

# Transcriptomics of microbial cultures, visualization, and GitHub

luke.thompson@noaa.gov  
@lukethompson

# Microbial Transcriptomics Example 1

## Gene expression patterns during light and dark infection of *Prochlorococcus* by cyanophage



### Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism

Luke R. Thompson<sup>a,1</sup>, Qinglu Zeng<sup>b,1</sup>, Libusha Kelly<sup>b</sup>, Katherine H. Huang<sup>b</sup>, Alexander U. Singer<sup>c</sup>, JoAnne Stubbe<sup>a,d,2</sup>, and Sallie W. Chisholm<sup>a,b,2</sup>

Departments of <sup>a</sup>Biology and <sup>b</sup>Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>c</sup>Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario Canada M5G 1L6; and <sup>d</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by James M. Tiedje, Michigan State University, East Lansing, MI, and approved July 29, 2011 (received for review February 15, 2011)

RESEARCH ARTICLE

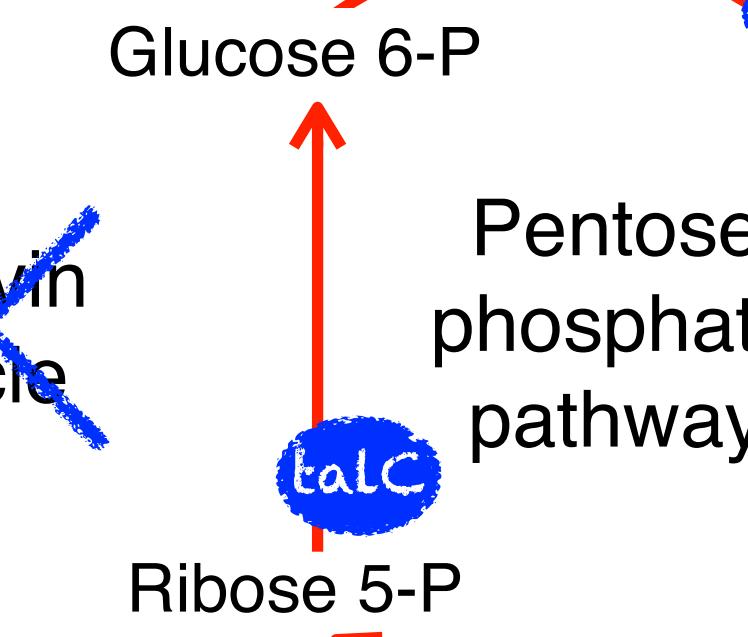
### Gene Expression Patterns during Light and Dark Infection of *Prochlorococcus* by Cyanophage

Luke R. Thompson<sup>1</sup><sup>✉a\*</sup>, Qinglu Zeng<sup>2</sup><sup>✉b\*</sup>, Sallie W. Chisholm<sup>1,2\*</sup>

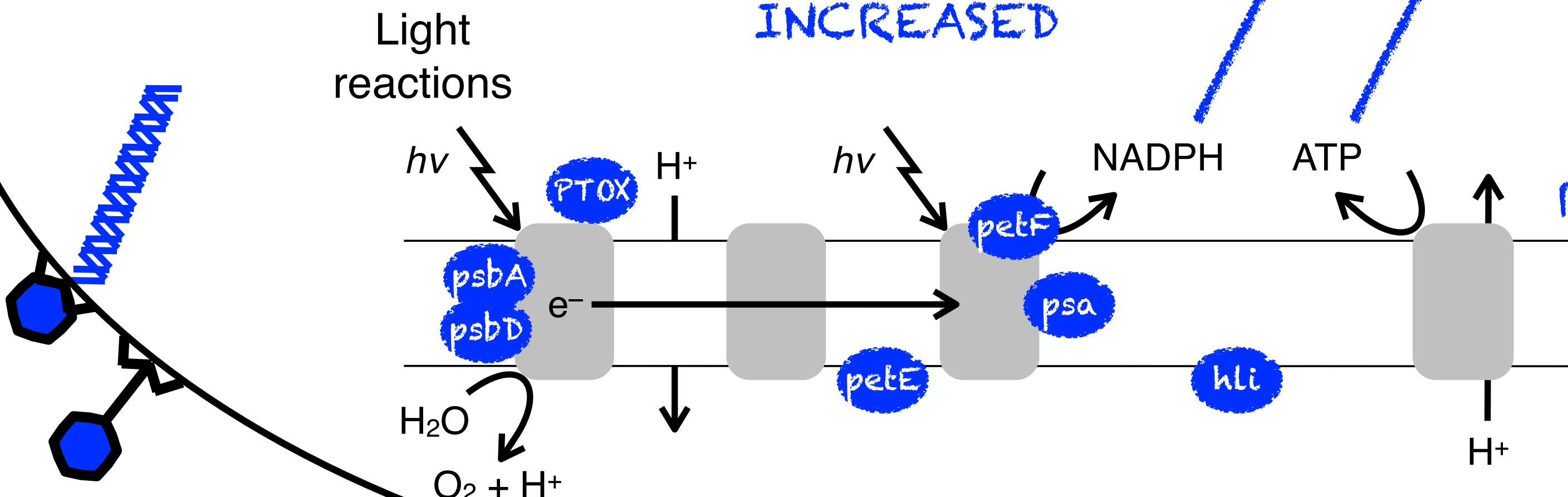
1 Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 2 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

# Infected host

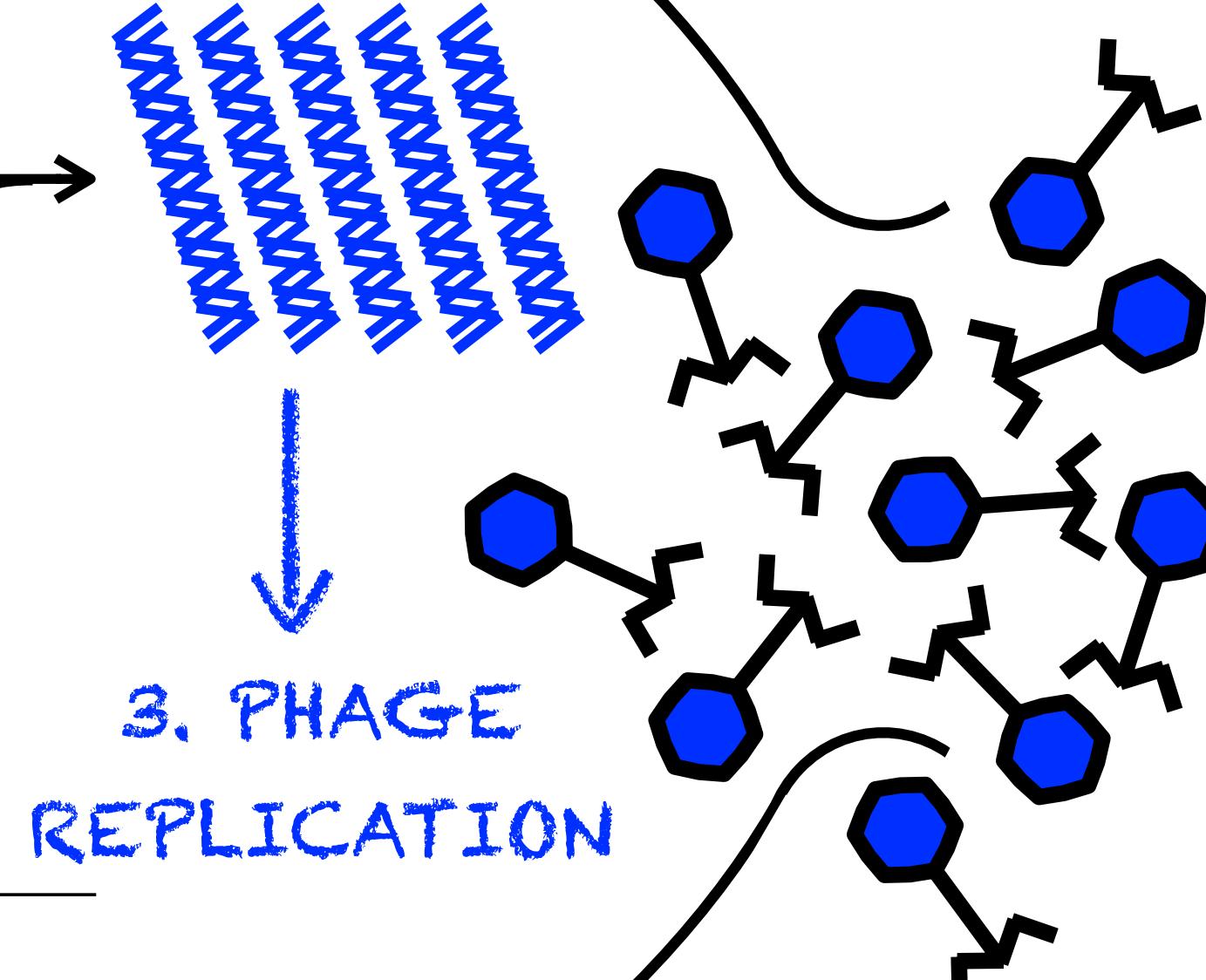
1. PHAGE GENES EXPRESSED



2. NADPH/NADP<sup>+</sup> INCREASED



3. PHAGE REPLICATION



Nucleotide biosynthesis

# Methods

## Mapping and counting RNA-Seq reads

- \* Map to reference genomes using the [BWA](#) (Burrows–Wheeler Aligner), generating SAM alignment files.
- \* Calculate the number of reads perfectly aligning to the sense and antisense strands of ORFs, rRNAs, tRNAs, and intergenic regions using [SAMtools](#) and [pysam](#).
- \* Paired reads were mapped separately; paired reads mapping to the same ORF were counted as one transcript for that ORF, whereas paired reads mapping to separate adjacent ORFs were counted once for each ORF.

## Normalization of phage and host transcript abundance

- \* Counts were normalized per sample using the [RPKM](#) method (reads per kbp gene length per million reads).
- \* Phage transcript counts were normalized to the total of phage plus host transcript counts. Host transcript counts were normalized to the total of host counts only.

