01-Research background

vCagAepitope: A Computational Investigation into the Microbiome's Role in Sustaining Pro-Tumorigenic T-Cell Responses

The goal of the vCagAepitope project is to identify the microbial or host-derived factors responsible for the chronic activation of pathogenic, CagA-specific CD8+ T-cells in the gut, which have been shown to drive intestinal tumorigenesis.

1. Research Background

Helicobacter pylori (*H. pylori*) is a bacterium that colonizes the human stomach and is a primary risk factor for gastric cancer. Its oncogenic potential is strongly linked to the expression of the CagA (Cytotoxin-associated gene A) protein, which is injected into host epithelial cells, disrupting cellular signaling and promoting carcinogenesis. While the role of CagA in gastric cancer is well-established, its influence on diseases of the lower gastrointestinal tract, such as colorectal cancer (CRC), is an area of active and critical investigation. The gut microbiome, the complex ecosystem of bacteria, viruses, and fungi residing in the intestines, is known to profoundly influence host immunity and cancer development. This project investigates the intersection of a specific bacterial virulence factor (*H. pylori* CagA), host genetics, and the gut microbiome (including virome) in the context of intestinal tumorigenesis.

2. Previous Findings & The Central Unanswered Question

Our collaborators have established a direct causal link between *H. pylori* infection and intestinal tumor development in a genetically susceptible mouse model (`Apc_1638N`). Their key findings are:

- **CagA-Dependent Tumorigenesis:** Infection with CagA-positive *H. pylori* strains promotes intestinal inflammation and significantly accelerates tumor development, an effect not observed with CagA-negative strains.
- **Causative Role of CD8+ T-Cells:** This pro-tumorigenic effect is mediated by CagA-specific CD8+ T-cells. These cells are primed in the stomach during the initial infection but subsequently migrate to and populate the intestinal lamina propria and tumors. Depletion of CD8+ T-cells (but not CD4+ T-cells) abrogates tumor development.
- **Persistence of Pathogenic T-Cells:** Crucially, these CagA-specific CD8+ T-cells persist in the gut and remain active long after the primary *H. pylori* infection has been eradicated by antibiotics. Their numbers directly correlate with intestinal tumor burden.

This leads to the central unanswered question of this project: **What factor within the gut is responsible for the persistent activation of these pathogenic, CagA-specific CD8+ T-cells in the absence of the original CagA antigen?**

Initial hypotheses centered on molecular mimicry, where a peptide from a gut microbe or virus would mimic a CagA epitope. However, preliminary BLAST-based searches have failed to identify a likely candidate, suggesting a more complex mechanism is at play.

3. Current Data & Experimental Design

This project will analyze a rich dataset of paired-end, short-read Next-Generation Sequencing (NGS) samples derived from a well-controlled mouse study.

- **Host Model:** C57BL/6 mice with two genotypes:
    a. Wild-type (`Apc_wt`)
    b. Cancer-susceptible mutant (`Apc_1638N`)
- **Infection Groups:**
    a. Uninfected (`control`)
    b. Infected with wild-type *H. pylori* PMSS1 (`hp_wt`, CagA-positive)
    c. Infected with a CagA-knockout mutant *H. pylori* PMSS1 (`hp_koCagA`, CagA-negative)
- **Samples:** NGS data was generated from two sample types: tissue-adherent microbes from the **`caecum`** and luminal microbes from the **`intestinal_content`**. The `caecum` samples represent the microbial community most likely to interact with the host epithelium and will be the primary subject of our comparative analyses. The `intestinal_content` samples will be used to supplement the metagenomic assembly, creating a more comprehensive catalog of microbial genes and genomes.
- **Key Comparison:** The most critical comparison for all quantitative analyses will be between the **`ApcMUT_HpWT`** and **`ApcMUT_HpKO`** groups within the **`caecum`** samples. This isolates the CagA-dependent effects at the site of host-microbe interaction.

4. Proposed Bioinformatic Analyses & Next Steps

Hypothesis: The "Functional Footprint" Hypothesis (Metabolic Cross-talk)

- **Concept:** The T-cell activation is driven by a non-peptidic molecule or metabolite produced by a specific microbial community whose structure was permanently altered by the initial CagA-positive infection.

- **Rationale:** This shifts the search from sequence mimicry to functional output. The microbiome's metabolic activity, rather than its genetic sequence alone, may be the key driver of inflammation.
- **Analysis Plan:**
    a. **Quality Control:** Process raw FASTQ reads from all samples using `fastp` to trim adapters and remove low-quality sequences.
    b. **Taxonomic Profiling:** Using only the **`caecum`** samples, determine the species-level composition of the microbiome using `Kraken2`. Perform differential abundance testing (e.g., with `ANCOM-BC`) to identify species enriched in the `ApcMUT_HpWT` caecal community.
    c. **Functional Profiling:** This is the key step. Using only the **`caecum`** samples, use `HUMAnN3` to profile the abundance of microbial metabolic pathways. Identify pathways that are significantly enriched in the `ApcMUT_HpWT` group compared to controls.
    d. **Strain-Level Analysis:** For key species identified in step 2, use `StrainPhlAn` on the **`caecum`** samples to investigate if specific strains are associated with the tumor-promoting environment.
    e. **Correlation:** Correlate the abundance of candidate species, strains, or metabolic pathways from the `caecum` with existing data on tumor burden and T-cell numbers for each mouse.