CCI [redacted]

Microbiome Analysis Plan
-Final-

Microbiological Analyses of Oral Samples in the study titled "A randomised, controlled, examiner-blind, methodology development study to evaluate the effect of a stannous fluoride toothpaste on the oral microbiome"

(Prepared by CCI [redacted] upon request by Haleon, PLC)

Study Title: A randomised, controlled, examiner-blind, methodology development study to evaluate the effect of a stannous fluoride toothpaste on the oral microbiome

Protocol No: 300101

Sponsor: Haleon (UK) St Georges Avenue, Weybridge, Surrey, KT13 0DE, United Kingdom (UK)

Clinical Site: Salus Research, Inc. PPD

Microbiological Analysis Site: CCI [redacted]
PPD

This document contains confidentiality statements that are not relevant for this publicly available version

Document History

Document	Version	Summary of Changes
Original protocol	1.0	Not applicable (N/A)
Amendment 1		
Amendment 2		

Amendments incorporate all revisions to date, including amendments made at the request of the Sponsor or those discussed as explained elsewhere in this document.

Approval Signatures

PPD :

PPD _____

PPD

Date

Sponsor Representative

PPD _____

PPD

Date

Purpose

The CCI will perform microbiological analysis for oral samples collected from subjects with gingivitis participated in a clinical trial with a primary objective of exploring the effect of stannous fluoride - formulation on modulation of oral microbial community.

The primary endpoint measure will be the changes in the oral microbial composition. Plaque samples will be collected from four different locations (supragingival plaque, subgingival plaque, saliva, and tongue) in the oral cavity from all subjects at baseline and at the end of the study (6 weeks after baseline).

Samples will be collected per previously established sampling protocols at each time point (CCI CCI will provide the sampling protocols and Swab and Saliva Collection Kits). A total of 400 samples will be collected as shown in Table 1. Sample analysis will be performed according to the CCI standard operating procedures (SOPs) or protocols in effect during the conduct of this study.

Sample Management: The samples are listed in Table 1.

Table 1. Number of samples expected*

	Number of subjects	Number of samples/subject	Microbiome (16S RNA)	Metatranscriptomics
Baseline	50	4	200	200
Week 6	50	4	200	200
TOTAL	50	8	400	400

*Actual sample numbers may change.

Samples will be collected at the Clinical Research site (Salus Research, Inc.) per protocol, processed and frozen according to previously established sampling protocols. The frozen clinical samples will be sent on dry ice to the CCI in batches (***if storage is not possible for the entire sample size***). Shipping protocols and forms will be provided to Salus Research by the CCI Institute prior to shipment arrangements.

- After clinical collection, all samples will be processed for preservation using CCI Buffer 150-300 µl/vial, depending on sample type) and frozen at -80°C at least overnight before shipping to CCI Samples will be inventoried by the Clinical Research site and a shipping manifest matching the order of samples will be electronically provided to CCI prior to shipment. A hard copy of the manifest will be included in the shipment container. Sample vials will be labeled using the format "Subject ID.Visit No.Sample Type"
  Examples:
  1001.V1.SAL (saliva); 1001.V1.SUBG (subgingival plaque); 1001.V1.TNG (tongue); 1001.V1.SUPG (Supragingival plaque)
  (Plaque) or sites can be numbered as shown: 1001.V1.ST1; 1001.V1.ST2, ….
- Shipments will be scheduled in contact with CCI staff and a tracking number will be provided along with other necessary information.
- After receipt of samples, CCI lab staff will do an inventory of received samples cross

referencing to the manifest provided. Any discrepancy will be noted and clarified with the clinical site.

- Samples will be stored in a dedicated space of a freezer located at the Clinical and Translational Research Center at the CCI until analyses.
- Samples will be inventoried at each site--Subject>Salus-will be inventoried at Salus and Salus>CCI >destruction will be tracked and inventoried by CCI -with verification of the use of entire sample or destruction of remaining sample. Chain of custody will be documented.

At the end of the study, all study samples will be sent to the CCI contracted sequencing facility CCI Prior to submission of samples, a 5-digit code will be obtained and the necessary forms, e.g., submission form and Hazardous Materials Declaration form, will be filled out. Shipment to CCI will be documented and performed according to established protocols. The protocol used by the CCI will be followed CCI The following processes will take place at the Sequencing facility:

- After sample inventory, samples will be processed for DNA/RNA purification using a co-extraction protocol per CCI SOPs.
- DNA samples will be subjected to sequencing using species-level identification with 16S rRNA Gene Amplicon Sequencing
- RNA samples will be subjected to metatranscriptomics for microbiome gene expression analysis with functional assays. RNA extraction will be tested with a small batch of samples (optimization). If RNA is problematic, metagenomic sequencing will be performed instead of metatranscriptomics. This can be for select sites that the samples yield low RNA and for the sites with high RNA, metatranscriptomics can still be applied.

