Supplementary Information

Machine learning reveals the transcriptional regulatory network and circadian dynamics of *Synechococcus elongatus* PCC 7942

Authors: Yuan Yuan, Tahani Al Bulushi, Anand V. Sastry, Cigdem Sancar, Richard Szubin, Susan S. Golden, Bernhard O. Palsson

SI Figures

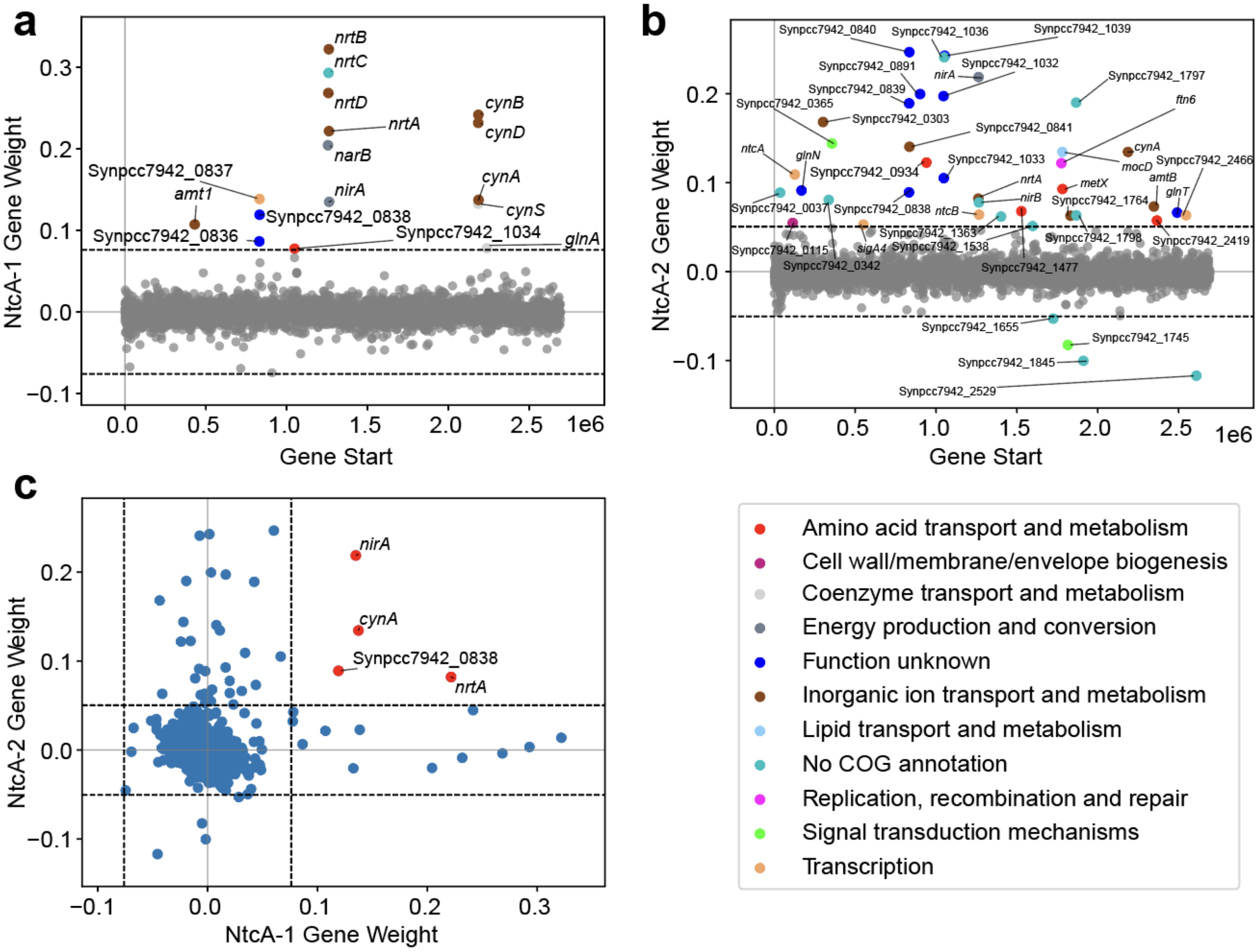

Fig. S1. The NtcA iModulons. (a) The gene weights plot for the NtcA-1 iModulon. (b) The gene weights plot for the NtcA-2 iModulon. (c) Four genes are shared between the NtcA-1 and NtcA-2 iModulons.

