GENE REGULATION

Integration of omic networks in a developmental atlas of maize

Justin W. Walley,[1,2*] Ryan C. Sartor,[1*] Zhouxin Shen,[1] Robert J. Schmitz,[3,4†] Kevin J. Wu,[1] Mark A. Urich,[3,4] Joseph R. Nery,[4] Laurie G. Smith,[1] James C. Schnable,[5] Joseph R. Ecker,[3,4,6] Steven P. Briggs[1‡]

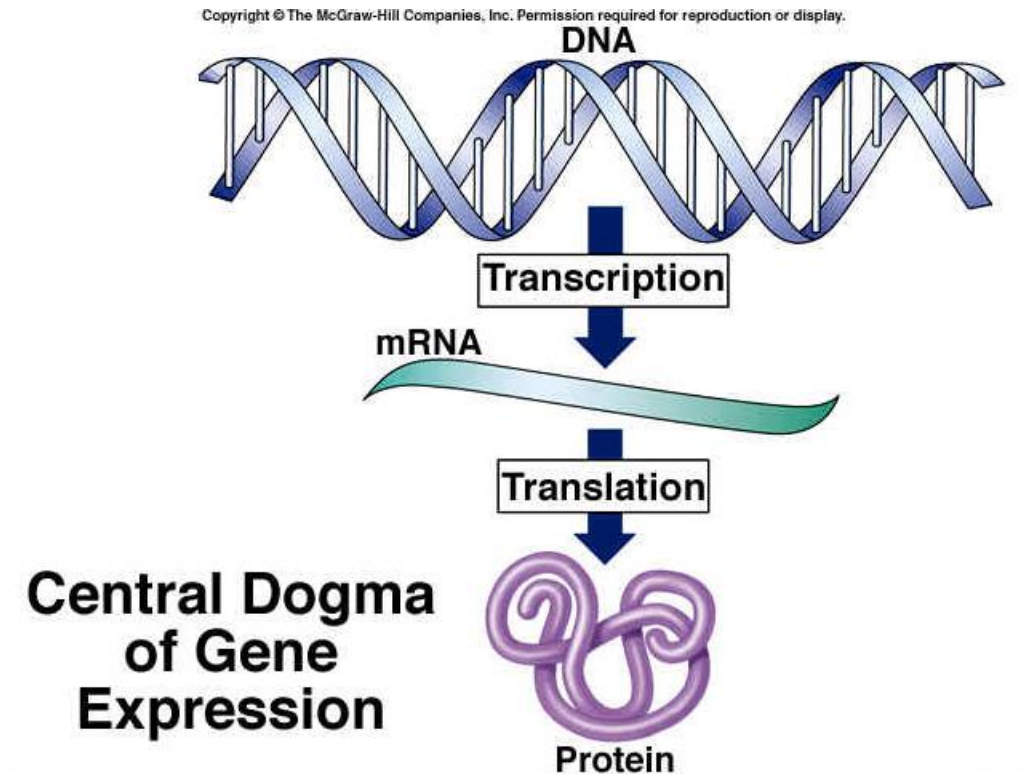
讲述人：周向宇

Background

Gene function

- GRNs
- Coexpression networks

Transcriptome and proteomics



Material and Method

- Materials
  - 23 tissues spanning vegetative and reproductive stages of maize development

- Methods
  - Transcriptome: mRNA-seq
  - Proteome:Electrospray ionization tandem mass spectrometry

Comparison of transcriptome and proteome data sets.

- **Reproducibility of the biological replicates**
  - Transciptome : 0.9
  - Proteome : 0.84
  - Phosphoproteome :0.7
- **The number of genes**
  - Transcripts were from 62,547 genes，
  - Proteins were from 6946 genes，
  - Phosphoproteins were from 5587 genes

Comparison of transcriptome and proteome data sets.

