UC Irvine
UC Irvine Previously Published Works

Title

Metagenomic Study of the MESA: Detection of Gemella Morbillorum and Association With Coronary Heart Disease

Permalink

https://escholarship.org/uc/item/9vn715pc

Journal

Journal of the American Heart Association, 13(19)

ISSN

2047-9980

Authors

Taylor, Kent D
Wood, Alexis C
Rotter, Jerome I
et al.

Publication Date

2024-10-01

DOI

10.1161/jaha.124.035693

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Journal of the American Heart Association

ORIGINAL RESEARCH

Metagenomic Study of the MESA: Detection of *Gemella Morbillorum* and Association With Coronary Heart Disease

Kent D. Taylor, PhD; Alexis C. Wood, PhD; Jerome I. Rotter, MD; Xiuqing Guo, PhD; David M. Herrington, MD; W. Craig Johnson, MS; Wendy S. Post, MD; Russell P. Tracy, PhD; Stephen S. Rich, PhD; Shaista Malik, MD, PhD

BACKGROUND: Inflammation is a feature of coronary heart disease (CHD), but the role of proinflammatory microbial infection in CHD remains understudied.

METHODS AND RESULTS: CHD was defined in the MESA (Multi-Ethnic Study of Atherosclerosis) as myocardial infarction (251 participants), resuscitated arrest (2 participants), and CHD death (80 participants). We analyzed sequencing reads from 4421 MESA participants in the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine program using the PathSeq workflow of the Genome Analysis Tool Kit and a 65-gigabase microbial reference. Paired reads aligning to 840 microbes were detected in >1% of participants. The association of the presence of microbe reads with incident CHD (follow-up, ~18 years) was examined. First, important variables were ascertained using a single regularized Cox proportional hazard model, examining change of risk as a function of presence of microbe with age, sex, education level, Life's Simple 7, and inflammation. For variables of importance, the hazard ratio (HR) was estimated in separate (unregularized) Cox proportional hazard models including the same covariates (significance threshold Bonferroni corrected $P<6\times10^{-5}$, 0.05/840). Reads from 2 microbes were significantly associated with CHD: *Gemella morbillorum* (HR, 3.14 [95% CI, 1.92–5.12]; $P=4.86\times10^{-6}$) and *Pseudomonas* species NFACC19-2 (HR, 3.22 [95% CI, 2.03–5.41]; $P=1.58\times10^{-6}$).

CONCLUSIONS: Metagenomics of whole-genome sequence reads opens a possible frontier for detection of pathogens for chronic diseases. The association of *G morbillorum* and *Pseudomonas* species reads with CHD raises the possibilities that microbes may drive atherosclerotic inflammation and that treatments for specific pathogens may provide clinical utility for CHD reduction.

Key Words: cardiovascular heart disease ■ *Gemella morbillorum* ■ MESA ■ metagenomics ■ *Pseudomonas*

Inflammation is a well-established feature of cardiovascular disease.[1,2] The hypothesis that pathogens contribute to this inflammation has seen acceptance and rejection since the demonstration in 1979 that an avian herpesvirus caused atherosclerotic-like lesions in chicken arteries and cholesterol accumulation in chicken smooth muscle cells.[3] The subsequent observation that *Chlamydia* infection increased atherosclerosis in mouse models led to clinical trials of antibiotics targeting *Chlamydia pneumoniae*, but results were uniformly negative for reduction of cardiovascular risk. Enthusiasm for the hypothesis therefore waned as equivocal results were obtained from studies of "candidate pathogens," including *Streptococcus* species, *Human papilloma virus, Helicobacter pylori*, and the viruses Epstein-Barr, hepatitis A, *Herpes* species, and influenza (for review, see previous studies[4,5]).

Correspondence to: Kent D. Taylor, PhD, Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation, 1124 W Carson St, Torrance, CA 90502. Email: ktaylor@lundquist.org

This manuscript was sent to Julie K. Freed, MD, PhD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

For Sources of Funding and Disclosures, see page 7 and 8.

© 2024 The Author(s). Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

RESEARCH PERSPECTIVE

What Is New?

- We conducted a metagenomic analysis of whole-genome sequencing data from the MESA (Multi-Ethnic Study of Atherosclerosis). Sequencing reads that do not align with the human reference were aligned with a microbial reference with bacterial, fungal, unicellular eukaryote, and viral sequences.
- We found an association between the presence of *Gemella morbillorum* and *Pseudomonas* reads and subsequent cardiovascular heart disease.
- These organisms were also detected in samples from atheroclerotic plaque.

What Questions Should Be Addressed Next?

- Because the detection of microbes by reanalysis of sequencing data from human blood samples is currently controversial, can these findings be confirmed by other approaches?
- Can these findings be confirmed by reanalysis of sequencing data in other cohorts?