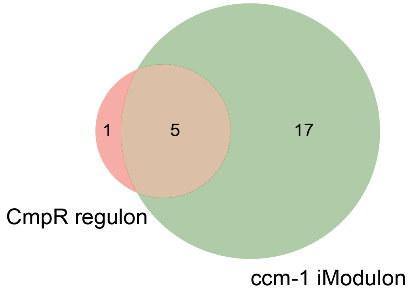

Fig. S2. The overlap between the ccm-1 iModulon and the CmpR regulon.

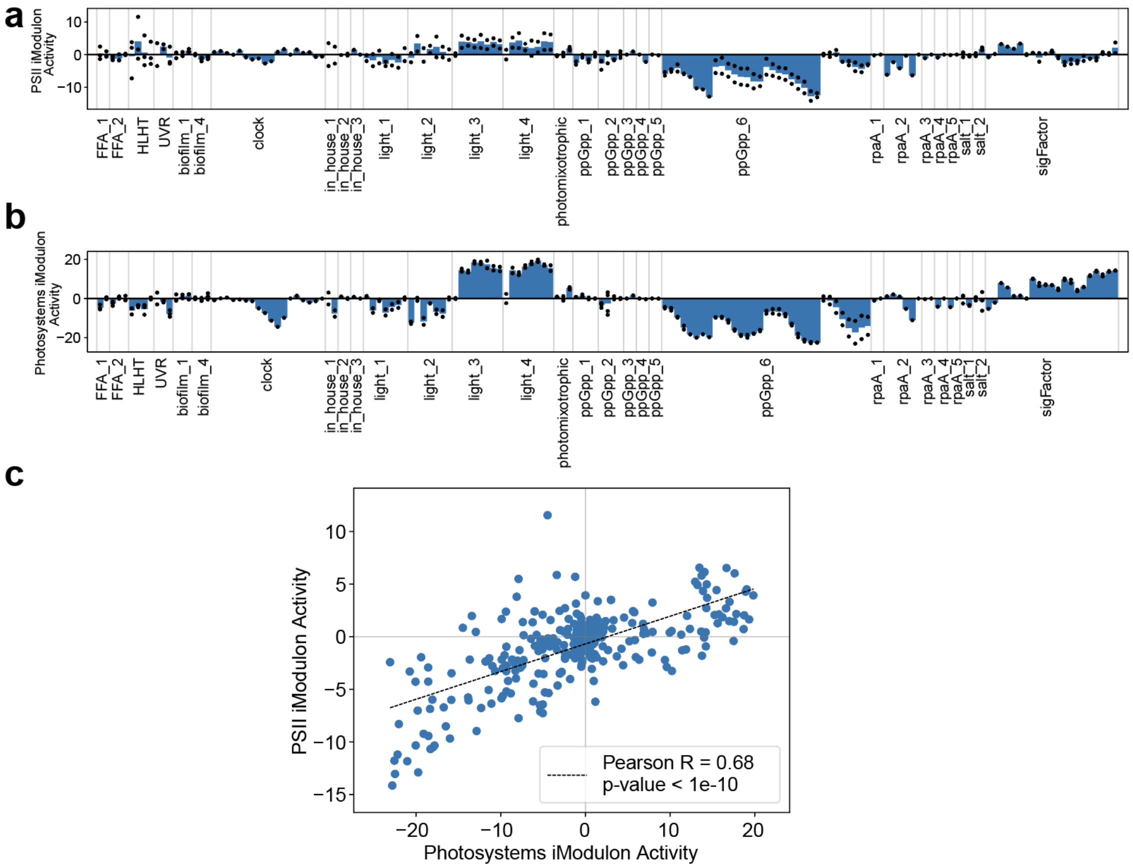

Fig. S3. The activity plots for the PSII and Photosystems iModulons. (a) Activity plot for the PSII iModulon across all the projects in the compendium. (b) Activity plot for the Photosystems iModulon across all the projects in the compendium. (c) Activity correlations between the two iModulons .

