

陈克乐 生物信息工程师

性别：男
民族：汉族
出生年份：1994 年
籍贯：江苏省连云港市

联系电话：18668129066
电子邮箱：chkl@zju.edu.cn
学历学位：硕士研究生
邮编：310030



教育经历

2016 年 9 月至 2019 年 3 月，就读于浙江大学医学院基础医学系生物化学专业，在徐立红教授课题组从事环境毒理研究，我的研究课题为微囊藻毒素对胰岛素通路的影响

2012 年 9 月-2016 年 7 月，就读于长春理工大学生命科学技术学院生物工程专业，主修高等数学、概率论与数理统计、线性代数、生物化学、细胞生物学，以专业第一获得推免资格进入浙大医学院

奖励及荣誉

- 研究生期间获浙江大学学业奖学金
- 本科期间获国家励志奖学金 2 次、省政府奖学金 1 次、校级奖学金 6 次（其中校长奖学金一次）
- 本科期间获吉林省生物技能比赛三等奖
- 本科期间校优秀学生干部、校三好学生

个人能力

专业能力

- 具备全外显子组、低覆盖测序数据如 NIPT 及其升级版 (NIPTPlus) 数据分析经验，掌握其中的技术细节
- 了解转录本测序、chip-seq、TCGA 及 GEO 分析、GWAS 分析
- 熟悉常用生物信息学软件：bcl2fastq、fastp、bwa、bowtie2、picard、gatk、samtools、varscan2、bamdst、annovar、SCiPE 等使用、了解 phenolyzer、spliceAI、plink
- 熟悉 linux 环境下工作，具备 shell 脚本编写能力
- 熟悉并主要使用 python 编程，具备 python 数据分析以及数据可视化的 (pandas 和 matplotlib) 能力，具备 python 多进程编程 (multiprocessing) 能力，可以编写 sql 语句
- 掌握 java、perl 以及 R 语言语法，可以分析他人编写的 java 程序、perl 以及 R 写成的脚本并将其转换为 python，能够在 R 包 help 文档信息的帮助下完成绘图，例如热图和火山图
- 熟悉 anaconda、jupyter、Rstudio 等工具
- 熟练运用 git 及 gitlab 进行版本控制
- 熟练运用 markdown
- 熟悉常见的生物信息学数据库，例如 NCBI、UCSC、OMIM、clingen、Decipher、TCGA、ebi 等
- 具备医学统计学知识，能够将其运用到新的研发或者优化需求，了解常见算法如回归分析、PCA、HMM
- 具备分子生物学科研思维以及实验技能，具备生物信息学新 R 包、python 模块、新软件的文档及文献调研能力

外语水平

- cet6
- 研究生期间有三个学期做留学生助教，为留学生分子医学实验课批改留学生提交的实验报告并提供答疑
- 具备生物信息学领域文献调研能力

写作能力

- 撰写中英文学术文章，领域涉及生化检测试剂研发、环境毒理、基础医学，已发表的文章在简历最后列出

交流组织能力

- 研究生期间进行若干次学术汇报
- 2019 年 4 月后工作期间主讲过部门 (约 20 人) 知识分享会两次：主题分别为靶向测序生物信息学数据分析流程、隐马尔可夫模型 (HMM)
- 工作期间公司研发与临检中心跨年派对 (中心约 100 人) 的主持人

工作经历

生物信息工程师 | 杭州甄元医学检验实验室有限公司 | 2019 年 4 月至今

- 人类基因组学数据库下载、爬虫与整理，维护公司历史致病位点信息
- 全外显子组测序数据分析流程中功能模块的构建，为流程的大版本更新制定转录本注释规则，视情况参与流程迭代后的测试工作
- 为临检中心以及研发中心实验室提供数据分析支持
- 为部门新的研发方向进行文献调研
- 独立承担全外下机数据单亲二倍体 (UPD) 提示程序的算法设计和编码开发
- 在公司间合作项目 NIPT 项目中，负责 NIPTPro 项目落地及后续研发工作的承接，包括硬件采购，系统调试，为完全承接研发工作从头编写代码文档