Nonstandard Abbreviations and Acronyms

MESA	Multi-Ethnic Study of Atherosclerosis

However, evidence continues to accumulate for the hypothesis that microbes contribute to inflammation in cardiovascular diseases. For example, in epidemiology studies, severe infection, defined as hospital admission for sepsis or pneumonia, was associated with increased risk for cardiovascular disease in the Swedish Military Conscription Register[6] (236 739 subjects), OptumLabs Data Warehouse (2 258 464 hospitalizations),[7] and UK BioBank (331 683 for discovery, 271 329 for replication).[8] In the laboratory, treatment of monocytes with microbes results in macrophages with an atherosclerotic phenotype as defined by cytokines and foam cell formation.[9,10] In biomarker studies, cell-free DNA in plasma has been associated with cardiovascular disease.[11,12] In addition, clinical trials of anti-inflammatory treatments, such as colchicine and anti–interleukin-1β (canakinumab), have been successful at reducing the recurrence of cardiovascular events after myocardial infarction.[13–15] These observations revive interest in the hypothesis that microbes contribute to cardiovascular disease and suggest that success with a clinical trial of antimicrobial therapy awaits identification of organisms more pathogenic for the cardiovascular system.

Although the blood has traditionally been considered to be free of microbes when healthy, microbes have been detected,[16] possible sources are the gut microbiome, particularly in dysbiosis, and the oral microbiome, particularly in periodontitis. One method for detection of microbial organisms is a "metagenomic analysis," the examination of short sequence reads from next-generation sequencing that do not align with the human reference and so are "left over" from a sequencing experiment.[17] Alignment of these "left overs" with a microbial reference file may identify known microbes,[18] and contig assembly of overlapping short reads may identify hitherto unculturable and unknown microbes.[19]

To detect microbes and test for association with incident coronary heart disease (CHD), we report a metagenomic study of short-read, whole-genome sequence data from 4424 participants of the MESA (Multi-Ethnic Study of Atherosclerosis), generated as part of the National Heart, Lung, and Blood Institute Trans-Omics of Precision Medicine program.[20]

METHODS

Data Access

We used data publicly available data from the National Center for Biotechnology Information with the following accessions: MESA, PRJNA396088; PRJNA991655[21]; and PRJNA242791.[22] We used PathSeq and the Genome Analysis Tool Kit as implemented in the BioData Catalyst without modification, except that we updated the human reference in the Genome Analysis Tool Kit bundle to GRCh38.p14 (see Metagenomics analysis, below).

Study Participants: MESA

MESA is a prospective, community-based cohort study of 6814 men and women aged 45 to 84 years, free of clinical cardiovascular disease at the time of enrollment in 2000 to 2002. Participants were recruited from 6 US regions: Baltimore, MD; Chicago, IL; Los Angeles, CA; New York NY; St. Paul, MN; and Winston-Salem, NC, and at baseline had the following sex and ethnicity (self-reported) distributions: 53% female, and 38% non-Hispanic White (self-report "White or Caucasian" and not "Spanish, Hispanic, or Latino" in 2000–2002), 28% African American, 12% Chinese, and 22% Hispanic (both Caribbean and Mexico/Central America). All participants provided written informed consent, including for genetic study. MESA has been approved by the institutional review boards of each field center, the Data Coordinating Center at the University of Washington (Seattle, WA), the Central Laboratory at the University of Vermont (Burlington, VT); and the Genetic Analysis Center and CT Reading Center at The Lundquist

Institute (Torrance, CA). The Institutional Review Board at The Lundquist Institute for Biomedical Innovation approved this specific project. Extensive details of MESA have been described.[23,24] At the baseline examination, clinical characteristics and anthropometric measurements were obtained by trained personnel using standardized protocols. A fasting blood sample was also drawn (after a minimum 8-hour fast) and stored at −80 ° C until DNA isolation. Questionnaires were administered to collect self-reported demographic data, including age, sex, race, dietary information, and health behaviors. Subsequently, participants participated in follow-up telephone calls at yearly intervals for the ascertainment of cardiovascular events, and in clinical examinations at ≈18-month intervals.

Coronary Heart Disease

CHD events were adjudicated from yearly telephone follow-up, medical records, and the National Death Index. CHD was defined by myocardial infarction, resuscitated cardiac arrest, and CHD death ("hard" CHD in the MESA protocols). Agatston score was determined by computed tomography.

Covariates

Age, sex, household income, highest education levels, and smoking status were obtained through in-person interviews with trained assessors. Participants' smoking information was categorized into current smokers compared with former/never smokers. Physical activity was assessed using a detailed, semiquantitative questionnaire adapted from the Cross-Cultural Activity Participation Study. Habitual dietary intake was assessed via a food frequency questionnaire that asked about the frequency of intake, and typical portion size, of 120 foods (including mixed dishes, such as chow mein) over the past 12 months. Height and weight were measured in duplicate by trained study staff. A mean of both measurements was used to calculate body mass index as weight in kilograms (kg) divided by height in meters (m) squared (kg/m^2). Interleukin-6 was measured via ultrasensitive ELISA (Quantikine HS Human IL-6 Immunoassay; R&D Systems, Minneapolis, MN). Health behavior and clinical information at baseline was summarized into the American Heart Association's Life's Simple 7 score; this score reflects proximity to "ideal cardiovascular health" by aggregating (1) smoking status; (2) body mass index; (3) physical activity; (4) healthy diet; (5) total cholesterol; (6) blood pressure; and (7) fasting plasma glucose.[25,26]

