Supporting Information:
Matrix-matched calibration curves for assessing analytical figures of merit in quantitative proteomics

Lindsay K Pino,[1] Brian C Searle,[1] Han-Yin Yang,[1] Andrew N Hoofnagle,[2] William S Noble,[1,3] Michael J MacCoss[1]*

[1] Department of Genome Sciences, University of Washington
[2] Department of Laboratory Medicine, University of Washington
[3] Department of Computer Science and Engineering, University of Washington
* To whom correspondence should be addressed; Email: maccoss@uw.edu

TABLE OF CONTENTS

Supplementary Figures

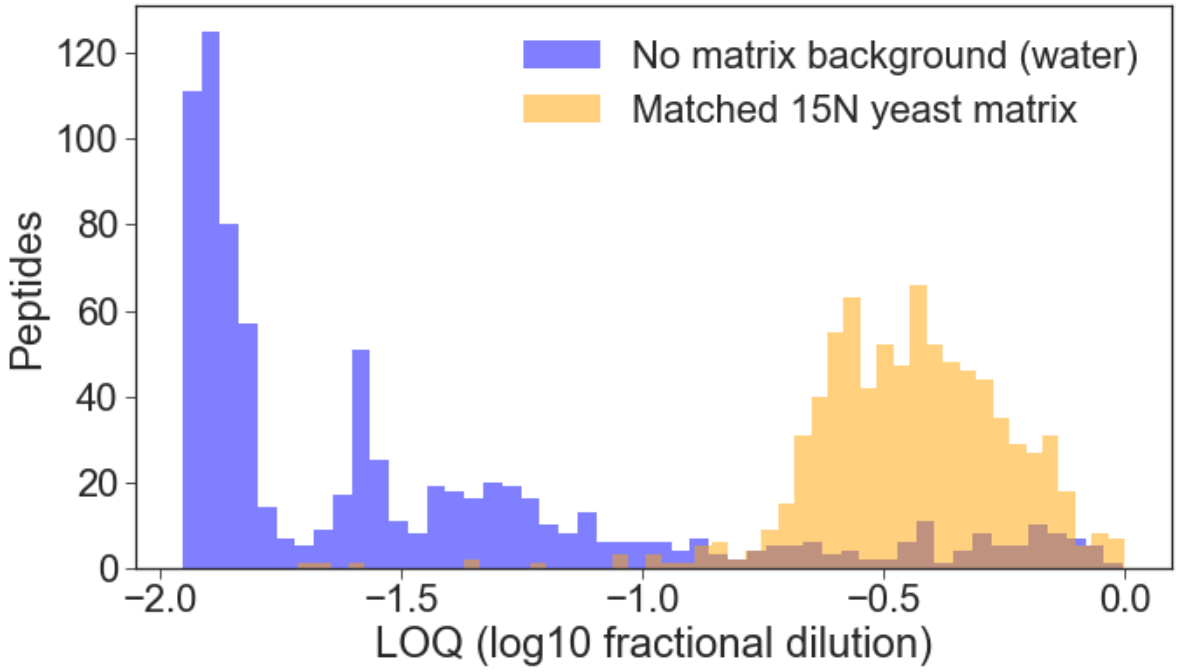

Fig S1. Reference materials must be diluted with a similarly complex material to preserve matrix properties. The yeast reference material was diluted in a matched matrix with 15N-shifted yeast (orange) or diluted in water with no matrix replacement (blue). Our curve-fitting model was then run on each data set to calculate the LOQs. The reported peptide LOQs are significantly more sensitive when the matrix complexity is not replaced, showing the importance of retaining matrix properties when building the calibration curve.

