**SuperDecode v1.0使用手册**

华南农业大学农学院，刘耀光实验室

2024年7月

SuperDecode是一款主要用于基因编辑靶点突变分析的本地版软件包，其包含3个模块，DsDecodeMS、HiDecode以及LaDecode，分别对应通过Sanger测序、二代测序（Next-generation sequencing, NGS）以及三代测序（Third-generation sequencing, TGS）数据，对包含编辑靶点的扩增子进行分析解码。DsDecodeMS为我们的网页版工具DsDecodeM (Xie et al., 2017)的基础上升级的本地版，可以直接读取靶点扩增子的Sanger测序文件的峰图，分析突变序列和类型，包括双等位、杂合突变等。HiDecode是对在靶点扩增子两端添加一对样本barcode的混合文库的二代测序文件进行解码分析的软件，可实现对多种类型的样本进行高通量的突变分析，包括二倍体和多倍体、细胞系，原生质体，以及嵌合突变等。LaDecode是利用单分子测序技术对目标DNA区段内包含多个靶点的扩增子进行高通量分析的工具，从而确定其突变序列类型（包括复杂结构变异），获得每个突变序列的单倍型。从时效和费用考虑，建议对较少样品采用Sanger测序和DsDecodeMS分析；每批较多样品（接近或多于96样品）采用NGS和HiDecode分析。

**一、软件下载和安装**

1、软件下载：

我们提供了Windows、MacOS的界面版本，以及Linux命令行版本，软件包已上传至GitHub（<https://github.com/xiexr/SuperDecode>）。此外，考虑到国内访问GitHub的网络受到限制，我们也将软件包上传到了奶牛快传（高速的数据传输工具）。

Windows版本的奶牛快传链接：

MacOS版本的奶牛快船链接：

Linux版本的奶牛快船链接：

此外，也可以通过TBtools-II的插件（Plugin）仓库直接使用SuperDecode。

2、软件安装：

软件为可以直接运行的程序，无需安装。下载后，将软件解压至本地磁盘（注意：**软件包所在路径不包含空格**）。对于界面版，进入SuperDecode文件夹后，直接双击“SuperDecode.exe”即可以打开软件。为便于使用，可以将“SuperDecode.exe”发送快捷方式到桌面。MacOS版本在第一次打开时需要加载相关库，时间会长些，后面再次打开就不会了。对于Linux版本，进入解压后（解码命令：tar -zxf superdecode\_linux.tar.gz）的文件夹后，通过直接执行相关模块的命令运行各模块（文件夹内提供了参考的脚本和运行测试数据的命令，用户可以参考测试文件的格式建立自己的输入文件）。

3、测试数据下载

SuperDecode为每个模块提供了练习测试文件数据，均保存于SuperDecode路径下的example文件夹中。为方便用户练习测试，进入每个模块的主界面后，可通过菜单栏“Help”下拉框中的“Load example files”加载测试数据。

**二、DSDecodeMS的扩增子样本制备和分析**

DSDecodeMS通过直接读取靶点扩增子的Sanger测序（ab1）文件峰图，分析靶点的突变序列和类型，主要用于少量二倍体样本中简单突变类型（纯合、杂合和双等位突变）的检测。

**1、靶位点的扩增**

因Sanger测序读长限制，目的扩增片段大小在700 bp以内，在靶点前后大约200～350 bp的位置设计特异性扩增引物T#-F以及T#-R（长度大约为18~21 nt，Tm值（GC% x 41 + 69.3）- 650/L（L引物碱基数）为58℃~60℃。使用NCBI的blast工具检测引物在目的参考基因组是否特异，特别是引物3’端的特异性）。扩增片段内可以包含1个或2个靶点。如果目的片段在基因组中存在同源性较高的序列，可以适当延长扩增片段，获得特异的扩增子。推荐的扩增体系如表1-1。

表1-1 DSDecodeMS样本扩增子PCR反应体系（25 μL）

|  |  |
| --- | --- |
| 2 × Taq mix (普通Taq) | 12.5 μL |
| T#-F (10 μM) | 0.5 μL (终浓度0.2 M) |
| T#-R（10 μM） | 0.5 μL |
| gDNA | 20~30 ng |
| ddH2O | 补足到25 μL |

PCR反应程序如表1-2所示。

表1-2 DSDecodeMS样本扩增子PCR反应程序

|  |  |  |
| --- | --- | --- |
| 预变性 | 94 ℃, 3 min |  |
| 变性 | 96℃, 15 s | 30~32 cycles |
| 退火 | 58 ℃, 15 s |
| 延伸 | 65℃10S, 68℃10S, 72 ℃ 10S min, 65℃10S （如果目的片段>1 kb，每节温度加5 s） |
| 终延伸 | 72 ℃, 5 min |  |

（注：延伸采用变温方式，以改善不同GC含量分布序列的扩增效率）

取3~4 μL的扩增产物进行琼脂糖凝胶（1%）电泳检测样本，如果样本浓度较低，可适当增加2~3个循环增加产物浓度。

扩增产物（~20 μL）直接送公司进行Sanger测序（国内多数公司Sanger测序包含PCR产物纯化服务）；也可以自己以琼脂糖凝胶电泳，切胶回收目的片段，然后使用凝胶回收试剂盒纯化回收产物，将回收产物送公司进行Sanger测序。为获得高质量的Sanger测序峰图，推荐在靶点上游或下游150～300 bp处设计内部引物进行测序，不建议使用PCR扩增用过的引物测序。如果扩增片段内包含2个靶点，在靶点1的上游和靶点2的下游分别设计测序引物进行Sanger测序。

**2、使用DSDecodeMS对Sanger测序峰图解码**

使用SuperDecode中的DSDecodeMS模块对Sanger测序获得的ab1文件进行突变检测。

1. 软件主界面介绍

DSDecodeMS的主界面主要包括：参数设置、参考序列输入、测序文件输入、峰图查看和解码结果查看模块（图1-1）。其中，参数设置模块主要包括菜单栏、打开、退出、窗口设置、截图工具、任务运行、测序荧光信号阈值的设置、序列反向互补、碱基窗口等；序列输入模块用于输入目标序列的野生型参考序列、靶位点序列、ab1测序文件输入；峰图查看模块用于实时显示所选阿布文件的峰图，便于检查峰图的质量；解码结果查看模块用于显示解码结果。

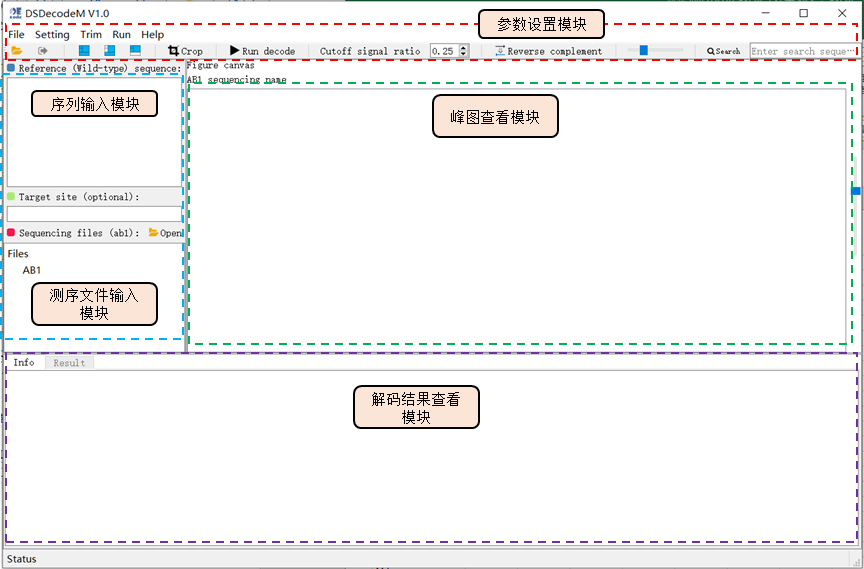


图1-1 DSDecodeMS主界面

（2）利用DSDecodeMS进行解码的操作

DSDecodeMS的使用主要分为3步，分别为参考序列以及测序文件的输参数设置以及解码运行，即可实现快速的检测样本突变类型，其详细操作步骤如下：

i）在序列输入模块框中粘贴输入扩增片段对应的野生型参考序列，即T#-F引物到T#-R引物之间的序列，输入类型为普通文本或Fasta格式的序列（与测序链方向相同或反向互补）。然后，可选地输入靶点序列（包含或者不包含PAM均可以），用于程序快速定位靶点序列附近的突变；一般情况下（测序质量好）可以不用输入靶点序列。点击“Open”按钮，从文件夹选择输入测序ab1文件，可以一批输入不限数目的对应同一个参考序列相同靶点的多个ab1文件。通过双击各测序文件名，可以在右侧查看峰图；建议在启动分析任务前逐个检查峰图质量。

ii）一般情况下，用软件预设的参数设置可以产生正确的分析结果，不需要做参数设置调整。如果某些质量稍差的测序文件不能产生正确的结果，可在参数设置模块中调整测序荧光信号阈值（默认值0.25），还可以通过菜单栏的“Setting”设置锚定序列AS、简并序列DS、和信号阈值后，再启动分析。

iii）点击“Run decode”按钮启动突变分析任务。

程序运行结束后，每个样本的突变序列和类型显示在底部的解码结果模块中。可以通过鼠标选择相应结果后，复制到Word或PPT中；也可以点击菜单栏“Run”下拉框中的“Save results to a txt file”或结果查看栏的右上角“Save”按钮保存结果至本地txt文件夹。

