

Concerns re: “An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process”

Claire D. McWhite* and Daniel R. Boutz

The University of Texas at Austin, Austin TX, USA

*Correspondence: claire.mcwhite@utexas.edu (mailto:claire.mcwhite@utexas.edu)

Dear Editorial Board of Scientific Reports,

We are writing to you with concerns about the proteomic data in the recently published paper “An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process”. This paper claimed major protein expression differences between non-genetically modified maize and Roundup-Ready maize both treated and untreated with Roundup. However, we found major errors with the analysis of proteomics data, mainly that fold changes of individual peptides are incorrectly represented as fold changes for full proteins. This underlying analysis error affects large parts of the analysis and conclusions.

In December of 2016, I (CDM) read the newly published paper and saw several discrepancies in the proteomics analysis presented in the file containing protein fold differences (Supplementary Table 5). Briefly:

- Some proteins have both positive and negative fold changes in the same comparison.
- The values in the column 'Mass (Da)' are too small to correspond to protein masses.
- A top fold enriched protein is from a corn smut fungus, not maize.

I received a file of peptide intensities from the authors (Additional Data File 1). Along with my colleague (DRB), I used this file to confirm that if any single peptide from a protein is enriched or depleted, the entire protein is counted as differently expressed across conditions. This is a large departure from all standards for protein quantification, as proteins are made up of multiple peptides.

Background on Peptide vs. Protein quantification.

Mass spectrometers collect spectra of component tryptic peptides of proteins. Protein fold changes are determined from integrated measurements of their component peptides. The enrichment of an individual peptide from a protein does not necessarily mean that the full protein is statistically enriched, as proteins are composed of multiple peptides. In Mesnage *et al*, if a single peptide is found to be enriched, the entire protein is counted as enriched. Likewise, if a single peptide is found to be depleted, the entire protein is counted as depleted.

Protein:

MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLIRAEELGQPGNLLGNDHIYNVIVTAHAFVMIFFMVMPIMIGGFGNW
LVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLASAMKVEAGAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGA
INFITTIINMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTD

Peptides:

MFADR

WLFSTNHK

DIGTLYLLFGAWAGVLGTALSLIR

AELGQPGNLLGNDHIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPR

MNNMSFWLLPPSLLLLASAMK

VEAGAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMK

PPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTD

Introduction to the conventional TMT fold change calculation

In isobaric labeling experiments such as TMT10plex used in this paper, proteins from each condition are cleaved into peptides with trypsin, their N-termini labeled with isobaric tags, mixed together, and identified by mass spectrometry. Upon fragmentation of differentially labeled peptides, each TMT variant will generate a unique reporter ion. The relative reporter intensities can be compared across conditions to find fold changes.

Enrichment of a protein is determined from the fold changes of its component peptides across conditions.

The first proof of concept paper on isobaric mass proteomics² describes the process of finding protein expression levels from peptide data. Particularly, that 1) proteins that are identified from only one peptide are discarded and 2) proteins with high standard deviation between peptide scores are discarded. A handbook for protein quantification³ also recommends discarding proteins identified by a single peptide and warns against counting a protein as enriched due to outlier peptides.

Though minor deviations in methods exist, standard isobaric labeling protein quantification methods involve integrating measurements over all distinct peptides per protein to get full protein fold changes²⁻⁵.

Claims of measuring protein expression differences

The text of Mesnage *et al* clearly implies that proteins are being quantified, as would be normal for a proteomics experiment measuring differences across conditions.

- “Changes in proteins and metabolites...”
- “we have performed proteomics and metabolomics analyses of NK603 (sprayed or unsprayed with Roundup) and isogenic maize kernels (Fig. 1). We used a TMT10plex™ isobaric mass tag labelling method and quantified proteins by Liquid chromatography-tandem mass spectrometry (LC-MS/MS)”
- “The projection of individual protein or metabolites on a 2-dimensional space”
- “Overall, the MCIA shows that the GM transformation process was the major contributor to variation in the protein and metabolite profiles...”
- “The list of proteins and metabolites having their levels significantly disturbed is given in Additional files 5 and 6, respectively.”
- “While only one protein is newly produced as a result of the transgene insertion, a total of 117 proteins and 91 metabolites have been altered in maize by the genetic transformation process”
- “One protein (B4G0K5) and 31 metabolites had their expression significantly altered...”
- “Among them, pyruvate kinase (B4F9G8), enolase (ENO1), and three glyceraldehyde-3-phosphate dehydrogenases (GAPC1, GAPC2, GAPC3) had their levels increased in NK603 maize.”
- “Additionally, while proteins associated with glycolysis were overexpressed...”

