1   **Comparison of subtyping approaches and the underlying drivers of microbial**

2   **signatures for chronic rhinosinusitis**

3   Kristi Biswas[a], Raewyn Cavubati[a], Shan Gunaratna[a], Michael Hoggard[b], Sharon Waldvogel-

4   Thurlow[a], Jiwon Hong[a,b], Kevin Chang[c], Brett Wagner Mackenzie[a], Michael W. Taylor[b],

5   Richard G. Douglas[a*]

6

7   [a]Department of Surgery, The University of Auckland, Auckland, New Zealand

8   [b]School of Biological Sciences & Maurice Wilkins Centre for Molecular Biodiscovery, The University of

9   Auckland, Auckland, New Zealand

10   [c]Department of Statistics, The University of Auckland, Auckland, New Zealand

11

12   **Running title:** Comparison of stratification approaches for CRS

13

14   ***Correspondence to**: Prof. Richard G. Douglas

15   Department of Surgery,

16   The University of Auckland,

17   Private Bag 92019,

18   Auckland 1142, New Zealand.

19   Email: Richard.douglas@auckland.ac.nz

20

21   **Keywords:** Epithelial barrier, microbiota, tight junctions, mucosal integrity, inflammation

22

Abstract

Chronic rhinosinusitis (CRS) is a heterogeneous condition characterised by persistent sinus inflammation and microbial dysbiosis. This study aimed to identify clinically relevant sub-groups of CRS patients based on distinct microbial signatures, with a comparison to the commonly used phenotypic subgrouping approach. The underlying drivers of these distinct microbial clusters were also investigated, together with associations with epithelial barrier integrity.

Sinus biopsies were collected from CRS patients (n=23), and disease controls (n=8). Expression of 42 tight junction genes was evaluated using quantitative PCR, together with microbiota analysis and immunohistochemistry for measuring mucosal integrity and inflammation.

CRS patients clustered into two distinct microbial sub-groups using probabilistic modelling Dirichlet (DC) multinomial mixtures. DC1 exhibited significantly reduced bacterial diversity, increased dispersion, and was dominated by *Pseudomonas*, *Haemophilus*, and *Achromobacter*. DC2 had significantly elevated B-cells, incidence of nasal polyps, and higher numbers of *Anaerococcus*, *Megasphaera*, *Prevotella*, *Atopobium*, and *Propionibacterium*. In addition, each DC exhibited distinct tight junction gene and protein expression profiles compared with controls. Stratifying CRS patients based on clinical phenotypic subtypes (absence or presence of nasal polyps (CRSsNP or CRSwNP respectively) or with cystic fibrosis (CRSwCF)) did account for a larger proportion of the variation in the microbial dataset compared with DC groupings. However, no significant differences between CRSsNP and CRSwNP cohorts were observed for inflammatory markers, beta-dispersion and alpha diversity measures.

46   In conclusion, both approaches used for stratifying CRS patients had benefits and pitfalls, but

47   DC clustering did provide greater resolution when studying tight junction impairment. Future

48   studies in CRS should give careful consideration into the patient subtyping approach used.

49

50

51   **Importance**

52   Chronic rhinosinusitis (CRS) is a major human health problem that significantly reduces

53   quality of life. While various microbes have been implicated, there is no clear understanding

54   of the role they play in CRS pathogenesis. Another equally important observation made for

55   CRS patients is that the epithelial barrier in the sino-nasal cavity is defective. Finding a

56   robust approach to subtype CRS patients would be the first step towards unravelling the

57   pathogenesis of this heterogeneous condition. Previous work has explored stratification based

58   on clinical presentation of the disease (with or without polyps), inflammatory markers,

59   pathology, or microbial composition. Comparing between the different stratification

60   approaches used in these studies has not been possible due to different cohorts, analytical

61   methods, or sample sites used. In this study, two approaches of subtyping CRS patients were

62   compared and the underlying drivers of the heterogeneity in CRS were also explored.

Introduction

As the first line of defence against inhaled antigens from the external environment, the upper airway epithelium represents an important physiological barrier (1). Maintaining the close cell to cell proximity that is central to the integrity of this barrier are tight junctions, which play a critical role in host defence (2, 3). Tight junctions comprise a range of transmembrane and scaffolding adaptor proteins that include occludin, claudins, junctional adhesion molecules, and zonula occludens (ZO) (4). These proteins limit the passage of macromolecules by sealing off the paracellular spaces between epithelial cells. On the other hand, in cases of tissue inflammation, the opening up of tight junctions assists in the release of tissue fluids and the influx of inflammatory cells and by doing so helps speed resolution. Accordingly, tight junctions are considered as the gatekeepers of inflammatory disease.

Recent studies have considered the role of epithelial barrier defects in a number of chronic inflammatory conditions, including chronic rhinosinusitis (CRS) (5, 6). CRS is an inflammatory condition of the upper respiratory tract persisting for over 12 weeks, affecting 5% of the general population (7). Symptoms include nasal discharge, facial pain, loss of smell and headaches. CRS is a complex and heterogeneous disease, with many underlying factors that present with similar symptoms, making it challenging to separate CRS into clinically relevant subtypes. Traditionally, classification of CRS has been based on the clinical phenotypes – the presence (CRSwNP) or absence (CRSsNP) of polyposis. Approximately 25-30% of CRS patients present with nasal polyps (8), and this cohort is considered to have a Th2-predominated inflammatory response compared with idiopathic CRS without nasal polyps (CRSsNP) that has Th1-type responses. However, this simplified view misrepresents the true complexities of CRS.

Recent efforts have investigated the pathogenesis of CRS based on inflammatory markers (9, 10), microbiota composition (11, 12), pathology or clinical factors (13, 14), but these studies

88   have produced little consensus on appropriate strategies to subtype patients. The role of

89   microbes in CRS remains unclear due to the frequent lack of resolution with antimicrobial

90   treatment. It has been suggested that a loss of overall microbiota diversity and deleterious

91   community changes (collectively termed 'dysbiosis') are more characteristic of CRS patients

92   than a single, disease-causing organism (15).

93   Stratification of patients based on probabilistic modelling of the bacterial communities in

94   lower respiratory diseases such as asthma and HIV-infected pneumonia patients has been

95   used successfully to classify immunological or clinical phenotypic variation across cohorts

96   (16, 17). Using a similar approach, a recent study subtyped CRS patients based on their

97   microbial community profiles (11). Each distinct microbial state was dominated by one

98   bacterial family and associated with a unique clinical and host immune response. Previous

99   studies have found that large interpersonal variations in the sino-nasal microbiome of CRS

100   patients (subtyped based on clinical diagnosis) can make classification difficult (12).

101   Accordingly, a distinct advantage of stratifying patients based on microbial community

102   profiles is that it will help to resolve the microbial heterogeneity of CRS and is a step towards

103   implementation of a precision medicine approach. The underlying drivers of these distinct

104   microbial states in CRS are yet to be investigated.

105   Several studies have investigated epithelial integrity in CRS mucosa and found that CRS

106   patients with nasal polyps (CRSwNP) have severely disrupted epithelia with decreased

107   expression of occludin and ZO-1 (6, 18, 19). These efforts also suggest a role for cytokines

108   (IFN-γ and IL-4) in disrupting tight junctions, whereas the influences of the microbes on the

109   host are less understood.

110   One common hypothesis is that a loss of tight junction integrity in CRS patients could lead to

111   the entry of environmental agents, including microbes, into host tissues (20). It remains

112   unclear whether the microbes are the cause of tissue damage or if microbial patterns or

113   associations are the consequence of a leaky epithelium, leading to a build-up of microbial

114   cells in the tissue. The mechanism by which bacteria penetrate the epithelial barrier is

115   unknown, but their presence in CRS tissue (in particular CRS with cystic fibrosis (CRSwCF))

116   presumably reflects a breakdown in mucosal integrity (21). Furthermore, any changes caused

117   to the micro-environment (such as lack of mucociliary clearance, damaged surfaces for

118   bacterial cell adherence) could have an impact on microbial composition and structure, as

119   observed previously in gut studies (22).