（3）其它功能

i）修剪峰图中低质量的两端

Sanger测序峰图在开始（约30～60 bp）和结束位置的质量一般较差，如果靶点序列距离峰图两端较近，会对分析产生影响。DSDecodeMS提供了去除峰图中低质量两端序列的功能。通过点击菜单栏“Trim”选择修剪方法，包括基于Richard Mott算法自动修剪和用户自定义修剪范围（图1-2），去除低质量的两端序列。

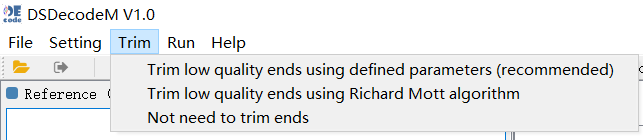


图1-2 DSDecodeMS修剪序列两端的方法

ii）截图功能

SuperDecode的每个模块都提供了截图功能，通过点击快速工具栏的“Crop”按钮，可以打开截图功能。拖动鼠标左键选择母本截图区域后，双击复制截图或者右键保存截图到指定的文件夹。

iii）子序列搜索功能

在快速工具栏的“Search”框输入一段需要查找的序列，按Enter键后，程序会在打开的峰图中自动搜索该序列。该功能支持简并序列的搜索。

iv）从ab1文件导出fasta格式的序列文件

通过点击菜单栏“Run”下拉框的“Export fasta”可以将每个ab1文件的序列输出为fasta格式的本地文件。

（4）解码失败的原因和解决办法

i）PCR产物可能含有微量的非特异产物，重复使用了扩增引物进行测序，可导致测序峰图质量较差。推荐在扩增序列内部针对每个靶点设计1条测序引物进行测序。

ii）目标扩增序列在基因组中为多拷贝重复序列，产物特异性很差，导致Sanger测序时产生大量的乱峰。重新设置目标序列拷贝的高特异性引物（可以往靶点两外侧延伸一定距离设计高特异的引物）。还可以设计2组巢式特异引物做2轮巢式PCR扩增提高产物的特异性。

iii）样本的突变为嵌合突变等复杂突变，或者是该区间存在多个靶点，靶点之间发生了片段缺失或倒位等结构变异。如果是嵌合突变，使用NGS和HiDecode解码。如果是多个靶点间的结构变异，使用TGS和LaDecode解码。

iv）输入的野生型参考序列与样本Sanger测序的结果序列不相符，无法将测序结果定位于参考序列中。检查输入的参考序列是否对应目标扩增序列。

v）在测序引物与靶点之间含有高GC区间、或者较长的polyA(T)、或较长的微卫星等的区域，导致在该区域出现假的套峰，影响解码。在不含此类区域的靶点另一侧设计测序引物，重新测序。

**三、HiDecode混合扩增子文库制备和分析**

当个人或团队的检测样本总量较大（>50）时，推荐使用NGS和HiDecode，可以降低检测费用。此方法尤其适用于二倍体的嵌合突变、多倍体、细胞系、原生质体或愈伤组织等样品的单靶点或者双靶点且靶点距离相近（150 bp以内）的突变检测。

**1、HiDecode混合扩增子文库构建**

HiDecode扩增子文库构建主要通过两轮PCR反应，在目的片段的扩增子两端添加一对特异的barcode序列，再混合为一个测序文库进行NGS（图2-1）。



图2-1 HiDecode短片段NGS文库的构建流程

（1）HiDecode引物设计

HiDecode文库构建涉及3对引物，包括：用于扩增靶序列的位点特异引物T#-F和T#-R，用于区分96孔PCR板的plateN-R和不同孔位置的posN-F，用于添加文库接头的引物Lib-F和Lib-R。构建完成的NGS文库序列构成如图2-2。



图2-2 NGS文库的序列构成

i）位点特异引物T#-F和T#-R

考虑到目前常用的NGS双端测序读长为150 bp，两端总长约为300 bp，因此，设计第一轮特异扩增的片段大小约200～260 bp（含引物两端的桥接序列共32 bp）。在检查特异引物T#-F和T#-R的位点特异配对部分的特异性后，在它们的5’端添加一段16 nt桥接序列。第一轮扩增的靶点特异性引物为：

靶序列特异性正向引物T#-F：5’-ctcggagtgatcgcacNNNNNN…NNNNN-3’（17～19 个N为基因组位点特异配对区，Tm = 56℃~58℃；5’端小写16 nt为与posN-F配对的接桥序列）;

靶序列特异反向引物T#-R：5’-ctgagaggctggatggNNNNNN…NNNNN-3’（17～19个N为基因组位点特异配对区, Tm = 56℃~58℃；5’端小写16 nt为与plateN-R配对的接桥序列）。

ii）通用引物posN-F和plateN-R

posN-F从5’到3’端含有与文库接头匹配的序列、指示各样品对应PCR板上96孔位置（A1~H12）的barcode碱基 (6 nt)、以及结合第一轮产物5’端16 nt桥接的序列，共96条；plateN-R含有与文库接头匹配的序列、指示PCR板编号的barcode碱基、以及结合第一轮产物3’端16 nt桥接的序列，一块PCR板使用1条plateN-R。用户可根据实验室实际需求合成plateN-R条数，本方法提供了96条plateN-R供选择。HiDecode中内置的posN-F和plateN-R为（具体序列见附表1-HiDecode所有引物序列.xlsx）：

posN-F：5’-CGCTCTTCCGATCTGTCANNNNNNctcggagtgatcgcac-3’（下划线序列为与文库引物Lib-F配对的接头序列，6 个N为区分PCR板上96孔位置（A1~H12）的特异性barcode序列，小写16 nt为结合第一轮产物正向桥接序列的碱基）。

plateN-R：5’-CAGACGTGTGCTCTTCCGATCTCTGTNNNNNNctgagaggctggatgg-3’（下划线序列为与文库引物Lib-R配对的接头序列，6 nt的N为区分不同PCR板的特异性barcode序列，小写16 nt为结合第一轮产物反向互补桥接序列的碱基）。

建议提前将96条posN-F（1.0 μM）分装20 μL至96孔PCR板，方便使用96孔针取样器或者多通道移液器取样，-20℃保存备用。

iii）文库接头引物Lib-F和Lib-R

NGS测序接头可以通过PCR方法引入扩增子中。接头中含有实验室(Laboratory)特异barcode（命名为Lab-barcode），用于标识不同测序用户（实验室）的文库。测序公司根据该Lab barcode将多用户混合测序的原始测序数据进行拆分给各客户。多数公司支持单Lab barcode或者双Lab barcode对数据进行拆分，如果是单Lab barcode，接头序列如下：

Lib-F：5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCA-3’（下划线序列为与posN-F配对的接头序列）

Lib-R：5’-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTT

CAGACGTGTGCTCTTCC-3’（下划线序列为与plateN-R引物配对的接头序列，8 N序列为Lab barcode；填写测序送样单时，填写的barcode序列需将Lib-R中的Lab barcode写成**反向互补序列即正链的序列**）。相同实验室（团队）的不同批次测序文库可以使用同一个Lab barcode，但同一批次有多个独立测序文库的情况下（公司可能将这些文库混合成同一批测序大样品），各文库应该使用不同的Lab barcode（如果使用了相同Lab barcode，向公司说明情况）。

如果使用双Lab barcode，应根据测序公司建议修改Lib-F和Lib-R的序列。

注意：此处为根据illumina接头设计，如果使用其它测序平台，需要依据对应测序平台的接头序列设计posN-F、plateN-R、和Lib-F、Lib-R。

（2）扩增步骤

1. 模板DNA工作液和各种引物的配制

所有引物和基因组DNA的工作溶液用0.3x TE配制（不要用1x TE，会带入过多EDTA抑制酶活性）。

特异引物T#-F, T#-R, plateN-R：稀释为10 M保存液，保存待用。

posN-F合成和工作液：合成96种posN-F（posA1-F~ posH12-F）,各5 OD，1 OD/管。每种引物其中1管以1x TE溶解成10 M保存液，其余的以干粉-20℃冷冻长期储藏。将每种posN-F（10 M）取2.0 L加入到一块96孔PCR板对应的孔（A1~H12），每孔加入18 L 0.3x TE，即为1.0 M posN-F工作液，冷冻保存（以下使用到剩余4~5L以下时丢弃）。

模板DNA工作液：取每样品模板DNA约200 ng加入到96孔PCR板对应的孔（A1~H12），每孔加入0.3x TE到一共10L（即20 ng/L）。冷冻保存。

ii）第一轮特异性扩增

HiDecode文库构建包含两步PCR扩增，第一步PCR扩增使用靶序列特异性引物T#F和T#R进行目标片段的扩增，根据样本数配制PCR总混合液，将PCR反应总混合液分装至96孔PCR板中，PCR反应总混合液体系配制如表2-1所示（样本数为*n*）。每孔分别加入15 μL反应液。