Presented data are for peptide fold changes, not protein fold changes.

Supplementary Figure 5 is described as a “List of proteins having their level significantly altered by the GM transformation process”. The fold changes from this file for the isogenic vs. Roundup-Ready strain nk603+Roundup comparison are plotted in Figure 1.

We found that this file actually describes fold change of individual peptides (Figure 1). If any one peptide from a protein falls above the cutoff, the entire protein is counted as enriched.

Multiple proteins described as enriched/depleted between samples have individual peptides with positive and negative fold changes. For example, the protein Q7M1Z8 has 4 entries in Supplementary Table 5, reproduced exactly in Table 1 below. Conventionally, each protein would have only one fold change measurement, as a single protein can only be enriched, unchanged, or depleted. Though only single peptides that show a log2 fold change above cutoffs (-0.5, +0.5) were given in Supplementary File 5, the raw peptide data show that multiple peptides from the presented proteins fall below this cutoff (Figure 2).

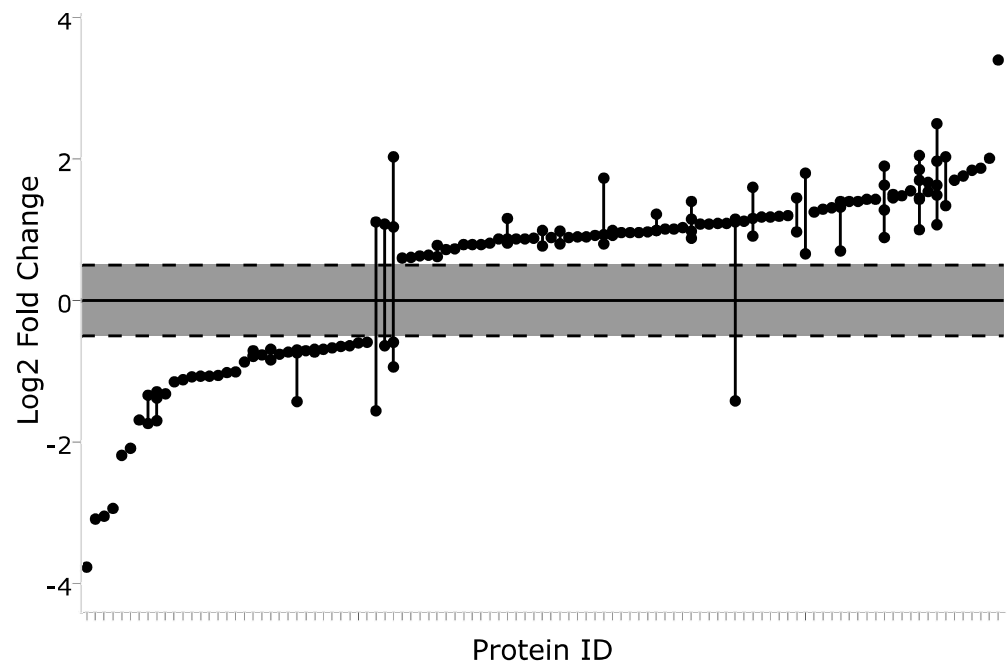


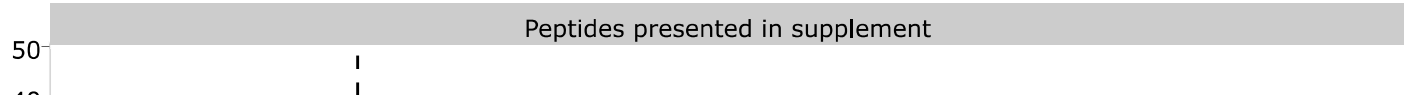
Figure 1. Each fold change given between isogenic and Roundup-treated Roundup ready maize is the fold change of an individual peptide. Vertical lines connect peptides from the same protein, showing many cases with conflicting fold changes. Most proteins plotted also have additional peptides that would fall in the gray area and are not differentially expressed, as shown in Figure 2

Table 1. One example from Supplementary Table 5 of the same protein showing both positive and negative fold change

Uniprot ID	Protein name	Mass (Da)	Log2 FC	P-adjusted values
Q7M1Z8	OS=Zea mays GN=Zm.3896	395.48	2.0333	0.0307
Q7M1Z8	OS=Zea mays GN=Zm.3896	661.31	1.0405	0.0423
Q7M1Z8	OS=Zea mays GN=Zm.3896	722.35	-0.5861	0.0247
Q7M1Z8	OS=Zea mays GN=Zm.3896	823.46	-0.9408	0.0001

Analysis of other peptides from proteins described as perturbed.