## Clustering of genes by expression pattern

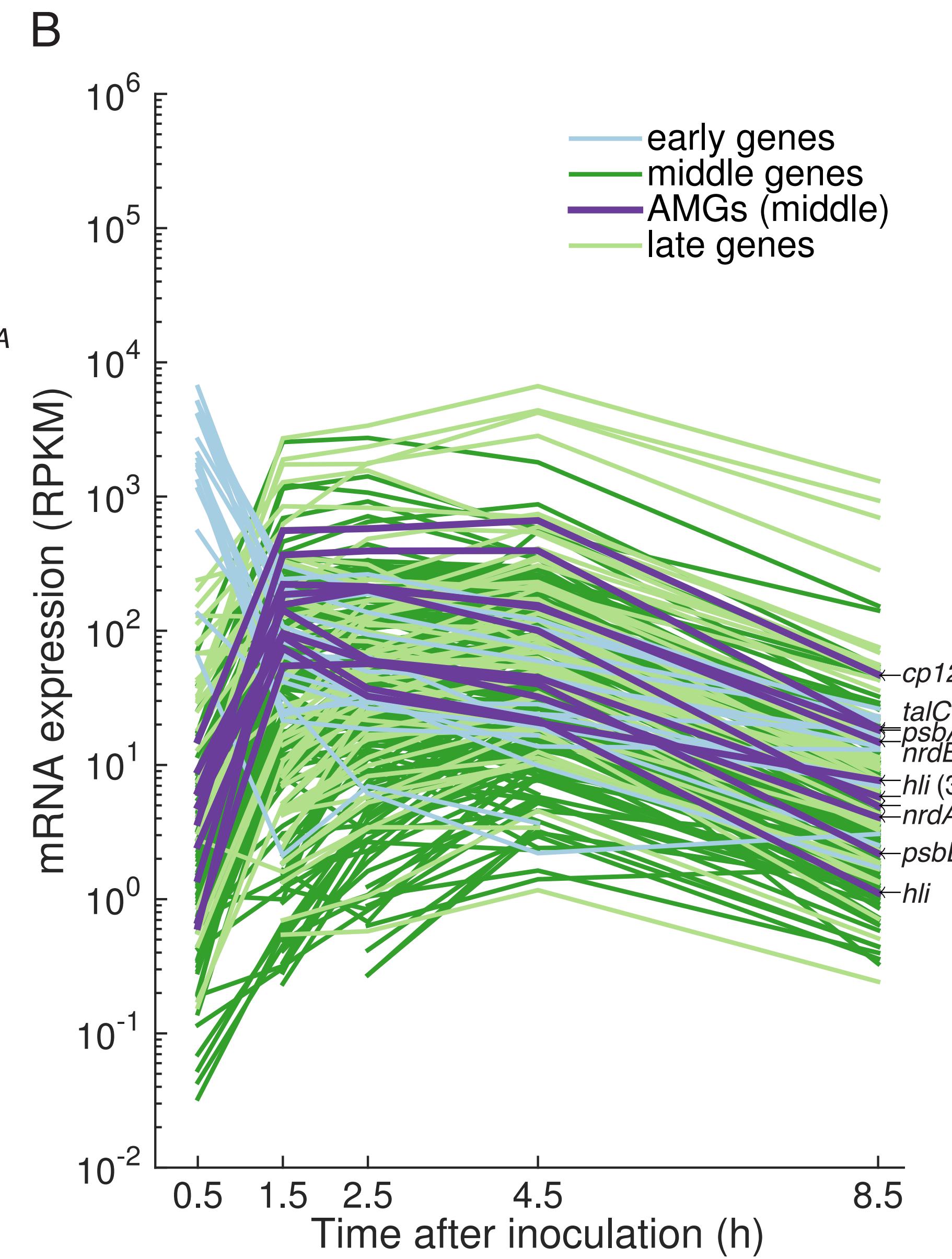
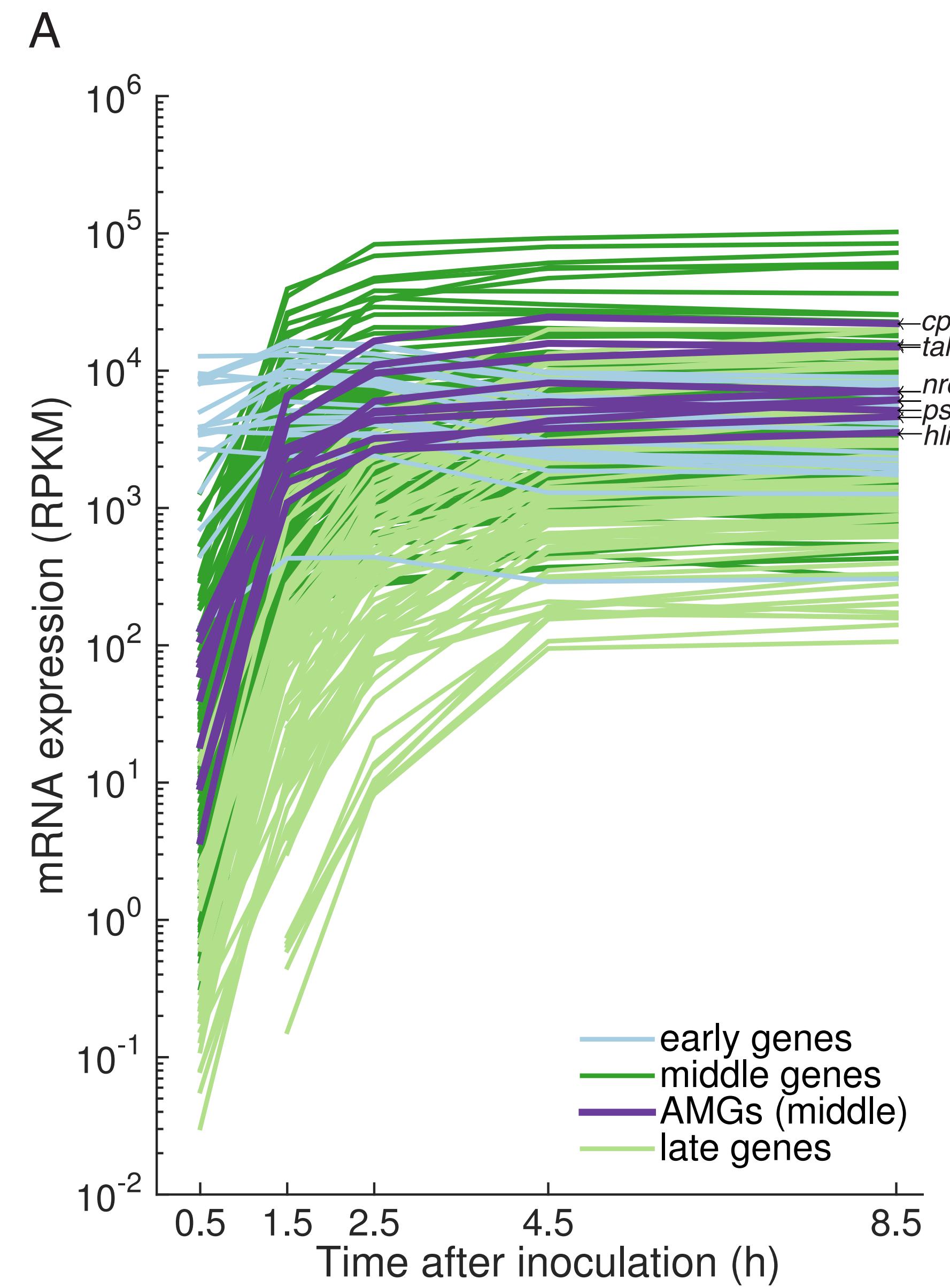
- \* Phage genes were clustered by transcript relative abundance patterns using two independent approaches: [partitioning around medoids](#) (PAM) and [hierarchical clustering](#).

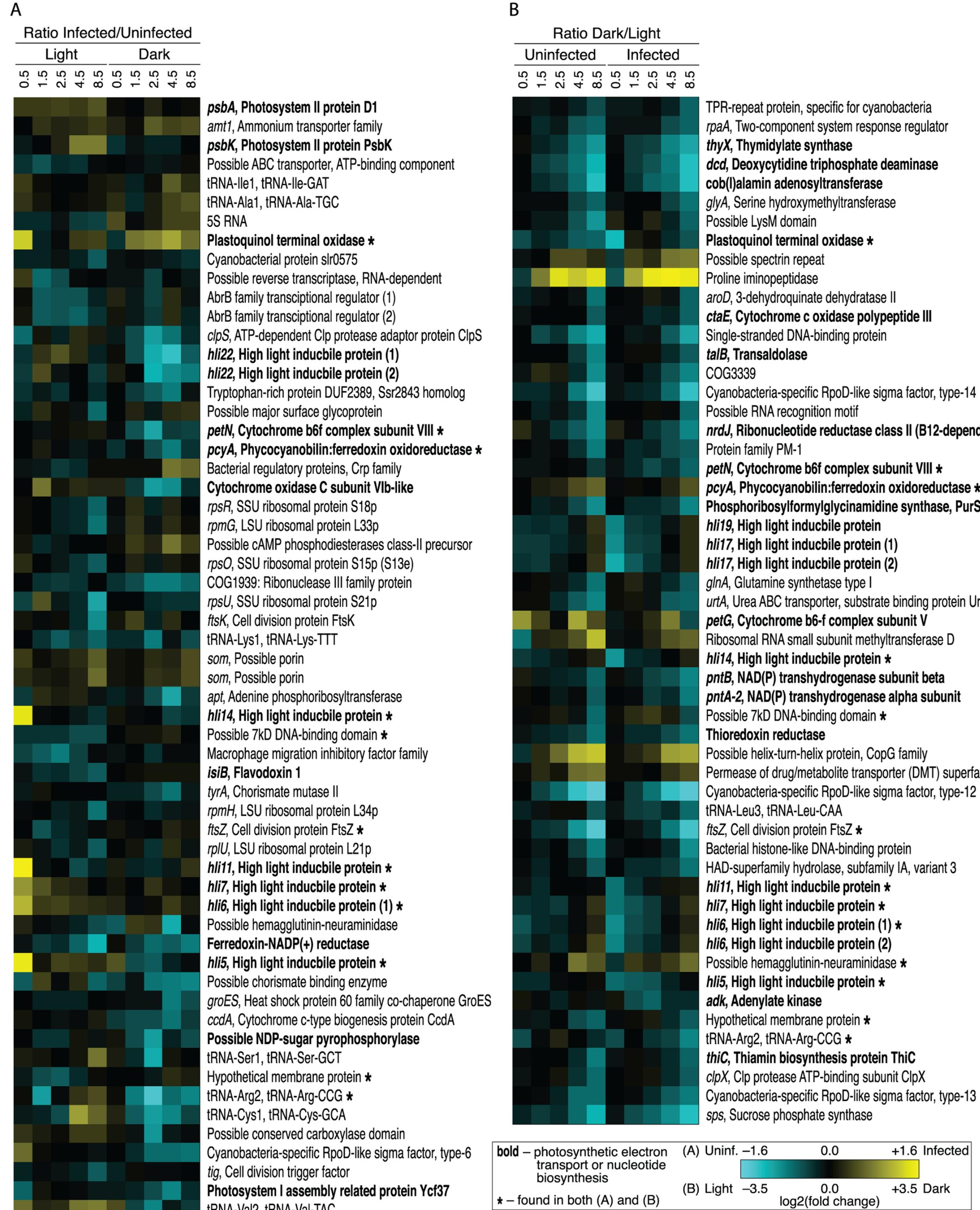
## Detection of differential gene expression

- \* Differentially expressed host transcripts (messenger RNAs and antisense RNAs) were identified using the R packages [DESeq2](#) and [NOISeq](#).
- \* Transcript abundances were analyzed at each timepoint separately (no proper t=0 control).
- \* Lists of differentially expressed genes (DEGs) were the intersection of lists derived from DESeq2 and NOISeq.

