

1 Supplemental material

2

3 Microbiome data analysis characteristics

4 After quality filtering, 39.6 million sequences were obtained for subsequent analysis
5 from 56 million paired-end raw reads with an average of 0.19-3.5 million reads per
6 sample. The rarefaction plots for the number of observed species, phylogenetic
7 diversity, Shannon and Chao Indices reached a plateau for all the samples rarified for
8 0.12 million sequences per sample (Supplementary Figure 1a). This plot indicates
9 sufficient sequencing depth to cover the maximum microbiome diversity, richness, and
10 observed species in each sample for further analysis (1, 2). A total of 21852 OTUs at
11 the 97% sequence similarity were obtained, with 13,650 OTUs in pre-cancer 19502 in
12 cancer and 7207 in the adjacent tumor group. The oral microbiome of the study was
13 assigned to a total of 51 phyla and 1055 genera, of which the core microbiome that
14 signifies taxa with the average relative abundance of equal to or more than 1% in any
15 group was identified at each of the hierarchical levels as follows (a) Phyla: 10 out of 51
16 (b) Genera: 50 out of 1055 (c) Species: 19 out of 67 identified from 10 genera having
17 a relative abundance of more than 5%. The predominant phyla (top 5) belonged to
18 *Firmicutes* (30.5%), *Proteobacteria* (27.3%), *Bacteroidetes* (12.8%) *Fusobacteria*
19 (6.2%) and *Actinobacteria* (5.2%) while predominant genera (top 5) were
20 *Streptococcus* (10.3%), *Capnocytophaga* (6.4%), unidentified genera of
21 *Enterobacteriaceae* (5.4%), *Neisseria* (3.3%) and *Leptotrichia* (3.2%) in all samples
22 combined (Supplementary table 1).

23 1. Mukherjee S, Mitra R, Maitra A, Gupta S, Kumaran S, Chakraborty A, Majumder PP.

24 2016. Sebum and Hydration Levels in Specific Regions of Human Face Significantly
25 Predict the Nature and Diversity of Facial Skin Microbiome. *Sci Rep* 6:36062.

26 2. Willis AD. 2019. Rarefaction, Alpha Diversity, and Statistics. *Front Microbiol* 10:2407.

27

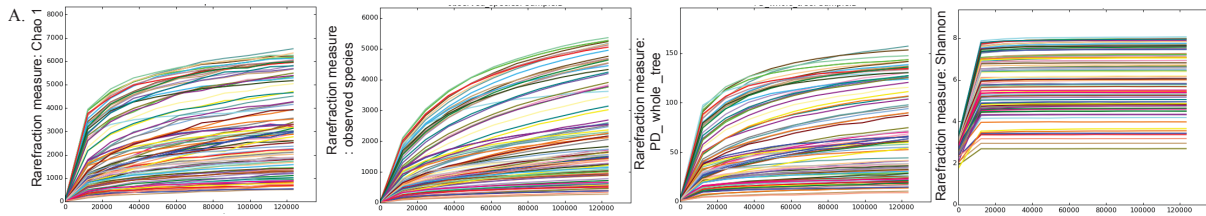
28

29

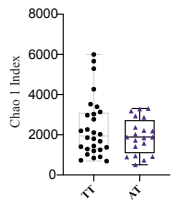
30

31 Supplemental figures:

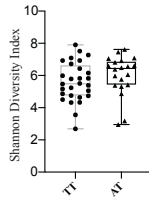
32



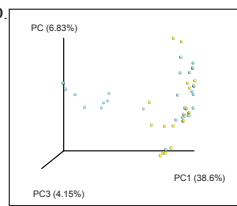
33 B.



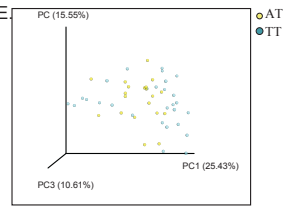
34 C.