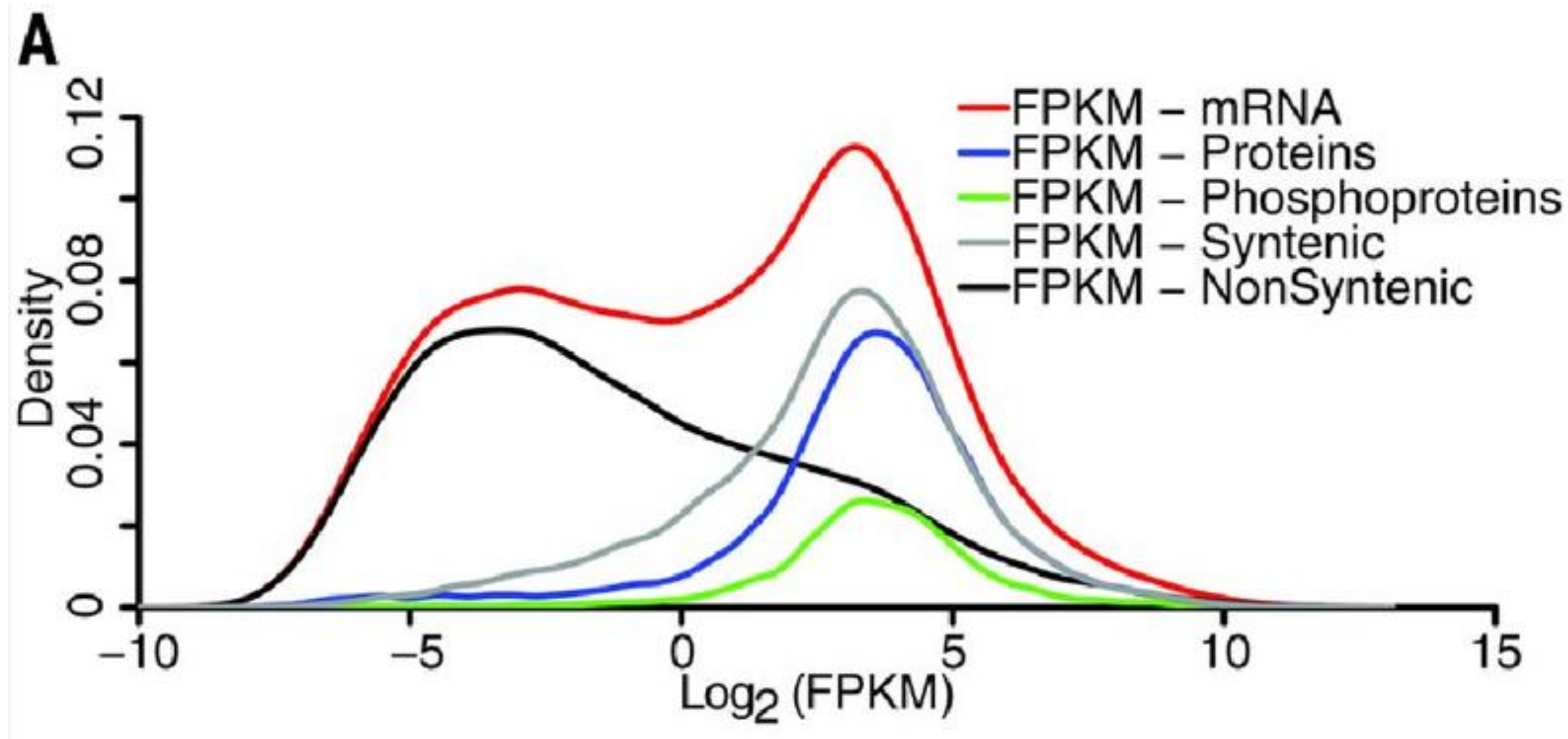


Fig. 1Comparison of transcriptome and proteome data sets.

(A) FPKM distribution of mRNA abundance (red). FPKM values of transcripts corresponding to quantified proteins (blue), phosphopeptides (green), syntenic genes conserved between maize and sorghum (gray), and nonsyntenic genes (black) are shown. Data are the average expression from the 23 tissues profiled.

