Mass Dynamics

Application Note

Differential Expression and Enrichment Analysis using Mass Dynamics™

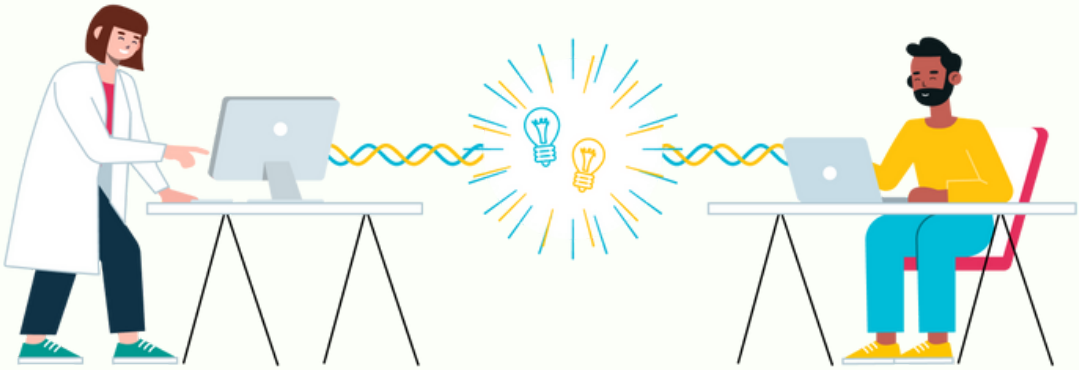
Last revised: February 2023

Why we built it

To accelerate the transition from Mass Spectrometry data to knowledge generation to inform validation studies, facilitate new hypotheses and enable more discoveries.

How we built it

- Seamless linking of LC-MS data to Gene Set Enrichment Analysis (GSEA) using CAMERA
- Integrating knowledge bases including Reactome and Gene Ontology (GO)
- Enabling better collaboration between multidisciplinary team members
- Designing for easy and quick interpretation of results

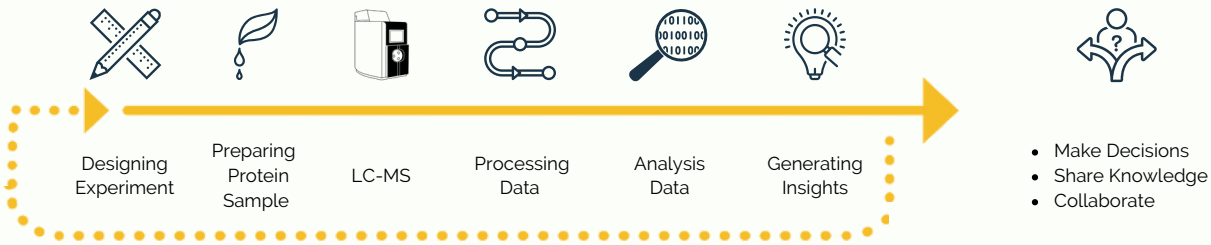


Background

In order to better understand the differences and mechanisms of diseases, a typical and standard process is to use Mass Spectrometry (MS) to make complex biological measurements of disease samples and compare these to healthy controls.

From a proteomics perspective, this encompasses a workflow that includes: design of experiments, preparing protein samples (protein purification, digestion with enzymes and sample clean up) and acquiring data using a liquid chromatography MS (LC-MS) system. Processing of the data results in a list of identified proteins and quantitative abundance changes between conditions.

