**科研 | Cell：人体肠道菌群的长期遗传稳定性和个人特异性**

**本文由小贤哥（胡世贤）编译**

**导读**

肠道菌群和宿主健康紧密相关，但是人体肠道菌群在长时间内是否稳定，菌群的遗传特征是否呈现个体化差异，以及肠道菌群的变化对宿主健康有什么作用？近日来自荷兰格罗宁根大学医学中心傅静远教授团队在Cell上发表了最新研究成果，为上述问题给出了答案。该研究通过对338人4年前后的51种表型，1183种血液代谢物，以及肠道菌群的大规模组学数据分析，发现肠道菌群存在宿主特异性，开发了高精度“人体微生物指纹”算法，实现用肠道微生物遗传背景准确区分不同宿主；并且进一步揭示了肠道微生物如何通过血液代谢来影响宿主心血管等疾病相关指标的机制。

关键词：肠道菌群；长期稳定性；代谢相关疾病；多组学

**论文ID**

原名：The long-term genetic stability and individual specificity of the human gut microbiome

译名：人体肠道菌群的长期遗传稳定性和个人特异性

期刊：Cell

IF：38.6

发表时间：2021/4/9 美东时间（北京时间为10号凌晨）

第一作者：Lianmin Chen (陈连民)

通讯作者：Jingyuan Fu (傅静远)

通讯作者单位：荷兰格罗宁根大学，大学医学中心，遗传系和儿科系

DOI号：https://doi.org/10.1016/j.cell.2021.03.024

**实验设计**

本研究主要纳入了荷兰LifeLines-DEEP大群体队列。在2013年，研究者对1135个个体的粪便样本进行了宏基因组测序 (Science,2016)，并对其中338人四年后再次进行了样本收集和宏基因组测序。研究者利用MetaPhlan2和HUMAnN2从宏基因组数据中得到了肠道菌群的菌种丰度和功能通路，用StrainPhlAn进行了菌群亚种（菌株）的分析，用SGVFinder分析了微生物的遗传结构。同时，研究者也收集了大量的表型数据，例如血细胞计数，血液生化指标，疾病情况和药物服用等。此外，对两个时间点的血液样本进行了流动注射时间飞行质谱检测以得到血液代谢物，并用人类代谢数据库（Human Metabolome Database）等进行了注释。以下为本研究最核心分析方法，1)用于利用微生物特征鉴定个体的高精度”微生物指纹”算法：将不同的微生物特征数据转化为不同类型的距离矩阵，如对微生物菌种丰度和功能通路丰度采用Bray-Curtis距离矩阵，对SNP单倍型采用Kimura距离矩阵，对SV采用Canberra距离矩阵。然后利用nearest neighbour聚类方法鉴定两个样本是否来自同一个体，从而计算准确鉴定率。进而通过接收者操作特征曲线（ROC）来检测方法的特异性和灵敏度；2）利用距离矩阵逐步结合选取最佳个体特异性的微生物特征：研究者采用双交叉验证，即随机选取总数据集的60%作为发现阶段，40%作为验证阶段。对所有距离矩阵标准化，然后采用1）中算法评估每个距离矩阵的个体鉴别准确度，然后逐步添加新的矩阵求取矩阵平均值，得到新的个体鉴别准确度，然后重复，直至计算所有矩阵的组合。整个交叉验证重复了10次，最优矩阵数量N取决于个体鉴别准确率不再提高；3）微生物特征和宿主表型及血液代谢物的关联分析：第一步采用一般混合线性模型，时间和个体作为随机效应，主变量作为固定效应，第二步采取差异分析法，即将两个时间点的变化差值放入一般线性模型中；4）中介分析：首先选取3）中与血液代谢和人体表型都相关的微生物特征，利用双向中介分析（mediate in R）探究因果关系。