生物信息工程师实习 | 杭州甄元医学检验实验室有限公司 | 2018 年 11 月-2019 年 3 月

- 维护部门本地人基因组突变位点数据库，为遗传代谢病的致病位点诊断分析提供最新数据支持
- HGVS 命名规则、ACMG 指南的学习
- 为临检中心以及研发中心提供数据分析支持
- 大型文本批量处理

项目经验

一、全外的生物信息学流程的维护和优化

- **项目描述：** 公司用 perl 构建了初步的全外分析的流程，我需要在接手流程维护工作前熟悉公司正在使用的分析思路和数据库，并承担后续的需求变更后的优化工作，以及关于生物信息学分析流程的培训工作（培训对象为公司内的其它部门）
- **项目时间：** 2019 年 4 月-6 月
- **涉及技术：** shell、perl、全外测序数据分析流程知识、全外分析流程中的生物信息学软件的使用，常用的转录本库、annovar 后台 hg19_refGenewithVer.txt 文件格式
- **项目职责：** 作为流程维护和对外培训负责人，绘制全外分析流程图，负责根据对全外流程的理解提出优化方案并将方案落地
- **项目收益：** 理解了公司的全外数据分析流程各部分软件和参数；维护过程中解决了流程中存在的对数据准确性有轻微影响的 bug；设计出了新的转录本注释规则，并根据该规则实现公司内部转录本数据的统一；为伴随新 ACMG 指南产生的新注释需求整理本地致病位点库，简化了历史位点信息注释流程

二、为全外流程构建新的功能模块

- **项目描述：** 模块 1：原全外流程 annovar 注释后的命名矫正依赖于一个国外网站 mutlyzer 的在线处理，而该网站访问速度不稳定，尤其 14:00-16:00 之间速度很慢，所以导致了流程到这一步速度的不稳定，为了使命名矫正这一步摆脱对国外网站的依赖，需要开发具有类似功能的本地命令行版；模块 2：由于新 ACMG 指南的产生，需要为全外流程增加一个当突变发生在起始密码子时，查看是否存在框内 ATG，如果存在，注释框内 ATG 之间的致病位点个数及详细信息的模块；模块 3：由于公司的遗传代谢病检测 panel、以及全外测序分析产生的 bam 文件越来越多，占用服务器越来越多的空间，而保留这些 bam 文件的目的是方便医学遗传组的同事们查看 igv，需要设计一个既满足 igv 查看需要，又能节约服务器空间的方案，为此设计了模块 3。
- **项目时间：** 2019 年 5-10 月
- **涉及技术：** 模块 1：HGVS 命名规则、annovar 产生的命名存在的问题、mutlyzer 网站的矫正规则、python 文本处理；模块 2：clinvar 数据库、本地位点致病性数据库、由参考基因组和 gff 文件得到转录本序列、由转录本序列中的碱基得到参考基因组坐标；模块 3：bam 以及 sam 文件格式、对计算机压缩算法的理解、samtools、pysam
- **项目职责：** 作为各模块开发负责人，解决以上 3 个问题
- **项目收益：** 编写了本地的命名纠正程序；编写了框内 ATG 间致病位点的注释程序；成功将 bam 文件在格式不变的情况下占用空间变成原来的 1/3 且不影响遗传分析师用 igv 查看；个人收获：对 HGVS 命名规则、annovar 产生的命名问题、mutlyzer 矫正规则、clinvar 数据库、参考基因组与转录本之间的坐标转换有了更好的理解和实战经验

三、软件调研工作

- **项目描述：** 调研工作 1：生信数据分析下游的遗传分析工作中需要将软件预测的可变剪接结果作为参考，为此需要寻找业内受到广泛认可的可变剪接预测软件；调研工作 2：我们的全外分析流程形成的 vcf 中突变有 4-5w 个，经过本地的各种过滤依然还会剩下数千个，为了提高遗传分析师的工作效率，需要按照送检医生登记的表型得到基因列表，遗传分析师根据列表判读，提高判读效率，调研前是人工处理，调研工作是为了找到可靠的自动工具取代人工
- **项目时间：** 2019 年 5 月和 10 月

