

Meta-analysis of available transcriptomic datasets reveals core pathways in the high-light stress response of *Arabidopsis thaliana*

Aleksandr V. Bobrovskikh^{1,2*} (avb@bionet.nsc.ru), Ulyana S. Zubairova^{1,2}, Alexey V. Doroshkov^{1,2}

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090, Russia

² Novosibirsk State University, Novosibirsk, 630090, Russia

INTRODUCTION & AIM

Big data to big insights. The large amount of transcriptome data on stress responses now enables the identification of conserved core mechanisms in plants [1];

From model to crops. *Arabidopsis thaliana* (Col-0) serves as a key model of plant genetics, with the largest available high-light stress-induced transcriptome datasets, and many mechanisms are expected to be species-conserved. This data needs to be analyzed integratively.

Introducing AraLightDEGs. We created a new integrated resource AraLightDEGs (www.sysbio.ru/aralightdegs) compiling 58 high-light (HL) conditions from 21 transcriptomic studies, revealing **differentially expressed genes (DEGs)** in photosynthetic tissue's response to various high-light condition;

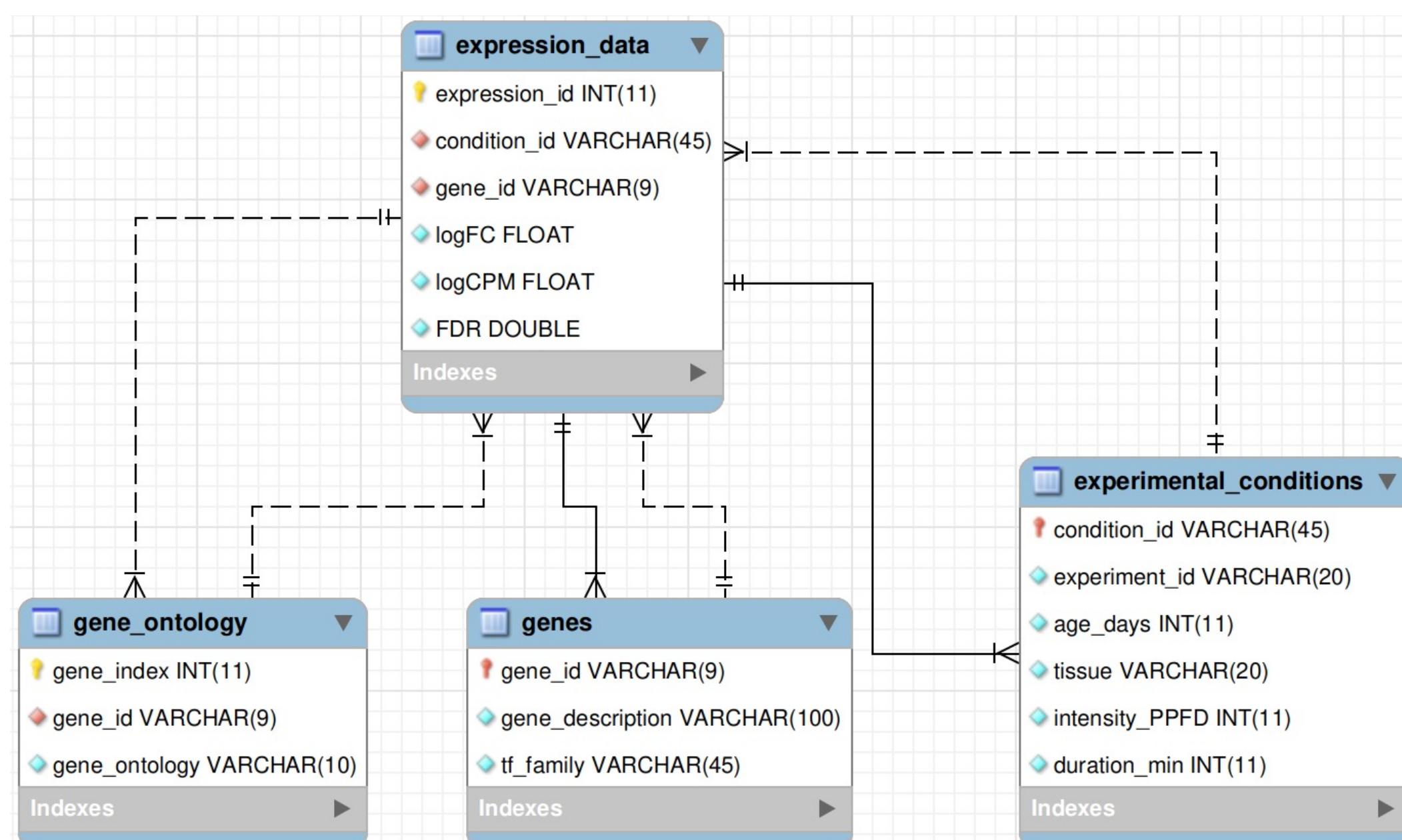
Breeding applications. Decoding such stress-induced gene regulatory networks can accelerate the development of stress-resistant crop varieties;

In our **case study**, we demonstrated the utility of AraLightDEGs in uncovering core pathways of *A. thaliana* high-light response.