**结果**

**1 LifeLines-DEEP长期随访人群**

研究者们在2013年和2017年两个时间点对来自荷兰LifeLines-DEEP群体的338个个体采集了51种表型数据，涵盖了人体测量学（如年龄，性别等），血液细胞计数，血液生化指标，疾病状态和药物服用。通过对比两个时间点的数据，研究者发现22个连续性性状中，19个有了显著变化。比如4年后人群的血液肌酐水平，血液收缩压和舒张压都显著提高（*P*Paired Wilcoxon=2.5x10-50; *P*Paired Wilcoxon=3.6x10-16; *P*Paired Wilcoxon=2.2x10-3）。一些人的生活方式，健康状态和药物服用也产生了变化，比如26个参与者4年后被诊断为肠易激综合症，6个参与者患有抑郁症。

**2 肠道菌群多样性和组成的变化**

为了研究人体肠道菌群的长期稳定性，研究者首先比较了两个时间点肠道微生物的组成和多样性。和4年前的基值相比，4年后人群的微生物多样性显著提高（菌种水平的shannon指数：*P*Paired Wilcoxon=2.4x10-7, 图 1A）。通过主成分分析（principal component analysis， PCA），菌群整体的种类和功能在4年后也发生了偏移（*P*PCoA1 Paired Wilcoxon=0.08; *P*PCoA2 Paired Wilcoxon=1.6x10-5）。同时也发现相同个体内的微生物整体的组成和功能差异比不同个体间的差异要小（*P*Wilcoxon=1x10-4，Figure 1B, C）。这个结果表明，相同个体，4年间肠道菌群的变化要小于人与人之间的差异。值得注意的是，个体内的差异变化是和该个体4年前基线的多样性显著相关的（rSpearman=-0.21, *P*=1.5x10-4，图1D），这与来自其它研究的假设“多样性越高的菌落越稳定”相一致。

在比较4年前后单个肠道菌群的菌种和生物通路时，59.9%的菌种和44.3%的生物通路的相对丰度都有发生了变化（FDR<0.05）,而且，同一菌属下的菌种变化程度高度一致，比如7中双歧杆菌（Bifidobacterium）属下的菌种相对丰度都有降低。年龄增长可能是导致该现象的部分原因，比如在之前的研究中就已经发现，双歧杆菌的菌种（B. adolescentis, B. bifidum， B. longum）的相对丰度会随年龄的增长而下降。

图1.肠道菌群的长期差异。A，基线（baseline，浅蓝）和跟进（follow-up，深蓝）的微生物alpha多样性的比较。Y轴表示菌种水平的香农指数。B,个体间两个时间点（基线，浅蓝；跟进，深蓝）和配对样本两个时间点（橙色）的Bray-Curtis距离（菌种组成）比较。C, 个体间两个时间点（基线，浅蓝；跟进，深蓝）和配对样本两个时间点（橙色）的Bray-Curtis距离（功能通路）比较。D,微生物组成的时间稳定性或取决于基线的多样性。X轴表示基线时期香农指数，Y轴表示配对个体内两个时间点的Bray-Curtis距离。图中包括了rank-based Wilcoxon检验以及Spearman关联分析的*P*值或相关系数。

**3 肠道菌群遗传稳定性“因种而定”**

微生物的遗传基因会随时间而变，比如发生遗传诱变和品系替换。肠道菌群的遗传特征可以影响相应的功能，进而可能影响宿主的表型改变。因此，研究者进一步比较了相同个体内4年前后肠道菌群的单核苷酸变异（single nucleotide polymorphism， SNP）和遗传结构变异（genomic structural variants， SVs）。

研究者发现，相同个体内的肠道菌群SNP和SV的差异，显著小于不同个体间肠道菌群的差异（Figure 2A, B）。富集显著SNP单倍型变异的菌种包括球菌Ruminococcus torques, 链球菌Streptococcus parasanguinis和费氏杆菌Faecalibacterium prausnitzii,然而双歧杆菌Bifidobacterium angulatum,甲烷短杆菌Methanobrevibacter smithii和另枝菌Alistipes putredinis则具有较低的遗传变异（P Wilcoxon<0.05，Figure 2A）。同时研究者在人类微生物计划（Human Microbiome Project, HMP）中的43个健康个体一年前后的数据观察到了一致的变化（Figure 2C），但是HMP中，富集变异的肠道微生物差异变化要低于本研究的LifeLines-DEEP队列，这可能是本研究的队列是4年的长期队列，而HMP仅有一年前后的对比数据（Figure 2D-F），因此，更长的时间跨度可能造成的更多的变异富集。

