

1 Title: The gut microbiome associated
2 with LGI1- and CASPR2-antibody
3 encephalitis.

4 Main Manuscript

5 **Authors:** Edmund Gilbert^{1,2}, Sophie Binks^{3,4}, Valentina Damato⁵, Christopher Uy⁶, Paula Colmenero⁶,
6 Mohamed Ibrahim Khalil^{1,2}, Marcus O'Brien⁷, Marcus Claesson^{7,8,9}, John F Cryan^{7,9}, Norman
7 Delanty^{1,2}, Sarosh R Irani^{3,10*}, Gianpiero L Cavalleri^{1,2*}

8

9 * Authors contributed equally.

10

11 **Affiliations:**

- 12 1. School of Pharmacy and Biomolecular Sciences, Royal College of Surgeons in Ireland, Dublin,
13 Ireland.
- 14 2. FutureNeuro SFI Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland.
- 15 3. Oxford Autoimmune Neurology Group, Nuffield Department of Clinical Neurosciences,
16 University of Oxford, Level 3, West Wing, John Radcliffe Hospital, Oxford, OX3 9DS, UK.
- 17 4. Dept of Neurology, John Radcliffe Hospital, Oxford, OX3 9DU.
- 18 5. Department of Neurosciences Drugs and Child Health, University of Florence, Florence, Italy.
- 19 6. Oxford Centre for Microbiome Studies, Kennedy Institute, University of Oxford, Oxford, OX3
20 7DQ, UK. SeqBiome Ltd., Moorepark Food Research Centre, Teagasc, Fermoy, P61 C996 Cork,
21 Ireland
- 22 7. APC Microbiome Ireland, University College Cork, T12 YT20 Cork, Ireland.
- 23 8. School of Microbiology, University College Cork, T12 YT20 Cork, Ireland.
- 24 9. Department of Anatomy and Neuroscience, University College Cork, T12 YT20 Cork, Ireland.
- 25 10. Departments of Neurology and Neurosciences, Mayo Clinic, Jacksonville, Florida, USA.

26

27 Abstract

28 Autoimmune encephalitis is a cause of brain inflammation characterised by auto-antibodies which
29 target cell surface neuronal proteins, and lead to neuronal dysfunction. In older people, common
30 forms are encephalitis with autoantibodies to leucine-rich glioma inactivated protein 1 (*LGI1*) and
31 contactin associated protein like 2 (*CASPR2*), whose presentation includes frequent focal seizures.
32 The exact cause of these autoantibodies remain unknown, but established predispositions include
33 overrepresented human leukocyte antigen (HLA) alleles. Yet, these alleles are themselves common in
34 the healthy ancestry-matched population. One potential aetiological hypothesis is that an
35 environmental trigger, such as the gut microbiome, interacts with a genetically predisposed
36 individual. To investigate this, we studied 47 patients with leucine-rich glioma-inactivated 1 (LGI1)- or
37 contactin-associated protein 2 (CAPSR2)-antibody encephalitis (LGI1/CASPR2-Ab-E) and 37
38 familial/environmentally matched controls, and performed metagenomic shotgun sequencing, to
39 describe compositional and functional differences in the gut microbiome. We observed that
40 LGI1/CASPR2-Ab-E gut microbiomes exhibited a significant reduction in the ratio of *Firmicutes* and
41 *Bacteroidetes* phyla, which associated with dosage of HLA susceptibility alleles in LGI1-Ab-E patients.
42 Furthermore, we identified differences in functional gene profiles in the gut microbiome that led to a
43 reduction of neuroinflammatory protective short-chain-fatty-acids (SCFA) in LGI1-Ab-E patients.
44 Taken together, our results suggest that a compositional shift in the gut microbiome of LGI1/CASPR2-
45 Ab-E associates with a neuroinflammatory state, possibly through the reduction of SCFA production.
46 Our study highlights the potential of the gut microbiome to explain some of the complex condition
47 and unravel aetiological questions. Validation studies with greater sample sizes are recommended.

48

49 Introduction

50 Autoimmune encephalitis (AE) is characterised by the presence of autoantibodies with pathogenic
51 potential which target cell surface neuronal proteins¹⁻³. Among the most common autoantibodies
52 detected in AE are those which bind leucine-rich glioma inactivated protein 1 (LGI1) and contactin
53 associated protein like 2 (CASPR2). Many of these patients typically have very frequent, focal seizures
54 in addition to altered mood, personality change, and cognitive impairment⁴. This semiology of these
55 seizures can be pathognomonic, as is well-established for patients with faciobrachial dystonic
56 seizures in association with LGI1-antibodies⁵⁻⁷. In addition, other unusual semiologies in LGI1-
57 antibody encephalitis (LGI1-Ab-E) include thermal, piloerection, autonomic and emotional events^{8,9}.
58 Patients with CASPR2-antibodies also have very frequent focal seizures¹⁰⁻¹². These clinical
59 observations imply that autoantibodies mediate forms of seizures which are likely to have
60 characteristic, and sometimes novel, underlying molecular mechanisms.

61 In addition to distinctive clinical features, these AE patients are often anti-seizure medication (ASM)
62 resistant, but show beneficial responses to immunotherapies, including corticosteroids, plasma
63 exchange, and intravenous immunoglobulins^{5,6,13,14}. Yet, despite immunotherapies and ASMs, more
64 than 70% of patients are left with long term deficits, which impair quality of life^{5,15,16}. In addition,
65 around 40% of people treated with immunotherapies experience adverse drug effects^{5,6}, and in the
66 case of LGI1-Ab-E up to 20% develop chronic epilepsy¹⁷. Hence there is an unmet need for
67 personalised immunotherapies that could be safe and effective in the long term.

68 Despite its potential to explain causation and address therapeutic needs, the aetiology of these
69 diseases is poorly understood. Recent works have demonstrated that more than 90% of patients with
70 LGI1-antibodies carry specific class-II HLA alleles, namely HLA-DRB1*07:01 (OR: 26.4 [CI=8.5-81.5]¹⁸,
71 OR: 85.5 [CI=10.4-699.6]¹⁹). Whereas CASPR2-antibody AE show a strong but distinct HLA profile with
72 an over-representation of HLA-DRB1*11:01²⁰ (OR: 9.4 [CI=4.6-19.3]²¹). A striking feature of these
73 genetic findings is the relatively common underlying population frequencies of the HLA risk alleles.
74 The HLA-DRB1*07:01 allele, for example, is also carried by 27% of healthy European-ancestry
75 populations^{18,21} (~12% globally²²) and HLA-DRB1*11:01 is present in 9% of the general European-
76 ancestry population²¹ (~5% globally, maximal in Europe²²). Whilst these strong HLA Class II
77 associations clearly implicate T cells in disease pathogenesis, their high general population
78 frequencies only partially explain disease causation. Hence, we examined potential contributory
79 environmental triggers.

80 The gut microbiome elegantly provides an interface to explain such an environmental trigger. Indeed,
81 in a variety of autoimmune central nervous system diseases, there is growing evidence for an
82 aetiological link between the gut microbiome and the brain, principally via innate immunity and T-
83 cells²³⁻²⁵. In several autoimmune diseases, shifts in gut microbial populations have been discovered
84 which may predispose to the illness^{23,26-28}. Further, in autoantigen-specific illnesses, a distinctive
85 microbial signature may provide a similarly elegant link to the LGI1 and CASPR2 autoantigens
86 through molecular mimicry.

87 Here, we investigated the gut microbiome in patients with LGI1- and CASPR2-Ab-E. We also collected
88 stool and salivary DNA from matched but unaffected relatives, close family members, or friends. We
89 sought to; 1) determine if species abundance or diversity in the gut microbiome was significantly
90 different between healthy and affected family members, 2) characterise potential functional
91 differences in these microbiotas, and finally 3) investigate the evidence of LGI1 sequence
92 homologues within the gut microbiota of LGI1-Ab-E patients.

93 Results

94 Recruitment and Cohort

95 We recruited 47 patients with LGI1- (n=43), CASPR2-antibody (n=3), or double positive LGI1/CASPR2-
96 antibody (n=1) encephalitis, henceforth collectively referred to as LGI1/-CASPR2-Ab-E. In parallel, we
97 recruited matched healthy controls (HC, n=37) through close family members who were typically a
98 domestic partner (see **Table 1**). We isolated both genomic DNA from blood and gut microbiome DNA
99 from stool (see Methods). All cases and HC were consistent with British- or Irish-like ancestry in
100 comparison to ancestry controls (**Supp Figure 1**). Tissue typing confirmed an enrichment of the HLA
101 DRB*07:01 allelotype in LGI1 patients (**Table 1**) compared to SNP genotype imputation of HLA alleles
102 in HC.

103 Gut Microbiome Composition

104 We first compared the composition of the gut microbiome in cases versus controls. We generated
105 gut microbiome genomic profiles from paired-end metagenomic shotgun sequencing (see Methods
106 and **Supp Figure 2-3**) and investigated composition changes with differential abundance and alpha
107 and beta diversity analysis. The shotgun sequencing generated an average of 8.7 million paired end
108 reads per sample where 84% of these reads passed the adapter trimming and quality filtering stage.
109 An average of 91% of reads passed quality control and were classified using DNA species abundance
110 estimators, *Kraken2* and *Bracken*.

111 The general compositional profile of cases and HC were consistent with the expected common
112 human gut microbiome species such as *Bacteroides* or *Faecalibacterium*²⁹ (**Figure 1A**). We tested for
113 differences in abundance of individual species and identified no statistically significant differences
114 that survived correction for multiple testing (**Table 2**). Given the rarity of this disease and the fine-
115 grained nature of individual species composition, the 16 most differentially abundant species are
116 reported with uncorrected p-values of < 0.05 (**Table 2, Supp Data 2**). Several of the species nominally
117 enriched in LGI1-CASPR2-Ab-E cases are also associated with the human oral and small intestinal
118 microbiome (e.g., *Streptococcus parasanguinis*, *Streptococcus salivarius*, *Bifidobacterium dentium*,
119 and various *Veillonella*).

120 To complement and extend this fine-grained species analysis, we simultaneously investigated phyla-,
121 genera-, and species-level enrichment between cases and controls using linear discriminant analysis
122 (LDA) (**Figure 2, Supp Data 3**). Supportive of previous species-level analysis, our LDA results also
123 provide evidence of human oral and small intestinal species of the *Veillonella* and *Streptococcus*
124 genera enriched in cases. In addition, our cases were enriched for the *Enterococcus* and *Haemophilus*
125 *D* genera, with depletion of the *Dorea* and *Coprococcus* genera, and the *Firmicutes* phylum (recently
126 renamed *Bacillota*^{30,31}). Beyond these depletions in HCs, we also observe a depletion specifically of
127 *Bifidobacterium longum* in cases (**Figure 2**).