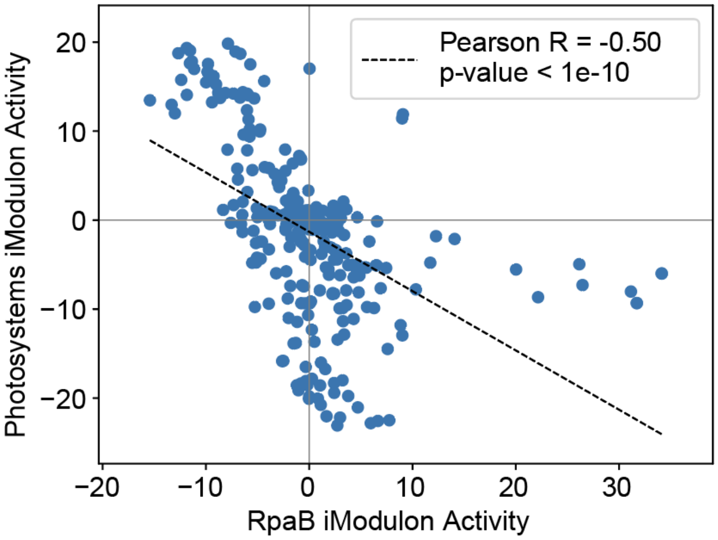

Fig. S4. Activity Correlation between the Photosystems and RpaB iModulons.

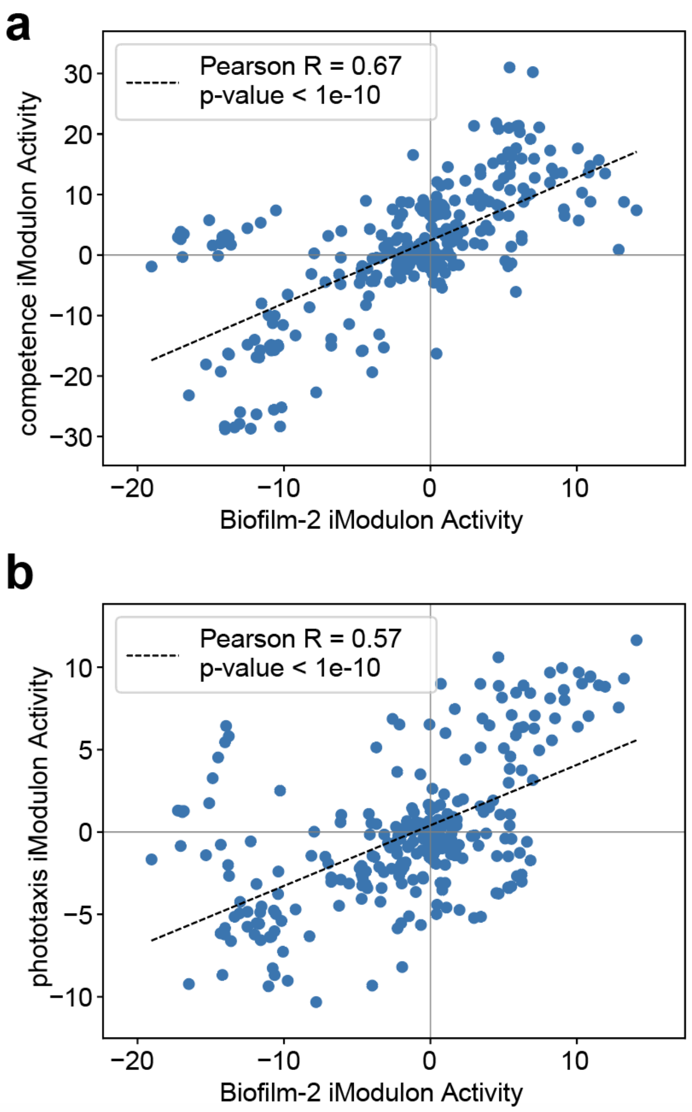

Fig. S5. Activity correlations between Biofilm-2 iModulon and two iModulons. (a) Correlation between the activity of the Biofilm-2 iModulon and the competence iModulon. (b) Correlation between the activity of the Biofilm-2 iModulon and the phototaxis iModulon.

