|  |  |
| --- | --- |
| 委托协议编号 |  |

技术服务（测试化验加工）委托协议

|  |  |  |  |
| --- | --- | --- | --- |
| 委托任务名称 | 基于过敏性鼻炎纵向队列的复杂免疫调控机制研究项目基础医学研究服务采购项目-03包：数据分析服务 | | |
| 委托方（甲方） | 北京市耳鼻咽喉科研究所（北京市耳鼻咽喉头颈外科研究中心） | | |
| 单位负责人 | 王成硕 | | |
| 项目负责人 | 李景云 | 联系电话 | 010-58265810 |
| 单位通讯地址 | 北京市东城区后沟胡同17号 | | |
|  | | | |
| 受委托方（乙方） | 北京市计算中心有限公司 | | |
| 单位负责人 | 刘彤 | | |
| 联系人 | 王雪峰 | | |
| 联系电话 | 15001065280 | | |
| 单位通讯地址 | 北京市海淀区丰贤中路7号3号楼北京市计算中心 | | |

**填 写 说 明**

1. 本协议适用于我院科研人员在项目研究过程中支付给外单位的检验、测试、化验及加工等费用时需要签署的协议。
2. 合同封面的委托任务名称指本合同的测试加工等具体内容，应用简明规范的专业术语明确概括所要完成的服务内容。
3. 本合同的委托方（甲方）和受托方（乙方）名称，须按单位公章及法人营业执照中所记载的详细名称填写，若涉及外文名称，首次出现时应写明全称及简称。
4. 本协议书未尽事项，可由当事人附页另行约定，并可作为本协议的组成部分。如协议研究内容涉及国家秘密或重大商业秘密的，双方应另行签署保密协议，承担保密义务。
5. 使用本协议书时约定无须填写的条款，应在该条款处注明“无”等字样。
6. 协议书要求A4纸打印，一式4份，左侧装订，正文内容所用字型应不小于5号字，协议正本中所涉及与本协议约定事项有关的技术资料及其指定附件备齐后应合装成册，其规格大小应与协议书一致。
7. 受委托方需提供测试化验加工的原始数据，委托方务必保留原始数据10年以上以备审计抽查。

依据《中华人民共和国民法典》及本协议书相关的科研项目、经费管理办法规定，为完成甲方承担的研究任务，经双方协商一致，各方在真实、充分地表达各自意愿的基础上，就本协议书中所描述的委托内容、经费支付、保密内容、知识产权等问题达成如下协议，签订本合同并由签约双方共同恪守。

1. **委托工作的主要内容、加工方式和要求**

1、测试加工内容

（1）单细胞质控；（2）机器学习数据清洗；（3）标准化与降维；（4）统计学方法；（5）机器学习方法；（6）线性模型；（7）非线性模型；（8）集成学习；（9）性能指标；（10）验证策略（11）可解释性分析；（12）多组学整合；（13）数据质控；（14）批次校正；（15）数据合并；（16）聚类与可视化；（17）细胞注释；（18）差异分析；（19）功能富集；（20）拟时序分析；（21）免疫组库分析（22）细胞通讯（23）整合分析；（24）遗传关联；（25）结果整理；（26）代码交付；（27）迭代优化；（28）培训。

2、测试加工方式和要求

委托进行数据分析；乙方在规定时间内完成上述服务内容，保留完整的原始实验记录和数据。实验结束后将最终报告交付甲方。

1. **考核指标及验收标准、方式和验收时间**
2. 考核指标

|  |  |  |
| --- | --- | --- |
| **序号** | **产品** | **数量（天）** |
| 1 | 单细胞质控 | 6 |
| 2 | 机器学习数据清洗 | 5 |
| 3 | 标准化与降维 | 4 |
| 4 | 统计学方法 | 4 |
| 5 | 机器学习方法 | 12 |
| 6 | 线性模型 | 6 |
| 7 | 非线性模型 | 6 |
| 8 | 集成学习 | 14 |
| 9 | 性能指标 | 7 |
| 10 | 验证策略 | 5 |
| 11 | 可解释性分析 | 6 |
| 12 | 多组学整合 | 20 |
| 13 | 数据质控 | 4 |
| 14 | 批次校正 | 5 |
| 15 | 数据合并 | 6 |
| 16 | 聚类与可视化 | 6 |
| 17 | 细胞注释 | 8 |
| 18 | 差异分析 | 7 |
| 19 | 功能富集 | 6 |
| 20 | 拟时序分析 | 8 |
| 21 | 免疫组库分析 | 10 |
| 22 | 细胞通讯 | 8 |
| 23 | 整合分析 | 12 |
| 24 | 遗传关联 | 4 |
| 25 | 结果整理 | 15 |
| 26 | 代码交付 | 10 |
| 27 | 迭代优化 | 10 |
| 28 | 培训 | 2 |
|  |  | 216 |

