**附件1**

**数据标准**

　　所有项目产生的数据须有完备的患者临床信息及匹配的高通量组学数据(数据类型见下文)。患者临床信息包括个体健康体检信息和疾病诊断、治疗、预后等随访信息。项目负责人必须依照指南要求标准提供数据。项目申请书须描述样本数量及年度数据产生计划，数据的样本数量与数据规范将被纳入评审和考核标准。

　　**一、样本收集要求**

　　肿瘤样本须有肿瘤组织和血液样本。样本须有准确的临床病理信息，包括诊断、治疗、预后等随访信息，须配备病理切片H&E染色结果。若有剩余样本，可冷冻保存以供后期实验验证。其他疾病如自身免疫病等疾病样本根据疾病特点和检测手段取样，具体要求可参考肿瘤样本。要求如下：

　　1. 样本采集标准

　　1.1 组织标本：组织标本应尽量去除坏死组织、血渍污物等可能影响检测结果的成分，根据不同的研究目的进行OCT包埋、福尔马林固定、液氮速冻、组织解离等初步处理，以上步骤应在标本离体60分钟内完成。肿瘤标本应包括无坏死的肿瘤组织，其他类型组织标本如淋巴结须选取有代表性的病变组织区域，组织标本的质量须至少满足基础型数据的实验需求；

　　1.2 外周血样本：按照不同的实验要求收集相应量的外周静脉血标本，采集量须满足基础型数据实验需求。依据实验需要对样本进行分离、冻存等处理，血浆或血清的提取须在30分钟内完成，-80℃保存。其他相应操作如免疫细胞分选等应在采样后12小时内完成。

　　2. 样本的质量控制

　　2.1 肿瘤组织：肿瘤组织质量不低于0.5克，肿瘤样品纯度不低于60%，肿瘤组织坏死率小于20%。肿瘤组织须有病理切片以准确判断肿瘤组织类型和组织学特征。其他组织要求参考肿瘤组织的质量控制标准；

　　2.2 血液样本：无凝血和溶血等情况，如进行免疫细胞分选，每例分离出的PBMC数量应大于8×106，T细胞数量大于3×106，B细胞数量大于1×106，细胞存活率大于95%。

　　3. 样本的运输：使用专用保存液、干冰或者液氮进行组织样本及外周血标本的运输。

　　4. 样品的保存：冻存于-80℃冰箱或者液氮中，须分装保存，避免反复冻融。

　　5. 样本必须提供详细的临床病理信息，具体要求如下：

　　5.1样本采集对象的可追溯唯一ID（如住院号等）；

　　5.2 样本采集对象的病理类型、病理分期（如TNM分期）、病理切片编号、诊断时间、实验室检查结果、影像学报告等；

　　5.3 样本采集时间、地点、采集方式（活检/手术切除等）、样本运输保存方式（冷冻组织、冰冻切片、石蜡切片等）、样本量（克、毫升）、样本纯度、组织坏死率；

　　5.4组织样本及血液样本所作检测名称、检测技术路径、检测数据报告。

　　**二、临床信息数据类型**

　　可追溯唯一识别ID、性别、年龄、民族、地区、血型、身高、体重、过敏史、恶性肿瘤家族史、免疫疾病家族史、感染史、吸烟史、饮酒史等；实验室检查（样本采集时）、影像学检查、病理检查报告、治疗情况和预后情况等，特别是随访信息和治疗响应情况。

　　**三、高通量组学数据类型**

　　1.外周血

　　TCR-Seq、BCR-Seq、单细胞多组学检测技术（scRNA-seq、scATAC-seq、scTCR-seq、scBCR-seq），鼓励获取TCRαβ链配对数据、BCR轻重链配对数据、以及对应抗原的匹配数据，鼓励通过CyTOF等方法获得淋巴细胞精细分群数据。

　　2.组织样本

　　TCR-Seq、BCR-Seq、WES（包括组织样本和对应个体外周血）、RNA-Seq、单细胞多组学检测技术（scRNA-seq、scATAC-seq、scTCR-seq、scBCR-seq），鼓励获取TCRαβ链配对数据、BCR轻重链配对数据、以及对应抗原的匹配数据，鼓励通过CyTOF等方法获得淋巴细胞精细分群数据。

　　**四、数据质控标准**

　　1.bulk TCR/BCR测序数据标准**：**外周血单样本clonotypes平均值不低于100,000种，组织样本clonotypes平均值不低于10,000种，重链和轻链各不少于1,000,000读数，单样本测序量不低于2Gbp （其中有效clonotype的定义为：能够组装出全长氨基酸序列且不包含终止密码子，具有明确的胚系V和J基因信息，同时CDR3区域含有典型的保守序列）。

　　2.bulk RNA-Seq测序标准**：**DV200不低于50%，检测覆盖不低于50M reads。

　　3.WES测序标准**：**单碱基覆盖率不低于100X，检测盒覆盖不低于32MB。

　　4.单细胞检测技术指标：

　　4.1 scRNA-seq：测序指标（有效barcodes不低于75%，有效UMIs不低于75%，RNA read Q30不低于75%）；细胞指标（单个样本细胞数不少于5000个，平均每个细胞的测序read pairs数大于20000，单个细胞基因中位数大于1000）；

　　4.2 scTCR-seq/scBCR-seq：平均每个细胞所测得的VDJ文库read pairs数不低于5000，RNA read Q30 不低于75%，有效barcodes不低于75%；细胞指标（3000-80000个细胞/芯片）；

　　4.3 scATAC-seq: 测序指标（测序read pairs不小于25000，有效barcodes不低于75% ，barcode Q30大于65%，read1/2 Q30大于65%，样本指数的Q30大于90%）；细胞指标（单个样本细胞数不少于5000个，高质量测序片段大于40%）。

　　**五、其他要求**

　　1. 申请人须严格遵守《中华人民共和国生物安全法》《中华人民共和国人类遗传资源管理条例》等有关规定，并严格遵守医学伦理和患者知情同意等规范要求。涉及人类遗传资源研究的项目申请书中应提供所在单位或上级主管单位伦理委员会的审核证明（电子申请书应附扫描件），否则将不予受理。

　　2. 获资助的项目须按要求进行相关科学数据采集，并将原始数据每半年上传至本计划指定的数据平台（https://ngdc.cncb.ac.cn/idp/）。