对扩增的sgRNA文库进行下一代测序以确定sgRNA分布

时间 3-5天

32 NGS文库PCR。我们提供了使用Illumina接头序列扩增sgRNA靶区域的NGS引物（表3）。要为NGS准备sgRNA文库，请为10个NGS-Lib-Fwd引物和1个NGS-Lib-KO-Rev或NGS-Lib-SAM-Rev条形码引物中的每一个设置一个反应，如下所示：

**表3 扩增sgRNA文库和NGS的引物序列。**

10种不同的NGS-Lib-Fwd引物各自在Illumina Read 1测序引物后含有1-10个额外的核苷酸，旨在增加NGS文库的多样性。扩增NGS的sgRNA文库时需要所有10种NGS-Lib-Fwd引物。8种NGS-Lib-Rev引物提供独特的条形码（粗体）用于区分混合测序运行中的不同sgRNA文库（例如，4个筛选生物代表的实验和控制条件），所需的NGS-Lib-Rev引物数量取决于混合运行中要一起测序的不同sgRNA文库的数量。由于GeCKO和SAM文库之间的sgRNA骨架不同，因此为每个文库提供了单独的NGS-Lib-KO-Rev和NGS-Lib-SAM-Rev引物。

**表20**

| **Component** | **Amount per reaction (μl)** | **Final concentration** |
| --- | --- | --- |
| NEBNext High Fidelity PCR Master Mix, 2× | 25 | 1× |
| Pooled sgRNA library template from Step 31 | 1 | 0.4 ng μl−1 |
| NGS-Lib-Fwd primer (unique) | 1.25 | 0.25 μM |
| NGS-Lib-KO-Rev or NGS-Lib-SAM-Rev primer (barcode) | 1.25 | 0.25 μM |
| UltraPure water | 21.5 |  |
| Total | 50 |  |

**关键一步**

为每个文库使用具有唯一条形码的不同反向引物，可以在单个NextSeq或HiSeq运行中汇集和测序不同的文库。这比在多个MiSeq运行中运行相同数量的文库更高效且更具成本效益。

**关键一步**

为了最大限度地减少扩增sgRNA的错误，使用高保真聚合酶（如NEBNext）非常重要。其他高保真聚合酶，如PfuUltra II（安捷伦）或Kapa HiFi（Kapa Biosystems），可用作替代品。

33 使用以下循环条件进行PCR：

**表 21**

| **循环次数** | **变性** | **退火** | **延长** |
| --- | --- | --- | --- |
| 1 | 98℃，3分钟 |  |  |
| 2–23 | 98℃，10秒 | 63℃，10秒 | 72℃，25秒 |
| 24 |  |  | 72℃，2分钟 |

34 反应完成后，汇集PCR反应产物，并根据制造商的说明使用QIAquick PCR 纯化试剂盒纯化PCR产物。

35 定量纯化的PCR产物，并在2% ( wt /vol)琼脂糖凝胶上电泳2 μg产物。成功的反应应产生约260至270 bp的产物用于敲除文库，约270至280 bp的产物用于激活文库。根据制造商的说明，使用QIAquick凝胶提取试剂盒进行凝胶提取。

**暂停点**

凝胶提取的样品可以在−20°C 下保存数月。

36 根据制造商的说明，使用Qubit dsDNA HS检测试剂盒对凝胶提取样品进行定量分析。

根据Illumina用户手册在Illumina MiSeq或NextSeq上对样本进行测序，读取1（正向）循环80次，索引1循环8次。我们建议在MiSeq上使用5% PhiX对照或在NextSeq上使用20% PhiX对照进行测序，以提高文库多样性；我们建议以文库中每个sgRNA的覆盖率>100次读取为目标。

38使用count\_spacers.py分析测序数据。我们提供了一个Python脚本，用于分析sgRNA表示的NGS结果。安装 Python 2.7和biopython 。准备一个包含引导间隔序列的.csv文件，每行对应一个序列。

39 要确定间隔物分布，请使用以下可选参数运行 python count\_spacers.py：

**表 22**

| **Flag** | **Description** | **Default** |
| --- | --- | --- |
| '-f' | .fastq file containing NGS data for analysis | NGS.fastq |
| '-o' | Output .csv file with guide spacer sequences in the first column and respective read counts in the second column | library\_count.csv |
| '-i'′ | Input .csv file with guide spacer sequences | library\_sequences.csv |
| '-no-g' | Indicates the absence of a guanine before the guide spacer sequence | Guanine is present |

运行 count\_spacers.py 后，间隔读取计数将写入输出.csv文件。相关统计数据（包括完美匹配的指导序列数量、不完美匹配的指导序列数量、没有密钥的测序读取数量、处理的读取数量、完美匹配的指导序列百分比、未检测到的指导序列百分比和倾斜率）将写入 statistics.txt。理想的sgRNA文库应具有>70%完美匹配的指导序列、<0.5%未检测到的指导序列和小于10的倾斜率。

**关键一步**

人类SAM文库在引导间隔序列前没有鸟嘌呤，因此在分析这些文库时请确保使用参数-no-g运行脚本。