1. 验收标准

（1）单细胞质控模块应用Seurat软件，进行细胞过滤（nFeature\_RNA＞100 & ＜6000，根据实际调整）和基因过滤（去除低表达/高dropout基因）进行数据清洗。通过设定线粒体基因表达比例阈值（如percent.mt ＜ 20）过滤受损细胞，最终得到高质量的单细胞数据。

（2）机器学习数据清洗模块应用KNN插值方法进行缺失值填充，使用ComBat和Scanorama算法进行批次校正。KNN插值中通常选择k值为3-7的邻居进行插值，通过验证批次效应去除效果并生成QC报告，确保数据的一致性和可比性。

（3）标准化与降维模块应用SCTransform或LogNormalize方法进行数据标准化，使用PCA或UMAP算法进行降维。PCA通常选择前50个主成分，通过结合特征选择模块优化降维参数，提取数据的关键特征并降低数据维度，便于后续分析和建模。

（4）统计学方法模块应用DESeq2或edgeR软件进行差异表达基因筛选，使用Wilcoxon秩和检验比较两组间的基因表达差异。通过设定P＞0.05，log2 fold change绝对值大于1，输出火山图和热图展示关键基因。高变异基因筛选（HVG）利用Seurat的FindVariableFeatures函数，识别高变异性基因。

（5）机器学习方法模块基于LASSO方法进行特征压缩，使用随机森林算法进行重要性排序。通过特征权重值，实现特征降维和关键特征提取。

（6）线性模型模块应用glm函数构建逻辑回归模型，使用survival包的coxph函数进行Cox回归分析生存数据。输出系数表及风险比（HR）评估模型的预测能力和风险因素的影响。

（7）非线性模型模块应用向量机（RBF核）模型和多层感知机（MLP）进行建模。通过优化超参数，提高模型的泛化能力和预测准确性。

（8）集成学习模块应用randomForest或xgboost包进行随机森林/XGBoost分析建模，通过特征重要性排序及模型可解释性分析（LIME/SHAP）增强模型的可解释性。

（9）性能指标模块应用pROC和MLmetrics包计算分类任务的F1score和ROCAUC，使用Metrics包计算回归任务的RMSE和R²。通过caret包生成混淆矩阵及校准曲线，全面评估模型的性能。

（10）验证策略模块应用5折交叉验证方法，将数据集分为训练集和验证集进行多次训练和验证。

（11）可解释性分析模块应用shapr或DALEX package包计算SHAP值，生成SHAPsummary图，直观呈现各特征对模型预测结果的影响。通路富集与模型特征关联分析借助clusterProfiler包，对关键基因进行GO/KEGG富集分析，生成特征通路网络图。

（12）多组学整合模块应用MOFA+框架进行单细胞 + 转录组联合建模和蛋白组数据融合。

（13）数据质控模块应用FastQC软件对测序数据进行初步质量评估，使用Seurat的AddModuleScore函数进行细胞周期基因校正。过滤低质量细胞时，设定质量控制指标（如nFeature\_RNA、nCount\_RNA等）的阈值，保证数据的质量和可用性。

（14）批次校正模块应用Harmony算法进行批次效应去除。通过计算校正前后的方差解释比例或使用t-SNE/UMAP降维可视化进行评估，确保数据的一致性和可比性。

（15）数据合并模块应用Seurat的Merge函数进行跨项目样本合并，并进行批次效应复检。

（16）聚类与可视化模块应用Seurat的FindClusters函数进行聚类分析，使用UMAP/tSNE算法进行降维可视化。通过交互式调整分群策略，确定合适的分辨率（如0.5/0.8/1.0），生成分群热图。

（17）细胞注释模块应用Seurat的FindAllMarkers函数识别细胞类型特异性基因标记，使用特定细胞类型的marker基因进行非目标细胞过滤。双标基因验证（iEOS/rEOS评分）利用Seurat的AddModuleScore函数，生成气泡图和小提琴图展示细胞类型特征和分布。