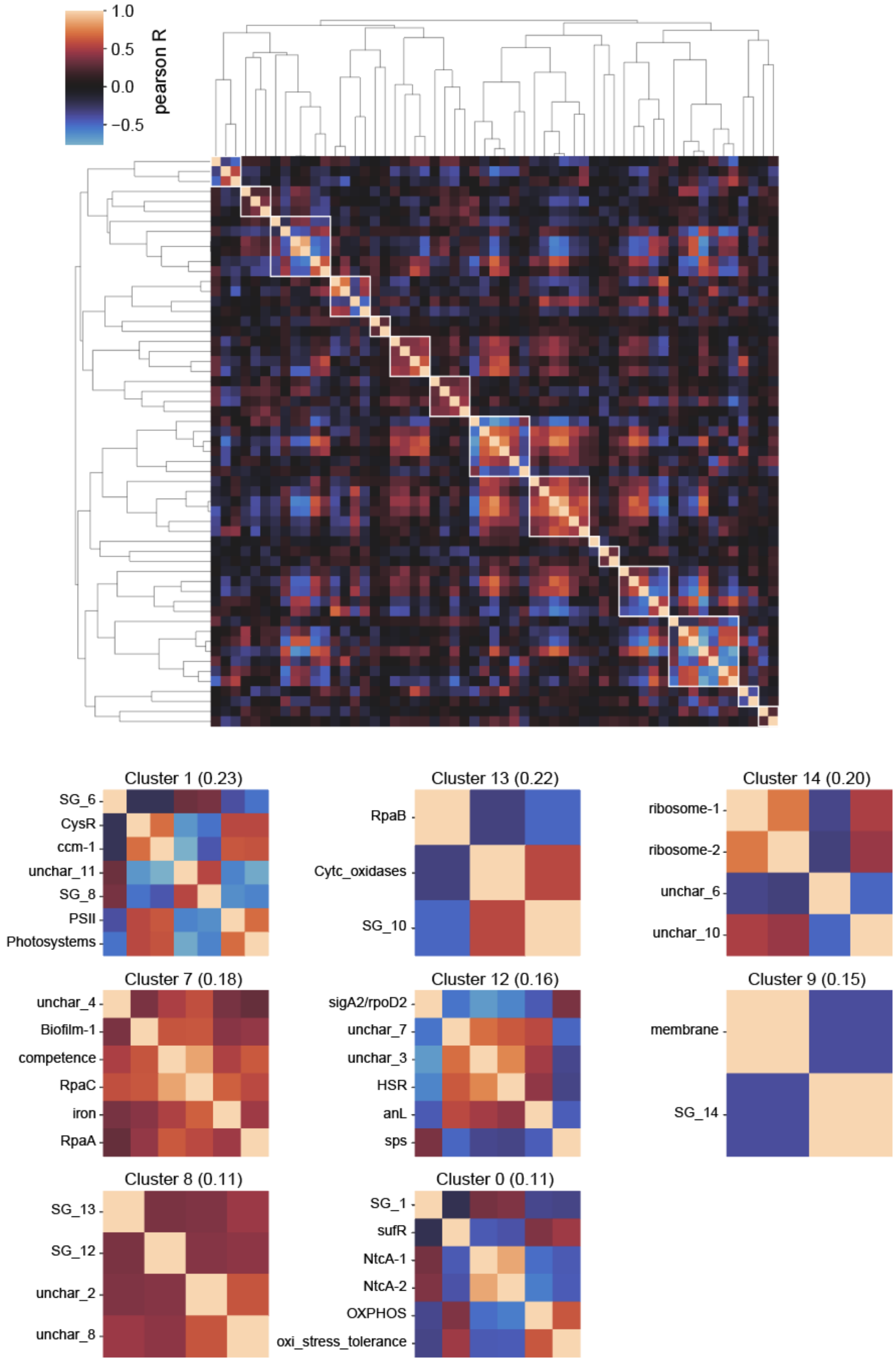

Fig. S6. Activity cluster plot and best clusters. The clusters are computed with the approach described in **SI Methods**.