Comparison of transcriptome and proteome data sets

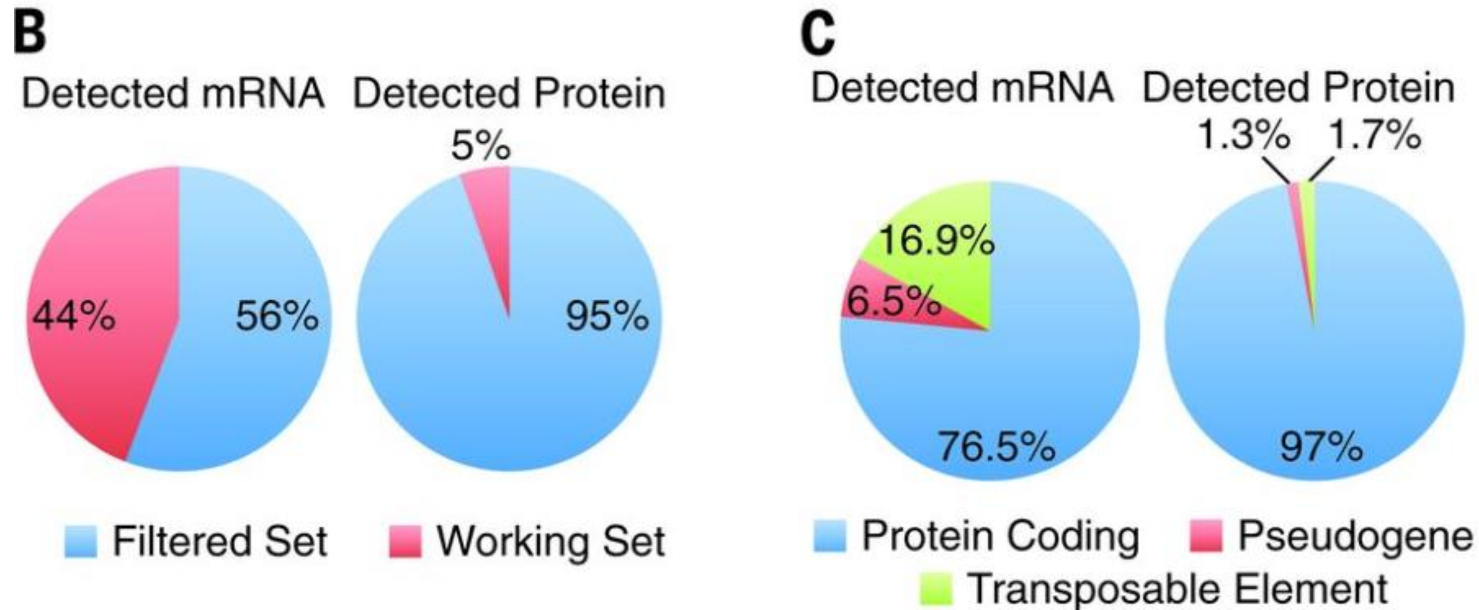


Fig B
Percentage of quantified mRNA and proteins in the annotated filtered (high-confidence gene models) and working (all gene models) gene sets.

Fig C
Breakdown of detected mRNA and proteins, based on annotations.

Conclusion:
transcripts from many genes may not produce proteins

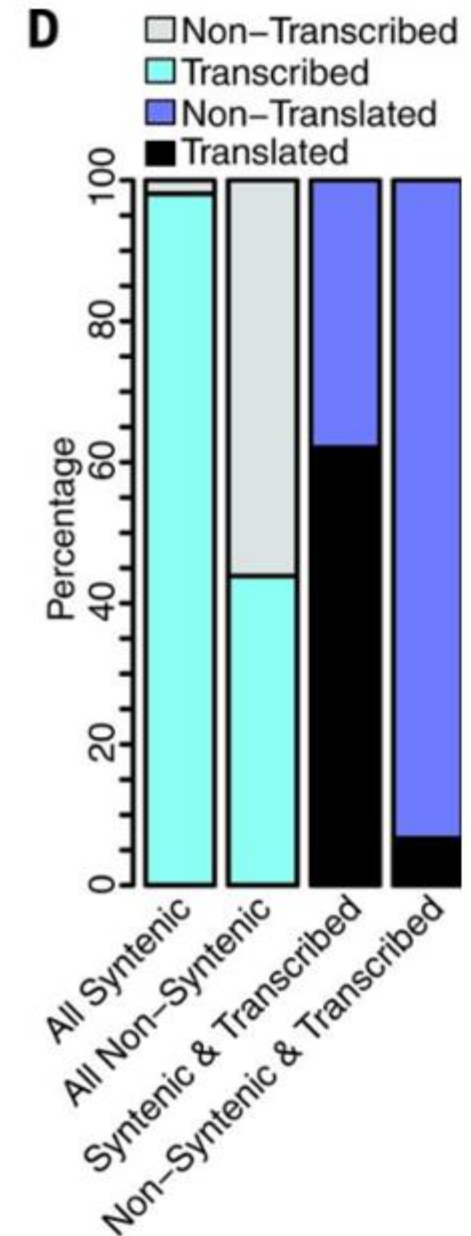
Comparison of transcriptome and proteome data sets

Fig D

- Percentages of all annotated genes that are transcribed and percentages of all transcribed genes that are translated, for both the syntenic and nonsyntenic gene sets.