表2-1 HiDecode第一轮PCR反应总混合液体系（每反应15 μL）

|  |  |
| --- | --- |
| 2 × Taq mix (普通Taq) | 7.5 μL × *n* |
| T#-F (10 μM) | 0.3 μL × *n*（终浓度0.2 μM） |
| T#-R (10 μM) | 0.3 μL × *n*（终浓度0.2 μM） |
| ddH2O | 7.5 μL × *n* |

使用96孔针取样器（或者多通道移液器）点取~1μL的DNA到分装了PCR反应液的孔中。PCR反应程序如表2-2。

表2-2 HiDecode样本建库第一轮PCR反应程序

|  |  |  |
| --- | --- | --- |
| 预变性 | 94 ℃, 3 min |  |
| 变性 | 96℃, 10 s | 10 cycles |
| 退火 | 58℃, 20 s |
| 延伸 | 65℃, 5 s |
| 延伸 | 72℃, 10 s |
| 延伸 | 65℃, 5 s |
| 变性 | 96℃, 10 s | 18~20 cycles |
| 退火/延伸 | 65℃, 15 s |
| 延伸 | 72℃, 10 s |
| 延伸 | 65℃, 5 s |
| 终延伸 | 68℃, 1 min | 1 cycle |

（注：延伸采用变温方式，以改善不同GC含量分布序列的扩增效率）

如果目的片段在基因组中存在相似度较高的序列导致T#-F以及T#-R无法特异性扩增目的片段，可以靶点上下游更大范围内设计第一组高特异引物exT#-F和exT#-R（Tm = ~60℃）。先进行巢式第一轮PCR，获得高特异的扩增产物。PCR反应体系如表2-3所示。

表2-3 HiDecode样本建库巢式第一轮PCR反应体系（每反应15 μL）

|  |  |
| --- | --- |
| 2 × Taq mix （普通Taq） | 7.5 μL × *n* |
| exT#-F （10 μM） | 0.15 μL × *n*（终浓度0.1 μM） |
| exT#-R （10 μM） | 0.15 μL × *n*（终浓度0.1 μM） |
| ddH2O | 7.5 μL × *n* |

PCR反应扩增程序如表2-4所示：

表2-4 HiDecode样本建库巢式PCR第一轮反应程序

|  |  |  |
| --- | --- | --- |
| 预变性 | 93℃, 3 min |  |
| 变性 | 96℃, 10 s | 15 cycles |
| 退火 | 60℃, 20 s |
| 延伸 | 65℃, 10 s |
| 延伸 | 72℃, 30 s |
| 延伸 | 65℃, 10 s |

将T#-F以及T#-R引物(原10 μM)等量混合，再以ddH2O稀释成各3 μM（共> n μL）。以上PCR反应结束后，各孔加入1 μL的3 μM T#-F/T#-R(终浓度为0.2 μM)。继续进行以下PCR扩增反应，获得目的片段的特异性扩增产物。PCR反应程序如表2-5所示。

表2-5 HiDecode样本建库巢式PCR第二轮反应程序

|  |  |  |
| --- | --- | --- |
| 变性 | 96℃, 10 s | 17-18 cycles |
| 退火/延伸 | 58℃, 20 s |
| 延伸 | 65℃, 5 s |
| 延伸 | 72℃, 10 s |
| 延伸 | 65℃, 5 s |
| 终延伸 | 68℃, 1min |  |

以上扩增产物片段大小为200~260 bp。每块板可随机取几个样本的扩增产物进行琼脂糖凝胶（1.5%）电泳进行检测。

iii）第二轮barcoding扩增

第二轮PCR扩增反应的目的是引入barcode序列以及添加测序接头。根据样本数配制第二轮PCR总混合液，将PCR反应总混合液分装至96孔PCR板。PCR反应总混合液体系如表2-6（样本数为*n*）：

表2-6 HiDecode样本建库第二轮PCR反应体系（每样品15 μL）

|  |  |
| --- | --- |
| 2 × Taq mix | 7.5 μL × *n* |
| plateN-R（10 μM） | 0.08 μL × *n* (终浓度0.05 μM) |
| Lib-F, Lib-R（10 μM） | 各0.3 μL × *n* (终浓度分别0.2 μM) |
| ddH2O | 7.5 μL × *n* |

使用96孔针取样器或者多通道移液器，分别点取~1μL稀释分装在96孔板的1.0 M posN-F，加入到各PCR孔反应液中（每种posN-F终浓度约0.06M）。

将第一轮PCR扩增产物加入30 μL ddH2O稀释3倍，使用96孔针取样器或者多通道移液器，分别点取~1μL稀释产物加入到各PCR孔反应液中（产物总稀释约45x）。

PCR反应程序如表2-7。

表2-7 HiDecode样本建库第二轮PCR反应体系

|  |  |  |
| --- | --- | --- |
| 变性 | 96℃, 10s | 8 cycles |
| 退火 | 55℃, 15 s |
| 延伸 | 65℃, 5 s |
| 延伸 | 68℃, 5 s |
| 延伸 | 72℃, 5 s |
| 变性 | 96℃, 10 s | 10~12 cycles  （可增减调整） |
| 退火/延伸 | 65℃, 15 s |
| 延伸 | 68℃, 5 s |
| 延伸 | 72℃, 5 s |
| 终延伸 | 68℃, 1 min | 1 cycle |

第二轮扩增后的文库片段长度约为340~400 bp（含接头和barcode等）。每板可随机取几个样本PCR产物进行琼脂糖凝胶（1.3%）电泳进行检测。如果扩增产物浓度较低，可增加2～3个循环，提高产物浓度。

将每块96孔板的建库产物为1组各取出3 μL混合（即每组样近300 μL），使用1.3%琼脂糖凝胶电泳，切胶回收目的片段，使用DNA纯化试剂盒纯化产物。检测每组纯化产物的浓度后，将各组样本产物再等量混合（不足96样品的组按比例减少混合量）。总混合产物终浓度至少为50 ng/μL，产物体积至少为60 μL。

**2、NGS测序**

将混合后的建库样本送测序服务公司，通过Illumina、X-Ten或Nova-seq等平台进行测序（如果使用其它平台，需要根据平台提供的接头序列，重新设计posN-F和plateN-R、Lib-F和Lib-R）。由于测序公司要求的最低测序量为1 Gb，推荐1～10块PCR板的混合产物测序量为1 Gb。按产物长度约350 bp估算，平均每个样品的深度为30000x ~ 3000x。

多数测序公司（如武汉康圣达医学检验所、海普洛斯、吉因加科技）的服务可以在送样后一周内交付测序数据。

**3、使用HiDecode软件分析突变**

（1）软件主界面介绍

HiDecode主界面包括：菜单栏、参数设置、文件输入、Barcode设置、孔板信息查看以及解码结果显示模块（图2-3）。其中，菜单栏提供了常用的设置功能、插件等，在快速菜单栏设置突变位点检出阈值、CPU线程数等。文件输入模块用于输入野生型参考序列文件、靶点序列文件以及测序文件。Barcode设置中选择使用的孔和板。样本孔板信息查看模块主要用于交互式显示各样本的分析结果。