Metagenomic Analysis

The scoring of the presence/absence of microbes was ascertained via metagenomic analysis of whole-genome sequence data from MESA as part of the Trans-Omics for Precision Medicine program of the National Heart, Lung, and Blood Institute. Sequencing was by the Broad Institute Genomics Platform (Stacey Gabriel, principal investigator), reads transferred to the Trans-Omics of Precision Medicine Informatics Research Center (University of Michigan), and aligned to the human reference with stringent machine-learning–based quality metrics.[20] Average read length was 151 nucleotides. Quality of the sequencing data can also be seen by the mean and median read depth of 38× and the high proportion of the human genome covered with a depth of >10× (mean, 0.979; median, 0.987). Sequence files ("Cram files") were transferred from the Sequence Read Archive to a Terra/BioDataCatalyst workspace via Gen3 (The National Heart, Lung, and Blood Institute BioData Catalyst, Zenodo.).[27] The PathSeq workflow in the Genome Analysis Tool Kit, and as developed for BioData Catalyst,[28–31] was applied in 3 steps: first, to remove human reads with low complexity, low quality, and duplication; second, to align the remaining short reads to a microbial reference, consisting of 65 gigabases of sequence from archaea, bacteria, fungi, protozoa, and viral genomes; then third, to perform a taxonomic classification, matching aligned reads with known microbes. A microbial-aligned read was considered "detected" if the alignment had an identity score > 90% with both read pairs aligning to the same microbe reference fasta file in the microbial reference; these are considered "unambiguous" by the PathSeq workflow. Because of zero inflation for the population distribution of transcript read frequency for several microbes, metagenomic reads for each were converted to indicate the presence or absence of each microbial organism in the sample.

Statistical Analysis

Sample Characteristics

Demographic information and health characteristics, stratified by quartiles of total number of microbes detected via transcript reads, were calculated as total number (N) and percentage (%). Crude differences between the quartiles (ie, without controlling for covariates or correcting for multiple testing) were conducted using linear regression on transformed outcomes for continuous variables (*P* trend interpreted) and a χ^2 test for binary or other categorical traits.

Associations Between the Detected Reads and Incident CHD

All microbes with presence or absence of at least 1 read in >1% of participants were included in tests of association with incidence of CHD. Although HIV was included in our microbial reference, there were no

groups of patients with HIV in this study. Furthermore, MESA ancillary studies have indicated that <1% of participants are positive for either hepatitis B or hepatitis C, and so these microbes were not included in our analyses.

Variable Selection

Because of the prevalence of detected reads being as low as 1% for some microbes, we first selected variables of importance using a single penalized Cox proportional hazard ("survival") model. Our rationale was that penalized models can minimize the increased likelihood of false positives seen with sparse data, particularly when there are few events.[32–34] Elastic net was used to apply a penalty function to the Cox proportional hazards models, to account for the sparse nature of the data structure for some microbes. Optimal penalty parameters for the penalty value (mixing percentage; α) and the strength of the penalty (regularization penalty; λ) were ascertained via the R package "caret" using cross validation. Briefly, data in the full data set were randomly assigned to training (2/3 of participants) and test (1/3 of participants) data sets. Parameter selection was conducted via bootstrapped estimates (25 repeats) of models for all values of λ between 0 and 1 (inclusive) at intervals of 0.05. Optimization was reached via feature-wise normalization change in successive coordinate descent iterations.[35] Model performance was judged on the basis of root mean square error of approximation, with α and λ parameters giving rise to the minimum mean cross-validated error used to generate new coefficients for the association of plasma microbe transcripts with incident events.

Effect Size Estimation

As penalized models do not provide confidence estimates around associations, for all microbes selected as variables of importance in the penalized models, the association was also parameterized using standard (nonpenalized) Cox proportional hazard models, modeling how the risk of experiencing a CHD event changes as a function of the presence of each microbe over time. Each microbe selected as a variable of importance in the regularized models was included in a separate model, and all models adjusted for were age, sex, self-reported race and ethnicity, highest education (as a proxy for socioeconomic status), Life's Simple 7 score, and interleukin-6 levels (as a proxy for inflammation). Microbial transcriptome-wide significance for these models was set at a Bonferroni-corrected $P<5.95\times10^{-6}$ (0.05/840 included microbes).

Sensitivity Analyses

For microbes detected in plaque, we repeated the detected reads-event analyses, after stratifying participants on baseline Agatston score=0 or >0, to probe for the possibility that the presence of microbes followed, or occurred simultaneously with, the onset of pathology given that Agatston score provides a general adjustment for the observed confounding of coronary artery calcium.

RESULTS

Sample Characteristics

After data merge, the total number of participants in the complete data set for this report was 4421. Over an average of 15.61±3.45 follow-up person years (total person years=69 007), we observed 333 CHD events (251 myocardial infarctions, 2 resuscitated cardiac arrests, and 80 CHD deaths). The mean number of nonhuman microbes detected by metagenomic sequencing in each participant was 79.15±20.05 (range, 30–262; Table 1). The number of microbes detected differed between groups defined by self-reported race and ethnicity (X_2=30.39, df=9, $P=3.8\times10^{-4}$, Table 1), and those who went on to develop CHD (X_2=9, df=3, P=0.02). No association was found between the average number of microbes detected and baseline age, sex, total cholesterol, systolic blood pressure, high-density lipoprotein cholesterol, current smoking status, or diabetes status (all P>0.05; Table 1).

