2 研究方法

2.1数据的获得与整理

得益于动态血糖检测技术，我们首次有望为患妊娠期糖尿病的每位孕妇个体都建立属于自己的精却的营养学建议，从而进一步保障母亲与后代的健康水平。

与杭州妇女医院合作的西湖大学生命科学院自2019年以来就开始了相关数据的收集。如果病人被确证为妊娠期糖尿病，符合研究条件的患者将在今后的观察期内佩戴运动传感器以及实时血糖仪，进而获得患者的实时血糖数据和运动情况。再搭配记录饮食的手机APP，我们可以获得每个个体对不同饮食的生理反应，进而运用计算机和统计学对患者进行多组学分析，来帮助她们找到适合自己的糖尿病治疗方案[1]。

2.2 数据的分析

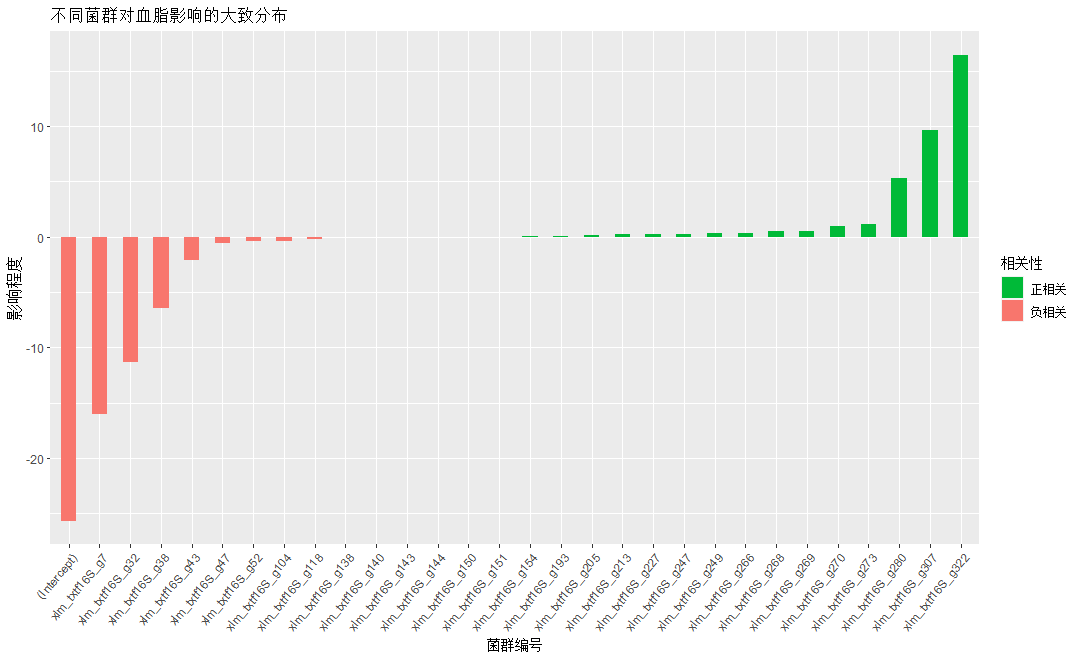
鄞州高级中学科技新苗团队的成员收到了来自西湖大学郑矩圣实验室的部分患者数据。我们在郑教授与助手苟望龙老师的指导下，使用R语言对数据进行了初步处理。

在分析时，我们把目光聚焦在了容易被大众忽视的胆固醇上，并把重心放在了肠道菌群与胆固醇之间的作用关系上。通过热力图，我们很难直接发现胆固醇与饮食或肠道菌群之间的直接联系。但是进一步挖掘后，我们成功发现了在数字背后隐藏的关键信息。

由于我们获得的数据中，有205个潜在的影响因素，但只有100位患者的数据可供分析。从数学角度上分析，若是希望获得胆固醇与影响因素的关系，则必须有足够的方程组可供联立。而此时的自变量明显多于数据量，因此无法直接建立数学模型。所以我们首先需要初步筛选出较显著的菌群才能进行下一步计算。

为了迈出第一步，我们使用岭回归对众多变量进行了初步筛选。岭回归是一种通过对数据进行正则化，来提高回归模型对变量的识别能力的数学方法。通过该方法，我们能够快速筛选出与血脂有较显著关联的肠道菌群种类。

经过初步筛选后，我们将目标菌群范围缩小到了29个，足够建立一个线性回归模型了。为了更直观地看到菌群与胆固醇的相关性，我们建立了一个普通的线性回归模型，并用计算结果绘制了以下柱状图：



