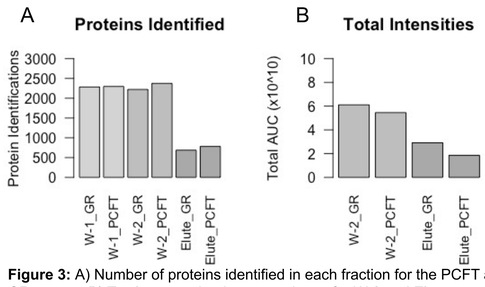

Figure 3: A) Number of proteins identified in each fraction for the PCFT and GR groups, B) Total area under the curve shown for W-2 and Elute Fractions for the PCFT and GR groups.

The number of protein identifications in the Elute fractions are much lower than in the Wash fractions as expected

- Wash fractions mainly contain background proteins
- Elute fractions contain proteins that interact with folate transporter plus an unknown percentage of false positives

The total protein abundance drops after each wash

- Data not shown for W-1 fractions as the Total AUC is over 3 orders of magnitude higher than the W-2 fractions.
- Many background proteins are washed out before the biotinylated proteins are eluted in the final fraction.

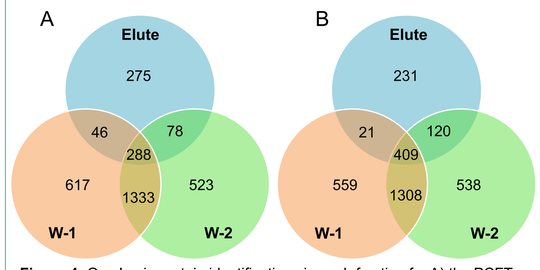

Figure 4: Overlap in protein identifications in each fraction for A) the PCFT group or B) the GR group.

The GR group has almost double the overlap between the Elute and Wash fractions as the PCFT group.

Protein Abundances

The hyperLOPIT2017 dataset contains 72 – 85% of identified proteins in the BioID experiment.

Fraction	Total Protein IDs	Total LOPIT Proteins	GR Protein IDs	GR LOPIT Proteins	PCFT Protein IDs	PCFT LOPIT Proteins
W-1	2330	1986	2297	85.42%	2284	85.42%
W-2	3249	2403	2375	79.83%	2222	79.48%
Elute	1220	887	781	75.93%	687	72.78%

Table 1. The total number of proteins identified for and contained in the LOPIT plot is shown for each fraction and group.

Abundances of proteins identified and quantified in the Wash-2 (W-2) and Elute of two groups in a BioID experiment were summed and binned before mapping to the LOPIT plot.

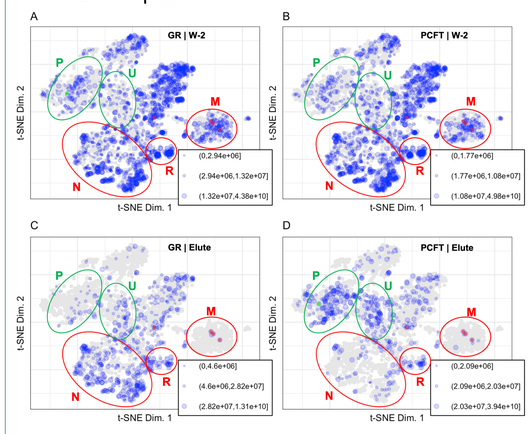

Figure 5: Binned protein abundances are mapped to the LOPIT data for A) W-2 GR (1896 proteins), B) W-2 PCFT (1766 proteins), C) Elute GR (593 proteins), and D) Elute PCFT (500 proteins). Size indicates protein abundance. Folate Transporter is shown in green; carboxylases are shown in red. Regions for plasma membrane (P) and Undefined (U) are shown in green ellipses to represent enrichment of in the PCFT elute. Regions for nuclear (N), ribosomal (R) and mitochondrial (M) are shown in red ellipses to represent background in for the PCFT elute.

GR group (Free BirA Biotin Ligase)

- Expected to display promiscuous biotin labeling of proteins in multiple subcellular locations driven by cyclosis in the cell.
- High representation of proteins in the nucleus is observed in the Elute fraction.
- Pattern of protein localization is similar in the Wash and Elute fractions.
- Lack of proteins localized to the plasma membrane and cytosol.

PCFT group (BirA-Folate Transporter Fusion)

- Expected to display localized biotin labeling of proteins that are potential interaction partners with Folate Transporter.
- High representation of proteins is observed in the plasma membrane and unknown assignment in the Elute fraction.
- Pattern of protein localization differs markedly between the Wash and Elute fractions.
- High density of proteins localized to the plasma membrane and unknown region may be true interaction partners with the BirA-Folate Transporter Fusion protein.

Fold Changes

Carboxylases are positive controls because they are the endogenously biotinylated proteins.

- Carboxylases should always be detected in the Elute and when they are detected in the W-2 it gives an indication of the loss of specific binding, i.e. loss of true positives.
- Carboxylase values can be used as a standard to establish a minimum Elute:Wash ratio necessary to have specific binding to streptavidin.

Fold change for each comparison is calculated, binned and mapped on the LOPIT plot.

- A cutoff of 3 times higher area in the Elute than in the W-2 samples is based on the ratio of carboxylases in the Elute and W-2.
- The proteins retained likely includes many false positives.

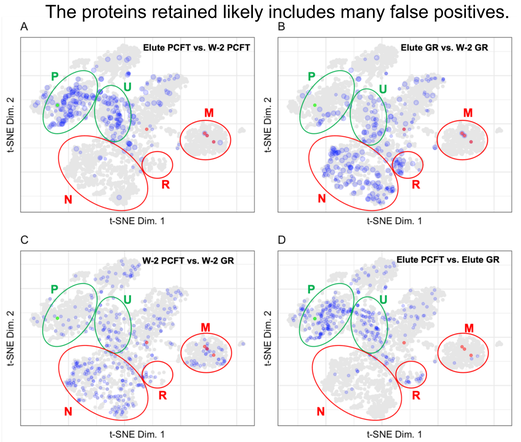

Figure 6: Plots displaying the fold change with a 3X cut-off for A) Elute PCFT vs. W-2 PCFT (238 proteins), B) Elute GR vs. W-2 GR (264 proteins), C) W-2 PCFT vs. W-2 GR (350 proteins), and D) Elute PCFT vs. Elute GR (310 proteins). PCFT is highlighted in green. Carboxylases are highlighted in red. Regions for plasma membrane (P) and Undefined (U) are shown in green ellipses to represent enrichment of in the PCFT elute. Regions for nuclear (N), ribosomal (R) and mitochondrial (M) are shown in red ellipses to represent background in for the PCFT elute.

Conclusions

The number of proteins of interest is decreased from 1220 to 310 in the resultant set of proteins that should be compared within the analysis group.

- Most proteins of interest are localized to the plasma membrane and the unknown assignment area in the LOPIT map.

Acknowledgements

National Cancer Institute (R01 CA053535)

National Institute of Health (P30 ES020957, P30 CA022453, and S10 OD010700)

Wayne State University Funding

Data available at: 10.6019/PXD032297