A total of 5335 (of 5379 or 99.2%) MESA "cram" files were successfully processed by the PathSeq workflow. Mean number of nonhuman reads per participant was ~350 000 of a mean of ~860 million total reads. The mean percentage of these that subsequently aligned on the microbial reference was 2.3%. The number of microbes detected by at least 1 read was 4460, and the number of microbes detected in >1% of the participants was 840.

Association With Incident CHD

Our rationale for beginning with a regularized Cox proportional hazard model is that this approach is useful for the sparse and high dimensional data in this report.[32,33] In regularized Cox proportional hazard models that included the presence/absence of transcripts for 840 microbes, along with age, sex, self-reported race and ethnicity, highest education level, Life's Simple 7 score, and interleukin-6 values, the presence of gene transcripts for 10 microbes were selected as variables of importance for predicting the likelihood of experiencing a subsequent incident CHD event (Table 2). After a Bonferroni correction for multiple testing, 2 microbes met were significant for association with incident

Table 1. Sample Characteristics at Baseline, Stratified by Quartile of Number of Microbial Species Detected

Characteristic	Quartile			
	1 (N=1172)	2 (N=1057)	3 (N=1148)	4 (N=1044)
Demographic characteristics				
Age, mean±SD, y	60.92±9.80	61.01±9.84	61.25±9.89	60.77±9.82
Sex, N (%) male	626 (53.41)	566 (53.55)	580 (50.52)	513 (49.14)
Self-reported race and ethnicity, N (%)*				
Non-Hispanic White	442 (37.71)	456 (43.15)	453 (39.46)	440 (42.15)
Chinese	201 (17.15)	128 (12.11)	141 (12.28)	113 (10.83)
African American	257 (21.93)	239 (22.61)	295 (16.99)	262 (25.10)
Hispanic	272 (23.21)	234 (22.14)	259 (22.56)	229 (21.93)
Metagenomic sequencing				
Microbial species detected, mean±SD, N	56.48±7.82	73.33±3.11	84.01±3.45	105.16±17.2
Clinical/health characteristics				
Life's Simple 7 score, mean±SD	8.79±2.07	8.61±2.13	8.60±2.05	8.68±2.06
Outcomes				
CHD event, N (%)	78 (6.66)	63 (5.96)	99 (8.62)	93 (8.91)*

*$P<0.05$ in tests of difference.

CHD: *Pseudomonas* (hazard ratio [HR]=3.32±0.34, $P=1.58\times10^{-5}$; penalized HR=1.5; Table 2), and *Gemella morbillorum* (HR=3.14±0.25, $P=4.86\times10^{-6}$; penalized HR=2.40; Table 2).

Subsequently, we considered the role of overlapping infections in incident CHD. Because we detected both *Pseudomonas* and *G morbillorum* in only 1 individual, we did not estimate the association of joint infection with these microbes on incident CHD. However, there was overlap in infections across the 10 microbes selected as variables of importance in the penalized Cox proportional hazards model. We detected none of these 10 microbes in 56% of the sample, 1 of the 10 microbes in 36%, and >1 in 8%. Thus, the number of

Table 2. Frequencies and Percentages of Individuals With the Presence of Transcripts in the Plasma at Baseline, Stratified by Those Who Did/Did Not Experience an Incident CHD Event, for All Microbial Species Selected as Variables of Importance in Regularized Models for Incident CHD, and Parameter Estimates From Nonregularized Survival Models

Species	Frequencies		Parameter estimates	
	did not develop incident CHD (N=4088)	Developed incident CHD (N=333)	HR (±95% CIs)	*P* value
Mesorhizobium species LNJC395A00	1390 (34)	127 (38)	1.23 (0.94–1.54)	0.07
Pseudomonas species NFACC19-2	58 (1.42)*	17 (5.11)*	3.32 (2.03–5.41)*	1.58×10^{-6}*
Porphyromonadaceae KA00676	44 (1.08)	10 (3.00)	3.04 (1.56–5.91)	0.001
Streptococcus species HMSC071D03	42 (1.03)	9 (2.70)	3.14 (1.61–6.11)	7.56×10^{-4}
Gemella morbillorum	70 (1.71)*	18 (4.41)*	3.14 (1.92–5.12)*	4.86×10^{-6}*
Laccaria bicolor	63 (1.54)	12 (3.60)	2.16 (1.18–3.94)	0.01
Citrobacter braakii	48 (1.17)	12 (3.60)	2.24 (1.25–4.02)	0.007
Streptomyces griseorubens	238 (5.82)	28 (8.41)	1.39 (0.94–2.04)	0.10
Prevotella species oral taxon_306	45 (1.00)	9 (2.70)	2.49 (1.28–4.84)	0.007
Buttiauxella gaviniae	52 (1.27)	11 (3.30)	1.89 (1.03–3.46)	0.04
Anaerococcus provenciensis	45 (1.00)	10 (3.00)	2.84 (1.51–5.34)	0.001

All models controlled for age, sex, highest education level, income, Lifes's Simple 7 score, and interleukin-6. CHD indicates coronary heart disease; and HR, hazard ratio.