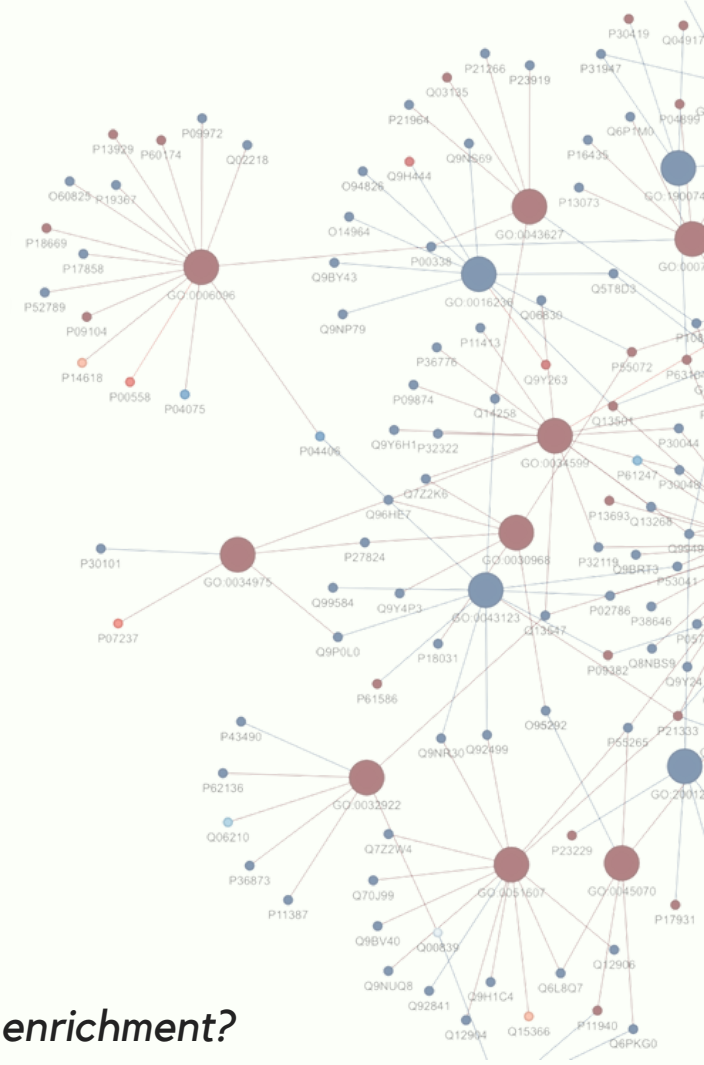
Proteins are then interrogated with statistical analyses, including differential expression, to deduce insights that describe the biological changes observed.



The final part of the workflow - generating insights - remains one of the most challenging, overlooked and time-consuming components of the workflow process. This is because an analysis would involve multiple steps, depending on the level of understanding of the MS expert and/or biologist/s undertaking the work, as well as the types of bioinformatics workflows adopted to achieve the insights. As a consequence, it requires multidisciplinary teams to identify novel outcomes, publish findings as well as design and iterate on new experiments.

This Application Note describes Mass Dynamics' approach to couple high-quality, streamlined processing of LC-MS data with the generation of insights and biological understanding through the use of enrichment analysis.

Enrichment



What is Enrichment?

Enrichment is a bioinformatics method that identifies enriched or over-represented gene sets among a list of genes, enabling scientists to gain deeper biological insight into MS analyses.

Mass Dynamics makes Enrichment Analysis more accessible by integrating multiple publicly available knowledge sources, removing the need to learn multiple tools and handling a variety of file formats.

How does Mass Dynamics perform enrichment?

Mass Dynamics identifies enriched terms from public databases based on comparing protein intensity measurements in your experiment's dataset using the Bioconductor package LIMMA [1] and the CAMERA algorithm to perform gene set enrichment using a competitive gene set method accounting for inter-gene correlation [2].

By using CAMERA, this enrichment analysis applies a modified form of the 'Gene Set Enrichment Analysis' (GSEA) [3] method adjusting for correlation in protein intensities and overlap in protein sets. Since both CAMERA and GSEA are ranked-based, a high absolute Average Fold Change (AFC) is not required for a protein set to be significantly enriched.

Construction of Gene and Protein Set libraries

Gene set libraries are created using data from many publicly available knowledge bases including UniProt, Gene Ontology (GO) and Reactome [4-6]. As the information Mass Dynamics surfaces is derived from proteomics assessment of samples, the approach assembles gene set libraries using proteins as terms, not genes. Enrichment analysis can be performed for human (txid: 9606), mouse (txid 10090), Chinese hamster - Cricetulus griseus (txid 10029) and yeast - Saccharomyces cerevisiae S288C (txid 559292) LC-MS analyses.

Characterising epithelial to mesenchymal transition and the link to drug resistance

Delving deeper to elucidate mechanisms of disease progression

PXD002057 ("HER2 dataset") consists of LC-MS data from 2 cancer cell lines, a parental SKBR3 cell line and another cell line derived from the first which is resistant to human epidermal growth factor receptor 2 (HER2)-targeted therapy (AZD8931-resistant). Label-free LC-MS was adopted to investigate regulators/markers of phenotype transition (epithelial-to-mesenchymal, EMT) observed in resistant cell lines [7].

The dataset contains Q-Exactive LC-MS raw data for an experiment with 2 samples, each with 3 replicates. We have previously used this dataset as a benchmark of our MD 1.0 Discovery service [8], which showed the capability to achieve reliable results for protein quantification, emulating Perseus on benchmark datasets over a wide dynamic range.

Enrichment: workflow

Example data process

Clicking "Create Experiment" allows the ability to upload raw or pre-processed files (e.g. MaxQuant txt files) to Mass Dynamics. Once complete, you will then be able to visualise your results and review the data quality, selecting modules or pre-defined templates, such as the 'Quality Control Report'.

