MANTI: Automated annotation of protein N-termini for rapid interpretation of N-terminome datasets

Fatih Demir¹²*, Jayachandran N. Kizhakkedathu³, Markus M. Rinschen¹⁴, Pitter F. Huesgen²⁵*

¹Aarhus University, Department of Biomedicine, Høegh-Guldbergsgade 10, 8000 Aarhus C, Denmark
²Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.
³Centre for Blood Research, Department of Pathology & Laboratory Medicine, School of Biomedical Engineering, Department of Chemistry, University of British Columbia, 251 - 2222 Health Sciences Mall, Vancouver V6T 1Z3, British Columbia, Canada.
⁴III. Department of Medicine, University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20251 Hamburg, Germany.
⁵Cologne Excellence Cluster Cellular Stress Response in Aging-Associated Diseases (CECAD), Medical Faculty and University Hospital, Institute of Biochemistry, Department of Chemistry, University of Cologne, Joseph-Stelzmann-Str. 26, 50931 Cologne, Germany

*Corresponding Authors
Fatih Demir: fatih.demir@biomed.au.dk, Tel.: +45 2066 5308.
Pitter F. Huesgen: p.huesgen@fz-juelich.de, Tel.: +49 246 61 6873

SUPPORTING INFORMATION

Figure S1. N-terminal peptide purification workflow
Figure S2. Screenshot of the Yoğurtlu MANTI interface to MANTI
Figure S3. Venn diagrams with overlap of canonical and noncanonical cleavages-containing proteins
Figure S4. Heatmaps depicting the frequency of the first amino acids for PAN N-termini
Figure S5. Heatmaps depicting the frequency of the first amino acids for WT1 N-termini
Figure S6. Correlation among the PAN and WT1 kidney disease models’ N-termini