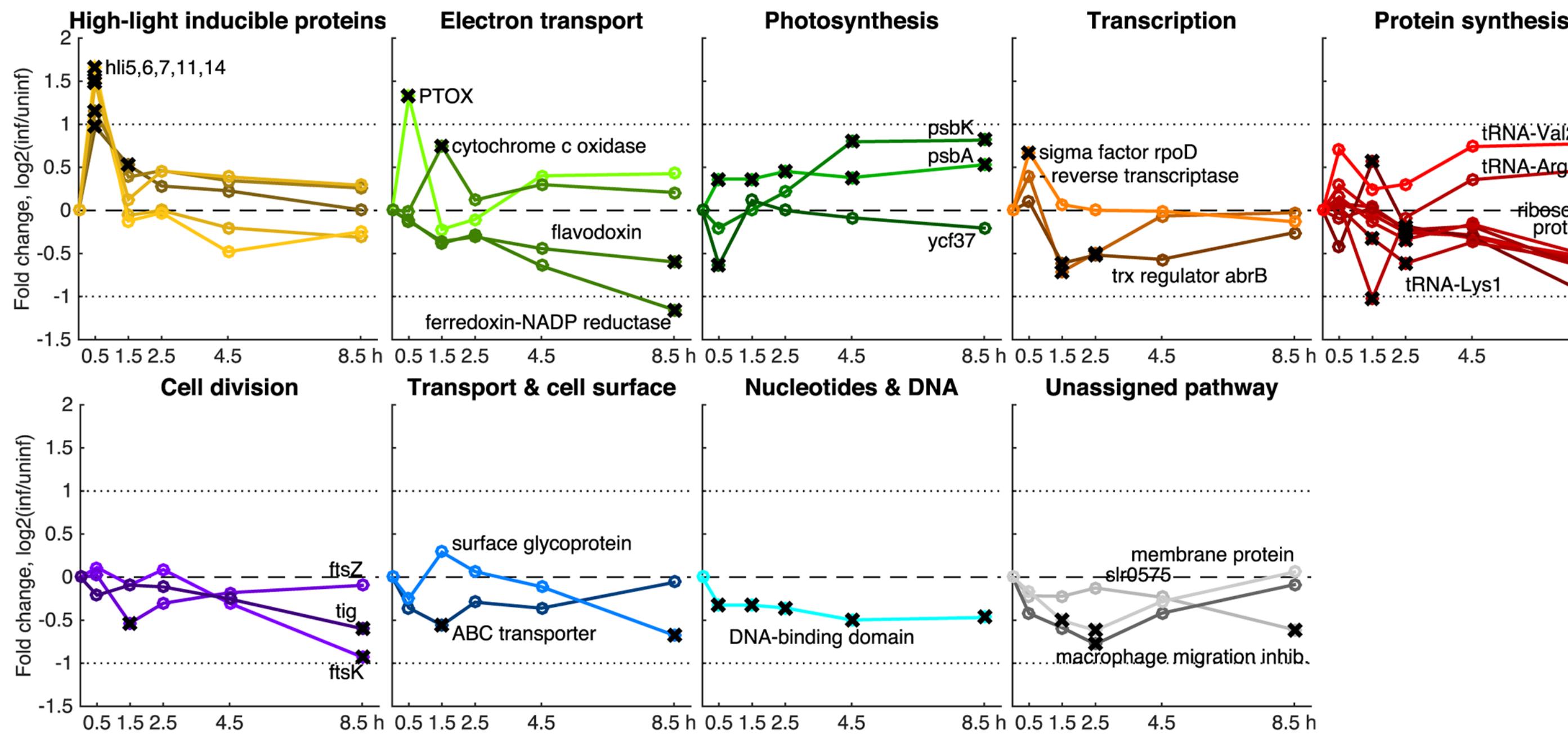
## Terminology of differential gene expression

- \* We avoided the terms “up-regulation” and “down-regulation” because we lack evidence of regulatory mechanisms and our data reflect relative but not absolute abundances. We instead favor the terms “enriched” and “diminished” in reference to relative transcript abundance.
- \* In some cases we used the common term “differentially expressed genes” (DEGs); we emphasize that “expression” in this sense refers to transcript abundance only and reflects the net result of transcription minus transcript degradation.