比较SNP单倍型和SV发现，这两者有显著的一致性，表明不同的遗传变异类型都可以表示部分菌种的相同遗传稳定性（Figure 2A, B）。有趣的是，遗传变异不稳定的菌种常常和人体疾病相关。比如，球菌Ruminococcus torques在克罗恩病人中有很高的相对丰度，链球菌Streptococcus parasanguinis在肠道感染的病人中的丰度也会升高。值得注意的是，研究者并没有发现肠道菌群的遗传变异和相对丰度变化相关，这表明两者相互独立。因此研究肠道菌群的遗传变异也是今后研究的新方向。

总结来说，在4年的时间跨度下，相同个体的肠道菌群差异要小于不同个体间的肠道菌群差异；不同的菌种的遗传稳定性不同。基于以上，研究者提出以下假设：肠道菌群的特异性可以用来区分宿主；而菌种的变化也可能和宿主表型变化相关。

图2.肠道菌群的SNP单倍型和结构变异的长期稳定性。A，配对个体之内（橙色）和不同个体之间（绿色）在主要菌种亚种（菌株）的SNP单倍型差异。每个菌种边的数字表示了配对样本数。B, 配对个体之内（橙色）和不同个体之间（绿色）在主要菌种亚种（菌株）的结构缺失（dSV）和结构变异（vSVs）差异。每个菌种边的数字表示了配对样本数。C,LLD和HMP两个队列之间，相同个体内的菌种SNP单倍型比较。每个点表示一种菌种。深色点表示菌种随时间的变化在LLD和HMP之间有显著差异（FDR<0.05）,浅灰色则表示两个数据中变化一致。D-F，球菌Ruminococcus torques（D）, 费氏杆菌Faecalibacterium prausnitzii(E)和真杆菌Eubacterium rectale（F）在LLD（蓝色）和HMP（红色）的差异比较。

**4 肠道菌群的遗传特异性可用来做宿主的“微生物指纹”**

研究者发现一些肠道菌群在不同人体间的遗传差异很大，但是在同一个体内能保持稳定，比如甲烷短杆菌Methanobrevibacter smithii （Figure 2A）。以100对碱基（bp）为单位，相同个体内不同时间点的M. smithii平均仅有0.11个bp的差异，然而在不同个体间差异达到2.77bp（PWilcoxon=3.6x10-64）。这也使得用肠道菌群来鉴定宿主成为了肯能。研究者从100名个体配对的两个时间点的M. smithii遗传距离，成功鉴定了94名个体，即准确率达到94%。同时，研究者也用其它数据进行了个体鉴别，例如代谢物，肠道菌群菌种，菌群的SNP等，除了高鉴定率的M. smithii，琥珀酸杆菌Phascolarctobacterium succinatutens的准确率也达到了88%。然而，肠道菌群的菌种和生物通路的相对丰度鉴定率仅为12%和5%，这表明了人体肠道菌群的遗传特征比其相对丰度更具有个体特异性。

研究者进一步采用了双阶段交叉验证的方法，将71个数据特征（包括所有菌种SNP,SVs等）一同纳入模型，得到了85%的鉴定率（Figure 3A），优化后的模型包含30个数据特征，鉴定率仍高达82%（Figure 3B）。在特异性和敏感度分析中，ROC曲线面积（AUC）可以达到95% （Figure 3C）。当选取0.46微生物距离作为阈值时，对应的特异性为99%，敏感度为88%（Figure 3D）。以此参数计算，研究者可以以93%的准确度鉴定298对两个时间点配对的相同个体。此外，研究者在HMP队列中做了验证，在43对配对的相同个体中，该“微生物指纹”算法可以对41对个体达到100%的准确鉴别率（Figure 3B）。