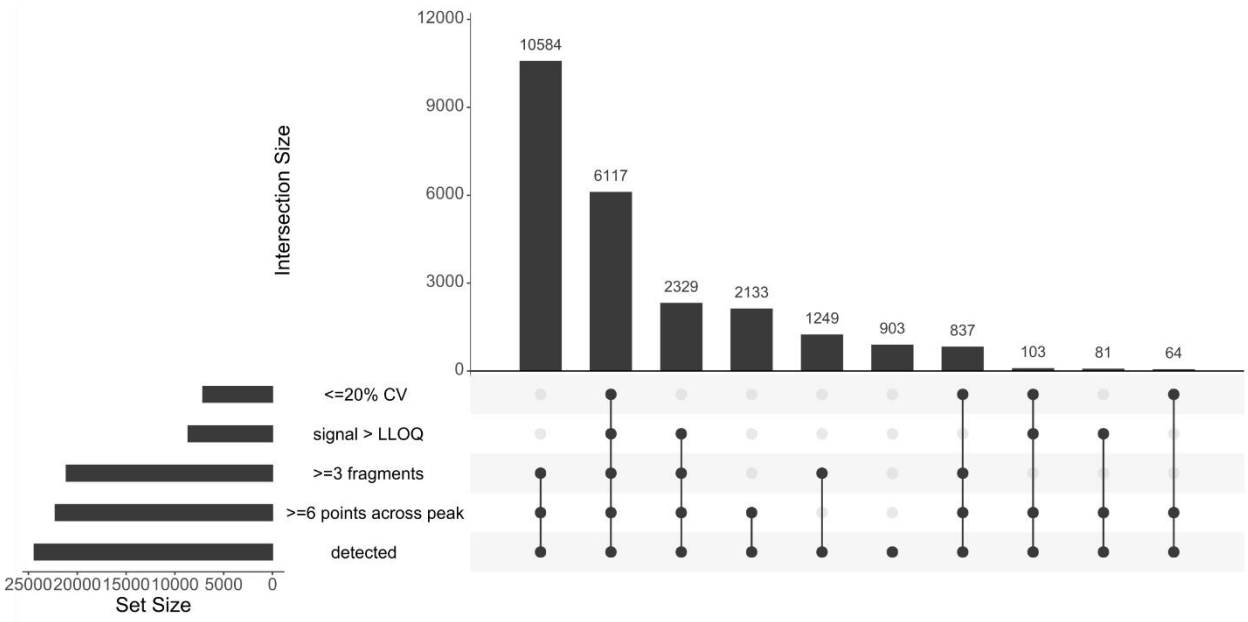

Fig S2. Curation of DIA data using common targeted proteomics criteria can filter for the highest quality peptides. Several common quantitative metrics for the yeast DIA data are used to filter the detected peptides: minimum of 3 interference-free fragments (as determined by the automated interference detection algorithm internal to EncyclopeDIA), at least 6 points across the MS2 chromatogram peak (as determined by Skyline), LOQ as determined by the proposed method here, and less than 20% coefficient of variation of the peak area (as calculated from TIC-normalized triplicate measurements of the undiluted yeast proteome over the entire acquisition period of the curve dataset). Nearly half of the peptides detected in yeast (10,584 peptides) have at least 6 points across the peak and at least 3 fragment ions, but do not have an LOQ as determined by our proposed approach nor do they meet a maximum of 20% peak area CV.