120   In this study, we aimed to stratify CRS patients based on their microbial community

121   composition using a probabilistic modelling approach and using the traditional phenotypic

122   approach. In addition, we investigated several possible underlying influences on these

123   microbial states by measuring gene and protein expressions of host tight junction, epithelial

124   integrity and inflammatory cells (T-cells, B-cells, and macrophages) in the sino-nasal tissue

125   biopsies.

126

127

128   **Results**

129   *Clinical parameters*

130   Non-parametric pairwise comparisons were made based on clinical factors in this cohort of

131   31 subjects (CRS = 23, disease controls = 8). Patient demographics are included in Table S1.

132   Dirichlet multinomial mixtures using probabilistic modelling in R were used to stratify

133   patients based on their microbial communities (operational taxonomic unit (OTU)-level) (23).

134   The Laplace approximation was used to find the model of best fit and determine the number

135   of clusters from the dataset. Unique microbial states were labelled Dirichlet clusters (DC).

136   None of the measured clinical factors were significantly different between the two DC groups

137   (DC1 and DC2) and disease controls (Table S2). A similar analysis was performed on

138   phenotypic subtypes of CRS (CRSsNP, CRSwNP, CRSwCF) and disease controls (Table

139   S3), where age and polyposis were found to be significant factors.

140

Overall bacterial community composition

142   Quality filtering resulted in 346,003 bacterial 16S rRNA gene sequences from 31 samples.

143   Samples that did not meet a rarefaction threshold of 677 were removed from further analysis,

144   including 3 CRSsNP samples. A final number of samples included in further analyses was

145   thus 8 disease controls and 20 CRS patients (CRSsNP = 5, CRSwNP = 8, CRSwCF = 7; DC1

146   = 13, DC2 = 7). The final, rarefied dataset included 200 taxonomically assigned OTUs at

147   97% sequence similarity (ranging from 4-45 OTUs per sample) and this OTU table was used

148   for all subsequent microbial-related analyses.

149   Large inter-personal variation in microbial community composition was observed between

150   individuals of each cohort (Fig. 1). The bacterial taxa *Staphylococcus* (OTU2), *Streptococcus*

151   (OTU7), *Propionibacterium* (OTU6) and *Corynebacteriaceae* (OTU10) were prevalent in

152   majority of the samples but only at low relative sequence abundances (Fig. 2A).

153

Bacterial community composition of DC groups

155   Samples were assigned into clusters based on bacterial community composition using

156   Dirichlet distributions (23). Dirichlet cluster 1 (DC1) comprised samples from all three CRS

157   sub-groups (CRSsNP = 4, CRSwNP = 2, CRSwCF = 7). In contrast, DC2 was dominated by

158   CRSwNP samples (n = 6) along with one CRSsNP sample.

159   A variety of alpha-diversity measurements revealed significant differences between DC1 and

160   controls, however no significant differences were noted between DC1-DC2 and DC2-controls

161   (Fig. 1B). The bacterial taxa *Pseudomonas* (average relative abundance 19% ± SD 37%),

162   *Achromobacter* (7.6% ± SD 27%), and *Haemophilus* (8.5% ± SD 27%) dominated DC1,

7

163   while DC2 was dominated by *Moraxella* (18% ± SD 31.5%) and *Stenotrophomonas* (6.7% ±

164   SD 17.7%), but these taxa all had low prevalence as shown in Fig. 2B. Interestingly, controls

165   were not dominated by any single bacterial taxon but instead had a low abundance of

166   multiple genera, some of which were highly prevalent.

167   Taxa that were significantly different between groups were investigated through multiple

168   pairwise comparisons at OTU- and genus-level (Table S4). OTUs that were significantly

169   elevated in DC2 compared with DC1 were *Anaerococcus* (OTUs 12, 19 and 251), *Prevotella*

170   (OTU200), *Megasphaera* (OTU179), and *Atopobium* (OTU203). Control samples were

171   differentiated from DC1 and DC2 by significant increases in *Streptococcus* (OTUs 7, 18),

172   *Veillonella* (OTU34), *Massilia* (OTU77)*, Peptoniphilus* (OTU11) and *Halomonadaceae*

173   (OTU109).

174   The dispersion of samples based on microbial community profiles in each DC was compared

175   by calculating the distances to the centroid in non-metric multidimensional scaling (nMDS)

176   analyses. DC1 samples were significantly more dispersed than control samples ($p = 0.011$,

177   Tukey's honest test) (Fig. 3B). In addition, Permutational Multivariate Analysis of Variance

178   (PERMANOVA) test confirmed that each of the clusters identified for the CRS cohort and

179   control group explained a significant proportion ($R^2 = 12.2\%$, $p = 0.014$) of the overall

180   variation in the bacterial dataset. However these PERMANOVA results should be interpreted

181   with caution in light of the different beta-dispersion patterns of the groups assessed (24).

182

183   ***Bacterial community composition of phenotypic groups***

184   Patients were also stratified based on clinical presentation of the disease, which includes the

185   presence or absence of nasal polyposis and comorbidity of cystic fibrosis.

186   Alpha diversity was significantly (p <0.001) lower in CRSwCF patients compared with

187   controls. Significant differences in diversity were also observed (p <0.05) between CRSwCF

188   and CRSwNP samples.

189   Similar to controls (as described above), CRSsNP did not show dominance of any single

190   bacterial taxon. However CRSwNP samples were dominated by *Moraxella* (average relative

191   abundance 16% ± SD 30%) and *Stenotrophomonas* (5.8% ± 16.5%), while CRSwCF samples

192   were dominated by *Pseudomonas* (35.4% ± 46%), *Staphylococcus* (24.5% ± 41.6%),

193   *Achromobacter* (14.1% ± 37%), and *Haemophilus* (13.9% ± 36.8%) (Fig. 2C).

194   Pairwise comparisons between control samples and CRS subtypes (CRSsNP, CRSwNP,

195   CRSwCF) were performed on OTU- and genus-level data. CRSwCF samples were

196   significantly reduced in *Propionibacterium*, *Corynebacterium*, *Anaerococcus* and

197   *Peptoniphilus* compared with controls. Furthermore, CRSsNP samples had significantly

198   lower abundance of *Peptoniphilus*, while CRSwNP were reduced in *Streptococcus* and

199   *Veillonella*, compared to controls. OTUs and genera that were significantly different between

200   CRS subtypes can be found in Table S5.

201   There were no significant differences in dispersion between phenotypic groups and controls

202   when calculating analysis of variance. However, PERMANOVA tests were able to explain a

203   larger proportion of the variation (R^2 = 23.4%, p = 0.001) by phenotypic subtyping methods

204   for CRS than with DC clustering.

205

206   ***Tight junction protein and gene expression patterns in sino-nasal tissue***

207   Based on our initial aim to investigate the underlying drivers of each subtype in CRS, the

208   expression of 42 tight junction genes in sino-nasal tissue biopsies was measured for each

209   patient and compared to that of disease controls. After results were normalised to a

9

210   housekeeping gene, fold-changes in gene expression were recorded for each DC and CRS

211   phenotypic subtype compared to controls.

212   Ten tight junction genes were significantly under- or over-expressed in DC1 and/or DC2

213   compared to controls (Fig. 4A). The gene ACTA1, which encodes a skeletal α-actin protein

214   that maintains the cytoskeleton and cell movement, was the only gene to be significantly

215   over-expressed in both DC1 and DC2. Each DC group exhibited unique tight junction gene

216   expression patterns. Three genes were significantly under-expressed in DC1: CSDA, TCF7

217   and PVRL1. In DC2, nine genes were significantly under-expressed compared with controls

218   (Fig. 4A).