35 D.



36 E.



37

38

39

40

41

42

43

44

45

46

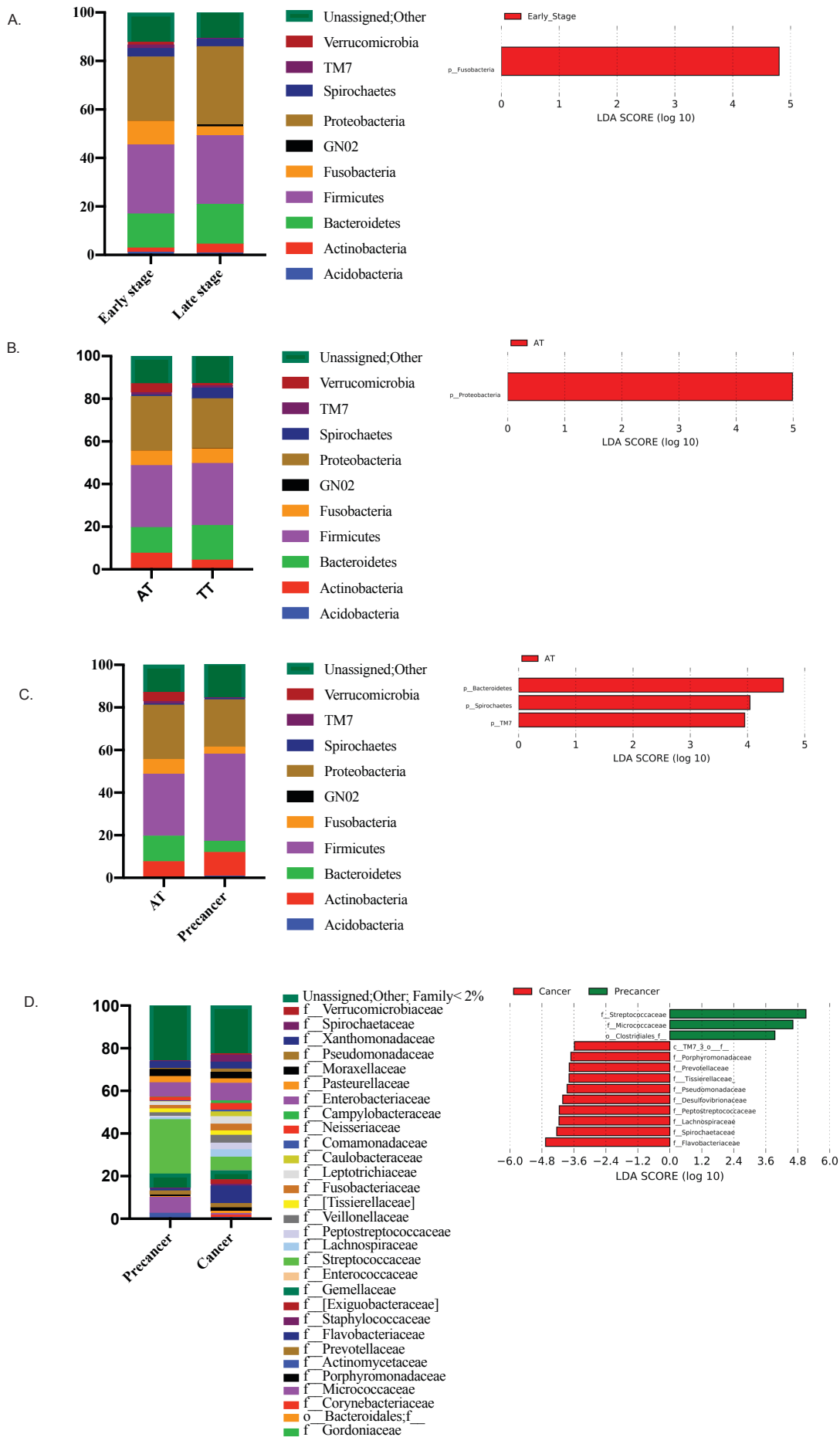
47

48

49

50

Supplemental figure 1: (A) Rarefaction analysis of 16S rRNA gene sequence obtained in present data set by four measures: Chao1, observed species, PD whole tree and Shannon indices. Bacterial alpha diversity indices among tumor tissue and adjacent tumor tissue samples measured by (B) Chao1 and (C) Shannon diversity indices. Statistical analysis was done by t-test, Mann-Whitney test. Bacterial beta diversity indices between tumor tissue and adjacent tumor tissue samples measured by (D) unweighted UniFrac distance (E) weighted UniFrac distance.