128 From the LDA, the differential balance of the *Firmicutes* and *Bacteroidetes* phyla in cases and
129 controls was of interest. Whilst changes in the *Firmicutes/Bacteroidetes* (F/B) ratio has been
130 implicated in some human health-related traits³²⁻³⁴, this remains debatable³². We investigated such
131 large-scale changes in the context of our modest sample size. As expected³³ these two Phyla account
132 for the majority of phyla represented in the analysis - 52% + 39.8% (91.2% - Cases) and 30.4% +
133 61.7% (92.1% - HC). Although the F/B ratio was higher in controls (2.35) than cases (1.21), but this
134 was not statistically significant (Wilcoxon Rank Sum p-value = 0.060) (**Supp Figure 4**).

135 Considering this F/B ratio difference between cases and controls we explored whether the dosage of
136 HLA-DRB1*07:01 impacts this F/B composition. Analysing the F/B ratio in HC and the LGI1-Ab-E cases

137 (as these were the largest case subtype) (**Supp Figure 4**), we found the F/B ratio to be significantly
138 lower in LGI1 positive cases relative to HCs (cases: 1.21, HCs: 2.35, Wilcox two-sided p-value: 0.032).
139 As expected, we found an enrichment of HLA-DRB1*07:01 heterozygotes in *LG11*-Ab-E cases versus
140 controls (HWE exact p-value: 3×10^{-5}) but not in controls (exact p-value: 0.16) (**Supp Figure 5**). We
141 observed the F/B ratio to decrease as a measure of HLA-DRB1*07:01 dosage (Pearson's product-
142 moment correlation = -0.31 (95% confidence intervals -0.56 to -0.006), p-value = 0.046). The same
143 pattern was no observed in the HCs (p-value = 0.95) (**Supp Figure 5**). Overall, this is the first hint of
144 gut composition differences between *LG11*-Ab-E cases and controls associated with HLA-DRB1*07:01
145 genotype.

146 We next investigated within-gut and between-gut diversity. Testing within-gut with Chao1, Simpson,
147 and Shannon alpha diversity measures we found evidence of subtle differences between case and HC
148 status – consistent with the observations from differential species analysis and LDA. We observed
149 Simpson alpha diversity to be significantly enriched in HC compared to cases (p-value = 0.0313), but
150 not Shannon (Shannon p-value = 0.1112) (**Figure 1B**, **Supp Figure 6**). Investigating between-gut
151 diversity we performed principal component analysis (PCA) of Aitchison distances of this abundance
152 data. We observed a borderline difference between cases and HC along principal component 2 (p-
153 value = 0.083) (**Figure 1C**, **Supp Figure 7**). We also observe significant differences between dietary
154 variables and these principal components (**Supp Data 4**), though the uneven sampling size or
155 extremes in scale (such as vitamin most B12 consumption is “infrequent” or “daily”) leads us to
156 caution against these dietary-based observations.

157 Functional Characteristics

158 We next sought to extend the suggestive findings of differences in case-control gut microbiome
159 profiles, characterising function through KEGG modules and pathways³⁵ using HUMAnN3³⁶ (see
160 Methods). We tested both high-level KEGG modules and lower-level pathways for differential
161 abundance using *Aldex2*³⁷, using an uncorrected p-value of 0.05 similar to our differential species
162 analysis (**Table 3**, **Supp Data 5-6**). We observed three KEGG modules to be significantly different in
163 cases and controls, all enriched in HC. These three modules are particularly interesting in the context
164 of epilepsy. Microbial coenzyme A (CoA) is a key source of short chain fatty acids (SCFA) in the gut²⁸,
165 where SCFAs are protective for neuroinflammation and seizure³⁸ and epilepsy in general³⁹. Ornithine
166 acts as a precursor to CoA in its acetyl-CoA form⁴⁰ and histidine acts as a precursor to ornithine⁴¹. Of
167 the KEGG pathways all but pentose and glucuronate interconversions (KEGG pathway map00040)
168 were also enriched in HCs. In PCA of these KEGG pathway differences, we observe significant
169 clustering of cases versus HC on principal component three (**Supp Figure 8**) (p-value = 0.044) -
170 though this significant clustering was not found in PCA of KEGG module differences (**Supp Figure 9**).

171 Finally, we tested the evidence that the gut microbiome of LGI1-Ab-E patients contains potential
172 bacterial homologues of the human *LG11*, searching the *RefSeq* database for *LG11* homologues with
173 *blastp* (see Methods for quality cutoffs). We identified 13 organisms to contain such homologues
174 (**Supp Data 7**), however the majority potential homologues are annotated as “leucine-rich” or
175 “hypothetical”, and nearly all organisms in which these homologues are observed are typically found
176 in marine environments. Mapping these potential homologues to the metagenomic data of our LGI1-
177 Ab-E cases, we do not observe any hits and no evidence of their presence in the gut microbiome of
178 LGI1-Ab-E cases.

179 Discussion

180 This metagenomic analysis of the gut microbiome in 47 patients with LGI1/-CASPR2-Ab-E, matched
181 with 27 HC with paired genomic data suggests a potential compositional bias in the gut microbiome
182 of AE. Together our results suggest subtle taxa compositional changes in cases with these differences
183 associated with HLA-DRB1*07:01 dosage. Further this compositional change was associated with
184 depletion of KEGG modules linked to SCFA generation in cases.

185
186 We primarily observed that the general gut microbiome of individuals with LGI1/-CASPR2-Ab-E
187 appeared largely consistent with a normal microbiome composition with high abundance of common
188 microbiome species²⁹. We therefore infer that LGI1-CASPR2-Ab-E is not characterised by a large
189 difference in microbiome composition driven by specific species. The most positive signal of differing
190 microbiome composition between cases and controls is the enrichment of *Bifidobacterium Longum*
191 in HC. Interestingly, this species has been observed to exert beneficial health effects, on the human
192 gut⁴², inflammation²⁷, and potentially reducing cytotoxic effects of certain ASMs⁴³.

193
194 Instead, linear discriminant analysis highlights several taxa potentially enriched in either our cases or
195 controls, with some of these taxa previously associated with traits relevant to LGI1-CASPR2-Ab-E.
196 Several genera enriched in our cases have been associated with neurological conditions and allows
197 us to place our results into the context of the interplay of the microbiome and neuroinflammation
198 and epilepsy. *Streptococcus*, our strongest signal, has been found to be associated with focal epilepsy
199 in children⁴⁴ where in a small sample (n=10) the authors posit a neuroinflammatory model. Indeed, a
200 mouse model of intestinal inflammation increases convulsant activity and ASM resistance⁴⁵. In
201 addition to *Streptococcus*, the taxa *Enterococcus*⁴⁶ and *Roseburia*⁴⁷ have been found to be associated
202 with intractable childhood or drug-resistant epilepsy (respectively), and *Roseburia* in another study
203 of LGI1-Ab-E⁴⁸. The genera *Coprococcus* and *Dorea* are both enriched in our HC. These have been
204 associated with drug-resistant epilepsy⁴⁷, but *Coprococcus* is also noted for its anti-inflammatory
205 associations³³ and is depleted in our cases. Our results would support larger, multi-centre studies of
206 the role of the gut microbiome in epilepsy healthcare, from ASM use to disease aetiology and
207 longitudinally from onset to chronic as has been achieved in the cases reported here.

208
209 In addition to subtle genera or species differences, our primary observation is the reduction of F/B
210 ratio in LGI1-Ab-E cases compared to HC. This reduction in the F/B ratio appears to be associated
211 with HLA DRB*07:01 allele dosage. The frequency of HLA DRB*07:01 in our cases (91% carriers)
212 matches that observed in cohorts of similar ancestry^{19,21}, and our AE cases were primarily LGI1
213 positive (n=43/47) – in keeping with literature¹⁰. To our knowledge this is the first observation linking
214 gut microbiome compositional changes to the major LGI1-Ab-E genetic risk locus. Encouragingly a
215 reduction of the F/B ratio and taxa diversity has been associated in another gut microbiome study of
216 LGI1-Ab-E in a cohort of 15 patients of Han Chinese ancestry compared to 25 matched controls using
217 16S rRNA sequencing⁴⁸. Although the Chinese investigation did not consider HLA genotypes, together
218 our studies show evidence that a subtle shift in taxa composition measured by the F/B ratio is
219 associated with LGI1-Ab-E and is consistent with an inflammatory disease model⁴⁸. That said, for
220 obesity, one of the first traits where the F/B ratio was suggested as a biomarker, results are
221 frequently discrepant³² and that phyla encompass a wide variety of taxa with differing functional
222 attributes²⁶. Wider and further studies are warranted in neuroinflammatory and epilepsy traits, to
223 investigate the veracity of our observation. These F/B ratio findings appear to be complemented by
224 our functional characteristics results. We find KEGG modules that are nominally depleted in cases,
225 and these modules are associated with CoA synthesis, itself leading to the synthesis of SCFA, which
226 are believed to have an anti-neuroinflammatory effect^{39,48}. Taken together with the enrichment of
227 species such as *Bifidobacterium Longum* in HC, our results suggest a model whereby compositional
228 changes in the gut microbiome are associated with functional shifts linked to a loss of
229 neuroprotective SCFA in a context of a specific HLA class II genotype.

230

231 In addition to these observed compositional and functional changes, we found in LGI1/-CASPR2-Ab-E
232 cases there is an enrichment of species that are normally associated with the oral or small intestinal
233 microbiomes. It has been observed that corticosteroids impact saliva quality⁴⁹ in a way that could
234 result in accelerated movement of matter through the digestive system, leading to abundance of oral
235 or small intestinal communities. Corticosteroids are a commonly prescribed treatment in of AE^{5,6} as
236 early immune therapy and a majority of our patients have been exposed to this treatment (**Table 1**) -
237 therefore it is a possibility that this observation is driven by treatment and not underlying aetiology.
238 In a linear discriminant analysis comparing cases that have a history of steroid treatment versus case
239 with no history, and versus HCs, we find evidence that this signal is driven by steroid use history
240 (**Supp Figure 10**). Furthermore, there is previous evidence ASMs may also impact the gut
241 microbiota⁴³ as does the anti-epileptic ketogenic diet^{50,51}. In the latter case, none of the LGI1-
242 CASPR2-Ab-E patients included in this study were or had been on a ketogenic diet.

243

244 The primary limitation of this study was cohort size. LGI1-CASPR2-Ab-E is a rare disease which can
245 impact cognitive function, and despite recruitment across the UK and Ireland, only 47 cases were
246 available for analysis. Further, we have studied chronic cases, which may have a different
247 composition to the microbiome of cases with an acute onset. We expect that the sample size
248 constraint impacted our ability to detect more subtle differences between LGI1-CASPR2-Ab-E cases
249 and controls, and limits the statistical power of sub-group analysis. Combining with other AE cohorts
250 of a similar dietary and pharmacological background may provide novel insights, though future care
251 will need to be taken in harmonising gut microbiome sequence generation methods to ensure inter-
252 study comparisons are possible. Compositional changes in the microbiome have been associated
253 with ageing⁵², and we lacked information on the ages of most of our HC. However, we expect our
254 cases and controls to be largely aged-matched (and background matched) as our HC are
255 predominantly spousal or sibling in relationship.

