DART-ID Increases Single Cell Proteome Coverage

Albert Chen[1,2], Alexander Franks[3], Nikolai Slavov[1,2]

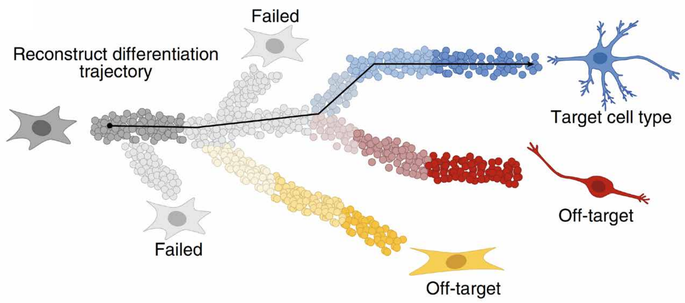
1 Department of Bioengineering, Northeastern University, Boston MA 02115, USA
2 Barnett Institute, Northeastern University, Boston MA 02115, USA
3 Department of Statistics and Applied Probability, UC Santa Barbara, CA 93106, USA

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY
DOI: 10.1101/399121

Summary

- Single cell proteomics by mass spectrometry can help answer may biological questions involving cell diversity and heterogeneity
- Peptide/protein identification is a noisy, stochastic process
- Incorporating peptide retention time (RT) within a principled Bayesian framework increases proteome coverage
- Global RT alignment method is more accurate and boosts coverage
- Manuscript: https://doi.org/10.1101/399121
- Code: https://github.com/SlavovLab/DART-ID

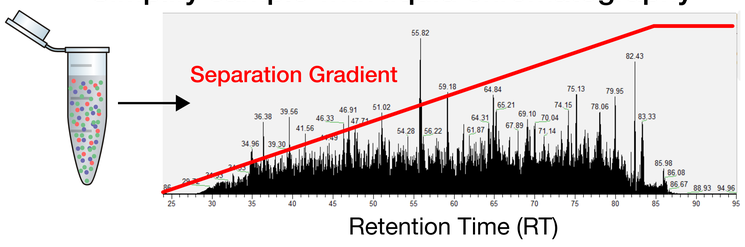
Improve Directed Cell Differentiation



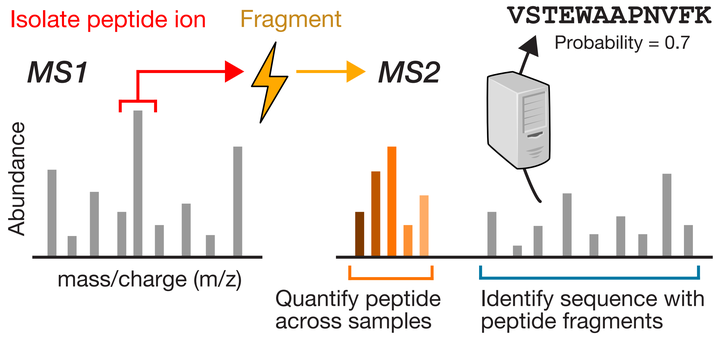
- Use single cell proteomics to profile differentiating cells
- Understand branching points, find biomarkers, and intervene to maximize differentiation efficiency

Peptide Identification by Mass Spectrometry

Simplify sample with liquid chromatography

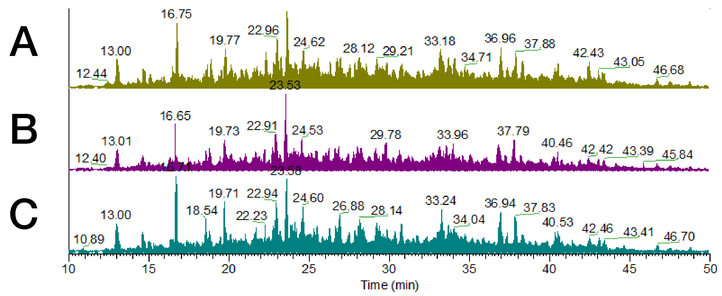


Identify and quantify peptides with fragments



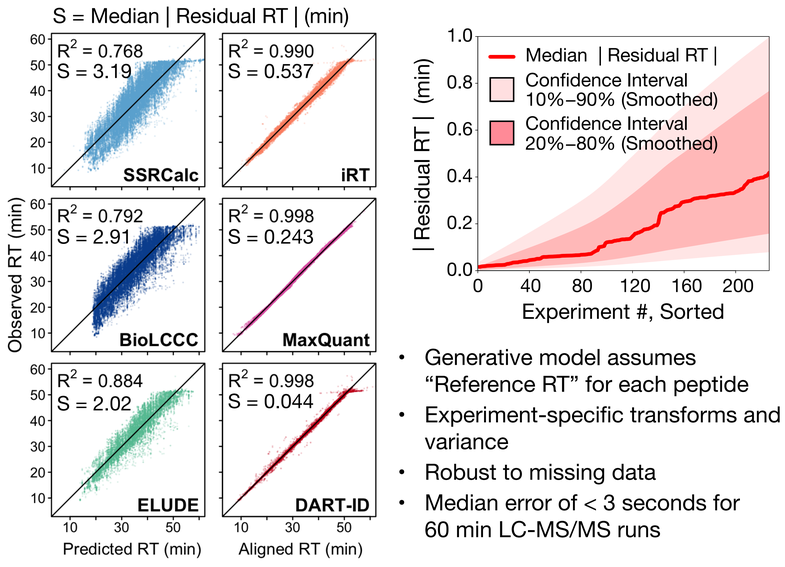
GOAL: Incorporate retention time into identification

