Characterization of salivary microbiome in patients with pancreatic cancer

Pedro J. Torres, Erin M. Fletcher, Sean M. Gibbons, Michael Bouvet, Kelly S. Doran, Scott T. Kelley

Clinical manifestations of pancreatic cancer often do not occur until the cancer has undergone metastasis resulting in a very low survival rate. In this study, we investigated whether salivary bacterial profiles might provide useful biomarkers for early detection of pancreatic cancer. Using high-throughput sequencing of bacterial small subunit ribosomal RNA (16S rRNA) gene, we characterized the salivary microbiota of patients with pancreatic cancer and compared them to healthy patients and patients with digestive or non-digestive diseases. A total of 146 patients were enrolled at the UCSD Moores Cancer Center where saliva and demographic data were collected from each patient. Of these, we analyzed the salivary microbiome of 108 patients: 8 had been diagnosed with pancreatic cancer, 13 with pancreatic disease, 53 with digestive diseases (including cancer), 12 with non-digestive diseases (including non-digestive cancer) and 22 were classified as non-diseased (healthy) controls. Bacterial 16S rRNA sequences were amplified directly from salivary DNA extractions and subjected to high-throughput sequencing (HTS). We found significant differences in the abundances of several bacterial genera between patients with pancreatic cancer and patients in other categories, including *Porphyromonas*, *Haemophilus,* and *Campylobacter*. In particular, we found the bacterial ratio of *Leptotrichia* to *Porphyromonas* was consistently higher in pancreatic cancer patients compared to those with pancreatic disease, other digestive diseases and healthy controls. Interestingly, we did not find lower abundances of *Neisseria elongata* and *Streptococcus mitis* or higher numbers of *Granulicatella adiacens* reported previously using microarray methods. While salivary bacterial biomarkers appear promising, larger sample studies and repeat sampling of individuals are needed to confirm the validity of particular markers.

1 Characterization of the Salivary Microbiome in Patients with Pancreatic Cancer

3 Pedro J. Torres[1], Erin M. Fletcher[1], Sean M. Gibbons[2,3], Michael Bouvet[4], Kelly S. Doran[1,5],

4 Scott T. Kelley[1]

6 [1] Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego, CA

7 92182-4164, USA

8 [2] Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA

9 [3] Institute for Genomics and Systems Biology, Argonne National Laboratory, 9700, S. Cass Ave.

10 Lemont, IL 60439, USA

11 [4] Department of Surgery, University of California, 3855 Health Sciences Drive, La Jolla, CA

12 92093-0987, USA

13 [5] Department of Pediatrics, University of California San Diego School of Medicine, La Jolla,

14 CA, 92093, USA.

16 **Running Title:** Salivary Microbial Biomarkers for Pancreatic Cancer

18 **#Correspondence to:** Dr. Scott T. Kelley, Department of Biology, San Diego State University,

19 5500 Campanile Drive, San Diego, CA 92182-4164, USA; Tel: 619-594-5371, Fax: 619-594-

20 5676, skelley@mail.sdsu.edu

21 **Word count:** 3,599

23 **Total number of figures and tables:** 7

30 **Abstract**

31 Clinical manifestations of pancreatic cancer often do not occur until the cancer has undergone

32 metastasis, resulting in a very low survival rate. In this study, we investigated whether salivary

33 bacterial profiles might provide useful biomarkers for early detection of pancreatic cancer. Using

34 high-throughput sequencing of bacterial small subunit ribosomal RNA (16S rRNA) gene, we

35 characterized the salivary microbiota of patients with pancreatic cancer and compared them to

36 healthy patients and patients with other diseases, including pancreatic disease, non-pancreatic

37 digestive disease/cancer and non-digestive disease/cancer. A total of 146 patients were enrolled

38 at the UCSD Moores Cancer Center where saliva and demographic data were collected from

39 each patient. Of these, we analyzed the salivary microbiome of 108 patients: 8 had been

40 diagnosed with pancreatic cancer, 78 with other diseases and 22 were classified as non-diseased

41 (healthy) controls. Bacterial 16S rRNA sequences were amplified directly from salivary DNA

42 extractions and subjected to high-throughput sequencing (HTS). Several bacterial genera differed

43 in abundance in patients with pancreatic cancer. We found a significantly higher ratio of

44 *Leptotrichia* to *Porphyromonas* in the saliva of patients with pancreatic cancer than in the saliva

45 of healthy patients or those with other disease (Kruskal-Wallis Test; P<0.001). *Leptotrichia*

46 abundances were confirmed using real-time qPCR with *Leptotrichia* specific primers. Similar to

47   previous studies, we found lower relative abundances of *Neisseria* and *Aggregatibacter* in the

48   saliva of pancreatic cancer patients, though these results were not significant at the P<0.05 level

49   (K-W Test; P=0.07 and P=0.09 respectively). However, the relative abundances of other

50   previously identified bacterial biomarkers, e.g. *Streptococcus mitis* and *Granulicatella adiacens,*

51   were not significantly different in the saliva of pancreatic cancer patients. Overall, this study

52   supports the hypothesis that bacteria abundance profiles in saliva are useful biomarkers for

53   pancreatic cancer though much larger patient studies are needed to verify their predictive utility.

54

55
56 **Introduction**
57
58   In the United States, approximately 40,000 people die every year from pancreatic

59 adenocarcinoma, making it the fourth leading cause of cancer related death. Patients diagnosed

60 in the early stage of pancreatic cancer have a 5-year survival rate of 24%, compared to 1.8%

61 when diagnosed in the advanced stage (1). Clinical manifestations of pancreatic cancer do not

62 appear until after the cancer has undergone metastasis (2), emphasizing the need for early

63 detection biomarkers. The etiology of pancreatic cancer remains elusive, with cigarette smoking

64 being the most established risk factor (3–6), although links have also been made to diabetes (7;

65 8), obesity (9), and chronic pancreatitis (10). Recent research has also shown that men with

66 periodontal disease have a two-fold greater risk of developing pancreatic cancer after adjusting

67 for smoking, diabetes, and body mass index (11).

68   The human oral cavity harbors a complex microbial community (microbiome) known to

69 contain over 700 species of bacteria, more than half of which have not been cultivated (12).