52 **Supplemental figure 2:** (A) Relative and differential abundances of core bacterial phyla among early
53 (T1 and T2) and late stages (T3 and T4) of cancer groups. (B) Relative and differential abundances of
54 core bacterial phyla among adjacent tumor tissue and tumor tissues. (C) Relative and differential
55 abundances of core bacterial phyla among adjacent tumor tissue and pre-cancer stage. (D) Relative
56 and differential abundances of bacteria at family level among pre-cancer and cancer groups. Differential
57 abundance analysis was done by LEfSe and represented through LDA plot.

58

59

60

61

62

63

64

65

66

67

68

69

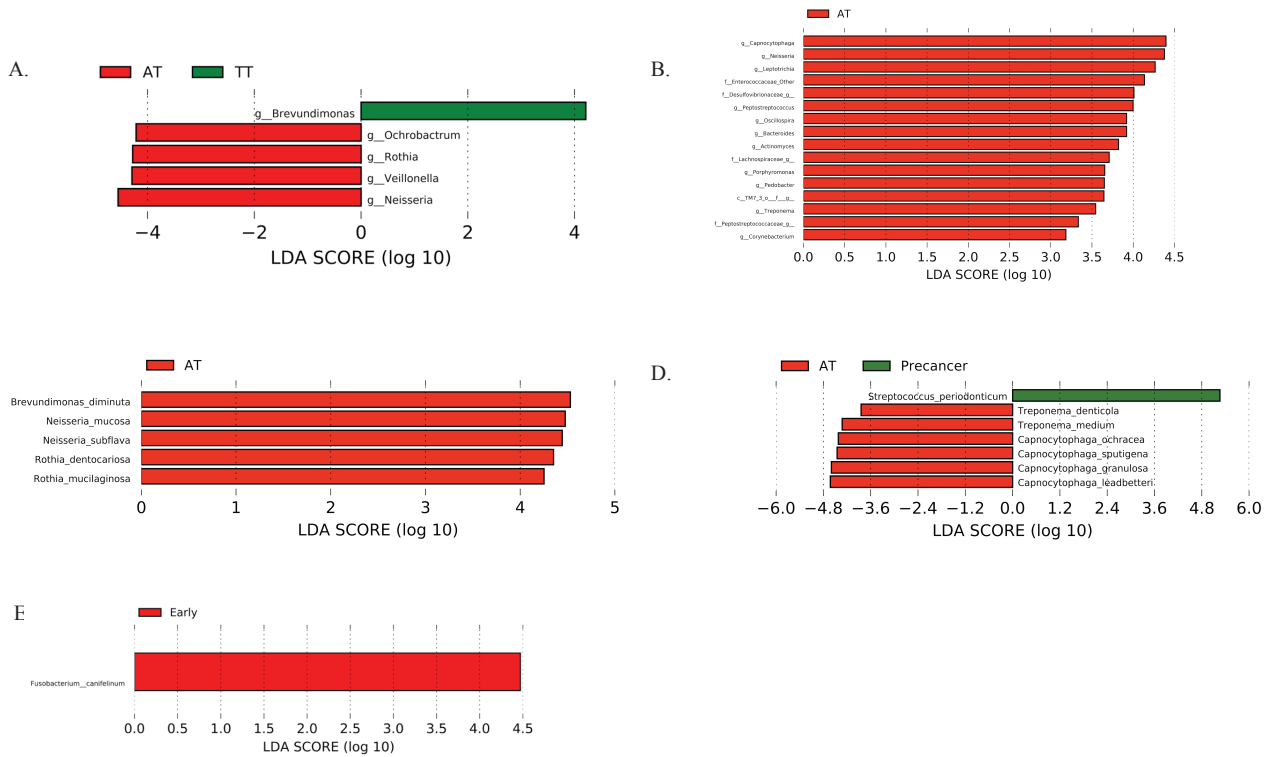
70

71

72

73