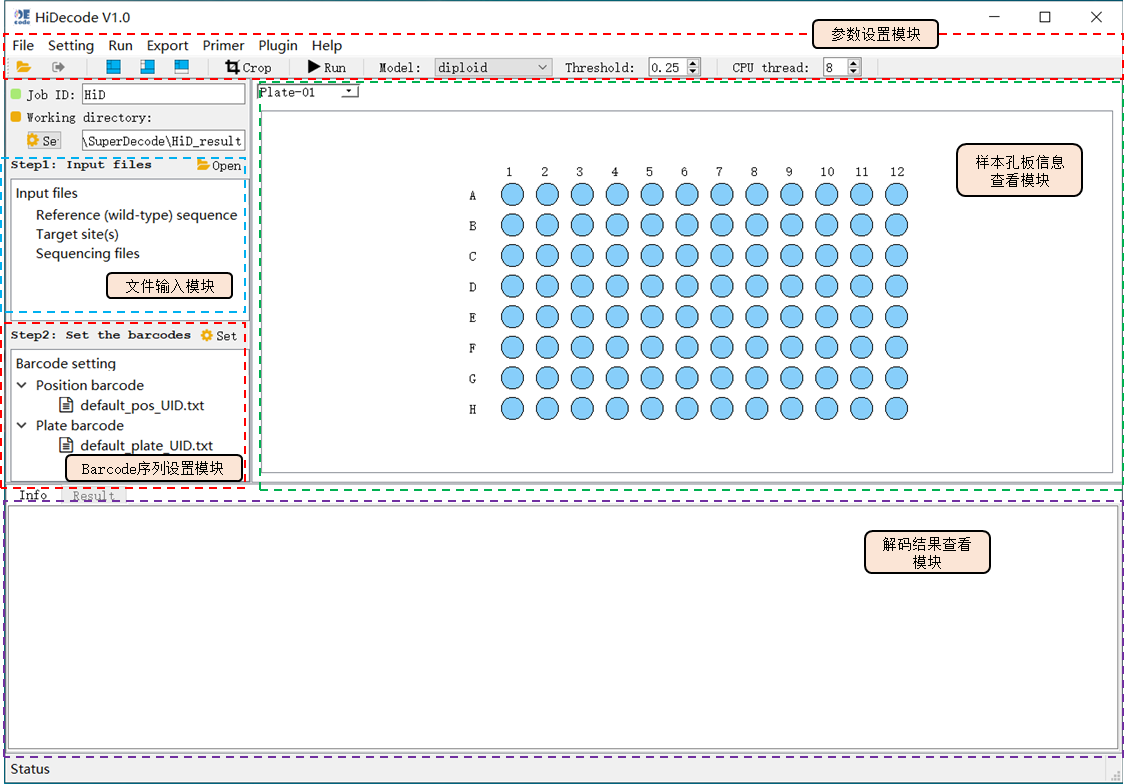


图2-3 HiDecode主界面

（2）基本分析操作

使用HiDecode软件从NGS混合扩增子文库分析靶点突变序列的主要步骤包括参考序列输入、测序文件输入、参数设置和解码。详细操作如下：

i）通过点击菜单栏或者左侧文件输入区的“File”按钮，打开文件输入对话框（图2-4），依次输入：Fasta格式的参考序列文件（覆盖T#-F到T#-R之间的野生型基因组参考序列）、可选的靶点序列文件（包含或者不包含PAM序列均可以，也可不输入靶点序列文件）、以及测序文件（双端reads文件，要求为过滤且去除接头后的Clean data）。参考序列、靶点序列文件的格式可以通过点击输入框下方的提示查看。

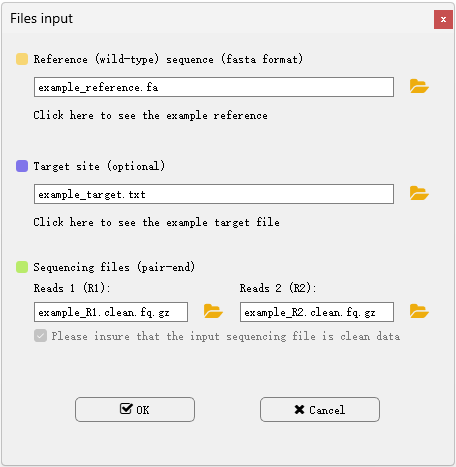


图2-4 HiDecode文件输入界面

ii）通过双击左侧的“default\_pos\_UID.txt”和“default\_plate\_UID.txt”设置扩增时使用的孔位置barcode和板编号barcode。打开设置框后，勾选使用的barcode。

iii）在HiDecode工具栏中设置数据分析对应的参数。程序提供了预设的3种突变检出阈值Threshold，分别为diploid（用于分析二倍体样本数据）、polyploid（用于分析多倍体样本数据）、以及low frequency（用于分析细胞系、原生质体、愈伤组织等样本数据）。除了预设的Threshold，用户可以根据需要自行调整Threshold。当样本中某变异位点的reads数占比大于设置的阈值时，程序才认为该基因型为真实的突变，从而输出该突变位点。

iv）点击“Run”按钮，启动分析任务。

（3）分析结果导出和结果解读

程序运行结束后，样本孔板信息模块中会显示每个孔的分析结果，黄色表示无结果，蓝色表示有结果，红色为当前选中显示的样本（图2-5）。选中某个样本时，底部的解码查看模块中会显示该样本的突变情况，包括对应的突变类型、覆盖的总reads数、参考序列、每条突变序列及其支持的reads数。通过拉动上方的滑动条，可以调整需要显示的序列范围。

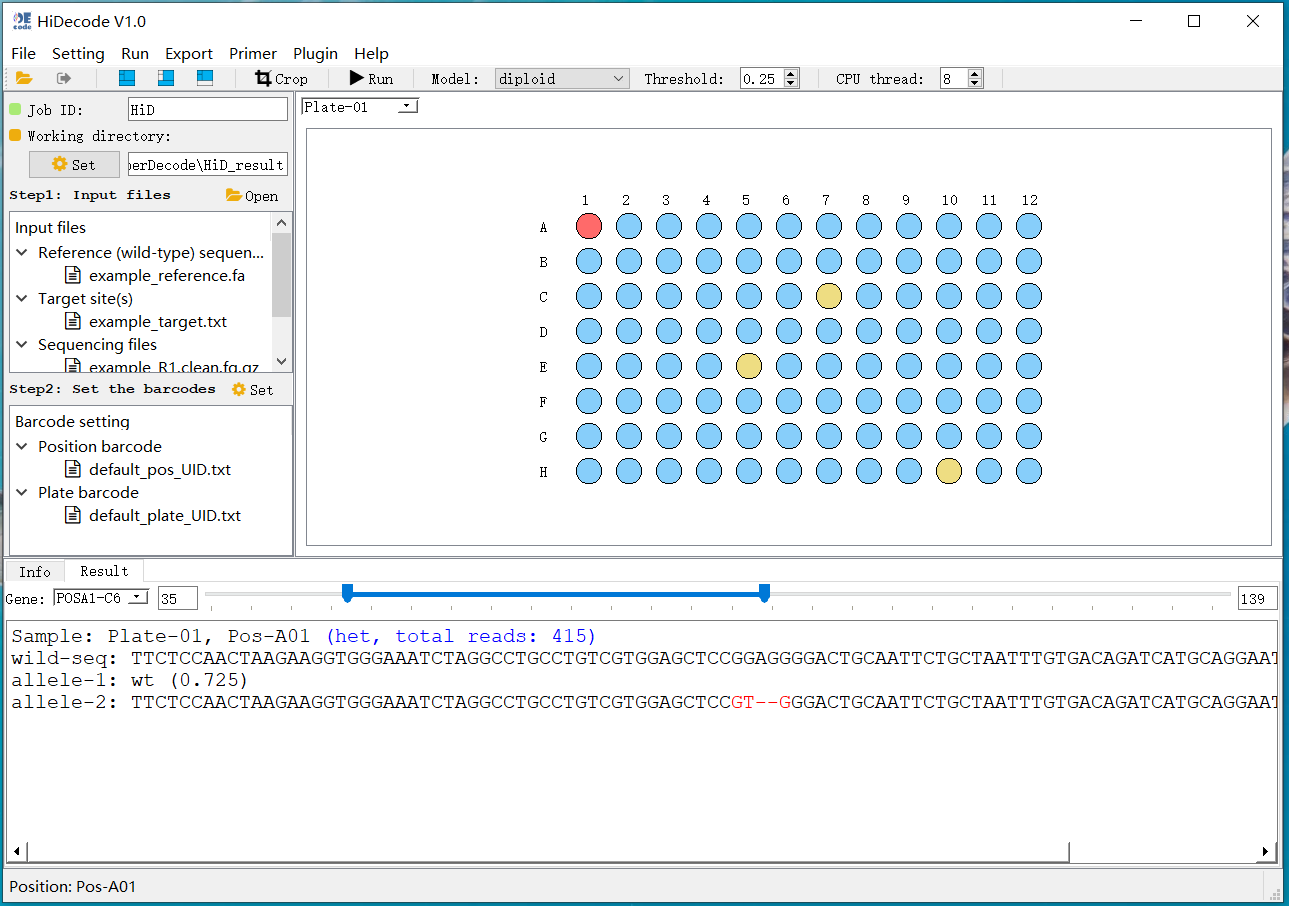


图2-5 HiDecode结果分析

解码后的结果可点击菜单栏的“Export”将结果输出至excel文件，可以输出两种类型的结果文件。“Export all results to excel file”为输出所有孔和板的解码详细结果（包括reads数、突变碱基、每条突变序列和野生型序列的比对等）（图2-6），“Fetch mutations to a single file”为仅输出含有数据的样本信息、突变碱基及其reads数。