Sample Assay and Analysis Schedule*:

- 4-6 weeks for sample processing and 16S sequencing (by July 15, 2024)
- 10-12 weeks for sample processing and metatranscriptomics analysis (by Sep 3, 2024)
- Upon receipt of data, 3-4 weeks for data processing and bioinformatics at CCI (by Oct 1, 2024)
- 3-4 weeks—Statistical analysis and final report (by Oct 29, 2024)

*Based on the date of sample receipt at CCI

CCI

Appendix I: Microbiological Assay Methods

I. Microbial Profiling and Comparison

16s rRNA Gene Amplicon Sequencing

The samples will be processed and 16S sequencing will be performed by CCI Bacterial DNA extraction from each sample will be performed using a CCI DNA Kit and will be PCR-amplified and purified using AMPure beads CCI The DNA samples will be prepared for targeted sequencing with the Quick-16S™ NGS Library Prep Kit. Quick-16S™ Primer Set V1-V3 will be used for sequencing. In comparison with V3-V4 region, V1-V3 is a better region for bacterial identification, especially the streptococci. V3-V4 does not differentiate between many species of Streptococcus. Overall, V1-V3 is better, and it is also widely used in oral microbiome research. These primers were custom designed by CCI to provide the best coverage of the 16S rRNA gene while maintaining high sensitivity.

The sequencing library will be prepared using an innovative library preparation process in which PCR reactions will be performed in real-time PCR machines to control cycles and therefore limit PCR chimera formation. The final PCR products will be quantified with qPCR fluorescence readings and pooled together based on equal molarity. The final pooled library will be cleaned up with the Select-a-Size DNA Clean & Concentrator™, then quantified with TapeStation® CCI and Qubit® (CCI). The library mixture, spiked with 20% Phix, will be run on a MiSeq using a MiSeq V2 Cartridge 600 cycles (CCI). In order to utilize the low abundance reads, a quality filtering (Q25) will be conducted to ensure the highest quality.

Bioinformatics Data Analysis: CCI will send the raw 16S sequences to CCI after sequencing described above. The 16S sequences will be quality-filtered and taxonomically assigned to the species level using the state-of-the-art CCI CCI pipeline uses CCI CCI algorithm to denoise the sequence reads and then compile the microbial profile to species level for each sample. The CCI algorithm will quality-filter, trim, denoise, pair reads and remove chimera from the raw reads and generate high-quality amplicon sequence variants (ASVs). ASVs will be taxonomically assigned to species level, with an open-reference 16S rRNA NGS reads taxonomy assignment pipeline using CCI well-curated 16S rRNA reference sequence set that includes the current version (V15.23) of the HOMD references and a set of full-length 16S rRNA sequences derived from The National Center for Biotechnology Information (NCBI). This comprehensive set of 16S rRNA reference sequences represent a total of 17,035 microbial species and can identify both human-oral and non-oral, as well as potential novel species. Species level read count table will be used for a variety of downstream comparative and statistical analyses including:

1. Visual Analysis - Interactive taxonomy bar plots
2. Compositional data analysis (CoDa) - centered log ratio (clr) data transformation
3. Distance/Dissimilarity metric measurement - Aitchison distance
4. Microbial profiles - heatmap and clustering
5. Phylogenetic trees with relative abundance
6. Microbial diversity analyses - alpha and beta diversities, core microbiome analysis
7. Variance-based compositional principal component (PCA) biplot for beta-diversity exploration
8. Network Correlation analysis - SPARCC

9. Differential Abundance Analysis – AncomBC2
10. LEfSe (Linear discriminant analysis Effect Size) determines the features (organisms, clades, operational taxonomic units, genes, or functions) most likely to explain differences between classes (non-compositional test)

The above-mentioned analyses will be done as within group analysis (for both treatment groups, separately, comparing baseline and 6-week time points) and between group analysis (comparing treatment groups at both time points-baseline and 6 week).

II. Functional Profiling by Metatranscriptomic Deep Sequencing

The meta transcriptome constitutes the genes expressed by microbes within any given sample. The study of this through the sequencing and analysis of RNA can give insight into how the microbiome or isolated bacteria behaves under conditions of interest, gaining a deeper understanding of the gene expression and associated functional outcomes.

RNA extraction and Library Preparation and Sequencing: The samples were processed and analyzed with the CCI Metatranscriptomics Sequencing Service CCI Specific details for the project can be found in the final report PDF.