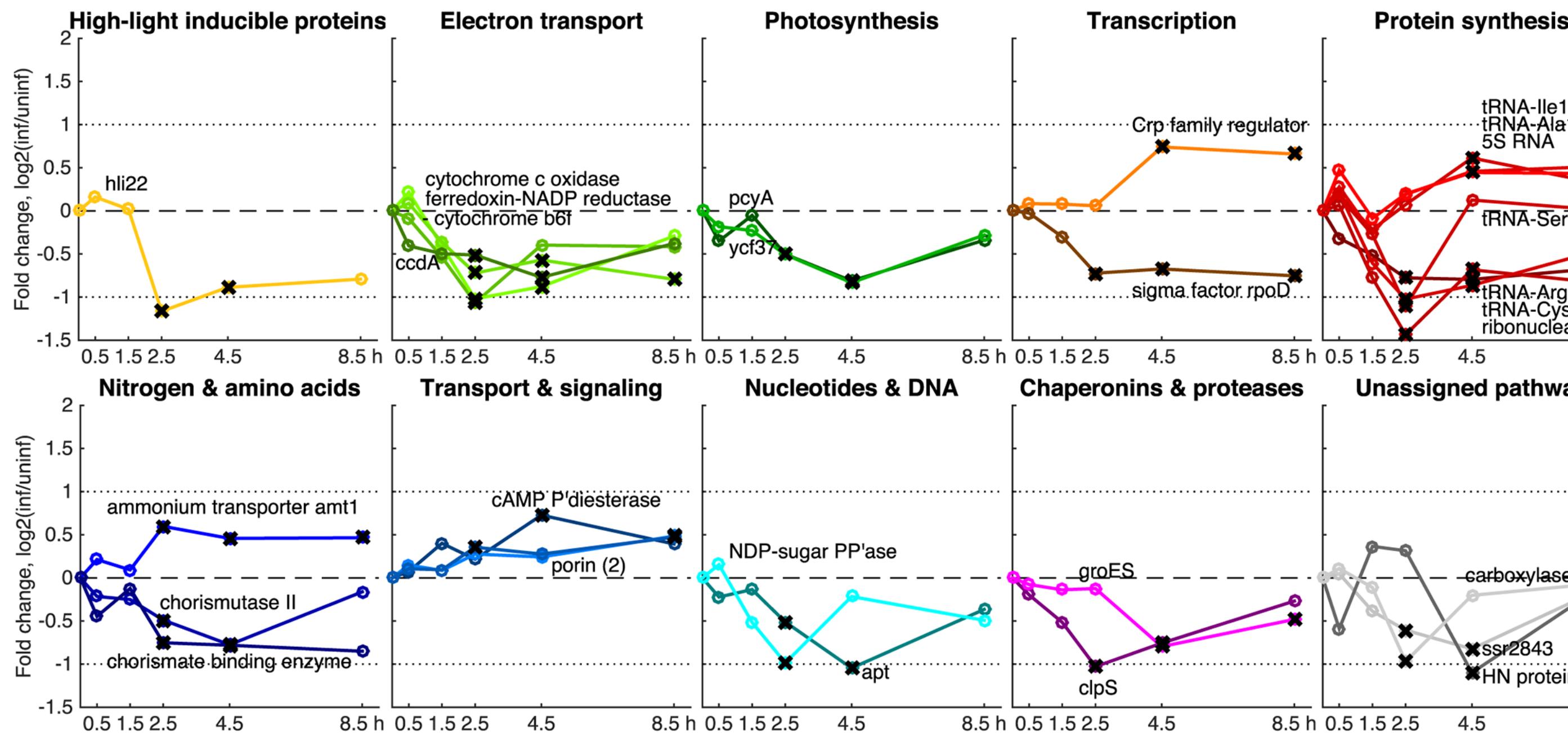


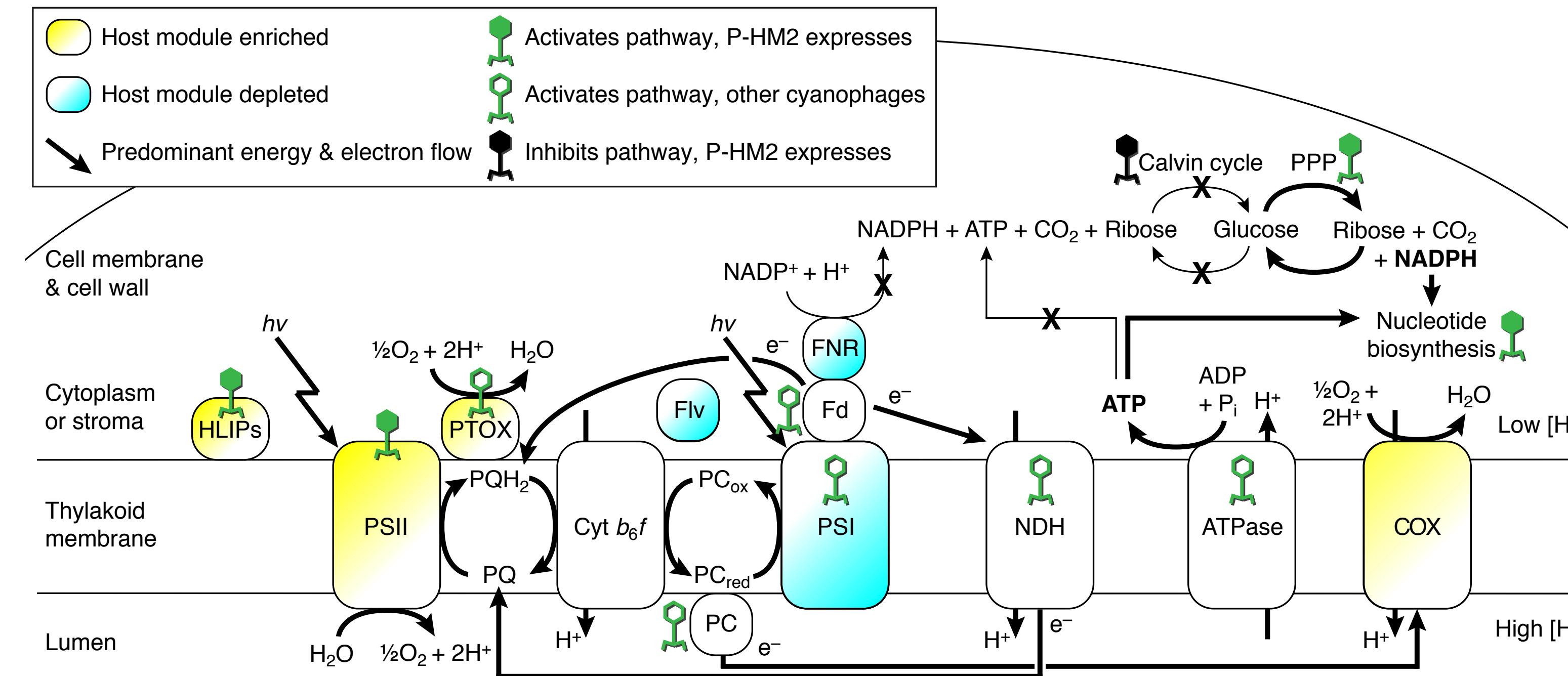
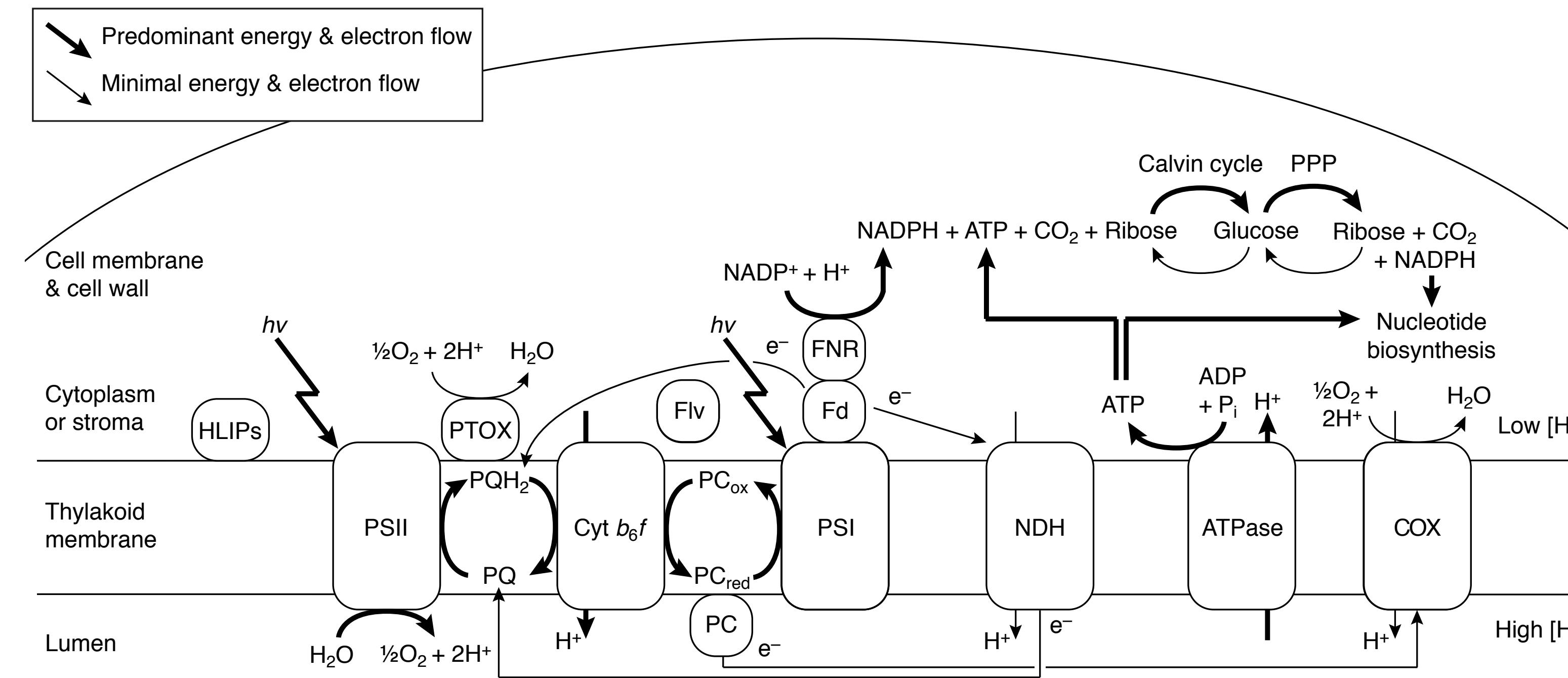


### A. Host differentially expressed genes – Light infection



### B. Host differentially expressed genes – Dark infection





# Microbial Transcriptomics Example 2

## Transcriptional characterization of *Vibrio fischeri* during colonization of juvenile *Euprymna scolopes*

**environmental  
microbiology**

Environmental Microbiology (2017) 00(00), 00–00 doi:10.1111/1462-2920.13684

**Transcriptional characterization of *Vibrio fischeri* during colonization of juvenile *Euprymna scolopes***

Luke R. Thompson,<sup>1†</sup> Kiel Nikolakakis,<sup>2†</sup> Shu Pan,<sup>2</sup> Jennifer Reed,<sup>2</sup> Rob Knight<sup>1</sup> and Edward G. Ruby<sup>2,3\*</sup>

<sup>1</sup>Department of Pediatrics, University of California, San Diego, CA, USA.

<sup>2</sup>Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI, USA.

<sup>3</sup>Pacific Biosciences Research Center, University of Hawaii, Manoa, HI, USA.

**Introduction**

Bacterial colonization of host organisms has been studied in many different model systems, and in the context of both pathogenesis and beneficial symbiosis (Bry *et al.*, 1996; Dedeine *et al.*, 2001; Russell and Rychlik, 2001; Hongoh, 2010; Gilbert *et al.*, 2012; Nyholm and Graf, 2012; Bulgarelli *et al.*, 2013; McFall-Ngai *et al.*, 2013; Almagro-Moreno *et al.*, 2015; Uzal *et al.*, 2015; Kao *et al.*,

# Methods

## Sequence read processing and mapping

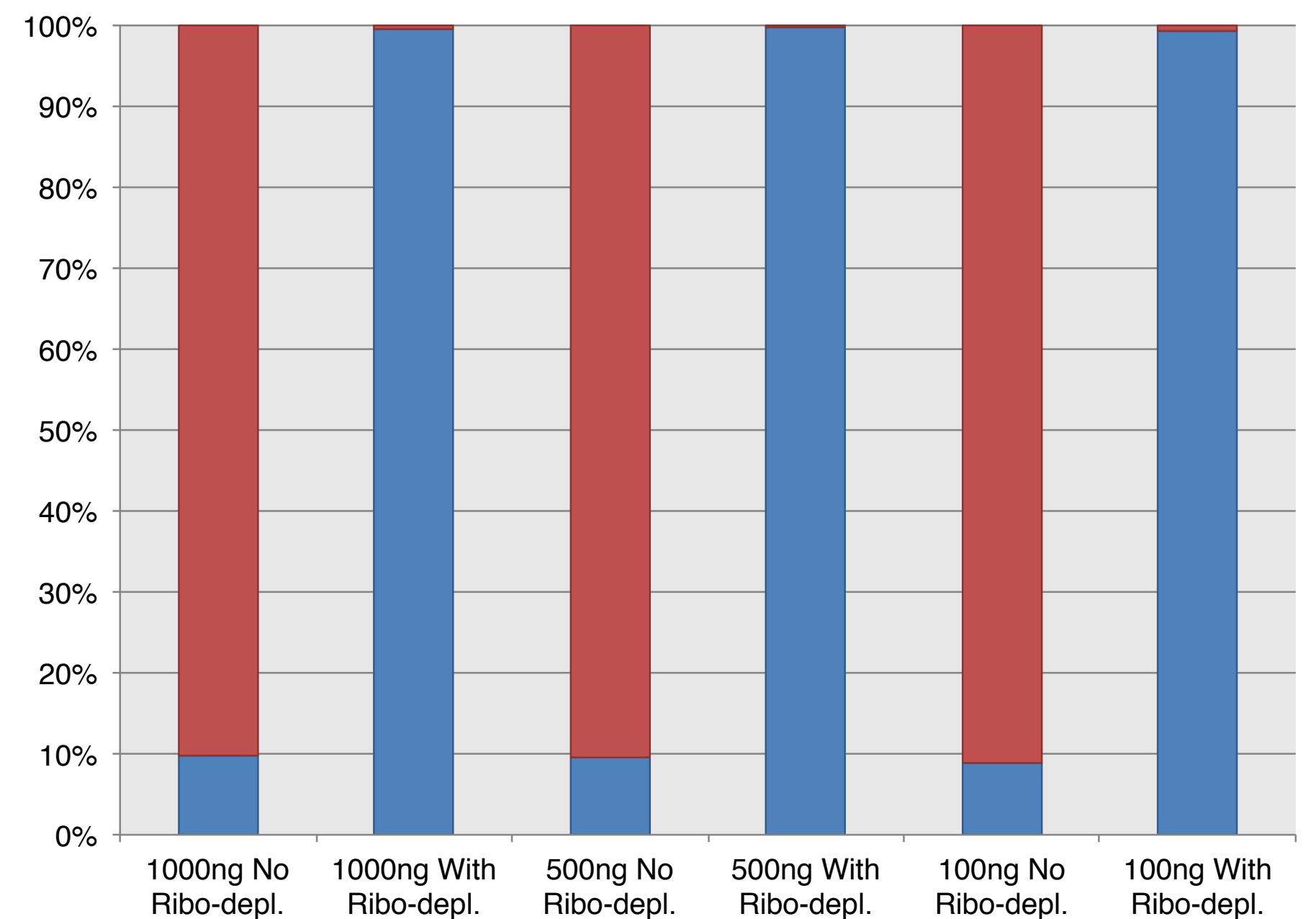
- \* Map to reference genomes using the BWA (Burrows–Wheeler Aligner), generating SAM alignment files, then processed with SAMtools.
- \* The numbers of reads mapping to protein-coding (CDS) or rRNA genes were calculated using the htseq-count command of HTSeq.