219   Subtyping patients based on phenotypic characteristics found 7 tight junction genes to be

220   under expressed and ACTA1 gene to be over expressed in the CRS cohorts compared with

221   the controls (Fig 4B). Genes PVRL1 and TCF7 were significantly under-expressed in all

222   three CRS subtypes. The expression profile of CRSsNP and CRSwCF were very alike, with

223   the only one exception (gene CSDA) which was under-expressed in CRSwCF patients

224   (Fig. 4B). Spearman correlation analysis of the dataset after adjustment for multiple

225   comparisons showed a significant negative association between tight junction gene MTDH

226   expression and members of the bacterial genus *Pseudomonas* (Fig. S1).

227   Staining of sino-nasal tissue biopsies identified ZO-1 and occludin proteins at apico-lateral

228   contact points of adjacent epithelial cells (Fig. S2). Claudin-1, by contrast, was

229   predominantly seen in subapical regions, with concentrated, continuous staining amongst the

230   mid-basal region of epithelial cells (Fig. S2). ZO-1 area staining was significantly ($p = 0.03$)

231   lower in DC2 (mean $0.5 \pm$ S.D 0.3%) compared with controls ($1.1 \pm 0.5\%$) (Table S3). There

232   were no significant differences in claudin-1 and occludin staining in disease controls relative

233   to DC1 and DC2. In addition, there were no significant differences in the staining of ZO-1,

234   claudin-1 and occludin proteins in the three CRS subtypes compared with controls.

235 Furthermore, no correlations were observed between tight junction proteins and abundance of

236 bacterial taxa (Fig. S1).

237

238 *Inflammatory state and mucosal integrity of tissue biopsies*

239 Inflammatory marker cells (T-cells, B-cells and macrophages) were enumerated in the tissue

240 biopsies to assess the inflammatory state of the patients. Although all three cell types were

241 reduced in the disease controls compared with DC1 and DC2 cohorts, B-cells (CD20+) were

242 the only identified inflammatory marker with a significant difference between the groups,

243 with higher numbers in DC2 (Table S2).

244 In contrast, by traditional CRS subtyping approaches, all three inflammatory marker cells

245 were significantly elevated only in CRSwNP cohort compared with controls. There were no

246 significant differences between the CRS sub-type cohorts, except for the elevated amounts of

247 macrophages (CD68+) in CRSwNP compared with CRSwCF.

248 The sino-nasal tissue biopsies were also assessed for mucosal integrity. Goblet cell

249 enumeration and cilia integrity in the biopsy specimens were significantly reduced in DC2

250 compared with DC1 and control (Table S2). Similarly, when subtyping the same group of

251 patients based on phenotypic approaches, goblet cell enumeration and cilia integrity were

252 significantly reduced in CRSwNP patients compared with controls and CRSsNP patients.

253 Interestingly, CRSwCF patients had no significant differences in mucosal integrity (collagen

254 content, goblet cell enumeration and cilia integrity) compared with controls. Furthermore,

255 Spearman correlation analysis showed a positive association between goblet cell counts and

256 tight junction gene LEF1 expression (Fig. S1). No other significant associations were found

257 between mucosal integrity, inflammatory marker cells and all other measured variables in this

258 study.

259

11

260

Discussion

262 Endotyping of CRS patients has been the subject of considerable recent research, as it is

263 hoped that a sub-classification of this condition will allow for more specific and effective

264 therapies to be administered. In this study, we defined microbial states for CRS using

265 probabilistic modelling, in which patients with similar microbial states were clustered

266 together. Furthermore, the same cohort of patients were also sub-typed based on phenotypic

267 presentation of the disease. We then sought to understand the underlying influences on these

268 cohorts by investigating sino-nasal mucosal integrity, tight junction gene/protein expression

269 and inflammatory status. The two approaches of clustering CRS patients will be compared

270 and discussed further.

271

Resolving the microbial heterogeneity of CRS

273 Phenotyping of CRS patients based on clinical factors can be subjective and provides little

274 information about microbes and their involvement in this disease. As shown previously for

275 gastrointestinal, lower and upper respiratory diseases (11, 16, 17, 23), distinct microbial

276 states were identified for CRS patients, allowing for stratification based on bacterial

277 composition. The advantage of the new clustering approach used in this study and by others

278 (11) is that it reflects a patient's microbial state at the time and places the patient into a

279 distinct microbial cluster type. Appropriate targeted treatment strategies could then be

280 prescribed for patients in the future based on their distinct microbial pattern.

281 In this study, two distinct microbial states of CRS patients were identified that were

282 significantly different to each other in diversity, beta-dispersion and the relative abundance of

283 members from the genus *Anaerococcus*. Although this novel way of classifying CRS samples

284 could explain some of the microbial variations in the dataset, there was still a large

12

285   proportion of variation that could not be accounted for. This is most likely due to the sizeable

286   inter-patient variation observed in this study and reported previously (15, 25). We anticipate

287   with larger cohort sizes these microbial states would be more pronounced and less obscured

288   by individual variation. While studies that use less invasive approaches to obtain samples

289   such as nasopharyngeal swabs and rinses are able to increase cohort sizes (26, 27), these

290   samples are not suitable for analyses of histology and mucosal integrity. In addition, tissue

291   biopsies are required to study host gene expression levels. For these reasons, tissue biopsies

292   were used in the analyses within this study.

293   Subtyping patients based on the phenotype of the disease (with or without nasal polyps or

294   with cystic fibrosis) has been the most common approach to date. The results of the microbial

295   communities from each of these subtypes is confirmatory to those previously described by

296   our group (28, 29). Interestingly, a greater amount of the variation (23.4%) observed in the

297   dataset from this study could be explained through this patient phenotypic clustering

298   approach rather than by the DC approach.

299

300   **Potential drivers of microbial signatures for CRS patients**

301   Understanding the underlying factors that shape the sino-nasal microbial community of a

302   CRS patient is essential for defining the pathogenesis of this complex disease and for

303   developing better treatment protocols. The most common hypothesis for the pathogenesis of

304   CRS is that the epithelial barrier is defective (30). Allergens (such as pollen), host genetics,

305   bacteria, viruses or inflamed tissue could all contribute to the disruption of the sino-nasal

306   epithelial barrier (31). In this study, we chose to investigate several of these potential host

307   factors that could contribute to CRS pathogenesis and determine their influence on the sino-

308   nasal microbial state. Out of the 42 measured tight junction genes, nine exhibited reduced

309   expression in CRS patients compared with controls. Reduced expression of tight junction

13

310   genes in CRS patients compared with controls have been observed previously (6). The reason

311   for the overexpression of ACTA1, which encodes for proteins involved in cell motility,

312   structure and integrity (www.uniprot.org) is unclear but presumably reflects the host's

313   response to an impaired epithelial surface. Each DC in this study had a unique tight junction

314   gene expression profile, suggesting that the two identified CRS DCs are functionally different

315   from each other. It remains unknown whether specific tight junction gene expression could

316   play a role in determining distinct microbial states in the nasal cavity, or vice versa. The

317   significant association between tight junction genes and bacterial community composition

318   observed in this study needs further validation with *in vitro* tests. However, this observation

319   does provide some evidence of interactions between host tight junction gene expression and

320   sino-nasal microbial communities.

321   Stratification of patients based on the traditional phenotypic approach were not able to clearly

322   separate out the tight junction gene expression profiles of CRSsNP and CRSwCF cohorts.

323   This lack of clarity, would suggest that future studies studying tight junction gene expression

324   profiles in CRS patients should consider using alternative patient stratification approaches.

325   Of the three measured tight junction proteins in this study, only ZO-1 was significantly

326   downregulated in DC2 relative to controls. ZO-1's adhesive function between transmembrane

327   proteins and the underlying actin skeleton denotes its critical role at the epicentre of the tight

328   junction complex (32, 33). Furthermore, depletion of ZO proteins in mammary epithelial

329   cells results in failure of tight junction strands to assemble and consequent loss of barrier

330   function (33). These findings suggest an important role for ZO-1 in the establishment of

331   functioning tight junction complexes. It is possible that reduced expression of ZO-1 protein

332   could underlie subsequent downregulation of integral tight junction proteins, resulting in a

333   loss of barrier properties. A loss of claudin-4, ZO-1, occludin and E-cadherin in tissue

334   biopsies of CRS patients has been found previously (6, 34). The lack of any significant

14

335   observations on the tight junction protein measurements in the phenotypic CRS sub groups

336   compared with controls, also emphasises the need to look at other ways to stratify CRS.

