Extended Discussion

Mass Spectrometry Meta-Analysis

Large-scale untargeted mass spectrometry analyses are often confounded by relatively low reproducibility of peptide identifications between experiments (1). The large number of phosphopeptides present, the stochastic nature of peptide sampling through data-dependent analyses, and variances in peptide ionization are the main contributors to this phenomenon. To investigate reproducibility of proteomic identifications in wild-type tissue, protein and phosphoisoforms (i.e., peptide(s) with a specific phosphorylation pattern) were binned according to the number of experiments in which they were identified (Figure S3A). Proteins were measured in an average of 4-5 ($\mu = 4.5$) experiments, while phosphoisoforms were only measured in an average of 2-3 experiments ($\mu = 2.5$). In addition to the reasons discussed above, lower reproducibility of phosphoisoform measurements is due to the generally low abundance of phosphorylation on target proteins and the necessity of assigning phosphoryl groups to a specific amino acid within a peptide. Enrichment of phosphopeptides enables identification of low-level species, but introduces variability since the overall pool of possible phosphopeptides may be extremely large, involving several hundred thousand in number, e.g., if each of the ca. 40,000 Medicago proteins is phosphorylated at several sites. Phosphopeptide quantification requires the assignment of the phosphoryl group to a specific amino acid residue within a peptide (i.e., localization). Simply identifying phosphopeptides does not ensure localization of phosphosites, as localization is dependent on MS/MS spectral quality. In quantitative studies, both identification and localization of phosphopeptides are required, increasing the challenges of reproducible protein phosphorylation measurements.

Due to multiplexing, it is difficult to directly correlate the number of experiments with the number of biological replicates (i.e., *N*). To determine the distribution of *N* on our measurements, and ultimately our ability to determine significance, the number of wild-type proteins or phosphoisoforms for each *N* was plotted (Figure S3B). Protein measurements are more reproducible and will have high statistical power, due to a large average *N* ($\mu = 7$). An average phosphoisoform has an *N* between 3 and 4 ($\mu = 3.5$), offering modest statistical power, but representing more replicates than similar studies of this scale. Recently, we have demonstrated that a large *N* (12) is sufficient to significantly measure subtle changes using isobaric labels (2). As our average *N* is much lower, we ensure changes are biologically relevant by significance testing (Student's t-test) and imposing a fold cutoff of 1.35.

To determine the distribution of ratios in all WT experiments we plotted the count of localized phosphoisoforms with a certain $\log_2$ ratio of +NF/-NF (Figure S4). These distributions clearly show the compression of the dynamic range of changes due to both the asynchronous nature of our samples, as well as the fact the cells that are changing will be a small population within the whole plant extract. This compression of quantitative data was consistent through all experiments and was a contributing factor in setting the fold cutoff to 1.35.

In addition to biological replicates, we explored cellular fractionation to obtain deep proteomic coverage. Of the 11 quantitative experiments, proteins were isolated from whole cell lysate, isolated microsomes, and the upper and lower phases of a two phase membrane separation (plasma membrane is enriched in the upper phase). To investigate the benefit of analyzing multiple cellular compartments we divided both protein and phosphoisoform identifications into five classes: whole cell (species which were in a combination of compartments, including the whole cell), microsome only, upper phase only, lower phase only, and not in whole cell (species

which were in a combination of compartments, not including the whole cell). For protein analysis, 5.3% of proteins were not detected in the whole cell fraction, suggesting a benefit from cellular fractionation. A greater number of phosphoisoforms, 9%, were not identified in whole cell fractions, suggesting added value in cellular fractionation for phosphopeptide detection. Phophorylation measurements are made at the peptide level, whereas protein measurements are made based on a collection of peptides. Thus, a unique peptide identified in a membrane fraction may be grouped with an existing protein, but a unique phosphopeptide identified in a membrane fraction constitutes a new measurement. Here, cellular fractionation increases identification of both proteins and phosphoproteins, but clearly phosphorylation measurements benefited more.

Phosphoproteomic comparison to Grimsrud et al. (2010)

We previously published a large-scale analysis of phosphoproteins in the roots of Medicago, identifying 3,404 unique phosphosites on 829 proteins. To compare the peptides identified here to those found in Grimsrud *et al* (3). we compiled all of the localized phosphopeptides from both studies and compared them based solely on their backbone amino acid sequence (Figure S5). We filtered in this way to negate any differences in localization, due to the evolution of our localization software since the original publication. This analysis reveals 539 phosphopeptides are overlapping between the two studies, leaving 1,827 unique phosphopeptides that were identified in Grimsrud *et al.* (3), but not in the current study. The large number of phosphopeptides identified only in Grimsrud *et al.* (3) can be explained by the differences in sample preparation and data acquisition. The original work employed multiple enzymes in addition to trypsin (e.g., LysC, AspN, GluC, etc.) to create a wider array of phosphopeptides that would not be present in a sample only digested by trypsin. In addition, the original description of the Medicago proteome used extensive ETD sequencing, while the current

paper consisted predominantly of HCD spectra. Together, these factors explain the discrepancy identified by the Venn diagram in Figure S5.

Supplemental Figures

Supplemental Figure S1. Schematic illustration of proteomic and phosphoproteomic workflow. Proteins were isolated from *Medicago* whole cell lysate or membrane enriched fractions and peptides were produced *via* tryptic digestion. To enable quantitation, peptides from multiple biological replicates were isobarically labeled and pooled. Samples were fractionated to decrease complexity before immobilized metal affinity chromatography (IMAC) phosphopeptide enrichment. Nano-HPLC-MS/MS analysis of enriched and non-enriched fractions yielded large scale quantitation of protein and phosphorylation changes in response to NF treatment

Supplemental Figure S2. Flow diagram of the data collection scheme.

A-D) Identification of a peptide belonging to the protein calreticulin. The scan cycle consisted of MS1 analysis in the Orbitrap (A), here we also show a zoom in of the isolation window and describe the quantmode scan sequence (B). The resulting MS2 identified the peptide AEAGED(pS)DETNK, belonging to the protein calreticulin (C). The reporter tags were corrected for isotopic impurities, normalized to total protein content of each channel and plotted along with the fold change and p-value (D, Student's t-test).

E-H) Identification of a peptide belonging to the protein nuclear steroid receptor. The scan cycle consisted of MS1 analysis in the Orbitrap (E), here we also show a zoom in of the isolation window and describe the quantmode scan sequence (F). The resulting MS2 identified the peptide HGLP(pS)PT(oM)NR, belonging to the protein nuclear steroid receptor (G). The reporter tags

were corrected for isotopic impurities, normalized to total protein content of each channel and plotted along with the fold change and p-value (H, Student's t-test).

Supplemental Figure S3. Analysis of replication and cellular fractionation in proteomic experiments

(A) Distribution of the number of experiments in which a unique protein or phosphoisoform was identified. On average a protein was identified in 4-5 experiments (μ=4.5), while on average a phosphoisoform was only identified in 2 or 3 experiments (μ=2.5).

(B) Distribution of the number of biological replicated for a unique protein or phosphoisoform. Multiplexing within experiments makes a direct correlation between the number of experiments and the number of biological replicates (i.e., *N*). On average a protein had an *N* of 7, while on average a phosphoisoform had an *N* of 3 or 4 (μ=3.5).