## Detection of differential gene expression

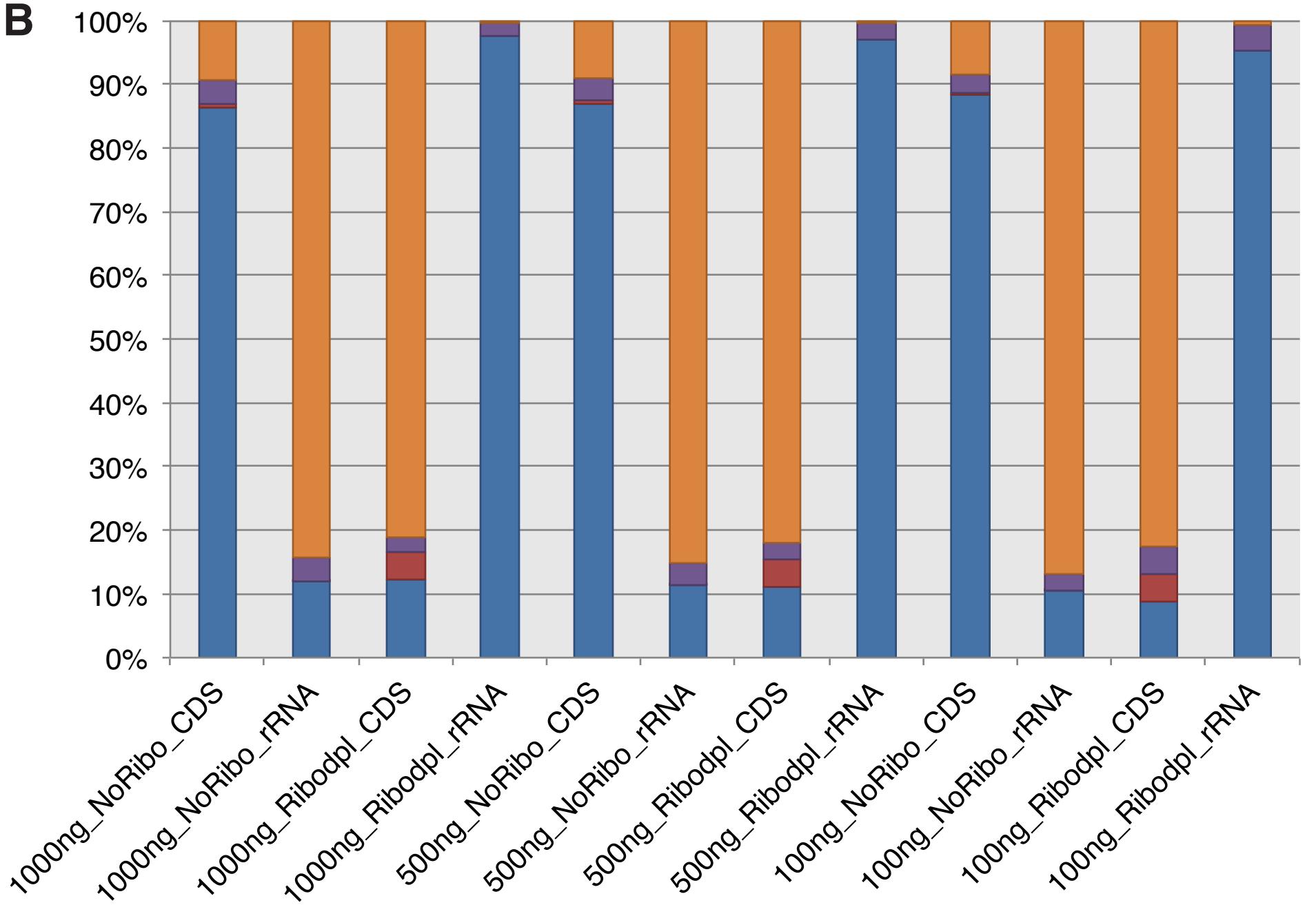
- \* Differentially expressed host transcripts (messenger RNAs and antisense RNAs) were identified using the R packages DESeq2 and NOISeq.
- \* NOISeq was used to filter low counts.
- \* DESeq2 was used to detect differential expression. The three conditions (squid-associated, planktonic, and cultured) were contrasted pairwise for all genes, and results exported as Benjamini–Hochberg adjusted p-values and log<sub>2</sub>(fold change). Differentially expressed genes were identified using tiered cutoffs of these values, with the most stringent cutoff being an adjusted p-value < 0.001 and abs(log<sub>2</sub>(fold change)) > 3.0 (three replicates per condition).

