

# SuperDecode (v1.0)使用手册

刘耀光实验室，华南农业大学农学院

2025 年 3 月

**SuperDecode (v1.0)**是一款主要用于基因编辑靶点突变分析的工具箱，包括本地版和在线网络版。此软件包由 3 个模块构成，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测序数据进行解码分析。DSDecodeMS 为我们之前的网页版工具 DSDecode 和 DSDecodeM (Liu et al., 2015; Xie et al., 2017, <http://skl.scau.edu.cn/>)的升级电脑本地版。它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，包括纯合、杂合和双等位突变等。HiDecode 是对在样本靶点扩增子两端添加一对 barcode 的多样本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，也适合于自然变异位点分析和细胞器 RNA 编辑效率的分析。LaDecode 是利用单分子测序技术对目标 DNA 区段内包含多个编辑靶点的扩增子进行高通量分析，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(二倍体双等位和杂合突变)采用 Sanger 测序和 DSDecodeMS 分析；每批较多样品(多于 40 或 50 个样本时)采用 NGS 和 HiDecode 分析。

本软件包公开供科研人员非盈利目的使用；公司或机构的商业盈利目的使用本软件包需要获得我们的授权许可。

本软件包相关信息：Li et al. SuperDecode: an integrated toolkit for analyzing mutations induced by genome editing. Molecular Plant. 2025, DOI: <https://doi.org/10.1016/j.molp.2025.03.002>

## 一、软件下载和安装

### 1、硬件要求

为了保证 SuperDecode 的正常运行，我们推荐在内存大于 4Gb 并且 CPU 至少含有 4 个核心的计算机上运行 SuperDecode 本地版软件包。

### 2、软件下载

本地版 **SuperDecode** 提供了 Windows 和 MacOS 界面版本，以及 Linux 命令行版本。此软件包已上传至 GitHub (<https://github.com/xiexr/SuperDecode>) 供下载。此外，考虑到国内访问 GitHub 网络受到限制，我们也将此软件包上传到奶牛快传(高速的数

据传输工具，链接为：<https://tbtools.cowtransfer.com/s/d69cdec128f64e>），请用户下载对应电脑系统的软件包。此外，也可以通过华南农业大学夏瑞实验室开发的 TBtools-II 插件（Plugin）仓库直接使用 **SuperDecode**。

测序数据量较小的分析任务，用户也可以使用 **SuperDecode 在线网络版**（网址为：[www.crispr-ge.com/superdecode/](http://www.crispr-ge.com/superdecode/)）。用户也可以通过该网站下载本地版工具包。

### 3、软件安装

本软件为可以直接运行的程序，无需在电脑安装。下载后，将软件解压至电脑本地磁盘（注意：软件包所在路径不包含空格和中文字符）。对于界面版，进入 SuperDecode 文件夹后，直接双击“SuperDecode.exe”即可以打开运行软件。为便于使用，可以将“SuperDecode.exe”发送快捷方式到电脑桌面。MacOS 版本在第一次打开时需要加载相关库，所需时间会长些，以后再次打开就很快进入运行状态了。在 MacOS 版本下第一次使用 HiDecode 和 LaDecode 时，可能会提示模块调用的软件未通过系统安全检测，用户在提示后需要在 MacOS 系统的“安全与隐私”中设置相应权限，即可正常使用。对于 Linux 版本，进入解压后（解压命令：`tar -zxvf 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）。

### 4、测试数据下载

SuperDecode 为每个模块提供了练习测试数据文件，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、用于 DSDecodeMS 的扩增子样本制备和分析

DSDecodeMS 主要用于少数二倍体样本的简单突变类型（纯合、杂合和双等位突变）的检测。它直接读取编辑靶点扩增子的 Sanger 测序（ab1 格式）文件峰图，分析靶点的突变序列和类型，

### 1、包含编辑靶点片段的扩增

因 Sanger 测序读长限制和考虑 PCR 扩增效率，建议目的扩增片段长度在 500~800 bp（过短片段反而不利于电泳分析和回收纯化），在靶点前后大约 200~400 bp 的位置设计特异性扩增引物 T#-F 以及 T#-R（长度大约为 18~22 nt， $T_m$  值 =  $(GC\% \times 41 + 69.3) - 650/L$ （L 引物碱基数）= 58°C~60°C。使用 NCBI 的 Primer-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 个靠近的靶点。如果目的片段在基因组中存在同源性较高的序列，可以适当延长或缩小扩增片段以设计特异选择性引物，获得特异的扩增子。扩增体系如表 1-1。

表 1-1 DSDecodeMS 样本扩增子 PCR 反应体系（每反应 25  $\mu\text{L}$ ， $n$  为反应数）

2 $\times$ Taq mix (普通 Taq)	12.5 $\mu\text{L} \times n$
T#-F (10 $\mu\text{M}$ )	0.5 $\mu\text{L} \times n$ (终浓度 0.2 $\mu\text{M}$ )
T#-R (10 $\mu\text{M}$ )	0.5 $\mu\text{L} \times n$
gDNA	20~30 ng
ddH <sub>2</sub> O	补足到总量 26 $\mu\text{L} \times n$ (每反应有 1 $\mu\text{L}$ 分装冗余)

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°C 3 min	
变性	96°C 15 s	28~30 cycles
退火	58~60°C 15 s	
延伸	65°C 10 s, 68°C 10 s, 72°C 10 s, 65°C 10 s (如果目的片段>1 kb, 每节温度加 3~5 s)	
终延伸	70°C, 1 min	

(注：采用变温延伸方式，可改善不同 GC 含量分布序列的扩增效率)

取 3  $\mu\text{L}$  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  $\mu\text{L}$ ) 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，然后使用凝胶纯化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，建议在靶点上游或下游 150~250 bp 处设计内部测序引物 (17~20 nt,  $T_m=56^\circ\text{C}\sim 58^\circ\text{C}$ ) 进行测序以获得高质量测序文件，不建议使用 PCR 扩增用过的引物再用于测序 (经常会产生杂峰信号)。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、使用本地软件包中的 DSDecodeMS 对测序峰图解码

使用 SuperDecode 本地软件包中的 DSDecodeMS 模块对 Sanger 测序 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、子序列搜索窗口等；参考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（可不输入）、ab1 测序文件输入（每批不限文件数目）；峰图查看模块用于实时显示所选 ab1 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。

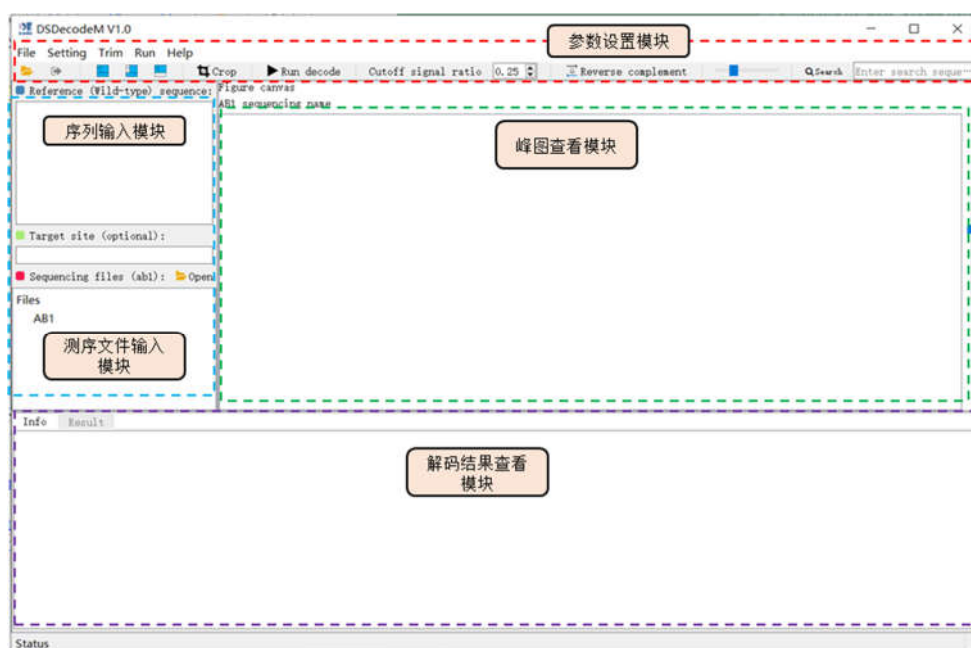


图 1-1 DSDecodeMS 主界面

### (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入、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（包含或不包含 PAM），用于程序快速定位靶点序列附近的突变；一般情况下（测序文件质量好）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（高质量测序文件），用软件预设的参数设置就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 长度、简并序列 DS 长度和信号阈值后，再启动分析。

iii) 点击“Run decode”按钮启动突变分析任务。

DSDecodeMS 快速完成分析任务，把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 30~50 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，可能会对分析产生影响。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。

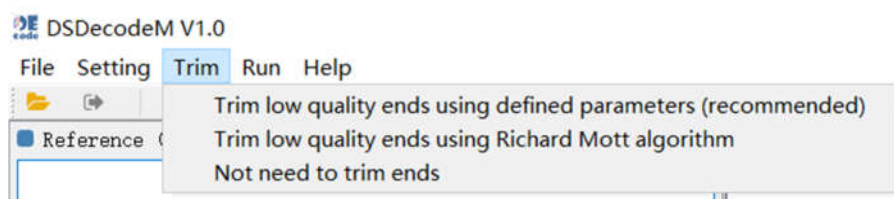


图 1-2 DSDecodeMS 修剪序列两端的方法

ii) 截图功能

DSDecodeMS 读取的 ab1 文件峰图具有较高的清晰度，可用于论文报告的发表。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（最少 4 碱基）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。

3、使用网络版在线工具的 DSDecodeM 对测序峰图解码

使用 SuperDecode 网络版在线工具的 DSDecodeM 模块对 Sanger 测序 ab1 文件进行突变检测。

(1) DSDecodeM 网络版在线工具界面介绍

与 SuperDecode 本地版中的 DSDecodeMS 模块不同，SuperDecode 网络版在线工具的 DSDecodeM 模块仅包含参考序列输入、测序文件输入和参数设置模块（图 1-3）。其中，参考序列输入模块用于输入目标序列的野生型参考序列；测序文件输入模块用于输入 ab1 测序文件(网络版在线工具单个分析任务最多可输入 20 个 ab1 测序文件)；参数设置模块用以设置锚定序列 AS、简并序列 DS、荧光信号阈值以及可选择输入的靶点序列。

SuperDecode

Home

DSDecodeM

HiDecode

LaDecode

Document

DSDecodeM: a tool for rapid decoding of multiple superimposed sequencing chromatograms.