We wondered if there were other quantifiable peptides from these proteins that fall below the threshold. We found that 24 of the 105 proteins described as perturbed only have evidence from a single peptide in the raw peptide data. These proteins should be discarded. Most other proteins have multiple other peptides whose fold changes must have fallen below the thresholds. As an extreme case, there are 4 peptide fold changes from P15590 shown in the supplement, while there are measurements for 48 peptides in the raw peptide data. The presence of many other peptides below threshold strongly suggest that many of these proteins would not show enrichment if all their peptide measurements were integrated, and listed enriched peptides are likely to be false positives.



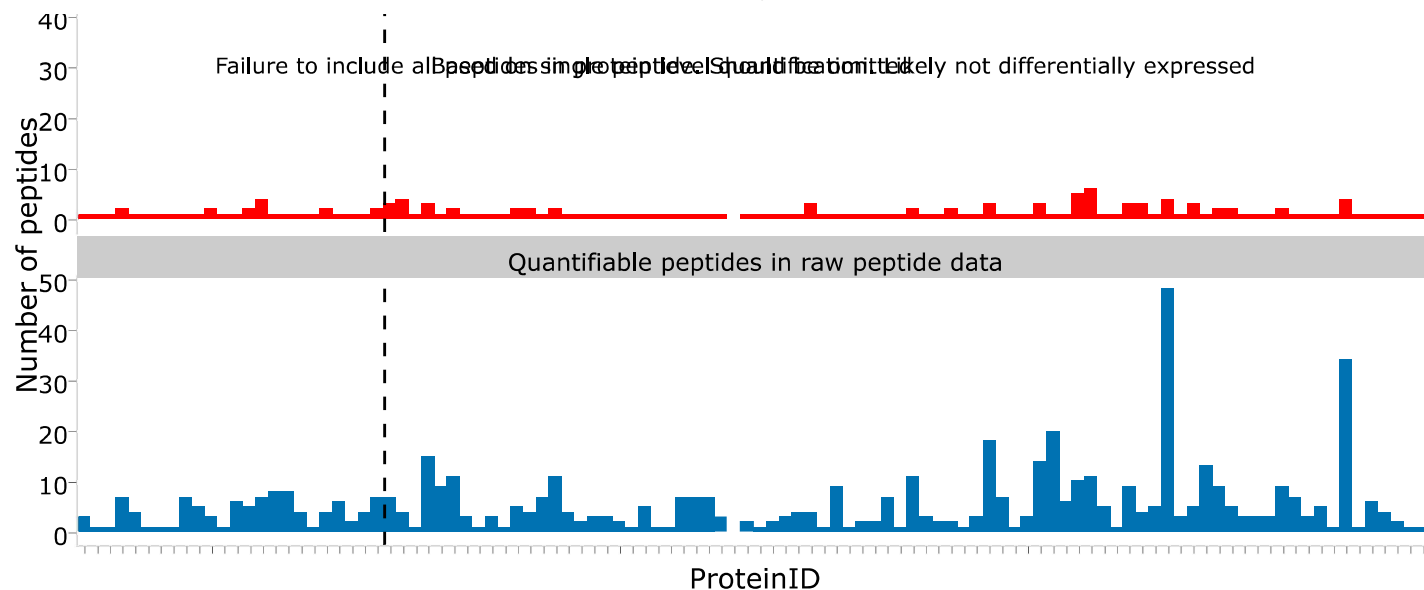


Figure 2. Many more quantifiable peptides for proteins flagged as differentially expressed are present in the raw data than in the presented data, suggesting that many peptides for these proteins are not differentially expressed. Thus, proteins largely seem to be flagged as differentially expressed based on outlier peptides or single peptide observations

The use of individual peptide fold changes instead of protein fold changes to obtain lists of proteins significant affects the conclusions in this paper. To highlight one example, the abstract states that “Changes in proteins and metabolites of glutathione metabolism were indicative of increased oxidative stress.”. The text describe three proteins as altered to support this oxidative stress conclusion: “The comparison between Roundup-sprayed NK603 and control samples revealed a similar pattern to that observed in unsprayed samples. However, glutathione metabolism (KEGG ID 480) showed a significant alteration in sprayed NK603. The proteins assigned to that pathway, glutathione S-transferase 1 and 6-phosphogluconate dehydrogenase (P12653 and B4FSV6 respectively) were more abundant in sprayed samples while another glutathione transferase isoform GST-5 (A0A0B4J3E6) was less abundant.”