70 Researchers have identified a core microbial community in healthy individuals (13) and shifts

71 from this core microbiome have been associated with dental carries and periodontitis (14). The

72 composition of bacterial communities in saliva seems to reflect health status under certain

73 circumstances (15), making the analysis of salivary microbiomes a promising approach for

74 disease diagnostics. A study by Mittal *et al.* (2011) found that increases in the numbers of

75 *Streptococcus mutans* and lactobacilli in saliva have been associated with oral disease prevalence

76 (16), while another study showed that high salivary counts of *Capnocytophaga gingivalis*,

77 *Prevotella melaninogenica* and *Streptococcus mitis* may be indicative of oral cancer (17).

78   A recent study by Farrell *et al.* (2012), suggested that the abundances of specific salivary

79 bacteria could be used as biomarkers for early-stage pancreatic cancer. Using the Human Oral

80   Microbe Identification Microarray (HOMIM), researchers observed decreased levels of *Neisseria*

81   *elongata* and *Streptococcus mitis* in patients with pancreatic cancer compared with healthy

82   individuals, while levels of *Granulicatella adiacens* were significantly higher in individuals with

83   pancreatic cancer (18). The HOMIM's ability to detect 300 of the most prevalent oral bacterial

84   species has made it a suitable method for assessing community profiles at the phylum level as

85   well as many common taxa at the genus level. However, the HOMIM microarray method fails to

86   detect approximately half of the bacterial species commonly present in saliva (19).

87   In this study, we applied high-throughput sequencing (HTS) of the bacterial small-

88   subunit ribosomal RNA (16S rRNA) genes to determine the salivary profiles of patients with and

89   without pancreatic cancer. The use of HTS to sequence 16S rRNA bacterial genes from entire

90   salivary microbial communities allows for a more comprehensive profile of the microbiome in

91   health and disease (20). During this study, we collected 146 saliva samples from patients at the

92   UCSD Moores Cancer Center. HTS was used to characterize the salivary microbiome of patients

93   with pancreatic cancer and compare them to patients with other diseases (including pancreatic

94   disease, non-pancreatic digestive disease/cancer and non-digestive disease/cancer) as well as

95   non-diseased (healthy) controls. This allowed us to test the hypothesis that patients with

96   pancreatic cancer may have a distinct microbial community profiles compared to non-diseased

97   controls and to other forms of digestive and non-digestive diseases. Our results demonstrated that

98   patients with pancreatic cancer had a significantly higher abundance ratio of particular bacterial

99   genera.

100

101   **Materials and Methods**

102   **Sample collection and patient information**

103   This study was approved by the University of California San Diego (UCSD) and San

104   Diego State University (SDSU) joint Institutional Review Board (IRB Approval #120101).

105   Patients recruited for the study were being clinically evaluated at the UCSD Moores Cancer

106   Center or were undergoing endoscopy procedures by UCSD Gastroenterologists in the Thornton

107   Hospital Pre-Procedure Clinic between May 2012 and August 2013. All patients were required to

108   fast for 12-hours prior to cancer evaluation and endoscopy procedures. To avoid bias during

109   enrollment, the research coordinator responsible for recruiting participants' was unaware of

110   patient diagnosis at time of sample collection. Consenting participants were provided with IRB-

111   approved consent forms, and HIPAA forms, as well as an optional, voluntary written survey in

112   which they could share relevant information about antibiotic, dental and smoking history. All

113   participants gave informed consent and their identities were withheld from the research team.

114   Each subject was free to withdraw from the study at any time. Participants were asked to give a

115   saliva sample into a 50 mL conical tube. If the amount of saliva exceeded 55 uL, 10 uL was

116   transferred into tube containing Brain-Heart Infusion media (BHI) and glycerol for future

117   culturing. The remaining saliva was broken up into 55 uL aliquots and stored in sterile cryovials.

118   Both BHI and saliva samples were then immediately stored at -80 °C until further processing.

119   Of the 146 participants, three subjects voluntarily withdrew and seven were not included

120   in the study due insufficient production of saliva (<55 uL) leaving 136 saliva samples. After

121   sample collection, the research coordinator accessed the participants' medical records

122   electronically for patient diagnosis information that was included under a novel subject ID

123   number. Diagnosis was used to determine health status and assess the stage of disease when each

124   sample was taken. The various diagnoses were grouped into the following categories: pancreatic

125   cancer, other disease (including pancreatic disease, non-pancreatic digestive disease/cancer and

126   non-digestive disease/cancer), and healthy (non-diseased) controls. Healthy individuals were

127   defined as participants with no documented chronic digestive or non-digestive disease, and a 5-

128   year resolution of any previously documented digestive or non-digestive disease. Exclusion

129   criteria included participants undergoing active chemotherapy or radiation therapy or use of

130   antibiotics two weeks prior to saliva collection as well as invasive surgery in the past year.

131

132   **DNA Isolation, PCR and 16S rRNA Sequencing**

133   Bacterial DNA was extracted directly from 50 uL of patient saliva using the MoBio

134   PowerSoil DNA Extraction Kit (Catalogue 12888-05, Mo Bio Laboratories, Carlsbad, CA, USA)

135   following the manufacturer's protocol. Genomic DNA was quantified using the NanoDrop™

136   Spectrophotometer and stored at -20 °C.

137   The 16S ribosomal RNA (rRNA) amplicon region was amplified using barcoded 'universal'

138   bacterial primer 515F (5'- AATGATACGGCGACCACCGAGATCTACAC TATGGTAATT

139   GT GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-

140   CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXXX AGTCAGTCAG CC

141   GGACTACHVGGGTWTCTAAT-3') (X's indicate the location of the 12-bp barcode) with

142   Illumina adaptors used by the Earth Microbiome Project ([http://www.earthmicrobiome.org/emp-](http://www.earthmicrobiome.org/emp-standard-protocols/16s)

143   [standard-protocols/16s](http://www.earthmicrobiome.org/emp-standard-protocols/16s)). The barcoded primers allow pooling of multiple PCR amplicons in a

144   single sequencing run. PCR was carried out using the reaction conditions outlined by the Earth

145   Microbiome Project. Thermocycling parameters were as follows: 94°C for 3 minutes

146   (denaturing) followed by amplification for 35 cycles at 94°C for 45 s, 50°C for 60 s and 72°C for

147   90 s, and a final extension of 72°C for 10 minutes (21). PCR amplicons were then sequenced on

148   the Illumina MiSeq platform at the Argonne National Laboratory Core sequencing facility

149   (Lemont, IL).