Sequence-specific nuclease-based genome targeting systems, i.e., TALENs and CRISPR/Cas9, often cause uniform biallelic and heterozygous mutations in diploid organisms. Direct sequencing of PCR products containing such mutations results in superimposed sequencing chromatograms. The Degenerate Sequence Decode (DSDecodeM) program serves to decode automatically sequencing chromatograms with biallelic, heterozygous, and homozygous mutations into allelic sequences.

For testing the program, five sequencing ab1 files and the corresponding reference are provided. Click [download](#) to obtain the test files.

Upload a reference sequence

Please enter an intact (wild-type) reference sequence of your target material. It must completely cover the sequence range of the sequencing ab1 file(s) for one or multiple editing targets.

reference sequence

The program can search the two strands of the reference, so, it not necessary to convert the complement of the reference sequence.

参考序列输入模块

Upload chromatogram files

Upload the multiple chromatogram (\*.ab1) files (up to 20 files):

选择文件

未选择任何文件

The multiple chromatogram (\*.ab1) files can be for the same target site or different targets covered by the reference sequence.

ab1测序文件输入模块

Optional parameter setting

Cutoff signal ratio (0.1-0.8)

0.30

Length of anchor sequence (10-20 nt)

20

Length of degenerate sequence (10-50 nt)

15

Target sequence of cleavage site (12-20 nt) \*

Usually it's not needed to enter a target sequence

\*For decoding failed at first time, entering the target sequence would be helpful and the chromatogram (\*.ab1) file(s) should belong to the same target.

参数设置以及靶点序列输入模块

Run DSDecodeM

图 1-3 DSDecodeM 网络版在线工具主界面

(2) 利用 DSDecodeM 进行解码的操作

DSDecodeM 的使用主要分为 4 步，分别为输入参考序列、上传 ab1 测序文件、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) 在参考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）。

ii) 在测序文件输入模块中点击“选择文件”按钮上传 ab1 文件（每批限 20 文件）。

iii) 在参数设置模块中设置锚定序列 AS 长度、简并序列 DS 长度、荧光信号阈值以及靶点序列（一般的高质量测序文件采用默认值即可，不必调整参数）。

iv) 点击“Run DSDecodeM”按钮启动突变分析任务。

程序运行结束后，解码结果会跳转至新的网页中显示（图 1-4），可以通过鼠标选择相应结果复制保存。

The screenshot displays the SuperDecode web application interface. At the top, there are navigation tabs: Home, DSDecodeM (selected), HiDecode, LaDecode, and Document. Below the tabs, a description states: "DSDecodeM: a tool for rapid decoding of multiple superimposed sequencing chromatograms." It also shows a job ID: 202501071whAeG and a note about copying results. The main content area shows two sample results, each with a sample name, anchor sequence, degenerate sequence, two alleles, and their alignments with the reference sequence.

**Sample: test\_file1.ab1**

Anchor sequence: GTTTAATGGCAAGGAGTTCC  
Degenerate sequence: [CG]T[CT][TG][CG]C[TA][CG][TG][CG][TC][CA]G[AT][CT]  
Allele1: GTTTAATGGCAAGGAGTTCCCTCGGCAGTCTCGAT  
Allele2: GTTTAATGGCAAGGAGTTCCGTTTCTCGGCAGTC  
Alignments of Allele1, Allele2, and reference sequence:  
Allele1: GTTTAATGGCAAGGAGTTCC-----CTCGGCAGTCTCGAT (deletion)  
Allele2: GTTTAATGGCAAGGAGTTCCGTTTCTCGGCAGTC (insertion)  
Reference: GTTTAATGGCAAGGAGTTCCG-TTCCTCGGCAGTCTCGATGAAACCTAGAGTTGAGATTGGTGGCGATGAT

**Sample: test\_file2.ab1**

Anchor sequence: AATGGCAAGGAGTTCCGTTT  
Degenerate sequence: [CT][CT][CG]G[GC][CA][GA][TG][CT][CT][CT][CG][AG][AT][TG]  
Allele1: AATGGCAAGGAGTTCCGTTTCTCGGCAGTCTCGAT  
Allele2: AATGGCAAGGAGTTCCGTTTCTCGGCAGTCTCGATG  
Alignments of Allele1, Allele2, and reference sequence:  
Allele1: AATGGCAAGGAGTTCCGTTTCTCGGCAGTCTCGAT (WT)  
Allele2: AATGGCAAGGAGTTCCGTTTCTCGGCAGTCTCGATG (deletion)  
Reference: AATGGCAAGGAGTTCCGTTTCTCGGCAGTCTCGATGAAACCTAGAGTTGAGATTGGTGGCGATGATTTTC

图 1-4 DSDecodeM 网络版在线工具解码结果

#### 4、DSDecodeMS/DSDecodeM 解码失败的原因和解决办法

i) PCR 产物可能含有非特异产物，如果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，引物对它们没有选择特异性，导致产物特异性差，Sanger 测序时产生重叠峰。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），利用目的序列与其重复序列之间的单碱基或多碱基差异位点，将引物 3' 端设计到差异位点上提高产物的特异性。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) 样本的靶突变为嵌合突变等复杂突变，或者是该区间存在多个编辑靶点，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必须是与样本编辑前的野生型一致的序列（确认从某品系的参考基因组下载的参考序列要与目的品系的编辑前野生型序列完全一致）。检查输入的参考序列是否对应目标扩增序列并且与目的样本编辑前的野生型序列一致。

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

当个人或团队同时检测的样本总量较大（多于 40 或 50 个样本）时，推荐使用 NGS 和 HiDecode，可降低检测费用。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。在本手册中，我们推荐了一种通过两轮 PCR 在各样本的扩增子末端添加一对样本 barcode 的方法，从而实现将多个样本的扩增子混合，建成一个测序文库。除了本手册推荐的 barcode，用户也可以自定义样本 barcode，建成混合扩增子测序文库。此外，HiDecode（和 LaDecode）也支持对单个样本的扩增子文库进行分析（即仅通过简单的扩增，不需引入样本 barcode）。

#### 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）。



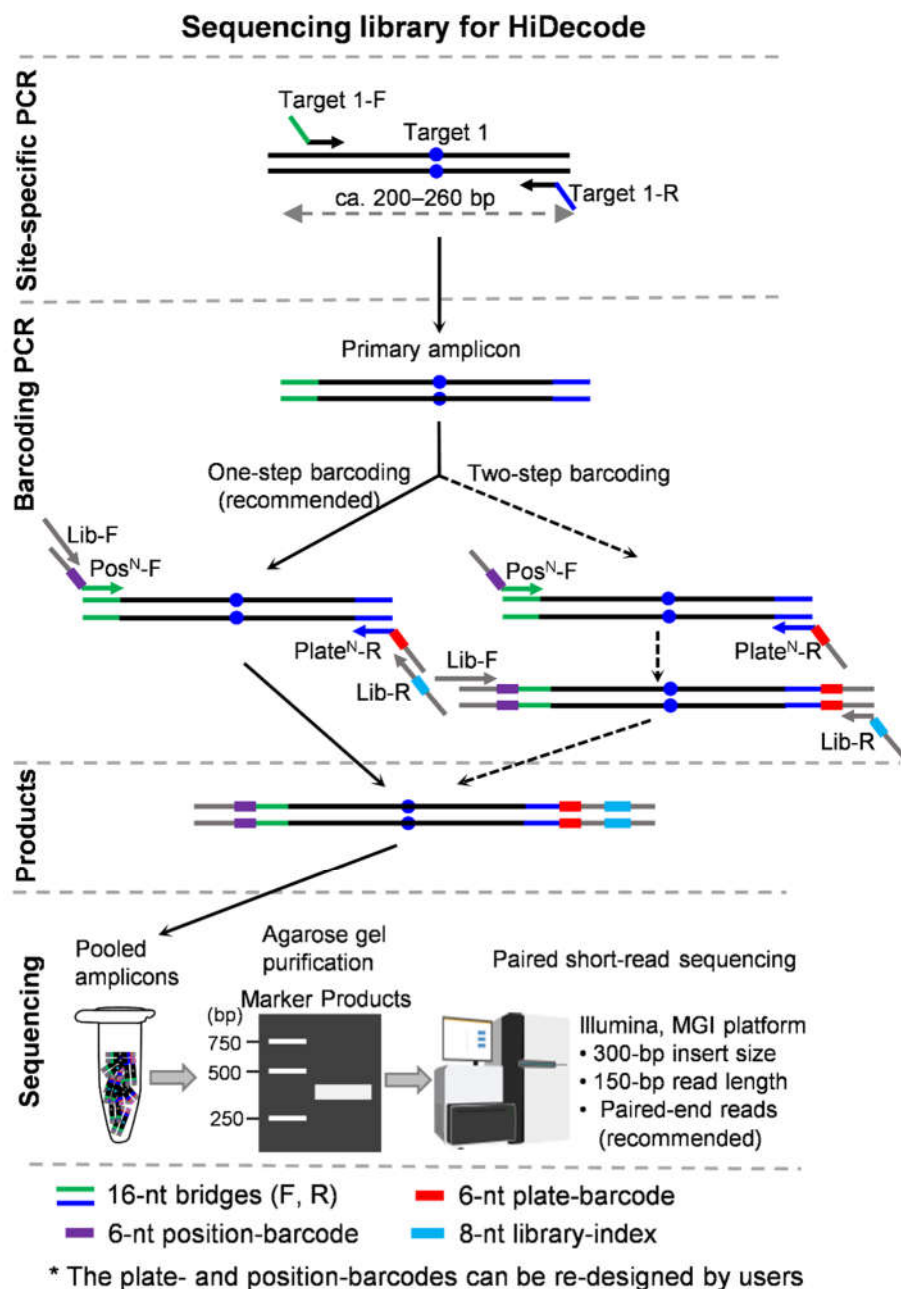


图 2-1 HiDecode 短片段 NGS 文库的构建流程

### (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，用于区分不同样本的 pos<sup>N</sup>-F/col<sup>N</sup>-F 和 plate<sup>N</sup>-R/row<sup>N</sup>-R（超过 96 个样本的文库构建，使用区分 96 孔 PCR 板的 plate<sup>N</sup>-R 和不同孔位置的 pos<sup>N</sup>-F，少于 96 个样本的文库构建，可选择使用区分 96 孔 PCR 板不同行位置的 row<sup>N</sup>-R 和不同列位置的 col<sup>N</sup>-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。