## RNA-Seq from ribo-depleted, low-biomass samples

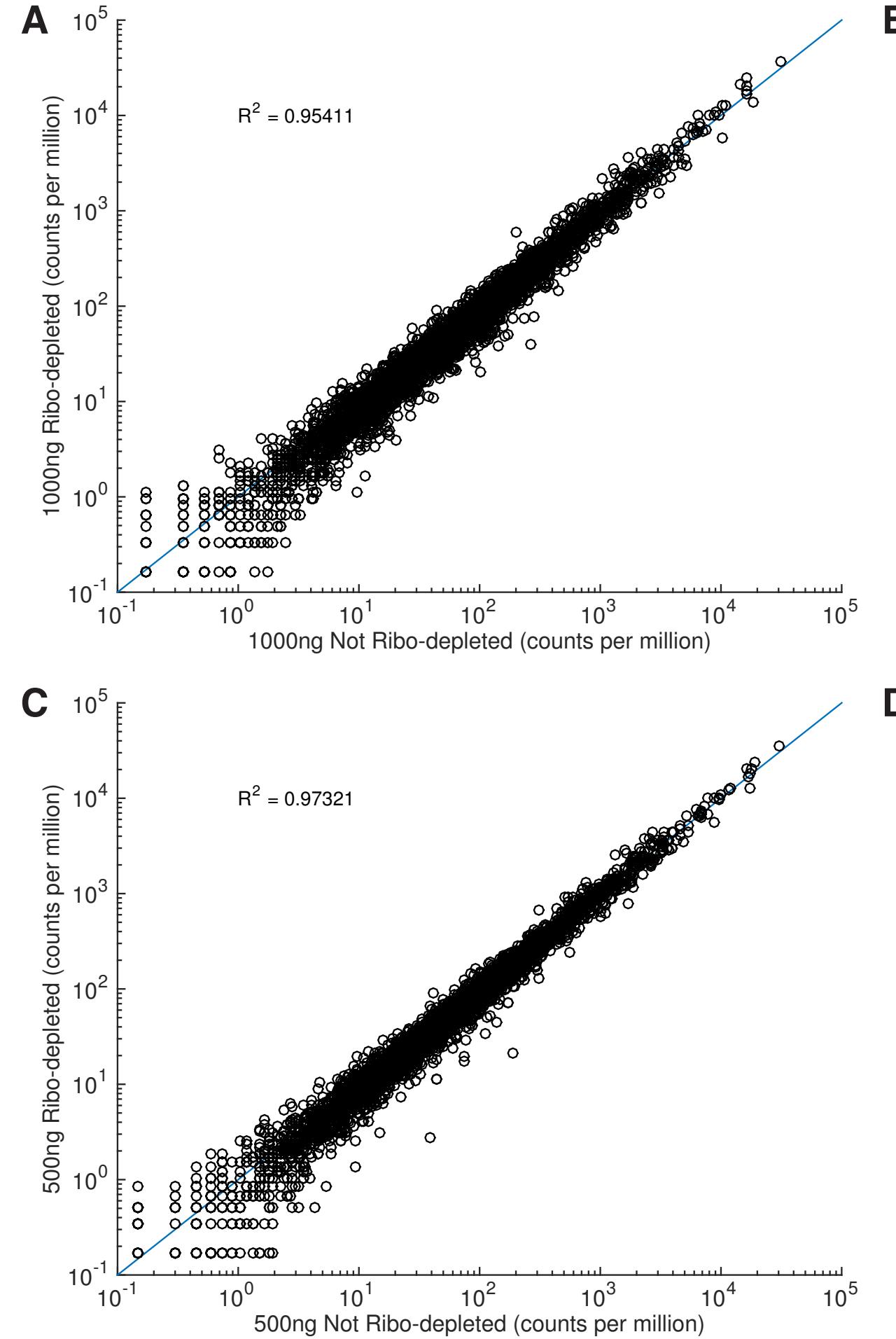
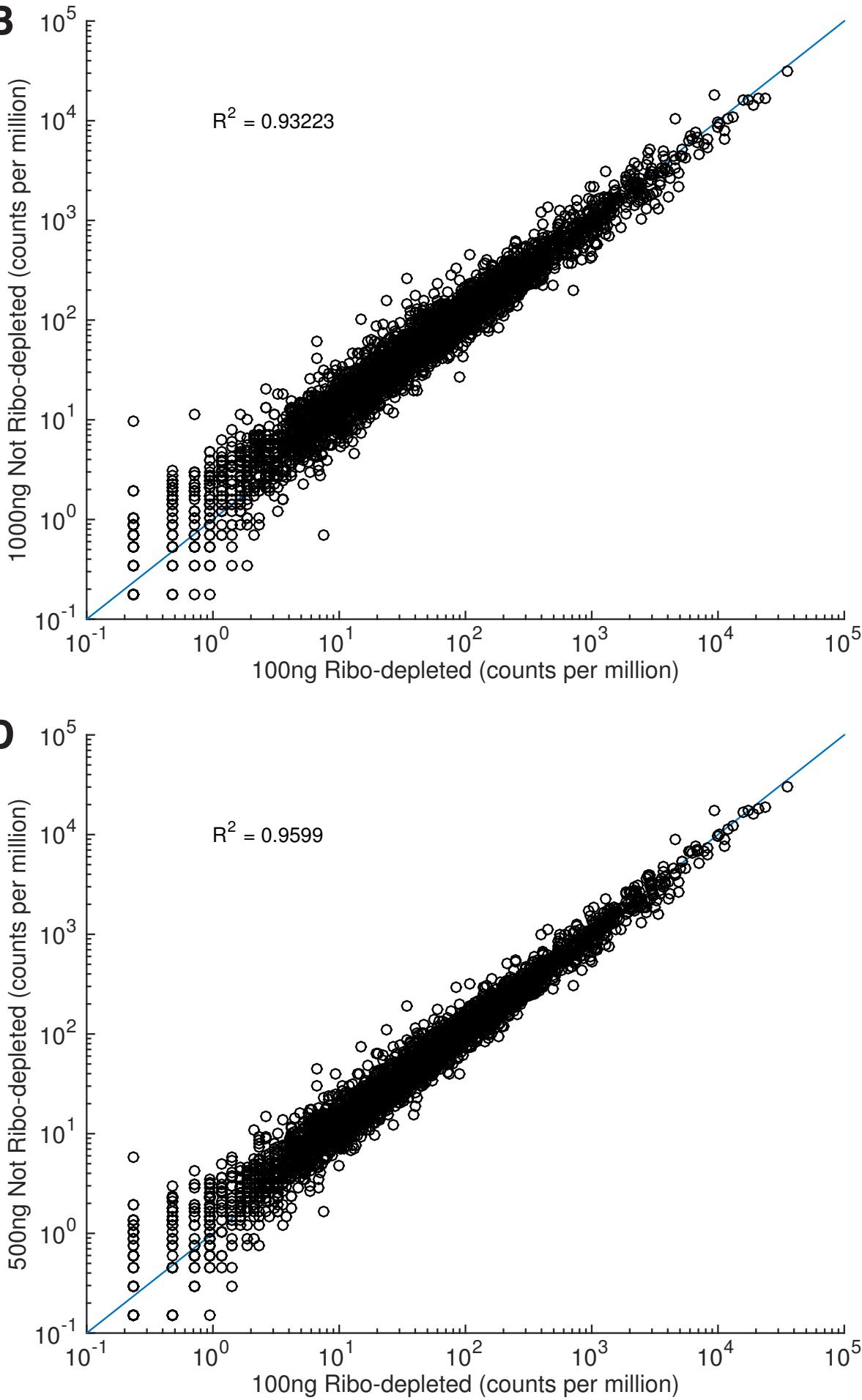
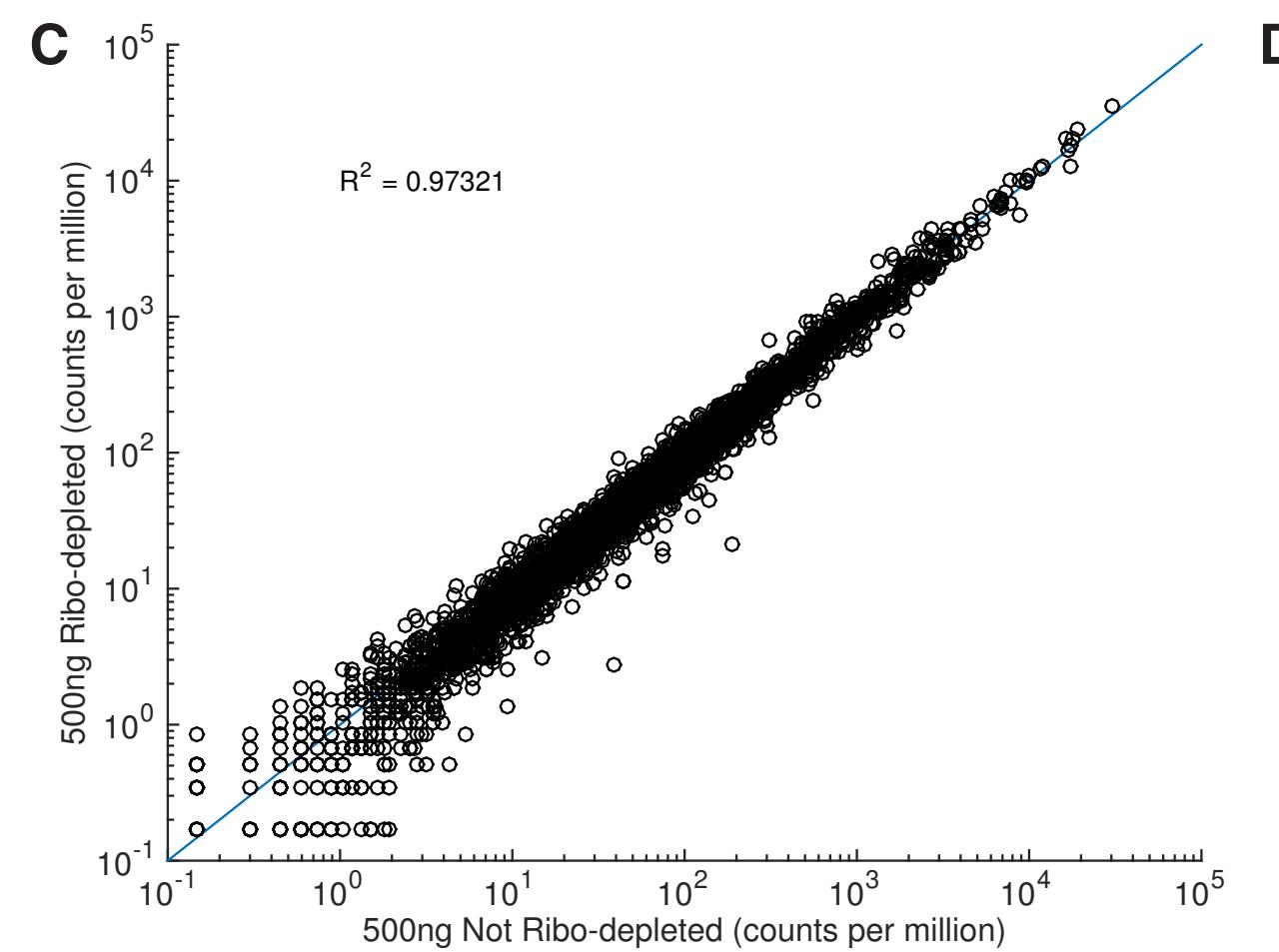
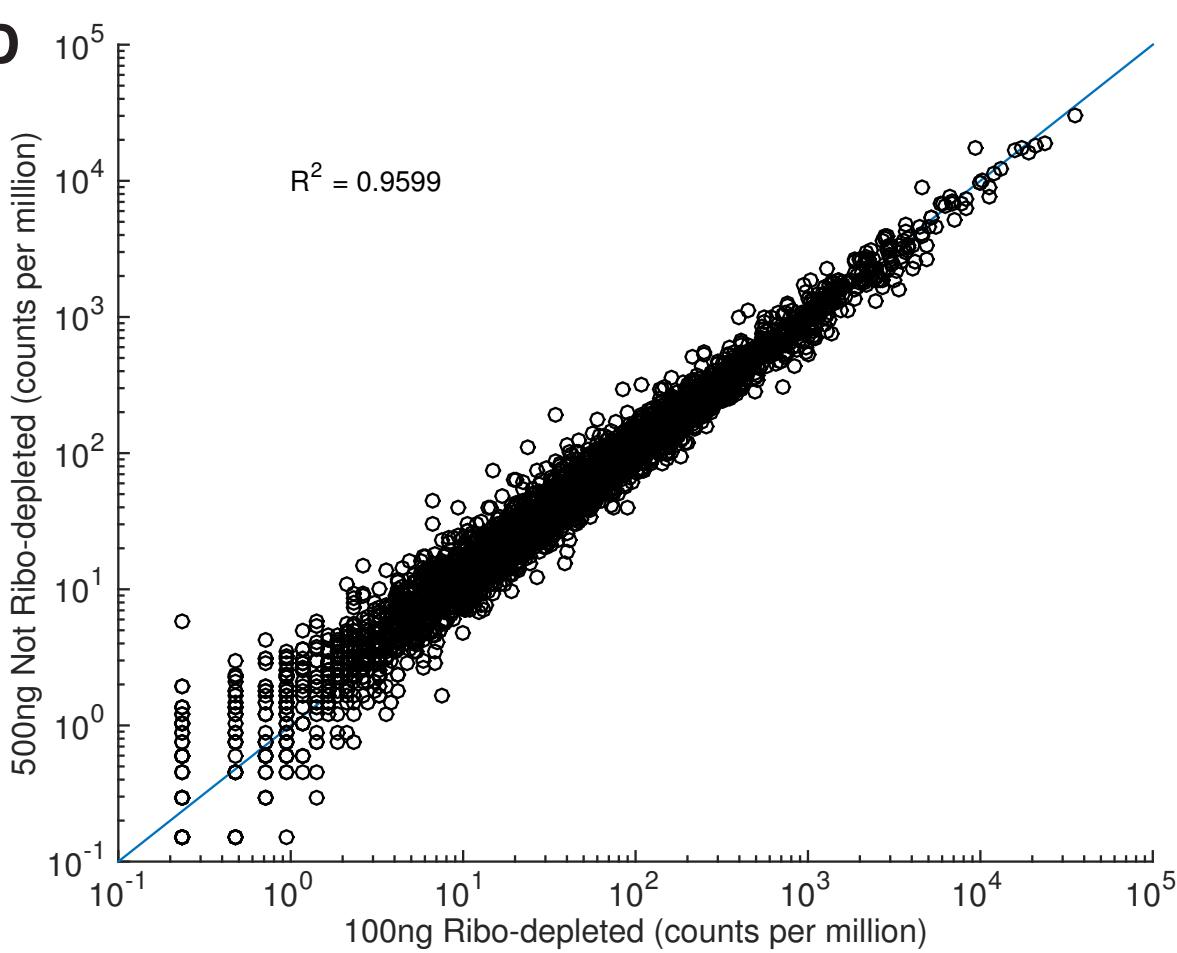
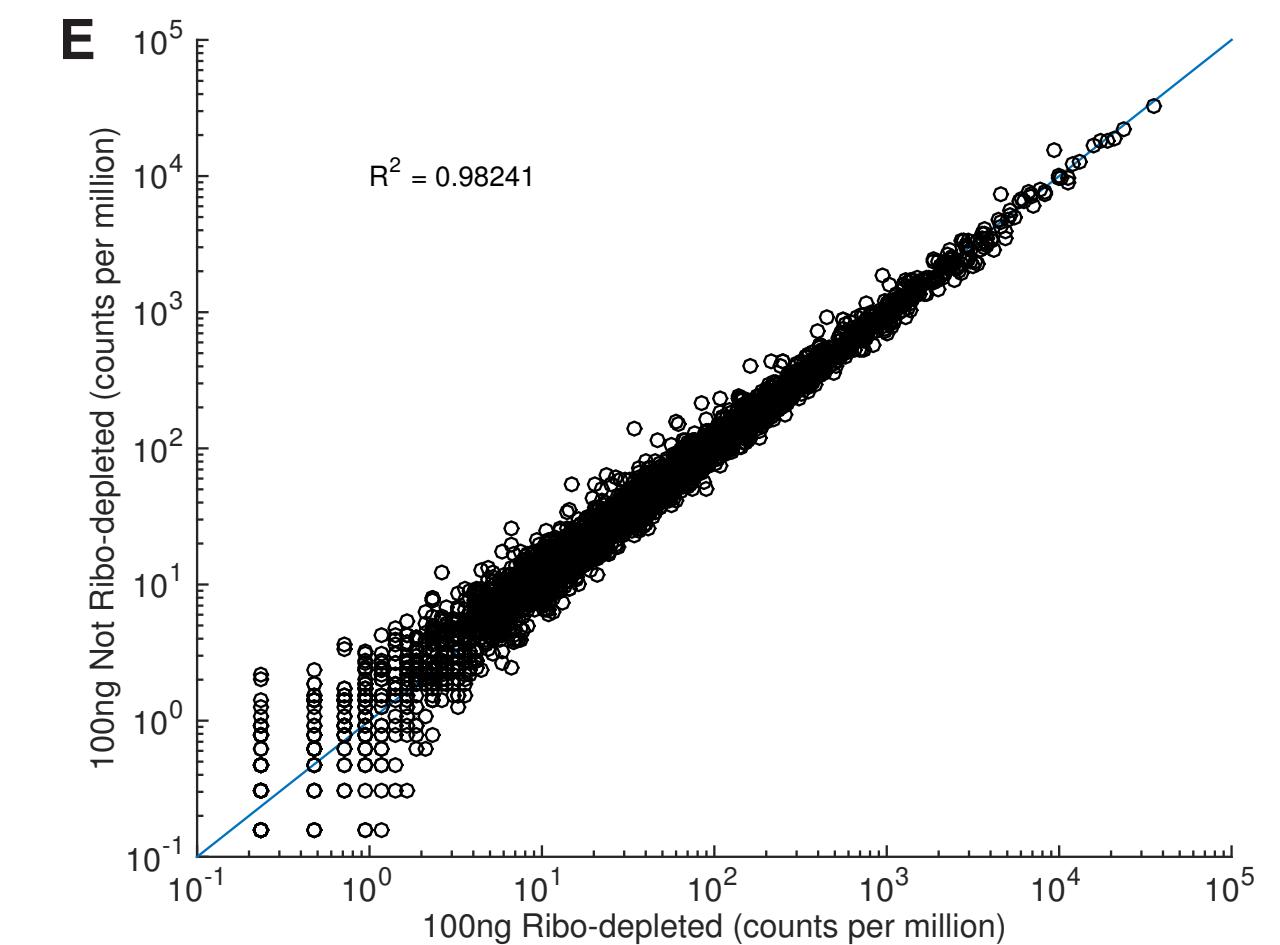
- \* We determined the lower limit of ribo-depleted RNA that would produce robust results when following the standard protocol of the TruSeq RNA Sample Preparation Kit (Illumina).
- \* Using a single sample of *V. fischeri* total RNA, we made TruSeq libraries from three amounts of non-ribo-depleted total RNA (1000, 500, and 100 ng), and nine amounts of ribo-depleted RNA (1000, 500, 100, 50, 25, 10, 5, 2.5, and 1 ng).
- \* Ribo-depletion using the Ribo-Zero Gold Epidemiology Kit before library prep reduced the percentage of rRNA in the sample from ~90% to ~1%.
- \* Ribo-depletion did not appreciably affect the relative abundance of individual mRNAs detected. DESeq2 did not identify any genes significantly differentially abundant between ribo-depleted and non-ribo-depleted samples (FDR-adjusted p-value < 0.05).
- \* Genes with the lowest relative mRNA abundance (across all samples) were those most likely to be undetected in the low-input RNA samples.
- \* Input total RNA could be reduced to 50 ng without loss of coverage, reduced further to 10 ng without loss of mRNA relative abundance fidelity, reduced further to 2.5 to 5 ng with ~10% reduction in coverage and mRNA relative abundance fidelity.