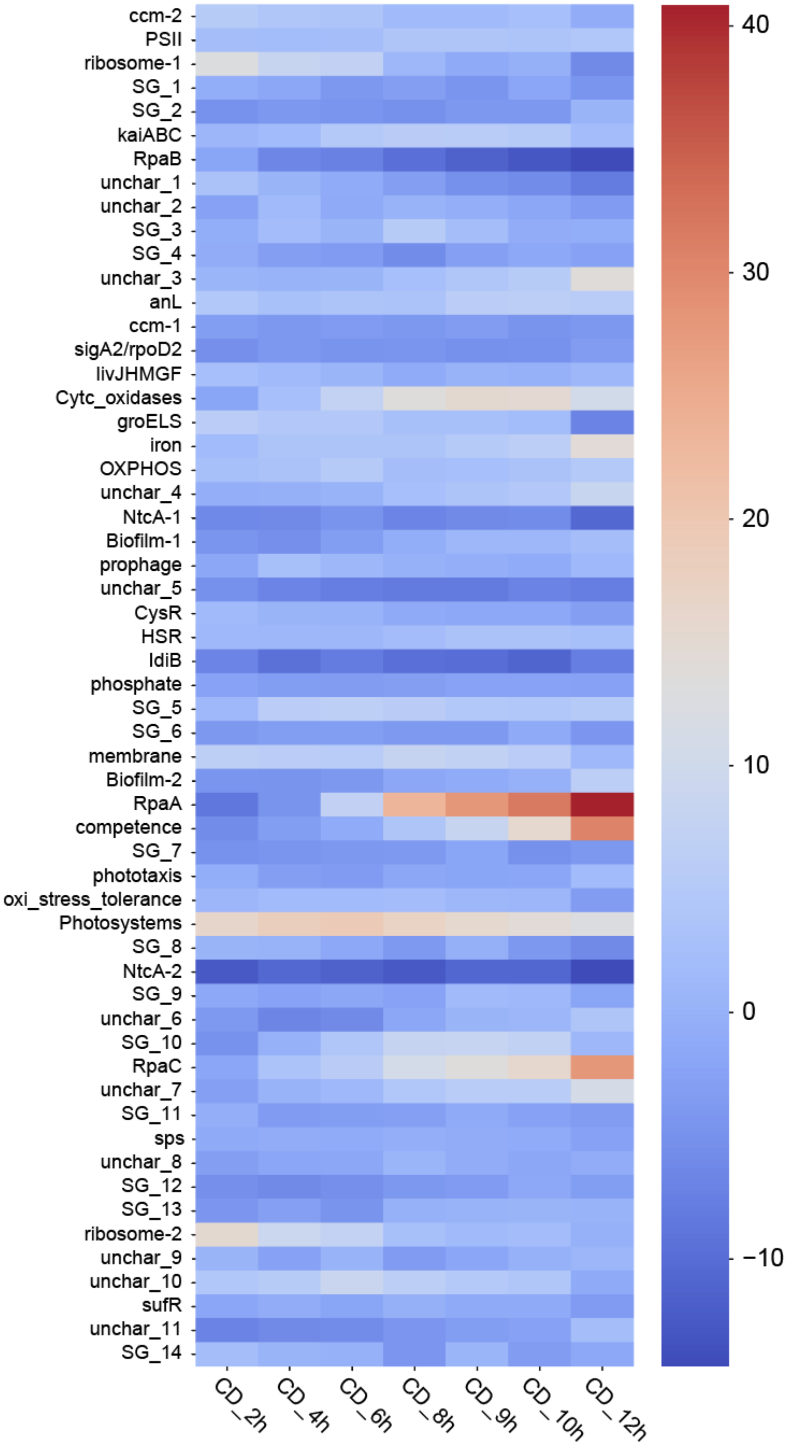

Fig. S7. Activity patterns of all iModulons under the “Clear Day” condition.

