**Power Analysis**

To ensure the study is adequately powered to detect meaningful effects, we performed sample size calculations for the two co-primary endpoints: (1) changes in the gut microbiome composition in human participants following dietary interventions, and (2) the associated impact on breast cancer risk in a preclinical mouse model using fecal microbiota transplantation (FMT). Below, we outline the parameters, assumptions, statistical tests, and operating characteristics (power versus effect sizes) for each endpoint. Calculations aim for a minimum power of 80% (β = 0.20) and a significance level of α = 0.05, adjusted for multiple comparisons where applicable. The endpoint most difficult to detect statistically is identified, and separate calculations are provided to ensure robustness.

**1. Power Analysis for Changes in the Gut Microbiome**

**Objective**: Detect significant changes in gut microbiome composition (e.g., differential abundance of key taxa such as *Akkermansia muciniphila* or SCFA-producing bacteria) across five dietary interventions in a crossover design.

**Statistical Test**: Linear Discriminant Analysis Effect Size (LEfSe) and DESeq2 will be used to identify differentially abundant microbial features, with Bray-Curtis dissimilarity matrices analyzed via Analysis of Similarities (ANOSIM) for community composition differences. For sample size purposes, we base calculations on a paired t-test (or Wilcoxon signed-rank test for non-normal data) to compare microbiome metrics (e.g., relative abundance of a key taxon) within individuals across interventions, as this simplifies power estimation while aligning with the crossover design.

**Parameters and Assumptions**:

* **Effect Size**: We assume a moderate Cohen’s d of 0.5 (standardized mean difference in microbial abundance), reflecting a biologically meaningful shift based on prior studies of dietary interventions (e.g., PHGG increasing *Akkermansia muciniphila* abundance, Slavin et al. [1]). Smaller (d = 0.3) and larger (d = 0.8) effect sizes are also considered for operating characteristics.
* **Standard Deviation**: Variability in microbial abundance is assumed to be approximately equal to the mean change (σ ≈ Δ), a common assumption in microbiome studies due to high inter-individual variation, adjusted for within-subject correlation in a crossover design.
* **Correlation**: A within-subject correlation of 0.6 is assumed due to repeated measures, reducing variability compared to a parallel design.
* **Significance Level (α)**: 0.05, adjusted to 0.01 (Bonferroni correction for 5 interventions: 0.05/5) to account for multiple comparisons.
* **Power (1-β)**: Minimum of 80%.
* **Design**: Crossover with 20 participants, each receiving all 5 interventions, yielding 20 paired observations per intervention comparison.

**Sample Size Calculation**: Using the formula for a paired t-test:

​

Where:

* = 2.576 (for = 0.01, two-tailed),
* =0.841 (for 80% power),
* =1 (standardized),
* Δ=0.5 (effect size),
* =0.6 (correlation).

Adjusting for the crossover design, where each participant provides paired data, the per-intervention sample size is reduced by the correlation factor. However, since we need only 20 participants total (each contributing to all 5 interventions), we calculate the effective sample size per comparison:

 comparisons per participant.

With 20 participants, we have 20 observations per intervention, sufficient for paired analyses. Using G\*Power for a paired t-test with = 0.5, = 0.01, and power = 80%, the required sample size is approximately 17 participants. Accounting for a 20% dropout rate (anticipating 40 consents to achieve 20 completers), we propose recruiting 20 participants as sufficient.

**Operating Characteristics**:

* Effect Size = 0.3: Power ≈ 50% (n = 20),
* Effect Size = 0.5: Power ≈ 85% (n = 20),
* Effect Size = 0.8: Power ≈ 99% (n = 20).

A graph with a blue line

AI-generated content may be incorrect.

Fig.1 Power vs Effect Size for Gut Microbiome Changes

The analysis employs a paired t-test, which is suitable for the crossover study design, with a sample size of = 20 and a significance level of = 0.01 (adjusted using the Bonferroni correction for five interventions). Effect sizes ranging from 0.2 to 1.0 were tested to evaluate statistical power. The results are presented in Figure 1, which displays a power curve illustrating the relationship between effect size and statistical power. A vertical line marks the assumed effect size ( = 0.5), while a horizontal line indicates the 80% power threshold. The analysis confirms that a sample size of = 17 is required to detect an effect size of = 0.5 with 80% power, demonstrating that the chosen sample size ( = 20) is sufficient.

**2. Power Analysis for Associated Breast Cancer Risk in Preclinical Model**

**Objective**: Assess whether FMT from participants under different dietary interventions reduces mammary tumorigenesis (e.g., tumor volume or incidence) in syngeneic mice grafted with E0771 mammary tumor cells.

**Statistical Test**: One-way ANOVA to compare tumor outcomes across 5 FMT groups (maltodextrin, cow’s milk + maltodextrin, cow’s/soy milk + PHGG, soy milk + maltodextrin, PHGG), followed by post-hoc tests. For sample size estimation, we use a two-sample t-test comparing the control (maltodextrin) to the most promising intervention (e.g., PHGG) as a conservative approach.