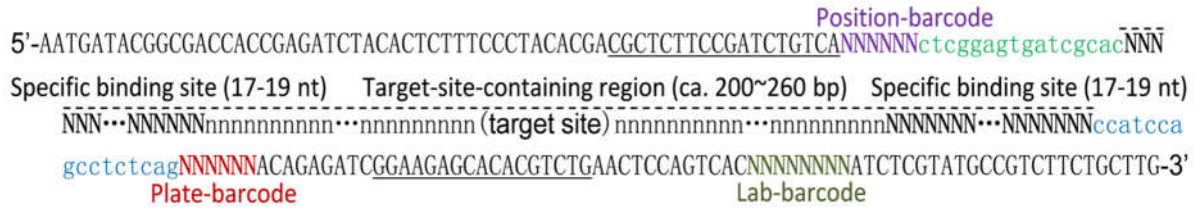


图 2-2 NGS 文库的序列构成

#### 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'-ctcggagtgatcgacNNNNNN...NNNNN-3'（17~19 个 N 为基因组位点特异序列（ $T_m = 56^{\circ}\text{C} \sim 58^{\circ}\text{C}$ ）；5'端小写 16 nt 为给与 pos<sup>N</sup>-F/col<sup>N</sup>-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 为基因组位点特异序列（ $T_m = 56^{\circ}\text{C} \sim 58^{\circ}\text{C}$ ）；5'端小写 16 nt 为给与 plate<sup>N</sup>-R/row<sup>N</sup>-R 配对的接桥序列）。

#### ii) 通用引物 pos<sup>N</sup>-F/col<sup>N</sup>-F 和 plate<sup>N</sup>-R/row<sup>N</sup>-R

对于不同样本数的文库构建，HiDecode 提供了两种文库构建的策略，分别是使用 pos<sup>N</sup>-F 和 plate<sup>N</sup>-R 构建样本数大于 96 的文库（根据实验需求合成 96 条 pos<sup>N</sup>-F 和对应 PCR 板数目的 plate<sup>N</sup>-R），使用 col<sup>N</sup>-F 和 row<sup>N</sup>-R 构建样本数等于或少于 96 的文库（根据实验需求合成对应数目的 col<sup>N</sup>-F 和 row<sup>N</sup>-R，一个文库最多使用 12 条 col<sup>N</sup>-F 和 8 条 row<sup>N</sup>-R）。

对于构建待检测样本数大于 96 的文库，使用 pos<sup>N</sup>-F 和 plate<sup>N</sup>-R 引物对每个样本添加特异性 barcode，pos<sup>N</sup>-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；plate<sup>N</sup>-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 plate<sup>N</sup>-R。用户可根据实验室实际需求合成 plate<sup>N</sup>-R 条数，本方法提供了 96 条 plate<sup>N</sup>-R 供选择。HiDecode 中内置的 pos<sup>N</sup>-F 和 plate<sup>N</sup>-R 为（具体序列见附表 1-HiDecode 所有引物序列.xlsx）：

pos<sup>N</sup>-F: 5'-CGCTCTTCCGATCTGTCANNNNNNctcggagtgatcgac-3'（下划线序列为与文库引物 Lib-F 配对的接头序列，NNNNNN 为区分 PCR 板上 96 孔位置（A1~H12）的特异性 barcode 序列，小写 16 nt 为此 barcoding PCR 引物结合第一轮产物的正向桥

接序列)。

plate<sup>N</sup>-R: 5'-CAGACGTGTGCTCTTCCGATCTCTGTNNNNNNNctgagaggctggatgg-3' (下划线序列为与文库引物 Lib-R 配对的接头序列, NNNNNN 为区分不同 PCR 板的特异性 plate-barcode 序列, 小写 16 nt 为此 barcoding PCR 引物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 pos<sup>N</sup>-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°C 保存备用。

对于构建 12 x 8 barcode 模式的文库, 可选择使用 col<sup>N</sup>-F 和 row<sup>N</sup>-R 引物对每个样本添加特异性 barcode。col<sup>N</sup>-F 从 5' 到 3' 端分别含有与文库接头引物 (Lib-F, Lib-R) 匹配的序列、指示各样品对应 PCR 板上 12 列位置 (1~12) 的 barcode 碱基 (6 nt)、以及结合第一轮产物 5' 端 16 nt 的桥接序列, 共 12 条; row<sup>N</sup>-R 含有与文库接头匹配的序列、指示各样品对应 PCR 板上 8 行位置 (A~H) 的 barcode 碱基 (6 nt)、以及结合第一轮产物 3' 端 16 nt 的桥接序列, 共 8 条。本方法提供了 12 条 col<sup>N</sup>-F 和 8 条 row<sup>N</sup>-R 供选择。HiDecode 中内置的 col<sup>N</sup>-F 和 row<sup>N</sup>-R 为 (具体序列见附表 1-HiDecode 所有引物序列.xlsx):

col<sup>N</sup>-F: 5'-CGCTCTTCCGATCTGTCANNNNNNctcggagtgatgcac-3' (下划线序列为与 Lib-F 配对的接头序列, NNNNNN 为区分 PCR 板上 12 列位置 (1~12) 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)。

row<sup>N</sup>-R: 5'-CAGACGTGTGCTCTTCCGATCTCTGTNNNNNNNctgagaggctggatgg-3' (下划线序列为与 Lib-R 配对的接头序列, NNNNNN 为区分 PCR 板上 8 行位置 (A~H) 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)。

建议提前将 12 条 col<sup>N</sup>-F (2.0 μM) 和 8 条 row<sup>N</sup>-R (2.0 μM) 各吸取 10 μL 至 96 孔 PCR 板对应的孔中混合, 以方便使用 96 孔针取样器或者多通道移液器取样, -20°C 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 方法引入到扩增子中。接头中含有实验室特异的文库 barcode (Lib-barcode)。测序公司根据不同客户提供的 Lib barcode, 将多用户样品混合测序的原始测序数据包进行标识拆分给各客户。相同实验室 (团队) 的同一批次的多个测序次级文库可以使用不同的 Lib barcode, 也可使用相同 Lib barcode, 后者的次级文库之间则以指示不同 PCR 板编号的 plate<sup>N</sup>-barcode 进行区分 (例如次级文库 1 是第 1~10 号板, 次级文库 2 是 11~20 号板)。多数测序公司支持单 Lib barcode 或者双 Lib barcode 对数据进行拆分, 如果是单 Lib barcode, 接头序列如下:

Lib-F: 5'-  
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT

CTGTCA-3'（下划线序列为与 pos<sup>N</sup>-F/col<sup>N</sup>-F 配对的接头序列）

Lib-R: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTT  
CAGACGTGTGCTCTTCC-3'（下划线序列为与 plate<sup>N</sup>-R/ row<sup>N</sup>-R 引物配对的接头序  
列，8 N 序列为 Lib barcode。填写测序送样单时，需将 Lib-R 中的 Lib barcode 写成反  
向互补序列即正链的序列）。

如果使用双 Lib 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 pos<sup>N</sup>-F/col<sup>N</sup>-F、plate<sup>N</sup>-R/ row<sup>N</sup>-R、和 Lib-F、Lib-R。

**(2) 扩增步骤**

i) 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 0.3× TE 配制（建议不要用 1× TE，会带入  
过多 EDTA 抑制酶活性）。

特异引物 T#-F, T#-R, plate<sup>N</sup>-R: 稀释为 10 μM 保存液，保存待用。

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

col<sup>N</sup>-F 和 row<sup>N</sup>-R 的合成以及工作液配置：根据实验需求合成对应条数的 col<sup>N</sup>-F(单  
个文库构建最大使用数为 12 条, col<sup>1</sup>-F~ col<sup>12</sup>-F) 以及对应条数的 row<sup>N</sup>-R (单个文库最  
大使用数为 8 条, row<sup>A</sup>-R~ row<sup>H</sup>-R)，各 5 OD, 1 OD/管。每种引物其中 1 管以 1× TE  
溶解成 10 μM 保存液，其余的以干粉-20℃冷冻长期储藏。将每种 col<sup>N</sup>-F (10 μM) 和  
row<sup>N</sup>-R 取 2.0 μL 加入到一块 96 孔 PCR 板对应的孔，每孔加入 16 μL 0.3× TE，即为 **1.0  
μM col<sup>N</sup>-F/row<sup>N</sup>-R 工作液**，冷冻保存（以下使用到剩余 4~5 μL 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 0.3× TE 到一共 10 μL（即 20 ng/μL）。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, *n* 为反应数）

2 × Taq mix (普通 Taq)	7.5 μL × <i>n</i>
----------------------	-------------------