(C) Cellular distribution of proteins and phosphoisoforms in whole cell, microsome, and plasma membrane fractions. "In whole cell" fraction comprises any combination of sub-cellular fractions which contain whole cell. "Not in whole cell" fraction comprises any combination of sub cellular fractions that does not contain whole cell. Analysis of sub-cellular fractions increased the number of identifications by 5.3% and 9% for proteins and phosphoisoforms, respectively.

Supplemental Figure S4. Distribution of +Nod Factor / -Nod Factor ratios for all WT experiments.

For all 9 experiments containing a WT measurement the Log2 ratio was calculated for the +NF/-NF comparison and plotted as a histogram.

Supplemental Figure S5. Comparison to previous Medicago phosphoproteome.

The localized phosphopeptides from Grimsrud et al. (red) and the data outlined in this publication (blue) were compared based on the unique backbone amino acid sequence. This was done to negate any differences imparted by slightly altered localization algorithms. The results were plotted as a Venn diagram, with overlap displayed in purple.

Supplemental Figure S6. Functional group of proteins which showed differential phosphorylation at significant level in the roots of Medicago within one hour of NF treatment in wild-type plants.

Supplemental Figure S7. Time course analysis of NF-induced phosphorylation.

Time course phosphorylation data was clustered hierarchically and grouped using K-means clustering. The first two nodes are depicted in figures 4B and 4C. The remaining eight nodes are depicted here. Functional analysis determined no enrichment for gene ontology terms within these nodes.

Supplemental Figure S8. A global view of the distribution of differentially expressed genes in wild-type plants into functional categories based on Gene Ontology.

Terms from the Gene Ontology Plant GOslim were mapped to the set of differentially expressed genes in wild-type plants as well as the full set of Mt3.5 gene models. Shown are the terms represented in the differentially expressed gene list as a fraction of the total, along side the proportion of the same term within the full set of Mt3.5 genes. Bar heights for the 'unknown' category are not to scale. Actual values are listed above each bar. Values for the full Mt3.5 gene

list represent genes which were expressed in all samples at an average of 5 fragments/kilobase or higher.

Supplemental Figure S9. Quantitative RT-PCR validation of expression levels of selected genes in wild-type plants, *nfp* and *dmi3*.

NF-induced transcriptional changes in the candidate genes in wild-type plants (A), *nfp* (B) and *dmi3* (C) One hour after treatment with NF. Candidates genes were selected for qRT-PCR validation based on transcript abundance, biological relevance and fold-change in RNA sequencing data. *Mt1g043220* which was down-regulated in *dmi3* upon NF-treatment was also validated in wild-type and *nfp* to be down-regulated by NF-treatment. Similarly, *Ctg_82104, Mt4g100380, ctg_72999* which were upregulated in *nfp* were also up-regulated in wild-type plants upon NF-treatment. Error bars represent SD. Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *.

Supplemental Figure S10. Quantitative RT-PCR detection of GSE-induced upregulation of candidate genes in wild-type (A), *nfp* mutant (B) and *dmi3* mutant (C)

One hour after treatment with GSE. *Contig_82104* and *Mt4g100380* were upregulated in both wild-type plant and *nfp* mutant, but not in *dmi3* mutant, indicating that the expression of these candidate genes are upregulated by GSE in DMI3 dependent manner. Contig_72999 was not significantly upregulated in wild-type plant, *nfp* and *dmi3* by GSE-treatment. However, it was upregulated by NF in wild-type and *nfp* plants. *Mt1g043320* expression was down-regulated by GSE-treatment in all the genotypes tested. Error bars represent SD. Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *.

Supplemental Figure S11. Effect of silencing *LYR1* on NF-induced differential regulation of candidate genes *Ctg_82104* in wild-type plants and *nfp* mutant.

Expression level of *Ctg_82104* one hour after NF-treatment in wild-type plants and wild-type plants transformed with *LYR1-RNAi* silencing construct (A), and *nfp* mutant and *nfp* mutant transformed with *LYR1-RNAi* silencing construct (B). (C) The validation of RNAi based silencing of the expression level of *LYR1* by semi-quantitative RT-PCR. Error bars represent SD. (D-E) Transgenic roots of Medicago wild-type plants expressing *LYR1*-RNAi construct were selected using DsRed1 visible reporter (E), Bright field image (D). Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *.

Supplemental Tables

Supplemental Table S1. Complete list of localized phosphorylation sites and changes in their phosphorylation levels upon NF-treatment in Medicago.

Supplemental Table S2. Complete list of proteins identified in the current study and quantification of changes upon NF-treatment.

Supplemental Table S3. Complete list of phosphosites which show significant change upon NF-treatment.

Supplemental Table S4. Summary of Medicago RNA-Sequence read mapping

Supplemental Table S5. Normalized read counts per gene for all sample sets, sorted by normalized expression level.

Supplemental Table S6. Complete list of genes which show significant change at the transcript expression level upon NF-treatment.

Supplemental Table S7. Primers used for quantitative RT-PCR validation of RNA sequencing data

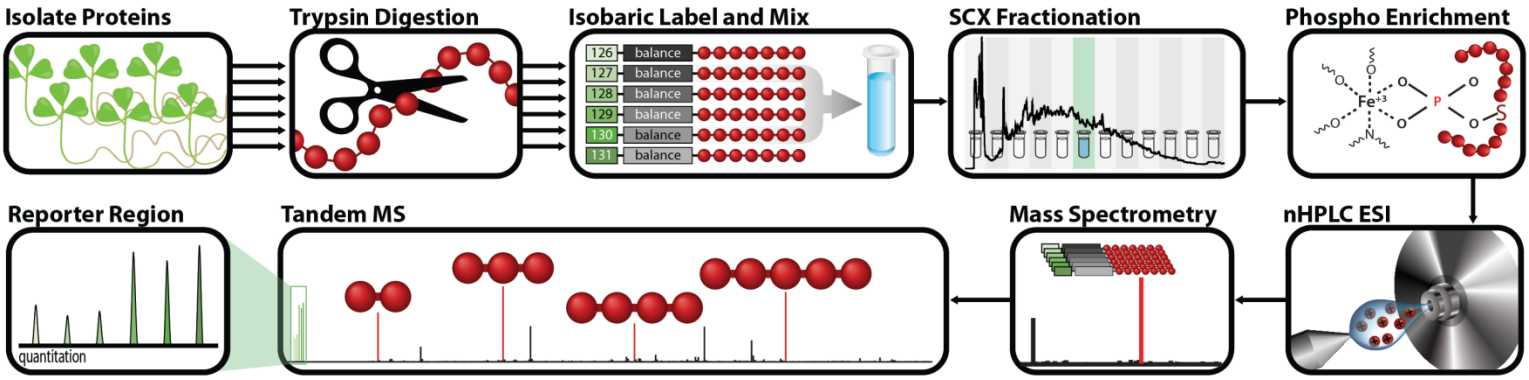
Supplemental References

1. Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C., and Spiegelman, C. (2010) *Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry.* J. Proteome Res. 9, 761-776

2. Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J., Swaney, D. L., Tervo, M. A., Bolin, J. M., Ruotti, V., Stewart, R., Thomson, J. A., and Coon, J. J. (2011) *Proteomic and phosphoproteomic comparison of human ES and iPS cells.* Nat. Methods 8, 821-827