337   Another potential driver of microbial states in CRS is mucosal integrity. Several studies have

338   reported significant changes in the structure and composition of the mucosa in CRS (35-37).

339   CRS subjects exhibited polarising results in regards to evidence of mucus hypersecretion,

340   with a significantly higher goblet cell count compared with controls. Interestingly, DC2

341   cohorts had significantly lower goblet cell counts and reduced integrity of cilia compared

342   with DC1 and controls. This observation was also made for subgroup CRSwNP patients

343   compared with other CRS cohorts and controls. Mucociliary function in CRS is essential for

344   the physiological function and immunity of the nose (38). A loss in function promotes

345   formation of biofilms and bacterial infections. Certain species of *Pseudomonas*, *Haemophilus*

346   and *Streptococcus* produce ciliostatic or ciliotoxic agents that could result in loss of ciliary

347   function (39). In this study, correlations with cilia integrity and goblet cell numbers and

348   bacterial taxa were not observed. However, further research into toxin production by

349   signature members of each cohort that results in cilia loss is required.

350   Previous studies have shown pro-inflammatory cytokines (IFN-γ and IL-4) to disrupt

351   epithelial integrity *in vitro*, which in turn could cause changes to the sino-nasal micro-

352   environment by disrupting surfaces for attachment or increase the permeability of microbes

353   to the underlying tissue (6, 40). In addition, Cope et al. (11) found distinct microbial clusters

354   to have unique patterns of immune response. Evidence from these studies suggests that

355   microbial states are influenced by various host factors, possibly including cellular junctions.

356   Furthermore, consistent with findings from other groups, CRS patients in this study based on

357   microbial states (DC1 and DC2) or phenotypic subtyping (CRSwNP) had higher abundance

358   of B-cells in the mucosa, compared with controls. Previous efforts have shown a proliferation

359   of B-cells in the sino-nasal mucosa when exposed to antigens (41). Inflammatory cytokine

15

360   IL-13 is a key factor in stimulating antibody production (such as IgE and IgA) by B-cells in

361   response to antigen exposure in airway inflammatory diseases (42). IgE levels are elevated in

362   eosinophilic CRSwNP patients (43). Accordingly, the elevated levels of B-cells in CRS

363   patients is a good indication of the inflammatory status of the sino-nasal tissue of CRS

364   patients.

365

366   **Stratification approach for CRS**

367   The traditional approach of stratifying CRS patients based on their phenotypic presentation of

368   the disease is standard practice, however such approaches are struggling to unravel the

369   complexities of the disease. This study showed that, while phenotypic subtyping approaches

370   can help explain some of the variation in the microbial community dataset, alternative

371   methods using microbial signatures can also have some benefit. Each microbial state of CRS

372   in this study was found to have a unique tight junction expression profile, which was not

373   observed for the phenotypic subtyping approach. Accordingly, it remains unclear whether

374   microbial states influence, or are being influenced by, barrier impairment. To answer these

375   questions, *in vitro* studies will need to be undertaken in the future.

376

377

378   **Materials and Methods**

379   *Patient recruitment and sample collection*

380   Twenty-three adult patients undergoing functional endoscopic sinus surgery for CRS by a

381   single surgeon (RD) were recruited for this study. Diagnosis and subsequent recruitment of

382   CRS patients were based on EPOS 2012 guidelines (44). These included CRS patients with

383   polyps (CRSwNP; n = 8), CRS patients without polyps (CRSsNP; n = 8), and CRS patients

384   with cystic fibrosis (CRSwCF; n = 7). Control subjects (n = 8) without any signs of sinus

385   disease undergoing endoscopic surgery for the removal of pituitary tumours or medial orbital

386   decompression were also recruited. Control patients exhibited no evidence of mucosal

387   inflammation on endoscopy or computed tomography scans. Exclusion criteria included

388   patients being administered systemic corticosteroids or antibiotics within four weeks prior to

389   surgery, age less than 18 years, immunodeficiency, pregnancy, and other comorbidities (apart

390   from cystic fibrosis). This study was approved by the Health and Disability Ethics Committee

391   of New Zealand (NTX/08/12/126). Prior to sample collection, informed written consent was

392   obtained from all patients.

393   Patients completed a symptom score sheet, prior to surgery, in which the following CRS

394   symptoms were rated on a scale of 0-5: nasal obstruction, anterior nasal discharge, posterior

395   nasal discharge, facial pain or fullness, and loss of smell. These scores were summated to

396   give the 'Symptom Severity' score in Table S1. Lund-Mackay scoring was used to quantify

397   radiological disease severity (45). Tissue biopsies were collected intraoperatively from the

398   ethmoidal sinus under general anaesthesia prior to administration of topical vasoconstrictors

399   or intravenous antibiotics. Biopsied specimens were rinsed in sterile saline and partitioned.

400   Samples for quantitative PCR and bacterial community analysis were fixed in RNA*later* for

401   24 h, then stored at -20°C. Samples for immunohistochemistry analysis were fixed in

402   Carnoy's solution (60% ethanol, 30% chloroform, 10% glacial acetic acid) before paraffin

403   embedding.

404

405   ***Bacterial community analysis***

406   *DNA extraction*

407   DNA was extracted from tissue biopsies using sterile Lysing Matrix E bead tubes (MP

408   Biomedicals, Australia) and the AllPrep DNA/RNA Isolation kit (Qiagen, Germany) as

409   previously described (46). A negative DNA extraction control using 200 µL sterile water was

410   carried out simultaneously.

411

412 *Bacterial community sequencing*

413 The V3-V4 region of the bacterial 16S rRNA gene was amplified using primers 341F and

414 806R (47), with Nextera DNA library Prep Kit adapters attached. PCR reactions,

415 amplification conditions and purifications were carried out as previously described (12). In

416 brief, genomic DNA (~100 ng) from each sample was amplified in duplicate PCR reactions

417 of 35 cycles then pooled to a final volume of 50 μL. Negative PCR controls were included in

418 all PCR reactions and yielded no detectable amplicons. Eluent from the negative extraction

419 control was also subjected to PCR amplification and yielded no detectable product.

420 Purification using Agencourt AMPure magnetic beads (Beckman Coulter Inc., USA) was

421 carried out as per the manufacturer's instructions. Purified PCR products were quantified

422 using Qubit dsDNA High-Sensitivity kits (Life Technologies, New Zealand), standardised to

423 ~5 ng per sample, and submitted to the University of Auckland Genomics Centre for library

424 preparation and sequencing using Illumina MiSeq (2 x 300 bp paired-end reads). Raw

425 sequence reads were deposited into the SRA-NCBI database (BioProject ID: PRJNA482256).

426

427 *Bioinformatics*

428 Sequences were merged and quality filtered in USEARCH (version 8.0) with default settings

429 as previously described (12). OTU clustering based on a 97% 16S rRNA gene sequence

430 similarity threshold was performed using the UCLUST algorithm in USEARCH (48). Each

431 OTU was taxonomically assigned in QIIME (49) using the RDP classifier 2.2 against the

432 SILVA 16S rRNA gene database (version 128) (50, 51). Sequences mapping to the human

433 genome were removed from subsequent analyses. Samples were rarefied to an even

434 sequencing depth of 677 reads. Alpha-diversity (including Shannon, Simpson, and richness

435 (observed OTUs)) and beta-diversity (including weighted and unweighted UniFrac distances,

436   and Bray-Curtis dissimilarity) were calculated in QIIME. Graphical outputs were created in R

437   (version 3.4.1) (52) and GraphPad Prism (version 7.03).

