Modular End-to-End TMTpro™ Proteomics Analysis with Channel-Set Normalization and Reference-Based Control of Deuterated Channels

Farhad Shakeri[1]*, Stefan Selzer[1], Andreas Buness[2], Marc Sylvester[3], Sasa Koncarevic[1], David Van Meter[1], Ian Pike[4]

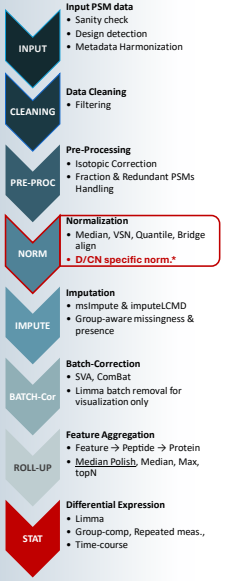
[1]Proteome Sciences R&D GmbH & Co. KG, Frankfurt am Main, Germany; [2]Core Unit for Bioinformatics Data Analysis, University of Bonn, University Hospital Bonn, Bonn, Germany; [3]Core Facility Analytical Proteomics, University of Bonn, University Hospital Bonn, Institute of Biochemistry and Molecular Biology, Bonn, Germany; [4]Proteome Sciences plc, Coveham, Surrey, UK, Cobham, Surrey, United Kingdom

Introduction:

Since Proteome Sciences' launch of TMTpro 35plex reagents at ASMS 2024 via Thermo Fisher Scientific, few open-source workflows explicitly address the systematic technical divergence between **deuterated (D)** and **non-deuterated (CN)** reporter channels. Without specific correction, these chemical disparities can introduce technical bias that interferes with biological interpretation. As the owners of TMT®, we have developed a modular R package which implements a pragmatic solution using subset-specific **Variance Stabilizing Normalization (VSN)** and peptide-level reference anchoring to harmonize D and CN channels. This strategy is integrated into a complete end-to-end pipeline that automates the transition from raw Proteome Discoverer™ PSMs to protein- and peptide-level statistics, ensuring robust downstream analysis for complex experimental designs.

Modular TMT® Workflow:

The analysis follows a strictly ordered trajectory to ensure data integrity and statistical validity.



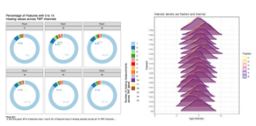
- **TMT-Specific Logic:** Each module is fully flexible, offering multiple integrated methods to accommodate diverse high-plex experimental designs.
- **Hierarchical Scaling:** The workflow supports seamless data transition and summarization across Feature, Peptide, and Protein levels.
- **Algorithm Selection:** Users can systematically compare different statistical algorithms within the same modular framework.
- **Customizable Parameters:** Individual modules allow for fine-tuned parameterization to optimize TMTpro channel-set alignment.
- **Experimental Versatility:** The pipeline natively handles complex study designs, including group-comparisons, time-course series, and repeated measures.
- **Advanced Statistical Modeling:** Supports the generation of sophisticated contrasts, such as interaction effects (double deltas) and group-averaged comparisons.

Multi-Dimensional Data Visualization

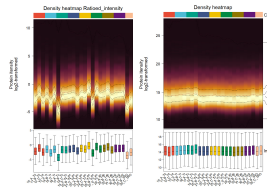
The modular pipeline integrates a diverse array of high-resolution visualizations designed to monitor data integrity and biological signal at every stage of the workflow.

- **Quality Control & Harmonization**: Fraction-level QC plots provide immediate feedback on labeling efficiency and technical consistency across multiplexed runs.
- **Missingness & Normalization**: Advanced schematics validate group-aware imputation strategies and confirm the successful removal of systematic bias.
- **Intensity & Z-score Heatmaps**: Heatmaps visualize Intensity distributions to identify technical outliers, while Z-score heatmaps reveal fine-grained expression patterns and cluster consistency across experimental groups.
- **Exploratory Data Analysis**: Multidimensional scaling via PCA demonstrates robust sample clustering and the effective removal of chemistry-induced artifacts.
- **Statistical & Functional Insight**: High-contrast Volcano plots and functional enrichment maps provide a direct look at differentially expressed proteins and their biological relevance.
- **Feature Overlap & Identification**: Venn diagrams and UpSet plots monitor the consistency of top hits across contrasts.