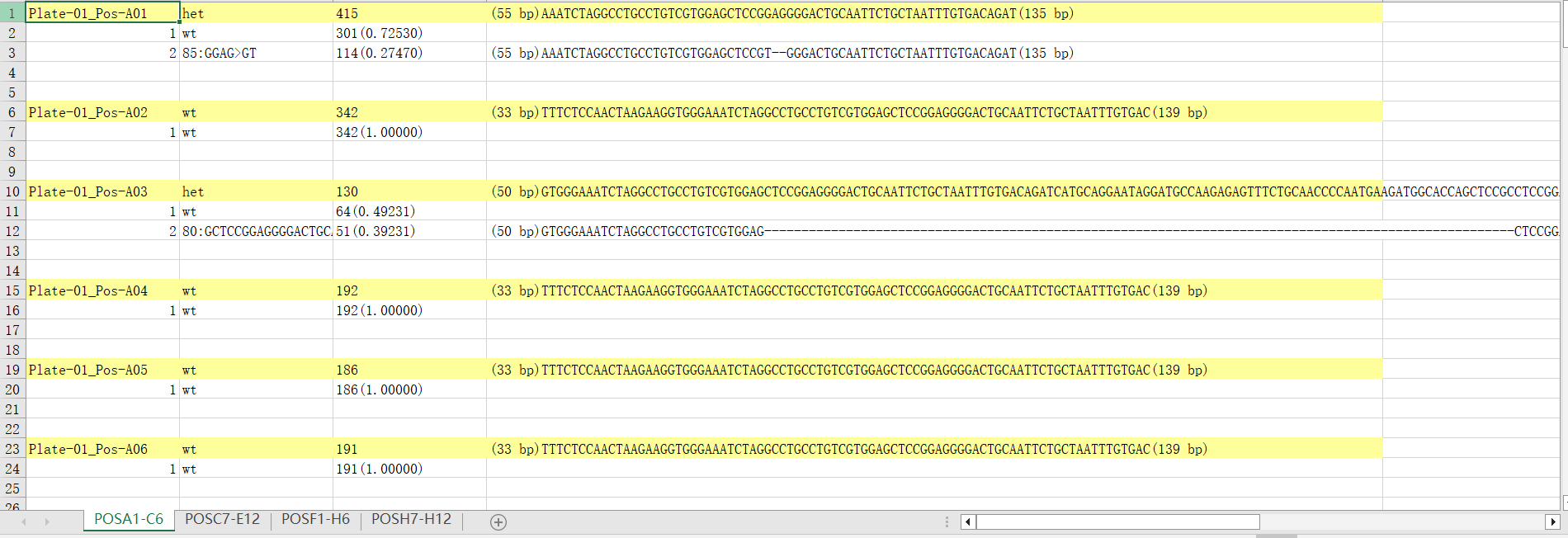


图2-6 基于HiDecode解码输出的详细突变结果表

另外，“Export”中的“Plot mutations of all samples”可以将不同样本的基因型以图形的形式输出，方便用户直观查看样本的突变结果（图2-7）。

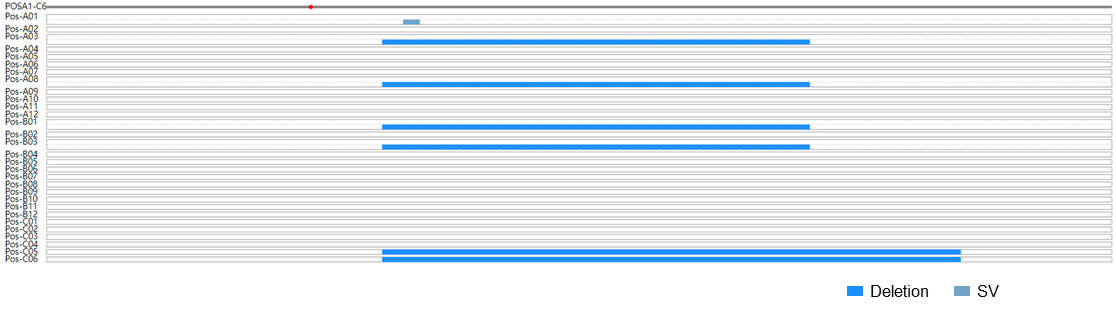


图2-7 基于HiDecode解码的输出的突变图形图

（4）其它实用功能

i）特异引物设计

HiDecode的“Primer”中提供了特异引物设计功能，用于在靶序列特异引物T#-F和T#-R上自动添加桥接序列（图2-8）。输入桥接序列、样本目的靶点名称以及目的片段的特异性匹配序列（5’-3’），点击“Generate primers”键，可批量生成含有桥接序列的靶序列特异性扩增引物，用于NGS建库的第一轮PCR反应。注意，该功能不具有对参考基因组匹配特异性的检测功能，建议使用NCBI primer-blast或者CRISPR-GE primerDesign (http://skl.scau.edu.cn/)等工具。

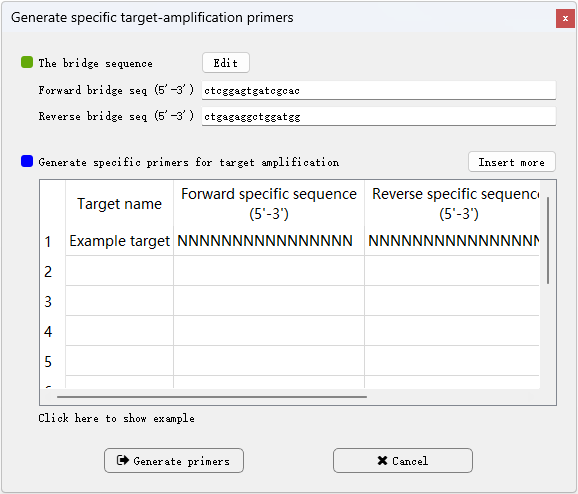


图2-8 HiDecode特异性引物设计输入界面

ii）自定义barcode

虽然HiDecode提供了96个孔位置barcode和96个板编号barcode，但用户可以根据自己需求重新定义PCR中使用的barcode序列。打开barcode选择框后，通过双击各barcode序列（图2-9），或者点击菜单栏“Setting”下拉框的“Setting the default barcodes sequence”（批量修改），可以设置用户自己定义的barcode。

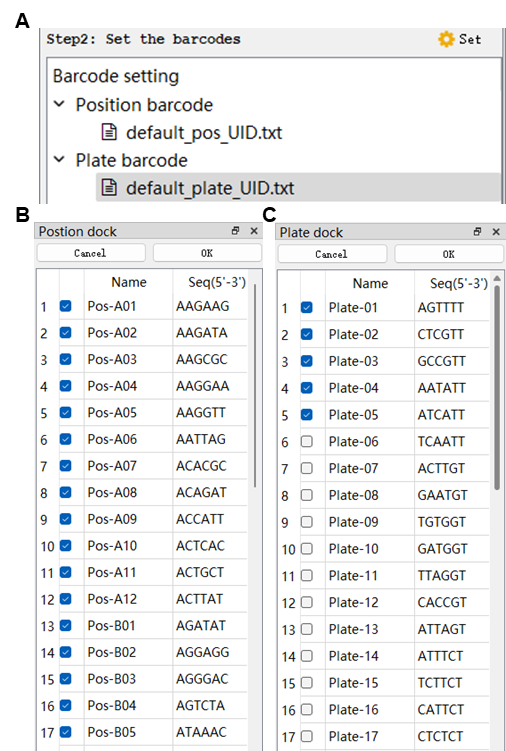


图2-9 通过barcode选择框修改barcode序列

ii）barcode在扩增子中的位置设定

为提高程序的运行速度，HiDecode采用了根据barcode在扩增子所在位置提取其序列，从而判定reads对应样本。如果用户修改了posN-F和plateN-R中barcode的位置，可以通过“Setting”下拉框的“Setting the barcodes position and length”的对话框中重新设置barcode所在位置和长度（图2-10）。

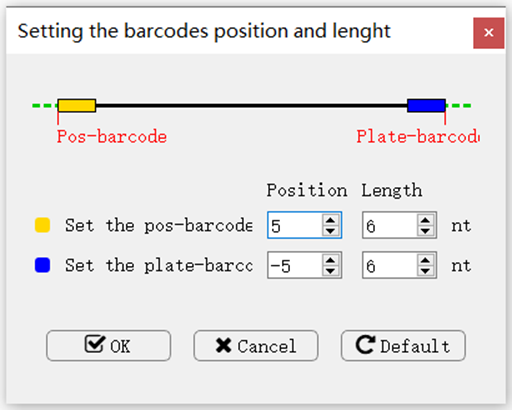


图2-10 barcode序列位置设定界面

iii）原始测序文件去除接头和过滤

绝大多数测序服务公司会提供经去除接头和过滤后的Clean reads文件。如果提供的是Raw data数据，HiDecode提供了测序数据质控以及去除接头序列的功能，通过菜单栏“Plugin”下拉框中的“Quality control for sequencing files”功能，打开该功能（图2-11），程序会自动对输入的测序文件进行低质量reads的过滤，以及依据用户提供的接头序列去除接头，输出Clean reads和质控报告。

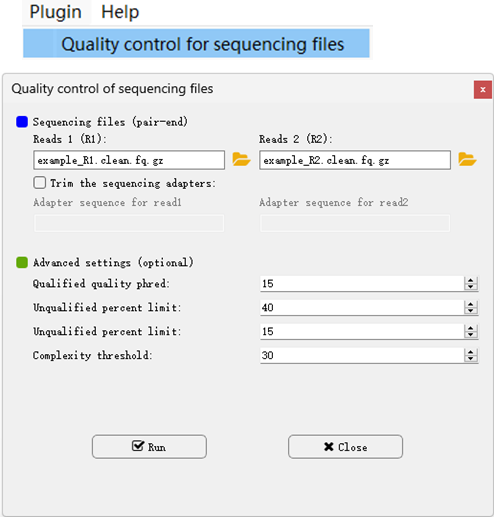


图2-11 HiDecode质控以及接头去除设置界面

iv）单样本文库检测

对于单个样本建库（即扩增靶点序列后直接建库，不进行混样）的测序数据，HiDecode提供了从单个样本的文库分析突变的功能，通过菜单栏的“Plugin”下拉框中的“Mutation analysis for single sample”，打开该功能（图2-12）。程序会根据输入的参考序列、靶点序列（可选）以及测序文件，快速分析样本的突变情况。

