Linking genotype to phenotype in *Streptococcus pneumoniae*: Genome-wide approaches for profiling genetic interactions

Clement Ng[1#], Yifan Li[1#], Justin J. Zik[1], Morgan N. Price[2], Keisha Hanifa Alma Mayra[1], Audrey A. Santoso[1], Adam P. Arkin[2,3], Adam M. Deutschbauer[2,4], Lok-To Sham[1*]

1 Infectious Diseases Translational Research Programme and Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. 2 Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. 3 Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA. 4 Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. #These authors contributed equally to this work. *Corresponding author. Email: lsham@nus.edu.sg

NUS National University of Singapore | Infectious Diseases Translational Research Programme, Yong Loo Lin School of Medicine | Department of Microbiology and Immunology, Yong Loo Lin School of Medicine

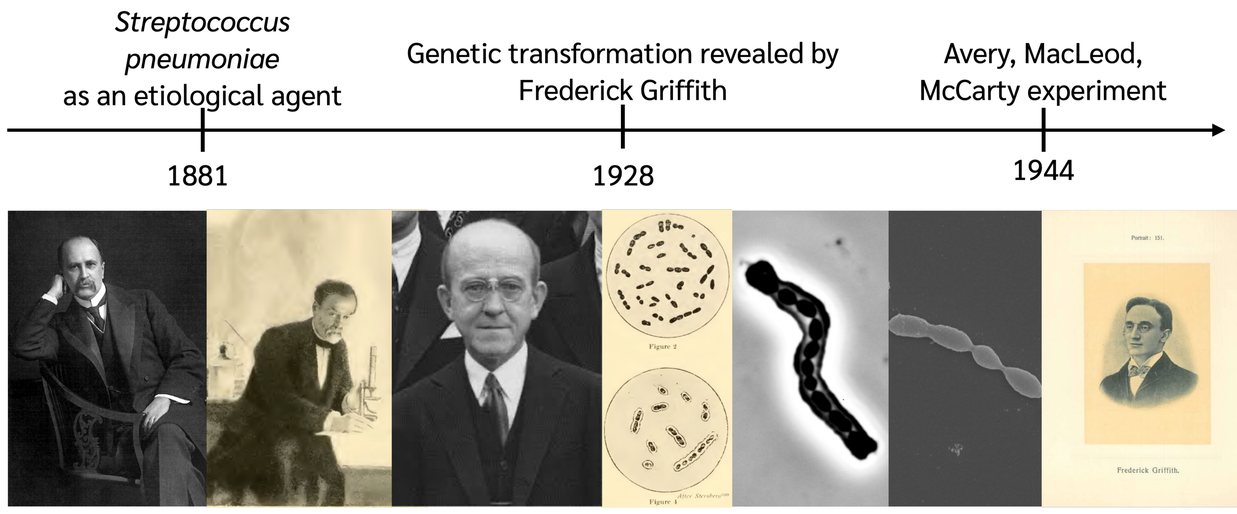
Abstract

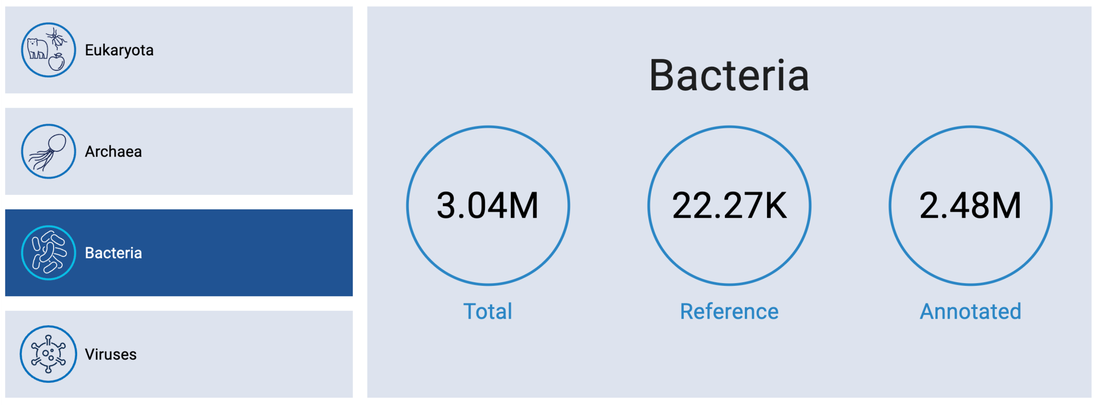
Next-generation DNA sequencing (NGS) has significantly increased the volume of sequencing data available in public databases. However, millions of these sequences remain as genetic "dark matter." For example, about one-third of bacterial genes in the model organism *Escherichia coli* lack known functions, partly due to challenges like gene redundancy. Transposon sequencing (Tn-seq) is a widely used, unbiased approach for studying gene function. This technique tracks changes in the abundance of randomly inserted, gene-inactivating transposons within a saturated mutant pool, helping to identify genes essential for survival under various conditions. However, traditional Tn-seq is not designed to map genetic interaction networks involving more than a few genes. My laboratory recently developed dual Tn-seq and Sup-seq to uncover negative and positive genetic interactions, respectively. These approaches expedite functional genomics, open new research opportunities, and accelerate drug discovery.