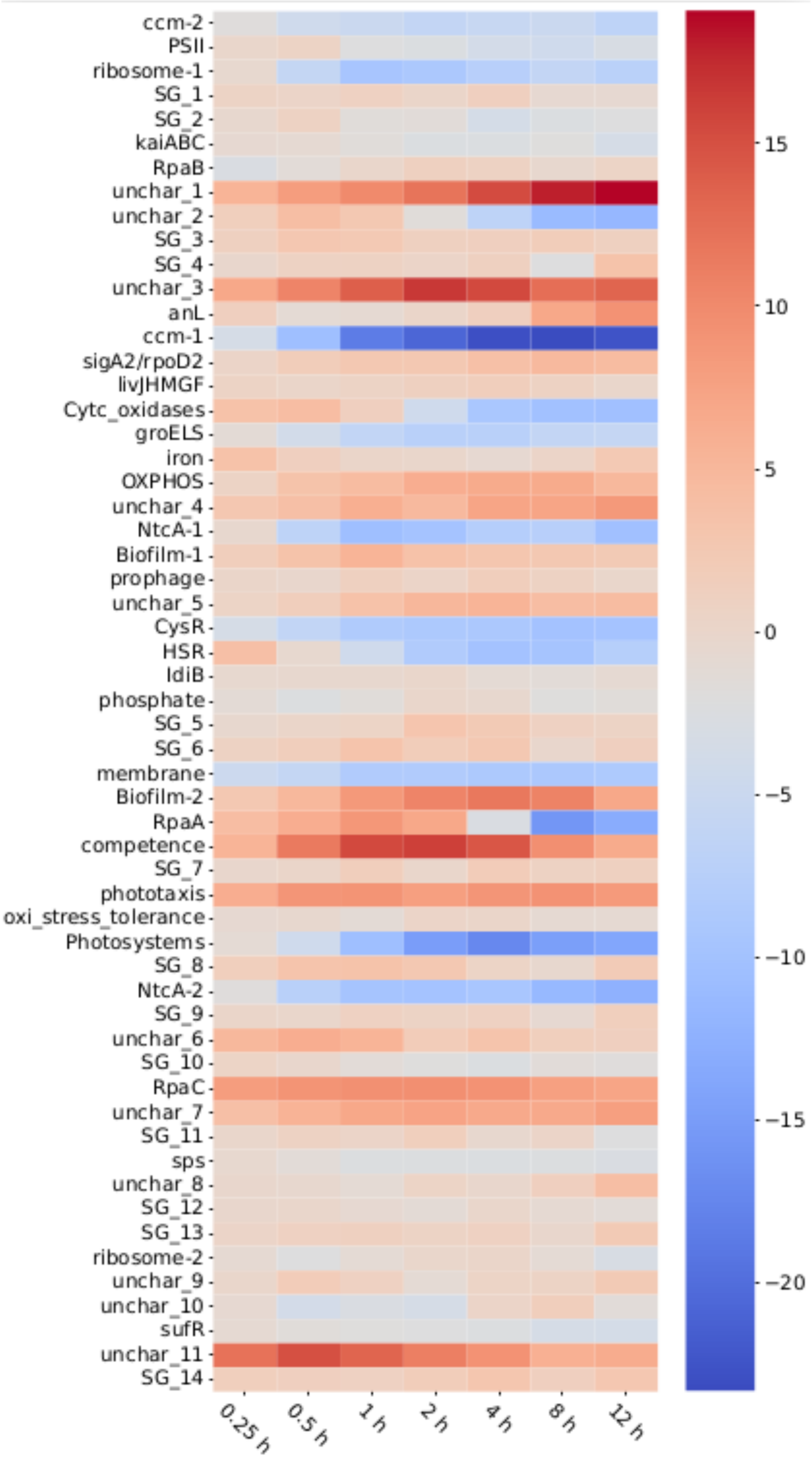

Fig. S8. Activity patterns of all iModulons at night.