（18）差异分析模块应用Seurat的FindMarkers函数或limma包进行组间差异分析和亚群特异性差异基因筛选。设定P＞0.05，log2 fold change阈值为0.25-1。

（19）功能富集模块应用clusterProfiler包进行GO/KEGG富集分析和GSEA分析。对上下调基因分开分析，设定pvalueCutoff（如0.05）和qvalueCutoff（如0.2），并注释文献相关通路。

（20）拟时序分析模块应用Monocle软件进行细胞轨迹分析，使用Slingshot进行伪时序热图分析。通过降维（如t-SNE或UMAP）和图构建算法（如DDRTree或ICA）确定细胞状态转变路径，标注细胞状态转变统计。

（21）免疫组库分析模块应用scRepertoire包进行TCR/BCR克隆扩增分析和免疫组库多样性统计。设定clonotype定义参数，计算克隆频率和扩增指数，调整克隆阈值并可视化，深入探究免疫细胞的克隆特征和多样性。

（22）细胞通讯模块应用CellChat软件构建多组细胞互作网络，筛选显著的细胞间互作关系，得到网络图、热图和气泡图。

（23）整合分析模块应用GSVA或ssGSEA算法进行单细胞 + 转录组联合分析，使用MCP-counter算法进行免疫浸润评分。

（24）遗传关联模块应用SMR和TWAS工具进行共定位分析，使用LDBlockShow工具进行LD Block绘图。输入基因表达数据和GWAS summary数据，计算基因表达与性状之间的共定位概率（如SMR p-value ＜ 0.05）。

（25）结果整理模块应用R语言的data.table包处理差异基因表格和富集结果，使用ggplot2包制作高分辨率矢量图（PDF/PNG）。确保表格包含基因名、log2 fold change、p-value、adj. p-value、富集通路等信息，并按期刊要求的尺寸、字体和颜色规范进行格式化。

（26）代码交付模块提供Python/R完整分析流程，Seurat/Monocle完整分析流程，以及自定义函数注释。使用Git进行版本控制和Docker环境配置，确保代码的可用性和可复现性。

（27）迭代优化模块应用自动化调参工具进行模型调参（贝叶斯优化），支持参数调整（如分群分辨率/差异阈值）和结果复现与答疑。设定参数搜索空间（如学习率范围0.001-0.1、树深度3-10等），并利用交叉验证评估参数组合性能，按需支持不超过3轮。

（28）培训模块提供代码及流程交付使用培训，确保受培训人员理解并掌握交付的代码和分析流程。

1. 验收方式

完成数据分析服务后，甲、乙双方通过对服务项目成果文件的浏览，共同审查其是否符合本合同所列出的所有服务要求款项。如果双方对服务项目成果均无异议，则服务项目成果经双方确认后通过验收。