T#-F (10 $\mu$ M)	0.3 $\mu$ L $\times n$ (终浓度 0.2 $\mu$ M)
T#-R (10 $\mu$ M)	0.3 $\mu$ L $\times n$ (终浓度 0.2 $\mu$ M)
ddH <sub>2</sub> O	补足到 15.5 $\mu$ L $\times n$ (每反应有 0.5 $\mu$ L 分装冗余)

使用 96 孔针取样器（或者多通道移液器）点取~1 $\mu$ L 的 DNA 到已分装了 PCR 反应液的孔中（96 孔针取样器每次点取样品后需使用清水冲洗 3 次，再使用无水乙醇或 95%工业酒精（用喷壶）喷洗 1 次，点燃，去除在 96 孔针取样器上残留的 DNA）。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°C 3 min	
变性	96°C 10 s	10 cycles
退火	58°C 20 s	
延伸	68°C 5 s	
延伸	72°C 5 s	
延伸	68°C 5 s	
变性	96°C 10 s	18~20 cycles
退火/延伸	68°C 5 s	
延伸	72°C 5 s	
延伸	68°C 5 s	
终延伸	70°C 1 min	1 cycle

（注：采用变温延伸方式，以兼顾不同 GC 含量序列的扩增效率）

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 难以特异性扩增目的片段，可在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（ $T_m = \sim 60^\circ\text{C}$ ）。先进行巢式 PCR 预扩增富集高特异产物模版。PCR 反应体系如表 2-3 所示。

表 2-3 HiDecode 样本建库巢式 PCR 预扩增体系（每反应 15  $\mu$ L）

2 $\times$ Taq mix （普通 Taq）	7.5 $\mu$ L $\times n$
exT#-F (10 $\mu$ M)	0.15 $\mu$ L $\times n$ (终浓度 0.1 $\mu$ M)
exT#-R (10 $\mu$ M)	0.15 $\mu$ L $\times n$ (终浓度 0.1 $\mu$ M)
ddH <sub>2</sub> O	补足到 15.5 $\mu$ L $\times n$

PCR 反应扩增程序如表 2-4 所示：

表 2-4 HiDecode 样本建库巢式 PCR 预扩增程序

预变性	93°C 3 min	
变性	96°C 10 s	15 cycles
退火	60°C 20 s	
延伸	68°C 10 s	
延伸	72°C 20~30 s	
延伸	68°C 10 s	

（延伸时间根据产物长度适当调整）

将 T#-F 以及 T#-R 引物(原 10  $\mu\text{M}$ )等量混合，再以 ddH<sub>2</sub>O 稀释成各 3  $\mu\text{M}$ 。以上 PCR 结束后，各孔加入 1  $\mu\text{L}$  的 3  $\mu\text{M}$  T#-F/T#-R(终浓度为 0.2  $\mu\text{M}$ )。继续进行以下 PCR 扩增反应（2 轮共 32~33 循环），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 程序

变性	96°C 10 s	17-18 cycles
退火/延伸	58°C 20 s	
延伸	68°C 5 s	
延伸	72°C 5 s	
延伸	68°C 5 s	
终延伸	70°C 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 PCR 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  $\mu\text{L}$ ）

2 × Taq mix	7.5 $\mu\text{L} \times n$
plate <sup>N</sup> -R (10 $\mu\text{M}$ ) （注：对于检测样本数小于 96 且使用 col <sup>N</sup> -F/row <sup>N</sup> -R 引物构建的文库，体系中不需	0.08 $\mu\text{L} \times n$ (终浓度 <u>0.05 <math>\mu\text{M}</math></u> )

加入 plate <sup>N</sup> -R)	
Lib-F, Lib-R (10 μM)	各 0.3 μL × <i>n</i> (终浓度分别 <u>0.2 μM</u> )
ddH <sub>2</sub> O	补足到 15.5 μL × <i>n</i>

注：plate<sup>N</sup>-R（和下面添加的 pos<sup>N</sup>-F）只在 PCR 前几个循环起到桥接目的片段和 Lib-F/Lib-R 的作用，要使用较低浓度（~0.05 μM），过高浓度会干扰后期循环 Lib-F/Lib-R 的正常扩增。

对于使用 pos<sup>N</sup>-F/plate<sup>N</sup>-R 添加特异性 barcode 的文库构建，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 μM pos<sup>N</sup>-F，加入到各 PCR 孔反应液中（每种 pos<sup>N</sup>-F 终浓度约 0.06 μM）。

对于使用 col<sup>N</sup>-F/row<sup>N</sup>-R 添加特异性 barcode 的文库构建，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 μM col<sup>N</sup>-F/row<sup>N</sup>-R 混合液，加入到各 PCR 孔反应液中（每种 col<sup>N</sup>-F 和 row<sup>N</sup>-R 终浓度约 0.06 μM）。

往第一轮 PCR 扩增产物加入 30 μL ddH<sub>2</sub>O（稀释 3×），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 45×）。

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°C 10 s	8 cycles
退火	57°C 15 s	
延伸	68°C 5 s	
延伸	72°C 10 s	
延伸	68°C 5 s	
变性	96°C 10 s	~8 cycles (可增减调整)
退火/延伸	68°C 10 s	
延伸	72°C 10 s	
延伸	68°C 5 s	
终延伸	72°C 1 min	1 cycle

注：由于第一轮 PCR 产物模板只稀释了几十倍，此轮扩增循环数不必过多。

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（1.2%）电泳进行检测。

将每块 96 孔板的建库产物为 1 组各取出 3 μL 混合成一亚组（~300 μL），使用 1.2% 琼脂糖凝胶电泳（约 20 min），切胶回收目的片段，使用 DNA 纯化试剂盒纯化产物。检测每亚组纯化产物的浓度后，将各亚组样本产物再等量混合（不足 96 样品的亚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。



除了以上的靶向编辑样品的测序建库，同样的方法也适合于自然变异位点分析，以及细胞器 RNA 编辑效率的分析（构建 cDNA 测序库）。

## 2、NGS 测序

将混合后的建库样本送测序公司，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 pos<sup>N</sup>-F 和 plate<sup>N</sup>-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 3000× ~ 30000×。

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、文件输入和 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。

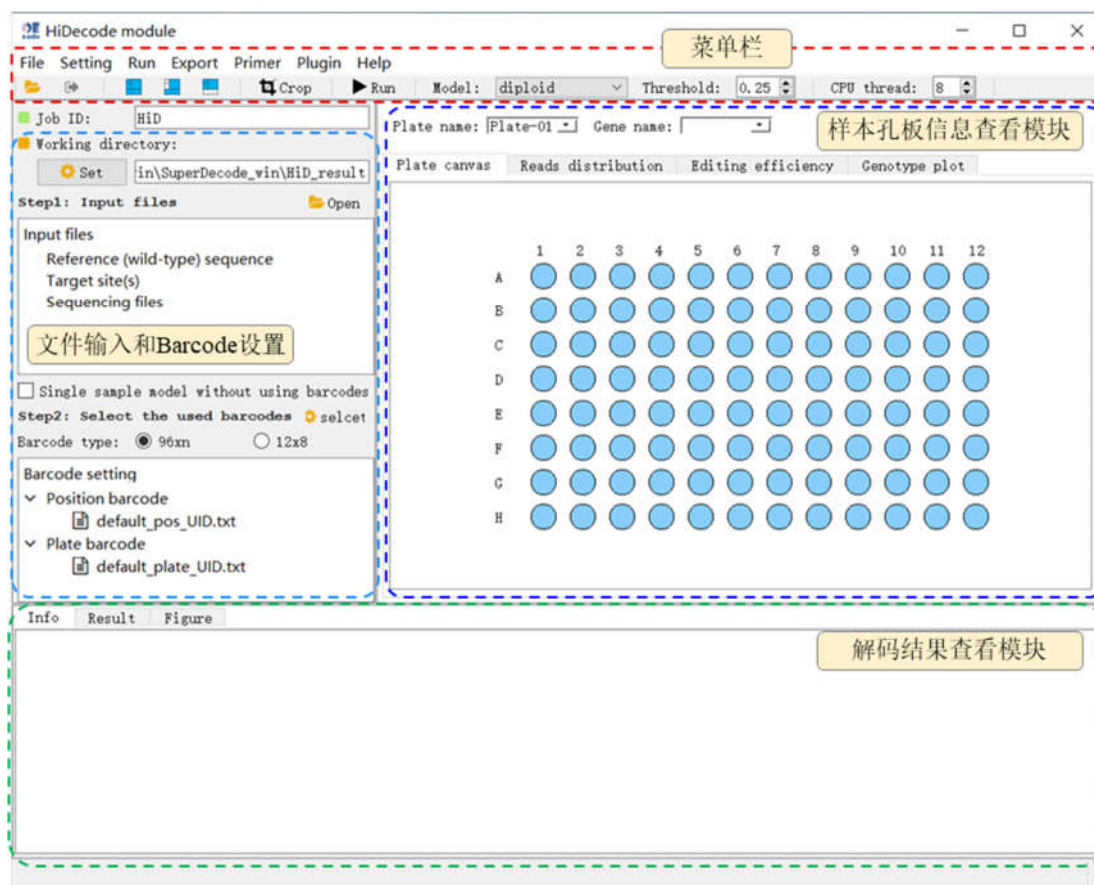


图 2-3 HiDecode 主界面

## (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) 通过点击菜单栏的“File”按钮或者左侧文件输入区的“Open”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 之间的受体品系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。