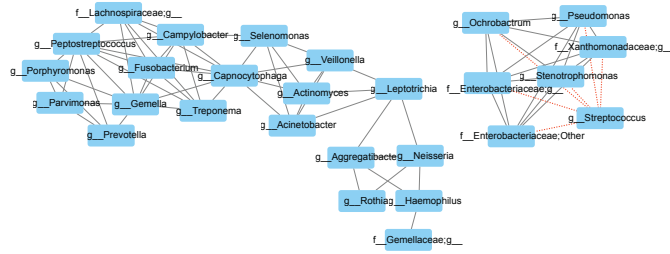
74



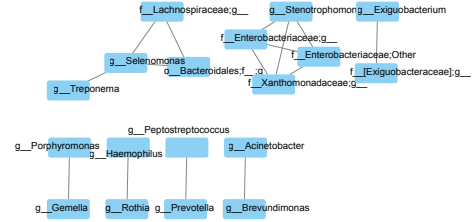
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93

Supplemental figure 3: Differential abundances of core bacterial genera among (A) adjacent tumor tissue (AT) and tumor tissues (TT). (B) pre-cancer and adjacent tumor tissue. Differential abundances of bacteria at species level among adjacent tumor tissue (AT) and (C) tumor tissue (TT) (D) and pre-cancer groups. (E) Differential abundances of bacteria at species level between early (T1 and T2) and late cancer stage (T3 and T4) groups. Differential abundance analysis was done by LEfSe and represented through LDA plot.

A.



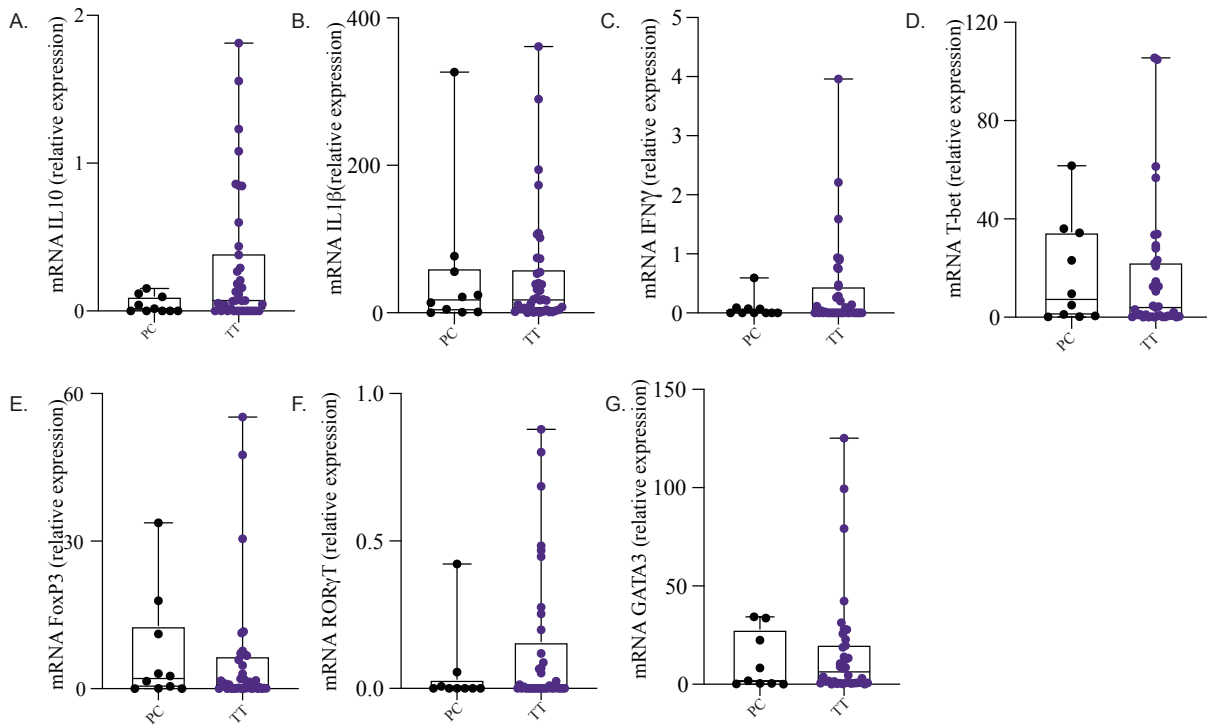
B.