Displaying the peptide evidence for these three proteins (Figure 3), we can see that conclusions about A0A0B4J3E6 and P12653 are based on single peptides. B4FSV6 has three other peptides with fold enrichments below threshold, suggesting that the full protein would not show changes across conditions. Differential expression of these proteins is not supported by the data.

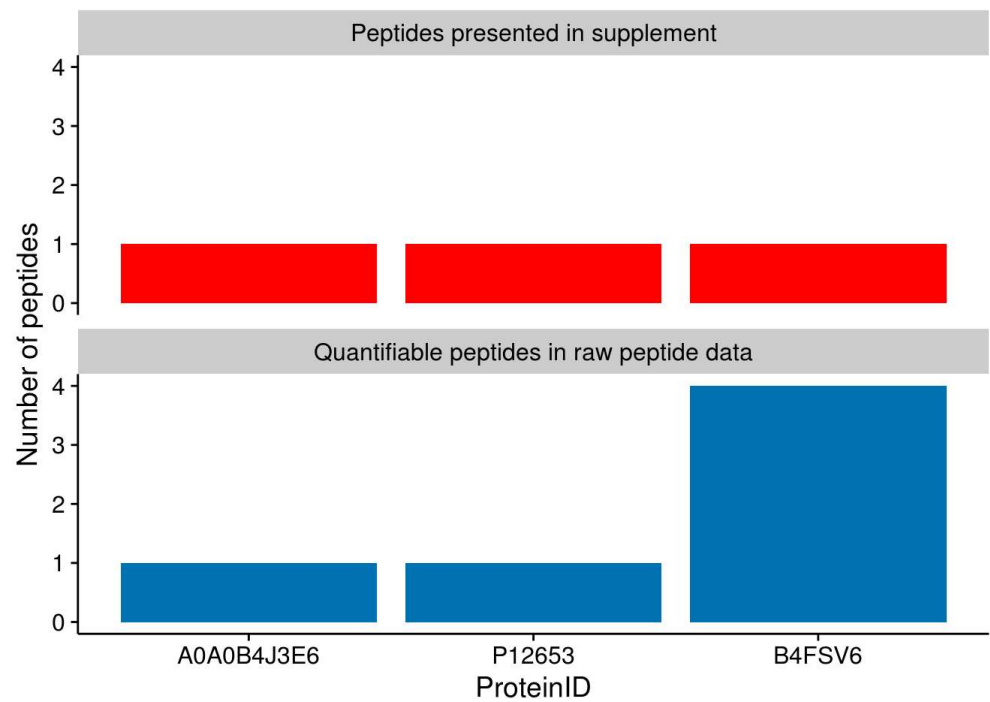


Figure 3. Specific proteins described as having altered expression in the text are poorly supported by the proteomics data. Proteins A0A0B4J3E6 and P12653 are based on single peptide observations, while most B4FSV6 peptides are not differentially expressed. Subset of Figure 2 data

Further demonstration that fold change values in Supplementary File 5 described as being for proteins correspond directly to the peptide fold changes in Additional Data File 1

The first few rows of Supplementary Table 5 are reproduced in Table 2. Notably:

- 1. The top fold depleted protein between the control strain and the Roundup Ready strain is a fungal tubulin, not a maize protein, suggesting the control strains were potentially infected.
- 2. The values given in the column 'Mass (Da)' do not correspond to the masses of the full proteins (Table 3). Full proteins have masses in the 10's of thousands of Daltons. Instead, the values in the 'Mass (Da)' column correspond to the Mass to Charge ratio in the peptide TMT file (Table 4), a value calculated from mass and charge state of an individual spectrum.