256

257 To conclude, our work suggests that LGI1-CASPR2-Ab-E is associated with specific compositional
258 changes in the gut microbiome, shifting the balance between the predominant taxa *Firmicutes* and
259 *Bacteroidetes* in favour of *Bacteroidetes*. This ratio change is associated with the significant HLA risk
260 allelotype DRB*07:01 in cases only and matches previous studies of LGI1-Ab-E cases. This profile is
261 also associated with a shift in functional microbial activity linked to reduced SCFA levels. Given our
262 low sample size, reflective of this rare condition and its effect on statistical power, the exact effect
263 the microbiome plays in disease pathogenesis and/or progression remains unclear. However, our
264 results show a role in a subtle compositional and functional change that is consistent with a
265 neuroinflammatory model and case-specific effect in the HLA DRB*07:01 allelotype. Further work
266 with expanded sample sets and coordinated sequencing protocols appears to be the next step in
267 elucidating the potential trigger for this rare and refractory epilepsy.

268 Methods

269 Ethics Approval

270 This study was approved by ethics committees at the Royal College of Surgeons in Ireland (under
271 protocol REC1631), and Oxford University (under protocol number 16/YH/0013). Ethics for this study
272 was approved in the UK by Yorkshire & The Humber - Leeds East Research Ethics Committee
273 (16/YH/0013). All patients gave written informed consent.

274 Recruitment

275 Eligible patients were aged >18 with a diagnosis of autoimmune encephalitis/epilepsy and a positive
276 test for LGI1 and/or CASPR2 antibodies in serum and/or CSF. HCs were spouses, relatives or friends
277 attending clinic or subsequently recruited from the community, preferentially recruiting genetic
278 relatives, or co-habiting relatives. All participants gave informed consent or assent was given by next
279 of kin. Saliva and stool samples were gathered in the home environment and posted via secure
280 biokit. Stool was collected using a Omnigene Gut stool collection kit (DNA Genotek, OM-200). Saliva
281 was collected using Isohelix GeneFix saliva collection kits (Isohelix, GFX-02). Samples were posted to
282 the lab using an overnight tracked-return service and stored at -80°C on arrival. At the time of
283 sampling study participants also completed a questionnaire covering toileting routine, bowel habits,
284 dietary and medication intake (see **Supplementary Data 1** for this questionnaire).

285 DNA Extraction

286 DNA was extracted from stool and saliva samples within two weeks of their arrival at the lab. Saliva
287 extractions were performed using the Isohelix GeneFix Saliva-Prep DNA Kit and stool extractions
288 using the QIAamp PowerFecal Pro DNA Kit (Qiagen, 51804). DNA quality was assessed using a
289 Nanodrop One spectrophotometer.

290 SNP Genotyping

291 Salivary DNA samples were genotyped on the Illumina OmniExpress chip at Edinburgh Genomics,
292 according to manufacturer's instructions. Genomic ancestry analysis was performed using principal
293 component analysis (PCA) from this SNP genotype data. Case and HC genotypes were merged with
294 Irish and British ancestry controls from the Irish DNA Atlas⁵³ (n=193), Irish Trinity Student⁵⁴ (n=2,228),
295 and People of the British Isles⁵⁵ (n=2,039) using PLINK v1.9^{56,57}. We filtered for samples and variants
296 with a missingness <5%, and variants with a minor allele frequency <1% and a p-value signifying
297 deviation from Hardy-Weinberg Equilibrium <1e⁻⁹. We performed PCA on a set of SNPs independent
298 in terms of linkage disequilibrium, using the PLINK --indep-pairwise command with a window of
299 1000, step size of 50 and linkage filtering threshold of 0.2. In the case of HC, we performed HLA
300 imputation from these SNP genotypes using SNP2HLA⁵⁸

301 MB Sequencing

302 Gut microbiome DNA samples were shipped to SeqBiome Ltd, Fermoy, County Cork, Ireland. Library
303 construction and shotgun sequencing of autoimmune cases and controls were carried out at
304 SeqBiome Ltd using Nextera library preparation⁵⁹ and Illumina NovaSeq sequencing (2 × 150 bp per
305 sample, Illumina, Hayward, California, USA).

306 Gut microbiome sequence data was then processed initially with *FastQC*⁶⁰, where data quality was
307 visually inspected. All samples were of generally good quality with minimal quality drop off at the
308 end of sequences. *Trimmomatic*⁶¹ was used for trimming and quality filtering using the following
309 parameters: SLIDINGWINDOW:5:22, MINLENGTH:75. The data was then passed to *Kneaddata*, a
310 pipeline incorporating *Trimmomatic*, *Bowtie 2*⁶², and other elements designed for contaminant
311 removal. Taxonomic assignment was then performed using *Kraken2* using a confidence of 0.1 and the

312 *Kraken2*^{63,64} GTDB bacterial database⁶⁵. *Kraken2* is a taxonomic classification system using exact k-
313 mer matches to achieve high accuracy and fast classification speeds, using a customised version of
314 the GTDB database, the results of which were analysed using the R based phyloseq package⁶⁶. GTDB
315 is a database which clusters available genomes based on Average Nucleotide Identity (ANI), and
316 assigns taxonomy based on the National Center for Biotechnology Information (NCBI) classifications.
317 If a sequence does not meet clustering requirements (95% ANI), but does not have a unique
318 taxonomic classification, a suffix will be added to indicate this (e.g., *Escherichia coli* A). This allows for
319 the most accurate taxonomic classifications, even in the case of organisms which have yet to be
320 formally identified.

321 Pathway and gene family assignment was then performed using *humann3*³⁶. The standard Uniref90
322 annotation of gene families were collapsed into Kyoto Encyclopedia of Gene and Genomes (KEGG)³⁵
323 orthology annotation and further collapsed into KEGG module and then pathway annotation.

324 Comparative Analysis

325 We used the phyloseq package⁶⁶ in R⁶⁷ v4.1.3 for comparative alpha diversity, beta diversity and
326 taxonomic analysis, using the package *ALDEx2*³⁷, to determine if any species were differentially
327 abundant (at group level) based on a Kruskal-Wallis test, using 512 Monte-Carlo instances drawn
328 from the Dirichlet distribution. P-values were initially corrected for Benjamini-Hochberg multiple
329 testing, and all were > 0.05. We therefore focussed discussion on the species with P-values significant
330 prior to multiple testing correction (< 0.05 prior to correction).

331 To account for the compositional structure of the Next Generation Sequencing (NGS) generated data
332 and to avoid the likelihood of generating spurious correlations, we first imputed the zeros in the
333 abundance matrices using the count zero multiplicative replacement method (cmultRepl,
334 method="CZM") implemented in the *Compositions* package and applied a centred log-ratio
335 transformation (CLR) using the *codaSeq.clr* function in the *CoDaSeq* package.

336 Linear discriminant analysis (LDA) effect size estimation was carried out using the *run_lefse()*
337 function *microbiomeMarker* package in Bioconductor on the species abundance matrix, selecting
338 CPM normalisation method, 100 bootstrap replicates, a Kruskal-Wallis and Wilcoxon p-value cutoff of
339 0.05 and LDA effect size cutoff of 2. To test HLA-DRB1*07:01 divergence from Hardy-Weinberg-
340 Equilibrium (HWE) expectations we utilised the R package *HardyWeinberg* and the function
341 *HWExact()* to calculate an exact test for the HWE. To test correlations between HLA dosage and F/B
342 ratio we used the R function *cor.test()*.

343 Alpha and Beta diversity analysis was carried out using the R packages *vegan* and *phyloseq* and
344 statistical comparison was tested using Kruskal-Wallis tests. Principle Coordinate Analysis (PCA) was
345 carried out using the *PCA* function in R using Aitchison distance matrix (CLR + Euclidean distances). To
346 test the significance of groupings in PCA space, Permutational Multivariate Analysis of Variance
347 (PERMANOVA) test was then performed using 999 permutations with significance cut-off value of
348 0.05. Dispersion homogeneity was tested using the *betadisper* function from *vegan*.

349 Functional Comparison

350 KEGG pathways and modules were assessed for differential abundance using *Aldex2* based on a
351 Kruskal-Wallis test using 512 Monte-Carlo instances drawn from the Dirichlet distribution. P-values
352 were initially corrected for multiple testing, and all were > 0.05. We therefore focussed discussion on
353 the pathways and modules with P-values significant prior to multiple testing correction (< 0.05 prior
354 to correction). KEGG pathways and KEGG modules were analysed by computing PCAs using Aitchison
355 distances. To test the significance of groupings, PERMANOVA tests were performed on these

356 distances, using 999 permutations, with a p-value cut-off of 0.05. Dispersion homogeneity was tested
357 using the *betadisper* function from *vegan*.

358 To assess the presence of potential bacterial homologs of the human *LG11* protein, a *blastp* search of
359 the *LG11* protein was run using a database of 64,695 bacterial genomes acquired from the *RefSeq*
360 database from which proteomes were generated using *Prokka*⁶⁸. Potential homologs were then
361 filtered from these results using cutoffs of >40% identity, >10% coverage and an e-value <1e-10.
362 These potential homologs were then mapped back to the metagenomic data using *blastx* with
363 cutoffs of >80% identity, >80% coverage and an e-value of <1e-6.

364 Acknowledgements

365 We would like to thank all the participants of this study, without whom the work could not have
366 been completed. We would like to thank the help of Bullers and Mrs Rachel Feeney from the OCMS
367 for their help with kit assembly and sample extractions.

368 Open access

369 For the purpose of Open Access, the author has applied a CC BY public copyright licence to any
370 Author Accepted Manuscript (AAM) version arising from this submission.

371 Funding

372 This work was funded by a grant from the Medical Research Charities Group, the Health Research
373 Board, Ireland, and Epilepsy Ireland (grant code: MRCG-2018-005), and Science Foundation Ireland
374 (SFI) under Grant Number 16/RC/3948 and co-funded under the European Regional Development
375 Fund and by FutureNeuro industry partners. A senior clinical fellowship (to SFI) from the Medical
376 Research Council [MR/V007173/1], Wellcome Trust Fellowship [104079/Z/14/Z], the National
377 Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) (The views expressed
378 are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of
379 Health)

380 Competing Interests

381 SRI has received honoraria/research support from UCB, Immunovant, MedImmune, Roche, Janssen,
382 Cerebral therapeutics, ADC therapeutics, Brain, CSL Behring, and ONO Pharma, and receives licensed
383 royalties on patent application WO/2010/046716 entitled 'Neurological Autoimmune Disorders'. And
384 has filed two other patents entitled "Diagnostic method and therapy" (WO2019211633 and US-
385 2021-0071249-A1; PCT application WO202189788A1) and "Biomarkers" (PCT/GB2022/050614 and
386 WO202189788A1).