图3.人体宿主中肠道菌群作为生物指纹的表现。A， 整合所有微生物的遗传特征和组成菌种可以在LLD队列中338个个体的纵向样本达到85%的识别准确度。在LLD中最佳优化后生物指纹模型包含30个微生物特征，准确度达82%。横坐标表示在逐步特征选择中的特征数量。灰色点表示模型的准确度。蓝色线表示在所有模型中的准确度拟合线。B，在LLD（蓝色）和HMP（红色）队列中生物指纹模型的准确度比较。C，在LLD（蓝色）和HMP（红色）队列中指纹模型的ROC分析，同时给出了两个队列曲线下面积和95%的置信区间。D，优化后模型中，相同个体内和不同个体间差异矩阵分布。在LLD队列中，在距离矩阵0.46的阈值下，模型识别度可以达到0.99的特异性和0.88的敏感度。在相同阈值下，在HMP的队列中，特异性和敏感度分别为1和0.95.

**5 肠道菌群相对丰度和遗传变异与宿主表型相关**

为了探究肠道菌群在人体健康中的作用，研究者对4年间肠道菌群的变化和宿主表型的变化进行了关联，采用两步分析策略。第一步，研究者采用混合线性模型对两个时间点的菌群特征和51个宿主表型数据进行分析。第二步，选取第一步FDR<0.05的结果，进一步采用差异（delta）关联分析，例如两个时间点微生物以及宿主表型变化量的关联。结果发现了169个菌群特征和34个宿主表型的190个显著关联（第一步FDR<0.05，第二步P<0.05，且变化趋势一致）。其中，人体的BMI，血压，HbA1C和抑郁症所关联的菌群特征数量最多（Figure 4A）。比如，人体的收缩压和链霉菌Lachnospiraceae bacterium呈正相关（betadelta=0.24, Pdelta=1.1x10-5, Figure 4B）；糖化血红蛋白（HbA1C）和黄素合成负相关（betadelta=-0.22, Pdelta=4.9x10-5, **Figure 4C**）。再如，菌群的SV变化与人体免疫相关联。布劳特氏菌Blautia obeum中的一个SV（3019-3020kb）与人体血液淋巴细胞负相关（betadelta=-0.29, Pdelta=6.5x10-4, **Figure 4D**）。在疾病方面，研究者发现了15个和抑郁症关联，3个和肠易激综合症关联，3个和哮喘关联。例如科林斯菌的SV（1164到1165kb，编码组氨酸激酶）和抑郁症关联最显著（**Figure 4E**）。

图4.人体表型和肠道菌群的差异（delta）关联。A，肠道微生物和表型的关联。每个条形柱表示了一个表型，并按照关联的数量进行排序。不同的颜色代表不同的微生物因素，例如，红色表示生物通路的相对丰度，蓝色表示菌种的相对丰度，橙色表示结构缺失，绿色表示结构变异。B,收缩压和毛螺菌（Lachnospiraceae bacterium）的正差异相关。C，血液HbA1C和黄素生物合成通路的负差异相关。D，较高丰度的布劳特氏菌的vSV (3019-3020 kb) 与血液淋巴细胞较低的差异（delta）相关。差异（delta）表示两个时间点微生物因素和表型间的差值。E, 柯林斯菌4\_8\_47FAA dSV (1164-1165 kb)片段的缺失和抑郁症病发相关。相关性用2x2联表表示。

**6肠道菌群相对丰度和遗传变异与宿血液代谢物相关**

为了进一步分析肠道菌群在人体疾病中的作用，研究者假设人体血液代谢物是其中重要参与者。通过对群体两个时间点的1183种血液代谢物的比较，研究者发现近27%的代谢物发生了显著改变（FDR<0.05）。