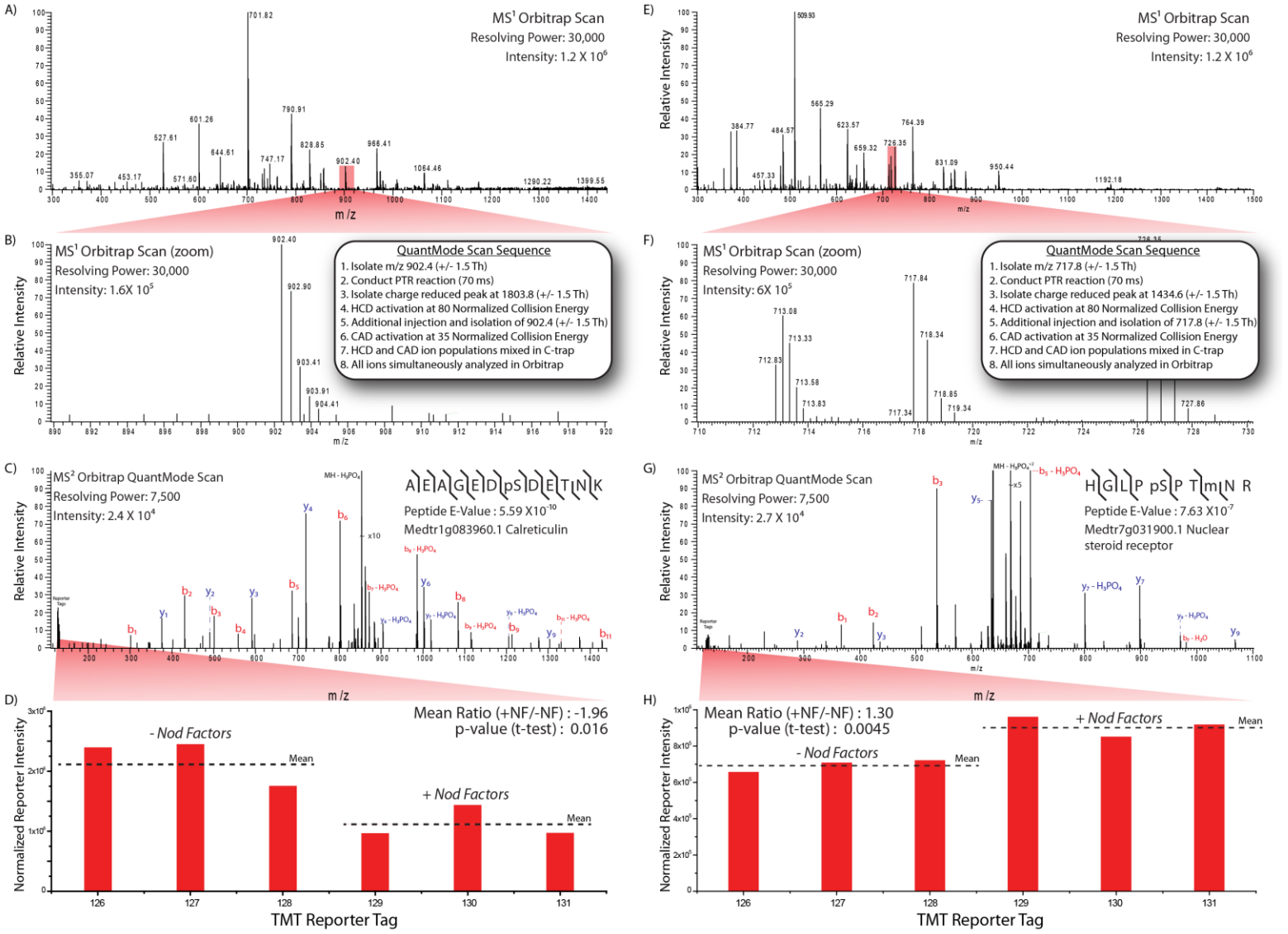
3. Grimsrud, P. A., den Os, D., Wenger, C. D., Swaney, D. L., Schwartz, D., Sussman, M. R., Ané, J. M., and Coon, J. J. (2010) *Large-scale phosphoprotein analysis in Medicago truncatula roots provides insight into in vivo kinase activity in legumes.* Plant Physiol. 152, 19-28

Rose *et al* Supplemental Fig. S1



Supplemental Fig. S1. Schematic illustration of proteomic and phosphoproteomic workflow. Proteins were isolated from *Medicago* whole cell lysate or membrane enriched fractions and peptides were produced *via* tryptic digestion. To enable quantitation, peptides from multiple biological replicates were isobarically labeled and pooled. Samples were fractionated to decrease complexity before immobilized metal affinity chromatography (IMAC) phosphopeptide enrichment. Nano-HPLC-MS/MS analysis of enriched and non-enriched fractions yielded large scale quantitation of protein and phosphorylation changes in response to NF treatment