150

151   **Sequence Analysis**

152   16S rRNA sequences were de-multiplexed using the Quantitative Insights Into Microbial

153   Ecology (QIIME v. 1.8.0, http://www.qiime.org) pipeline. Sequences were grouped into

154   operational taxonomic units (OTUs) at 97% sequence similarity using the Greengenes reference

155   database. OTUs that did not cluster with known taxa at 97% identity or higher in the database

156   were clustered *de novo* (UCLUST (22)). Representative sequences for each OTU were then

157   aligned using PyNast (23), and taxonomy was assigned using the RDP classifier (Version 2.2)

158   (24). A phylogenetic tree was built using FastTree (25). Before performing downstream analysis,

159   patient samples were rarefied to 10,0000 sequences per sample, singletons and OTUs present in

160   <25% of samples were removed prior to rarefaction. Chimeric sequences were identified using

161   ChimeraSlayer in QIIME, as well as with DECIPHER (26), and subsequently removed. Alpha

162   diversity metrics were computed using QIIME. Beta diversity distance between samples

163   (weighted and unweighted UniFrac) were computed and used to account for both differences in

164   relative abundance of taxa and phylogeny (27). Beta diversity comparisons were done using

165   analysis of similarities (ANOSIM). We also tested whether there were significant differences in

166   abundance ratios of particular genera between our different categories with GraphPad Prism

167   version 6.0 using the Kruskal-Wallis test followed by Dunn's multiple comparison correction.

168   Statistical significance was accepted at a $p < 0.05$. Analysis and identification of potential

169   contaminants was done using SourceTracker (28).

170   **Quantitative PCR (qPCR)**

171   *Leptotrichia* abundance was determined using qPCR. Briefly, for each sample we

172   estimated *Leptotrichia* abundance using *Leptotrichia* specific 16S primers and normalized their

173   values to overall bacterial abundance estimated using qPCR with universal bacterial 16S primers

174   (5′-TCCTACGGGAGGCAGCAGT-3′ forward primer, and 5′-

175   GGACTACCAGGGTATCTAATCCTGTT-3′ reverse primer) developed by Nadkarni et al.

176   (2002) (29). qPCR was performed on a Bio-Rad CFX96 Touch ™ Real-Time PCR Detection

177   Instrument. The maximum C_t (threshold cycle) for the universal 16S primers was set to 35 cycles

178   and C_t levels above this threshold were considered background noise. Genus-specific primers for

179   amplification of *Leptotrichia* were designed using 16S rRNA sequences obtained from the RDP

180   classifier (Version 2.2) (24). Primer3 online software was used for primer selection, and

181   conditions were settled following the recommendations of Thorton and Basu, 2011 (30). The

182   *Leptotrichia* forward primer sequence (5'- GGAGCAAACAGGATTAGATACCC - 3') and the

183   *Leptotrichia* reverse primer sequence (5'- TTCGGCACAGACACTCTTCAT- 3') generated an

184   amplicon of 87 bp. The PCR reaction contained 1 uM of both forward and reverse *Leptotrichia*

185   primers with thermocycling parameters of 50°C for 2 min, 95°C for 10 min and 40 cycles of

186   95°C for 15 sec and 62.5°C for 1 min. The amplification reactions for the universal primers and

187   *Leptotrichia* primers were carried out in at least duplicate using 25 uL of SYBR Green Master

188   Mix (Bio-Rad) and 0.85 ng/uL of extracted DNA as template. Various online tools, including *In*

189   *silico* PCR Amplification (31) and Ribosomal Database Project (24) were used to check the

190   specificities of the oligonucleotide primer sequences for the target organism. A saliva sample

191   was sequenced (Eton Bioscience, San Diego, CA) using our novel primers and primer specificity

192   was further confirmed with a 16S rRNA database BlastN search.

Results

Salivary microbial diversity profiles were generated for a total of 108 patients. 8 patients were diagnosed with pancreatic cancer (P), 78 were diagnosed with other diseases (including cancer) (O), and 22 were considered healthy (non-diseased) controls (H). Table 1 details the individual clinical characteristics, including, gender and ethnicity. Of the 108 patients, 23 patients in pancreatic, digestive, and non-digestive disease categories were diagnosed with having cancer. Table 2 details the types of cancer, as well as category groupings and the mean age of the cancer patients in each category.

Illumina sequencing yielded approximately 6.8 million sequences across all samples. The sequences are available on FigShare (http://dx.doi.org/10.6084/m9.figshare.1422174) along with the mapping file (http://dx.doi.org/10.6084/m9.figshare.1422175). An analysis of potential sample contamination using SourceTracker (28) identified some evidence of human skin and/or environmental contamination. The sequences associated with OTUs identified as contaminants, mostly *Staphylococcus* (skin) and *Cyanobacteria* (chloroplasts), were removed from all subsequent analyses. From these data, we identified a total of 12 bacterial phyla and 139 genera. Proteobacteria, Actinobacteria, Bacteroidetes, Firmicutes, and Fusobacteria were the 5 major phyla, accounting for 99.3% of oral bacteria (Figure 1). The mean relative abundance of Proteobacteria was lower in pancreatic cancer patients relative to other sample categories, while Firmicutes tended to be higher, though these were not significant after adjusting for multiple comparisons (FDR). The pancreatic cancer group also had higher levels of *Leptotrichia,* as well as lower levels of *Porphyromonas,* and *Neisseria* (Figure 2). In general, multi-level taxonomic profiles of the healthy group resembled the 'other' disease group, while the pancreatic cancer group was readily distinguishable (Fig. S1). However, there were no significant differences

217 among the three main groupings (H, O, and P) in either beta diversity (ANOSIM; P=0.1) or

218 alpha diversity (Chao1, K-W test; P=0.6; Faith's PD, K-W test; P=0.56).