研究者首先探究了哪些血液代谢物在携带不同群菌亚种的个体中有差异，如果有差异，进一步检验菌群亚种的改变是否和这些代谢物的变化相关。研究者总共观察到63个代谢物和5种菌种的亚种簇。例如，在292个配对的样本中，有两种截然不同的费氏杆菌Faecalibacterium prausnitzii的亚种簇和15种代谢物相关(**Figure 5A**)。最显著相关的代谢物如利索异黄烷，吡咯，对甲酚硫酸盐的水平都在费氏杆菌Faecalibacterium prausnitzii亚种簇2中显著降低。同时，在24个菌群亚种簇改变（Faecalibacterium prausnitzii，由1到2）的个体中，发现这些代谢物水平降低；在13个个体中（Faecalibacterium prausnitzii，由2到1），这些代谢物水平升高（**Figure 5B-E**）。

研究者也观察到了133个肠道菌群特征（菌种和功能丰度，SVs）和81个代谢物变化相关（FDRjoint<0.05, Pdelta<0.05, **Figure 6A**）。其中很多关联发现都是已知的和肠道菌群相关的代谢物。和硫胺素最相关的菌种是理研菌Alistipes senegalen (**Figure 6B**)和拟杆菌属Bacteroidales bacterium以及微生物TCA循坏。理研菌Alistipes senegalen也是已知的富含硫胺素合成的细菌。

另外种重要的代谢物是蛋白结合性尿毒症毒素，是已知的和肠道菌群氨基酸代谢以及多种慢性病相关。研究者探究了58种尿毒症毒素，发现它们和微生物显著富集性相关（**Figure 6A**）。比如最显著和对甲酚硫酸盐相关的微生物是拟杆菌属Bacteroidales bacterium ph8 （**Figure 6C**）。值得注意的是，22.6%的菌群-代谢物关联（455中的103）是在布劳特氏菌属Blautia wexle的SV中发现（**Figure 6A**），其中27种是和尿毒症毒素相关（**Figure 6D**）。而这些SV也是编码膜结构，氨基酸酶，尿酶和蛋白结合基因。

图5. 费氏杆菌Faecalibacterium prausnitzii 亚种的替换与血液代谢物改变相关。A.基于SNP单倍型分析，得出两种不同的费氏杆菌Faecalibacterium prausnitzii亚种，图中用红蓝区分。B. 费氏杆菌Faecalibacterium prausnitzii的亚种替换。携带亚种由1到2的个体表示为绿色，由2到1的个体表示为橙色。C-E,亚种的替换和不同血液代谢物改变的关联，利索异黄烷A(C)，1,2,5-三甲基-1H-吡咯(D)和P-硫酸甲酚(E)。

图6. 人体血液代谢物差异（delta）和肠道菌群的关联。A,455个显著的肠道菌群和血液代谢的关联。B，硫胺素和理研菌Alistipes senegalen的差异正向相关。C，来自微生物的P-硫酸甲酚(E)和拟杆菌属Bacteroidales bacterium丰度变化相关。D,多种布劳特氏菌属Blautia wexle的vSVs和菌群来源的尿毒症毒素关联热图。

**7肠道菌群通过血液代谢影响宿主表型**

169个肠道菌群特征和宿主临床表型相关，122个菌群特征和血液代谢物相关，这其中29个和两者都相关（**Figure 7A**）。为了揭示微生物是否通过血液代谢来影响宿主表型，研究者采用了双向中介分析，发现了21个中介关联（FDRmediation<0.05 and Pinverse mediation>0.05, **Figure 7B**）。绝大多数都是微生物通过介导硫胺素（9个）和乙酰基-N-甲酰基-5-甲氧基犬尿氨酸（AFMK，9个）来影响人体血压。其中，硫胺素对人体心血管疾病的影响已经在临床随机对照试验中验证。AFMK是褪黑素的一种降解产物，可以通过抑制前列腺素合成来降低血压。中介分析发现多种肠道细菌可以参与到这些通路中。例如微生物硫酸盐还原生物通路可以通过提高血浆硫胺素的水平来降低血压（21%, Pmediation=6.0x10-3, **Figure 7C**）；微生物脂多糖的合成也可以通过影响血浆AFMK来调节血压（16%, Pmediation=6.0x10-3, **Figure 7D**）。此外，研究者也发现了血液代谢物也介导了肠道菌群对人体血脂和血糖的影响。例如酪醇4-硫酸盐，一种尿毒症毒素，介导了球菌Ruminococcus的SV对人体低密度脂蛋白的影响（17%, Pmediation=0.017, **Figure 7E**）。