1. 验收时间

在甲方提供完整数据后90个自然日。

**第三条 测试化验加工细目：**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **测试化验加工的内容** | **测试结果的呈现方式** | **计量单位** | **单价**  **（万元/单位）** | **数量** | **金额（万元）** |
| **1** | **单细胞质控** | **数据分析结果文件（电子版）** | **天** | **0.1** | **6** | **0.6** |
| **2** | **机器学习数据清洗** | **数据分析结果文件（电子版）** | **天** | **0.1** | **5** | **0.5** |
| **3** | **标准化与降维** | **数据分析结果文件（电子版）** | **天** | **0.1** | **4** | **0.4** |
| **4** | **统计学方法** | **数据分析结果文件（电子版）** | **天** | **0.08** | **4** | **0.32** |
| **5** | **机器学习方法** | **数据分析结果文件（电子版）** | **天** | **0.14** | **12** | **1.68** |
| **6** | **线性模型** | **数据分析结果文件（电子版）** | **天** | **0.15** | **6** | **0.9** |
| **7** | **非线性模型** | **数据分析结果文件（电子版）** | **天** | **0.15** | **6** | **0.9** |
| **8** | **集成学习** | **数据分析结果文件（电子版）** | **天** | **0.15** | **14** | **2.1** |
| **9** | **性能指标** | **数据分析结果文件（电子版）** | **天** | **0.1** | **7** | **0.7** |
| **10** | **验证策略** | **数据分析结果文件（电子版）** | **天** | **0.1** | **5** | **0.5** |
| **11** | **可解释性分析** | **数据分析结果文件（电子版）** | **天** | **0.12** | **6** | **0.72** |
| **12** | **多组学整合** | **数据分析结果文件（电子版）** | **天** | **0.1** | **20** | **2** |
| **13** | **数据质控** | **数据分析结果文件（电子版）** | **天** | **0.14** | **4** | **0.56** |
| **14** | **批次校正** | **数据分析结果文件（电子版）** | **天** | **0.1** | **5** | **0.5** |
| **15** | **数据合并** | **数据分析结果文件（电子版）** | **天** | **0.08** | **6** | **0.48** |
| **16** | **聚类与可视化** | **数据分析结果文件（电子版）** | **天** | **0.12** | **6** | **0.72** |
| **17** | **细胞注释** | **数据分析结果文件（电子版）** | **天** | **0.12** | **8** | **0.96** |
| **18** | **差异分析** | **数据分析结果文件（电子版）** | **天** | **0.1** | **7** | **0.7** |
| **19** | **功能富集** | **数据分析结果文件（电子版）** | **天** | **0.1** | **6** | **0.6** |
| **20** | **拟时序分析** | **数据分析结果文件（电子版）** | **天** | **0.15** | **8** | **1.2** |
| **21** | **免疫组库分析** | **数据分析结果文件（电子版）** | **天** | **0.12** | **10** | **1.2** |
| **22** | **细胞通讯** | **数据分析结果文件（电子版）** | **天** | **0.12** | **8** | **0.96** |
| **23** | **整合分析** | **数据分析结果文件（电子版）** | **天** | **0.15** | **12** | **1.8** |
| **24** | **遗传关联** | **数据分析结果文件（电子版）** | **天** | **0.12** | **4** | **0.48** |
| **25** | **结果整理** | **数据分析结果文件（电子版）** | **天** | **0.1** | **15** | **1.5** |
| **26** | **代码交付** | **代码及使用说明1套** | **天** | **0.1** | **10** | **1** |
| **27** | **迭代优化** | **代码及使用说明1套** | **天** | **0.1** | **10** | **1** |
| **28** | **培训** | **培训相关文件1套** | **天** | **0** | **2** | **0** |
|  | **合计** | | | | | **24.98** |

**第四条 经费支付方式：**

1. 甲方应支付委托费用共计24.98万元（包括全部费用及税费，甲方无需向乙方支付其他费用）。

2. 支付方式：一次支付受委托方 （按以下第①种方式）：

①一次总付：24.98万元，乙方在完成测试服务通过验收且开具增值税发票后60个工作日内支付全部测试费。

②分期支付：

第一次支付 万元，甲方在本合同签订后且乙方开具当期金额增值税发票后的

个工作日内支付。

第二次支付 万元，甲方在乙方通过全部测试技术服务验收且开具当期金额增值税发票后的 个工作日内支付。

③其它方式：

1. 乙方账户信息：

户名：北京市计算中心有限公司

开户行：宁波银行北京分行营业部

账号：77010122001379612

4.甲方开具发票的要求及信息：

单位开户名称：北京市耳鼻咽喉科研究所（北京市耳鼻咽喉头颈外科研究中心）

单位开户银行：北京银行前门文创支行

纳税人识别号：12110000400686363A

1. **知识产权归属**
2. 双方在履行本协议之前各自所获得的知识产权及相应权益均归各自所有，不因共同申请本课题而改变。
3. 本协议所产生的所有成果的知识产权全部归属于委托方（甲方），受委托方（乙方）不得利用测试结果单独申报任何形式的成果。
4. 在履行本协议过程中各自向对方提供的相关信息,不构成向对方授予任何关于知识产权的许可行为。
5. 本合作协议不视为委托方给受委托方相关授权，未授权受委托方可使用委托方名义对外宣传或与第三方发生关系。

**第六条 保密条款**

1. 受委托方保证不向委托方以外的人员提供或披露本合同的委托内容及未公开的信息和资料，包括但不限于本协议的委托内容及结果，保密人员范围为乙方全体知悉或可能知悉的相关人员。保密期限自本合同生效之日起至约定的保密内容依法成为公开信息之日或双方确认的解密之日止。本保密条款不因本协议的无效、终止、解除、撤销而失去效力。
2. 双方保证采取一切合理和必要措施和方式对委托中知悉的对方商业秘密进行保密。