**Parameters and Assumptions**:

* **Effect Size**: We assume a moderate Cohen’s d of 0.65 for tumor volume reduction, based on preclinical studies linking gut microbiome changes to cancer outcomes (e.g., FMT reducing tumor growth by 30-40% [2]). Smaller ( = 0.4) and larger ( = 1.0) effect sizes are explored.
* **Standard Deviation**: Tumor volume variability is assumed as σ= 50 mm³, with a mean difference () of 32.5 mm³ ( = 0.65 × 50), derived from prior mouse studies.
* **Significance Level ()**: 0.05, adjusted to 0.0125 (Bonferroni correction for 4 comparisons vs. control: 0.05/4).
* **Power (1-)**: Minimum of 80%.
* **Design**: 500 mice total (100 per group, with 5 groups × 20 participants providing FMT), but we calculate per-group size for the primary comparison.

**Sample Size Calculation:**

Using the formula for a two-sample t-test:

​

Where:

* = 2.2.241 (for = 0.0125, two-tailed),
* =0.841 (for 80% power),
* =50 mm³,
* Δ=32.550 mm³,

mice per group

With 100 mice per group (5 groups × 20 FMT samples), the study exceeds this requirement. Using G\*Power for a t-test with = 0.65, = 0.0125, and power = 80%, approximately 42 mice per group are needed. The proposed 100 mice per group accounts for variability in FMT efficacy and potential losses (e.g., 10% mortality).

**Operating Characteristics**:

* Effect Size = 0.4: Power ≈ 45% (n = 100),
* Effect Size = 0.65: Power ≈ 95% (n = 100),
* Effect Size = 1.0: Power ≈ 99% (n = 100).

A graph with a line going up

AI-generated content may be incorrect.

Fig.2 Power vs Effect Size for Breast Cancer Risk in Mice

The power analysis for breast cancer risk was conducted using a two-sample t-test to compare the control and intervention groups. The study design included = 100 mice per group and a significance level of = 0.0125, adjusted with the Bonferroni correction for four comparisons. Effect sizes ranging from 0.2 to 1.2 were examined to assess statistical power. The results are shown in Figure 2, which presents a power curve depicting the relationship between effect size and detection power. A vertical line indicates the target effect size ( = 0.65), while a horizontal line marks the 80% power threshold. The analysis confirmed that only = 42 mice per group are required to achieve 80% power for = 0.65, demonstrating that the selected sample size ( = 100) provides ample statistical power.

**Endpoint Most Difficult to Detect**

The breast cancer risk endpoint in mice is likely more difficult to detect statistically due to greater variability (e.g., tumor growth influenced by FMT efficacy, mouse genetics, and experimental conditions) compared to microbiome changes in humans, which benefit from the controlled crossover design and within-subject comparisons. Thus, the sample size of 100 mice per group ensures >80% power even for moderate effects, while 20 human participants suffice for microbiome changes.

**Summary**

* **Human Microbiome**: 20 participants provide >80% power for a moderate effect ( = 0.5) on microbiome composition.
* **Mouse Breast Cancer Risk**: 100 mice per group provide >90% power for a moderate effect ( = 0.65) on tumorigenesis.
* **Total Sample**: 20 women and 500 mice (100 per group) are proposed, with recruitment of 40 women to account for dropouts.

**Reference**

[1] Z. Yasukawa, R. Inoue, M. Ozeki, T. Okubo, T. Takagi, A. Honda, Y. Naito, Effect of Repeated Consumption of Partially Hydrolyzed Guar Gum on Fecal Characteristics and Gut Microbiota: A Randomized, Double-Blind, Placebo-Controlled, and Parallel-Group Clinical Trial, Nutr. 2019, Vol. 11, Page 2170 11 (2019) 2170. https://doi.org/10.3390/NU11092170.

[2] B. Routy, E. Le Chatelier, L. Derosa, C.P.M. Duong, M.T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M.P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J.R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M.D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors, Science 359 (2018) 91–97. https://doi.org/10.1126/SCIENCE.AAN3706.

Response to Reviewer:

“Our sample sizes (20 participants, 100 mice per group) exceed the minimum requirements from our power analysis (17 participants, 42 mice per group) for 80% power at moderate effect sizes (d = 0.5 for microbiome, d = 0.65 for breast cancer risk). This provides 85% and 95% power, respectively, ensuring robust detection of effects. Figures 1 and 2 illustrate power across various effect sizes, confirming adequacy for both co-primary endpoints, with the mouse endpoint—likely the most challenging—well-powered at n = 100 per group.”

The R code used to generate these calculations and figures is available upon request, enabling reviewers to repeat the analysis with the specified parameters.