*Significant associations ($P<1.5\times10^{-5}$ after a Bonferoni correction).

microbes present, from all microbes selected as variables of importance to incident CHD, was associated with incident CHD (HR=1.51±1.05, P=2.07×10^{-9}), with each additional microbe increasing the risk of incident CHD by 1.5 times. In comparison, when considering all microbes present in at least 1% of the sample, total microbial burden (ie, the number of microbes detected) was not significantly associated with incident CHD (HR=1.00±1.00, P=0.04).

Detection in Plaque Samples

To find additional support for the presence of *G morbillorum* and *Pseudomonas* species NFACC19-2, we applied PathSeq to data available in the Sequence Read Archive. Our rationale for repeating an analysis of these data was that the original articles did not present results at the microbe level and that we wanted to increase our chance of detection by selecting data sets with CHD-related characteristics. First, we applied PathSeq to 12 files in the Sequence Read Archive (SRA040611), composed of DNA from carotid atherosclerotic plaque obtained from endarterectomy of 7 patients after cerebral ischemia or stroke and from autopsy of 5 patients with cause of death unrelated to cardiovascular disease.[23] *G morbillorum* reads were detected in 4 of 5 of the autopsy and 2 of 7 of the endarterectomy samples. Notably, 50 *Gemella* reads were detected in 1 of the autopsy samples, a number 5 times the highest number observed in MESA participants. *Pseudomonas* species NFACC19-2 was detected in 1 autopsy sample. Second, we applied PathSeq to 27 RNA-sequencing files generated by SMART sequencing[24] (BioProject PRJNA991655; GSE236610), composed of RNA samples retrieved directly from balloons in percutaneous coronary interventions; *G morbillorum* reads were detected in 6 of 13 patients with stable CAD and 10 of 14 with acute coronary syndrome and reads from *Pseudomonas* strain NFACC19-2 were observed in 2 of 13 patients with stable CAD and 9 of 14 patients with acute coronary syndrome.

Sensitivity Analyses

We stratified analyses by Agatston score, a measure of calcium deposit in the coronary artery,[36] to probe for the possibility that atherosclerotic processes predispose individuals to infection of the putative microbes, in part driving the link between baseline infection and incident CHD. Of those with an Agaston score of 0, N=29 experienced an incident CHD event, of whom 2 showed the presence of bacteria at baseline for each of *G morbillorum* or *Pseudomonas* species NFACC19-2. This precluded an analysis of these microbes within Agaston strata individually. Of those with an Agaston score of 0 at baseline, who subsequently experienced incident CHD, N=18 showed the presence of at least 1 microbe selected as a variable of importance to incident CHD. Analysis of these data did not suggest modification of the associations (Table 3).

DISCUSSION

We conducted a metagenomic study of the short whole-genome sequencing reads from the blood samples of 4421 participants in the MESA and report a significant association between subsequent CHD events and detection of sequence reads from 2 bacteria, *G morbillorum* and *Pseudomonas* strain NFACC19-2. With the same workflow, we detected these 2 bacteria in publicly available data from endarterectomy samples and from RNA retrieved from balloons from percutaneous coronary interventions.

There are numerous reports of infection attributed to *G morbillorum*, for example, endocarditis,[37–39] osteoarticular infection,[40] necrotizing soft tissue infection,[29,30] and septic shock in the immunocompromised.[41] *G morbillorum* has been detected in 95% of oral samples (Human Oral Microbiome Database V3.1), and 1% of gut samples (the Human Microbiome Project I; microbiomedb.org); this observation suggests that the oral microbiome is the source of *G morbillorum*. The fact that the genus *Gemella* harbors virulence factors from multiple *Gemella* strains may explain inconsistencies in the designation of microbes associated with infection.[42]

The reference sequence for *Pseudomonas* species NFACC19-2 (GCF_900119125.1) has not been further assigned to a species as of this report. The lack of assignment points to some of the problems with microbial studies: (a) large numbers of microbial species are poorly characterized, (b) some have not been cultured,

Table 3. Parameters From Cox Proportional Hazard Models for the Association of Microbial Transcripts (Present Versus Absent) With Incident CHD, Stratified by Baseline Agaston Score (0 Versus >0), for All Microbes Significantly Associated With Incident CHD in the Whole Population

Variable	Agatston score=0 (N=1466)		Agatston score >0 (N=2828)	
	HR (+/– 95% CIs)	*P* value	HR (+/– 95% CIs)	*P* value
Pseud*omonas* species NFACC19-2	3.27 (1.91–5.60)	1.67×10^{-5}	6.45 (1.50–27.74)	0.01
Gemella morbillorum	4.46 (1.09–19.99)	0.04	3.05 (1.81–5.15)	2.95×10^{-5}

All models controlled for age, sex, highest education level, income, Lifes's Simple 7 score, and interleukin-6. CHD indicates coronary heart disease; and HR, hazard ratio.

and (c) and species assignment changes, making comparisons to older literature difficult. Furthermore, earlier microbiome studies used sequencing of the ribosomal region and do not have the resolution of later studies. However, the related *Pseudomonas aeruginosa* is a model organism for *Pseudomonas* infection, an opportunistic pathogen well known for infections in immunocompromised hosts, for nosocomial infections, and for its ability to develop multidrug resistance.