Table 2. First few rows of Supplementary Table 5

UniProt ID	Protein name	Mass (Da)	Log2 FC	P-adjusted values
W7LNM5	Tubulin alpha chain OS=Gibberella moniliformis (strain M3125 / FGSC 7600) GN=FVEG_00855	626.81	-3.7702	0.0011
B6SIZ2	Oleosin OS=Zea mays GN=LOC100280642	401.21	-3.0929	0.0019

UniProt ID	Protein name	Mass (Da)	Log2 FC	P-adjusted values
Q41784	Tubulin beta-7 chain OS=Zea mays GN=TUBB7	761.09	-3.0487	0.0058

Table 3. Actual protein masses are two orders of magnitude larger than the masses in Supplementary Table 5 [cite ExPasy

UniProt ID	Mass (Da)
W7LNM5	50378.66
B6SIZ2	18332.85
Q41784	50094.36

Table 4. Proteins described as having significant fold changes in Supplementary Table 5 are actually peptides, with Mass to Charge Ratio (m/z) mislabeled as Mass (Da)

Peptide	UniProt ID	Modifications	m/z	charge state
eDAANNYAR	W7LNM5	N-Term(TMT6plex)	626.81	2
tPDYVEEAHRR	B6SIZ2	N-Term(TMT6plex)	401.21	2
eLHIQGGQcGNQIGAk	Q41784	N-Term(TMT6plex); C10(Carbamidomethyl); K17(TMT6plex)	761.09	2

Conclusion

We found that -in contrast to the authors' assertions- peptides, not proteins, are quantified in this paper, and that this likely introduces false positives in measuring differential protein expression.. This same type of analysis was used in another recent paper by the same lead authors in Scientific Reports, "Multiomics reveal non-alcoholic fatty liver disease in rats following chronic exposure to an ultra-low dose of Roundup herbicide"⁶. Both papers incorrectly use peptide fold changes as a proxy for full protein differences, and thus their conclusions are based on misinterpretation of the data.

Conflict of Interest Statement

Neither CDM nor DRB have any conflicts of interest with the subject matter of this manuscript

References

- [1] <https://www.nature.com/articles/srep37855> (<https://www.nature.com/articles/srep37855>) "An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process", Robin Mesnage, Sarah Z. Agapito-Tenfen, Vinicius Vilperte, George Renney, Malcolm Ward, Gilles-Eric Seralini, Rubens O. Nodari & Michael N. Antoniou, Scientific Reports 6, Article number: 37855 (2016), doi:10.1038/srep37855 (doi:10.1038/srep37855)
- [2] <http://www.mcponline.org/content/3/12/1154.full> (<http://www.mcponline.org/content/3/12/1154.full>) "Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents", Philip L. Ross, Yulin N. Huang, Jason N. Marchese, Brian Williamson, Kenneth Parker, Stephen Hattan, Nikita Khainovski, Sasi Pillai, Subhakar Dey, Scott Daniels, Subhasish Purkayastha, Peter Juhasz, Stephen Martin, Michael Bartlett-Jones, Feng He Allan Jacobson and Darryl J. Pappin, Molecular & Cellular Proteomics, 3, 1154-1169 (2004), doi: 10.1074/mcp.M400129-MCP200 December 1, 2004
- [3] <https://tools.thermofisher.com/content/sfs/brochures/AN-63410-Quantitation-of-TMT-Labeled-Peptides-Velos-Pro-Proteomics.pdf> (<https://tools.thermofisher.com/content/sfs/brochures/AN-63410-Quantitation-of-TMT-Labeled-Peptides-Velos-Pro-Proteomics.pdf>) "Quantitation of TMT-Labeled Peptides Using Higher-Energy Collisional Dissociation on the Velos Pro Ion Trap Mass Spectrometer, Roger G. Biringer, Julie A.

Horner, Rosa Viner, Andreas F. R. Hühmer, August Specht, Thermo Fisher Scientific, San Jose, California, USA

[4] https://link.springer.com/protocol/10.1007%2F978-1-60761-780-8_12 (https://link.springer.com/protocol/10.1007%2F978-1-60761-780-8_12) "Quantification of Proteins by iTRAQ", Richard D. Unwin, LC-MS/MS in Proteomics, Volume 658 of the series Methods in Molecular Biology pp 205-215 (2010)

[5] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4261935/> (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4261935/>) "Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics", Navin Rauniyar and John R. Yates, III*, Journal of Proteome Research, 13(12): 5293–5309 (2014), doi:10.1021/pr500880b (doi:10.1021/pr500880b)

[6] <https://www.nature.com/articles/srep39328> (<https://www.nature.com/articles/srep39328>) "Multiomics reveal non-alcoholic fatty liver disease in rats following chronic exposure to an ultra-low dose of Roundup herbicide", Robin Mesnage, George Renney, Gilles-Eric Séralini, Malcolm Ward & Michael N. Antoniou, Scientific Reports 7, Article number: 39328 (2017), doi:10.1038/srep39328 (doi:10.1038/srep39328)