这幅图的结果未考虑饮食等因素的影响，但是能让我们大致了解菌群与血脂的相关性。为了得到更准确的结果，我们还需要继续研究。

我们将得到的第一批菌群与患者的运动情况等数据结合在一起作为自变量进行了传统的线性回归，并通过筛选模型返回的P-value等统计学指标进行二次筛选。最后我们将饮食数据与菌群数据一起进行第三次线性回归，并且得到了最终的结果。第三次回归也是最后一次回归，目的在于防止我们遗漏饮食因素对血脂的影响，从而保障结果的严谨和全面。

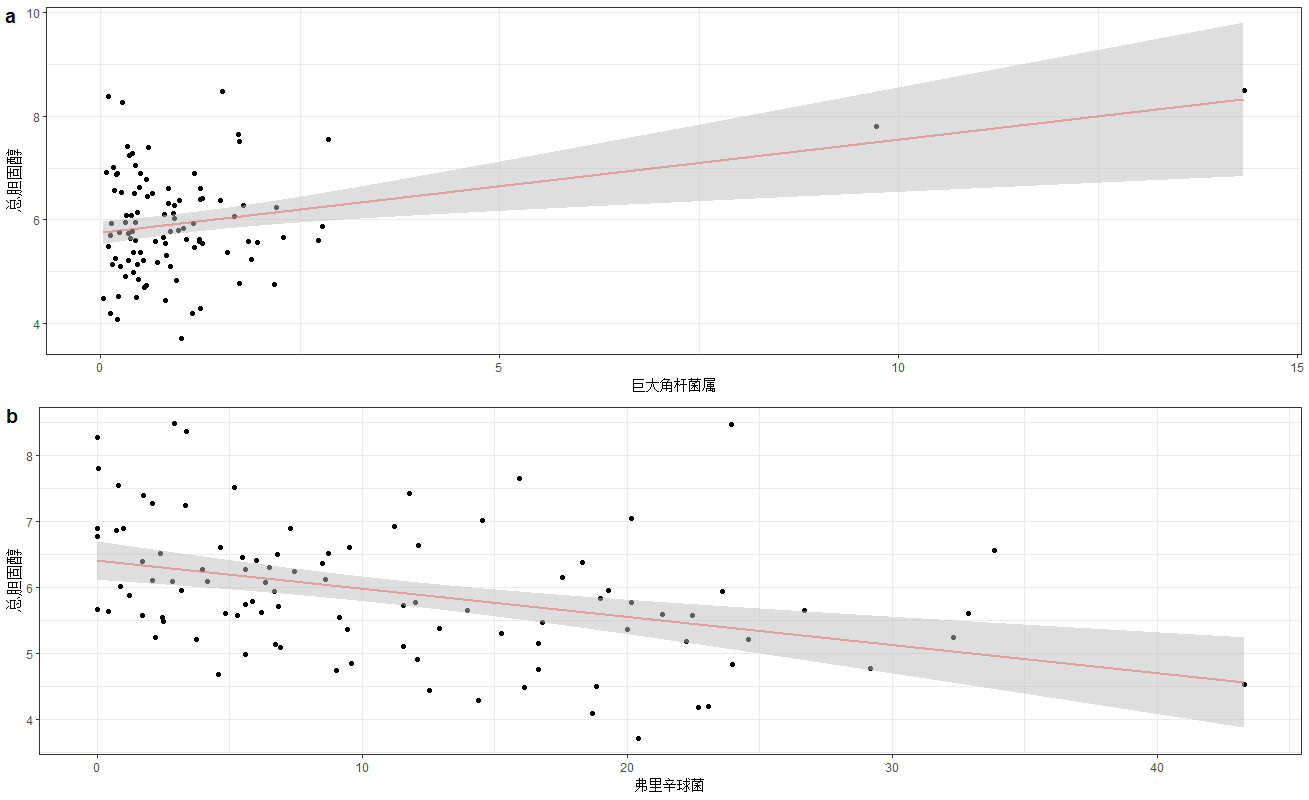
2.3 分析结果

最终结合多方面的因素，在考察了回归模型的P值等统计学参数后，有以下两种肠道菌群被我们认定对血脂有较显著影响：

d\_\_Bacteria;p\_\_Actinobacteriota;c\_\_Coriobacteriia;o\_\_Coriobacteriales;f\_\_Coriobacteriaceae;g\_\_Enorma（中文名：巨大角杆菌属，P-value = 0.03563885，与TC、LDL-C水平呈正相关）；

d\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Lachnospirales;f\_\_Lachnospiraceae;g\_\_Frisingicoccus（中文名：弗里辛球菌，P-value = 0.003784949，与TC、LDL-C水平呈负相关）；