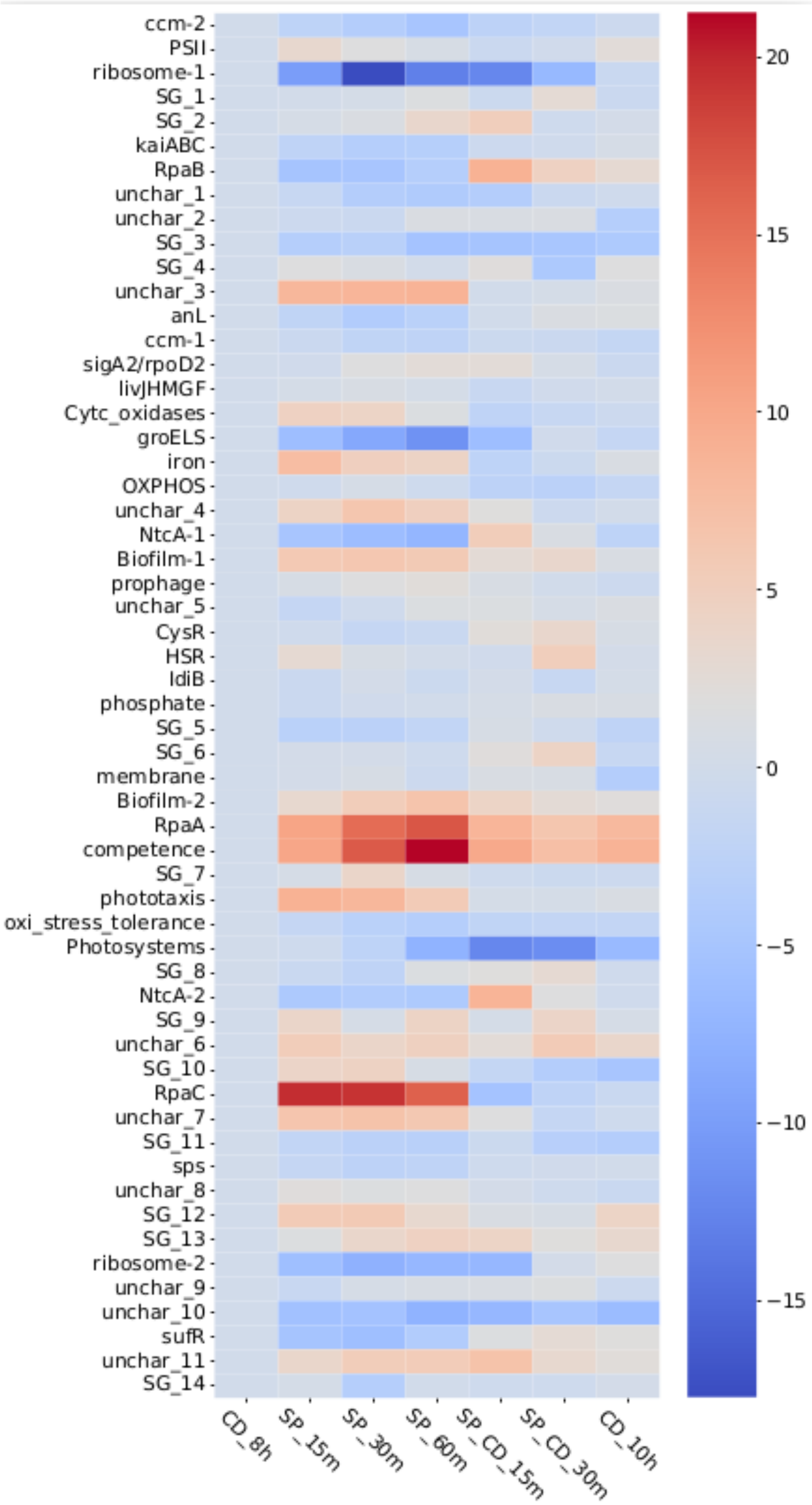

Fig. S9. Activity patterns of all iModulons under shade pulse.