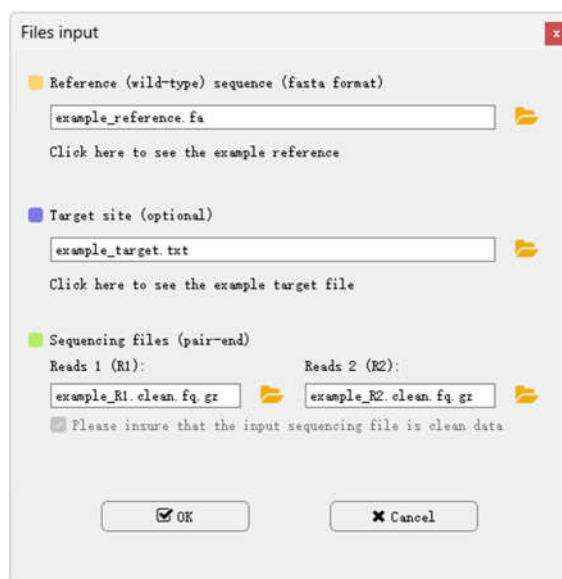


图 2-4 HiDecode 文件输入界面

ii) 根据文库构建方式, 在 barcode 设置区域中选择对应的 barcode 类型。使用  $\text{pos}^N\text{-F/plate}^N\text{-R}$  添加特异性 barcode 的文库, 选择  $96 \times n$  类型 barcode, 使用  $\text{col}^N\text{-F/row}^N\text{-R}$  添加特异性 barcode 的文库, 则选择  $12 \times 8$  类型 barcode。对于  $96 \times n$  类型 barcode, 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。对于  $12 \times 8$  类型 barcode, 通过双击左侧的 “default\_col\_UID.txt” 和 “default\_row\_UID.txt” 设置扩增时使用的列位置 barcode 和行位置 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, 预设 Threshold 为 0.25)、polyploid (用于分析多倍体样本数据, 预设 Threshold 为 0.1)、以及 low frequency (用于分析细胞系、原生质体、愈伤组织等多样低频变异样本数据, 预设 Threshold 为 0.0025)。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。

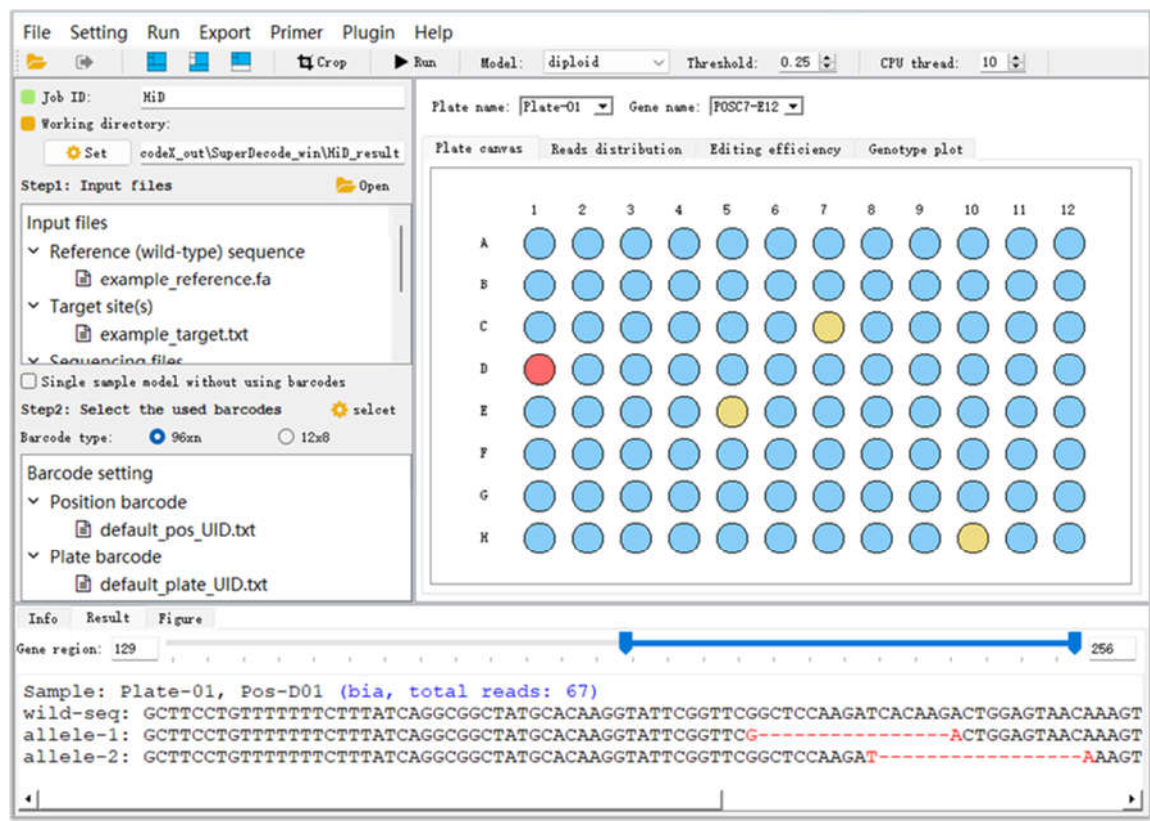


图 2-5 HiDecode 结果分析

同时，HiDecode 可通过交互模式查看总体分析结果。对于所选的板以及对应的基因，点击样本孔板信息模块上方的“Reads distribution”、“Editing efficiency”以及“Genotype plot”可分别显示该板中所有样本的 Reads 分布、突变频率以及基因型图等结果（图 2-6）。

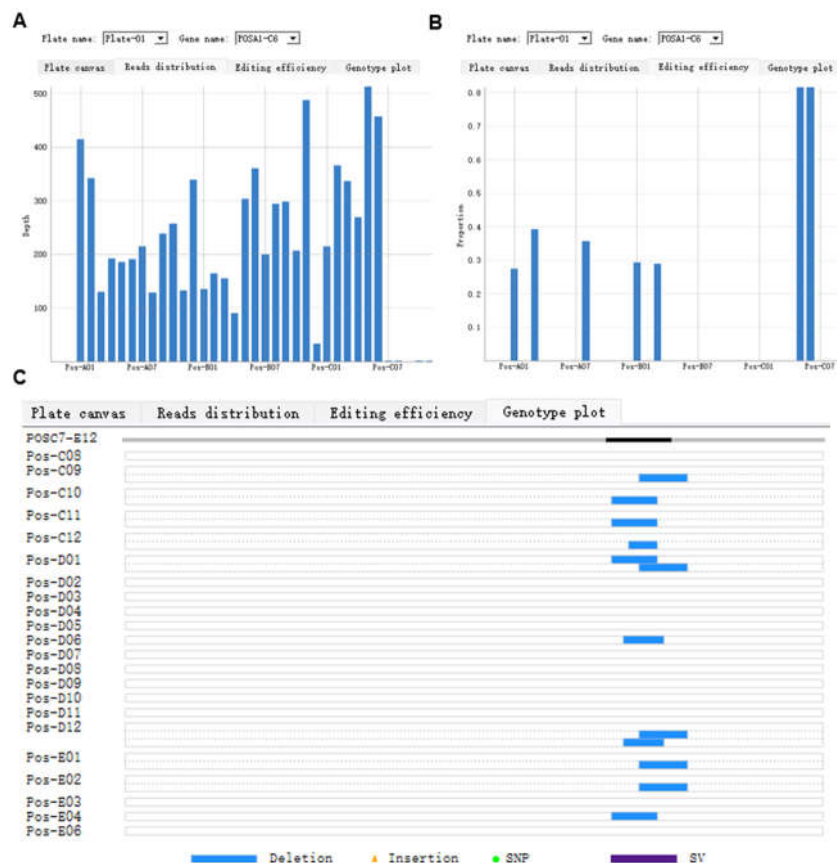
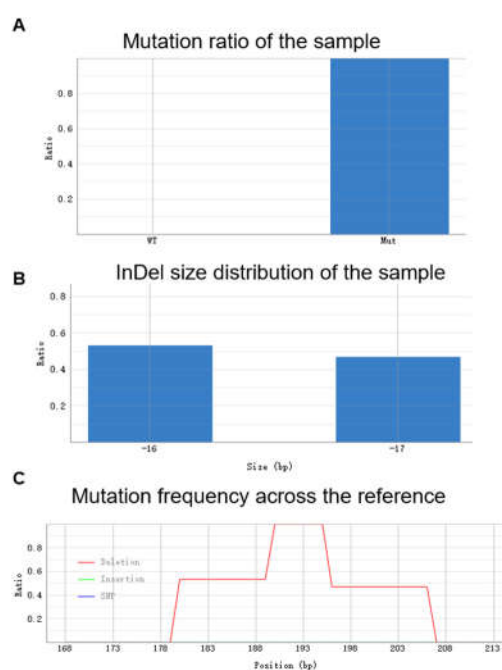


图 2-6 目标板中的 Reads 分布、突变频率以及基因型分析

对于选中的某个样本，点击解码查看模块上方的“Figure”按钮，可查看当前样本的突变频率、InDel 片段大小的分布以及所选样本在参考序列中每个位置的突变频率（图 2-7）。



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 2-8），“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 2-8 基于 HiDecode 解码输出的详细突变结果表

另外，“Export”中的“Export mutation genotypes 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 2-9）。



### i) 特异引物设计

21



primerDesign (<http://skl.scau.edu.cn/>)等工具。

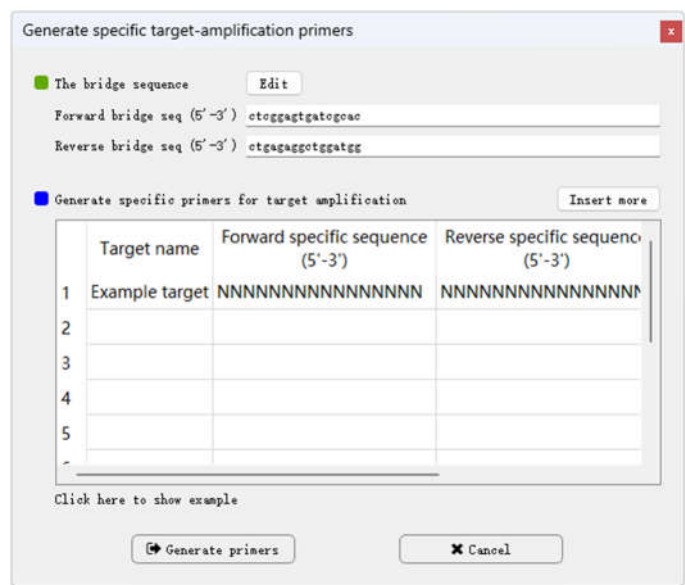


图 2-10 HiDecode 特异性引物设计输入界面

ii) 自定义 barcode 在扩增子中的位置以及序列

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。

对于使用 pos<sup>N</sup>-F/plate<sup>N</sup>-R 构建的文库，如果用户修改了 pos<sup>N</sup>-F 和 plate<sup>N</sup>-R 中 barcode 的位置或者序列长度，可以通过“Setting”下拉框的“Reset the default 96×n barcodes (define your own barcodes)”的对话框中重新设置 barcode 所在位置和长度（图 2-11）。