94
95

96 **Supplemental figure 4: Co-occurrence analysis.** Co-occurrence analysis of top 30 bacterial genera
 97 in (A) pre-cancer and (B) cancer groups with Pearson's correlation analysis presented through network
 98 analysis. Black line: positive interaction; red dotted line: negative interaction.

99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116



117

118 **Supplemental figure 5:** Relative expression of cytokine genes and T cell associated genes in pre-
 119 cancer and cancer group (A) *IL10* (B) *IL1β* (C) *IFNγ* (D) *T-bet* (T_{H1}) (E) *FoxP3* (Treg) (F) *RORγT* (T_{H17})
 120 (G) *GATA3* (T_{H2}). n=10 (Pre-cancer group); n=40 (Cancer group). Statistical analysis was done t-test
 121 followed by Mann-Whitney test. PC: pre-cancer, TT-tumor tissue

122

123

124

125

126

127

128

129

130

131

132

133

134

135 **Supplemental table 2:** Dominant species having more than 2% of relative abundance.

136

Species	Relative abundances
<i>Rothia__mucilaginosa</i>	6.169
<i>Capnocytophaga__granulosa</i>	3.464
<i>Capnocytophaga__leadbetteri</i>	7.468
<i>Capnocytophaga__ochracea</i>	2.476
<i>Capnocytophaga__sputigena</i>	5.868
<i>Exiguobacterium__aurantiacum</i>	2.568
<i>Streptococcus__dysgalactiae</i>	2.992
<i>Streptococcus__periodonticum</i>	21.652
<i>Fusobacterium__nucleatum</i>	7.040
<i>Brevundimonas__diminuta</i>	5.058
<i>Neisseria__mucosa</i>	4.865
<i>Neisseria__subflava</i>	3.158
<i>Haemophilus__parainfluenzae</i>	5.236
<i>Treponema__medium</i>	3.525
<i>Akkermansia__muciniphila</i>	3.672

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154 **Supplemental table 3:** Summary of the area under ROC curve (AUC) indices using
155 differential genus in precancer and cancer samples.

156

Predictive Genus	AUC	95% CI	P value
<i>Streptococcus</i>	0.7874	0.6501 - 0.9246	0.0006
<i>Rothia</i>	0.7529	0.6267 - 0.8791	0.0027
<i>Capnocytophaga</i>	0.8103	0.6964 - 0.9243	0.0002
<i>Treponema</i>	0.7092	0.5758 - 0.8425	0.0130
<i>Leptotrichia</i>	0.6782	0.5260 - 0.8303	0.0343

157 95% CI;95% confidence interval

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172 **Supplemental table 4 -Significantly correlated TILs with most abundant genera**
 173 **of tumour microenvironment.**

174

Tumour infiltrating lymphocytes	Genus	Pearson correlation coefficient (r)	P value
CD19	<i>Porphyromonas</i>	0.5847	0.0043
	<i>Streptococcus</i>	0.4902	0.0205
	<i>Oscillospira</i>	0.4607	0.031
	<i>Aggregatibacter</i>	0.4717	0.0267
CD 4	<i>Brevundimonas</i>	0.4939	0.0372
CD 4 Naive	<i>Rothia</i>	0.6261	0.0294
	<i>Porphyromonas</i>	0.8424	0.0006
	<i>Prevotella</i>	0.5869	0.0449
	<i>Gemellaceae;g__</i>	0.833	0.0008
	<i>Streptococcus</i>	0.6721	0.0167
	<i>Oscillospira</i>	0.6423	0.0243
	<i>Aggregatibacter</i>	0.6193	0.0318
CD 4 CM	<i>Porphyromonas</i>	0.7811	0.0027
	<i>Prevotella</i>	0.7172	0.0087
	<i>Gemellaceae;g__</i>	0.62	0.0315
	<i>Oscillospira</i>	0.5767	0.0496
CD 4 TEMRA	<i>Fusobacterium</i>	0.8277	0.0009
	<i>Leptotrichia</i>	0.7631	0.0039
	<i>Capnocytophaga</i>	0.7192	0.0084
	<i>c__BD1-5;o</i>	0.9427	<0.0001
CD 4 EM	<i>Porphyromonas</i>	-0.6775	0.0155
	<i>Prevotella</i>	-0.647	0.023