Tabs can be customised with multiple modules to interrogate the results, including volcano plots for pairwise analysis (Figure 1) to outline the differential expression results of proteins.

The volcano plot allows the user to interactively explore the results of their experiment, generate protein lists, and interrogate proteins in more detail using other modules, such as the violin plot. Generated protein lists can be used for over-representation analysis to identify pathways relevant to the study. The entire dataset can also be assessed using Enrichment Analysis by selecting the 'Gene Set Enrichment Analysis (GSEA)' template.

Figure 1. Customised Experiment View with Volcano Plot.
Selected proteins in the volcano plot are automatically displayed in other modules that have been brought into the tab, such as the violin and list table modules.

HER2 dataset results

What are gene sets?

A gene set usually consists of a list of genes that share common functions or are annotated in the same pathway or network [8].

Links are provided to gene set/pathway origin databases to enable users to gain a detailed understanding of the biological meaning of the differentially expressed proteins from the analysis.

How to read the Enrichment Results

Enrichment results yield a series of gene set statistics represented in a linked table/plot format. Each table is specific to the selected comparison - such as AZD8931-resistant (up) vs parental SKBR3 (down) in relation to specific GO databases (e.g. biological processes) or Reactome. The statistical analysis ranks the expression of proteins in each set relative to all proteins in the dataset, indicating sets that are enriched.

The protein-set volcano plot shows the $\log_{10}$(p-value) on the y-axis and Average Fold Change (AFC) on the x-axis for each protein set. Researchers can sort the corresponding table according to p-value or select directly in the plot. Unlike a protein-centric volcano plot, the protein-set plot can show significant protein sets with low average fold change, as protein sets may contain a mix of expression levels but still many that are ranked highly.

Figure 2. Typical Enrichment Analysis output

Biological insights into Epithelial to Mesenchymal Transition (EMT) linked drug resistance

Whilst HER2 targeted therapies can result in significant tumour regression in HER2 positive breast cancer, drug resistance is common (70% developing resistance within 2 years) for multiple therapeutics, including monoclonal antibody and small molecules targeting HER2. Understanding the mechanisms behind resistance is imperative to define appropriate therapeutic response, such as the use of combined treatment strategies to mitigate resistance responses. The last 20 years has seen a major research focus on this, as highlighted by the number of articles in PubMed.

Analysis of the HER2 dataset using Mass Dynamics allows the interrogation of enriched pathways observed due to AZD8931 resistance, which was previously linked to EMT transition. Enrichment Analysis of AZD8931-resistant (up) vs parental SKBR3-sensitive (down) LC-MS data, using the Reactome database, visualised the 123 significant sets involved in cancer progression/EMT transition, including the 'Regulation of expression of SLITs and ROBOs (R-HSA-9010553)', NEDDylation (R-HSA-8951664)', and Wnt signalling (positive regulation of canonical Wnt signaling pathway (GO:0090263).

Protein Set Library	Significant Sets (FDR < 0.05)
GO: BP	23
GO: CC	11
GO: MF	0
Reactome	89

GO: Biological process (BP)

NAME	AFC	ADJ P VALUE	P VALUE	
translational initiation	0.02	0.0068	1.7769e-06	View
post-translational protein modification	0.3	0.0114	0.0	View
Wnt signaling pathway, planar cell polarity pathway	0.27	0.0114	0.0	View
pre-replicative complex assembly	0.57	0.0114	0.0	View
NIK/NF-kappaB signaling	0.36	0.0114	0.0	View
positive regulation of canonical Wnt signaling pathway	0.61	0.0114	0.0	View
protein deubiquitination	0.21	0.0114	7.8074e-06	View

Reactome

NAME	AFC	ADJ P VALUE	P VALUE	
Axon guidance	0.2	5.3926e-06	5.5388e-09	View
Nervous system development	0.21	5.3926e-06	7.0584e-09	View
Formation of a pool of free 40S subunits	0.2	0.0001	4.3422e-07	View
L13a-mediated translational silencing of Ceruloplasmin expression	0.14	0.0001	4.2787e-07	View
Signaling by ROBO receptors	0.07	0.0001	4.4375e-07	View
Developmental Biology	0.01	0.0001	3.448e-07	View
Regulation of expression of SLITs and ROBOs	0.1	0.0001	4.6455e-07	View

GO: Cellular components (CC)

