Chủng VK

hiện tại mình có trình tự giải NGS của gần 90 chủng này, đang làm đề tài Nghiên cứu tác động của các chủng Helicobacter pylori Việt Nam lên hệ phiên mã và cơ chế điều hòa phiên mã trong tế bào biểu mô dạ dày, trong đề tài dự kiến làm 3 chủng việt nam, 1 chủng ATCC, về phân tích liên kết đặc ddiemr phiên mã của tb chủ với đặc điểm di truyền các chủng có các hướng nào có thể thực hiện, đề xuất ý tưởng

ChatGPT said:

Updated saved memory

Rất tuyệt vời — bạn đang có một bộ dữ liệu quý giá gồm gần 90 chủng H. pylori được giải trình tự, và một đề tài tập trung vào tác động lên hệ phiên mã tế bào chủ (cụ thể là biểu mô dạ dày). Dưới đây là những hướng nghiên cứu tiềm năng và đề xuất cụ thể để liên kết hệ phiên mã chủ với đặc điểm di truyền vi khuẩn, phù hợp với định hướng hiện tại của bạn.

✅ 1. Phân loại hệ gene vi khuẩn (comparative genomics)

→ Mục tiêu: xác định các đặc điểm di truyền đặc trưng ở mỗi chủng

Bạn đang có NGS của ~90 chủng, có thể làm:

Pan-genome analysis: xác định core genes (tồn tại ở tất cả) vs. accessory genes (chỉ có ở một số chủng)

Annotate gene liên quan đến độc lực: cagA, vacA, babA, oipA, iceA, T4SS (type IV secretion system), các gene nội độc tố, biến đổi biểu sinh (DNMT-like…)

SNP/InDel analysis trong các gene này hoặc genome-wide

👉 Phần này giúp nhóm các chủng HP thành các kiểu độc lực khác nhau → nền tảng để chọn 3 chủng đại diện.

✅ 2. Phân tích biểu hiện RNA-seq của tế bào chủ sau nhiễm

Bạn đã nói rõ rằng muốn phân tích:

Tác động lên hệ phiên mã tế bào biểu mô dạ dày (giả sử là GES-1 hoặc AGS...)

Sử dụng RNA-seq (đã hoặc sẽ thực hiện)

Các bước cần:

So sánh DEGs (Differentially Expressed Genes) giữa tế bào nhiễm các chủng khác nhau

Phân tích pathway/GO (NF-κB, EMT, apoptosis, DNA damage…)

Xác định các marker phiên mã đặc hiệu chủng (ví dụ: chỉ chủng A gây tăng IL8, chủng B gây tăng CDKN1A)

✅ 3. Liên kết đặc điểm phiên mã chủ ↔ đặc điểm di truyền vi khuẩn

Đây là phần tâm điểm khoa học, bạn có thể triển khai theo 3 hướng lớn:

🔸 A. Phân tích tương quan gene-to-gene (genotype–transcriptome association)

➤ Mục tiêu:

Tìm mối tương quan giữa đặc điểm gene/biến thể của HP và phản ứng phiên mã của chủ

Cách làm:

Biến đặc điểm vi khuẩn thành dữ liệu phân loại (có/không): ví dụ chủng có cagA, vacA s1, T4SS+, hay SNP tại vị trí X

Với từng đặc điểm, làm so sánh biểu hiện gene chủ:

IL8 cao hơn khi nhiễm chủng cagA+

CDH1 thấp hơn khi chủng có vacA s1m1

STAT1 chỉ tăng khi có T4SS

→ Dùng test thống kê (DESeq2 + annotation nhóm vi khuẩn)

🔸 B. Phân nhóm chủng theo gene set vi khuẩn → phân tích transcriptome chủ

➤ Mục tiêu:

Tìm nhóm chủng HP có mô hình độc lực giống nhau → xác định hiệu ứng hệ phiên mã chủ

Cách:

Gom nhóm chủng theo:

Cấu hình gene độc lực (cluster bằng PCA, hierarchical clustering)

Pan-genome group hoặc core SNP phylogeny

Với mỗi nhóm, xác định:

DEGs đặc trưng

Pathway bị kích hoạt rõ nhất

Liên kết với tính chất sinh học (viêm, EMT, apoptosis…)

🔸 C. Mô hình học máy (machine learning) hoặc mô hình thống kê đa biến

Nếu số lượng mẫu đủ (nhiễm nhiều chủng trên cùng dòng tế bào), bạn có thể:

Dùng đặc điểm vi khuẩn (cagA status, vacA type, SNP pattern...) làm biến đầu vào

Dùng biểu hiện gene chủ (hoặc PCAs của transcriptome) làm đầu ra

Áp dụng:

Random forest → tìm đặc điểm vi khuẩn nào quan trọng nhất

Partial least squares regression (PLS-R)

LASSO / ElasticNet

✅ 4. So sánh chủng Việt Nam và chủng ATCC

ATCC 43504 là chủng chuẩn CagA+ VacA s1m1, phổ biến làm control

So sánh DEGs giữa ATCC vs. 3 chủng Việt Nam:

Chủng nào gây kích hoạt mạnh hơn pathway nào?

Chủng nào ức chế miễn dịch, gây mất kết dính, tăng biểu hiện ZEB1, MMP9…?

→ Nếu có chủng Việt Nam gây biểu hiện đặc thù hơn cả ATCC, có thể là hướng mới!

🎯 Gợi ý tổng thể mô hình phân tích:

text

Copy

Edit

1. Phân tích gene & SNP vi khuẩn → nhóm chủng theo đặc điểm di truyền

2. Phân tích transcriptome chủ → nhóm phản ứng chủ theo DEGs/pathways

3. Liên kết 1 ↔ 2 bằng:

- So sánh từng gene độc lực ↔ gene biểu hiện

- Gom nhóm chủng ↔ DEGs pathway

- Tương quan / ML mô hình

🧪 Gợi ý kỹ thuật đi kèm:

Mục tiêu Công cụ gợi ý

Annotate genome HP Prokka, Roary, Panaroo

Pan-genome BPGA, Roary, PanX

SNP calling Snippy, CSIPhylogeny

Phân tích RNA-seq STAR + DESeq2 / edgeR

Pathway chủ GSEA, IPA (hoặc Metascape miễn phí)

Tương quan R (ggpubr, corrplot), Python (seaborn), mlr3