Background and the unmet medical needs

Streptococcus pneumoniae

- Kills more than **0.5 million people** worldwide annually.
- A leading cause of lower respiratory infections.
- More than 100 serotypes have been identified. **Licensed vaccines cover at most 23 types**.
- Naturally competent. AMR is on the rise.





Positive genetic interaction

Epistasis / Synthetic viable / suppression

Substrate → **Enzyme A** → **TOXIC** → **Enzyme B (essential)** → Non-essential product

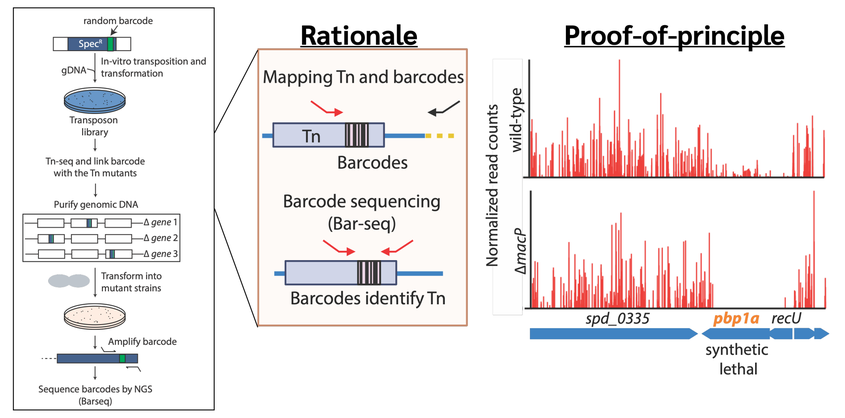
Negative genetic interaction

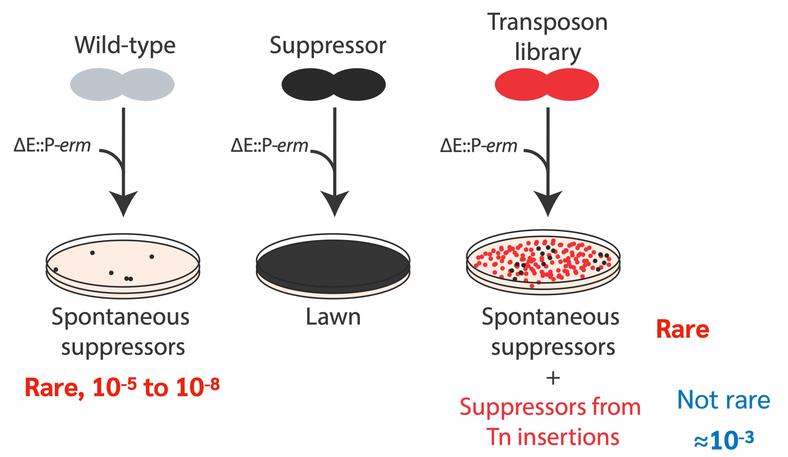
Synthetic lethal / sickness

Substrate → **Enzyme A** + **Enzyme B** → **Product required for survival**

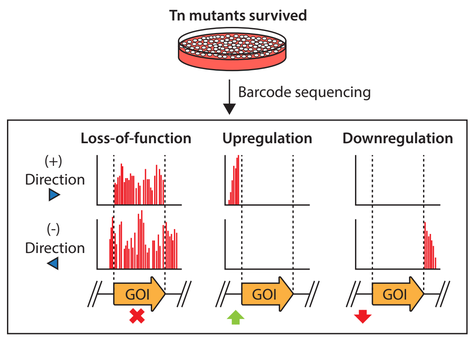
Profiling genetic interactions with Sup-seq