219   As in previous studies by Farrell *et al.* (2012) and Lin *et al.* (2013) (32), we saw lower

220 relative abundances of *Neisseria* and *Aggregatibacter*, although these differences were not

221 significant (K-W test; P=0.07 and P=0.09 respectively). *Bacteriodes* was more abundant in

222 pancreatic cancer patients compared to healthy individuals, similar to what Lin *et al.* observed,

223 although this too was not significant (K-W test; P= 0.27). We did not see any difference in the

224 relative abundance of *Streptococcus* or *Granulicatella*, which were shown to differ in a prior

225 pancreatic cancer study (18). Additional analytical targets were based on a preliminary study

226 consisting of our first 61 saliva samples (including 3 from pancreatic cancer patients), which

227 showed significantly higher *Leptotrichia* and lower *Porphyromonas* in pancreatic cancer patient

228 saliva.

229   The abundance ratio of *Leptotrichia*, specifically two OTUs (arbitrarily named OTU

230 31235 and OTU 4443207), to *Porphyromonas* was significantly higher in pancreatic cancer

231 patients (Figure 3). A BLAST comparison of these OTUs to the 16S sequence in the Human Oral

232 Microbiome database (33) (HOMD RefSeq Version 13.2) found OTU 31235 to be 100% similar

233 to *Leptotrichia sp. Oral taxon 221,* while OTU 4443207 was 99.3% similar to *Leptotrichia*

234 *hongkongensis*. We found a strong positive correlation (Pearson's correlation r=0.903,

235 P=0.0000001) between *Leptotrichia* abundances obtained from 16S rRNA sequencing (OTU

236 relative abundances) and from real-time qPCR (Figure 4).

238   **Discussion**

239   Our analysis of salivary microbial profiles supports prior work suggesting that salivary

240   microbial communities of patients diagnosed with pancreatic cancer are distinguishable from

241   salivary microbial communities of healthy patients or patients with other diseases, including non-

242   pancreatic cancers. At the phylum level, pancreatic cancer patients tended to have higher

243   proportions of Firmicutes and lower proportions of Proteobacteria (Figure 1). At finer taxonomic

244   levels, we observed differences in the mean relative abundances of particular genera in

245   pancreatic cancer patients compared to other patient groups (Figure 2). For instance, there was a

246   higher proportion of *Leptotrichia* in pancreatic cancer patients, while the proportion of

247   *Porphyromonas* and *Neisseria* were lower in these patients.

248   The most striking difference we found between the microbial profiles of pancreatic

249   cancer patients and other patient groups was in the ratio of the bacterial genera *Leptotrichia* and

250   *Porphyromonas* (LP ratio) (Figure 3). The LP ratio had been identified as a potential biomarker

251   from a preliminary analysis and an analysis of the full dataset found significantly higher LP ratio

252   in pancreatic cancer patient saliva than in other patient groups. To verify these differences using

253   another method, we cross-validated the relative abundances of *Leptotrichia* (Figure 4).

254   Interestingly, during the analysis of the 16S rRNA data, we successfully used the LP ratio to

255   reclassify one of the patients in the non-pancreatic cancer disease group. This particular

256   individual had been initially diagnosed as having an unknown digestive disease, but the patient's

257   high LP ratio suggested pancreatic cancer (Figure 3). Subsequently, the patient was re-evaluated

258   and diagnosed with pancreatic cancer, supporting the notion that the LP ratio may serve as a

259   pancreatic cancer biomarker.

260   Despite the small cohort of patients in this study, we believe our results are especially

261   noteworthy because we were able to distinguish between patients with pancreatic cancer and

262   patients with a variety of other diseases (including non-pancreatic cancer), in addition to healthy

263   controls. Other researchers have proposed the use of ratios of bacterial taxa previously.

264   Galimanas *et al.* (2014) suggested using salivary bacteria abundance ratios as a means for

265   differentiating between healthy and diseased patients (34). Taxonomic ratios have been used to

266   differentiate between subjects in studies of obesity (35), diabetes (36), and periodontal disease

267   (37). Ratio comparisons also help to control for high levels of taxonomic variability among

268   individuals (38–41).

269   A review of the literature revealed that *Leptotrichia*'s role in oral health remains elusive.

270   However, these bacteria have been found in the bloodstream of immune-compromised patients

271   (42) and co-occur significantly with colorectal tumors (43). *Leptotrichia* have been isolated from

272   cardiovascular and gastrointestinal abscesses, from systemic infections, and are thought to be

273   pathogenic (44). In regards to *Porphyromonas*, antibodies to *Porphyromonas gingivalis* have

274   been directly associated with pancreatic cancer (45). A European cohort study measured plasma

275   antibodies to 25 oral bacteria in pre-diagnostic blood samples from 405 pancreatic cancer

276   patients and 416 matched controls and found a >2-fold increase in risk of pancreatic cancer

277   among those with higher antibody titers to a pathogenic strain of *P. gingivalis* (45). At first

278   glance, it appears contradictory that individuals with higher *Porphyromonas* antibody titers

279   would have lower oral abundances. However, studies of systemic immunization of animals to

280   particular periopathogens including *Porphyromonas* have shown reduced colonization of these

281   bacteria in the mouth and a reduction of periodontitis (46–48). Similarly, higher *Porphyromonas*

282   antibody titers in individuals with pancreatic cancer may decrease their oral abundance, though

283   this connection needs to be formally tested.

284   Shifts in salivary microbial diversity could also be a systematic response to pancreatic

285   cancer. Pancreatic cancer is known to weaken the immune system (49), which could lead to

286   overgrowth of oral bacteria and a shift towards systemically invasive periodontal pathogens. The

287   proliferation of bacterial pathogens could assist cancer progression through systemic

288   inflammation (50) or immune distraction (51). Thus, an initial increase in *Porphyromonas* might

289   be followed by a decrease due to systemic invasion and antibody production. Indeed,

290   inflammation is thought to play a significant role in the development of pancreatic cancer (52).