通过进一步对这些菌群进行研究，我们有望研制出更具有针对性的降血脂药物，从而更好地治疗妊娠期糖尿病。

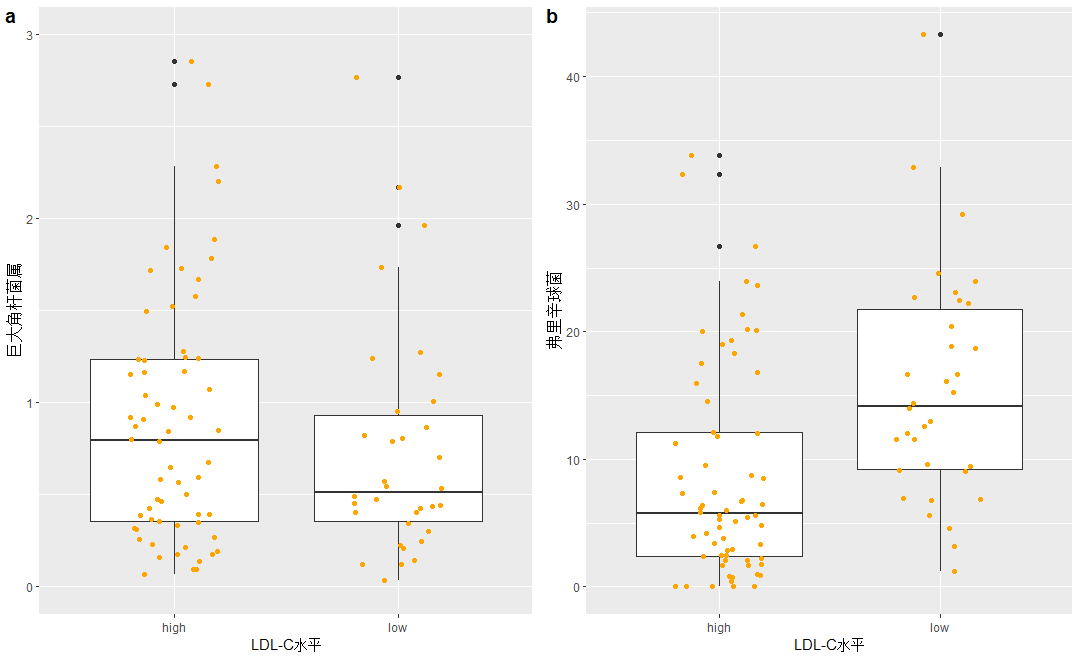


3 讨论

3.1血脂水平与相关菌群丰度存在显著关联

按照血液中LDL-C的水平，我们以2.6 mmol/L的LDL-C浓度为标准，将数据集分为高水平组与低水平组[2]。高水平LDL-C个体更容易患心血管疾病[3]，因此有必要研究对LDL-C有影响的菌群。

通过箱图，我们不难看出，菌群丰度与LDL-C水平之间有明显的关联，且与在上文中得出的结论互相印证。因此，我们也许可以为高血脂疾病、心血管疾病的治疗和预防提供新的思路。