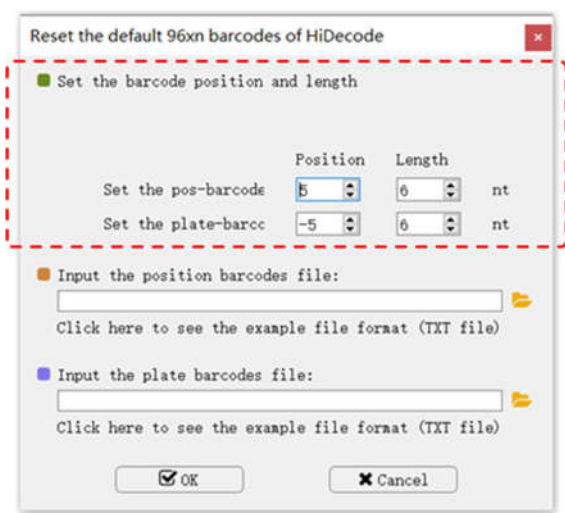


图 2-11 通过 barcode 设置框修改 96×n barcode 序列的位置和长度

虽然 HiDecode 提供了 96 个孔位置 barcode 和 96 个 PCR 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 2-12），或者点击菜单栏“Setting”下拉框的“Reset the default 96×n barcodes (define your own barcodes)”（批量修改），可以设置用户自己定义的 barcode。

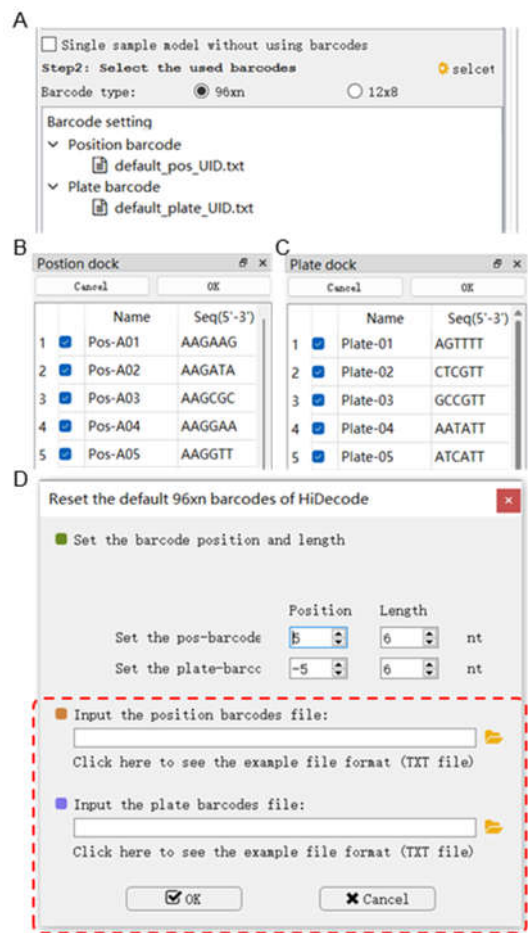


图 2-12 通过 barcode 选择框修改 96×n 的 barcode 序列

对于使用  $\text{col}^N\text{-F}/\text{row}^N\text{-R}$  引物构建的文库，用户同样可以通过“Setting”下拉框的“Reset the default 12×8 barcodes (define your own barcodes)”的对话框中重新设置 barcode 所在位置和长度（图 2-13）。

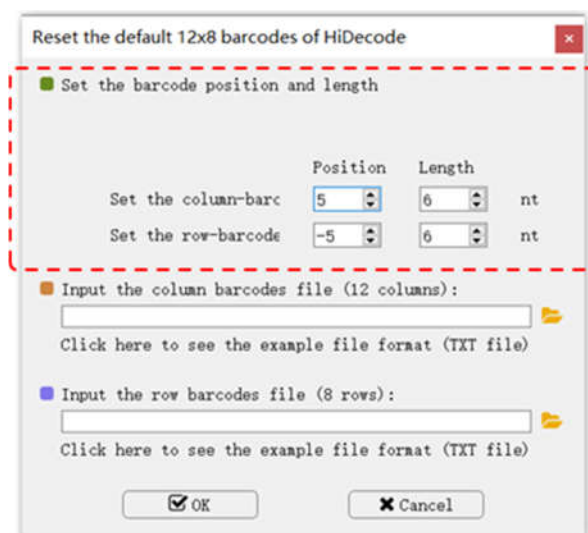


图 2-13 通过 barcode 设置框修改  $12 \times 8$  barcode 序列的位置和长度

用户同样可以根据自己需求重新定义 PCR 中使用的  $12 \times 8$  的 barcode 序列。打开 barcode 选择框后，通过双击各 barcode 序列（图 2-14），或者点击菜单栏“Setting”下拉框的“Reset the default  $12 \times 8$  barcodes (define your own barcodes)”（批量修改），可以设置用户自己定义的 barcode。

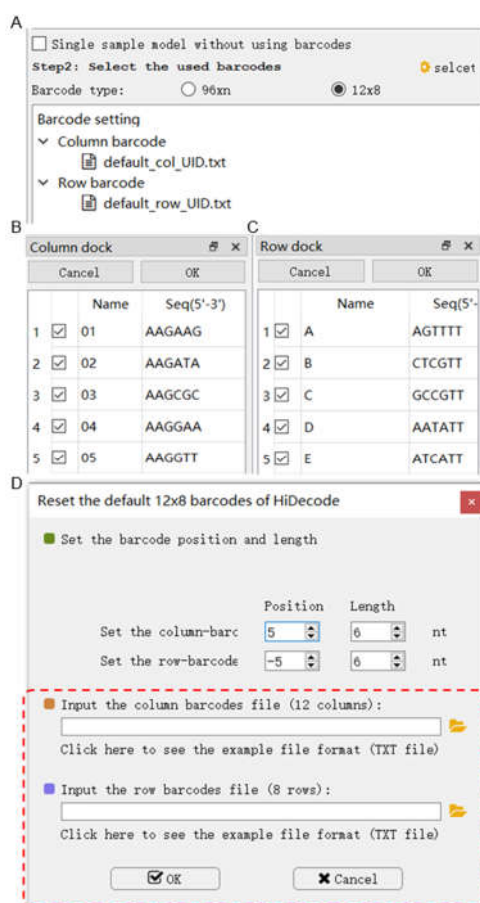


图 2-14 通过 barcode 选择框修改  $12 \times 8$  的 barcode 序列

HiDecode 可用于所有 PCR 扩增子建库的二代测序数据分析，例如：使用 HiDecode 分析 Hi-TOM 数据：

Hi-TOM 使用  $12 \times 8$  的建库模式，且 barcode 序列长度均为 4 nt，因此需在 HiDecode 中设置相应的参数。首先在 barcode 设置区域中选择  $12 \times 8$  的 Barcode type，随后点击菜单栏中“Setting”键下的“Reset the default  $12 \times 8$  barcodes (define your own barcodes)”设置 barcode 的序列及 barcode 长度（图 2-15）。

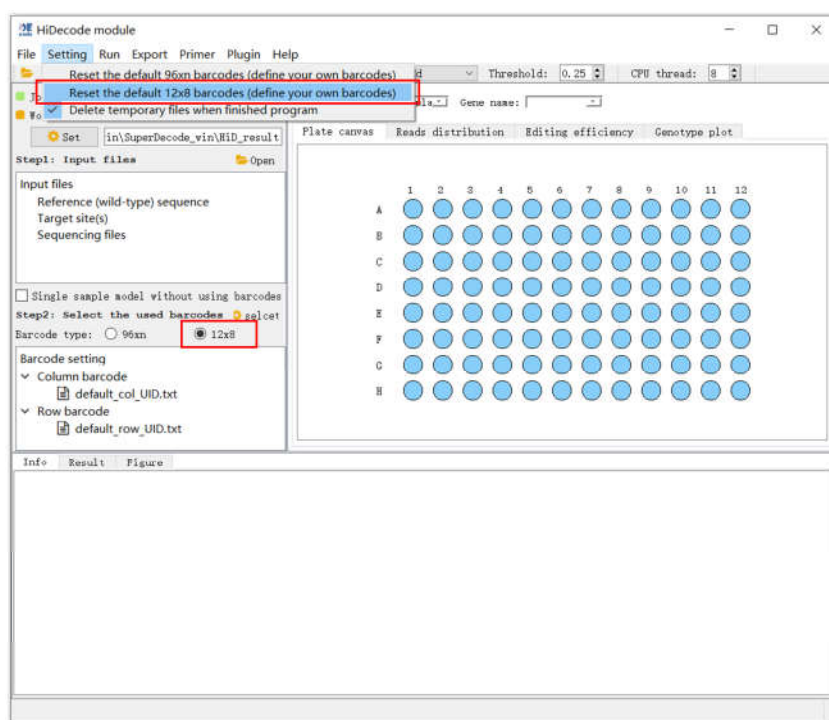


图 2-15 HiDecode 中设置 Hi-TOM 数据的 barcode 类型以及 barcode 序列

点击“Reset the default  $12 \times 8$  barcodes (define your own barcodes)”后，在 barcode 设置框将 barcode 长度设置为 4，并输入 Hi-TOM 数据使用的 column 和 row barcode 序列文件（图 2-16）