Conclusion:

- A greater frequency of protein expression is a possible mechanistic explanation for the eightfold enrichment of genes responsible for visible mutant phenotypes among syntenically conserved genes in maize



- Whether transcriptome-based networks predict the same relationships as proteome-basednetworks？

- NEXT： coexpression networks and GRNs

coexpression networks

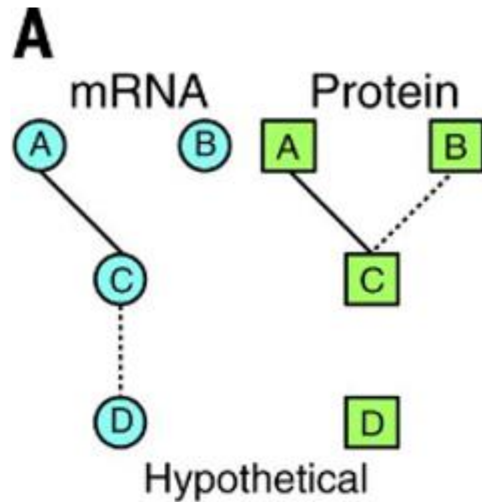


Fig A
Hypothetical undirected coexpression subnetwork showing conserved (solid lines) and nonconserved (dotted lines) coexpression edges between mRNA and protein networks.

- Feature:undirect
- Node:genes connected on the basis of highly correlated expression patterns
- Method:Spearman correlations， WGCNA
- Threshold---correlation score >0.75

Compare the mRNA and protein based coexpression networks

- Calculate edge conservation
- Found 6.1% edges were conserved in both networks

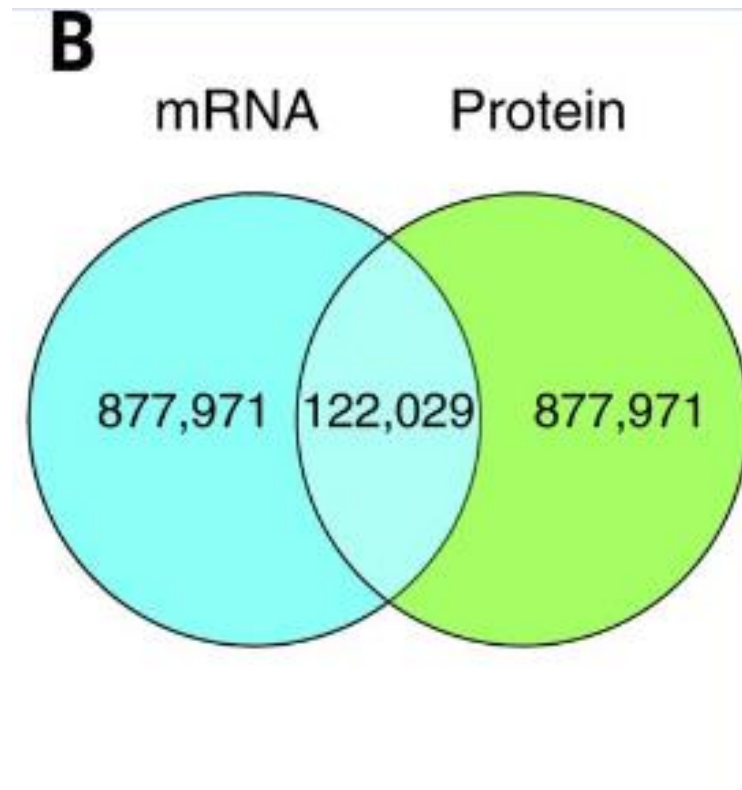
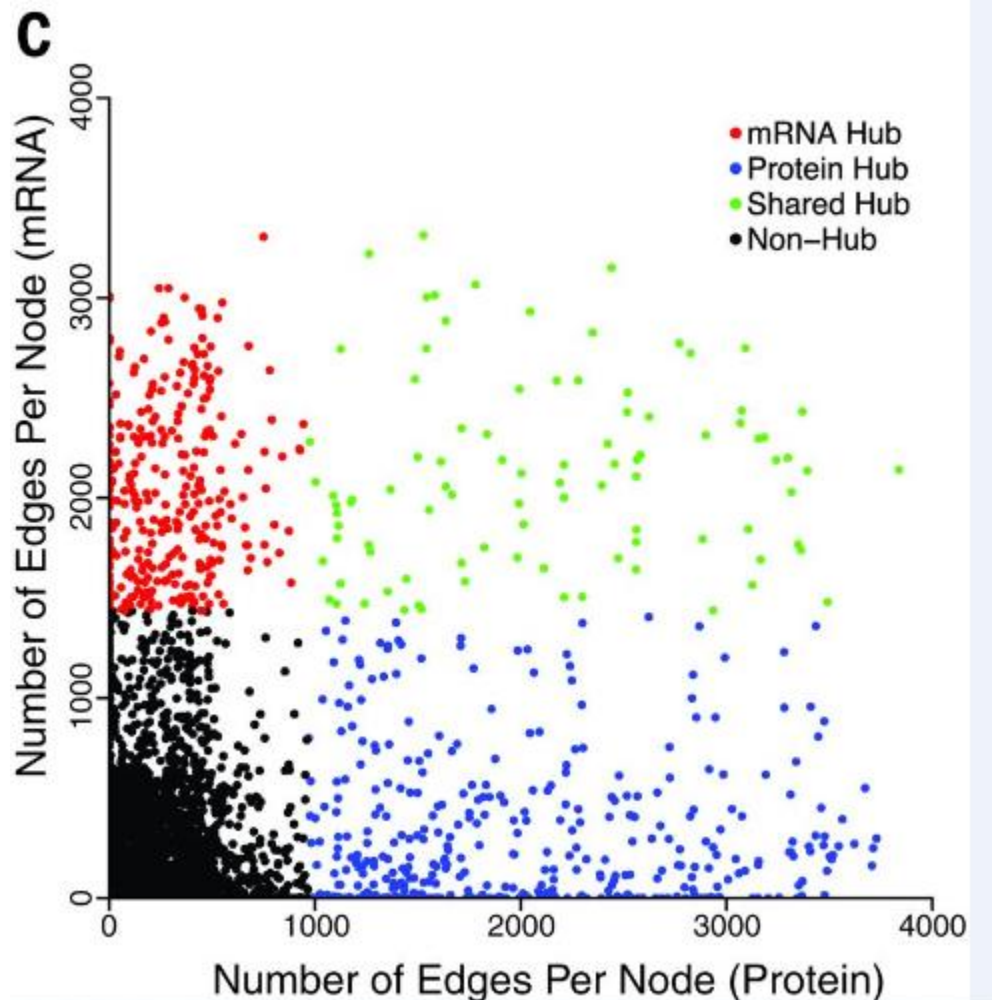