- RNA Extraction: One of the extraction kits CCI RNA Kit CCI or CCI RNA Miniprep Kit CCI, CCI will be used depending on the sample type and sample volume. Both kits start with mechanical lysis of microbial samples using CCI Lysis Tube (0.1 mm and 0.5 mm) to ensure efficient lysis of bacteria, archaea, and fungi.
- RNA Library Preparation: The RNA-Seq library will be prepared using the CCI Total RNA Library Kit CCI CCI with 500 ng RNA as input. All libraries will be quantified with TapeStation CCI and then pooled in equal abundance. The final pool will be quantified using qPCR.
- Sequencing: The final library will be sequenced on either the CCI or the CCI CCI

Bioinformatics Data Analyses: Initial QC, adapter trimming and preprocessing of metagenomic sequencing reads are done using CCI The quality-controlled reads are then subjected to a translated search against a comprehensive and non-redundant protein sequence database, CCI The CCI database, provided by CCI represents a clustering of all non-redundant protein sequences in CCI such that each sequence in a cluster aligns with 90% identity and 80% coverage of the longest sequence in the cluster. The mapping of metagenomic reads to gene sequences are weighted by mapping quality, coverage and gene sequence length to estimate community wide weighted gene family abundances based on an approach for species-level functional profiling of metagenomes and metatranscriptomes. Gene families are then annotated to CCI reactions (Metabolic Enzymes) to reconstruct and quantify CCI metabolic pathways in the community. Furthermore, the CCI gene families are also regrouped to Enzyme Commission enzymes, Pfam protein domains, CAZy enzymes and GO Terms in order to get an exhaustive overview of gene functions in the community. Lastly, to facilitate comparisons across multiple samples with different sequencing depths, the abundance values are normalized using total-sum scaling (TSS) normalization to produce CPM (copies per million) units, analogous to TPMs (transcript per million) in RNA-Seq analysis. The human reads will be filtered out by CCI against the

CCI This is done automatically by the pipeline we are using right after reads uploading to the server and prior to CCI process.

III. Compliance, QC, and Data Management

***Compliance**:* Raw data analysis will be performed according to Standard Operating Procedures developed at the CCI i.e., microbial identification, for Next Generation Sequencing are standard in the microbiological field as noted by many publications including by the CCI Director.

Quality Control: As stated above, sample bioinformatic analysis follows the standard bioinformatics tools available and reported in the scientific literature. The cut off standard read number is 5000 raw read pairs. After CCI denoising and pairing, there will be at least 1000 merged reads per sample. The V1V3 length is approximately 480 nt. After the read pairs are denoised by CCI and merged, there is a step in the analysis that excludes ASV lengths of less than 300 nt (merged length). An evaluable microbiome result is defined as 1) sufficient DNA/RNA extracted from the sample and 2) after sequencing, at least 5000 16S rRNA amplicon reads per sample.

Data Output**:* Sample groups will be compared using the measures described above. In comparisons of the groups and time points, ***alpha diversity will illustrate the different numbers of bacterial species (graphs); ***beta diversity*** will illustrate the overall differences in entire bacterial profiles (graphs); AncomBC2 and LEfSe will illustrate the statistically different bacterial species with respect to the comparison groups. Interpretation from these analyses will be provided in Analysis Report with appropriate graphs and plots comparing groups at each time point with respect to changes in microbiological profiles. When performing multiple pairwise comparisons, the mixed directional false discover rate (mdFDR) will be taken into account.

Excel spreadsheets of raw data with bacterial identification with respect to each group at all time points will be provided to Haleon. This data can be used for more in-depth biostatistical analysis (i.e., PERMANOVA, regression etc.).

Data Storage and Archiving Plan after Completion of the Study:

16S rRNA sequence data and analysis results: Raw 16S rRNA sequence reads will be stored and made accessible to Haleon through the CCI storage service. CCI has multiple back-up storage and redundancy strategies and data will be kept up to 6 months after the completion of the study**.** Data analyses are also provided in a bioinformatic report that is downloadable. Data interpretations will be provided in a separate report as mentioned above.

For the data uploaded to and stored in CCI the data is stored on the hub securely for 6 months after completion of study without cost. After 6 months, there will be an annual maintenance cost CCI to access the data.

Raw sequences data will also be stored on the CCI server securely up to 6 months after the completion of study. After 6 months, there will be an annual maintenance cost CCI to access the data stored on CCI

Metatranscriptomic sequence data and analysis results: Raw sequence reads and functional profiling results of the metatranscriptomic study will be stored and made accessible to Haleon through the CCI up to 6 months after the completion of the study without cost. After 6 months, there will be an annual maintenance cost CCI to access the data stored on CCI both 16S and metagenomic studies.