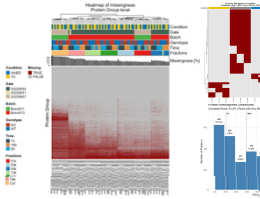
Integrated Data Visualization Showcase



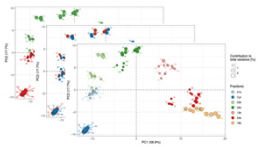
Fraction-Level Quality Control
- **Intensity Distributions:** Monitoring signal stability across fractions.
- **Fraction Overlap:** Visualizing peptide carry-over to validate fractionation efficiency.



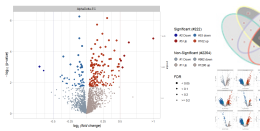
Normalization & Bias Correction:
- Systematic Evaluation: "Before vs. After" visualizations demonstrating the removal of technical variance.



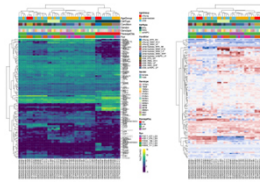
Missingness Analysis:
- **Global & Group-Aware:** Analysis of missing value patterns across the entire dataset AND exclusive per group.
- **Distribution Profiles:** Histograms of missingness used to guide the selection of optimal imputation algorithms.



Exploratory PCA
- **Multivariate Clustering:** PCA plots showing sample relationships and batch-effect mitigation.
- **Covariable Mapping:** Color-coded layouts mapped to metadata variables to confirm relevant clustering,



Statistical Contrast Summary
- **Differential Expression:** Volcano plots per contrast.
- **Hit Consistency:** Venn Diagram and UpSet plots showing the overlap of top hits between different contrasts.



Expression Heatmaps
- **Global Overview:** Intensity heatmaps visualizing absolute abundance and technical outliers.
- **Pattern Recognition:** Z-score heatmaps highlighting fine-grained expression trends.

Functional Enrichment
- **Pathway Mapping:** Biological insights derived from integrated functional analysis.
- **Regulation Profiles:** Comparative enrichment across down-regulated, up-regulated, and total regulated protein sets per contrast.

TMT-35plex: D/CN Channel Normalization

Experimental Design

TMT-35plex: 35 reporter ions across two isotopically distinct series: 18 CN channels and 17 D channels. CN and D reporters overlap at the same nominal masses in pairs, enabling direct within-plex comparison of the two series.

Idea:

To isolate technical bias from biological signal, the same pooled sample was loaded into all 35 channels. With **no biological variability by design**, any observed D–CN intensity difference is **purely a technical artifact** of the labelling chemistry and instrument response.

Note:

Subset to pairs with the same nominal mass: Three CN channels (126, 127C, 127N) and two D channels (127D, 135CD) have no counterpart in the opposite series and are therefore excluded from pairwise bias analysis. The assessment focuses on the 15 matched CN/D pairs yielding a clean, per-mass comparison.

Why bias correction is non-trivial

The D–CN bias is not a uniform offset; it is intensity- and variance-dependent, and its absolute direction varies by instrument setup:

- **Instrument-Dependent Scale:** D channels can read lower or higher than CN channels depending on fragmentation and chromatography.
- **Intensity-Dependent Divergence:** The gap narrows or expands across the dynamic range, often reversing sign at the high end.
- **Non-Linear Trend:** This manifests as a tilted LOESS line in MA plots in *figure 1*, not a flat shift.
- **Failure of Global Shifts:** Methods like median subtraction only removes the average component; the intensity-dependent slope remains.
- **Physical Basis:** The slope reflects distinct variance distributions between deuterated and non-deuterated features that emerge across the concentration range.