291   We also compared the relative abundances of several other bacterial genera that were

292   indicated as potential biomarkers in previous work by Farrell *et al.* (2012). Like Farrell *et al.*, we

293   found a lower proportion of *Neisseria* in pancreatic cancer patient saliva compared with the

294   healthy and other disease category, though this trend was not significant. However, we did not

295   find the same results as Farrell *et al.* for the other bacterial genera they identified. Our data also

296   showed an increase in *Bacteroides* and decrease in the abundance of the bacterial genus

297   *Aggregatibacter* in patients with pancreatic cancer, supporting the results of a pilot study by Lin

298   *et al.* (2013), though neither trends were significant.

299   Methodological differences between our study and the Farrell *et al.* study in particular,

300   may partially explain our divergent results. For instance, the inability of the V4 region of the16S

301   rRNA gene to discriminate *Streptococcus mitis* from other *Streptococcus* species may have

302   prevented us from detecting difference in this species' abundance (18). Additionally, our study

303   had a broader array of patient categories and cancers were not always confined to the pancreas at

304   the time of sampling.

305   Interestingly, since the completion of our study, Mitsuhashi *et al.* (2015) reported the

306   detection of oral *Fusobacterium* in pancreatic cancer tissue (53). A retrospective review of our

307   abundance data also found a lower relative abundance of *Fusobacterium* in pancreatic cancer

308   patients compared to other patient categories (Figure 2; K-W test, P=0.03 prior to FDR

309   correction) suggesting the processes driving differences in *Fusobacterium* may be similar to our

310   proposed mechanism for *Porphyromonas*. Although the result was not significant after adjusting

311   for multiple-comparisons (FDR), we suggest *Fusobacterium* abundance should be considered as

312   a potential biomarker target for future studies with larger patient cohorts.

313   Overall, our study suggests that members of the salivary microbiome have promise as

314   potential pancreatic cancer biomarkers and we may have uncovered an important new prospect

315   in this regard (i.e., the LP ratio). However, our relatively small number of samples from

316   pancreatic cancer patients and the discrepancies between our findings and previous work indicate

317   that much larger patient cohorts will be needed to determine whether salivary biomarkers are

318   diagnostically useful. Future studies should focus on improved metadata collection, including

319   diet and oral health information (i.e. periodontal disease), which would make it possible to run

320   statistical analyses that control for multiple factors involved in shaping oral microbial diversity.

321   It will also be important to sample the same individual's saliva over time to assess whether we

322   can distinguish between disease stages and also to control for intra-individual variation. Further,

323   it is possible that single biomarkers may never be able to consistently identify pancreatic patients

324   from other conditions. Thus, we may need more complex metrics that combine the abundances

325   of multiple salivary bacteria, metabolite profiles, and detailed patient metadata. Effective

326   diagnostic biomarkers for pancreatic cancers have been difficult to find, but are sorely needed

327   and have the potential to save thousands of lives each year.

328

329 **Acknowledgements**

330 We thank S. Owens and J. Marcel for their help in our sequencing runs. We also thank M.

331 Wachter for assistance in enrolling patients into the study. Thanks to all the patients who

332 volunteered to be a part of this study.

333

334

335 **References:**

336 1.   Li D, Xie K, Wolff R, Abbruzzese JL. Pancreatic cancer. *Lancet* 2004;**363**:1049-57.

337 2.   Holly EA, Chaliha I, Bracci PM, Gautam M. Signs and symptoms of pancreatic

338 cancer: a population-based case-control study in the San Francisco Bay area. *Clin Gastroenterol*

339 *H* 2004;**2**:510-17.

340 3.   Vrieling A, Bueno-de-Mesquita HB, Boshuizen HC, Michaud DS, Severinsen

341 MT, Overvad K, Olsen A, Tjønneland A,Clavel-Chapelon F, Boutron-Ruault MC, Kaaks

342 R, Rohrmann S, Boeing H, Nöthlings U, Trichopoulou A, Moutsiou E, Dilis V, Palli D, Krogh

343 V, Panico S, Tumino R, Vineis P, van Gils CH, Peeters PH, Lund E, Gram IT, Rodríguez

344 L, Agudo A,Larrañaga N, Sánchez MJ, Navarro C, Barricarte A, Manjer J, Lindkvist B, Sund

345 M, Ye W, Bingham S, Khaw KT, Roddam A, Key T, Boffetta P, Duell EJ, Jenab M, Gallo

346 V, Riboli E.Cigarette smoking, environmental tobacco smoke exposure and pancreatic cancer

347 risk in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer*

348 2010;**126**:2394–2403.

349 4.   Nakamura K, Nagata C, Wada K, Tamai Y, Tsuji M, Takatsuka N, Shimizu H. Cigarette

350 Smoking and Other Lifestyle Factors in Relation to the Risk of Pancreatic Cancer Death: A

351   Prospective Cohort Study in Japan. *Jpn J Clin Oncol* 2011;**41**:225–231.

352   5.   Fuchs CS, Colditz GA, Stampfer MJ. A prospective study of cigarette smoking and the

353   risk of pancreatic cancer. *Arch Intern Med* 1996;**156**:2255-60.

354   6.   Zheng W, McLaughlin JK, Gridley G, Bjelke E. A cohort study of smoking, alcohol

355   consumption, and dietary factors for pancreatic cancer (United States). *Cancer Cause Control*

356   1993;**4**:477-82.

357   7.   Haugvik S-PP, Hedenström P, Korsæth E, Valente R, Hayes A, Siuka D, Maisonneuve

358   P, Gladhaug IP, Lindkvist B,Capurso G.Diabetes, Smoking, Alcohol and Family History of

359   Cancer as Risk Factors for Pancreatic Neuroendocrine Tumors: A Systematic Review and Meta-

360   Analysis. *Neuroendocrinology* 2015 Jan 22. [Epub ahead of print].

361   *8.*   Liu X, Hemminki K, Försti A, Sundquist K, Sundquist J, Ji J. Cancer risk in patients with

362   type 2 diabetes mellitus and their relatives. *Int J Cancer* 2015 Jan 21. [Epub ahead of print].

363   9.   Bracci PM. Obesity and pancreatic cancer: overview of epidemiologic evidence and

364   biologic mechanisms. *Mol Carcinogen* 2012;**51**:53-63.

365   10.   Malka D, Hammel P, Maire F, Rufat P, Madeira I. Risk of pancreatic adenocarcinoma

366   in chronic pancreatitis. *Gut* 2002;**51**:849-52.