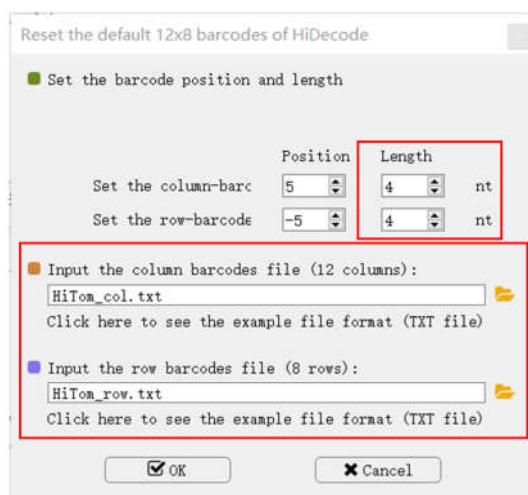


图 2-16 通过 barcode 设置框设置 Hi-TOM 的 barcode 序列

点击文件输入区域中的“Open”键，打开文件输入对话框，并输入参考序列、靶点序列以及测序文件（其中参考序列和测序文件必须输入，靶点序列文件可不输入），根据测序样本设置对应参数后，点击“Run”键开始分析 Hi-TOM 数据（图 2-17）。

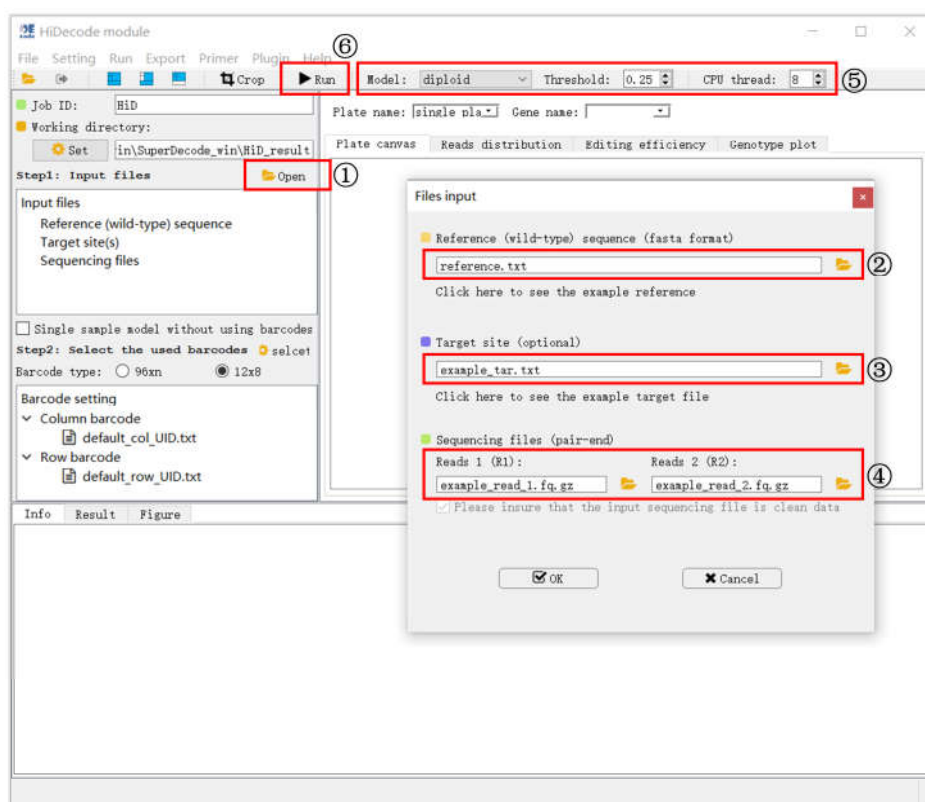


图 2-17 HiDecode 分析 Hi-TOM 数据的文件输入流程

HiDecode 分析结束后可点击对应样本查看分析结果，或点击菜单栏中的“Export”键输出全部样本的分析结果。

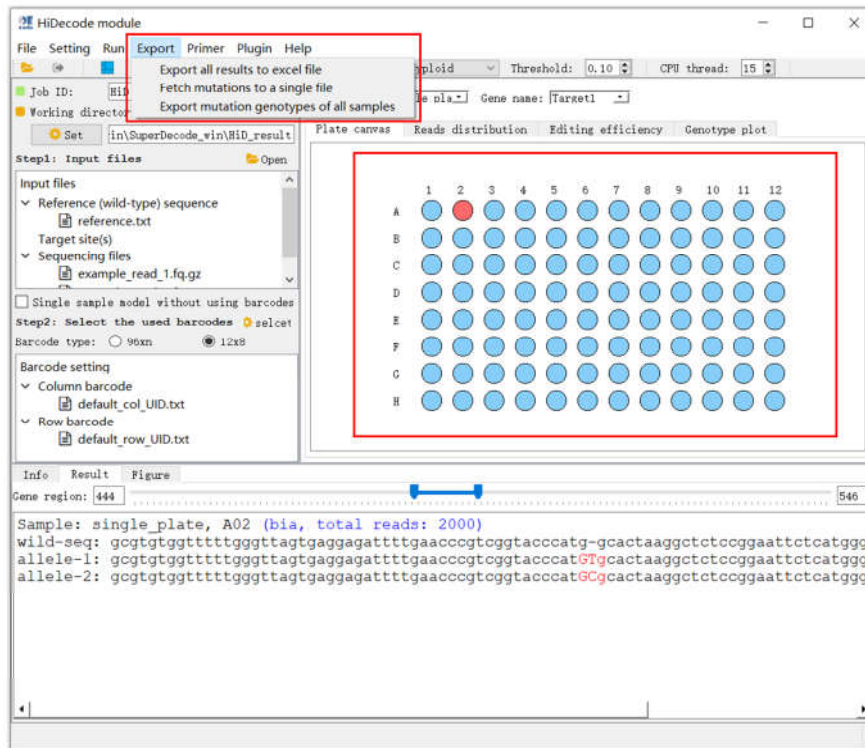


图 2-18 Hi-TOM 数据分析结果查看以及输出

结果比较：使用 Hi-TOM 和 HiDecode 对 Hi-TOM 提供的测试数据进行解码分析，结果显示只有 A01、B05、G10 和 H12 样本的结果有差异，其余样本分析的基因型结果一致（图 2-19）。

A									
Sort	Reads	mut Ratio	Left variant	Left variant	Right variant	Right variant	Left reads	Right reads	seq
A01	1	886	74.83%	WT	WT	-	-	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	2	166	14.02%	WT	WT	-	-	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	3	132	11.15%	WT	WT	-	-	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
A02	1	451	35.68%	II	II	C	C	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	2	442	34.97%	II	II	T	T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	3	125	9.89%	II	II	T	T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	4	92	7.28%	II	II	C	C	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	5	90	7.12%	II	II	C	C	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	6	64	5.06%	II	II	T	T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
A03	1	649	48.15%	SNP	SNP	C->T	C->T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	2	372	27.60%	WT	WT	-	-	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	3	111	8.23%	SNP	SNP	C->T	C->T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	4	82	6.08%	WT	WT	-	-	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	5	74	5.49%	SNP	SNP	C->T	C->T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
A04	1	1198	87.38%	II	II	A	A	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	2	105	7.66%	II	II	A	A	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
A05	1	571	45.61%	II	II	T	T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	2	539	43.05%	II	II	G	G	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
	3	81	6.47%	II	II	T	T	TTTGGGT	TTTGGGTAGTGAGGAGATTITG
B									
single_plate_A01		chi	316	(515 bp)	ctcatgggtgagtcaggtttgtttgtctccatagttccggtcg				
		1 wt	235 (0.80696)	(515 bp)	ctcatgggtgagtcaggtttgtttgtctccatagttccggtcg				
		2 545:ct>C	26 (0.08228)	(515 bp)	ctcatgggtgagtcaggtttgtttgtctccatagttccggtcg				
		3 549:at>A	19 (0.06013)	(515 bp)	ctcatgggtgagtcaggtttgtttgtctccatagttccggtcg				
		4 548:ca>C	16 (0.05063)	(515 bp)	ctcatgggtgagtcaggtttgtttgtctccatagttccggtcg				
single_plate_A02		bia	2000	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		1 494:p>GT	1021 (0.51050)	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		2 494:p>GC	973 (0.48650)	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
single_plate_A03		bet	1133	(468 bp)	agattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		1 498:C>T	1000 (0.88261)	(468 bp)	agattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		2 wt	133 (0.11739)	(468 bp)	agattttgaacccgtcggtaccctatg-gcactaaggctctcc				
single_plate_A04		hom	2000	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		1 494:p>GA	1985 (0.99250)	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
single_plate_A05		bia	1995	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		1 494:p>GT	981 (0.49173)	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				
		2 495:p>GC	969 (0.48571)	(464 bp)	gaggaattttgaacccgtcggtaccctatg-gcactaaggctctcc				

图 2-19 Hi-TOM 与 HiDecode 的分析结果比较

iii) 在原始测序文件去除接头和过滤

大多数测序公司会提供去除接头和过滤后的 Clean reads 文件。如果提供的是 Raw data 数据，也可以使用 HiDecode 的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 2-20），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。