387 Data Availability

388 **The raw metagenomic sequence data will be made accessible on the European Genome-Phenome**
389 **Archive (EGA) upon publication and will be referred to here.**

390

391 References

- 392 1. Scheffer, I.E. *et al.* ILAE classification of the epilepsies: Position paper of the ILAE Commission
393 for Classification and Terminology. *Epilepsia* **58**, 512-521 (2017).
- 394 2. Graus, F. *et al.* A clinical approach to diagnosis of autoimmune encephalitis. *Lancet Neurol*
395 **15**, 391-404 (2016).
- 396 3. Steriade, C. *et al.* Acute symptomatic seizures secondary to autoimmune encephalitis and
397 autoimmune-associated epilepsy: Conceptual definitions. *Epilepsia* **61**, 1341-1351 (2020).
- 398 4. Brenner, T. *et al.* Prevalence of neurologic autoantibodies in cohorts of patients with new and
399 established epilepsy. *Epilepsia* **54**, 1028-35 (2013).
- 400 5. Thompson, J. *et al.* The importance of early immunotherapy in patients with faciobrachial
401 dystonic seizures. *Brain* (2017).
- 402 6. Irani, S.R. *et al.* Faciobrachial dystonic seizures precede Lgi1 antibody limbic encephalitis. *Ann*
403 *Neurol* **69**, 892-900 (2011).
- 404 7. Irani, S.R. *et al.* Faciobrachial dystonic seizures: the influence of immunotherapy on seizure
405 control and prevention of cognitive impairment in a broadening phenotype. *Brain* **136**, 3151-
406 62 (2013).
- 407 8. Aurangzeb, S. *et al.* LGI1-antibody encephalitis is characterised by frequent, multifocal
408 clinical and subclinical seizures. *Seizure* **50**, 14-17 (2017).
- 409 9. Finke, C. *et al.* Evaluation of Cognitive Deficits and Structural Hippocampal Damage in
410 Encephalitis With Leucine-Rich, Glioma-Inactivated 1 Antibodies. *JAMA Neurol* **74**, 50-59
411 (2017).
- 412 10. Gadoth, A. *et al.* Expanded phenotypes and outcomes among 256 LGI1/CASPR2-IgG-positive
413 patients. *Ann Neurol* **82**, 79-92 (2017).
- 414 11. Bien, C.G. *et al.* Anti-contactin-associated protein-2 encephalitis: relevance of antibody titres,
415 presentation and outcome. *Eur J Neurol* **24**, 175-186 (2017).
- 416 12. Irani, S.R. *et al.* Antibodies to Kv1 potassium channel-complex proteins leucine-rich, glioma
417 inactivated 1 protein and contactin-associated protein-2 in limbic encephalitis, Morvan's
418 syndrome and acquired neuromyotonia. *Brain* **133**, 2734-48 (2010).
- 419 13. Shin, Y.W. *et al.* VGKC-complex/LGI1-antibody encephalitis: clinical manifestations and
420 response to immunotherapy. *J Neuroimmunol* **265**, 75-81 (2013).
- 421 14. Flanagan, E.P. *et al.* Basal ganglia T1 hyperintensity in LGI1-autoantibody faciobrachial
422 dystonic seizures. *Neurol Neuroimmunol Neuroinflamm* **2**, e161 (2015).
- 423 15. Binks, S.N.M., Klein, C.J., Waters, P., Pittock, S.J. & Irani, S.R. LGI1, CASPR2 and related
424 antibodies: a molecular evolution of the phenotypes. *J Neurol Neurosurg Psychiatry* (2017).
- 425 16. Binks, S.N.M. *et al.* Fatigue predicts quality of life after leucine-rich glioma-inactivated 1-
426 antibody encephalitis. *Ann Clin Transl Neurol* (2024).
- 427 17. Smith, K.M., Dubey, D., Liebo, G.B., Flanagan, E.P. & Britton, J.W. Clinical Course and Features
428 of Seizures Associated With LGI1-Antibody Encephalitis. *Neurology* **97**, e1141-e1149 (2021).
- 429 18. van Sonderen, A. *et al.* Anti-LGI1 encephalitis is strongly associated with HLA-DR7 and HLA-
430 DRB4. *Ann Neurol* **81**, 193-198 (2017).
- 431 19. Kim, T.J. *et al.* Anti-LGI1 encephalitis is associated with unique HLA subtypes. *Ann Neurol* **81**,
432 183-192 (2017).
- 433 20. Muniz-Castrillo, S. *et al.* Anti-CASPR2 clinical phenotypes correlate with HLA and
434 immunological features. *J Neurol Neurosurg Psychiatry* **91**, 1076-1084 (2020).
- 435 21. Binks, S. *et al.* Distinct HLA associations of LGI1 and CASPR2-antibody diseases. *Brain* **141**,
436 2263-2271 (2018).
- 437 22. Creary, L.E. *et al.* High-resolution HLA allele and haplotype frequencies in several unrelated
438 populations determined by next generation sequencing: 17th International HLA and
439 Immunogenetics Workshop joint report. *Hum Immunol* **82**, 505-522 (2021).
- 440 23. Tremlett, H., Bauer, K.C., Appel-Cresswell, S., Finlay, B.B. & Waubant, E. The gut microbiome
441 in human neurological disease: A review. *Ann Neurol* **81**, 369-382 (2017).

- 442 24. Sharon, G., Sampson, T.R., Geschwind, D.H. & Mazmanian, S.K. The Central Nervous System
443 and the Gut Microbiome. *Cell* **167**, 915-932 (2016).
- 444 25. McKay, K.A. *et al.* From bugs to brains: The microbiome in neurological health. *Mult Scler*
445 *Relat Disord* **12**, 1-3 (2017).
- 446 26. van der Vossen, E.W.J., de Goffau, M.C., Levin, E. & Nieuwdorp, M. Recent insights into the
447 role of microbiome in the pathogenesis of obesity. *Therap Adv Gastroenterol* **15**,
448 17562848221115320 (2022).
- 449 27. Yao, S., Zhao, Z., Wang, W. & Liu, X. Bifidobacterium Longum: Protection against
450 Inflammatory Bowel Disease. *J Immunol Res* **2021**, 8030297 (2021).
- 451 28. Silva, Y.P., Bernardi, A. & Frozza, R.L. The Role of Short-Chain Fatty Acids From Gut Microbiota
452 in Gut-Brain Communication. *Front Endocrinol (Lausanne)* **11**, 25 (2020).
- 453 29. Piquer-Esteban, S., Ruiz-Ruiz, S., Arnau, V., Diaz, W. & Moya, A. Exploring the universal
454 healthy human gut microbiota around the World. *Comput Struct Biotechnol J* **20**, 421-433
455 (2022).
- 456 30. Oren, A., Arahal, D.R., Rossello-Mora, R., Sutcliffe, I.C. & Moore, E.R.B. Emendation of Rules
457 5b, 8, 15 and 22 of the International Code of Nomenclature of Prokaryotes to include the
458 rank of phylum. *Int J Syst Evol Microbiol* **71**(2021).
- 459 31. Oren, A. & Garrity, G.M. Valid publication of the names of forty-two phyla of prokaryotes. *Int*
460 *J Syst Evol Microbiol* **71**(2021).
- 461 32. Magne, F. *et al.* The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in
462 Obese Patients? *Nutrients* **12**(2020).
- 463 33. Sorboni, S.G., Moghaddam, H.S., Jafarzadeh-Esfehani, R. & Soleimanpour, S. A
464 Comprehensive Review on the Role of the Gut Microbiome in Human Neurological Disorders.
465 *Clin Microbiol Rev* **35**, e0033820 (2022).
- 466 34. Stojanov, S., Berlec, A. & Strukelj, B. The Influence of Probiotics on the
467 Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease.
468 *Microorganisms* **8**(2020).
- 469 35. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on
470 genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361 (2017).
- 471 36. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse
472 microbial communities with bioBakery 3. *Elife* **10**(2021).
- 473 37. Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G. & Gloor, G.B. ANOVA-like differential
474 expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One* **8**, e67019 (2013).
- 475 38. Ding, M., Lang, Y., Shu, H., Shao, J. & Cui, L. Microbiota-Gut-Brain Axis and Epilepsy: A Review
476 on Mechanisms and Potential Therapeutics. *Front Immunol* **12**, 742449 (2021).
- 477 39. Kim, S., Park, S., Choi, T.G. & Kim, S.S. Role of Short Chain Fatty Acids in Epilepsy and
478 Potential Benefits of Probiotics and Prebiotics: Targeting "Health" of Epileptic Patients.
479 *Nutrients* **14**(2022).
- 480 40. Neis, E.P., Dejong, C.H. & Rensen, S.S. The role of microbial amino acid metabolism in host
481 metabolism. *Nutrients* **7**, 2930-46 (2015).
- 482 41. Peterson, C.T. *et al.* Short-Chain Fatty Acids Modulate Healthy Gut Microbiota Composition
483 and Functional Potential. *Curr Microbiol* **79**, 128 (2022).
- 484 42. Mills, S., Yang, B., Smith, G.J., Stanton, C. & Ross, R.P. Efficacy of Bifidobacterium longum
485 alone or in multi-strain probiotic formulations during early life and beyond. *Gut Microbes* **15**,
486 2186098 (2023).
- 487 43. Ilhan, Z.E., Brochard, V., Lapaque, N., Auvin, S. & Lepage, P. Exposure to anti-seizure
488 medications impact growth of gut bacterial species and subsequent host response. *Neurobiol*
489 *Dis* **167**, 105664 (2022).
- 490 44. Zhou, C. *et al.* Changes and significance of gut microbiota in children with focal epilepsy
491 before and after treatment. *Front Cell Infect Microbiol* **12**, 965471 (2022).

- 492 45. De Caro, C. *et al.* Intestinal inflammation increases convulsant activity and reduces
493 antiepileptic drug efficacy in a mouse model of epilepsy. *Sci Rep* **9**, 13983 (2019).
- 494 46. Lee, K., Kim, N., Shim, J.O. & Kim, G.H. Gut Bacterial Dysbiosis in Children with Intractable
495 Epilepsy. *J Clin Med* **10**(2020).
- 496 47. Chatzikonstantinou, S. *et al.* The gut microbiome in drug-resistant epilepsy. *Epilepsia Open* **6**,
497 28-37 (2021).
- 498 48. Ma, X. *et al.* Clinical Features and Gut Microbial Alterations in Anti-leucine-rich Glioma-
499 Inactivated 1 Encephalitis-A Pilot Study. *Front Neurol* **11**, 585977 (2020).
- 500 49. Mohiti A, A.M., Amooei M,. The Effect of Systemic Corticosteroid Use on the pH and Viscosity
501 of Saliva. *Shiraz E-Medical Journal* **22**, e101710 (2020).
- 502 50. Lindefeldt, M. *et al.* The ketogenic diet influences taxonomic and functional composition of
503 the gut microbiota in children with severe epilepsy. *NPJ Biofilms Microbiomes* **5**, 5 (2019).
- 504 51. Attaye, I., van Oppenraaij, S., Warmbrunn, M.V. & Nieuwdorp, M. The Role of the Gut
505 Microbiota on the Beneficial Effects of Ketogenic Diets. *Nutrients* **14**(2021).
- 506 52. Ghosh, T.S., Shanahan, F. & O'Toole, P.W. The gut microbiome as a modulator of healthy
507 ageing. *Nat Rev Gastroenterol Hepatol* **19**, 565-584 (2022).
- 508 53. Gilbert, E. *et al.* The Irish DNA Atlas: Revealing Fine-Scale Population Structure and History
509 within Ireland. *Sci Rep* **7**, 17199 (2017).
- 510 54. Desch, K.C. *et al.* Linkage analysis identifies a locus for plasma von Willebrand factor
511 undetected by genome-wide association. *Proc Natl Acad Sci U S A* **110**, 588-93 (2013).
- 512 55. Winney, B. *et al.* People of the British Isles: preliminary analysis of genotypes and surnames
513 in a UK-control population. *Eur J Hum Genet* **20**, 203-10 (2012).
- 514 56. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
515 datasets. *Gigascience* **4**, 7 (2015).
- 516 57. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage
517 analyses. *Am J Hum Genet* **81**, 559-75 (2007).
- 518 58. Jia, X. *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLOS ONE* **8**,
519 e64683 (2013).
- 520 59. Christie, W., Yadin, R., Ip, K. & George, K.W. Highly Multiplexed, Semiautomated Nextera
521 Next-Generation Sequencing (NGS) Library Preparation. *Methods Mol Biol* **2205**, 91-104
522 (2020).
- 523 60. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. (2010).
- 524 61. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
525 data. *Bioinformatics* **30**, 2114-20 (2014).
- 526 62. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
527 357-9 (2012).
- 528 63. Wood, D.E. & Salzberg, S.L. Kraken: ultrafast metagenomic sequence classification using
529 exact alignments. *Genome Biol* **15**, R46 (2014).
- 530 64. Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome*
531 *Biol* **20**, 257 (2019).
- 532 65. Parks, D.H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a
533 phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic*
534 *Acids Res* **50**, D785-D794 (2022).
- 535 66. McMurdie, P.J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and
536 graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
- 537 67. Team., R.C. R: A language and environment for statistical computing. *R Foundation for*
538 *Statistical Computing*. (2017).
- 539 68. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-9 (2014).