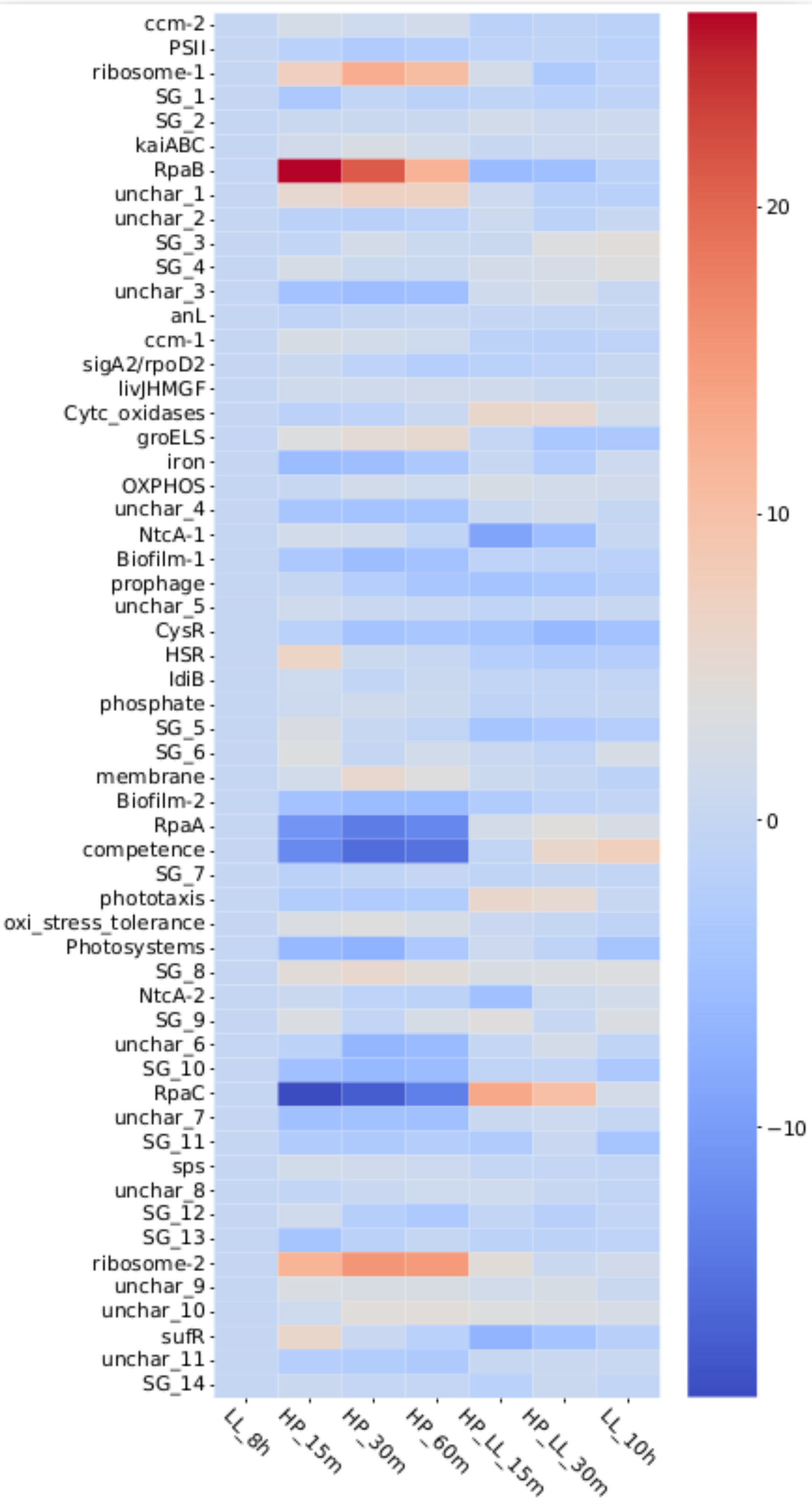

Fig. S10. Activity patterns of all iModulons under high light pulse.

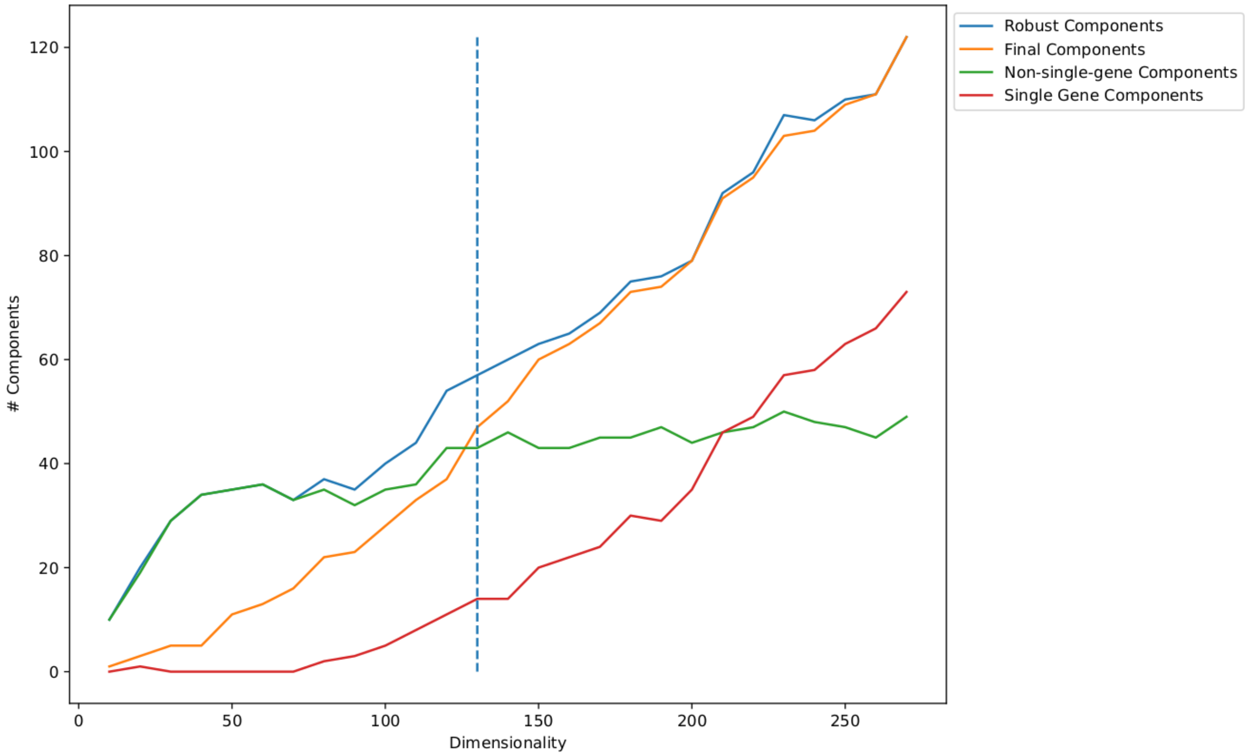

Fig. S11. Selection of the optimal dimensionality.

SI Datasets

Dataset S1. Gene weights for the RpaA iModulon.

Dataset S2. Gene weights for the Photosystems iModulon.

Dataset S3. Gene weights for the PSII iModulon.

SI Methods

Calculating iModulon activity clusters.

Global iModulon activity clustering was performed using the *clustermap* function from the Python Seaborn package[1]. The pairwise distance for each iModulon was calculated using Pearson R correlation as the distance metric. The optimal clustering threshold was determined by testing various distance thresholds and selecting the one that maximized the silhouette score.

SI References

1. Waskom, M. *et al.* mwaskom/seaborn: v0.8.1 (September 2017). Zenodo https://doi.org/10.5281/zenodo.883859 (2017).