Rose *et al* Supplemental Fig. S2

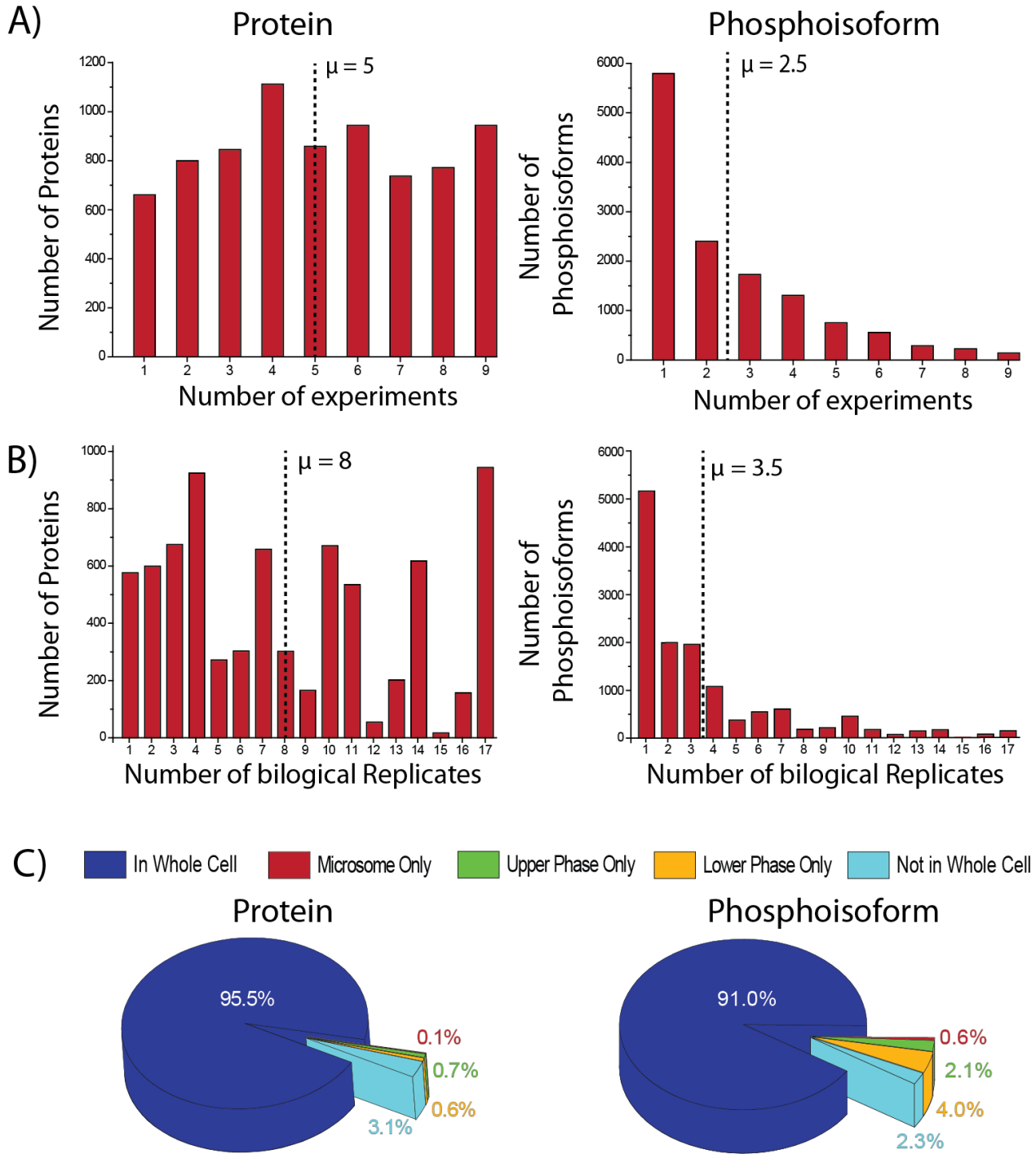


Supplemental Fig. S2. Flow diagram of the data collection scheme.

A-D) Identification of a peptide belonging to the protein calreticulin. The scan cycle consisted of MS1 analysis in the Orbitrap (A), here we also show a zoom in of the isolation window and describe the quantmode scan sequence (B). The resulting MS2 identified the peptide AEAGED(pS)DETNK, belonging to the protein calreticulin (C). The reporter tags were corrected for isotopic impurities, normalized to total protein content of each channel and plotted along with the fold change and p-value (D, Student's t-test).

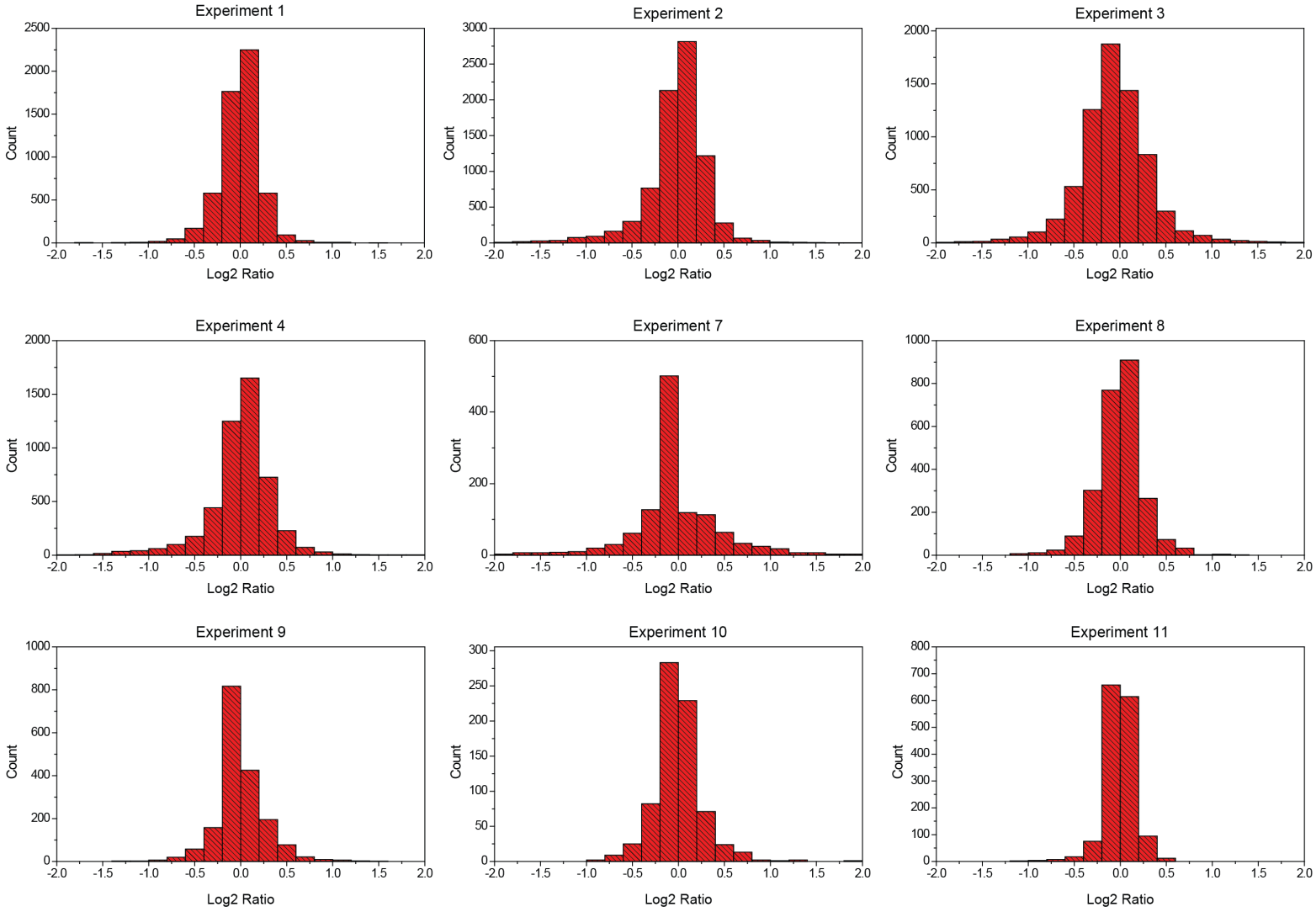
E-H) Identification of a peptide belonging to the protein nuclear steroid receptor. The scan cycle consisted of MS1 analysis in the Orbitrap (E), here we also show a zoom in of the isolation window and describe the quantmode scan sequence (F). The resulting MS2 identified the peptide HGLP(pS)PT(oM)NR, belonging to the protein nuclear steroid receptor (G). The reporter tags were corrected for isotopic impurities, normalized to total protein content of each channel and plotted along with the fold change and p-value (H, Student's t-test).

