"A High Throughput Methods for Identifying Novel Genes that Influence Metabolic Pathways Reveals New Iron and Heme Regulation in *Pseudomonas aeruginosa*" Glansville et al., 2021.

The tetrapyrrole heme is a critically important molecule that bacteria synthesize it de novo or acquire from the environment. *Pseudomonas aeruginosa*, causing cystic fibrosis and infections in burn victims, uses both ways to utilize heme. The heme uptake systems in this pathogenic organism are well studied; in contrast, the influence of heme biosynthesis on its uptake and its intracellular maintenance was not be considered for a long time. The essentiality of heme biosynthesis pathway and its genes has precluded its study. This research was conducted to discover novel genes and metabolic pathways that involve in intracellular heme level maintenance.

Glansville et al. (2021) attempted to create a new approach to identify genes and pathways involved in a metabolic process. The authors' new approach involved metabolite-coupled transposon sequencing and make use of three established technologies. Specifically, the approach, named 'Met-Seq', combines the use of biosensors, fluorescent-activated cell sorting (FACS), next-generation sequencing (NGS) techniques. In order to verify whether this new approach worked as intended, they used the Gram-negative bacteria *P. aeruginosa* as a model organism and tried to identify genes involved in the heme uptake and metabolism pathways. First, the authors make a heme biosensor that utilizes phytochrome light receptor proteins. Next, they used existing, previously characterized methods called FAST-INseq, TraDISort, and Persister-FACSSeq in sequence to provide the identification of many genes affecting the levels of heme. Figure 1A from the text is an apt summary of the Met-Seq technique. These processes, together with FACS enrichment, make up the new method for specific metabolite affecting gene identification, Met-Seq.

Generally, this study can be classified into two major areas, which are:

1) Devising a heme biosensor based on phytochrome light receptor protein architecture.
2) Tn-coupled "FlowSeq"-based studies to identify genes *en masse* that affect the levels of a directed metabolite, called by authors metabolite-coupled Tn-sequencing (Met-Seq).
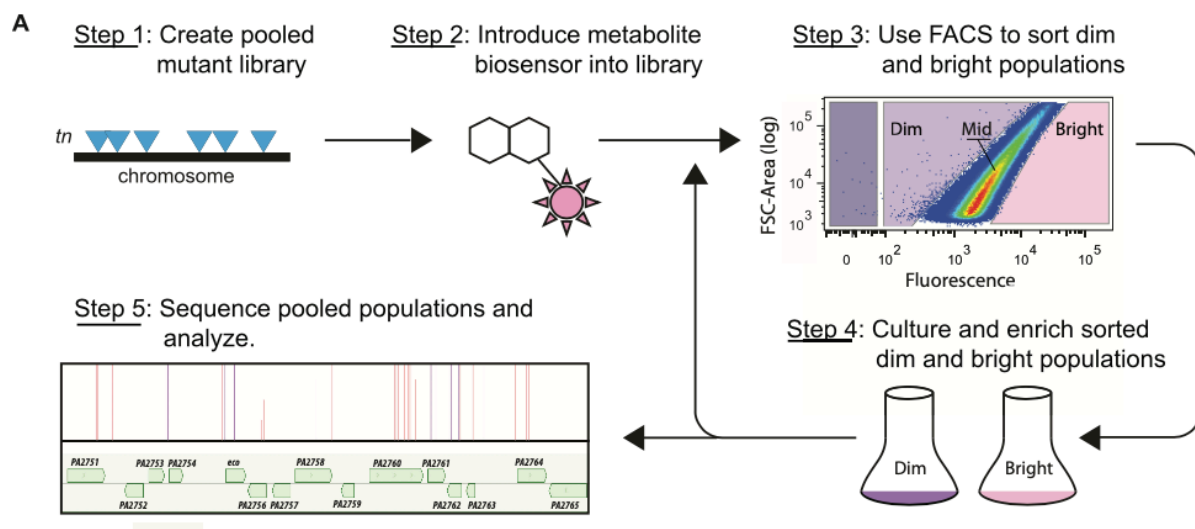

Figure 1A. Glanville et al., 2021.

Replication of data from Glanville et al. (2021) for this project was done using the Met-Seq data publicly available. Specifically, our group chose to replicate Figures 2E, 2F, 3A, and 3C. Figure 2E depicts the number of unique Tn insertions by the thousand on the Y-axis and the enrichment number on the X-axis. Enrichment 0 represents the Tn library and biosensor prior to FACS enrichment. The enrichment of the 'Dim' populations was completed three times and are represented in the graph as 1, 2, and 3. Data labels above the black lines indicate unique insertion percentages in open reading frames, known and predicted.

The results of this study can be summarized as follows:

1) After three rounds of FACS enrichment, 188 genes that significantly diminish the biosensor signal were identified. Know iron/heme regulatory genes (e.g., *dnr* that regulates heme biosynthesis) were detected. However, a majority were not previously thought to be involved in the heme metabolism pathways of *P. aeruginosa*; these groups of genes were associated with heme regulation, small RNAs/riboswitches, central metabolic pathways, and virulence delivery systems.
2) Dnr and four previously unknown genes were identified to affect intracellular levels of heme in *P. aeruginosa*.

What we have done

- For most of the results (tables and figures) presented in this article, data are not available online; thus, we have been obliged to reproduce only four figures (Figures 2 E and F, as well as Figures 3 A and C) for which data have been presented.
- Data used to reproduce Figures 2E, 2F, and 3A was Tn-seq data, which was downloaded from the NCBI website using the accession numbers. We have conducted genome mapping and analysis using PATRIC, bowtie2, and TSAS. The figures have been reproduced in Rstudio using ggplot2.
- Some of the data were downloaded from the supplementary materials, formatted, and analyzed in Rstudio using ggplot2.

General conclusion

- Despite being published early this year, this paper is not well documented. A Github repository link (Tn data) and NCBI accession numbers were provided in the paper; however, most of the data, as well as commands used for genome mapping and plotting the figures, were not deposited in the repository.
- Good documentation does not only help to reproduce the work but also increases the credibility of the researchers. Therefore, it should be taken seriously!