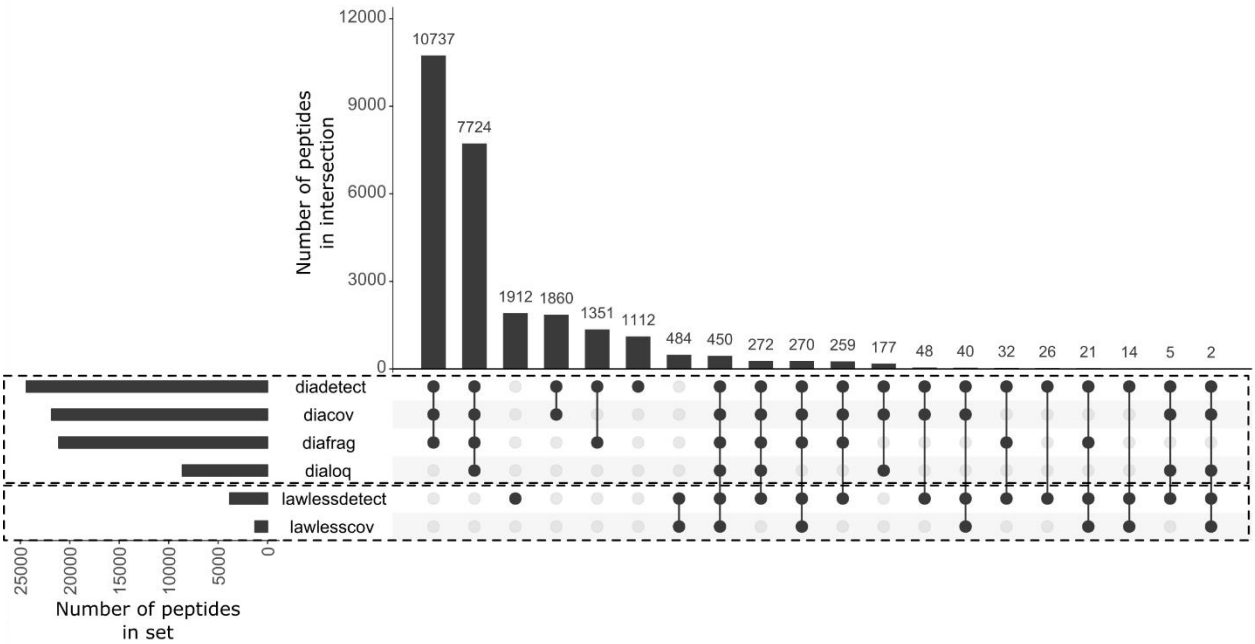

Fig S3. Matched matrix calibration curves can assess more candidate targets than conventional approaches without predetermining targets. The UpSet plot compares yeast peptides detected and deemed quantitative by DIA-MS and matched matrix calibration curves (this work, legend "dia*") with the peptides detected and validated by QConCat SRM-MS in Lawless *et al* 2016 (*12*). Of all peptides detected in the wide-window DIA-MS at 1% FDR ("diadetect", 24,400 peptides), only 6,117 peptides (25%) display all three desirable quantitative traits ("diacov", peptides with <20% CV in the undiluted yeast sample; "diafrag", peptides with >3 interference-free fragment ions; "dialoq", peptides with a defined LOQ). Lawless *et al* 2016 assessed over 4,000 total peptides ("lawlessdetect", QConCat peptides tested by Lawless *et al* 2016), of which 1,281 peptides (50%) displayed the desired quantitative properties ("lawlesscov", QConCat peptides tested by Lawless *et al* 2016 with <20% CV). Overall, the proposed framework assessed 6x more candidate peptides and defined 4.7x more peptides as quantitative. The quantitative peptides in the proposed approach map to 1,629 proteins; the quantitative peptides by the QConCat approach map to 644 proteins. Both approaches include quantitative peptides for 520 proteins, and the QConCat approach includes peptides for 124 proteins not represented by the proposed approach, while the proposed approach includes 1,109 proteins not represented by the QConCat approach.