图7. 肠道菌群，宿主代谢物和宿主表型间的中介分析。A，肠道菌群和宿主表型及代谢物相关的venn图。B，21个显著的中介分析结果。左边部分表示菌群因素，中间为血液代谢物，右边为宿主表型。C,菌群硫酸盐降解通路可以通过改变硫胺素来影响舒张压(Pmediation=0.006, 21%)。D，菌群脂多糖通过改变AFMK影响收缩压(Pmediation=622 0.006, 16%)。E，球菌Ruminococcus的vSV(300-305 kb)通过改变硫酸酪醇来影响血液LDL。

**8肠道菌群抗药性的增强**

细菌的抗药性给临床治疗感染性疾病带来了巨大的挑战，细菌的毒力因子也是肠道菌群能稳定存在于人体消化道以及对抗人体免疫的保证。研究者最后比较了两个时间点肠道菌群29个抗药性基因和59个毒力因子基因的相对丰度。发现抗性基因总量显著提高（P=1.1x10-9）而毒力因子基因总量下降（P=5.1x10-4）。单基因水平而言，55.2%（29个中的16个）的抗药性基因水平和18.6%（59个中的11个）的毒力因子基因水平发生了显著变化（FDR<0.05）。

荷兰的抗生素人体使用量在欧洲是最低的，但是四环素，氨基糖苷和林可酰胺是广泛的兽用药。研究者又对人体肉类食用量和菌群抗药性基因的改变做了关联分析，发现随着肉类食用量的提高，肠道菌群的氨基糖苷（rSpearman=0.18, P=9.2x10-4）和林可酰胺（rSpearman=0.15, P=5.5x10-3）抗药性基因水平也随之提高。这说明了当前畜牧业大量抗生素的使用，可能会造成人体肠道菌群耐药性的增加。

图S7. 肠道菌群的抗药性基因和毒力因子的长期改变。A,在基线（浅蓝）和后续（深蓝）两个时间点的总体菌群抗药性基因总量的比较。B, 在基线（浅蓝）和后续（深蓝）两个时间点的总体菌群毒力因子总量的比较。C,菌群丰度和抗药性基因丰度的正向差异变化。D，肠道菌群氨基糖苷抗性基因和肉食类食用的关联。E，肠道菌群林可酰胺抗性基因和肉食类食用的关联。

**结论**

人体肠道菌群的改变和人体诸多疾病相关联，这一结论在很多研究中都以发现。然而，大部分研究停留在横断面队列研究（cross-sectional），但肠道菌群在人体内是否长期稳定，人体内肠道菌群的变化是否可以导致宿主表型变化，无法从横断面队列研究中获得答案。格罗宁根大学医学中心在对338个个体4年前后的肠道微生物宏基因组，血液代谢组，表型组等配对多组学数据整合研究后，为上述问题提供了答案。1）4年的时间段内，同一个体内的肠道菌群相对丰度，遗传结构等相对稳定，并且菌群的个人特异性，可以用来作为“微生物指纹”；2）通过4年的纵向队列研究（longitudinal），研究者也揭示了肠道微生物，宿主血液代谢和宿主疾病的潜在因果关系；3）研究者发现人体肠道微生物中抗药性基因的增加和肉类食用量成正相关，表明了畜牧业中抗生素的使用可能助长了肠道细菌的耐药性。

**短评**

该研究基于普通群体大队列，4年前后两个时间点的肠道微生物数据，并结合其他多组学数据，深入探究了人体肠道菌群的组成和功能长期稳定性，揭示了肠道菌群和人体血液代谢物的关联变化，提供了大量潜在的代谢疾病病因。特别突出的是，该研究发现了人体部分菌群存在长期的稳定性，并有高度的个体特异性的这一普遍规律。同时，也为后续的肠道菌群longitudinal的研究和实验验证提供了重要的参考价值。