- **涉及技术**：论坛调研、文献调研、linux 本地运行实测、R 语言
- **项目职责**：作为调研工作负责人，解决以上问题
- **项目收益**：对目前主流的可变剪接预测算法有了了解，成功找到了可用于可变剪接预测的开源软件，并成功部署到本地；调研了 phenolyzer 软件，现部署到本地用于根据表型产生 gene list

四、网络数据库的爬虫整理

- **项目描述**：该网络数据库是全外分析注释中常用的基因表型关系数据库，该数据库一天更新一次，为了维持本地的数据能够处于比较新的状态，我们需要按季度更新
- **项目时间**：2019 年 11 月至今
- **涉及技术**：python 的 request、beautifulsoup 模块、应对反爬虫策略
- **项目职责**：数据库维护工作的负责人，按季度更新该数据库
- **项目收益**：成功更新了数据库，对 python 的 request、beautifulsoup 模块和该数据库爬取有了实战经验

五、全外家系样本 UPD 提示程序开发

- **项目描述**：单亲二倍体可能具有危害性，为了对全外家系样本提供更加全方位的检测结果提示，需要开发针对全外家系样本的 UPD 提示程序
- **项目时间**：2019 年 12 月
- **涉及技术**：单亲二倍体的类型以及特征，python 处理 vcf 文件，划窗统计算法、python 散点图和折线图的绘制
- **项目职责**：作为 UPD 提示程序的研发负责人，负责全部的算法设计和编码实现
- **项目收益**：开发出了第一个版本的 UPD 提示程序，以可视化形式提示本地收到的阳性样本，并显示是父源母源以及属于 iUPD 还是 hUPD 等信息

六、NIPT 升级版的项目落地和研发承接

- **项目描述**：公司的基础版 NIPT 受到市场的好评，升级版的预测加入了对胎儿 cnv 的预测
- **项目时间**：2020 年 1 月至今
- **涉及技术**：python、R、pandas、NIPT、生物信息学分析算法软件以及流程、医学统计学
- **项目职责**：作为项目落地与研发承接的负责人
- **项目收益**：负责硬件的采购、软件的安装调试、稳定性以及准确性的检验，承接研发工作，为 NIPT 升级版系统从头编写完整代码文档，掌握升级版 NIPT 的生物信息学分析技术细节。

已发表学术文章

- [1]（共同一作）Huang P, [Chen K](#), Ma T, et al. The effects of short-term treatment of microcystin-LR on the insulin pathway in both the HL7702 cell line and livers of mice. Environmental Toxicology. 2020;1-11.
- [2] Ma T, Cao N, [Chen K](#), et al. Microcystin-LR exposure disrupts the insulin signaling pathway in C2C12 mice muscle cell line. Environmental Toxicology. 2019;1-9.
- [3] Liu J, Xu C, Zhang S, Li H, [Chen K](#), Huang P, et al. Microcystin-LR disrupts insulin signaling by hyperphosphorylating insulin receptor substrate 1 and glycogen synthase. Environmental toxicology 2018; 33: 16-22.
- [4] 刘丹,刘照慧,[陈克乐](#),徐美玲,干高龙,于源华.三种生化试剂干片的制备及检验方法的建立[J].中国预防医学杂志,2016,17(06):409-413.
- [5] 杨惠宇,张昊,徐美玲,干高龙,[陈克乐](#),于源华.干式生化膜片的制备及其应用[J].应用化学,2015,32(08):963-968.
- [6] [陈克乐](#),黄朴,徐立红.蛋白磷酸酶 2A 影响胰岛素调节的研究进展[J].中国细胞生物学学报,2019,41(01):150-156.
- [7] [陈克乐](#).微囊藻毒素对人肝细胞系 HL7702 及小鼠肝组织胰岛素信号通路相关蛋白影响研究[D].浙江大学,2019.