367   11.   Michaud DS, Joshipura K, Giovannucci E, Fuchs CS. A prospective study of periodontal

368   disease and pancreatic cancer in US male health professionals. *J Natl Cancer I* 2007;**99**:171–5.

369   12.   Aas JA, Paster BJ, Stokes LN, Olsen I. Defining the normal bacterial flora of the oral

370   cavity. *J Clin Microbiol* 2005;**43**:5721-32.

371   13.   Zaura E, Keijser B, Huse SM. Defining the healthy "core microbiome" of oral microbial

372   communitites. *BMC Microbiol* 2009;**9**:259.

373   14.   Berezow AB, Darveau RP. Microbial shift and periodontitis. *Periodontol 2000*

374   2011;**55**:36-47.

375   15.   Yamanaka W, Takeshita T, Shibata Y, Matsuo K. Compositional stability of a salivary

376   bacterial population against supragingival microbiota shift following periodontal therapy. *PLoS*

377   *One* 2012;**7**:e42806.

378   16.   Mittal S, Bansal V, Garg S, Atreja G, Bansal S. The diagnostic role of Saliva-A Review.

379   *J Clin Exp Dent* 2011;**3**:e314–20.

380   17.   Mager D, Haffajee A, Devlin P, Norris C, Posner M, Goodson J. The salivary microbiota

381   as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and

382   oral squamous cell carcinoma subjects. *J Transl Med* 2005;**3**:27.

383   18.   Farrell J, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, Paster BJ, Joshipura K, Wong

384   DT. Variations of oral microbiota are associated with pancreatic diseases including pancreatic

385   cancer. *Gut* 2012;**61**:582–8.

386   19.   Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z. Oral microbiome profiles: 16S

387   rRNA pyrosequencing and microarray assay comparison. *PLoS One* 2011;**6**:e22788.

388   20.   Kuczynski J, Lauber CL, Walters WA. Experimental and analytical tools for studying

389   the human microbiome. *Nat Rev Genet* 2011;**13**:47-58.

390   21.   Caporaso J, Lauber C, Walters W, Lyons-Berg D, Lozupone CA, Turnbaugh PJ, Fierer

391   N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per

392   sample. *P Natl Acad Sci USA* 2011;**108**:4516–22.

393   22.   Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*

394   2010;**26**:2460-1.

395   23.   Caporaso JG, Bittinger K, Bushman FD. PyNAST: a flexible tool for aligning

396   sequences to a template alignment. *Bioinformatics* 2010;**26**:266-7

397   24.   Cole JR, Chai B, Marsh TL, Farris RJ. The Ribosomal Database Project (RDP-II):

398   previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy.

399   *Nucleic Acids Res* 2003;**31**:442-3

400   25.   Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with

401   profiles instead of a distance matrix. *Mol Biol Evol* 2009;**26**:1641–50.

402   26.   Wright ES, Yilmaz LS, Noguera DR. DECIPHER, A Search-Based Approach to Chimera

403   Identification for 16S rRNA Sequences. *Appl Environ Microbiol* 2012;

404   doi:10.1128/AEM.06516-11.

405   27.   Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing

406   high-throughput microbial community data. *Gigascience* 2013;**2**:16.

407   28.   Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman

408   FD, Knight R, Kelley ST. Bayesian community-wide culture-independent microbial source

409   tracking. *Nat Methods* 2011;**8**:761–3.

410   29.   Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by

411   real-time PCR using a broad-range (universal) probe and primers set. *Microbiol-UK*

412   2002;**148**:257–66.

413   30.   Thornton B, Basu C. Real-time PCR (qPCR) primer design using free online software.

414   *Biochem Mol Biol Edu* 2011;**39:**145–54.

415   31.   Bikandi J, San Millán R, Rementeria A, Garaizar J. In silico analysis of complete

416   bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* 2004;**20**:798–

417   9.

418   32.   Lin IH, Wu J, Cohen, SM, Chen C, Bryk D, Marr M, Melis M, Newman E, Patcher HL,

419   Alekseyenko AV, Hayes RB, Ahn J. Abstracts 101: Pilot study of oral microbiome and risk of

420   pancreatic cancer. *Cancer Res* 2013; doi: 10.1158/1538-7445.AM2013-101.

421   33.   Chen T, Yu W-HH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human

422   Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic

423   and genomic information. *Database* 2010;**2010**:baq013.

424   34.   Galimanas V, Hall MW, Singh N, Lynch MD, Goldberg M, Tenenbaum H, Cvitkovitch

425   DG, Neufeld JD,Senadheera DB. Bacterial community composition of chronic periodontitis and

426   novel oral sampling sites for detecting disease indicators. *Microbiome* 2014;**2**:32.

427   35.   Lazarevic V, Whiteson K, Gaïa N, Hernandez D, Farinelli L, Osterås M, François

428   P, Schrenzel J.Analysis of the salivary microbiome using culture-independent techniques. *J Clin*

429   *Bioinforma* 2012;**2**:4.

430   36.   Zhang Y, Zhang H. Microbiota associated with type 2 diabetes and its related

431   complications. *Food Sci Hum Wellness* 2013;**2**:167-72.

432   37.   Moolya NN, Shetty A, Gupta N, Gupta A, Jalan V, Sharma R. Orthodontic bracket

433   designs and their impact on microbial profile and periodontal disease: A clinical trial. *J Orthod*

434   *Sci* 2014;**3**:125–31.

435   38.   Ding T, Schloss PD. Dynamics and associations of microbial community types across

436   the human body. *Nature* 2014;**509**:357-60.

437   39.   Grice EA, Segre JA. The human microbiome: our second genome. *Annu Rev Genom*

438   *Hum G* 2012;**13**:151-70.

439   40.   Schwarzberg K, Le R, Bharti B, Lindsay S, Casaburi G, Salvatore F, Saber MH,

440   Alonaizan F, Slots J, Gottlieb RA, Caporaso JG, Kelley ST. The Personal Human Oral

441   Microbiome Obscures the Effects of Treatment on Periodontal Disease. *PLoS One*

442   2014;**9**:e86708.