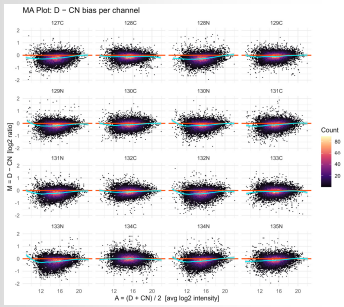


Fig 1: MA plot: raw D–CN bias per channel pair
- LOESS consistently below M = 0 → systematic negative offset in all 15 pairs
- Positive LOESS slope → intensity-dependent bias, not a flat additive shift
- Justifies variance stabilization over simple global shift normalization

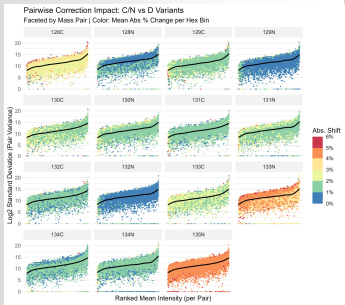


Fig 2: Distribution of Correction Magnitude.
- Correction magnitude varies markedly across CN/D pairs
- N-channel pairs (128N–134N): small, uniform corrections — stable
- C-channel pairs (128C–134C): intensity-specific hot-spots — unstable
- Unstable pairs show larger corrections at low intensity / high variance features

Observation:

Figure 2 reveals:
- **Systematic N vs. C Split:** Data reveals a clear divergence in correction behavior between channel series.
- **Hot-Spot Pattern:** **C-channel pairs** exhibit disproportionate corrections in low-abundance regimes, leaving a **heteroscedastic residual.**
- **Limitation**: Global shift normalization **cannot address the non-uniform noise** left after isotopic deconvolution.
- **Stabilization**: VSN explicitly models the mean–variance relationship, making it the **essential follow-up to impurity correction**.

Normalization Strategy by Experimental Design *(figure 3)*

Case 1 — Randomized design + bridge (pooled QC) samples (recommended)
- VSN fitted on CN channels → generalized-logarithmic (glog) scale anchored to the CN intensity distribution, preserving absolute intensity scale
- CN-derived parameters applied as a fixed reference to D channels → D channels transformed onto the same glog scale as CN
- Bridge/pooled QC channels present in both series are used for plex-wise alignment in multiplexed experiments

Case 2 — Randomized design, no bridge/pool samples
- Same vsn_ref approach: CN fit applied as reference for D
- Inter-series alignment relies on the CN grand median as anchor, valid under randomization because group differences average out

Case 3 — Blocked / non-randomized design (e.g. all cases in D, all controls in CN)
- Independent VSN per series with calib = "none" + explicit median alignment
- Only recommended when design forces it; make the scale loss explicit to downstream analysts

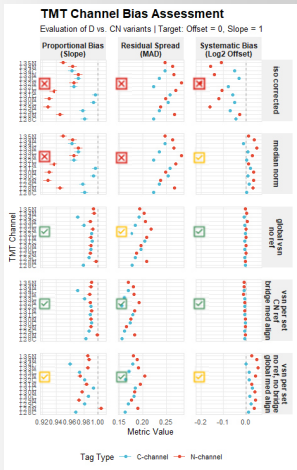


Fig 3, Forest plot: method comparison across normalization strategies
Per-channel D/CN bias across five normalization strategies. Each row is a method; columns show systematic offset (target: 0), proportional bias/slope (target: 1), and residual spread (MAD). N- and C-channel pairs colored separately. Iso correction reduces offset; VSN with CN reference achieves the best joint correction of offset and slope while preserving the intensity scale.

Conclusions

The D/CN bias in TMT-35plex is systematic and intensity-dependent, meaning a simple shift correction is insufficient. As demonstrated by the joint correction of offset and slope in Figure 3 (row 4), **stratifying the variance stabilization by label and anchoring the D channels to the CN reference effectively neutralizes both additive and proportional bias** while preserving the absolute intensity scale. The choice of alignment strategy is determined by experimental design: plex-wise via bridge samples when available, grand-median otherwise. Blocked designs require independent per-series normalization to avoid confounding biology with technical correction.