CCI is an online software solution that enables incredibly fast and easily accessible analysis of complex metagenomic data. This web interface gives scientists user-friendly access to numerous pipelines developed by CCI including amplicon and shotgun data analysis. In addition to having access to version-controlled, validated pipelines, the software enables continuous splicing of microbiome data for novel discoveries and appreciation of important analyses such as Principal Component Analysis, Alpha Diversity, Beta Diversity as well as significance testing between groups. The core benefits of the CCI pipeline includes but not limited to:

1. Characterize the functional potential of the microbiome community using CCI Pathways, Enzyme Commission (EC), Pfam, CAZy and Go Terms.
2. Strain-level taxonomic classification with phylogenetic inference of novel organisms meaning users is not relying on what has only been characterized in public databases. For example, if a novel strain is identified, the pipeline will call its nearest neighbor and inform users how to interpret the strain-level statistics.
3. Multi-kingdom ID & characterization with one single pipeline: Bacteria, Viruses, Phages, Fungi, Protists, Bacterial MAG's, AMR genes & Virulence Factors.
4. Precision Filters for confidence – advanced machine learning filters for differentiating signal from noise, as demonstrated in various third-party challenges, meaning that the results are less impacted by false positives.
5. Sample-type agnostic analysis. Regardless of sample origins, e.g., stool, skin, cheese, wastewater, or coral reefs, the CCI database is built agnostically to any specific sample type which gives users the flexibility to analyze multiple sample types across multiple projects over time and have the data continuity required for long-term and continuous R&D.

Shotgun Whole Genome Metagenomics (Optional or alternative to metatranscriptomics if RNA yield is low)

The samples will be processed and analyzed with the CCI Shotgun Metagenomic Sequencing Service for Microbiome Analysis CCI

DNA Extraction: One of three different DNA extraction kits will be used depending on the sample type and sample volume. In most cases, the CCI DNA Kit CCI CCI is used to extract DNA using an automated platform. In some cases, CCI CCI Kit CCI is used. For low biomass samples, such as skin swabs, the CCI Kit CCI is used as it permits for a lower elution volume, resulting in more concentrated DNA samples.

Shotgun Metagenomic Library Preparation: Genomic DNA samples will be profiled with shotgun metagenomic sequencing. Sequencing libraries will be prepared with either the KAPA™ HyperPlus

CCI

Library Preparation Kit CCI with up to 100 ng DNA input following the manufacturer's protocol using internal single-index 8 bp barcodes with TruSeq® adapters CCI CCI or the Nextera® DNA Flex Library Prep Kit CCI with up to 100 ng DNA input following the manufacturers protocol using internal dual-index 8 bp barcodes with Nextera® adapters CCI All libraries will be quantified with TapeStation® CCI CCI and then pooled in equal abundance. The final pool will be quantified using qPCR.

Sequencing: The final library will be sequenced on either the CCI or the CCI CCI

Bioinformatics Analysis: Raw sequence reads and functional profiling results of the metatranscriptomic study will be stored and made accessible to Haleon through the CCI up to 6 months after the completion of the study. CCI is an online software solution that enables incredibly fast and easily accessible analysis of complex metagenomic data. This web interface gives scientists user-friendly access to numerous pipelines developed by CCI including amplicon and shotgun data analysis. In addition to having access to version-controlled, validated pipelines, the software enables continuous splicing of microbiome data for novel discoveries and appreciation of important analyses such as Principle Component Analysis, Alpha Diversity, Beta Diversity as well as significance testing between groups. Strain-level abundance information was extracted from the Centrifuge outputs and further analyzed: (1) to perform alpha- and beta-diversity analyses; (2) to create microbial composition barplots with CCI (Caporaso et al., 2012); (3) to create taxa abundance heatmaps with hierarchical clustering (based on Bray-Curtis dissimilarity); and (4) for biomarker discovery with LEfSe (Segata et al., 2011) with default settings ($p>0.05$ and LDA effect size >2).

Statistical Analysis: Data will be provided for Microbiome Population (defined as: all randomized subjects who complete at least one use of study product and with an evaluable microbiome result obtained from at least one sampling area at both baseline and Week 6). Statistical calculations will be performed using the Python scikit-learn and StatsModels libraries. Differences in alpha diversity (Observed features and Shannon diversity) will be analyzed using a ranked one-way ANOVA followed by a Tukey honest significant difference test for multiple post-hoc comparisons. Permutational multivariate analysis of variance (PERMANOVA) will be used Bray-Curtis and Jaccard distance measures to assess microbial community compositional differences between groups (999 permutations).

Final Report: A static final report per analysis will be provided. For the 16S taxonomy analysis, the report will be provided online and can be downloaded. A sample 16S taxonomy report can be viewed here: CCI.

Functional meta-transcriptomic analysis report will be provided in both static and interactive formats. Haleon can have access to a web-based interactive tool (through CCI) and visualize the data and analytical results interactively. Reports can be downloaded for individual analysis as well as a final static report containing multiple analysis.
Reports can also be provided through a secure data sharing platform provided by the Sponsor.

CCI

CCI

Appendix II. CCI

Table 2. CCI