443   41.   Wang J, Qi J, Zhao H, He S, Zhang Y, Wei S, Zhao F. Metagenomic sequencing

444   reveals microbiota and its functional potential associated with periodontal disease. *Sci Rep*

445   2013;**3**:1843.

446   42.   Eribe E, Olsen I. Leptotrichia species in human infections. *Anaerobe* 2008;**14**:131–7.

447   43.   Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe

448   E, Holt RA. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*

449   2013;**1**:16.

450   44.   Han YW, Wang X. Mobile Microbiome Oral Bacteria in Extra-oral Infections and

451   Inflammation. *J Dent Res* 2013;**92**:485-91.

452   45.   Michaud DS, Izard J, Wilhelm-Benartzi CS, You DH, Grote VA, Tjønneland A, Dahm

453   CC, Overvad K, Jenab M, Fedirko V,Boutron-Ruault MC, Clavel-Chapelon F, Racine A, Kaaks

454   R, Boeing H, Foerster J, Trichopoulou A, Lagiou P, Trichopoulos D, Sacerdote C, Sieri S, Palli

455   D, Tumino R, Panico S, Siersema PD, Peeters PH, Lund E, Barricarte A, Huerta JM, Molina-

456   Montes E, Dorronsoro M, Quirós JR, Duell EJ, Ye W, Sund M, Lindkvist B, Johansen D, Khaw

457   KT, Wareham N, Travis RC,Vineis P, Bueno-de-Mesquita HB, Riboli E. Plasma antibodies to

458   oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. *Gut*

459   2012;**62**:1764-70.

460   46.   Evans RT, Klausen B, Sojar HT, et al. Immunization with Porphyromonas (Bacteroides)

461   gingivalis fimbriae protects against periodontal destruction. *Infect Immun* 1992;**60**:2926–35.

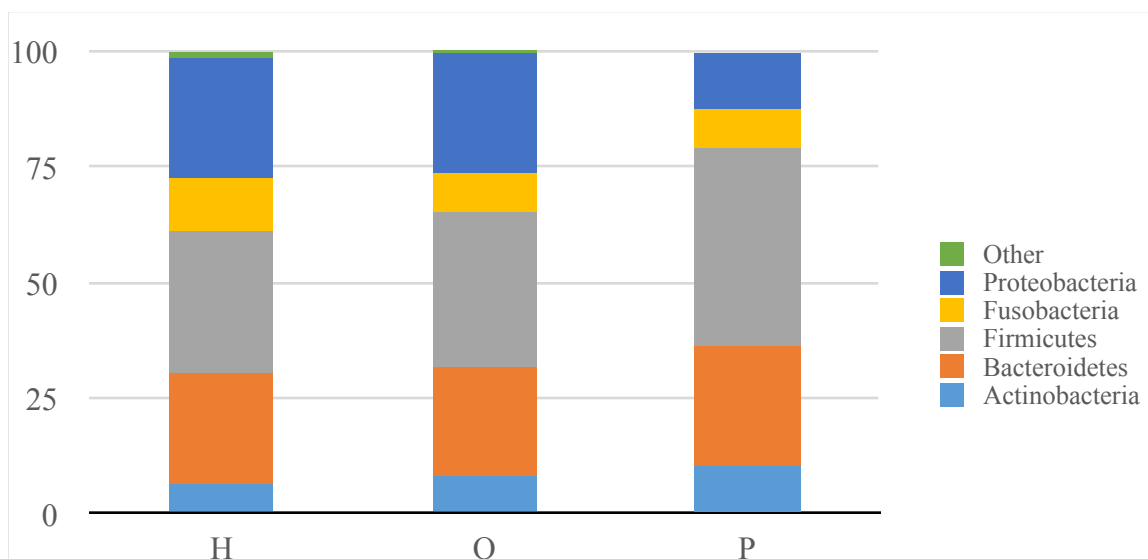
462   47.   Persson GR, Engel D, Whitney C, Darveau R, Weinberg A, Brunsvold M, Page RC.

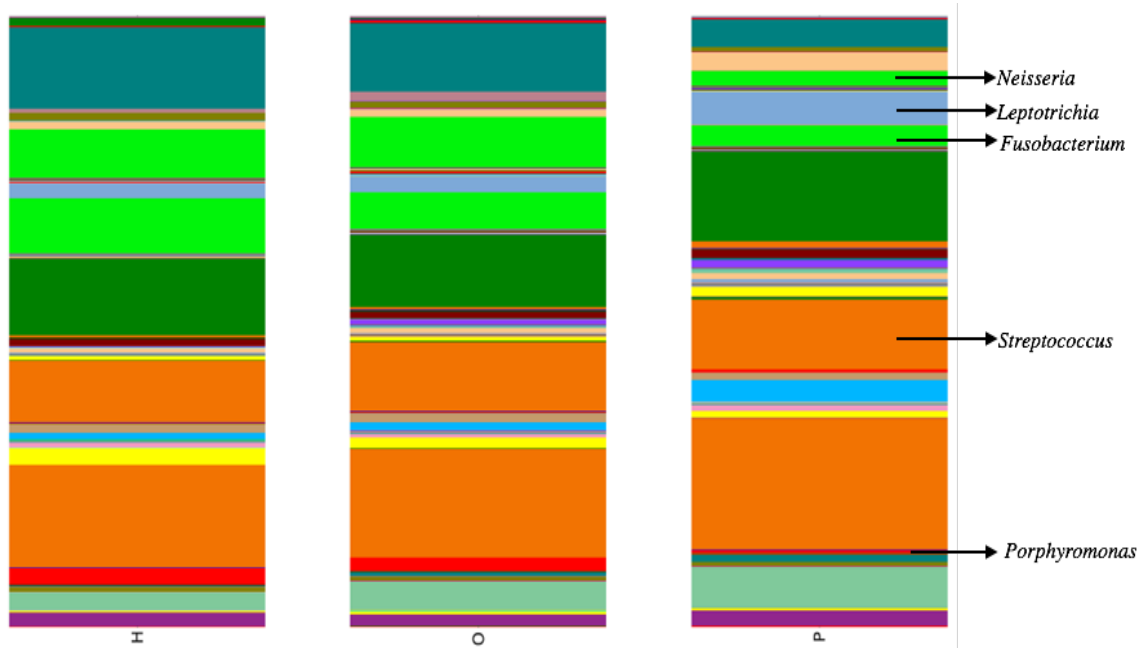
463   Immunization against Porphyromonas gingivalis inhibits progression of experimental