540

541 **Figures and Tables**

542 **Table 1**

	Cases			Healthy controls
	LGI1	CASPR2	LGI1/CASPR2	
No. of patients*	43	3	1	27
Control type:	NA	NA	NA	
Spouse/partner				15/27 (56%)
Sibling				1/27 (4%)
Child				1/27 (4%)
Other				3/27 (11%)
Unknown				6/27 (22%)
Mean (median) age at sampling**	66 (66)	69 (69)	60 (NA)	NA
Sex, female, %*	8/43 (19%)	3 (100%)	1 (100%)	NA
Clinical syndrome, %:				NA
Epilepsy	2/43	0	0	
Limbic encephalitis	35/43	3 (100%)	0	
Morvan's syndrome	0/43	0	1 (100%)	
Neuromyotonia/pain	4/43	1 (33%)	0	
NA	2/43	0	0	
Medications at time of sampling, %:				NA
Anti-seizure:				
Carbamazepine	1/43 (2%)	1 (33%)	0	
Clobazam/Clonazepam/Lorazepam	2/43 (5%)	0	0	
Gabapentin/Pregabalin	3/43 (7%)	2/3 (66%)	1 (100%)	
Lacosamide	1/43 (2%)	0	0	
Lamotrigine	5/43 (12%)	0	0	
Levetiracetam	15/43 (35%)	2/3 (66%)	0	
Phenytoin	1/43 (2%)	0	1 (100%)	
Valproate	6/43 (14%)	0	0	
None	10/43 (23%)	0	0	
NA	3/43 (7%)	0	0	
Immunotherapy:				
Azathioprine	1/43 (2%)	0	0	
Corticosteroids	23/43 (53%)	3/3 (100%)	1 (100%)	
Mycophenolate Mofetil	2/43 (5%)	0	0	
None	15/43 (35%)	0	0	
NA	3/43 (7%)	0	0	
HLA status, %:				
DRB1*07:01 carrier	39/43 (91%)	1/3 (33%)	0	11/27 (41%)
DRB1*07:01 homozygote	4/43 (10%)	0	0	1/27 (4%)
DRB1*07:01 negative	3/43 (7%)	2/3 (66%)	1 (100%)	4/27 (15%)
Unknown	0	0	0	11/27 (41%)* **
*No antibody status or sex for one patient **No age at sampling for three patients *** HLA imputation was used for HC and some imputed HLA genotypes did not pass quality control for sufficient carrier status to be called.				

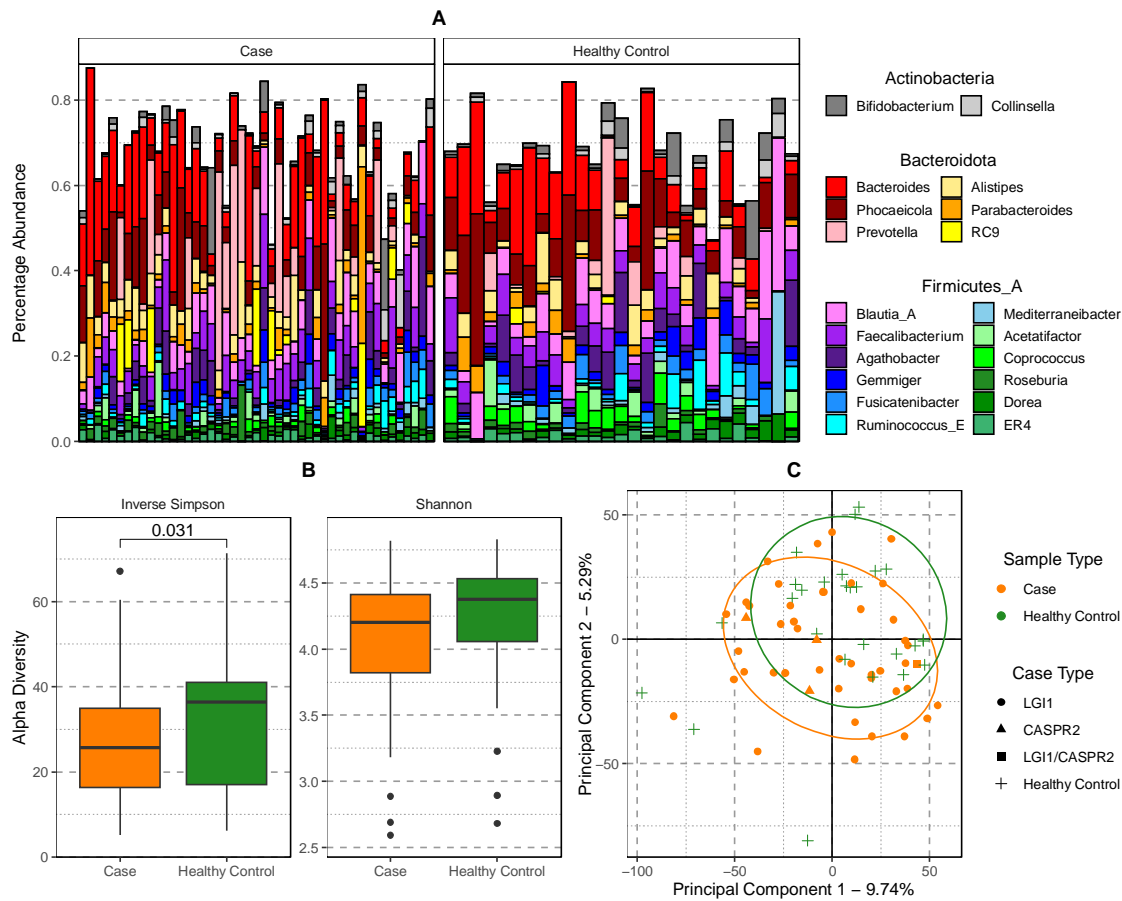
543

544

545 **Study Cohort Characteristics** – Shown are the cohort descriptors for 47 LGI1-CASPR2-Ab-E cases and
546 27 matched health controls. Cohort is characterised by; genetic sex (Sex), auto-antibody status if a
547 case (Case Type), whether an individual is inferred to be a HLA DRB*07:01 carrier through imputation
548 (HLA DRB Status), and what relationship the health control has with matched case (HC Type).

549

550 Figure 1



551
552

553 **Comparison of the gut microbiome in *LGI1-CASPR2-Ab-E* cases and healthy controls –** (A) Bar chart
 554 of the 20 most abundant Genera, grouped by Phylum and separating cases (left) and healthy control
 555 (HC) samples (right) in separate panels. Each vertical bar represents the compositional microbiome
 556 profile of an individual. Note: Firmicutes_A and similar names are placeholder taxa which have been
 557 classified as such in the NCBI database but does not meet the clustering requirements for GTDB.
 558 Unused space represents other species not in the top 20 most abundant. (B) The distribution of
 559 alpha diversity values between cases (orange) and HC (green) using two measures, the Inverse
 560 Simpson and Shannon to retain comparative direction of effect between the two. Pairwise p-values
 561 shown are Wilcoxon test p-values significant after Holm correction for multiple testing (Shannon not
 562 significant). (C) Principal Component Analysis of beta-diversity estimates between cases and HC.
 563 Ellipses shown are 80% confidence intervals assuming a multivariate t-distribution.
 564