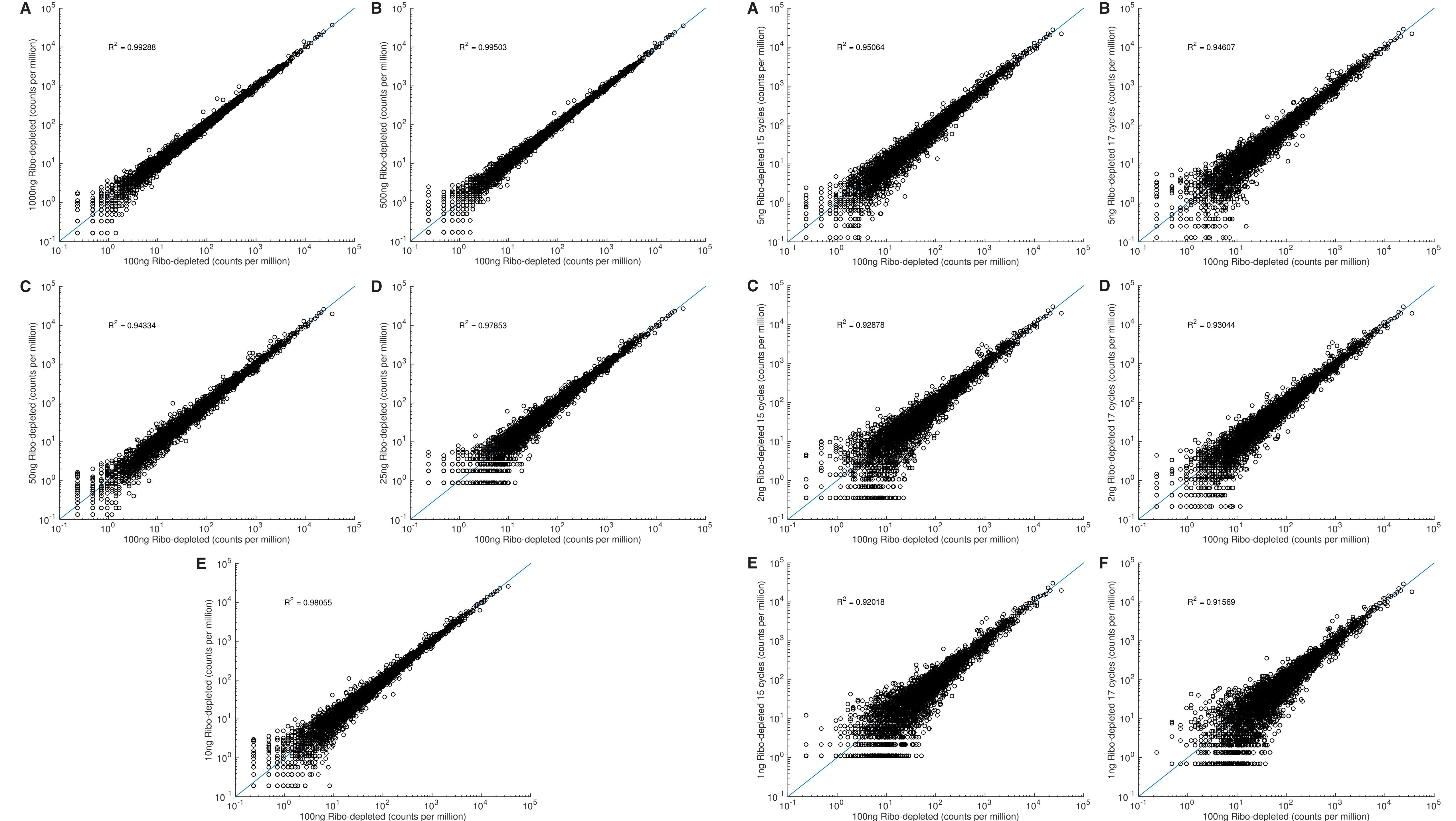
**A**

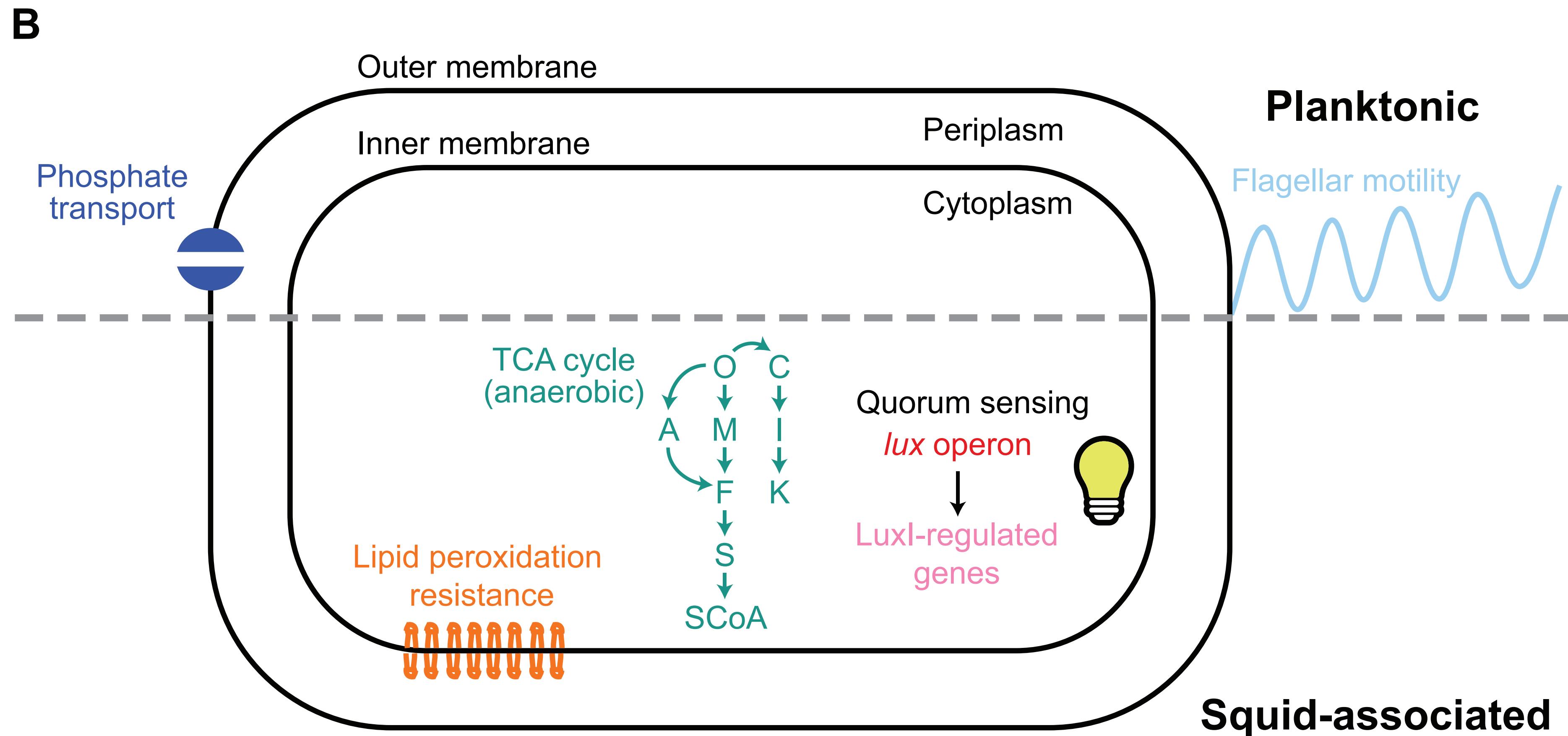
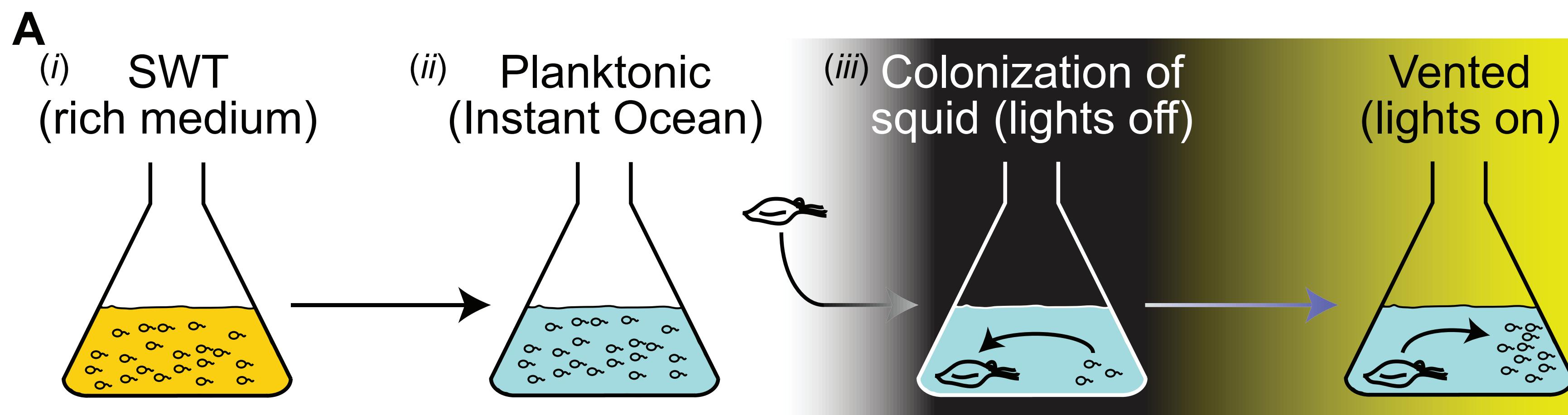
■ rRNA  
■ CDS

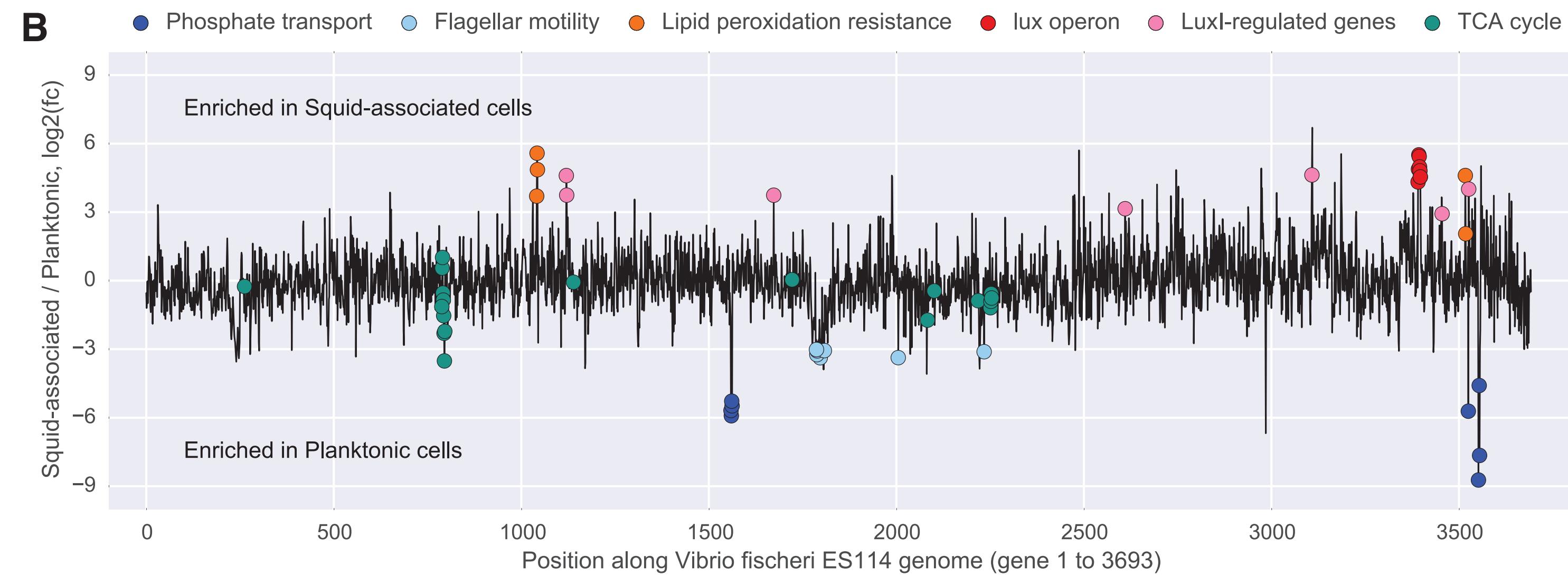
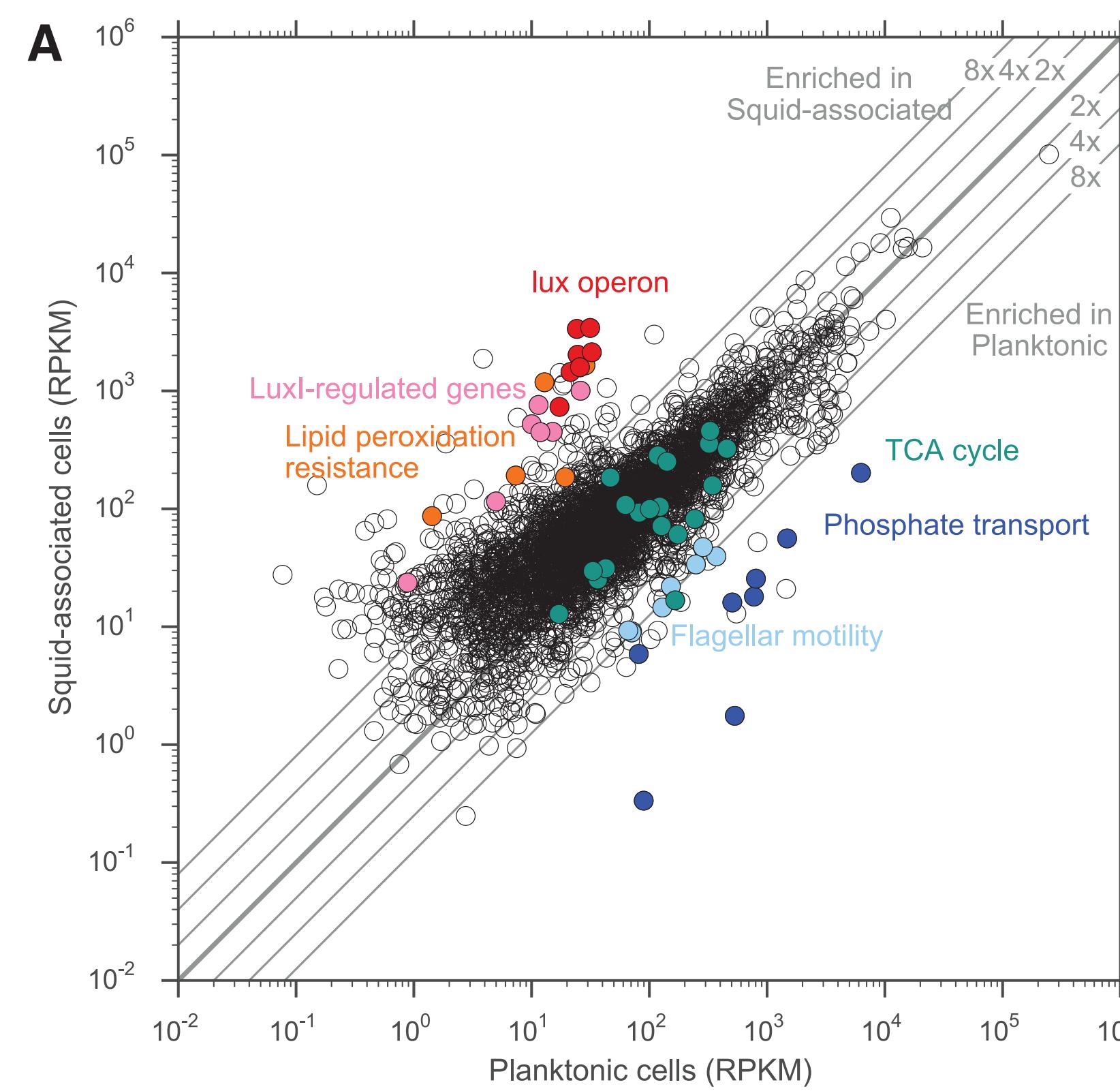
**B**

■ sum\_good\_alignments  
■ not\_aligned  
■ ambiguous  
■ no\_feature

**A****B****C****D****E**







**Thompson, L. R., Zeng, Q. & Chisholm, S. W.** Gene expression patterns during light and dark infection of *Prochlorococcus* by cyanophage. *PLOS ONE* 11, e0165375 (2016). doi:10.1371/journal.pone.0165375

**Thompson, L. R., Nikolakakis, K., Pan, S., Reed, J., Knight, R. & Ruby, E. G.** Transcriptional characterization of *Vibrio fischeri* during colonization of juvenile *Euprymna scolopes*. *Environ Microbiol* 19, 1845–1856 (2017). doi:10.1111/1462-2920.13684

@lukethompson