438

439   *RNA extraction and quantitative PCR (qPCR)*

440   RNA extraction of biopsied tissue was carried out in parallel to the DNA extraction using the

441   AllPrep DNA/RNA Isolation Kit (Qiagen) as per the manufacturer's instructions. Extracted

442   RNA was transferred to a collection tube and re-eluted in 30 µL DNA/RNA-free sterile

443   water.

444   Recovered RNA was treated with DNaseI (Invitrogen) for selective degradation of

445   contaminant DNA as per the manufacturer's instructions. The quantity and quality of RNA

446   were measured using a Nanodrop 3000 spectrophotometer. Successful removal of genomic

447   DNA from DNase-treated RNA samples was demonstrated by PCR targeting the human beta-

448   actin gene (53). DNase-treated RNA (standardised to ~100 ng/µL) was converted into cDNA

449   using iScriptTM Reverse Transcription Supermix for RT-qPCR (Bio-Rad, New Zealand) as

450   per manufacturer's instructions.

451   Pre-designed, 384-well qPCR arrays (TJ H384) were obtained from Bio-Rad Laboratories

452   Inc. (Auckland, New Zealand) for analysis of 42 known tight junction genes and one house-

453   keeping gene (GAPDH). Each sample was assessed for PCR performance, reverse

454   transcription efficiency, DNA contamination, and RNA quality. Analysis of results was

455   carried out using the ABI Prism 7900HT detection system (version 2.4). Each sample was

456   run in duplicate on separate qPCR arrays and averaged for further analysis. Tight junction

457   gene expression was normalised to the house-keeping gene GAPDH. Mean fold-change of

458   gene expression in CRS-affected patients compared to control was calculated using the

459   equation $2^{-\Delta\Delta Ct}$ (54).

460

461   *Histological analysis of sino-nasal tissue*

462   Paraffin-embedded tissues were prepared into 4 μm-thick sections then mounted on

463   Superfrost Plus positively-charged microscope slides (Thermo Fisher Scientific). Detailed

464   procedures for staining and analysis of tight junctions (ZO-1, claudin-1 and occludin), goblet

465   cells, cilia, collagen and inflammatory cells (CD4: T cells, CD20: B cells, CD68:

466   macrophages) can be found in the Supplementary Material.

467

468   *Statistical analysis*

469   Statistical analyses were carried out using GraphPad Prism software (CA, USA) and R.

470   Dirichlet multinomial mixtures using probabilistic modelling in R was used to stratify

471   patients based on their microbial communities (OTU-level) (23). The Laplace approximation

472   was used to find the model of best fit and determine the number of clusters from the dataset.

473   Unique microbial states were labelled Dirichlet clusters (DC). Sample dispersion between

474   groups was compared using analysis of variance, permutation test (PERMDISP), and Tukey's

475   honest significant differences in R. The software package 'adonis' PERMANOVA using

476   distance matrices (Bray-Curtis dissimilarity (weighted and unweighted), UniFrac distances

477   (weighted and unweighted)) was used to compare between centroids of groups.

478   Multiple non-parametric pairwise comparisons for categorical variables (diagnosis, gender,

479   ethnicity, polyposis, smoking status, antibiotic and steroid usage, and co-morbidities) were

480   tested using Fisher's exact test, with Bonferroni adjustment for multiple comparisons.

481   Pairwise comparisons between continuous variables for bacterial diversity, age, Lund-

482   Mackay scores, symptom severity scores, tight junction gene expressions, histological

483   analyses, bacterial OTUs and genera were tested using Dunn's test, with Bonferroni

484   adjustment for multiple comparisons. Heat maps depicting Spearman correlations, with

485   significance calculated using Spearman coefficients, including multiple comparisons

486   adjustment using Benjamini & Hochberg False Discovery Rate (BH-FDR) with hierarchical

487   clustering of correlation coefficients were generated in R. Factors with significant ($p < 0.05$)

488   positive or negative correlations were plotted.

489

490

491   **Acknowledgments**

492   The authors would like to thank the Garnett Passe and Rodney Williams Memorial

493   Foundation for its generous support of this research. Authors declare no conflict of interest.

494   Authors K.B., M.W.T, R.G.D contributed to the study design. Authors K.B., R.C., S.G., M.H,

495   S. W-T., J.H., K.C., B.WM contributed to data collection and analysis. Authors K.B., R.C.,

496   M.W.T., R.G.D. contributed to manuscript preparation.

References

1. Steed E, Balda MS, Matter K. 2010. Dynamics and functions of tight junctions. Trends Cell Biol 20:142-149.

2. Förster C. 2008. Tight junctions and the modulation of barrier function in disease. Histochem Cell Biol 130:55-70.

3. Kojima T, Go M, Takano K-i, Kurose M, Ohkuni T, Koizumi J-i, Kamekura R, Ogasawara N, Masaki T, Fuchimoto J. 2013. Regulation of tight junctions in upper airway epithelium. BioMed Res Int 2013:Article ID 947072.

4. Chiba H, Osanai M, Murata M, Kojima T, Sawada N. 2008. Transmembrane proteins of tight junctions. BBA-Biomembranes 1778:588-600.

5. Tieu DD, Kern RC, Schleimer RP. 2009. Alterations in epithelial barrier function and host defense responses in chronic rhinosinusitis. J Allergy Clin Immunol 124:37-42.

6. Soyka MB, Wawrzyniak P, Eiwegger T, Holzmann D, Treis A, Wanke K, Kast JI, Akdis CA. 2012. Defective epithelial barrier in chronic rhinosinusitis: the regulation of tight junctions by IFN-γ and IL-4. J Allergy Clin Immunol 130:1087-1096. e10.

7. Bhattacharyya N. 2003. The economic burden and symptom manifestations of chronic rhinosinusitis. Am J Rhinol Allergy 17:27-32.

8. Stevens WW, Schleimer RP, Kern RC. 2016. Chronic rhinosinusitis with nasal polyps. J Allergy Clin Immunol: In Practice 4:565-572.

9. Tomassen P, Vandeplas G, Van Zele T, Cardell L-O, Arebro J, Olze H, Förster-Ruhrmann U, Kowalski ML, Olszewska-Ziąber A, Holtappels G. 2016. Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers. J Allergy Clin Immunol 137:1449-1456. e4.

521   10.   Akdis CA, Bachert C, Cingi C, Dykewicz MS, Hellings PW, Naclerio RM, Schleimer

522   RP, Ledford D. 2013. Endotypes and phenotypes of chronic rhinosinusitis: a

523   PRACTALL document of the European Academy of Allergy and Clinical

524   Immunology and the American Academy of Allergy, Asthma & Immunology. J

525   Allergy Clin Immunol 131:1479-1490.

526   11.   Cope EK, Goldberg AN, Pletcher SD, Lynch SV. 2017. Compositionally and

527   functionally distinct sinus microbiota in chronic rhinosinusitis patients have

528   immunological and clinically divergent consequences. Microbiome 5:53.

529   12.   Hoggard M, Biswas K, Zoing M, Wagner Mackenzie B, Taylor MW, Douglas RG.

530   2017. Evidence of microbiota dysbiosis in chronic rhinosinusitis. Int Forum Allergy

531   Rhinol 7:230-239.

532   13.   Tan NW, Cooksley CM, Roscioli E, Drilling AJ, Douglas R, Wormald PJ, Vreugde S.

533   2014. Small-colony variants and phenotype switching of intracellular *Staphylococcus*

534   *aureus* in chronic rhinosinusitis. Allergy 69:1364-1371.

535   14.   Lal D, Rounds AB, Rank MA, Divekar R. 2015. Clinical and 22-item Sino-Nasal

536   Outcome Test symptom patterns in primary headache disorder patients presenting to

537   otolaryngologists with "sinus" headaches, pain or pressure. Int Forum Allergy Rhinol

538   5:408-416.