Fig B
Venn diagram depicting edge conservation (solid lines in Fig. 2A) between the two coexpression networks.

Coexpression network analyses.



- Categorize the hub genes as the nodes in the top 10th percentile for most edges
  - The majority（85%） were not shared between the mRNA and protein coexpression networks

Fig C
Number of edges a given gene (node) has in the protein (*x* axis) and mRNA (*y* axis) coexpression networks. Nodes above the 90th percentile for the number of edges are considered hubs and are colored according to whether they are hubs in the protein (blue) or mRNA (red) network or both (green). Black dots represent nonhub nodes.

Categorical enrichment analysis of coexpression modules.

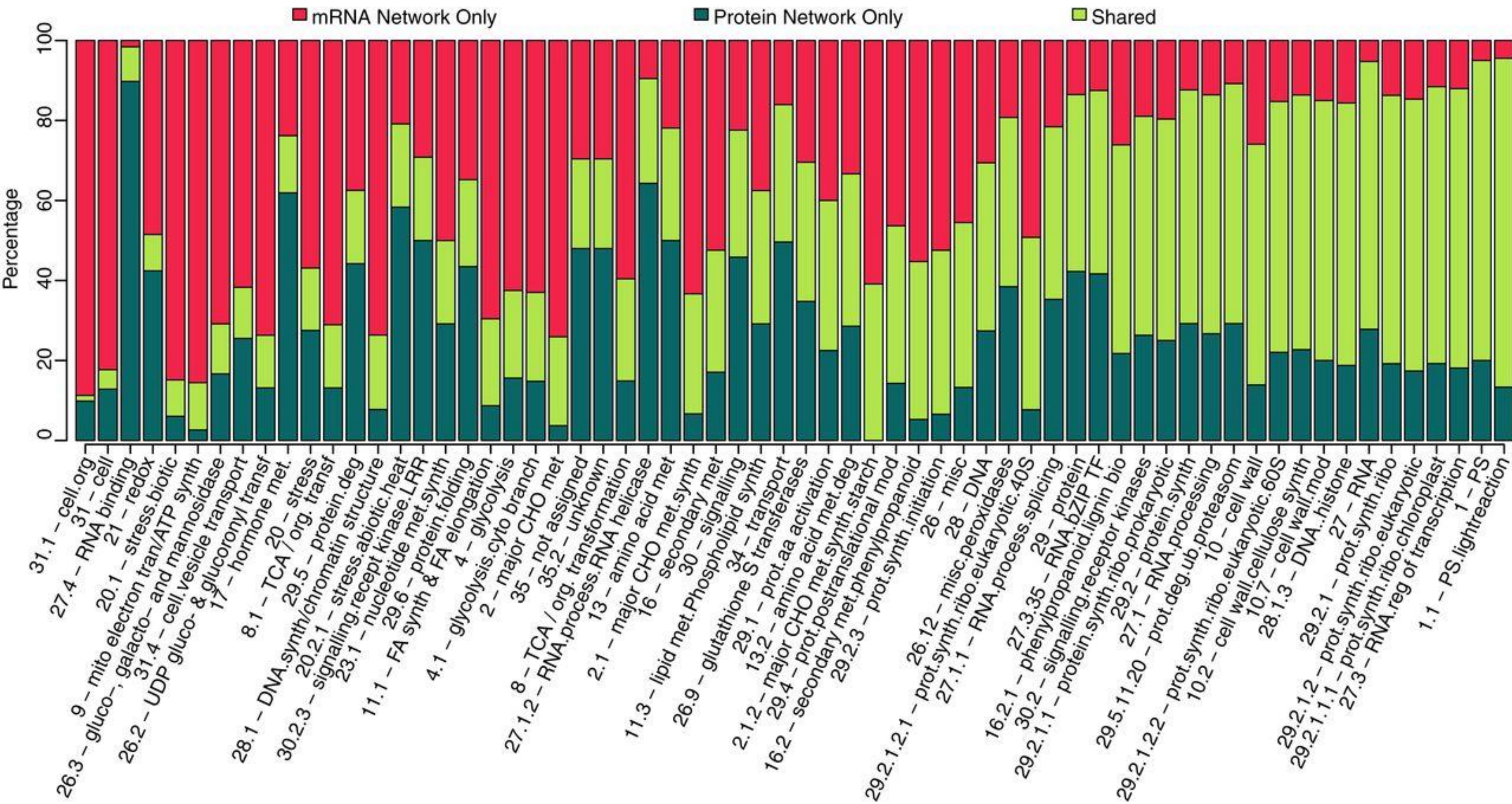


Fig
Coexpression modules were determined by WGCNA and functionally annotated using MapMan categories. Categories enriched (Benjamini-Hochberg adjusted *P* value ≤ 0.05) in one or more modules are represented by vertical bars and labeled with the bin number and name. For each category, the genes accounting for the enrichment were extracted separately from mRNA and protein modules. Only functional categories with at least 20 genes