METHODS

Data collection ▸ **Alignment, quantification, quality control** ▸ **DEGs**
NCBI SRA HISAT2, featureCounts, RNA-SeQC 2 edgeR

AraLightDEGs creation: MariaDB, MySQL Workbench, Flask, HTML, JS

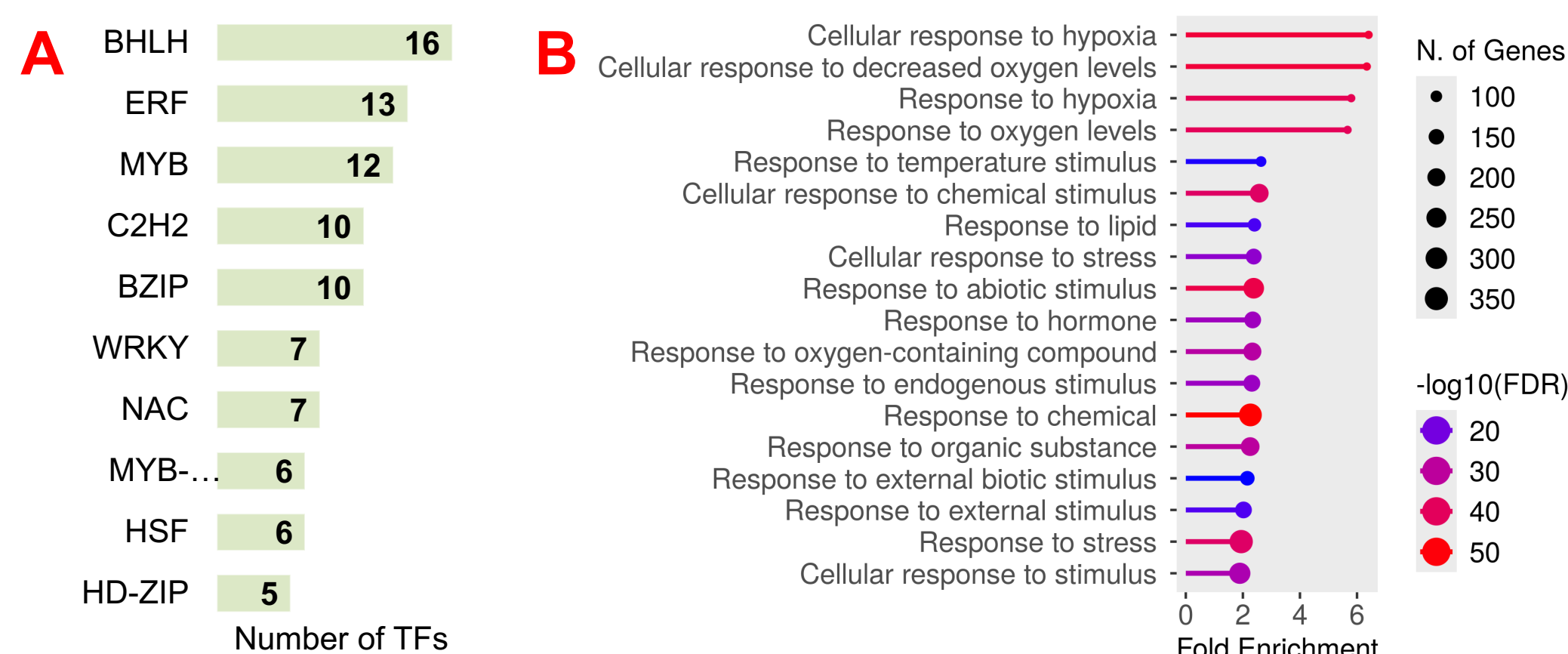


AraLightDEGs, EER diagram. Database includes 4 tables:

1. **expression_data** (218,000 DEGs in individual conditions);
2. **gene_ontology** contains Gene Ontology terms to identified DEGs;
3. **genes** contains descriptions of DEGs and information about transcription factors;
4. **experimental_conditions** contains details about high-light conditions (intensity, exposure time) and the plants characteristics (age, tissue).

RESULTS & DISCUSSION

Case study: we extracted a set of frequently occurring genes and TFs (in at least 50% of HL experiments) from AraLightDEGs. The composition of frequently occurring TFs is shown in Figure A; Figure B shows the annotations in GeneOntology terms describing this set of genes and TFs; Section C is a summary of the specific processes associated with the regulation patterns of these genes; and Figure D illustrates the core gene regulatory network of high-light responsive TFs (according to String-db).



The edges were extracted from String-db and visualized in Cytoscape. We used four interaction sources: textmining, experiments, databases, co-expression and medium confidence threshold (0.4).

CONCLUSION

Core genes of HL response. We identified 1,019 robust DEGs, of which 117 are TFs from bHLH, ERF, MYB, C2H2, bZIP and other families.

Functional enrichment: Upregulation of stress-responsive genes, ROS and high-light responsive genes, cellular defense processes. Downregulation of various metabolic and biosynthetic processes.

Our results **add important details** (uncircled nodes in Figure D) of the high-light stress response mechanisms in plants and can be used for further experimental validation.

FUTURE WORK / REFERENCES

1. Bobrovskikh, A. V., Zubairova, U. S., Bondar, E. I., Lavrekha, V. V., Doroshkov, A. V. (2022). Transcriptomic data meta-analysis sheds light on high light response in *Arabidopsis thaliana* L. International Journal of Molecular Sciences, 23(8), 4455.
2. **Coming soon!** Bobrovskikh, A. V., Zubairova, U. S., Doroshkov, A. V. (summer 2025) Preliminary title: "Identification of Key Differentially Expressed Genes in *Arabidopsis thaliana* Under Short- and Long-Term High Light Stress"