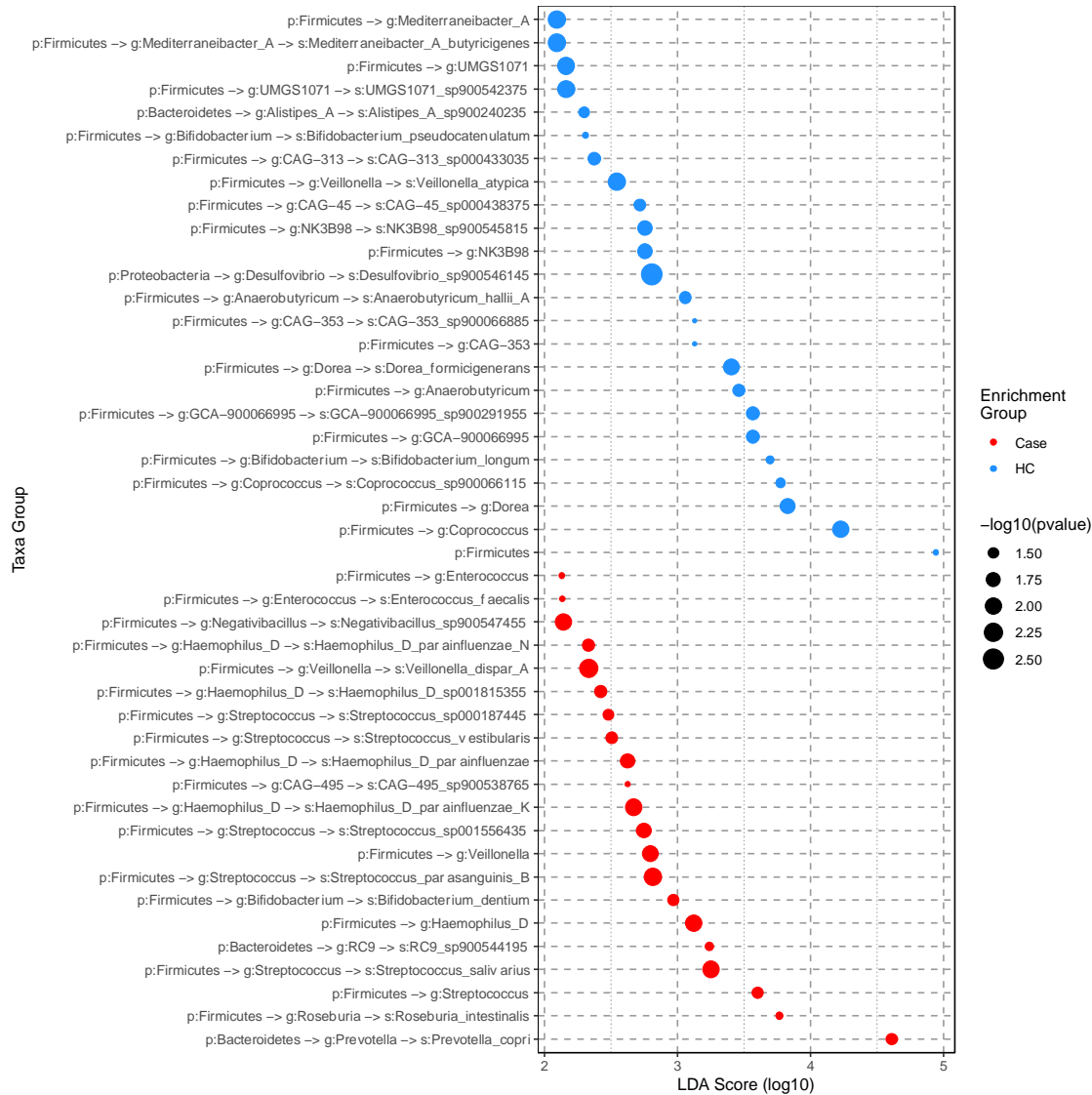
565 Table 2

GTDB Species	NCBI Species	Effect	Uncorrected P-value
Streptococcus parasanguinis_B	Streptococcus parasanguinis	-0.347	0.007
Streptococcus salivarius	Streptococcus salivarius	-0.302	0.013
Veillonella atypica	Veillonella atypica	-0.43	0.015
NK3B98 sp900545815	uncultured Eubacteriales bacterium	0.307	0.015
Haemophilus_D parainfluenzae	Haemophilus parainfluenzae	-0.302	0.017
GCA-900066995 sp900291955	Pseudoruminococcus massiliensis	0.29	0.019
Haemophilus_D parainfluenzae_K	Haemophilus parainfluenzae	-0.403	0.021
UMGS1071 sp900542375	uncultured Oscillospiraceae bacterium	0.354	0.024
Faecalibacterium prausnitzii_A	Faecalibacterium prausnitzii	-0.234	0.025
Streptococcus sp001556435	Streptococcus salivarius	-0.334	0.027
Veillonella dispar_A	Veillonella dispar	-0.392	0.028
Alistipes_A sp900240235	uncultured Alistipes sp.	0.397	0.032
Coprococcus sp900066115	Coprococcus sp.	0.281	0.041
COE1 sp001916965	Roseburia sp.	-0.28	0.042
Streptococcus vestibularis	Streptococcus vestibularis	-0.358	0.048
Bifidobacterium dentium	Bifidobacterium dentium	-0.366	0.048

566
567
568
569
570
571
572
573

Differential gut microbiome species between *LG11-CASPR2*-Ab-E cases and healthy controls –
Shown are the 16 differentiated species which have significant, uncorrected for multiple testing, p-values using the Kruskal-Wallis test using 512 Monte-Carlo instances drawn from the Dirichlet distribution. A positive value indicating an enrichment in HC samples.

574 **Figure 2**



575

576 **Differential enrichment of microbiome taxa in LGI1-CASPR2-Ab-E cases and healthy controls - Taxa**
 577 significantly (adjusted p-value < 0.05) enriched in either healthy controls (HC) or cases. Points are
 578 colour coded according to the sample type the taxon is enriched in, and the $\log_{10}(\text{LDA})$ score (effect
 579 size) is shown along the x-axis. Labels along the y-axis are in the format of "p:" to indicate phylum,
 580 "g:" to indicate genus, and "s:" to indicate species with "-" joining parent and child taxa.
 581

582 Table 3

Table 3			
Differential KEGG Components			
	Name	Effect	Uncorrected P-value
KEGG Module	Coenzyme A biosynthesis, pantothenate => CoA	0.468	0.0033
	Ornithine biosynthesis, glutamate => ornithine	0.369	0.023
	Histidine biosynthesis, PRPP => histidine	0.314	0.037
KEGG Pathway	Pentose and glucuronate interconversions	-0.472	0.0031
	Ribosome biogenesis in eukaryotes	0.358	0.0082
	Parkinson disease	0.369	0.021
	NOD-like receptor signaling pathway	0.346	0.022
	Necroptosis	0.325	0.039

583

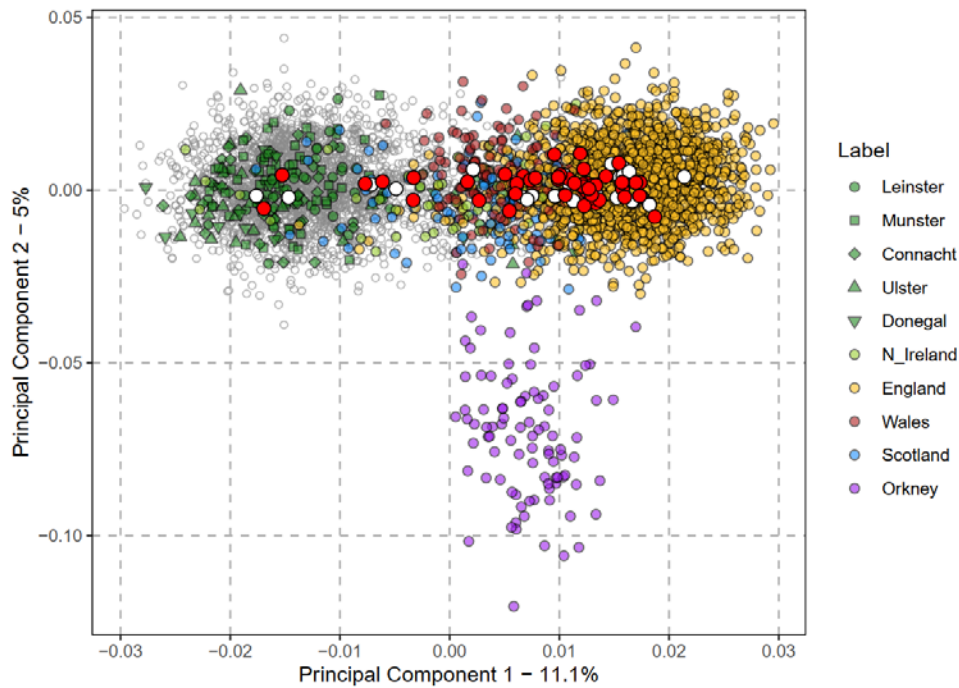
584

585 **Differential KEGG modules between *LG11-CASPR2-Ab-E* cases and healthy controls** - Shown are the
586 three differentiated KEGG Modules, and the five KEGG Pathways which have significant, uncorrected
587 for multiple testing, p-values using the Kruskal-Wallis test using 512 Monte-Carlo instances drawn
588 from the Dirichlet distribution. A positive value indicating an enrichment in healthy control samples.

589

590 **Supplementary Materials**

591

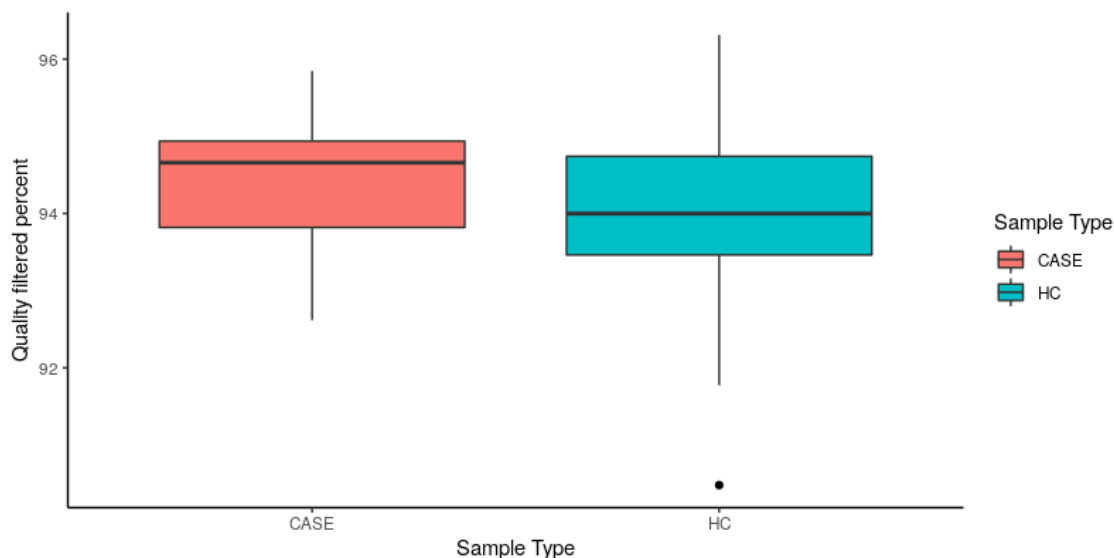


592

593

594 **Supplementary Figure 1 – Ancestry PCA of *LG11-CASPR2*-Ab-E Cases and Controls.** The principal
595 component decomposition of a genetic relationship matrix of 4,460 Irish and British ancestry
596 controls and 47 cases and 27 health controls (HC). Points are colour and shape coded according to
597 genetic ancestry group. Filled circles with red indicate the genetic position of a case and filled circles
598 with white indicate the genetic position of a HC.

599

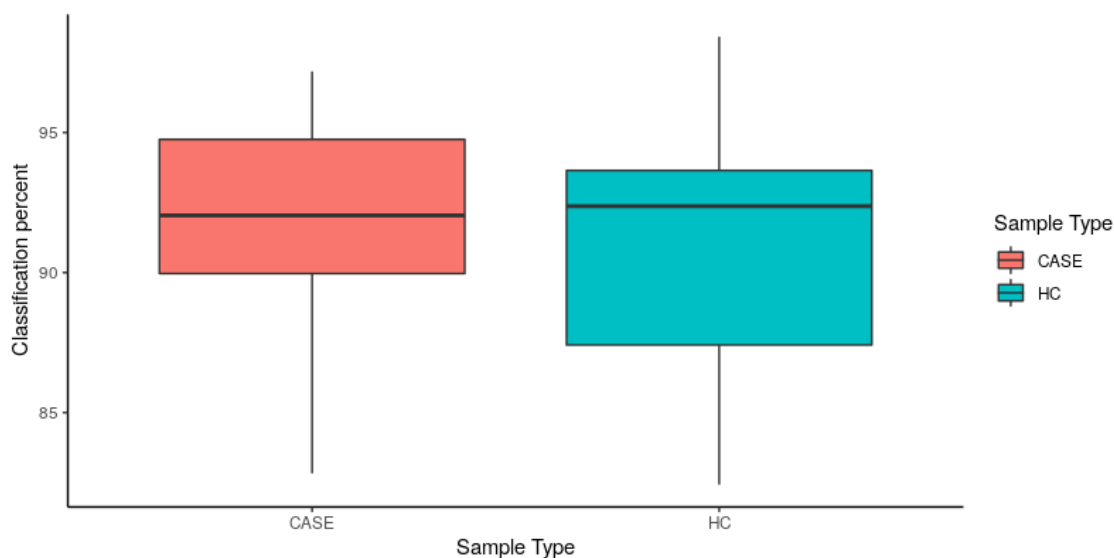


600

601 **Supplemental Figure 2 -Filtering quality of *LG11-CASPR2*-Ab-E case and healthy control gut**
602 **microbiome samples.** Boxplot representation of percentage of reads which pass the quality filtering
603 step, coloured by sample type. Boxplots show the median value, with lower and upper hinges
604 showing the 1st and 3rd quartiles. Whiskers show the largest value no further than 1.5 x the
605 Interquartile Range (IQR) from that range. Data points beyond these whiskers are plotted separately
606 as black points.