539   15.   Wagner Mackenzie B, Waite DW, Hoggard M, Douglas RG, Taylor MW, Biswas K.

540   2017. Bacterial community collapse: a meta-analysis of the sinonasal microbiota in

541   chronic rhinosinusitis. Environ Microbiol 19:381-392.

542   16.   Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker

543   ML, Hollams E. 2015. The infant nasopharyngeal microbiome impacts severity of

544   lower respiratory infection and risk of asthma development. Cell Host Microbe

545   17:704-715.

546   17.   Shenoy MK, Iwai S, Lin DL, Worodria W, Ayakaka I, Byanyima P, Kaswabuli S,

547   Fong S, Stone S, Chang E. 2017. Immune response and mortality risk relate to distinct

548   lung microbiomes in patients with HIV and pneumonia. Am J Respir Crit Care Med

549   195:104-114.

550   18.   De Benedetto A, Rafaels NM, McGirt LY, Ivanov AI, Georas SN, Cheadle C, Berger

551   AE, Zhang K, Vidyasagar S, Yoshida T. 2011. Tight junction defects in patients with

552   atopic dermatitis. J Allergy Clin Immunol 127:773-786. e7.

553   19.   Lambrecht BN, Hammad H. 2012. The airway epithelium in asthma. Nat Med 18:684.

554   20.   Kim R, Freeman J, Waldvogel-Thurlow S, Roberts S, Douglas R. 2013. The

555   characteristics of intramucosal bacteria in chronic rhinosinusitis: a prospective cross-

556   sectional analysis. Int Forum Allergy Rhinol 3:349-354.

557   21.   Kim RJT, Park L, Wood AJ, Yin T, Jain R, Douglas RG. 2015. Chronic rhinosinusitis

558   and cystic fibrosis: the interaction between sinus bacteria and mucosal immunity. Int

559   Forum Allergy Rhinol 5:380-385.

560   22.   Tropini C, Earle KA, Huang KC, Sonnenburg JL. 2017. The Gut microbiome:

561   connecting spatial organization to function. Cell Host Microbe 21:433-442.

562   23.   Holmes I, Harris K, Quince C. 2012. Dirichlet multinomial mixtures: generative

563   models for microbial metagenomics. PloS One 7:e30126.

564   24.   Warton DI, Wright ST, Wang Y. 2012. Distance-based multivariate analyses

565   confound location and dispersion effects. Methods Ecol Evol 3:89-101.

566   25.   Biswas K, Hoggard M, Jain R, Taylor MW, Douglas RG. 2015. The nasal microbiota

567   in health and disease: variation within and between subjects. Front Microbiol 6:134.

568   26.   Chalermwatanachai T, Vilchez-Vargas R, Holtappels G, Lacoere T, Jáuregui R,

569   Kerckhof F-M, Pieper DH, Wiele T, Vaneechoutte M, Zele T. 2018. Chronic

570   rhinosinusitis with nasal polyps is characterized by dysbacteriosis of the nasal

571   microbiota. Sci Rep 8:7926.

572 27. Polzehl D, Weschta M, Podbielski A, Riechelmann H, Rimek D. 2005. Fungus

573   culture and PCR in nasal lavage samples of patients with chronic rhinosinusitis. J

574   Med Microbiol 54:31-37.

575 28. Hoggard M, Biswas K, Zoing M, Wagner Mackenzie B, Taylor MW, Douglas RG.

576   Evidence of microbiota dysbiosis in chronic rhinosinusitis, p 230-239. *In* (ed), Wiley

577   Online Library,

578 29. Biswas K, Hoggard M, Jain R, Taylor MW, Douglas RG. 2015. The nasal microbiota

579   in health and disease: variation within and between subjects. Frontiers in

580   microbiology 6:134.

581 30. Lam K, Schleimer R, Kern RC. 2015. The etiology and pathogenesis of chronic

582   rhinosinusitis: a review of current hypotheses. Curr Allergy Asthma Rep 15:41.

583 31. Yeo NK, Jang YJ. 2010. Rhinovirus infection-induced alteration of tight junction and

584   adherens junction components in human nasal epithelial cells. Laryngoscope 120:346-

585   352.

586 32. Bauer H, Zweimueller-Mayer J, Steinbacher P, Lametschwandtner A, Bauer HC.

587   2010. The dual role of zonula occludens (ZO) proteins. BioMed Res Int 2010:Article

588   ID 402593.

589 33. Fanning AS, Anderson JM. 2009. Zonula occludens-1 and-2 are cytosolic scaffolds

590   that regulate the assembly of cellular junctions. Ann NY Acad Sci 1165:113-120.

591 34. Schleimer RP. 2017. Immunopathogenesis of chronic rhinosinusitis and nasal

592   polyposis. Annu Rev Pathol- Mech 12:331-357.

593   35.   Yamauchi K, Inoue H. 2007. Airway remodeling in asthma and irreversible airflow

594   limitation—ECM deposition in airway and possible therapy for remodeling. Allergol

595   Int 56:321-329.

596   36.   Van Bruaene N, Bachert C. 2011. Tissue remodeling in chronic rhinosinusitis. Curr

597   Opin Allergy Clin Immunol 11:8-11.

598   37.   Ponikau JU, Sherris DA, Kephart GM, Kern EB, Gaffey TA, Tarara JE, Kita H. 2003.

599   Features of airway remodeling and eosinophilic inflammation in chronic

600   rhinosinusitis: is the histopathology similar to asthma? J Allergy Clin Immunol

601   112:877-882.

602   38.   Cohen NA. 2006. Sinonasal mucociliary clearance in health and disease. Ann Otol

603   Rhinol Laryngol 115:20-26.

604   39.   Gudis D, Zhao K-q, Cohen NA. 2012. Acquired cilia dysfunction in chronic

605   rhinosinusitis. Am J Rhinol Allergy 26:1.

606   40.   Steelant B, Farré R, Wawrzyniak P, Belmans J, Dekimpe E, Vanheel H, Van Gerven

607   L, Krohn IK, Bullens DMA, Ceuppens JL. 2016. Impaired barrier function in patients

608   with house dust mite–induced allergic rhinitis is accompanied by decreased occludin

609   and zonula occludens-1 expression. J Allergy Clin Immunol 137:1043-1053. e5.

610   41.   Kato A. 2015. Immunopathology of chronic rhinosinusitis. Allergol Int 64:121-130.

611   42.   Kato A, Hulse KE, Tan BK, Schleimer RP. 2013. B-lymphocyte lineage cells and the

612   respiratory system. J Allergy Clin Immunol 131:933-957.

613   43.   Gevaert P, Nouri-Aria KT, Wu H, Harper CE, Takhar P, Fear DJ, Acke F, De Ruyck

614   N, Banfield G, Kariyawasam HH. 2013. Local receptor revision and class switching

615   to IgE in chronic rhinosinusitis with nasal polyps. Allergy 68:55-63.

616   44.   Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, Cohen N, Cervin A,

617   Douglas R, Gevaert P. 2012. EPOS 2012: European position paper on rhinosinusitis

618   and nasal polyps 2012. A summary for otorhinolaryngologists. Rhinology 50:1-12.

619   45.   Hopkins C, Browne JP, Slack R, Lund V, Brown P. 2007. The Lund-Mackay staging

620   system for chronic rhinosinusitis: how is it used and what does it predict? Otolaryngol

621   Head Neck Surg 137:555-561.

622   46.   Kim RJT, Biswas K, Hoggard M, Taylor MW, Douglas RG. 2015. Paired analysis of

623   the microbiota of surface mucus and whole-tissue specimens in patients with chronic

624   rhinosinusitis. Int Forum Allergy Rhinol 5:877-883.

625   47.   Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF.

626   2011. Transitions in bacterial communities along the 2000 km salinity gradient of the

627   Baltic Sea. ISME J 5:1571.

628   48.   Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.