are shown. Colored bars represent the proportion of genes in each enriched category that are specific to one network (mRNA, red; protein, blue) or shared between the networks (green).

- 35% protein-specific, 27% mRNA-specific, and 38% shared

Coexpression network analyses.

- that transcript- and protein-based coexpression networks yield differing predictions of gene relatedness and function

the discrepancy between transcriptome and proteome coexpression networks

- the limited correlation between mRNA and protein abundance,
    - differing stabilities of mRNA and protein,
    - translational control
    - protein movement from the tissue of synthesis

GRNs

- Directed networks of TFs and their target genes

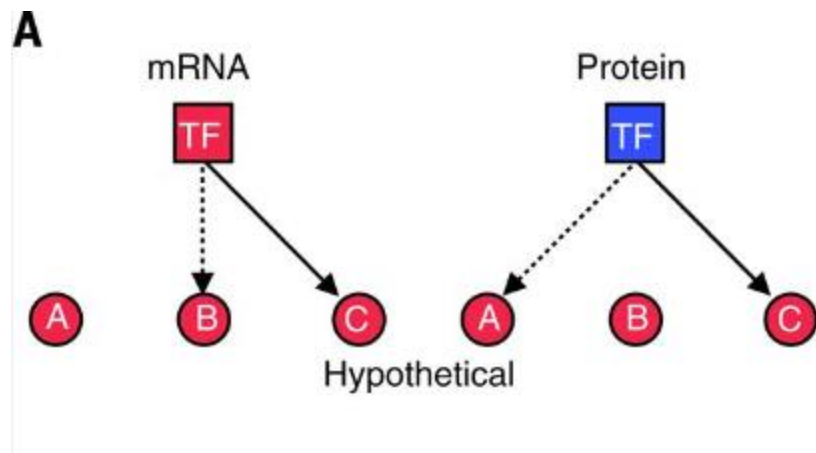


Fig A
Hypothetical GRN subnetwork depicting a TF regulator (square) and potential target genes (circle) quantified as mRNA (red) or protein (blue). GRN-specific and -conserved predictions are depicted by dotted and solid lines, respectively.

method
GENIE3 (an algorithm for the inference of GRN from expression data)

- random forest machine learning algorithm
- DREAM4 and -5 GRN econstruction challenges

GRN analyses.