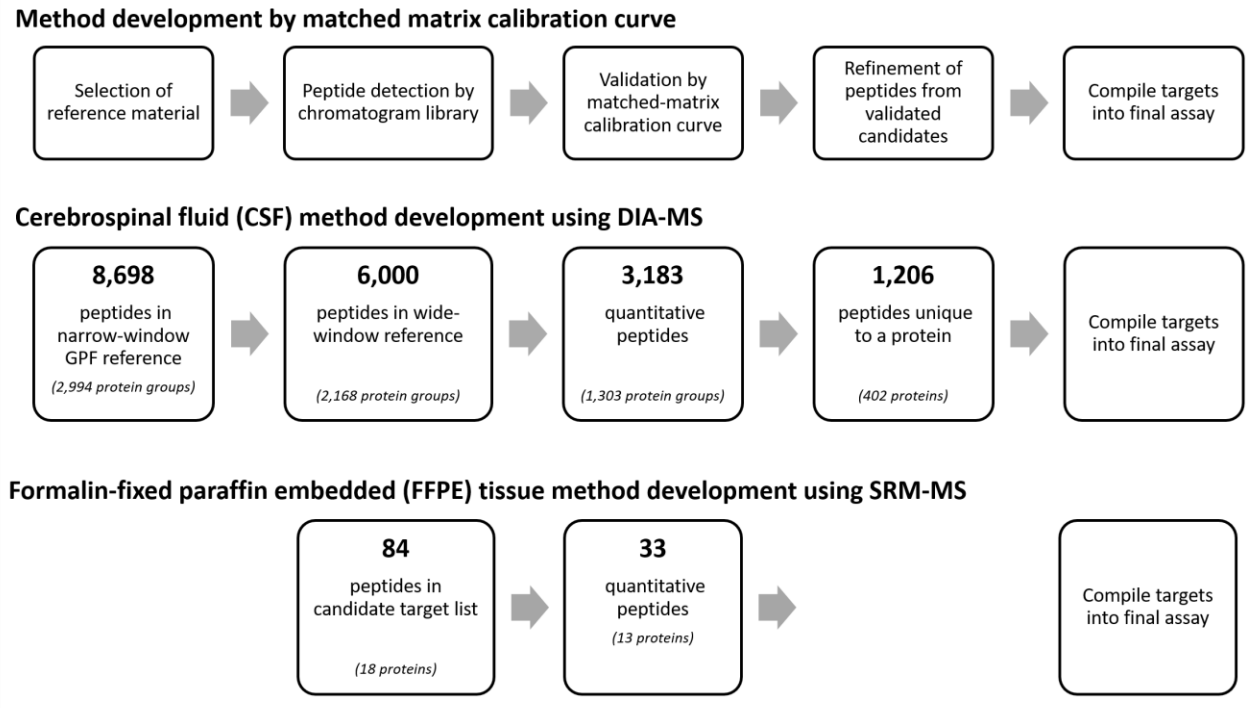

Fig S4. Matched matrix calibration curves can be used to rapidly develop targeted methods. Starting with all peptides detected in a gas-phase fractionated reference material as possible candidates, the first refinement step in this protocol discards any candidate target that cannot be detected in an unfractionated, single-shot acquisition of the reference material. Next, the matched matrix calibration curve framework is used to assess quantitative figures of merit, discarding any candidates whose abundance in the reference material is below the analyte LOQ. For most targeted quantitative proteomics work, targets that are unique to a protein are considered better candidates. Even if experimenters are not starting from DIA data, the matched matrix calibration curve approach can be used to quickly eliminate poor quantitative targets from an assay.

Supplementary Tables

Point	Yeast reference (fractional dilution)	Yeast matched matrix (fractional dilution)
A	1	0
B	0.7	0.3
C	0.5	0.5
D	0.3	0.7
E	0.1	0.9
F	0.07	0.93
G	0.05	0.95
H	0.03	0.97
I	0.01	0.99
J	0.007	0.993
K	0.005	0.995
L	0.003	0.997
M	0.001	0.999
N	0	1

Table S1. Dilution series for the yeast matched matrix calibration curves. Using a fractional dilution scheme, the reference material is diluted with the matched matrix to create a dilution series. In this 14-point design, the calibration standards span three orders of magnitude of the reference material and include a blank with only the matched matrix.

Point	Plasma reference (fractional dilution)	PBS (fractional dilution)
A	1	0
B	0.5	0.5
C	0.25	0.75
E	0.1	0.9
F	0.05	0.95
G	0.025	0.975
I	0.01	0.99
J	0.005	0.995
K	0.0025	0.9975
L	0.00125	0.99875
M	0.001	0.999
O	0	1

Table S2. Dilution series for the FFPE tissue block matched matrix calibration curves. Using a fractional dilution scheme, healthy donor plasma (reference) is diluted with PBS to create a dilution series. An equal volume of each calibration point was then mixed with a homogenate of chicken liver and prepared as individual FFPE tissue blocks.