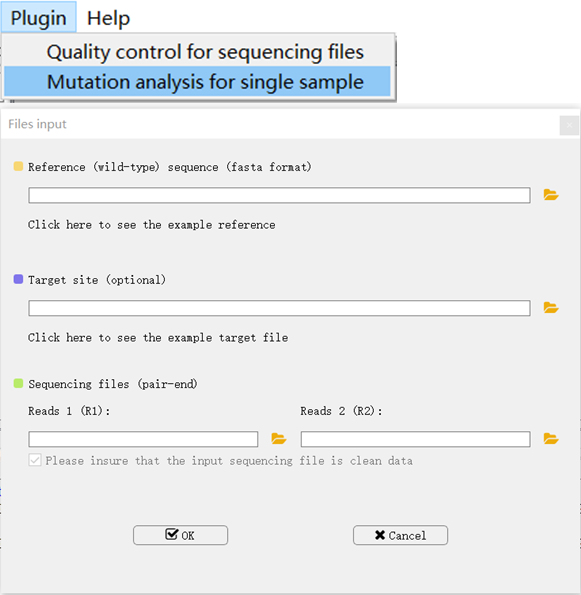


图2-12 HiDecode单样本文库输入界面

**四、LaDecode长片段混合扩增子文库的制备和分析**

对于特定目标序列内含有多个靶点同时编辑时，例如平铺删除、饱和突变等应用场景，多靶点之间往往会产生复杂的突变类型，导致利用Sanger测序或者NGS难以检测其突变。LaDecode利用三代测序（单分子测序，TGS）具有长读长的特点，对包含整个靶点区的长片段扩增子进行测序，从而获得靶点的突变序列和每个等位的单倍型。

**1、LaDecode长片段扩增子文库构建**

LaDecode的建库方法与HiDecode类似，通过两轮PCR反应，在包含多个靶点的目的片段的扩增子两端添加一对特异的barcode序列，混合所有样本的扩增产物后，进行TGS测序（图3-1）。但存在一些区别，包括：因扩增的靶点序列较长，第一轮特异引物的设计和扩增更为严格；因三代测序存在较高的错误率，因此barocde序列设置得更长；三代测序的接头比较特殊，由测序公司完成加接头的操作。

（1）LaDecode引物设计

与HiDecode建库类似，LaDecode长片段扩增子文库构建需要对样本的靶序列进行特异性扩增，并引入标识样本的孔barcode和板barcode序列，以将多个样本的扩增子混合为单个文库进行测序。文库构建涉及2对引物，包括：用于扩增靶序列的位点特异引物Target-F和Target-R，用于区分96孔PCR板的plateN-R和不同孔位置的posN-F。

i）特异引物Target-F和Target-R的设计

在第一个靶点上游约100 bp设计靶序列特异引物Target-F，在最后一个靶点下游100 bp左右的位置设计靶序列特异性引物Target-R，扩增片段控制在1～10 kb。引物的位点特异配对部分（18~20 nt, Tm值为60℃~63℃），需使用NCBI中的Blast工具检测引物特异性，并在引物5’末端添加用于第二轮扩增所需的桥接序列，序列为：

特异引物Target-F：5’- atcgcctggctccacgctccgagttNNNNNNNNNNNNNNNNNNNN-3’（NNN…为靶序列特异性引物序列，长度为18~20 nt，Tm值为58℃~62℃；小写字母为与posN-F引物特异性配对的桥接序列；下划线序列可以与扩增短片段的另一端互补配对形成茎环结构，以抑制引物二聚体和较短的非特异产物的扩增）。

特异引物Target-R：5’- tacgcctggctccacgctccctagaNNNNNNNNNNNNNNNNNNNN-3’（NNN…为靶序列特异性引物序列，长度为18~20 nt，Tm值为58℃~62℃；小写字母为与plateN-R引物特异性配对序列）。

ii）通用引物posN-F和plateN-R的设计

posN-F：5’- cttgNNNNNNNNNNNNNNNNNNNNNNNNatcgcctggctccacgctccgagtt -3’（NNN…为24 nt孔barcode序列，用于指示各样品对应96孔PCR板上的位置）

plateN-R：5’- cgacNNNNNNNNNNNNtacgcctggctccacgctccctaga -3’（NNN…为12 nt板barcode序列，用于指示不同的96孔PCR板编号，一个96孔PCR板对应1条plate-N引物）

具体序列见附表2-LaDecode所有引物序列.xlsx

将96条posN-F（3.0 μM）分装20 μL至96孔PCR板，方便使用96孔针取样器或者多通道移液器取样，-20℃保存备用。



图3-1 LaDecode长片段扩增子文库的构建流程

（2）扩增步骤

将每种LaDecode的posN-F中1管配制成3 M溶液，各取20 L加入到一块96孔PCR板对应的孔即为3 M posN-F工作液，冷冻保存

i）第一轮特异性扩增

使用靶序列特异性引物Target-F和Target-R对包含靶点的目标片段进行扩增，根据样本数配制PCR反应总混合液，将PCR反应总混合液分装至96孔PCR板中，其中PCR反应总混合液体系如表3-1所示（样本数为*n*）：

表3-1 LaDecode第一轮PCR反应总混合液体系（15 μL）

|  |  |
| --- | --- |
| 2 x Phanta max buffer | 7.5 μL × *n* |
| Target-F （10 μM） | 0.3 μL × *n*（终浓度0.2 μM） |
| Target-R （10 μM） | 0.3 μL × *n*（终浓度0.2 μM） |
| dNTPs (10 mM) | 0.3 μL × *n* |
| Phanta (1U/μL） | 0.3 μL × *n* |
| ddH2O | 6.3 μL × *n* |

（注：建议使用高保值酶扩增2 kb以上的片段）

将待检测样品的DNA（20 ng/μL）分装至96孔PCR板，并使用96孔针取样器或者多通道移液器吸取~1 μL的样品DNA（约20 ng）于PCR反应体系中进行样本目的基因扩增子的扩增，常规PCR反应程序如表3-2所示：

表3-2 LaDecode第一轮PCR反应程序

|  |  |  |
| --- | --- | --- |
| 预变性 | 95 ℃, 2 min |  |
| 变性 | 95 ℃, 30 s | 30~32 cycles |
| 退火 | 60 ℃, 20 s |
| 延伸 | (65 ℃ 10S, 68 ℃ 10S, 72 ℃, 15S) x n  （根据片段长度设置内循环数n，每1 kb/n） |
| 终延伸 | 70 ℃, 2 min |  |

（若序列片段较长，且片段的GC含量分布不均一，使用STI PCR方法（Zhao等2022）的延伸变温内循环进行片段扩增）

第一轮PCR反应结束后，每个板随机取几个样本的扩增产物进行琼脂糖凝胶（0.8%）电泳，检测扩增产物的浓度和特异性。如果扩增产物浓度较低，可增加2～3个循环，提高产物浓度。

ii）第二轮barcoding扩增

第二轮PCR扩增反应的目的是引入barcode序列。将第一轮PCR扩增产物稀释3倍（加入30 μL ddH2O）。根据样本数配制第二轮PCR总混合液（表3-3），将PCR反应总混合液分装15 μL至96孔PCR板中。

表3-3 LaDecode第二轮PCR反应总混合液体系

|  |  |
| --- | --- |
| 2 x Phanta max buffer | 7.5 μL × *n* |
| plateN-R（10 μM） | 0.3 μL × *n* (终浓度0.2 μM) |
| dNTPs (10 mM) | 0.3 μL × *n* |
| Phanta | 0.3 μL × *n* |
| ddH2O | 6.8 μL × *n* |

使用96孔针取样器或者多通道移液器从分装在96孔板的3μM posN-F取~1 μL加入到PCR反应液中（终浓度0.2 μM），以及取稀释的第一轮PCR产物~1 μL加入到PCR反应液中（最终稀释约45x）。PCR反应程序如表3-4。

表3-4 LaDecode第二轮PCR反应程序

|  |  |  |
| --- | --- | --- |
| 预变性 | 95 ℃, 2 min |  |
| 变性 | 95 ℃, 30 s | 15~18 cycles |
| 退火 | 60 ℃, 30 s |
| 延伸 | (65 ℃ 10S, 68 ℃ 10S, 72 ℃, 15S) x n  （根据片段长度设置内循环数n，每1 kb/n） |
| 终延伸 | 70 ℃, 2 min |  |

第二轮扩增后每个板可以随机取几个样本的PCR产物进行琼脂糖凝胶电泳进行检测。如果扩增产物浓度较低，可增加2-3个循环，提高产物浓度。

将每块96孔板为1组，将产物分别取出3 μL混合。以琼脂糖凝胶（0.8%）电泳，切胶回收目的片段，纯化产物。多靶点编辑产生的突变类型较多，其中包含大片段删除，为保证回收到所有片段，主要目的是去除引物以及未延伸完全的产物即可，因此电泳时间不宜过长。测定每组纯化产物的浓度后，每组样本的产物等量混合，要求DNA样品浓度>20 ng/μL，总量>2 μg。

