nature research

Corresponding author(s): Hao Chi, Si-Min He

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen based on preliminary experiments so as to provide sufficient power for statistical comparison (where appropriate).

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

The experimental findings in this manuscript are reliably reproduced. The sensitivity, precision and speed of Open-pFind are measured using six different datasets.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

This is not relevant to our study. For Open-pFind and other search engines, ALL MS/MS data extracted from RAW files are searched against protein databases.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

It is not relevant to our study because this study is mainly focused on the methods of MS/MS data analysis.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Nature Biotechnology: doi:10.1038/nbt.4236

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

1. Eight proteome search engines were used in this study.
1) Open-pFind (also implemented in pFind 3.1.2 as the default open search mode)
2) PEAKS 7.5
3) MODa 1.23
4) MSFragger v20170103
5) pFind 3.1 (restricted mode)
6) Comet 2016012
7) MS-GF+ v10072
8) Byonic 2.10

2. A semi-supervised learning algorithm is used to iteratively separate target PSMs from decoy ones based on the linear classification software package LIBLINEAR.

3. pXtract v2.0 (http://pfind.ict.ac.cn/software/pXtract/index.html) is used for converting RAW files to MS1 and MS2 files.

4. pQuant v2.0 (http://pfind.ict.ac.cn/software/pQuant/index.html) is used to check the quantitation ratio of each PSM identified.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

Not restrictions on availability apply.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell line were used. The datasets of Mann-Human-Velos, Gygi-Human-QE, Mann-Mouse-QEHF, and Pandey-Human-Elite were published previsouly. The detailed information is shown in Supplementary Table 1.

b. Describe the method of cell line authentication used.

No eukaryotic cell line were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell line were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No cell line listed by ICLAC was used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.

Nature Biotechnology: doi:10.1038/nbt.4236