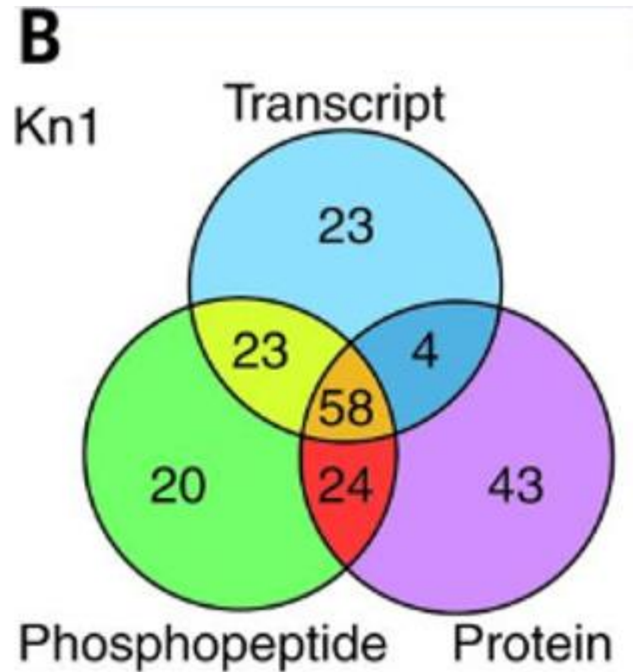


Fig B
Overlap of the true-positive predictions from the top 500 true GRN predictions for KN1 quantified as mRNA, protein, or phosphopeptide. True KN1 targets were identified by Bolduc *et al*.

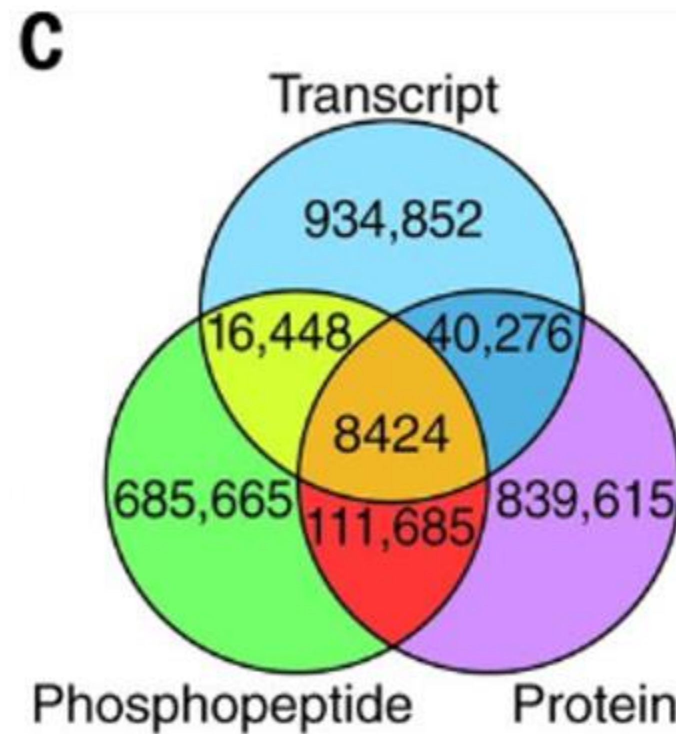
Benchmarks:

the homeobox TF KN1
the bZIP TF Opaque2

44% of all corrected targets were specific to a single type of GRN

GRN analyses.