NAME	AFC	ADJ P VALUE	P VALUE	
mitochondrial matrix	-2.49	6.9572e-06	8.3121e-09	View
mitochondrion	-1.85	0.0001	3.5181e-07	View
focal adhesion	0.24	0.0002	8.1221e-07	View
mitochondrial inner membrane	-2.11	0.0004	1.7723e-06	View
cytosolic ribosome	0.06	0.0033	0.0	View
proteasome accessory complex	0.48	0.0201	0.0002	View
cytosolic large ribosomal subunit	0	0.0201	0.0002	View

Biological insights into EMT linked drug resistance (cont)

Fast and streamlined Enrichment Analysis allows for rapid investigation into the pathways and processes involved in EMT transition and breast cancer progression. Through this approach, it is possible to identify other proteins to target for further investigations and/or define new hypotheses (based on the systemic measurements of the proteome) for subsequent experiments [9]. The strategy is also capable of detecting small up or down regulation in large sets or pathways of proteins which are unlikely to be obvious from a manual inspection. This was achieved without the need for researchers to manually process/export the results and thus will expedite visualising key insights and experiment iteration and validation studies.

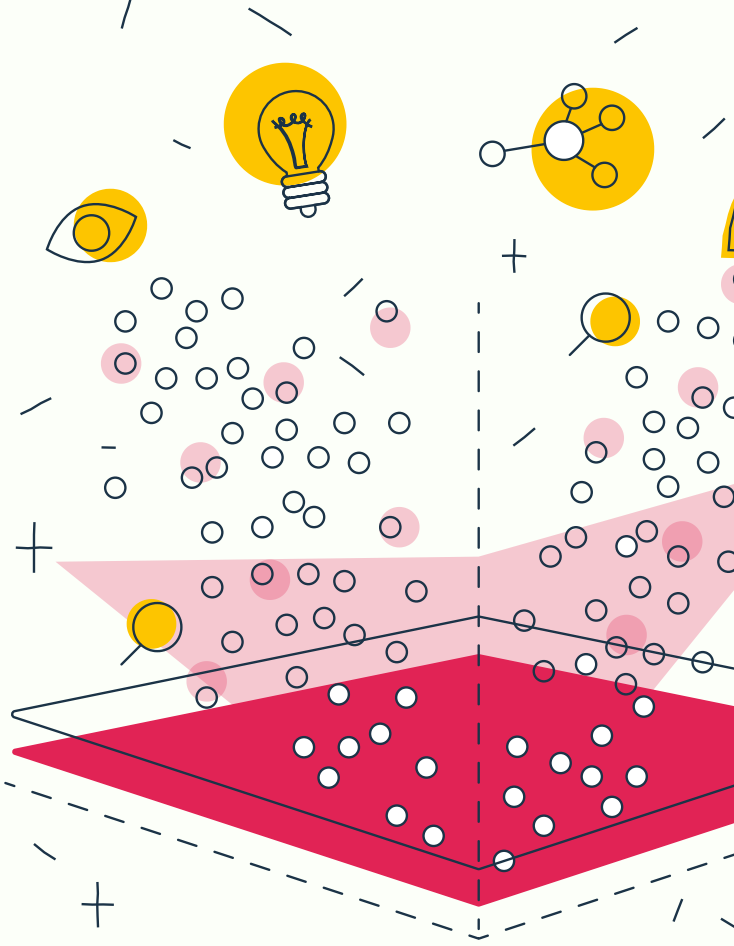
Enrichment of 'Regulation of expression of SLITs and ROBOs' (106 observed proteins out of 170), a regulator of cell function, was previously reported to be involved in breast cancer progression. Previous studies investigating the role of this pathway have injected exogenous SLIT2 expressing cells into nude mice, which reduced breast carcinoma size by 65% [10]. In contrast to these reports one study has implied that SLIT2 may act as a chemoattractant and induce brain metastasis of breast cancer cells [11]. Analysis with the Mass Dynamics enrichment feature rapidly highlighted the potential relationship of AZD8931 resistance to SLIT/ROBO function and provided key insights that can be further validated and used to create new novel hypotheses, accelerating scientific insights.

Similarly, NEDDylation has recently been shown to downregulate estrogen-related receptor beta (ERRβ), through regulation of the Skp, Cullin, F-box containing (SCF) complex. thus promoting tumeriogenesis and disease progression [12].

Enrichment Analysis also identified significant sets previously described as being linked to EMT transition. This includes the Wnt/β-catenin pathway, in which impaired signalling is known to contribute to EMT transition and cancer progression [12-15]. A recent study (Wang et. al. 2021, [16]) has shown that silencing of KIF3B, a sub-family member of Kinesin super family proteins (KIFs), could suppress breast cancer progression through regulation of the Wnt/β-catenin pathway and thus EMT transition.