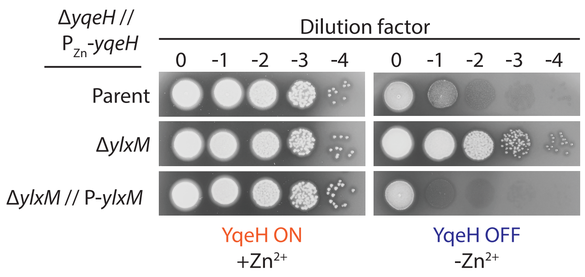
- Complement our recently published **dual Tn-seq** approach (*Science* 2025; PMID: 40997176).
- Randomly barcoded transposon sequencing (RB Tn-seq) enables rapid mapping of transposon mutants.
- Spontaneous suppressor mutants are **rare**. But a specific transposon mutant in a library that disrupts a specific gene is not.
- Transposon insertions that suppress an essential gene function are enriched.
- The process can be iterated, and the Tn mutants are sequenced.

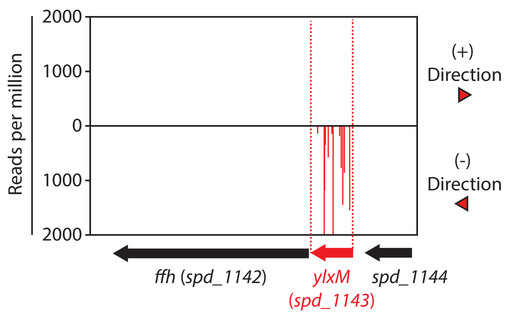


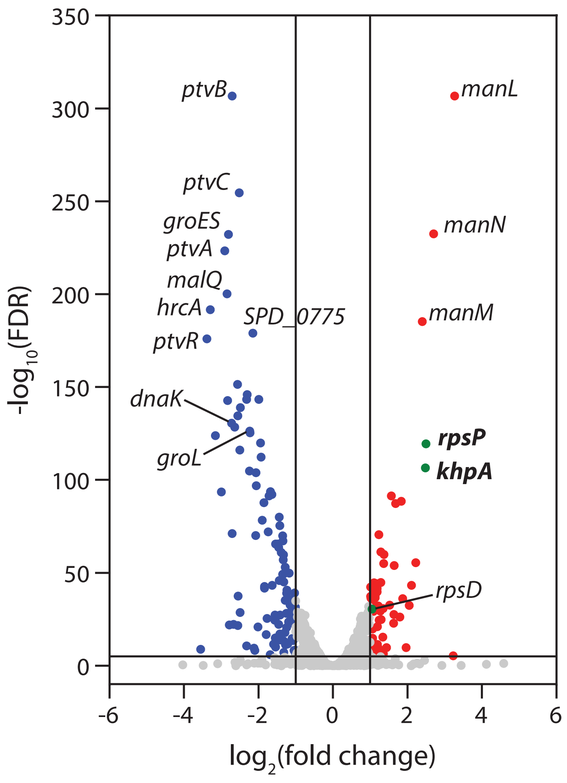


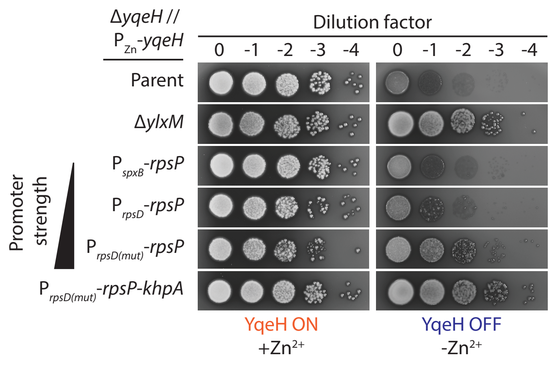
YqeH is suppressed by the loss of YlxM







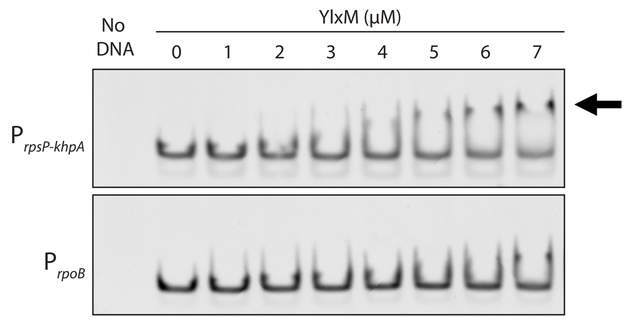


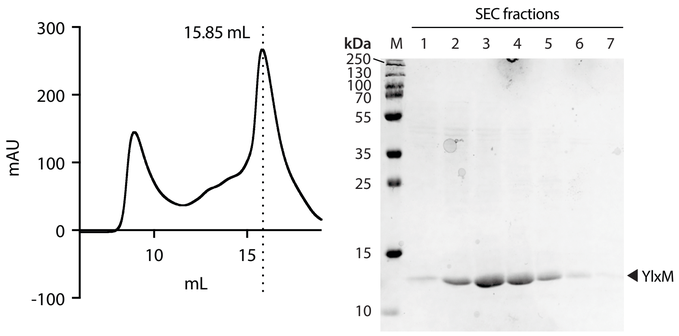


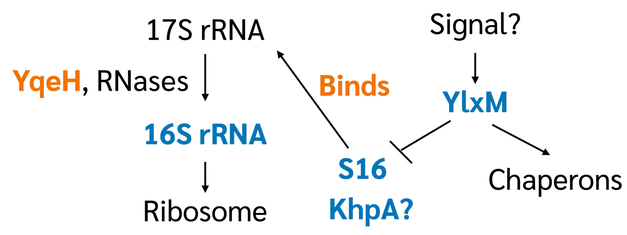
- YqeH is a widely conserved circular permutated GTPase required for **ribosome assembly**.
- Deletion of a small DNA-binding protein, **YlxM**, bypassed YqeH essentiality.
- RNA-sequencing experiments showed that ***rpsP-khpA*** is overexpressed in the Δ*ylxM* mutant, whereas **chaperon genes** are downregulated.
- **Overexpression of *rpsP* is sufficient** to alleviate the phenotype resulting from *yqeH* depletion
- Suppression is enhanced by khpA overexpression.

YlxM is a transcription factor that control translation

- YlxM was purified by Ni-NTA chromatography and size-exclusion chromatography.