629   Bioinformatics 26:2460-2461.

630   49.   Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,

631   Fierer N, Pena AG, Goodrich JK, Gordon JI. 2010. QIIME allows analysis of high-

632   throughput community sequencing data. Nat Methods 7:335.

633   50.   Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.

634   2012. The SILVA ribosomal RNA gene database project: improved data processing

635   and web-based tools. Nucleic Acids Res 41:D590-D596.

636   51.   Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid

637   assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ

638   Microbiol 73:5261-5267.

639   52.   R Core Team. 2015. R: A language and environment for statistical computing. R

640   Foundation for Statistical Computing.

641   53.   Wang X, Seed B. 2003. A PCR primer bank for quantitative gene expression analysis.

642   Nucleic Acids Res 31:e154-e154.

643   54.   Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-

644   time quantitative PCR and the 2− ΔΔCT method. Methods 25:402-408.

645

646

647 **Figure legends**

648 Fig. 1: Bacterial community composition and alpha diversity for the CRS cohorts and disease

649 controls. Grouped and patient-level profiles at genus-level are shown in the bar graphs for

650 (A) DC groupings and (C) phenotypic grouping. Box-and-whisker plots represent group

651 summaries for bacterial richness, Shannon diversity and Simpson diversity for (B) DC

652 groupings and (D) phenotypic groupings. Horizontal lines represent significant differences (p

653 < 0.05) between cohorts.

654

655 Fig. 2: Mean OTU relative abundance per sample (when present in a sample) plotted against

656 prevalence (occurrence) across (A) all samples, (B) each DC and control, and (C) each

657 phenotypic CRS subtype and control. All OTUs are plotted. The 14 most abundant OTUs are

658 color-coded, and all other OTUs are presented as black dots.

659

660 Fig. 3: Non-metric multidimensional scaling plot using Bray-Curtis dissimilarity distances

661 (weighted) for all samples for (A) DC clustering and (C) phenotypic subtyping. Ellipses

662 represent the 95% CI spread from centroids. Box-and-whisker plot of distances between each

663 subject to the centroid of their respective group (B) and (D). Beta-dispersion was

664 significantly different between DC1 and controls ($p = 0.0114$; Tukey's multiple comparisons

665 of means).

666

667 Fig. 4: Tight junction genes that were significantly different in expression for (A) DC groups

668 and (B) phenotypic subgroups compared with controls are displayed in the heatmap. Blue

669 represents genes that have low level expression, and red represents higher expression. Values

670 displayed are log-transformed mean fold-change ($2^{-\Delta\Delta Ct}$). * represents p values below 0.05

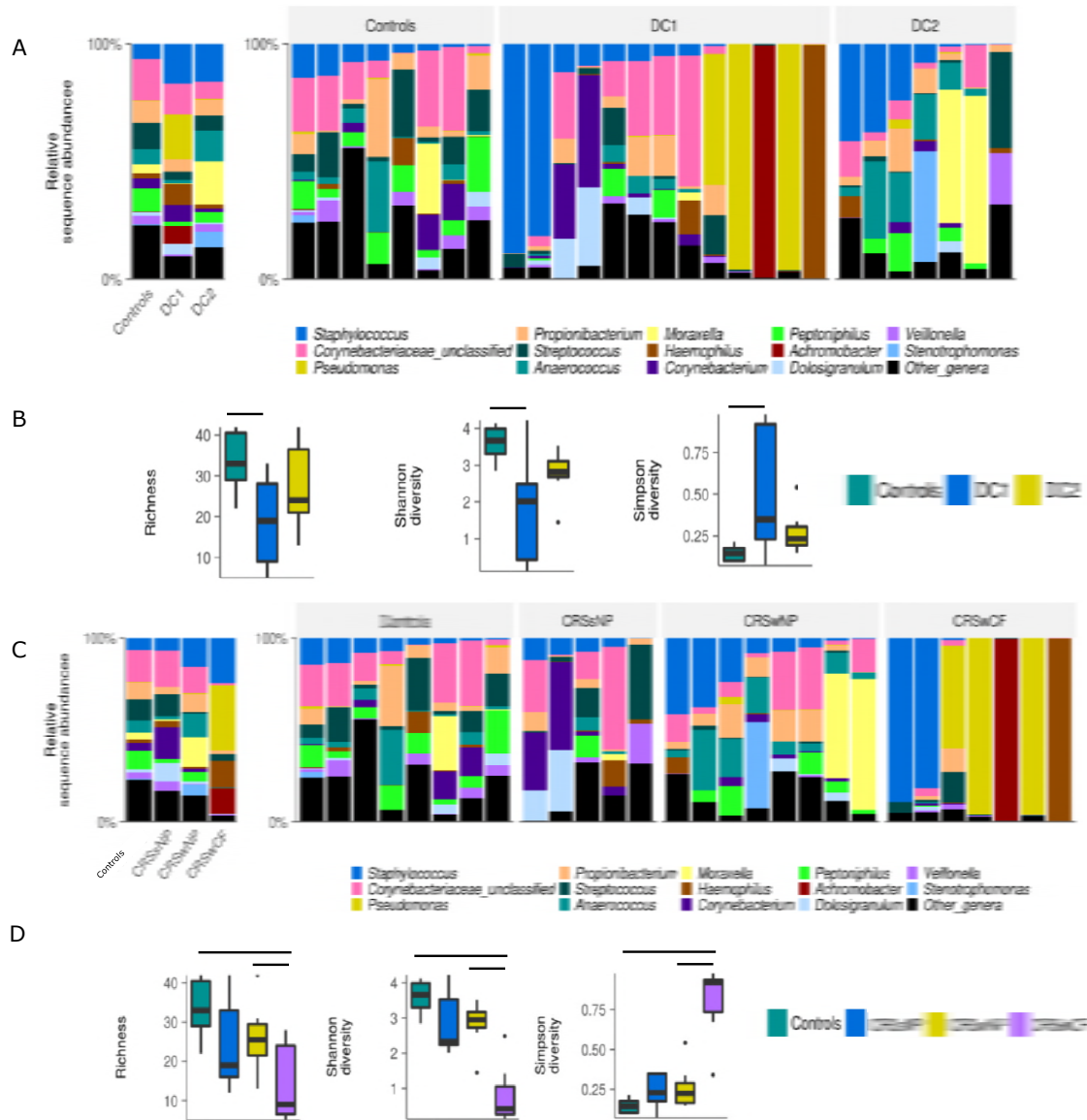671 and ** represents p values below 0.01.