607

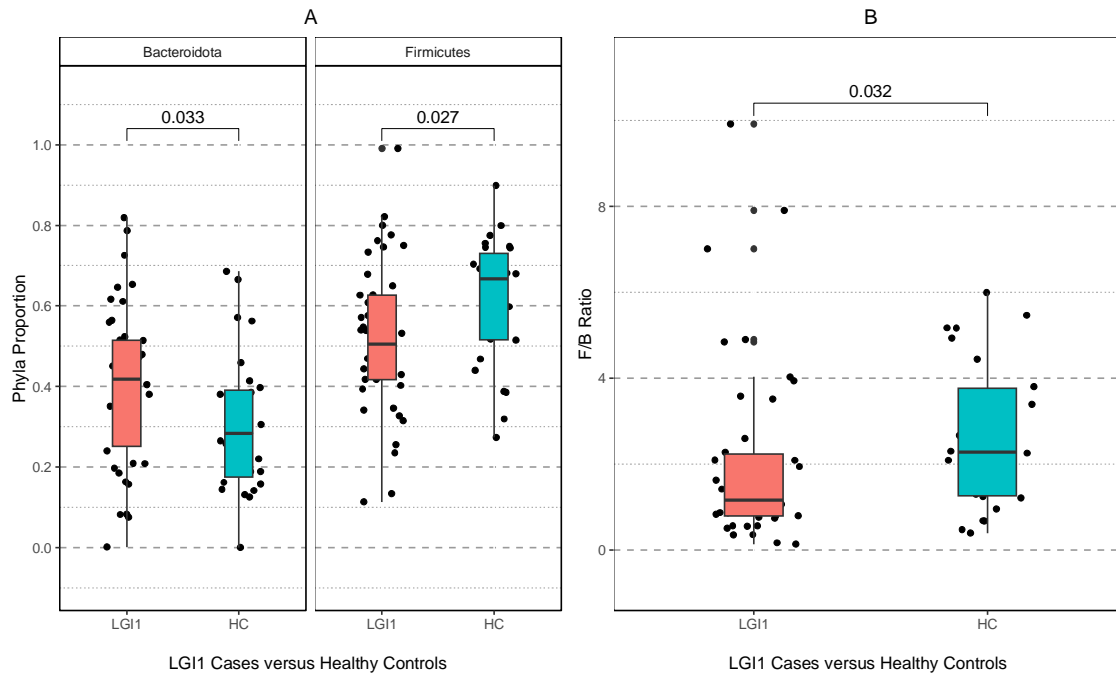
608



609

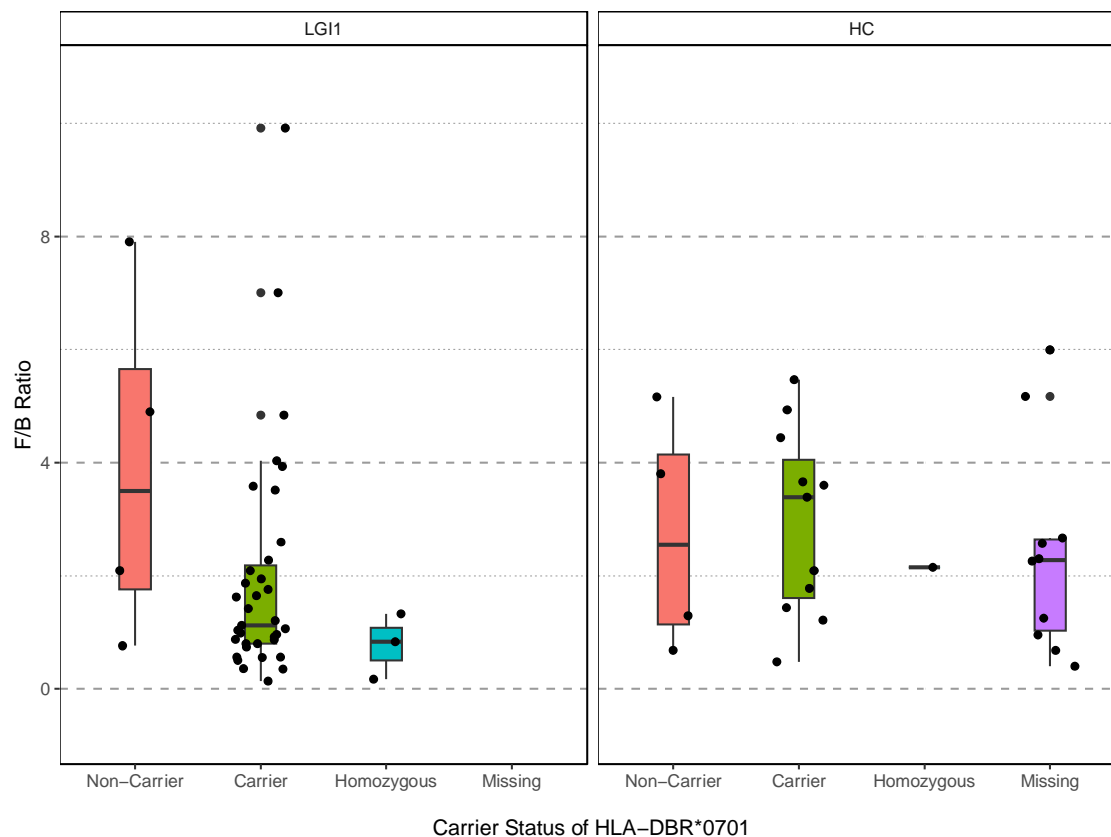
610 **Supplemental Figure 3: Kraken2 percent classification of *LG11-CASPR2*-Ab-E case and healthy**
611 **control reads.** Boxplots show the median value, with lower and upper hinges showing the 1st and
612 3rd quartiles. Whiskers show the largest value no further than 1.5 x the Interquartile Range (IQR)
613 from that range.

614



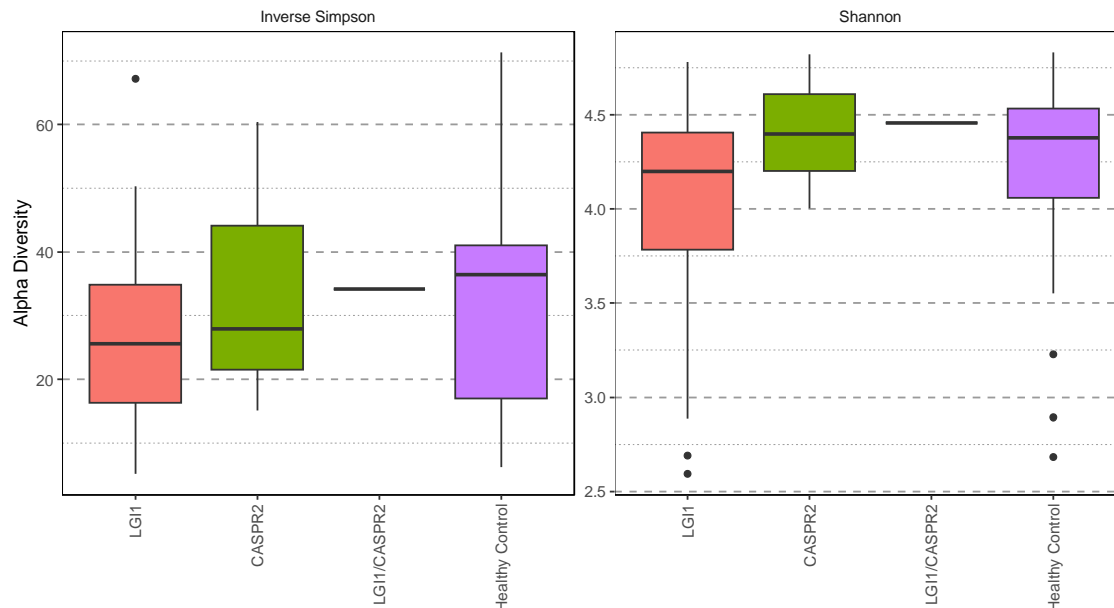
615

616 **Supplemental Figure 4: *Firmicutes/Bacteroides* ratio in *LGI1-CASPR2-Ab-E* cases and healthy**
617 **controls.** (A) The proportion of the two phyla *Firmicutes* and *Bacteroides* in cases and healthy
618 controls (HC). Comparison bars shown are the p value from Wilcox Signed Rank test. (B) The
619 *Firmicutes/Bacteroides* ratio between cases and HC. Individual data points are shown on top of
620 boxplots. Boxplots show the median value, with lower and upper hinges showing the 1st and 3rd
621 quartiles. Whiskers show the largest value no further than 1.5 x the Interquartile Range (IQR) from
622 that range.
623