	<i>Exiguobacterium</i>	0.6118	0.0345
	<i>Gemellaceae;g__</i>	-0.6679	0.0176
	<i>Fusobacterium</i>	-0.6682	0.0175
	<i>Campylobacter</i>	-0.8084	0.0015
CD 8 Naïve	<i>Porphyromonas</i>	0.7569	0.0044
	<i>F_Gemellaceae;g__</i>	0.8362	0.0007
	<i>Gemella</i>	0.6605	0.0194
	<i>Streptococcus</i>	0.8303	0.0008
CD 8CM	<i>Campylobacter</i>	0.5851	0.0457
CD 8 TEMRA	<i>Pedobacter</i>	0.6351	0.0265
	<i>Leptotrichia</i>	-0.58	0.0481
	<i>Capnocytophaga</i>	-0.5694	0.0533
CD 8 EM	<i>Porphyromonas</i>	-0.6252	0.0297
	<i>F_Gemellaceae;g__</i>	-0.6108	0.0349
	<i>Gemella</i>	-0.6287	0.0285
	<i>Streptococcus</i>	-0.7388	0.0061

175 CM; Central memory; EM; Effector memory; TEMRA: Effector memory cell re-expressing CD45RA

176

177 Tumour-infiltrating lymphocytes percentage value of flow cytometry data were correlated with relative
178 abundances of intra-tumoral bacteria at genus level with Pearson's correlation parameter. Values with
179 Pearson's co-relation coefficient (r) >0.5 and p value <0.05 are shown in the table.

180

181

182

183

184

185

186

187

188

189

190
191

Supplemental Table 5- List of antibodies used in flowcytometry and IHC.

Antibody	Clone	Sources
CD 19-FITC	HIB19	BD Biosciences
CD 3-PERCP	SK7	BD Biosciences
CD 4-FITC	RPA-T4	BD Biosciences
CD 8-PE	RPA-T8	BD Biosciences
CD 45RA-APC	HI100	BD Biosciences
CD 45RO-APC H7	UCHL1	eBioscience
CCR7- PE-Cyanine7	3D12	eBioscience
CD24-PE	ML5	BD Biosciences
CD27-PECY7	M-T271	BD Biosciences
CD38-APC	HIT2	BD Biosciences
CD20cy	L26	Dako
CD3	-	Dako
IgA	-	Dako

192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228

229 **Supplemental Table 6 - Primers used for RT- qPCR.**

230

Gene name	Primer sequence
IL-1 β	CGCCAATGACTCAGAGGAAG
	AGGGCGTCATTCAGGATCAA
IL 6	GTAGCCGCCCCACACAGACAGCC
	GCCATCTTTGGAAGGTTT
IFN γ	TCAGCTCTGCATCGTTTTGG
	GTTCCATTATCCGCTACATCTGAA
TNF α	TCTTCTCGAACCCCGAGTGA
	CCTCTGATGGCACCACCAG
IL 10	GTGATGCCCCAAGCTGAGA
	CACGGCCTTGCTCTTGTTTT
TGF- β	CAGCAACAATTCCTGGCGATA
	AAGGCGAAAGCCCTCAATTT
T-bet	CCCCTTGGTGTGGACTGAGA
	ACGCGCCTCCTCTTAGAGTC
GATA-3	GTCTCCCTGAGCCACATCT
	GTGGTCCAAGGACAGGCTG
ROR γ t	CGCTCCAACATCTTCTCC
	CTAACCAGCACCCTTCC
Foxp3	CAGCCATGATCAGCCTCACA
	GCACTGGGATTTGGGAAGGT
GAPDH	CCTGCACCACCAACTGCTTA
	GGCCATCCACAGTCTTCTGAG

231