Rose *et al* Supplemental Fig. S3



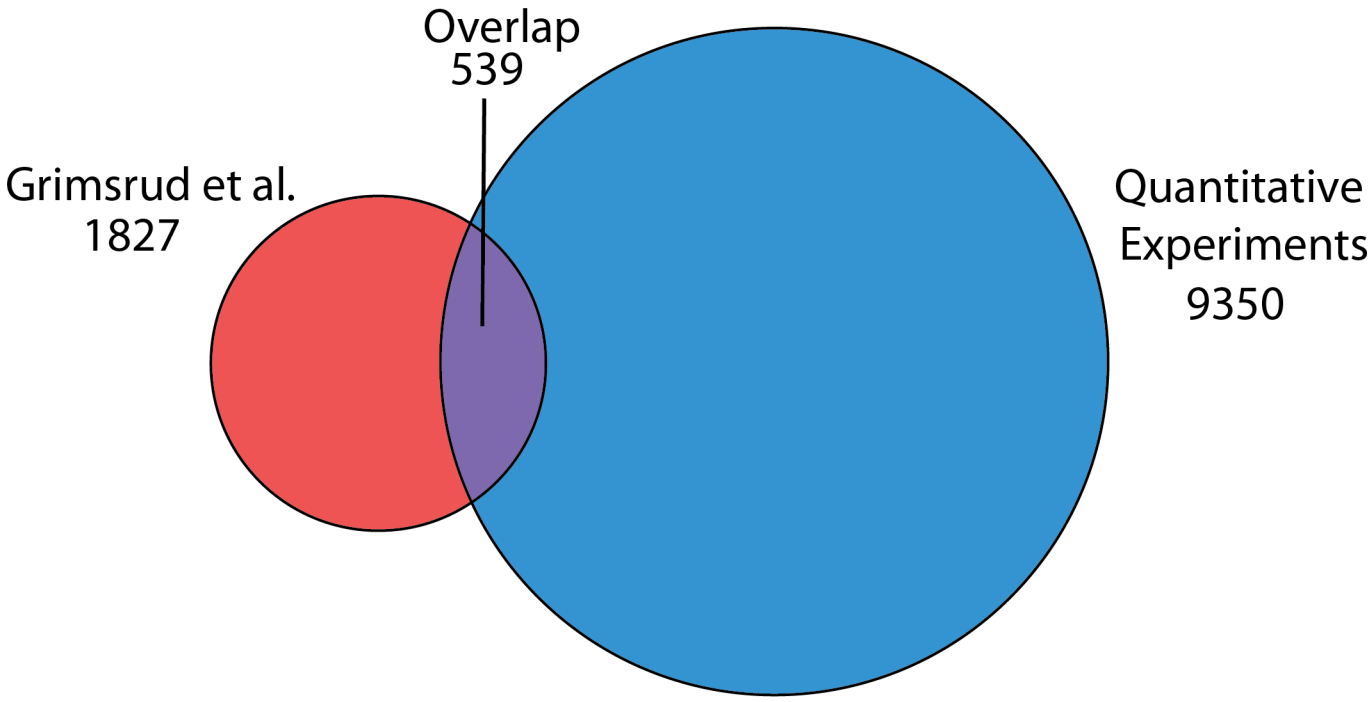
Supplemental Fig. S3. Analysis of replication and cellular fractionation in proteomic experiments
(A) Distribution of the number of experiments in which a unique protein or phosphoisoform was identified. On average a protein was identified in 4-5 experiments (μ=4.5), while on average a phosphoisoform was only identified in 2 or 3 experiments (μ=2.5).
(B) Distribution of the number of biological replicated for a unique protein or phosphoisoform. Multiplexing within experiments makes a direct correlation between the number of experiments and the number of biological replicates (i.e., *N*). On average a protein had an *N* of 7, while on average a phosphoisoform had an *N* of 3 or 4 (μ=3.5).
(C) Cellular distribution of proteins and phosphoisoforms in whole cell, microsome, and plasma membrane fractions. "In whole cell" fraction comprises any combination of sub-cellular fractions which contain whole cell. "Not in whole cell" fraction comprises any combination of sub cellular fractions that does not contain whole cell. Analysis of sub-cellular fractions increased the number of identifications by 5.3% and 9% for proteins and phosphoisoforms, respectively.

Rose *et al* Supplemental Fig. S4



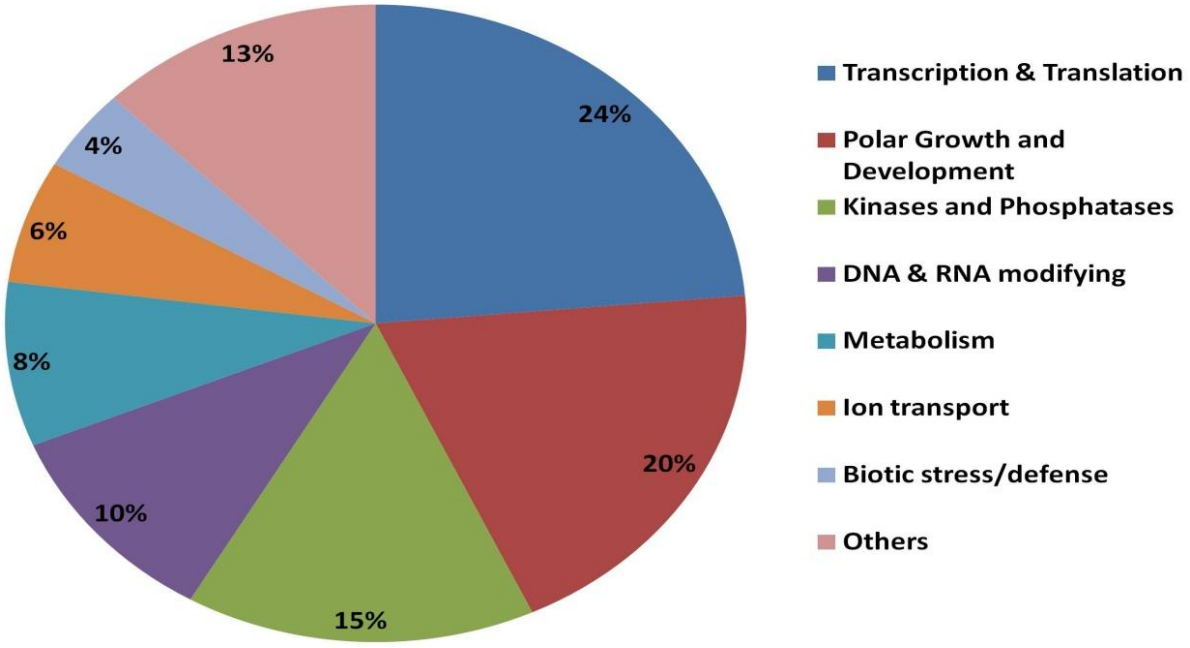
Supplemental Fig. S4. Distribution of +Nod Factor / -Nod Factor ratios for all WT experiments. For all 9 experiments containing a WT measurement the Log2 ratio was calculated for the +NF/-NF comparison and plotted as a histogram.