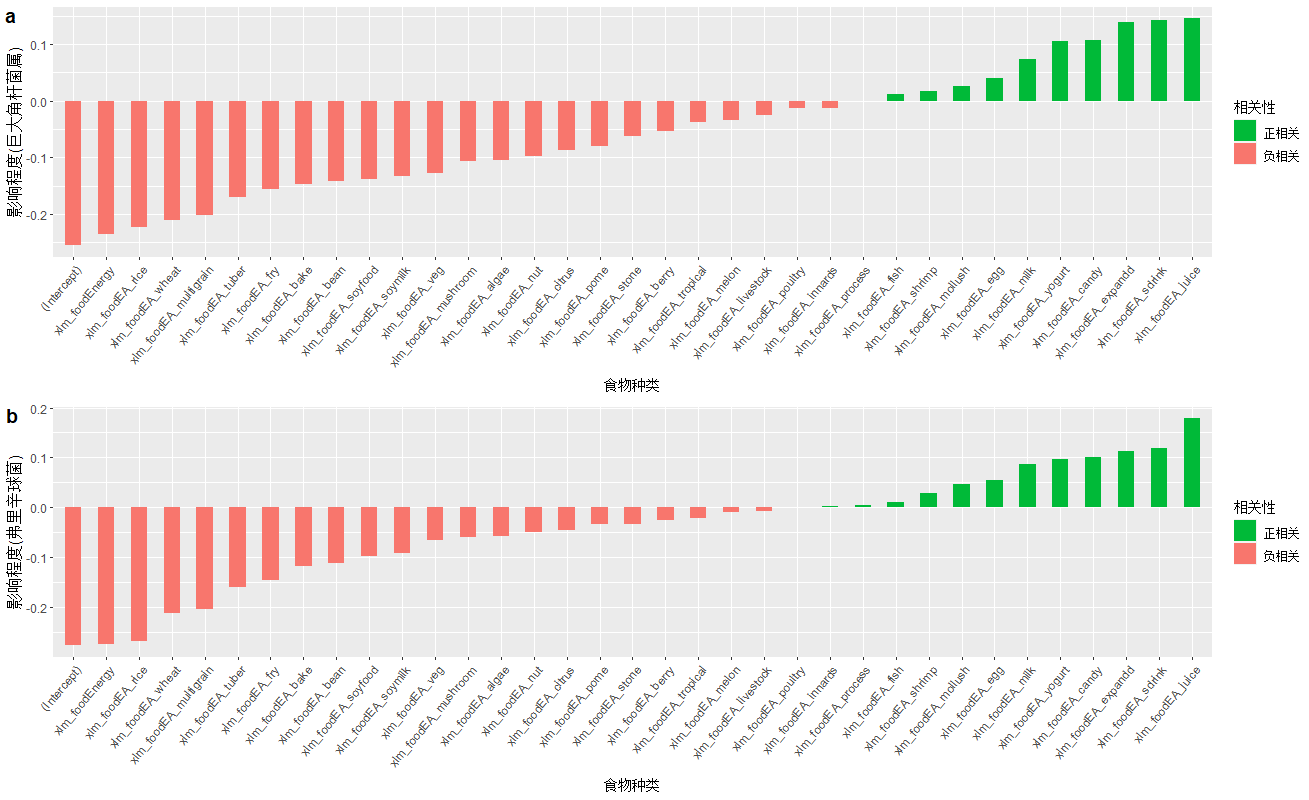


3.2 肠道菌群对生命活动的重要性

李兰娟院士在首届西湖生物大数据与健康创新论坛上曾做过一个关于肠道微生态概述、肠道微生态与宿主免疫代谢相关性的报告。在研究中，无菌大鼠的免疫力普遍低于非无菌大鼠，印证了肠道菌群在动物正常生命活动中的重要作用。研究肠道微生物，可能为人类的某些疾病找到全新的解决思路。

3.3 菌群与菌群，以及食物与菌群间的联系

食疗在糖尿病患者的治疗中占据非常重要的地位。通过探究食物对菌群以及血脂的影响，医生和营养学家可以糖尿病患者提供更科学的饮食建议。在此，我们可以简单地利用R语言进行探究。



出于时间因素，我们未能进一步进行深层次的研究。通过建立多层神经网络，我们有望进一步发现肠道菌群、食物、BMI、运动量、年龄等多重要素与血脂、血糖间的关系，并不断完善妊娠期糖尿病的数学模型。我们相信，在未来，针对每个患者的精确的多组学治疗方案将逐渐成为主流的疾病治疗和预防手段。届时，社会的医疗水平和人民的健康水平将达到新的高度。

1. Shengtao Gao , Jinfeng Wang, Maternal and infant microbiome: next-generation indicators and targets for intergenerational health and nutrition care, Protein & Cell, 2023;, pwad029, <https://doi.org/10.1093/procel/pwad029>
2. 中国成人血脂异常防治指南修订联合委员会. 中国成人血脂异常防治指南(2016年修订版)[J]. 中华心血管病杂志,2016,44(10)：833-853. DOI: 10.3760/cma.j.issn.0253-3758.2016.10.005
3. Verbeek R, Hoogeveen RM, Langsted A, Stiekema LCA, Verweij SL, Hovingh GK, Wareham NJ, Khaw KT, Boekholdt SM, Nordestgaard BG, Stroes ESG. Cardiovascular disease risk associated with elevated lipoprotein(a) attenuates at low low-density lipoprotein cholesterol levels in a primary prevention setting. Eur Heart J. 2018 Jul 14;39(27):2589-2596. doi: 10.1093/eurheartj/ehy334. PMID: 29931232; PMCID: PMC6287703.