Conventional Pairwise RT Alignments



- Conventional methods align pairwise to a "reference experiment"
- Choice of reference experiment not obvious
- Does not account for measurement errors in reference experiment

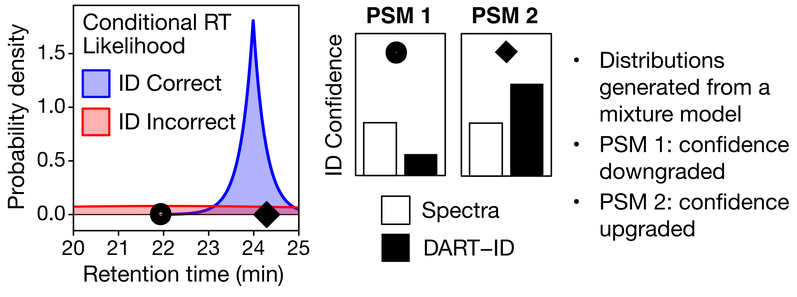
Global RT Alignment Reduces Error



- Generative model assumes "Reference RT" for each peptide
- Experiment-specific transforms and variance
- Robust to missing data
- Median error of < 3 seconds for 60 min LC-MS/MS runs

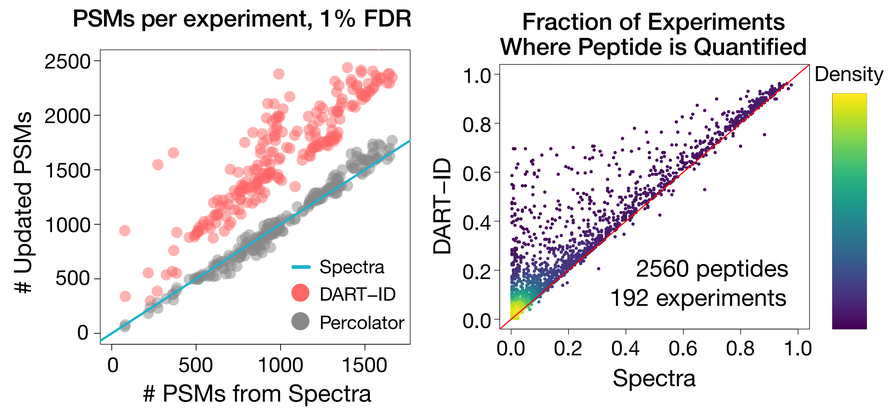
Incorporating RT Inferences into Identification with Bayes Theorem

$$P(\text{ ID correct } | \text{ RT }) = \frac{P(\text{ RT } | \text{ ID correct }) \times P(\text{ ID correct })}{P(\text{ RT })}$$

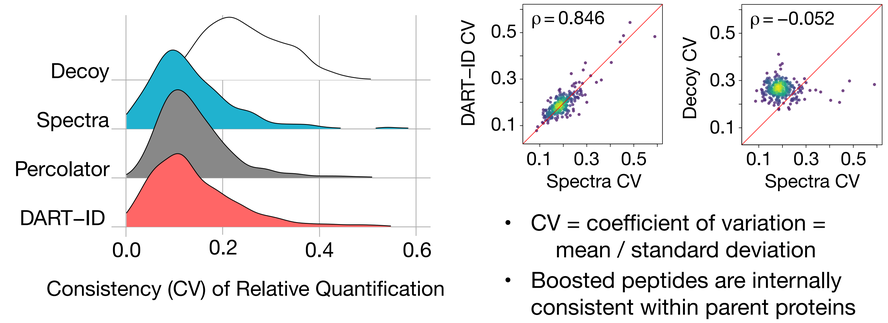


- Distributions generated from a mixture model
- PSM 1: confidence downgraded
- PSM 2: confidence upgraded

DART-ID Identifies More Peptides and Reduces Missing Data

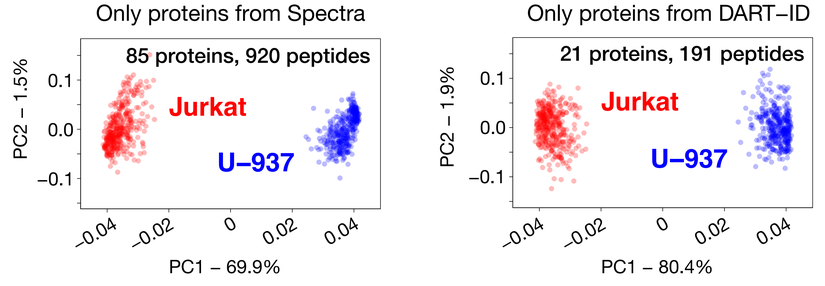


Boosted PSMs Give Consistent Protein Quantification



- CV = coefficient of variation = mean / standard deviation
- Boosted peptides are internally consistent within parent proteins

DART-ID Proteins Separate Cell Types



- Separation of blood cancer cell proteomes by Principal Component Analysis (PCA)

Conclusion

DART-ID takes advantage of reproducible retention times for peptide sequences within sets of LC runs to greatly increase the coverage of single cell proteomes.

Thanks to members of the Slavov Laboratory. This work was funded by the National Institute of Health to N.S. under Project Number 1DP2GM123497-01