Rose *et al* Supplemental Fig. S5



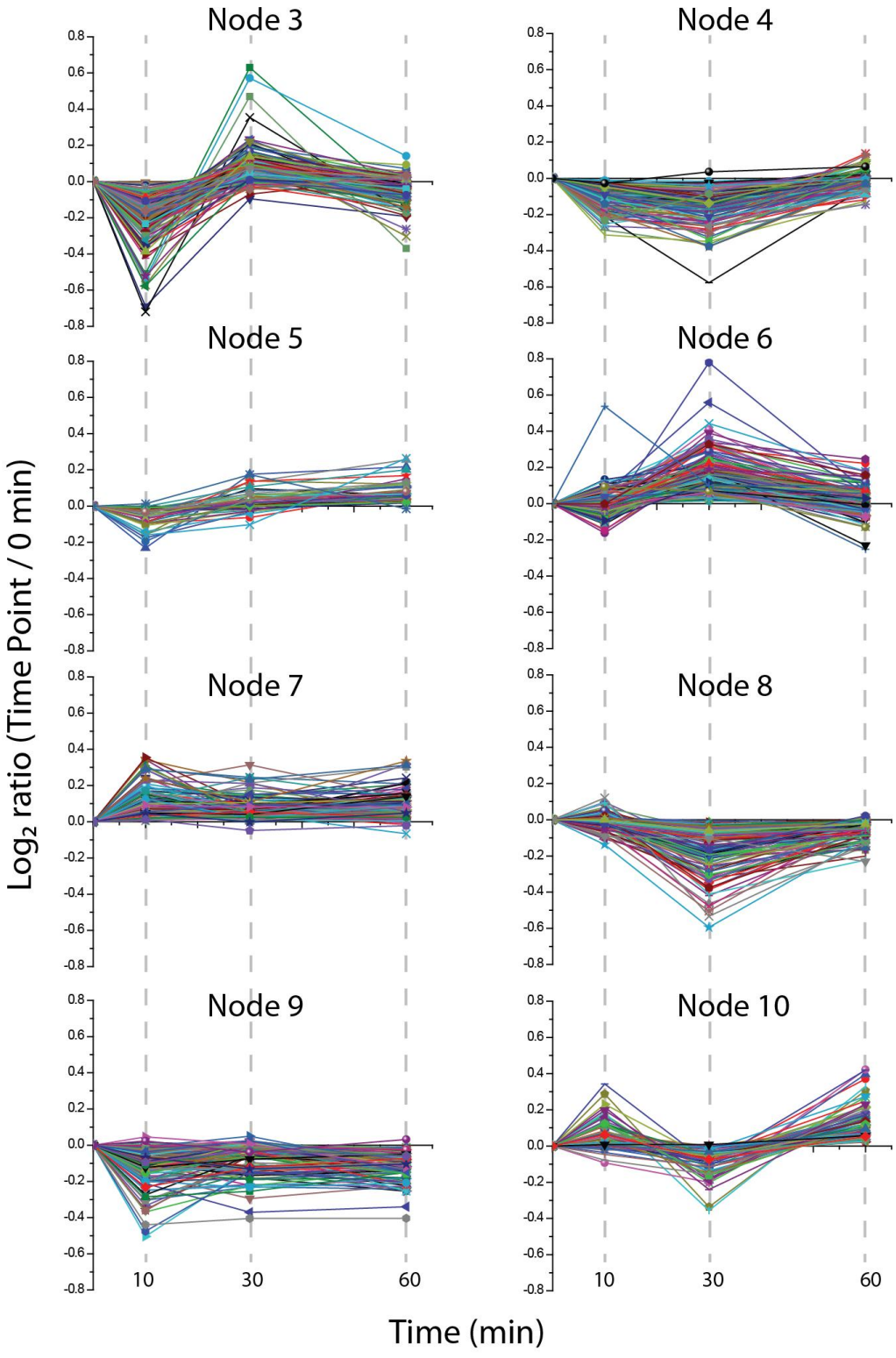
Supplemental Fig. S5. Comparison to previous Medicago phosphoproteome. The localized phosphopeptides from Grimsrud et al. (red) and the data outlined in this publication (blue) were compared based on the unique backbone amino acid sequence. This was done to negate any differences imparted by slightly altered localization algorithms. The results were plotted as a Venn diagram, with overlap displayed in purple.

Rose *et al* Supplemental Fig. S6



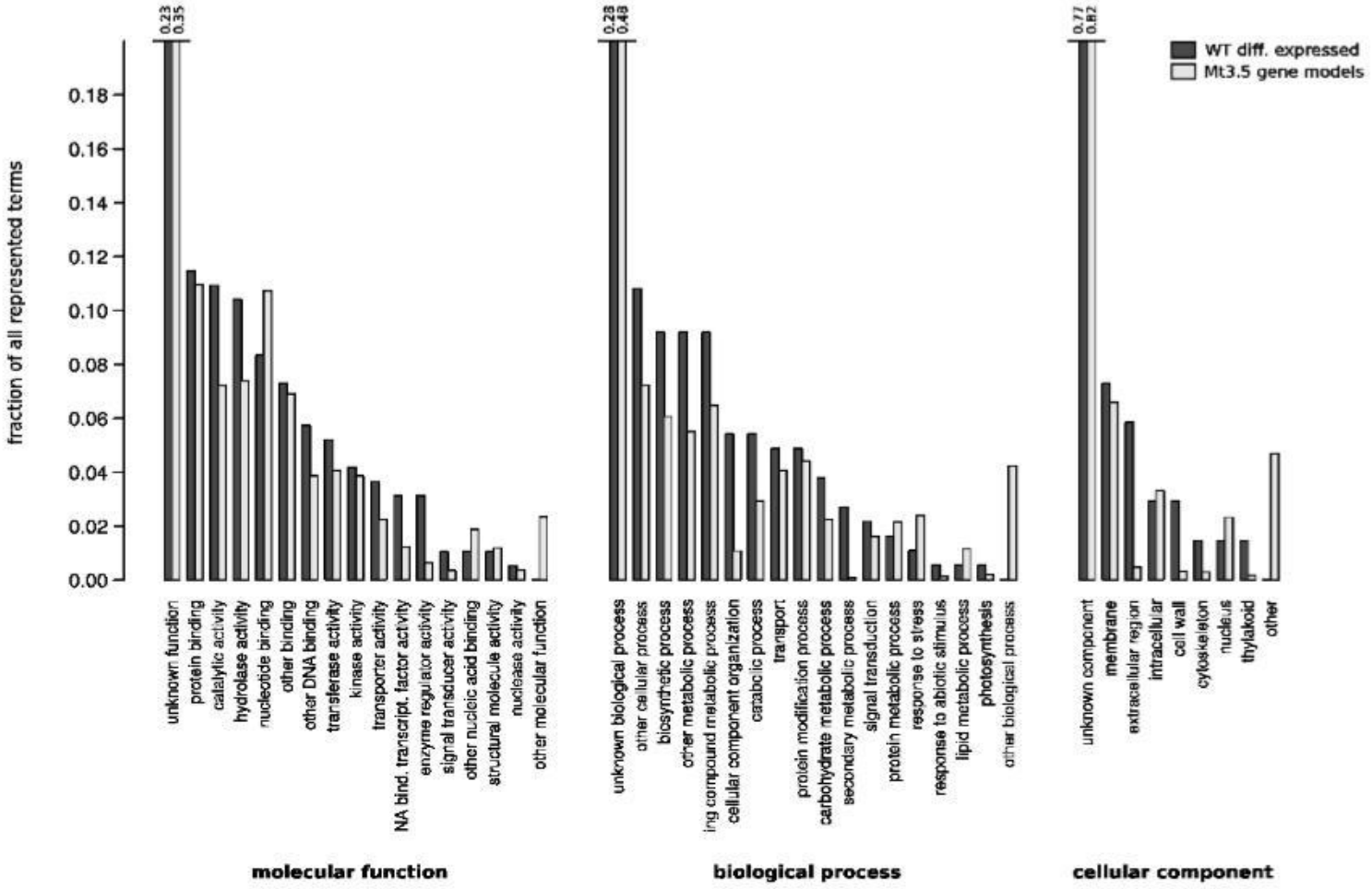
Supplemental Fig. S6. Functional group of proteins which showed differential phosphorylation at significant level in the roots of Medicago within one hour of NF treatment in wild-type plants.