Summary

- Seamless linking of LC-MS data to differential expression analysis with LIMMA and Gene Set Enrichment using CAMERA
- Faster evaluation of mechanisms behind EMT transition, breast cancer disease progression, and resistance
- Acceleration of knowledge generation by:
  - allowing fast interpretation and sharing of results to a multidisciplinary team
  - informing validation studies
  - facilitating new hypothesis generation



About Mass Dynamics™

Our mission is to free humanity and society from the burden of disease by unlocking the magic of Mass Spectrometry (MS) and the power of existing biological knowledge. We do this by delivering a powerful software platform that seamlessly connects multi-disciplinary life scientists to answer biological questions and understand the building blocks of life - better, faster and easier.

www.massdynamics.com

hello@massdynamics.com

github.com/massdynamics

Sign up for free membership, start today: app.massdynamics.com

References

1. Ritchie, ME., Phipson, B., Wu, D., Hu, Y., et. al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Research, Volume 43, Issue 7, Page e47, https://doi.org/10.1093/nar/gkv007

2. Wu, Di, and Smyth, GK. 2012. "Camera: A Competitive Gene Set Test Accounting for Inter-Gene Correlation." Nucleic Acids Research. https://doi.org/10.1093/nar/gks461.

3. Subramanian, A., Tamayo, P., Mootha, VK., Mukherjee, S., et.al. 2005. Gene Set Enrichment Analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS, 102 (43) 15545-15550; DOI: 10.1073/pnas.0506580102

4. Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., et al. 2007. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 18, 142. https://doi.org/10.1186/s12859-017-1559-2

5. The Gene Ontology Consortium. 2019. "The Gene Ontology Resource: 20 Years and Still GOing Strong." Nucleic Acids Research 47 (D1): D330–38.

6. UniProt Consortium. 2021. "UniProt: The Universal Protein Knowledgebase in 2021." Nucleic Acids Research 49 (D1): D480–89.

7. Creedon H., Gómez-Cuadrado L., Tarnauskaitė ., Balla J., Canel M., et al. 2016. Identification of novel pathways linking epithelial-to-mesenchymal transition with resistance to HER2-targeted therapy. Oncotarget. 2016; 7: 11539-11552. Retrieved from https://www.oncotarget.com/article/7317/text/

8. Bloom, J., Triantafyllidis, A., Burton (Ngov), B., et. al. 2021. Mass Dynamics 1.0: A streamlined, web-based environment for analyzing, sharing and integrating Label-Free Data. bioRxiv 2021.03.03.433806; doi: https://doi.org/10.1101/2021.03.03.433806

9. Chen, C.; Hou, J.; Tanner, J.J.; Cheng, J. 2020. Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. Int. J. Mol. Sci.21, 2873. https://doi.org/10.3390/ijms21082873

10. Prasad A, Fernandis AZ, Rao Y, Ganju RK. 2004. Slit protein-mediated inhibition of CXCR4-induced chemotactic and chemoinvasive signaling pathways in breast cancer cells. Journal of Biological Chemistry.279:9115–24.

11. Schmid BC, Rezniczek GA, Fabjani G, Yoneda T, et. al. 2007. The neuronal guidance cue Slit2 induces targeted migration and may play a role in breast metastasis of breast cancer cells. Breast Cancer Research Treatments. 106:333–42.

12. Naik, S.K., Lam, E.WF., Parija, M., Prakash, S., et al. 2020. NEDDylation negatively regulates ERRβ expression to promote breast cancer tumorigenesis and progression. Cell Death Dis 11, 703. https://doi.org/10.1038/s41419-020-02838-7

13. Ghahhari NM, Babashah S. 2015. Interplay between microRNAs and WNT/β-catenin signalling pathway regulates epithelial-mesenchymal transition in cancer. Eur J Cancer 51:1638–49. doi: 10.1016/j.ejca.2015.04.021

14. Shan S, Lv Q, Zhao Y, Liu C. et. al. 2015. Wnt/β-catenin pathway is required for epithelial to mesenchymal transition in CXCL12 over expressed breast cancer cells. Int J Clin Exp Pathol. Oct 1;8(10):12357-67. PMID: 26722422; PMCID: PMC4680367.

15. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D. et. al. 2010. Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. Am J Pathol.176:2911–20. doi: 10.2353/ajpath.2010.091125

16. Wang C., Zhang R., Wang X., Zheng Y., et. al. 2021. Silencing of KIF3B Suppresses Breast Cancer Progression by Regulating EMT and Wnt/β-Catenin Signaling. Frontiers in Oncology. 10: 3063. doi: 10.3389/fonc.2020.597464