**2、TGS测序**

纯化产物送至公司进行TGS建库测序，测序平台建议选用Pacbio HiFi测序平台。测序后要求公司交付去除测序接头后的Fasta、Fastq或者bam格式的reads文件。测序数据量根据片段大小以及样本数量设置，推荐1–4 kb、4–7 kb和7–10 kb对应分别测~1 Gb、2 Gb和3 Gb 的数据量/96样品。

**3、使用LaDecode软件分析突变**

（1）软件主界面

LaDecode软件的主界面与HiDecode相似，包括：菜单栏、参数设置、文件输入、Barcode设置、孔板信息查看以及解码结果显示模块（图3-2）。其中，菜单栏提供了常用的设置功能、插件等，在快速菜单栏设置突变位点检出阈值、CPU线程数等。文件输入模块用于输入参考序列文件、靶点序列文件以及测序文件。Barcode设置中选择使用的孔和板。样本孔板信息查看模块主要用于交互式显示各样本的分析结果。

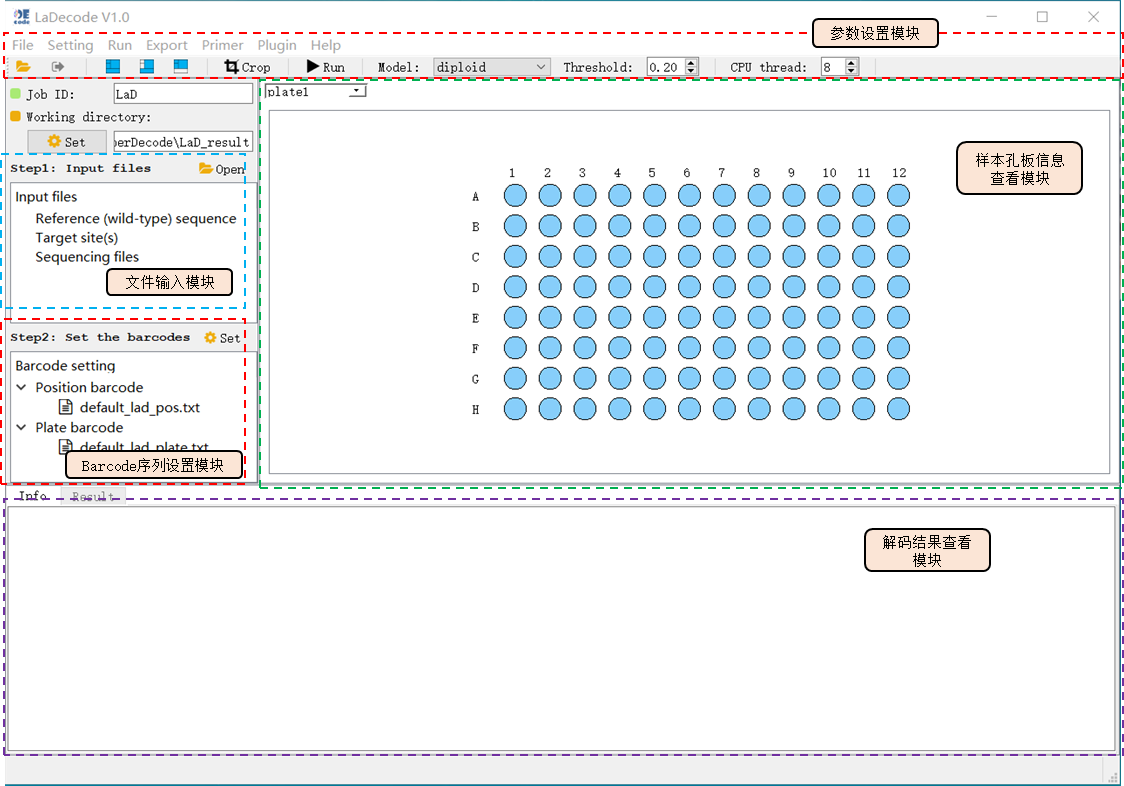


图3-2 LaDecode主界面

（2）基本分析操作

LaDecode的使用主要包括文件输入、参数设置和运行共3个主要步骤。其详细操作如下：

i) 通过工具栏或文件输入框中的“Open”按钮打开文件输入对话框，输入分析所需文件，包括：Fasta格式的参考序列文件（覆盖Target-F到Target-R之间的野生型基因组序列）、可选的靶点序列文件（包含或者不包含PAM序列均可以）、以及测序文件（Fasta、Fastq或者bam格式文件）（图3-3）。点击输入框下方的提示可以查看文件格式的示例。

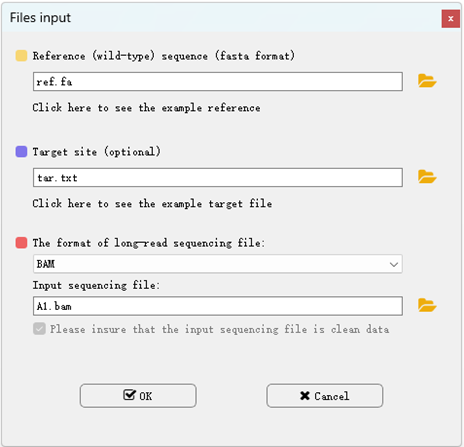


图3-3 LaDecode文件输入界面

ii）通过双击左侧的“default\_lad\_pos.txt”和“default\_lad\_plate.txt”设置扩增时使用的孔barcode和板barcode。打开设置框后，勾选使用的barcode。

iii）在LaDecode的工具栏中设置数据分析对应的参数。程序提供了预设的3种突变检出阈值Threshold，分别为diploid（用于分析二倍体样本数据）、polyploid（用于分析多倍体样本数据）以及low frequency（用于分析细胞系、原生质体、愈伤组织等样本数据）。除了预设的Threshold，用户可以根据需要自行调整。当样本中某变异位点的reads数占比大于设置的阈值时，程序才认为该基因型为真实的突变，从而输出该突变位点。

iv）点击“Run”按钮启动分析任务。

（3）分析结果导出和结果解读

LaDecode解码结束后，程序运行结束后，样本孔板信息模块中会显示每个孔的分析结果，黄色表示无结果，蓝色表示有结果，红色为当前选中的样本（图3-4）。选中某个样本时，底部的解码查看模块中会显示样本的突变情况，包括对应的突变类型、覆盖的总reads数、参考序列、每条突变序列及其支持的reads数。通过拉动上方的滑动条，可以调整需要显示的序列范围。