Rose *et al* Supplemental Fig. S7



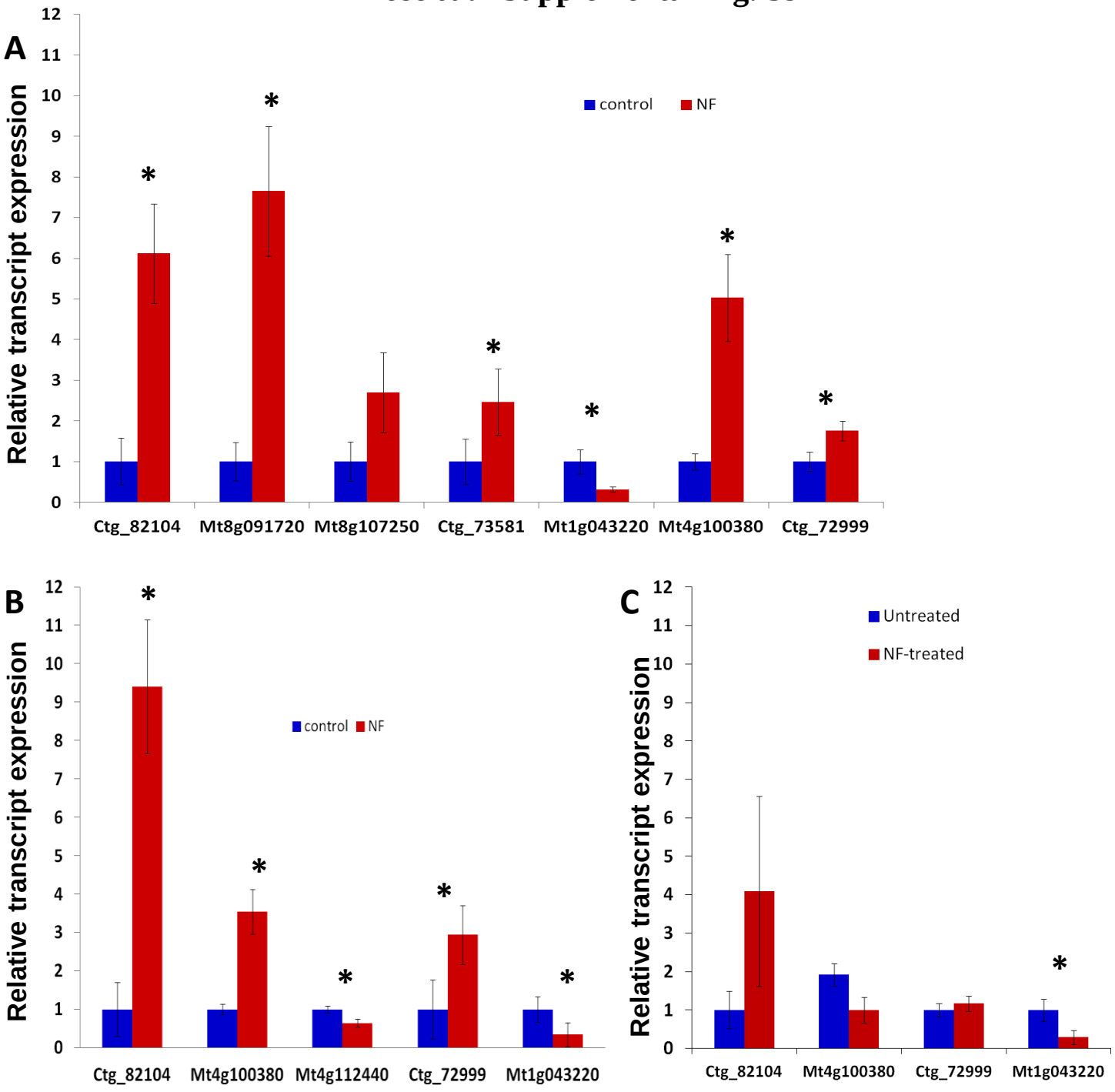
Supplemental Fig. S7. Time course analysis of NF-induced phosphorylation. Time course phosphorylation data was clustered hierarchically and grouped using K-means clustering. The first two nodes are depicted in figures 4B and 4C. The remaining eight nodes are depicted here. Functional analysis determined no enrichment for gene ontology terms within these nodes.

Rose *et al* Supplemental Fig. S8



Supplemental Fig. S8. A global view of the distribution of differentially expressed genes in wild-type plants into functional categories based on Gene Ontology. Terms from the Gene Ontology Plant GOslim were mapped to the set of differentially expressed genes in wild-type plants as well as the full set of Mt3.5 gene models. Shown are the terms represented in the differentially expressed gene list as a fraction of the total, along side the proportion of the same term within the full set of Mt3.5 genes. Bar heights for the 'unknown' category are not to scale. Actual values are listed above each bar. Values for the full Mt3.5 gene list represent genes which were expressed in all samples at an average of 5 fragments/kilobase or higher.