Fig. 1: Bacterial community composition and alpha diversity for the CRS cohorts and disease controls. Grouped and patient-level profiles at genus-level are shown in the bar graphs for (A) DC groupings and (C) phenotypic grouping. Box-and-whisker plots represent group summaries for bacterial richness, Shannon diversity and Simpson diversity for (B) DC groupings and (D) phenotypic groupings. Horizontal lines represent significant differences ($p < 0.05$) between cohorts.
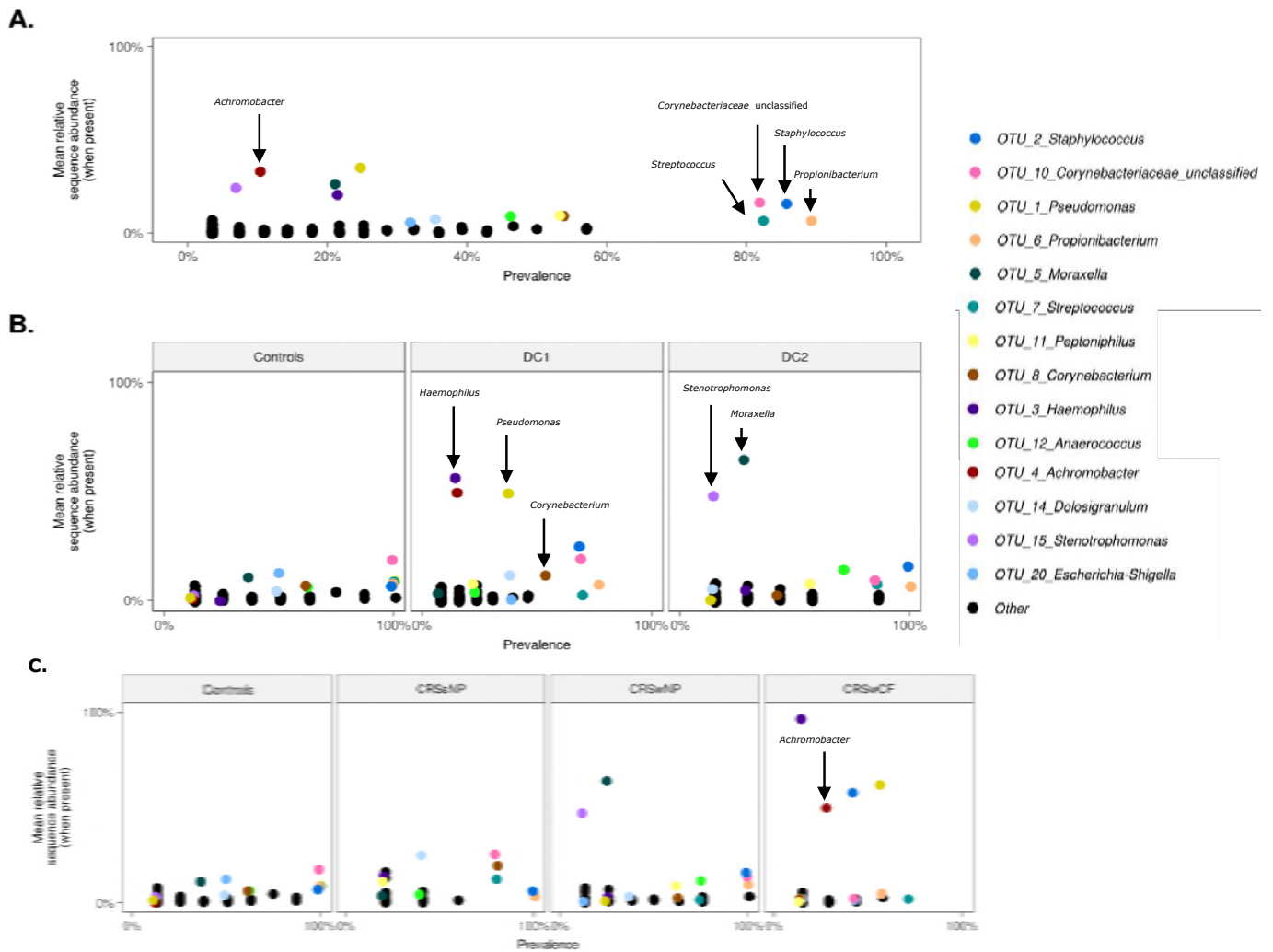
Fig. 2: Mean OTU relative abundance per sample (when present in a sample) plotted against prevalence (occurrence) across (A) all samples, (B) each DC and control, and (C) each phenotypic CRS subtype and control. All OTUs are plotted. The 14 most abundant OTUs are color-coded, and all other OTUs are presented as black dots.
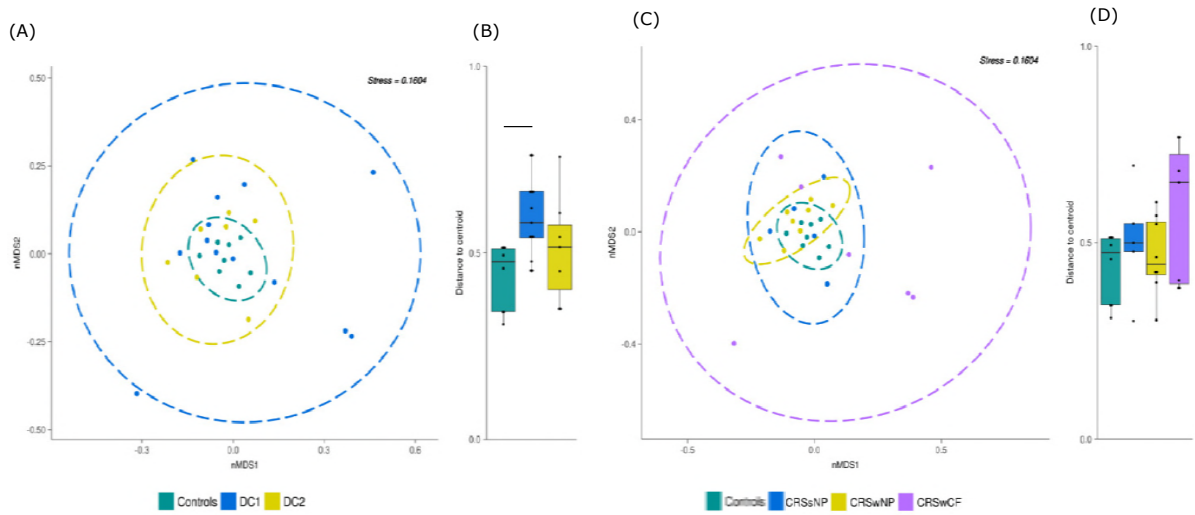
Fig. 3: Non-metric multidimensional scaling plot using Bray-Curtis dissimilarity distances (weighted) for all samples for (A) DC clustering and (C) phenotypic subtyping. Ellipses represent the 95% CI spread from centroids. Box-and-whisker plot of distances between each subject to the centroid of their respective group (B) and (D). Beta-dispersion was significantly different between DC1 and controls ($p = 0.0114$; Tukey's multiple comparisons of means).
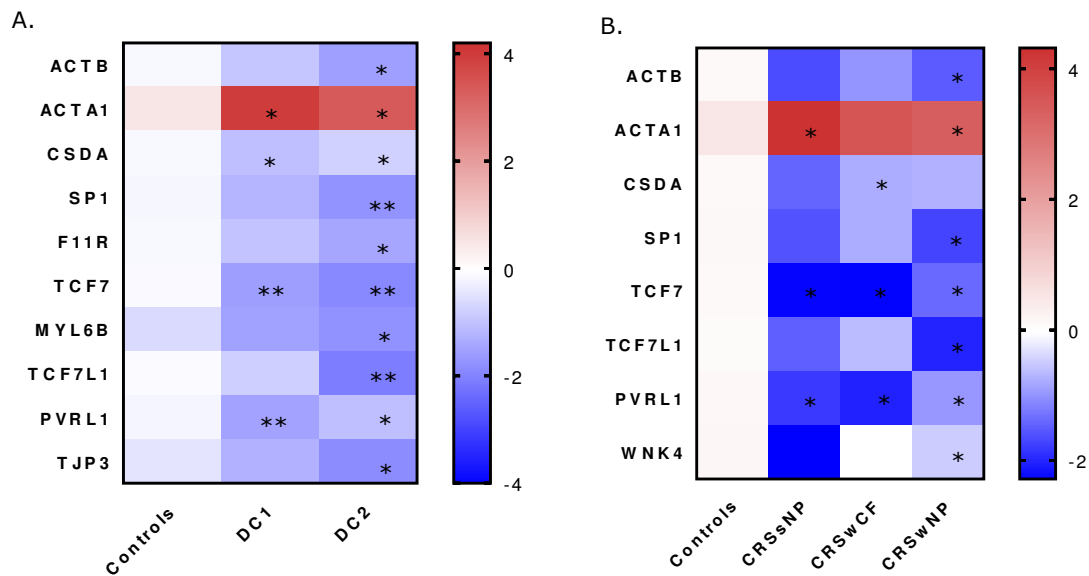
Fig. 4: Tight junction genes that were significantly different in expression for (A) DC groups and (B) phenotypic subgroups compared with controls are displayed in the heatmap. Blue represents genes that have low level expression, and red represents higher expression. Values displayed are log-transformed mean fold-change ($2^{-\Delta\Delta Ct}$). * represents p values below 0.05 and ** represents p values below 0.01.