- there was low edge conservation between the GRNs , with the vast majority of edges being present in a single GRN.
  - Considering one million edges, 93% were present in a single GRN
- the different accumulation patterns of mRNA, protein, and phosphorylation for a given TF result in disparate GRN predictions



Further validation

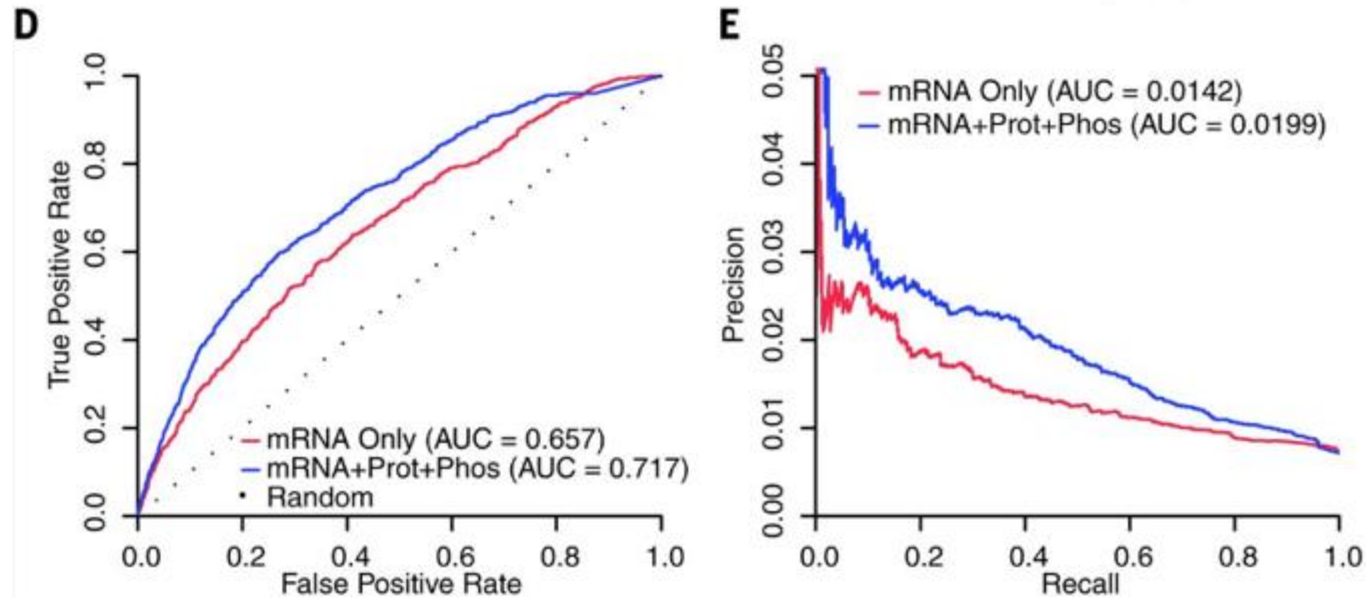
- Used 539 TFs regulators quantified as both mRNAs and proteins to reconstruct GRNs
- Different maize varieties (Mo17,B73)

Combine Multiple GRNs

To consolidate two or three networks, a new network was generated using the union of all TF expression data from the single networks as regulator inputs into the network and the same set of 41,021 target transcripts. This results in a network with redundancy at the gene level for TFs regulators that were quantified with multiple data types. To alleviate this redundancy, and obtain a combined score for each TF-Target edge, the product of all redundant edges was taken. When only a single edge existed (i.e. theTF was only quantified in one data type) when combining two data types, the square of the edge score was taken. For the final combined network consisting of all three data types, if only one edge was present ,the edge score was cubed. If two edges were present, the product of the two edges was multiplied by the average of the two edges.

For the phosphorylation data, the networks were constructed using phosphopeptide quantification but when combined, all phosphopeptides from a given protein were averaged in order to get phosphoproteinlevel information.

Evaluate the GRNs



ROC curves and precision-recall curves generated using known Kn1 and O2 target genes for a mRNA-only GRN (red) and a fully integrated GRN built by combining mRNA, protein, and phosphoprotein data into a single GRN (blue).

- Our comparison of transcriptome- to proteomebased dendrograms and coexpression networks showed little overlap at the gene level, even though the samples were classified similarly and had similar ontological enrichments.
- The discovery that most protein-expressing genes are conserved and syntenic also was unexpected
- Our findings highlight the importance of studying gene regulation at multiple levels.

My summary

- 此研究在各个角度说明转录组和蛋白组数据之间的低重合率，也进一步分析了这种现象的原因。然后又在各个角度证明各组学数据构建的网络分析得到的结果也是不完全一致的，具有互补性。十分具体的展现了不同组学整合的必要性。
- 然而此研究并未得到突破性的认识，后期应展示利用整合网络探索一些被遗漏的功能基因。
- 通过这篇文献，我对网络分析有了基本的认识。对于课题后期的数据深度分析有了简单的思路。在获取了蛋白组的数据后，参照此文献运用网络分析的方法来寻找功能基因。