**2024.02.29**

发现TwoSampleMR包在更新后，提取暴露、结局速度明显加快，开始加快这个课题的进度

**2024.03.30**  
1. 肝胆指标补充完善、心血管指标、心血管疾病IEU GWAS sumamry数据汇总补充

1. 进行肝胆指标、心血管指标、心血管疾病的暴露提取

（LDL-c、HDL-c、总胆固醇、TG、ApoB、γ-谷氨酰转移酶这种既可以认为是肝胆指标的也可以认为是心血管指标的，只放在肝胆指标这里了）

提取代码在这个路径下面：

2024.04.11

本地vcf文件进行MR的代码在这个路径

/home/linmiao/ZhangjunYu/MR/Code/local\_analysis

2024.05.05-2024.05.09

进行芬兰数据的清洗，截至目前未取得预期效果

芬兰数据中有

**方法一 MungeSumstats**

参考：<https://github.com/neurogenomics/MungeSumstats>

会考虑碱基是什么，只有完全匹配才会转化，但是只能转换单等位基因和SNP中ref和alt只有一个碱基的SNP，转化率非常低，X染色体注释不了。

**方法二 BSgenome**

参考：https://zhuanlan.zhihu.com/p/650938857

准确率很可以，和NCBI的SNP信息匹配的非常好，有些甚至比finngen原本的SNP匹配的还要好，X染色体也能注释，但是只有单等位基因，没有indel，是个很严重的问题。匹配过程不看碱基，只看染色体和pos。

**方法三 SumStatsRehab**

参考：https://github.com/Kukuster/SumStatsRehab

可以注释indel，但是总体准确率一般，有3w左右注释的不是很成功，X染色体注释不出来。匹配过程不看碱基，只看染色体和pos。

**方法四 Gwas2vcf**

参考：<https://mrcieu.github.io/gwas2vcf/install/>

Gh37的还不错，是OK的，运行效果各方面都很好，但是GH38的转化不出来，原因不明。

**方法五 Gwaslab**

非常不错的python包，完美契合目标，很不错