**第七条 承诺**

1. 如委托的任务涉及人类遗传资源采集、收集、买卖、出口、出境等，受委托方承诺遵照《人类遗传资源管理暂行办法》相关规定执行。
2. 如委托任务涉及动物实验，受委托方承诺自觉遵守《实验动物管理条例》，须具有动物生产和使用许可证，并严格选用符合要求的合格动物进行实验，保障动物福利。
3. 如委托任务的研究对象涉及人类受试者，受委托方承诺在签署协议前已经将委托任务的实施方案呈交受委托方伦理委员会讨论，并获得了伦理委员会批准。委托方也应在完成委托任务的过程中，自觉遵守国内外相关的医学伦理准则，保障保护受试者的安全和权益。
4. 在受委托方从事委托事项中发生的不可归责于委托方的人身、财产损害或侵权，由受委托方自行承担，委托方不承担责任。
5. 受委托方保证与委托方无直接经济利益关系，并保证委托关系及事项真实有效。
6. 受委托方保障委托内容的真实性和合法性，实验结果无虚假信息提供，确保可溯源。

**第八条 不可抗力**

1. 本协议所指不可抗力是指不能预见、不能避免并不能克服的客观情况，包括但不限于地震、火灾、水灾、战争、政府行为等。
2. 受委托方因不可抗力不能履行协议的，应当在不可抗力事件发生之日起七日内将不可抗力事由以书面方式通知委托方，并应当在合理期限内提供证明。
3. 因不可抗力不能履行本协议的，根据不可抗力的影响，部分或全部免除责任。受委托方延迟履行后发生不可抗力的，不能免除责任。

**第九条 违约责任**

1. 如无正当理由，委托方未能按期拨付工作经费，且经受委托方催促仍不能拨付或不能给出合理解释的，受委托方有权暂停履行受托任务。如委托方违约行为给受委托方造成损失的，委托方还应承担相应赔偿责任。
2. 如受委托方收到委托方拨款后，未能按时间和任务要求完成委托内容，或受委托方在完成委托工作时出现弄虚作假情况、不履行本协议或履行义务不符合要求的，委托方有权追回全部已拨经费。如受委托方违约行为造成委托方损失的，包括已发生的损失和可预见的损失，委托方有权要求赔偿并追究受委托方相关责任人员的法律责任。
3. 非因委托方违约或非因不可抗力，受委托方不能完成受托任务或受委托方逾期不能提交全部产出成果的，委托方有权解除本委托。委托解除后，受委托方应返还委托方已经拨付的项目经费。如受委托方的违约行为给委托方造成损失的，包括已发生的损失和可预见的损失，受委托方还应承担相应的赔偿责任。
4. 受委托方违反保密义务的，应当赔偿委托方所遭受的一切实际损失，委托方有权终止本协议并追回已经拨付的项目经费。

**第十条 协议的变更、终止及解除**

1. 本协议的变更应由双方协商一致后达成变更协议，并作为本协议的附件。
2. 本协议可由双方协商一致予以终止。

**第十一条 争议解决：**如在履行本协议的过程中发生争执，双方当事人应友好协商解决，如协商不成，任何一方可向委托方签署地（委托方所在地）有管辖权的人民法院提起诉讼。

**第十二条 其他约定事项（如无其他事项，请填“无”）**

**无**

**第十三条** 本协议一式四份，双方各执两份，具有同等法律效力。

与本协议约定事项有关的技术资料附件清单：**（如无其他附件，请填“无”）**

**无**

**第十四条 签字盖章页**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **委 托 方（甲 方）** | 单位名称 | 北京市耳鼻咽喉科研究所（北京市耳鼻咽喉头颈外科研究中心） （盖章) | | | |
| 法定代表人（授权人） | （签字） | 经办人  联系电话 | | 胡 旸 58265717  谢锦各 58265804 |
| 签订日期： | | | | |
|  | | | | | |
| **受委 托 方**  **（乙方）** | 单位名称 | 北京市计算中心有限公司  (盖章) | | | |
| 法定代表人（授权人） | （签字） | 经办人  联系电话 | 王雪峰  15001065280 | |
| 开户名称 | 北京市计算中心有限公司 | | | |
| 开户银行 | 宁波银行北京分行营业部 | | | |
| 银行账号 | 77010122001379612 | | | |
| 签订日期： | | | | |