The strengths of this study include the high quality of data available from both MESA and Trans-Omics of Precision Medicine for >4000 subjects with median follow-up of 17 years. Furthermore, we reduced false-positive results by using the PathSeq workflow, filtering to remove low-quality and duplicate reads, low-complexity sequence, human Y chromosome sequence, and residual human reads before alignment with the microbial reference. We confined our analysis to microbes detected in >1% of subjects and with a >90% alignment of both paired reads to the same reference (designated "unambiguous" by the software). The association of the *Gemella* and *Pseudomonas* remained significant after correction for the number of microbes (840 in >1% of participants). We further detected the same microbes in sequence reads from atherosclerosis-related tissues.

A major weakness of this study is that the microbes were not detected directly by microbiological culturing but indirectly by alignment of reads with a microbial reference. The inference of microbial viability in the blood from sequence reads without the presence of overt infection or confirmatory culturing is currently controversial, and the sensitivity and specificity of metagenomics to various microbes remains under investigation. A second concern of our study is that our analytical model was applied to a small number of CHD events with 6 adjustment variables and so parameter estimates may be unstable. Therefore, these results present an association warranting further investigation and replication, and should not been seen as clinically actionable. We think replication is required, with confirmation of these results to include: (a) additional metagenomic analyses of whole-genome sequencing files in major studies with cardiovascular phenotypes (a larger sample size); (b) identification of microbes in atherosclerotic plaque by means of microbiological methods (eg, culturing and biochemical testing); and (c) once greater support has been observed, testing the effect of identified microbes in animal models of cardiovascular disease (testing whether exposure to a given microbe causes or exacerbates CHD). The third major concern is the lack of longitudinal data on infection persistence, and on factors that could link infections to CHD, such as inflammation. Thus, we have refrained from either causal or mechanistic interpretations, as is appropriate for the observational nature of our findings, and await the results from a variety of possible further studies.

Given the recent success of clinical trials of inflammation therapies in reducing cardiovascular disease, our results suggest that the metagenomic exploration of microbial diversity in well-characterized cardiovascular disease cohorts may identify pathogens and zoonoses contributing to heart disease. This exploration should include both alignment with known organisms as assembled in a microbial reference, as in this report, and assembly into contigs for detection of hitherto unknown or unculturable microbes.[43] We speculate that microbe identification and characterization by metagenomics may generate a "short list" of organisms suitable for the development of immunologic therapies for cardiovascular disease, perhaps even to clinical trials of a "vaccine," to reduce or eliminate the constant low-grade blood infection fed from the oral microbiome.

ARTICLE INFORMATION

Received June 28, 2024; accepted August 22, 2024.

Affiliations

Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation, Torrance, CA (K.D.T., J.I.R., X.G.); Department of Pediatrics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA (K.D.T., J.I.R., X.G.); United States Department of Agriculture/Agricultural Research Service (USDA/ARS) Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX (A.C.W.); Department of Internal Medicine, Wake Forest University, Winston-Salem, NC (D.M.H.); Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA (W.C.J.); Division of Cardiology, Department of Medicine, Johns Hopkins University, Baltimore, MD (W.S.P.); Department of Pathology and Laboratory Medicine and Biochemistry, Larner College of Medicine at the University of Vermont, Colchester, VT (R.P.T.); Center for Public Health Genomics, University of Virginia, Charlottesville, VA (S.S.R.); Division of Cardiology, Department of Medicine, University of California Irvine, Irvine, CA (S.M.); and Susan Samueli Integrative Health Institute, Irvine, CA (S.M.).

Acknowledgments

The authors thank the MESA (Multi-Ethnic Study of Atherosclerosis) participants who provided biological samples and data for MESA, along with the MESA investigators and staff for their valuable contributions (a full list of participating investigators are listed on the MESA website). The authors acknowledge the contributions to the development of the National Heart, Lung, and Blood Institute BioData Catalyst Powered by Terra ecosystem, and in particular wish to thank Alisa Manning for initial advice on getting started, Pamela Bretscher, for help getting the workflow to run, and Mark Walker for clarifying definitions used by PathSeq. This article has been reviewed and approved by the MESA Publications and Presentations Committee.

Sources of Funding

Analysis at The Lundquist Institute is supported in part by National Clinical Assessment and Treatment Service Clinical and Translational Science Institute (NCATS CTSI) grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756. The MESA (Multi-Ethnic Study of Atherosclerosis) projects are conducted and supported by the NHLBI in collaboration with MESA investigators. MESA was supported by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165,

N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 from the NHLBI, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Support for additional participants (MESA Family Ancillary Study) was by R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176, and NCATS Grant UL1-TR-001881. This study was also supported in part by the NHLBI contracts R01HL151855 and R01HL146860. The Trans-Omics of Precision Medicine (TOPMed) program is supported by the NHLBI. Genome Sequencing for "NHLBI TOPMed: MESA" (phs001416) was performed at Broad Genomics (contracts 3U54HG003067-13S1 and HHSN268201500014C). Centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center at the University of Michigan (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center at the University of Washington (R01HL-120393; U01HL-120393; contract HHSN268201800001I). Support for the BioData Catalyst was provided by the National Institutes of Health, NHLBI, through the BioData Catalyst program (award 1OT3HL142479-01, 1OT3HL142478-01, 1OT3HL142481-01, 1OT3HL142480-01, 1OT3HL147154). Dr Wood is funded, in part, by the United States Department of Agriculture /Agricultural Research Service (USDA/ARS) (Cooperative Agreement 58-3092-5-001). Any opinions expressed in this document are those of the authors and do not necessarily reflect the views of the NHLBI, individual BioData Catalyst Consortium members, or affiliated organizations and institutions, nor reflect the views or policies of the USDA. The mention of trade names, commercial products, or organizations does not imply endorsement from the US government.