- **Purified YlxM binds to the *rpsP-khpA* promoter**, suggesting a direct transcriptional interaction.
- YlxM likely regulates protein folding.







Conclusions

- Identifying genetic interactions is a facile way to understand gene function.
- **Sup-seq is an unbiased approach** that enables genome-wide screens for genetic suppression events.
- As a proof of concept, we screened **7 essential query genes and identified 53 genetic interactions**.
- Functional genomics remains central for uncovering drug and vaccine targets.

7 essential query genes screened

58 candidate suppressors detected

53 confirmed by reconstruction

Acknowledgements

We thank Dr. Eun-Jin Lee and her team at Korea University for their guidance and insights into ribosome assembly. Special thanks to Ms. Dong Tianyun for constructing strain NUS6390 and to Mr. Colin Chung for collecting cells for mass spectrometry. We thank Dr. Luo Min for his assistance with protein purification. We would also like to thank members of the L.-T.S. laboratory for their helpful discussions. This work was supported by grants from the National Research Medical Council, Singapore (OFIRG23jan-0057 to LTS), the Ministry of Education, Singapore (MOE-T2EP30220-0012 and MOE-T2EP30224-0012 to LTS), and the National Research Foundation (NRF-F-CRP-2024-0002 to LTS)

Ministry of Education SINGAPORE | National Research Foundation Prime Minister's Office Singapore | NMRC National Medical Research Council Singapore