

西 安 交 通 大 学

系（专业） 1223 理科试验班（化学生物H）

系（专业）主任 张健保

批 准 日 期 2021 年 11 月 03 日

毕业设计(论文)任务书

生命科学与技术学院学院 1223 理科试验班（化学生物H 专业化生试验班 81 班学生郭骥瑞

毕业设计(论文)课题 基于统计学习的骨关节炎风险预测模型构建

毕业设计(论文)工作自 2021 年 9 月 11 日起至 2022 年 6 月 22 日止

毕业设计(论文)进行地点：生物医学信息与基因组学中心

课题的背景、意义及培养目标

背景与意义：骨关节炎是一种可引起不可逆关节病变，致残性较高的自身免疫性疾病。骨关节炎在中老年人群中发病率较高，60 岁以上人群患病率可达 60%，75 岁以上人群则可达到 80%。尤其是近年来，骨关节炎的发病率与致残率持续增长，对我国公共卫生系统构成了持续威胁。但由于初期症状不明显，诊断较为困难，骨关节炎患者往往会错过最佳治疗时机，导致疾病恶化，危害患者身体健康与生活质量。因此，构建一种能够对骨关节炎患者进行早发现与早治疗的方法对降低人群发病率有着积极的意义。培养目标：利用统计学习方法系统评价骨关节炎与个体基因、代谢物乃至环境之间的相关关系和因果关系，构建骨关节炎风险预测模型。掌握运用统计学习型研究复杂疾病的方法。

设计(论文)的原始数据与资料

1.目前中心已经建立了已经征集并获取了来自本省的数千例样本的 DNA 和表型信息，且样本量在持续增长中；2.本中心已申请获得了 UKB 约 17 万人群体的基因型数据，为本研究的进行打下了良好的数据基础；3.中心已建立了多种统计遗传学研究方法的分析体系，具有自行开发的遗传分析软件，可保障本课题的顺利实施。 

课题的主要任务

1.收集整理本中心自有群体和 UKB 约 17 万人群体表型数据；收集与骨关节炎相关的全基因组关联分析数据；2.利用统计学习方法分析与骨关节炎相关的基因位点或表型，并分析相关代谢通路；

3.构建骨关节炎风险预测模型并验证其效果。

课题的基本要求(工程设计类题应有技术经济分析要求)

1.课题进行前，要充分查阅相关文献资料，积累基本背景知识，学习相关的技术和方法；2.课题进行过程中，需要进行编程处理相关文件并对数据进行科学的统计、分析；3.对于出现的数据分析结果能够较全面地解释，理解其意义；4.通过课题研究，理解并掌握相关的专业基础理论知识，对研究结果能提出独特见解。

完成任务后提交的书面材料要求(图纸规格、数量，论文字数，外文翻译字数等)

1.按时提交论文一份（15000字），规格符合学校的毕业论文要求；2.翻译英文资料一份（3000字）；3.提供原始数据及相关文献资料。

主要参考文献

1.BoerCGetal.Decipheringosteoarthritisgeneticsacross826,690individualsfrom9populations.[J]Cell.2021Sep2;184(18):4784-4818.e17.2.OsvalAMontesinos-Lópezetal.NewDeepLearningGenomic-BasedPredictionModelforMultipleTraitswithBinary,Ordinal,andContinuousPhenotypes[B].G3GenesGenomesGenetics,Volume9,Issue5,1May2019,Pages1545–1556.

指导教师： 郭燕

接受设计(论文)任务日期： 2021年11月03日

（注：由指导教师填写）

学生签名：