Disclosures

None.

REFERENCES

1. Libby P. The changing nature of atherosclerosis: what we thought we knew, what we think we know, and what we have to learn. *Eur Heart J*. 2021;42:4781–4782. doi: 10.1093/eurheartj/ehab438
2. Libby P. Inflammation during the life cycle of the atherosclerotic plaque. *Cardiovasc Res*. 2021;117:2525–2536. doi: 10.1093/cvr/cvab303
3. Minick CR, Fabricant CG, Fabricant J, Litrenta MM. Atheroarteriosclerosis induced by infection with a herpesvirus. *Am J Pathol*. 1979;96:673–706.
4. VanEvery H, Franzosa EA, Nguyen LH, Huttenhower C. Microbiome epidemiology and association studies in human health. *Nat Rev Genet*. 2023;24:109–124. doi: 10.1038/s41576-022-00529-x
5. Campbell LA, Rosenfeld ME. Infection and atherosclerosis development. *Arch Med Res*. 2015;46:339–350. doi: 10.1016/j.arcmed.2015.05.006
6. Bergh C, Fall K, Udumyan R, Sjoqvist H, Frobert O, Montgomery S. Severe infections and subsequent delayed cardiovascular disease. *Eur J Prev Cardiol*. 2017;24:1958–1966. doi: 10.1177/2047487317724009
7. Jentzer JC, Lawler PR, Van Houten HK, Yao X, Kashani KB, Dunlay SM. Cardiovascular events among survivors of sepsis hospitalization: a retrospective cohort analysis. *J Am Heart Assoc*. 2023;12:e027813. doi: 10.1161/JAHA.122.027813
8. Sipilä PN, Lindbohm JV, Batty GD, Heikkilä N, Vahtera J, Suominen S, Väänänen A, Koskinen A, Nyberg ST, Meri S, et al. Severe infection and risk of cardiovascular disease: a multicohort study. *Circulation*. 2023;147:1582–1593. doi: 10.1161/CIRCULATIONAHA.122.061183
9. Leentjens J, Bekkering S, Joosten LAB, Netea MG, Burgner DP, Riksen NP. Trained innate immunity as a novel mechanism linking infection and the development of atherosclerosis. *Circ Res*. 2018;122:664–669. doi: 10.1161/CIRCRESAHA.117.312465
10. Saeed S, Quintin J, Kerstens HHD, Rao NA, Aghajanirefah A, Matarese F, Cheng S-C, Ratter J, Berentsen K, van der Ent MA, et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science*. 2014;345:1251086. doi: 10.1126/science.1251086
11. Destouni A, Vrettou C, Antonatos D, Chouliaras G, Synodinos JT, Patsilinakos S, Tzeli SK, Tsigas D, Kanavakis E. Cell-free DNA levels in acute myocardial infarction patients during hospitalization. *Acta Cardiol*. 2009;64:51–57. doi: 10.2143/AC.64.1.2034362
12. Cui M, Fan M, Jing R, Wang H, Qin J, Sheng H, Wang Y, Wu X, Zhang L, Zhu J, et al. Cell-free circulating DNA: a new biomarker for the acute coronary syndrome. *Cardiology*. 2013;124:76–84. doi: 10.1159/000345855
13. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, Fonseca F, Nicolau J, Koenig W, Anker SD, et al. Antiinflammatory therapy with Canakinumab for atherosclerotic disease. *N Engl J Med*. 2017;377:1119–1131. doi: 10.1056/NEJMoa1707914
14. Deftereos SG, Beerkens FJ, Shah B, Giannopoulos G, Vrachatis DA, Giotaki SG, Siasos G, Nicolas J, Arnott C, Patel S, et al. Colchicine in cardiovascular disease: in-depth review. *Circulation*. 2022;145:61–78. doi: 10.1161/CIRCULATIONAHA.121.056171
15. Tardif J-C, Kouz S, Waters DD, Bertrand OF, Diaz R, Maggioni AP, Pinto FJ, Ibrahim R, Gamra H, Kiwan GS, et al. Efficacy and safety of low-dose colchicine after myocardial infarction. *N Engl J Med*. 2019;381:2497–2505. doi: 10.1056/NEJMoa1912388
16. Castillo DJ, Rifkin RF, Cowan DA, Potgieter M. The healthy human blood microbiome: fact or fiction? *Front Cell Infect Microbiol*. 2019;9:148. doi: 10.3389/fcimb.2019.00148
17. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, Kawli T, Christians FC, Venkatasubrahmanyam S, Wall GD, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol*. 2019;4:663–674. doi: 10.1038/s41564-018-0349-6
18. Tan CCS, Ko KKK, Chen H, Liu J, Loh M, Chia M, Nagarajan N. No evidence for a common blood microbiome based on a population study of 9770 healthy humans. *Nat Microbiol*. 2023;8:973–985. doi: 10.1038/s41564-023-01350-w
19. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, Martin L, Neff NF, Okamoto J, Wong RJ, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. *Proc Natl Acad Sci USA*. 2017;114:9623–9628. doi: 10.1073/pnas.1707009114
20. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature*. 2021;590:290–299. doi: 10.1038/s41586-021-03205-y
21. Widlansky ME, Liu Y, Tumusiime S, Hofeld B, Khan N, Aljadah M, Wang J, Anger A, Qiu Q, Therani B, et al. Coronary plaque sampling reveals molecular insights into coronary artery disease. *Circ Res*. 2023;133:532–534. doi: 10.1161/CIRCRESAHA.123.323022
22. Mitra S, Drautz-Moses DI, Alhede M, Maw MT, Liu Y, Purbojati RW, Yap ZH, Kushwaha KK, Gheorghe AG, Bjarnsholt T, et al. In silico analyses of metagenomes from human atherosclerotic plaque samples. *Microbiome*. 2015;3:38. doi: 10.1186/s40168-015-0100-y
23. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacobs DR Jr, Kronmal R, Liu K, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156:871–881. doi: 10.1093/aje/kwf113
24. Olson JL, Bild DE, Kronmal RA, Burke GL. Legacy of MESA. *Glob Heart*. 2016;11:269–274.
25. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund K, Daniels S, Nichol G, Tomaselli GF, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic impact goal through 2020 and beyond. *Circulation*. 2010;121:586–613. doi: 10.1161/CIRCULATIONAHA.109.192703
26. Angell SY, McConnell MV, Anderson CAM, Bibbins-Domingo K, Boyle DS, Capewell S, Ezzati M, de Ferranti S, Gaskin DJ, Goetzel RZ, et al. The American Heart Association 2030 impact goal: a presidential advisory from the American Heart Association. *Circulation*. 2020;141:e120–e138. doi: 10.1161/CIR.0000000000000758
27. Ahalt S, Avillach P, Boyles R, Bradford K, Cox S, Davis-Dusenbery B, Grossman RL, Krishnamurthy A, Manning A, Paten B, et al. Building a collaborative cloud platform to accelerate heart, lung, blood, and sleep research. *J Am Med Inform Assoc*. 2023;30:1293–1300. doi: 10.1093/jamia/ocad048
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303. doi: 10.1101/gr.107524.110
29. Walker MA, Pedamallu CS, Ojesina AI, Bullman S, Sharpe T, Whelan CW, Meyerson M. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics*. 2018;34:4287–4289. doi: 10.1093/bioinformatics/bty501

30. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RGW, Getz G, Meyerson M. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol*. 2011;29:393–396. doi: 10.1038/nbt.1868
31. Heldenbrand JR, Baheti S, Bockol MA, Drucker TM, Hart SN, Hudson ME, Iyer RK, Kalmbach MT, Kendig KI, Klee EW, et al. Recommendations for performance optimizations when using GATK3.8 and GATK4. *BMC Bioinformatics*. 2019;20:557. doi: 10.1186/s12859-019-3169-7
32. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351:1–5. https://www.jstor.org/stable/26521529
33. Ambler G, Seaman S, Omar RZ. An evaluation of penalised survival methods for developing prognostic models with rare events. *Stat Med*. 2012;31:1150–1161. doi: 10.1002/sim.4371
34. Adhikary AC, Shafiqur Rahman M. Firth's penalized method in Cox proportional hazard framework for developing predictive models for sparse or heavily censored survival data. *J Stat Comput Simul*. 2021;91:445–463. doi: 10.1080/00949655.2020.1817924
35. Yuan G, Lu S, Wei Z. A line search algorithm for unconstrained optimization. *J Softw Eng Appl*. 2010;3:503–509. doi: 10.4236/jsea.2010.35057
36. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol*. 1990;15:827–832. doi: 10.1016/0735-1097(90)90282-T
37. Desai AK, Bonura EM. Multi-valvular infective endocarditis from Gemella morbillorum. *BMJ Case Rep*. 2021;14:e242093. doi: 10.1136/bcr-2021-242093
38. Shinha T. Endocarditis due to Gemella morbillorum. *Intern Med*. 2017;56:1751. doi: 10.2169/internalmedicine.56.8253
39. Gonzalez GN, Franco CD, Sinha T, Ramos EI, Bokhari SFH, Bakht D, Amir M, Javed MA, Ali K, Pineda Renté N. An emerging threat: a systematic review of endocarditis caused by Gemella species. *Cureus*. 2024;16:e58802. doi: 10.7759/cureus.58802
40. Saad E, Faris ME, Abdalla MS, Prasai P, Ali E, Stake J. A rare pathogen of bones and joints: a systematic review of osteoarticular infections caused by *Gemella morbillorum*. *J Clin Med Res*. 2023;15:187–199. doi: 10.14740/jocmr4891
41. Vasishtha S, Isenberg HD, Sood SK. Gemella morbillorum as a cause of septic shock. *Clin Infect Dis*. 1996;22:1084–1086. doi: 10.1093/clinids/22.6.1084
42. García López E, Martín-Galiano AJ. The versatility of opportunistic infections caused by *Gemella* isolates is supported by the carriage of virulence factors from multiple origins. *Front Microbiol*. 2020;11:524. doi: 10.3389/fmicb.2020.00524
43. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. Recovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–1542. doi: 10.1038/s41564-017-0012-7