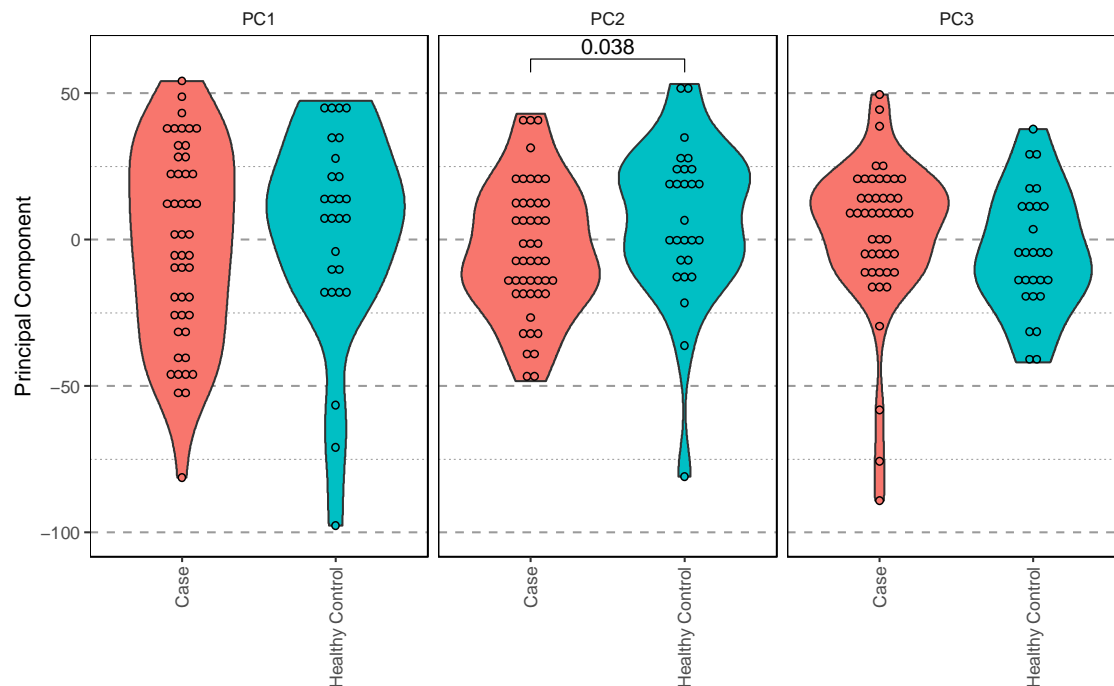
624

625 **Supplemental Figure 5: Firmicutes/Bacteroides ratio as a function of HLA-DRB1*0701 dosage.** The
626 ratio of the two phyla *Firmicutes* and *Bacteroides* in LGI1- Ab-E cases and healthy controls (HC)
627 grouped by HLA-DRB1*0701 dosage. As HC HLA-DRB1*0701 dosage was imputed from SNP genotypes
628 there is a proportion of missing data. Individual data points are shown on top of boxplots. Boxplots
629 show the median value, with lower and upper hinges showing the 1st and 3rd quartiles. Whiskers
630 show the largest value no further than 1.5 x the Interquartile Range (IQR) from that range.



631
632
633
634
635
636
637
638

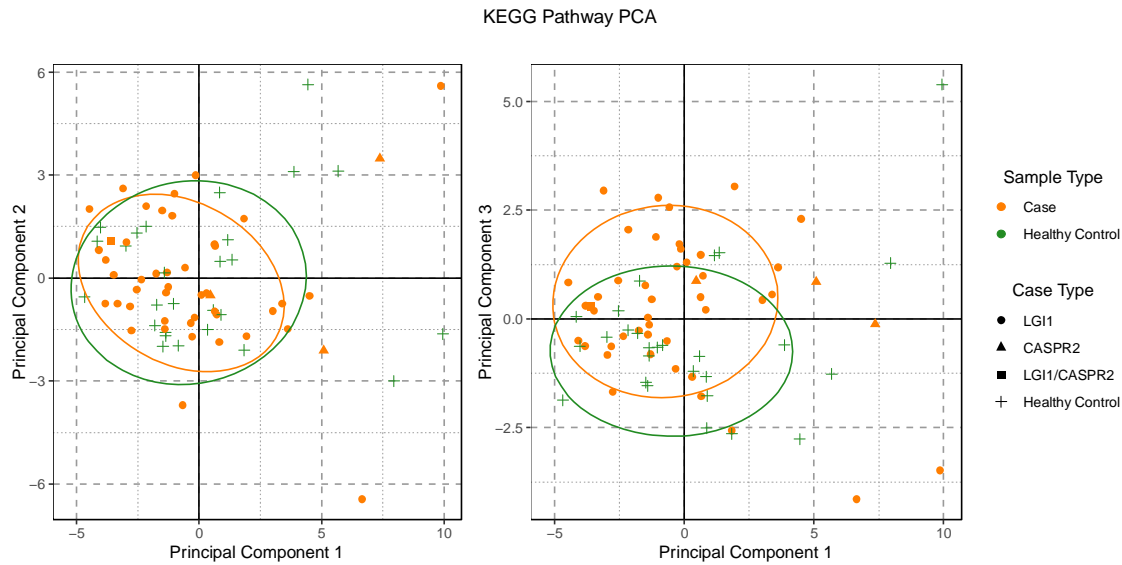
Supplemental Figure 6 - Measure of Alpha diversity stratified by case type. Shown are the Inverse Simpson and Shannon alpha diversity measurements. Distributions of alpha diversity values are shown in boxplots. Boxplots show the median value, with lower and upper hinges showing the 1st and 3rd quartiles. Whiskers show the largest value no further than 1.5 x the Interquartile Range (IQR) from that range. Data points beyond these whiskers are plotted separately as black points.



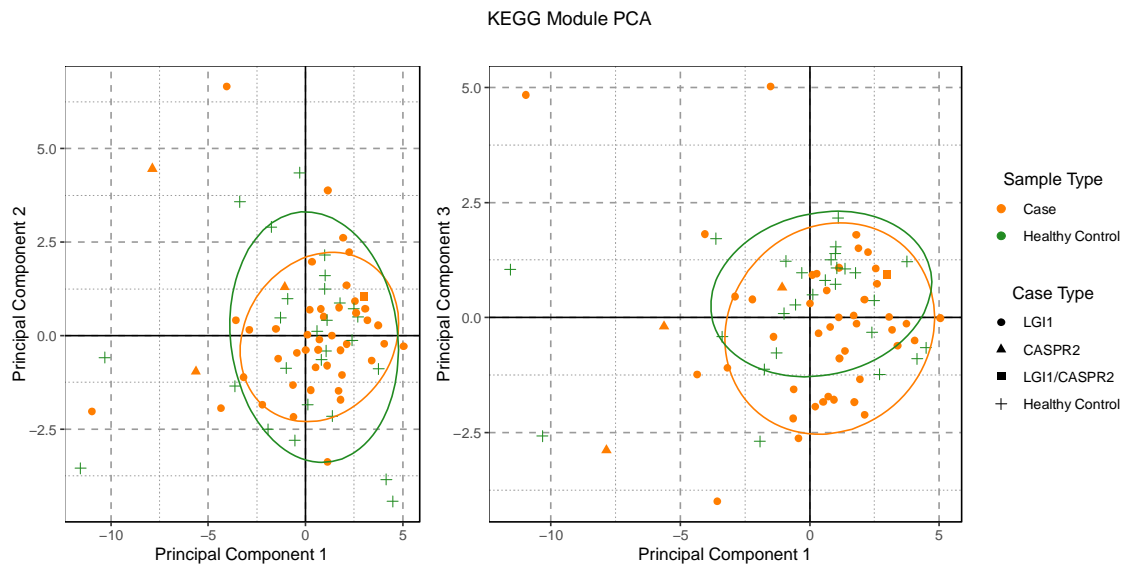
639
640
641
642
643
644

Supplemental Figure 7 – Distribution of LGI1-CASPR2-Ab-E cases and controls over beta-diversity principal components. Shown are the sample distributions across the first three Aitchison principal components of beta diversity, with violin plots showing overall distribution. Pairwise comparison values are significant p-values from Wilcox tests of significant differences. All comparison pairs were tested, and p-values adjusted with the Holm method.

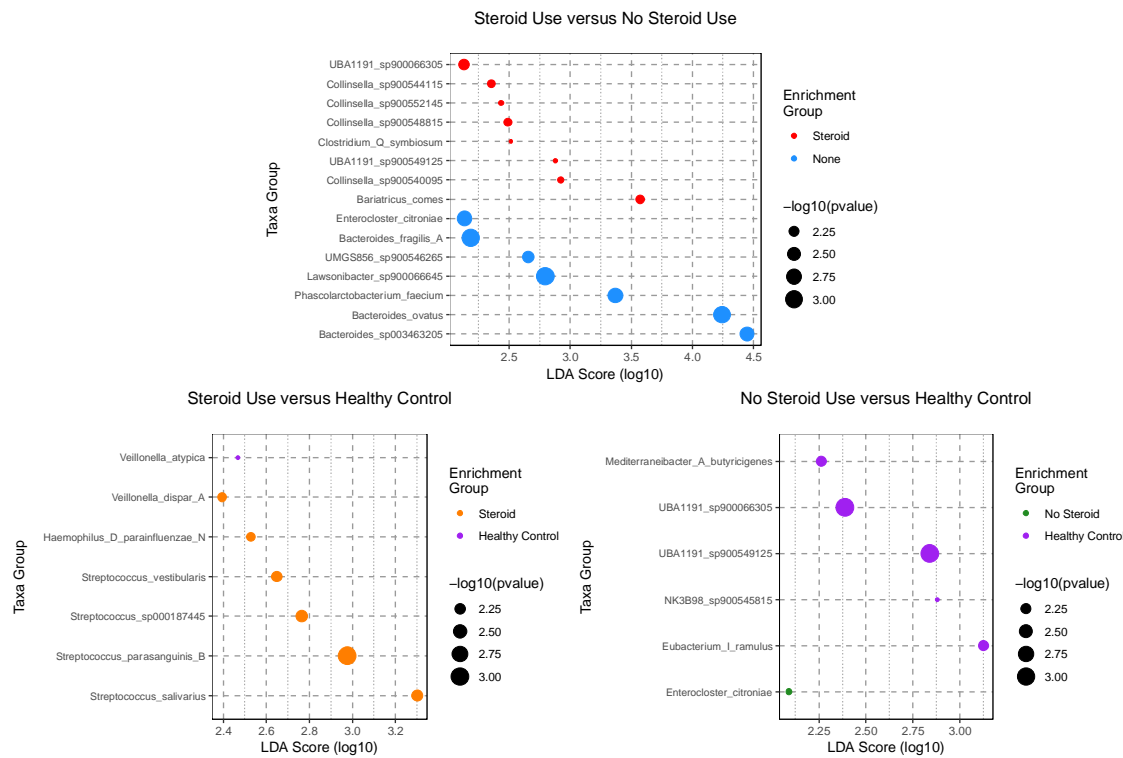
645
646



647
648 **Supplemental Figure 8 – Aitchison PCA of KEGG Pathway abundance.** Sample case and control type
649 are shown by colour and shape. Ellipses shown are 80% confidence intervals assuming a multivariate
650 t-distribution.
651



652
653 **Supplemental Figure 9 – Aitchison PCA of KEGG Module abundance.** Sample case and control type
654 are shown by colour and shape. Ellipses shown are 80% confidence intervals assuming a multivariate
655 t-distribution.
656



657

658

Supplemental Figure 10 – Differential enrichment of microbiome taxa in *LG11-CASPR2*-Ab-E cases stratified by steroid use history and healthy controls - Taxa significantly (adjusted p-value < 0.05) enriched in either healthy controls (HC) or cases. Points are colour coded according to the sample type the taxon is enriched in, and the log₁₀(LDA) score (effect size) is shown along the x-axis. Labels along the y-axis are in the format of “p:” to indicate phylum, “g:” to indicate genus, and “s:” to indicate species with “->” joining parent and child taxa.

664

665 **Supplemental Data 1 – Dietary and supplement questionnaire**

666 **Supplemental Data 2 – Species**

667 **Supplemental Data 3 - LDA Taxa results**

668 **Supplemental Data 4 – Dietary associations with beta-PCs**

669 **Supplemental Data 5 – All KEGG Module results**

670 **Supplemental Data 6 – All KEGG Pathway results**

671 **Supplemental Data 7 – Organism orthologues**