Frequently Asked Questions

1. Are most of the peptides observed in DIA/SWATH experiments really not useful for quantification?

A major thing to consider here is what is being used for the definition of "quantitative." We think this is an important point to consider before we can say whether the number of peptides is "radically small" or not. We admit that we are using a comparatively conservative and strict definition in this manuscript but we don't believe we should be criticised for that. Most proteomics laboratories, at best, have used a predetermined CV threshold to assess whether a peptide is quantitative. To do this a single preparation of a sample is used to make repeated measurements to compute the CV. However, this does not reflect a quantitative measurement. A real quantitative measurement must 1) have a change in signal between ***different*** samples that actually reflects the expected change in the quantity and 2) must have a sufficient precision to distinguish between abundances of ***different*** samples. This distinction is important because there can be a precise measurement that does not reflect the quantity ***AND*** the sample preparation variance will likely be larger than the analytical measurement variance.

Most detectable peptide can probably be used to demonstrate a difference in abundance. However, not every peptide's abundance will accurately reflect the change in quantity. In fact, we state this specifically in the introduction by saying, "Although a signal measured below the LLOQ may still be used to assess a difference between two conditions, when compared to a signal above the LLOQ, the magnitude of the difference in signal is not reflective of the true difference in analyte quantity."

We believe that all methods that are classified as "quantitative" should demonstrate the limit of quantification. Defining the Limit of Quantification has been a criteria for publication of quantitative methods for a long time but has been largely ignored in the field of proteomics. The strategy reported here can be used for MS1 quantification, fully targeted methods based on SRM, and even spectrum counting methods. We believe that for a peptide to be quantitative, that the signal must decrease when the peptide is diluted and the measurement across different preparations must be precise.

We are making a case in our manuscript that for a peptide to be quantitative, the signal must decrease by the expected amount when diluted in the matched matrix and there must be a point on the dilution curve where the CV, from different samples, is <20%. This experiment is analogous to a "recovery" or a "standard addition" experiment in a classical quantitative method validation experiment. In those cases a standard is added to the sample matrix and the method must demonstrate that the signal returned accurately reflects the change in quantity. However, because we don't have a standard to add to the matrix, we instead dilute the sample in a matched matrix and assess the opposite.

2. How much does the DIA/SWATH software you use influence whether a peptide is determined "quantitative" or not by these calibration curves?

The calibration curve method here does not point out the source of what is limiting the limit of quantification. The matrix matched dilution curve only reports the limit of quantification of the

overall method. The overall method reflects sample preparation, quantitative character of the peptides, data collection artifacts, software and human data analysis issues, etc… This is exactly what we are trying to capture. It is also why the use of an analytical measurement CV is not appropriate.

3. For background matrix, when should I use isotope labeling (18O, 15N, SILAC, etc) and when should I use a matrix of a different species (chicken liver)?

The choice of background matrix really comes down to two considerations: practicality and convenience. In our work, 15N was used for cell culture, 18O digest was used for human biofluids (which could not be cultured in heavy media to incorporate the isotope), and the chicken liver used for FFPE tissue block (background must be resilient to the paraffin-embedding process, e.g. 18O could not be used due to undesirable 16O back-exchange during the FFPE tissue blocking process). As far as advantages of one stable isotope labeled background matrix over another, certainly the background matrices using 15N (here shown with yeast) and SILAC (which we did not demonstrate in this work) have the big advantage of being more stable than 18O incorporation. Because a double-digest with 18O will back-exchange with 16O, curves made with 18O like the CSF demonstration in our work must be acquired as soon as possible following the 18O digest. Because the 15N and SILAC matrices are more stably incorporated, they are much easier to work with, when possible!

4. If I analyzed the same calibration curve data with different DIA software, would I get different LLOQ values for the same samples?

The specific LLOQ results we present here only assess the LLOQ for these particular samples, on this particular instrumentation (e.g. our QE-HF with our specific DIA method parameters), and with this signal processing process (e.g. Encyclopedia). It is our hope that the matrix-matched calibration curve approach proposed here will be used by the community to address these types of questions -- which methods are most sensitive? Which DIA software (e.g. fragment refinement) reports the best quantitative results? We hope that these questions can now be examined through the use of our proposed method.