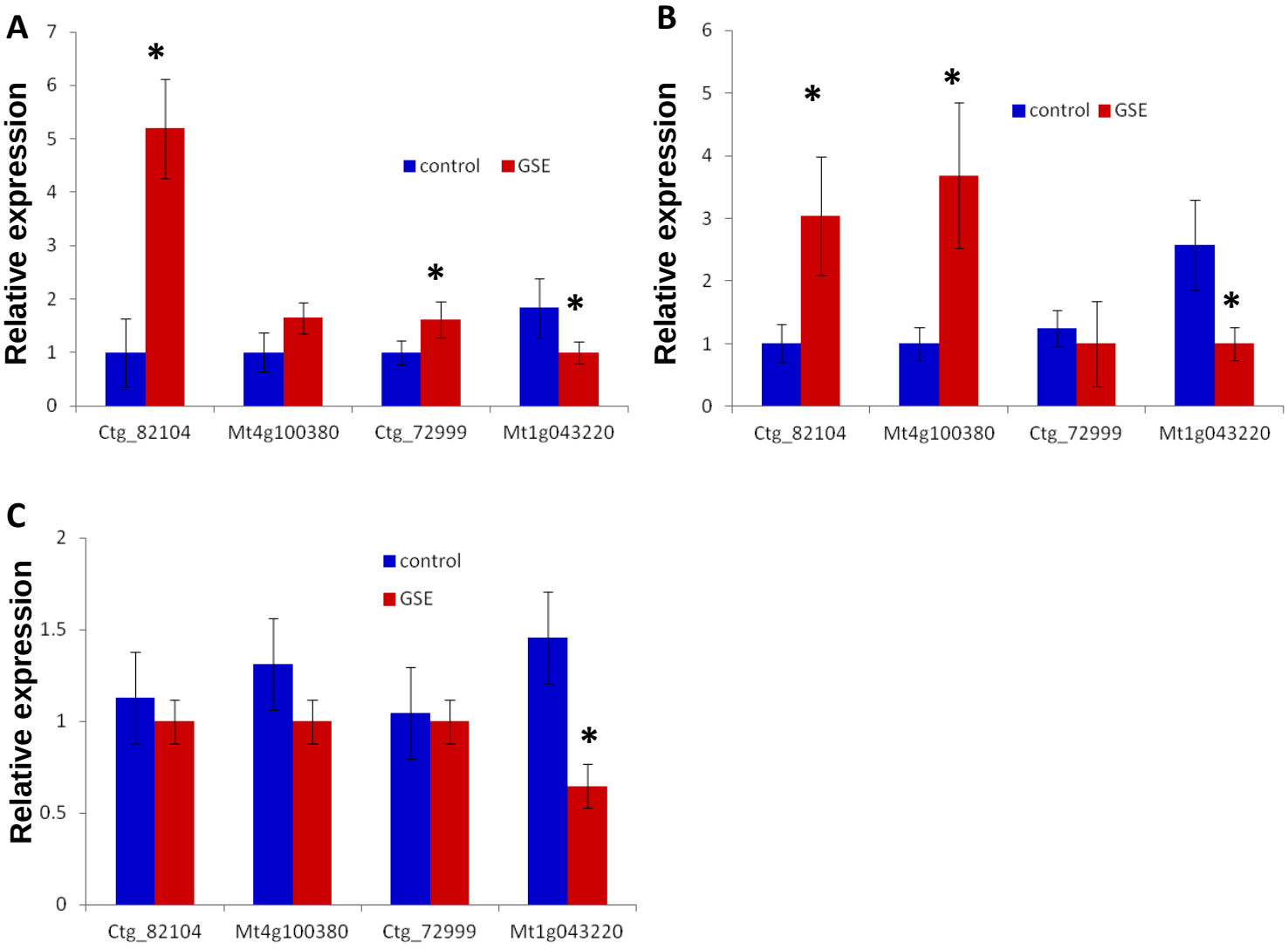
Rose *et al* Supplemental Fig. S9



Supplemental Fig. S9. Quantitative RT-PCR validation of expression levels of selected genes in wild-type plants, *nfp* and *dmi3*.

NF-induced transcriptional changes in the candidate genes in wild-type plants (A), *nfp* (B) and *dmi3* (C) One hour after treatment with NF. Candidates genes were selected for qRT-PCR validation based on transcript abundance, biological relevance and fold-change in RNA sequencing data. *Mt1g043220* which was down-regulated in *dmi3* upon NF-treatment was also validated in wild-type and *nfp* to be down-regulated by NF-treatment. Similarly, *Ctg_82104*, *Mt4g100380*, *ctg_72999* which were upregulated in *nfp* were also up-regulated in wild-type plants upon NF-treatment. Error bars represent SD. Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *.

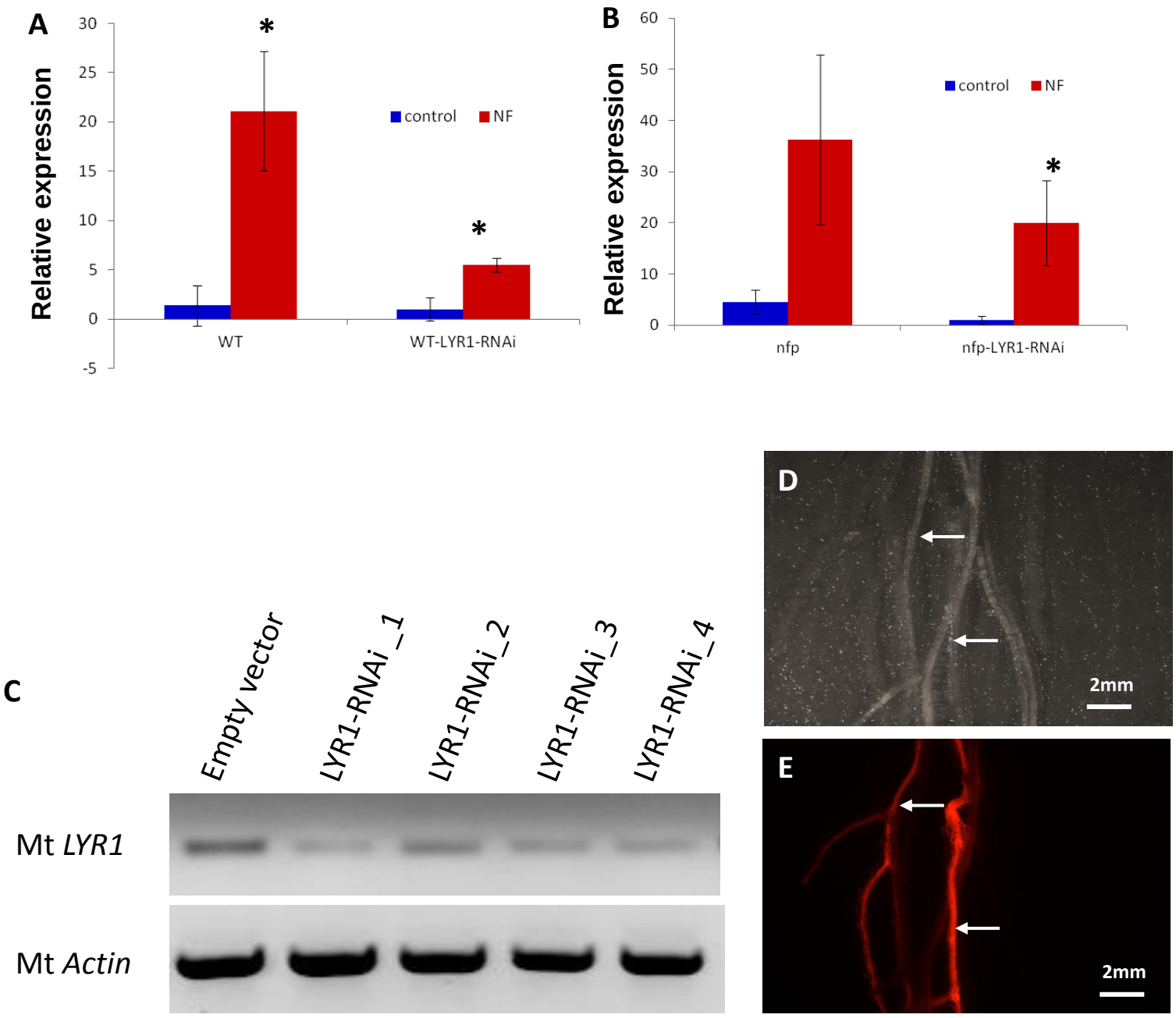
Rose *et al* *Supplemental* Fig. S10



Supplemental Fig. S10. Quantitative RT-PCR detection of GSE-induced upregulation of candidate genes in wild-type (A), *nfp* mutant (B) and *dmi3* mutant (C)

One hour after treatment with GSE. *Contig_82104* and *Mt4g100380* were upregulated in both wild-type plant and *nfp* mutant, but not in *dmi3* mutant, indicating that the expression of these candidate genes are upregulated by GSE in DMI3 dependent manner. Contig_72999 was not significantly upregulated in wild-type plant, *nfp* and *dmi3* by GSE-treatment. However, it was upregulated by NF in wild-type and *nfp* plants. *Mt1g043320* expression was down-regulated by GSE-treatment in all the genotypes tested. Error bars represent SD. Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *****.

Rose *et al* Supplemental Fig. S11



Supplemental Fig. S11. Effect of silencing *LYR1* on NF-induced differential regulation of candidate genes *Ctg_82104* in wild-type plants and *nfp* mutant.

Expression level of *Ctg_82104* one hour after NF-treatment in wild-type plants and wild-type plants transformed with *LYR1-RNAi* silencing construct (A), and *nfp* mutant and *nfp* mutant transformed with *LYR1-RNAi* silencing construct (B). (C) The validation of RNAi based silencing of the expression level of *LYR1* by semi-quantitative RT-PCR. Error bars represent SD. (D-E) Transgenic roots of Medicago wild-type plants expressing *LYR1*-RNAi construct were selected using DsRed1 visible reporter (E), Bright field image (D). Changes in the expression level for the candidates which are significant at $P<0.05$ are indicated by *.