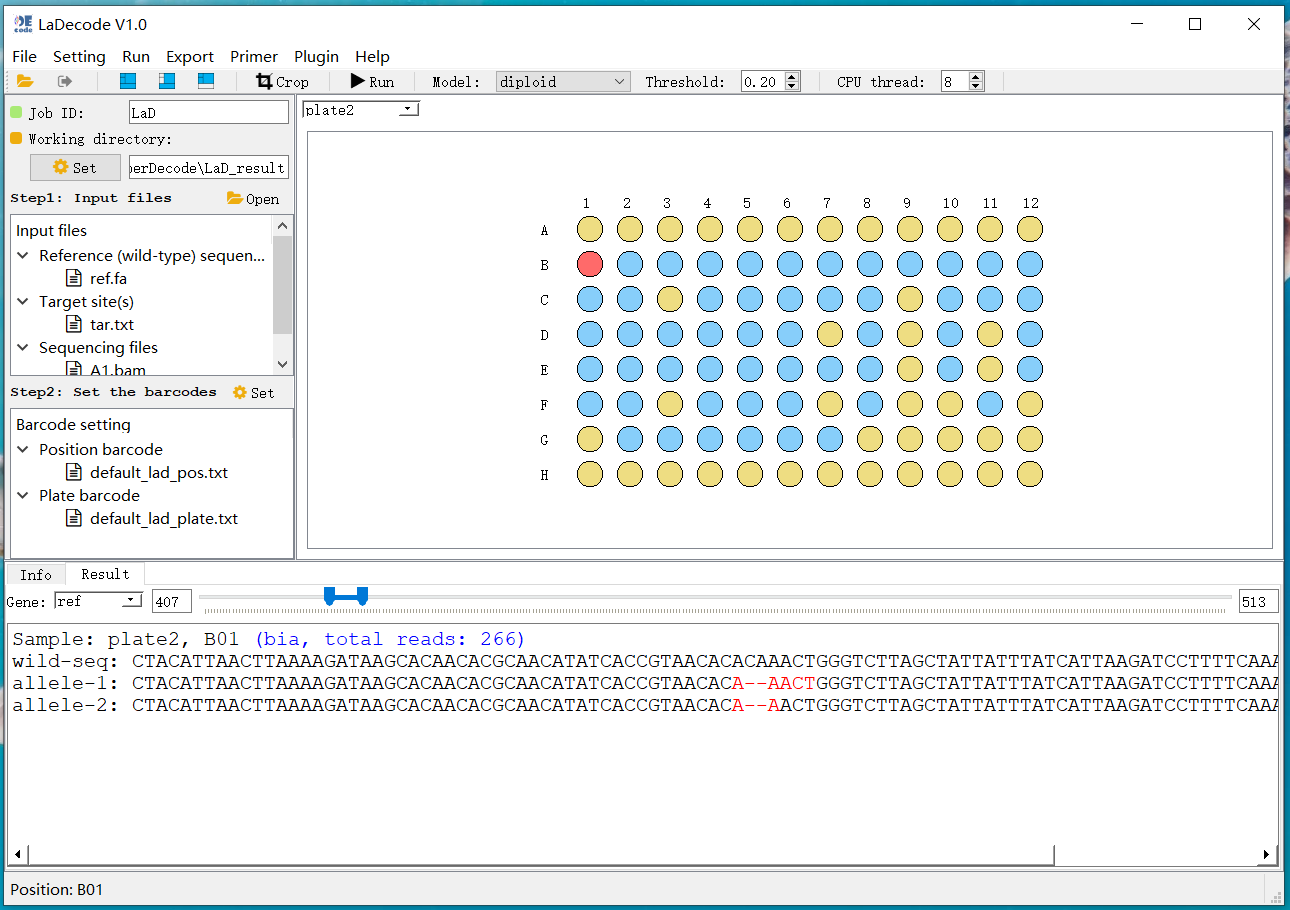


图3-4 LaDecode结果分析

解码后的结果可点击菜单栏的“Export”将结果输出至excel文件，可以输出两种类型的结果文件。“Export all results to excel file”为输出所有孔和板的解码详细结果（包括reads数、突变碱基、每条突变序列和野生型序列的比对等）（图3-5），“Fetch mutations to a single file”为仅输出含有数据的样本信息、突变碱基及其reads数。



图3-5 基于LaDecode解码的样本突变结果

另外，“Export”中的“Plot mutations of all samples”可以将不同样本的基因型以图形的形式输出，方便用户直观查看样本的突变结果（图3-6）。

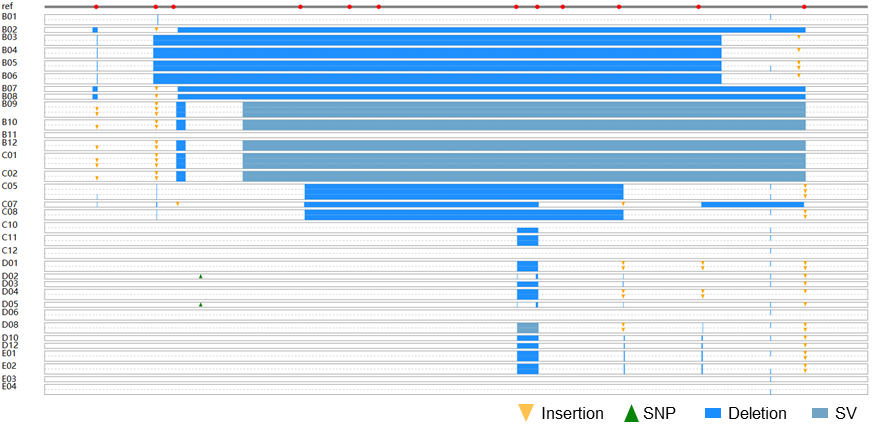


图3-6 基于LaDecode解码的输出的突变图

（4）其它实用功能

i）分析已拆分的测序文件

部分测序公司可以提供混合扩增子样本的拆分服务，根据客户提供的barcode序列，将拆分后的各样本reads文件提供给客户。为此，LaDecode提供了对已拆分数据进行分析的插件工具，点击菜单栏“Plugin”中的“Muataion analysis from demultiplexed fasta”，可进入该功能的分析界面，在对话框中输入野生型的参考序列文件、可选的靶点序列文件以及已拆分reads文件的存放路径（通常为Fasta格式），点击“OK”按钮，即可进行解码分析（图3-7），输出的结果文件自动保存至reads文件所在目录。

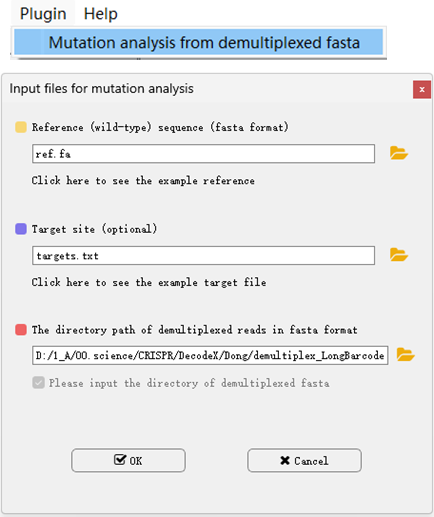


图3-7 基于LaDecode分析已拆分的测序数据

ii）特异引物设计

LaDecode的“Primer”中提供了特异引物设计功能，用于在靶序列特异正向引物Target-F和Target-R自动添加桥接序列（图3-8）。输入桥接序列、目的靶点名称以及不含桥接序列的目的片段特异扩增的匹配序列（5’-3’），点击“Generate primers”键，可批量生成含有桥接序列的靶序列特异性扩增引物，用于NGS混合建库的第一轮PCR反应。建议使用NCBI primer-blast或者CRISPR-GE primerDesign等工具检测所设引物在参考基因组的特异性。

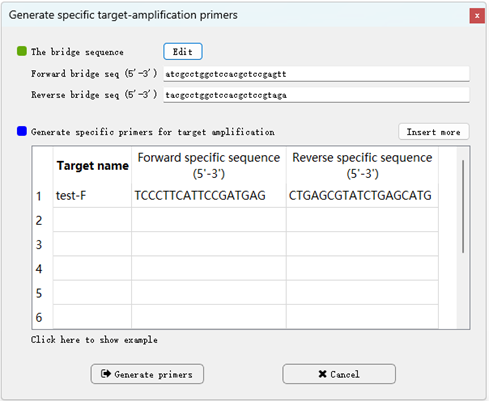


图3-8 LaDecode引物设计输入界面

iii）自定义barcode

虽然LaDecode提供了96个孔barcode和10个板barcode，但用户可以根据自己需求重新定义PCR中使用的barcode序列。打开barcode选择框后，通过双击各barcode序列（图3-9），或者点击菜单栏“Setting”下拉框的“Setting the default barcodes sequence”（批量修改，并保存为默认barcode），可以设置用户自己定义的barcode。

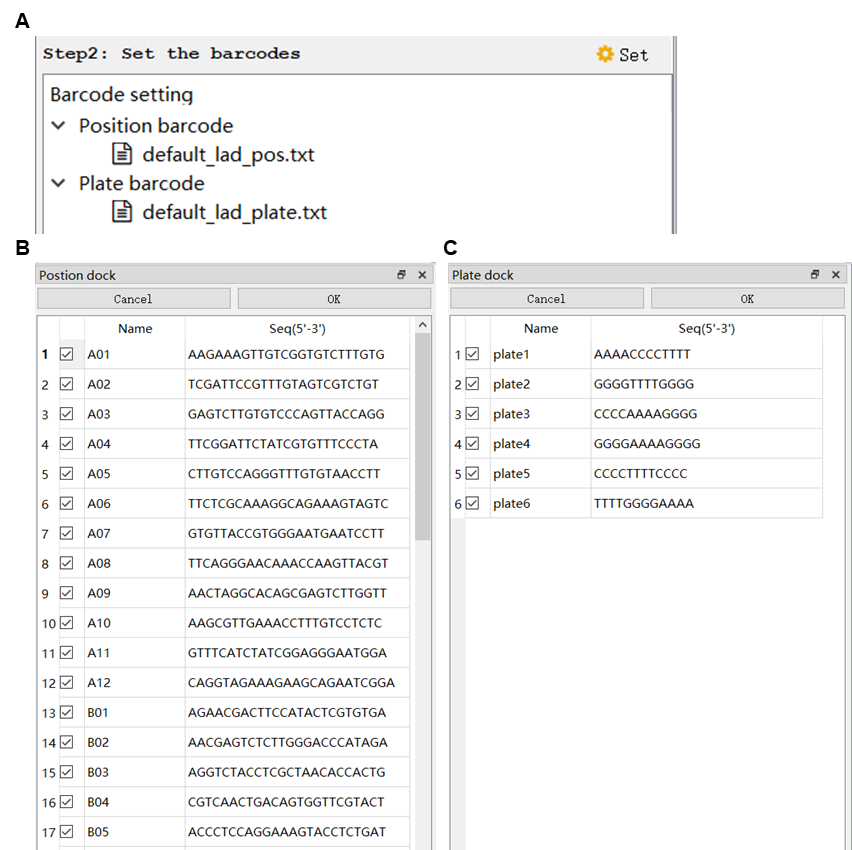


图3-9 LaDecode中barcode定义界面

**五、问题反馈**

谢先荣，xiexianrong@scau.edu.cn