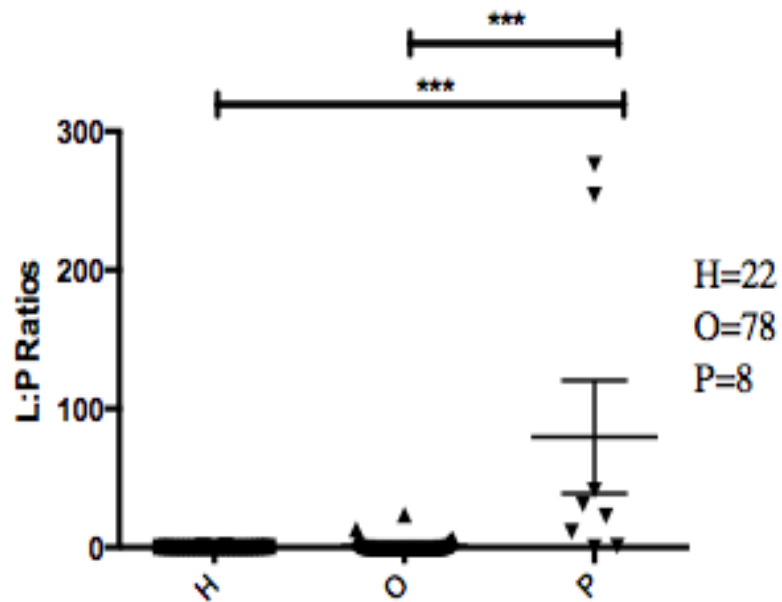
464   periodontitis in nonhuman primates. *Infect Immun* 1994;**62**:1026–31.

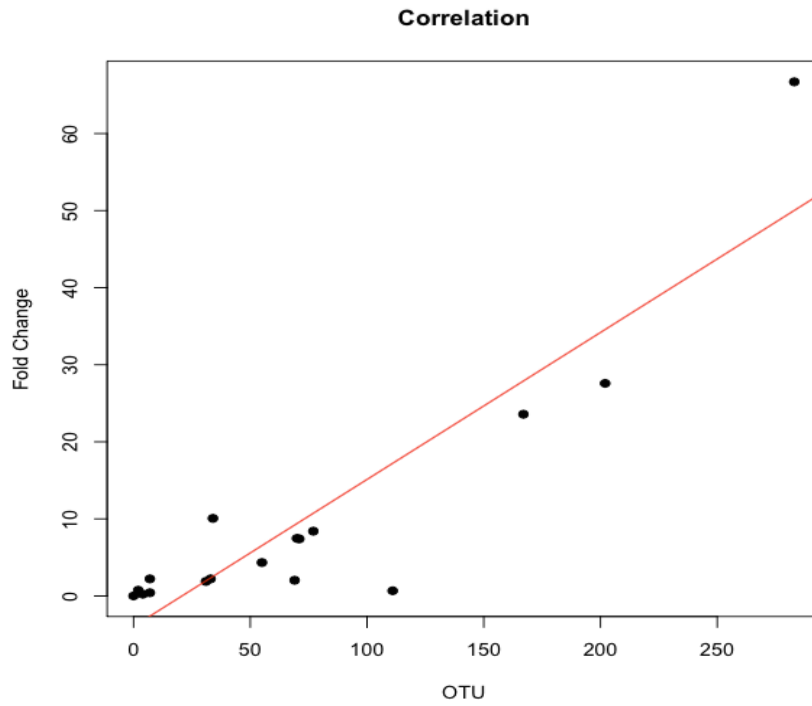
465   48.   Clark WB, Magnusson I, Beem JE, Jung JM, Marks RG, McArthur WP. Immune

Relative abundance of phyla identified in patient saliva summarized by diagnosis group.

Relative abundance of phyla in oral communities from 108 study patients summarized by diagnosis group (H= healthy control, O= other disease, and P= pancreatic cancer).

Mean relative abundances of particular genera in pancreatic cancer patients (P) compared to healthy (H) and other disease (O) patient groups.

Relative abundances of genera in oral communities from 108 patients. Arrows point to specific genera that showed interesting trends across diagnosis groups.[b]

Abundance ratio of *Leptotrichia* to *Porphyromonas* between different patient categories.

Each symbol represents the ratio of *Leptotrichia Oral Taxon 221* and *Leptotrichia hongkongenesis* to *Porphyromonas* for an individual patient (n=108). Patients are grouped into 3 different categories depending on their diagnosis: healthy control (H), other diseases (including cancer) (O), and pancreatic cancer (P). Horizontal bar and error bars represent the mean and SEM, respectively. ***p<0.001 (Kruskal-Wallis test followed by Dunn's multiple-comparison test).

Correlation between *Leptotrichia*abundance from 16S rRNA sequences and from real-time qPCR.

Cross validation of total *Leptotrichia* OTU abundance using real-time qPCR. After using 16S rRNA as a reference gene for normalization of the levels of *Leptotrichia* genus, data was normalized by fold change to three healthy controls with relatively low *Leptotrichia* OTU abundance. Each symbol represents a patient: P=6, and O=12. *Leptotrichia* OTU abundance was correlated with qPCR fold change according to Pearson's correlation (r=0.903).

Clinical characteristics of study sample (n=108)

Demographics	Pancreatic Cancer (P) n=8	Other Disease (O) n=78	Healthy Control (H) n=22	Total n=108
Sex				
Male	6	38	12	56
Female	2	40	10	52
Ethnicity				
Caucasian	6	56	15	77
Hispanic	2	6	5	13
Asian	0	4	1	5
Unknown	0	12	1	13

Types of identified cancers (n=23)

Cancer by Category	Age Mean	N
Pancreas (n=8)	71.1	
Pancreatic Cancer		8
Digestive (n=9)	64.7	
Ampullary		3
Esophageal		3
Stomach		1
Rectal		2
Non-Digestive (n=6)	54.8	
Breast		1
Skin		1
Testicular		1
Thyroid		3