Table S1. Main output directories and files generated by MANTI and their contents
Tables S2-8 Supplementary data tables containing full MANTI output file for PAN and WT1 datasets as well as the merged, combined N-terminal datasets and Enrichr GO BP term enrichment output are listed in a separate spreadsheet.
Figure S1. N-terminal peptide enrichment by negative selection. Protein N-termini and lysine side chain amines are endogenously modified or chemically labelled with an amine-reactive reagent before proteome digestion into small peptides. The proteome digest, usually performed with trypsin, generates new primary amines at all released peptides except the N-terminal peptides. As a side reaction, spontaneous pyro-Glu formation can occur. The reactive primary amine of digest-generated peptides are then captured by reaction with a water soluble, aldehyde-functionalized high molecular weight polymer (HPG-ALD). Subsequently, the polymer with covalently bound peptides is removed by filtration, whereas the inert, previously labelled or endogenously modified protein N-terminal peptides remain in the filtrate and are thereby selectively enriched and separated from the proteome digestion-generated peptides. Enriched N-terminal peptides are further desalted and analyzed by LC-MS/MS.
Figure S2. Screenshot of the Yoğurtlu MANTI graphical user interface. The main window of the graphical user interface (GUI) to MANTI, written in Perl/Tk, is shown. The interface is available as a Perl source file as well as an executable program for the Windows 10 64-bit platform which does not require any installed Perl on the user’s computer and comes bundled within each MANTI release. The graphical interface assures the definition of all the mandatory fields and enables a smooth user experience for the MANTI workflow. Many options can be adjusted in a user-friendly manner within this GUI, e.g. custom labels for the reporter ion channels.
Figure S3. Venn diagrams indicating the number of proteins with identified canonical and noncanonical N-terminal peptides, indicating previously annotated and proteolytic proteoforms, respectively. a) Overlap of protein IDs with canonical and noncanonical N-termini for PAN-treated rats, and b) for heterozygous WT1-deficient mice.
Figure S4. Heatmaps depicting the frequency of the first amino acids for PAN N-termini depending on modification and subcellular localization. First amino acids of the validated N-termini were categorized into canonical/noncanonical termini, separated by modification and localization as annotated by UniProt. a, b) Acetylated N-termini are mostly canonical cleavages with a strong bias on methionine and alanine as the first amino acids. c, d) Dimethylated N-termini display a generally broader distribution with a bias on alanine and serine as the first amino acids for noncanonical cleavages.
Figure S5. Heatmaps depicting the frequency of the first amino acids for WT1 N-termini. First amino acids of the validated N-termini were categorized into canonical/noncanonical termini, separated by modification and localization as annotated by UniProt for (a, b) acetylated and (c, d) dimethylated N-termini.
Figure S6. Correlation among the PAN and WT1 kidney disease models’ N-termini. a) Comparison of changes in the abundance of N-terminal peptides mapping to corresponding positions in homologous rat and mice proteins in the PAN and mouse WT1 datasets. b) Protein N-terminal peptides depleted in both disease models. c) N-terminal peptides specifically accumulating in rat glomeruli after PAN-treatment.
| File / directory       | Content                                                                                                                                                                                                 |
|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MANTI Masterfile      | Validated and aggregated information about N-termini including external data (limma, LOCALIZER, TargetP, TopFINDer, UniProt)                                                                                 |
| Unique_Ntermini       | Unique N-termini: N-terminal peptides grouped by position and modification for each protein. Collapses identified N-terminal peptides for multi-enzyme digestion strategy experiments or missed cleavages |
| Summary Masterfile    | MANTI parameters, files, experiments and the validated N-terminal peptides (modifications, peptide length, intensity distribution)                                                                          |
| validation.txt        | Validation strategy: why are peptides/proteins eliminated/not quantified                                                                                                                                  |
| 1st-aa-categories      | First amino acid occurrences in the peptide sequence for different categories/localizations                                                                                                              |
| 2d-aa-maps            | N-termini two-dimensional amino acid maps                                                                                                                                                                 |
|                       | categories                                                    | Separated by (non)canonical classification                                                                                                                                                                  |
|                       | localizations                                                | Separated byUniProt localization annotation                                                                                                                                                                 |
|                       | modifications                                                | Separated by modification                                                                                                                                                                                  |
| cleavages             | Cleavage positions/(non)canonical classifications                                                                                                                                                    |
| distributions         | Validated protein N-termini distributions                                                                                                                                                                                                                   |
|                       | amino-acids                                                 | Overall and specific amino acid compositions for first, second amino acids and first dipeptides                                                                                                             |
|                       | experiments                                                 | Overlap of N-terminal peptides (sequence + modification) among experiments                                                                                                                               |
|                       | fractions                                                   | Overlap of N-terminal peptides among fractions                                                                                                                                                           |
|                       | localizations                                               | Localization-specific amino acid distributions for first, second amino acids and first dipeptides                                                                                                         |
| executive-plots       | A selection of figures for the MANTI analysis including a one-page PDF summary of the N-terminal dataset                                                                                              |
| external-sources      | Data integrated from external sources used for processing                                                                                                                                                  |
| MaxQuant-txt          | Original MaxQuant txt folder files used                                                                                                                                                                                                                      |
| LOCALIZER / TargetP   | LOCALIZER / TargetP predictions                                                                                                                                                    |
| ProteinCorrelate      | Proteome-level information (e.g. shotgun proteome)                                                                                                                                                    |
| TopFINDer             | TopFINDer N-termini annotation                                                                                                                                                    |
| UniProt               | Downloaded annotation data from UniProt database                                                                                                                                                    |
| bufipretty            | Data assessment graphs (R script)                                                                                                                                                    |
| limma                 | Limma statistical results                                                                                                                                                    |
| polvops               | Prepared files for limma-based statistics (R script)                                                                                                                                                    |
| plots                 | All output graphs from bufipretty.R + polvops.R                                                                                                                                                    |
| Correlations          | Correlations of MaxQuant ratios among defined experiments and proteome-level to N-terminome ratio mapping                                                                                              |
| Distributions         | Distributions of MaxQuant ratios by the N-terminus position and histograms of N-terminus positions                                                                               |
| First_AA_Cats         | First N-terminal amino acids, plotted by phys.-chem. Properties, separated by the modifications                                                                                                          |
| First_AA_Locs         | First N-terminal amino acids, separated by modifications and the localizations in relative and absolute numbers; for absolute numbers a Fisher’s exact test whether one amino acid is overrepresented significantly in canonical vs.
| First_AA_Mods | First N-terminal amino acids are plotted for the different modifications in relative and absolute numbers – for the absolute numbers with a Fisher’s exact test whether one amino acid is overrepresented significantly in canonical vs. noncanonical N-termini. |
| General | The main set of plots for the MANTI processing including modification distribution, N-termini by localization, the overlaps among defined experiments and fractions as well as TargetP distances to N-termini, the positional annotation of N-termini and an UpSet plot showing the global overlap of the experiments. |
| Heatmaps | Heatmaps of amino acids for N-termini, especially the first amino acids separated by category, localization and modifications. |
| PrecedingAA_Distance | Plots for the distance of the N-terminus start to the corresponding preceding amino acid, combined with the mean log2FC. Additionally, a zoomed version to 20 amino acids preceding the N-terminus is supplied. |
| Quantifications | Boxplots and raincloud plots of quantifications |
| Volcanos | Volcano plots generated by polvops.R, separated by ratios and in multiple versions with GlobIDs and Genes |
| ratios-ids | Lists of protein IDs, genes and ratios |
| specially-encoded | Amino acid compositions for proteins encoded by the chloroplastidial and/or mitochondrial genome |

Table S1. Main output directories and files generated by MANTI and their contents