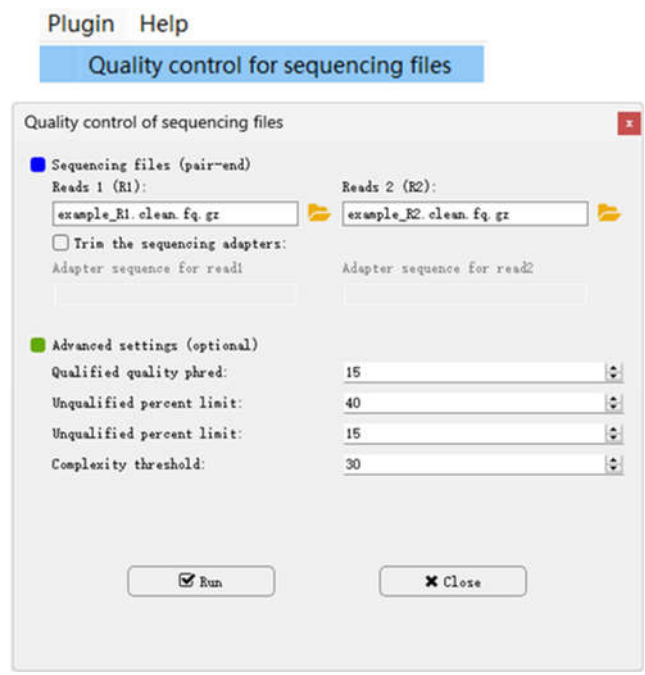


图 2-20 HiDecode 质控以及接头去除设置界面

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，点击文件输入模块下方的“Single sample model without using barcodes”即可打开该功能（图 2-21）。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。



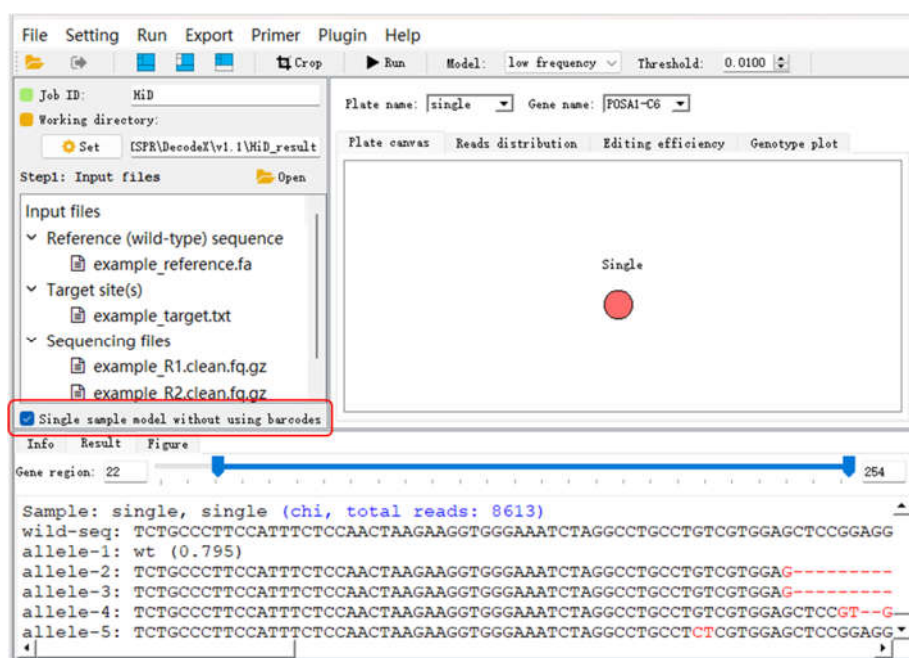


图 2-21 HiDecode 单样本文库突变分析

#### 4、使用网络版在线工具中的 HiDecode 模块分析突变

##### (1) 网络版 HiDecode 界面介绍

HiDecode 网络版主界面包括：文件输入模块和 Barcode 设置模块（图 2-22）。其中，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 设置模块用于设置使用的孔和板信息，包括 Barcode 序列、Barcode 序列的长度和 Barcode 在扩增子中的位置信息。



**SuperDecode**

Home DSDecodeM HiDecode LaDecode Document

HiDecode: decoding mutations from short-read sequencing data of pooled amplicons.

HiDecode is designed for large-scale mutation identification from one or two target-site-containing amplicons in a large number of samples using NGS. Various small mutations, including chimeric mutations, can be analyzed by HiDecode in diverse materials, such as diploid and polyploid organisms, culture cell lines, protoplasts, and calli. It not only can detect mutations from pooled amplicons, but also can identify mutations from amplicon of single sample. For testing the program, click [download](#) to obtain the test files.

**Input files**

Step 1: input reference sequence in fasta format

>seq\_name  
reference sequence

Or input a reference file:  未选择任何文件

Step 2: input target sequence file (optional):  未选择任何文件

Step 3: input paired sequencing file (in #.fq.gz format)

Read1 file:  未选择任何文件 Read2 file:  未选择任何文件

☐ Single sample (not a pooled amplicon)

Step 4: threshold setting <sup>?</sup>:  (0.001-1.00)

**Barcode settings**

☒ Using default barcodes of HiDecode

Select position barcode:  Select plate barcode:

☐ Define your barcode

Upload position barcode file:  未选择任何文件 Upload plate barcode file:  未选择任何文件

[Example\\_position\\_barcode.txt](#) [Example\\_plate\\_barcode.txt](#)

Position-barcode position:  Plate-barcode position:

Position-barcode length:  Plate-barcode length:

Your email (optional):

\*In case the webpage is closed during the program's execution, it is recommended to enter a valid email address to receive the analysis results.

图 2-22 HiDecode 网络版在线工具主界面

## (2) 基本分析操作

使用 HiDecode 网络版在线工具从 NGS 扩增子文库分析靶点突变序列的主要步骤包括文件输入、样本分析模式选择、参数设置和解码。详细操作如下：

i) 在文件输入模块中依次输入：**Fasta** 格式的参考序列或包含参考序列的文件（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）。

ii) 在文件输入模块中选择样本分析模式，如果为单个样本的文库分析，则点击“Single sample (not a pooled amplicon)”启动单样本文库分析模式。根据样本类型在文件输入模块中设置突变检测阈值 Threshold。

iii) 如果为多样本扩增子混合文库，在上传相关文件以及设置突变检测阈值 Threshold 后，需在参数设置模块中设置 Barcode 信息，包括 Barcode 序列、Barcode 序列的长度和 Barcode 在扩增子中的位置信息。如果为单样本文库的突变分析则不需设置 Barcode 信息。

iv) 点击“Run HiDecode”按钮启动分析任务。（注意：在启动分析前可在 email 输

入框中输入接收解码结果的邮箱，防止解码过程中网页关闭导致结果丢失）。

程序运行结束后，解码结果会跳转至新的网页中显示，可以通过鼠标选择相应样本查看对应解码分析结果，包括当前选择样本的序列与参考序列的比对情况、当前样本的总体突变频率、Indel 片段大小的分布以及所选样本在参考序列中每个位置的突变频率（图 2-23）。解码分析结果可点击网页下面的下载按钮进行保存。

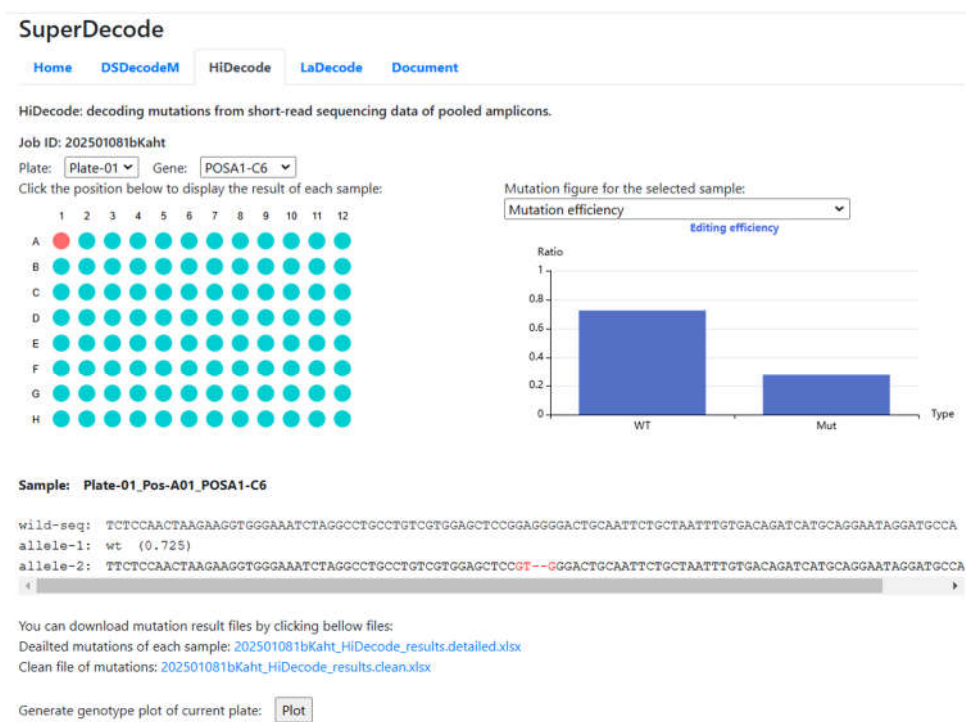


图 2-23 网络版 HiDecode 解码分析结果

#### 四、LaDecode 长片段扩增子混合文库的制备和分析

对特定目标序列区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 TGS 三代测序（单分子测序）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 barcode 序列设置得更长；三代测序的接头比较特殊，由测序公司完成扩增子加接头的操作。LaDecode 也同样支持分析用户自定义 barcode 建成的文库，以及分析单个样本的扩增子文库。

### (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 序列或列位置 barcode 和行位置 barcode 序列, 以便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, 用于区分不同样本的 pos<sup>N</sup>-F/col<sup>N</sup>-F 和 plate<sup>N</sup>-R/row<sup>N</sup>-R (超过 96 个样本的文库构建, 使用区分 96 孔 PCR 板的 plate<sup>N</sup>-R 和不同孔位置的 pos<sup>N</sup>-F, 少于 96 个样本的文库构建, 可选择使用区分 96 孔 PCR 板不同行位置的 row<sup>N</sup>-R 和不同列位置的 col<sup>N</sup>-F)。

### i) 特异引物 Target-F 和 Target-R 的设计

在第一个靶点上游约 100~200 bp 位置设计靶序列特异引物 Target-F, 在最后一个靶点下游约 100~200 bp 位置设计靶序列特异性引物 Target-R, 扩增片段约 1~10 kb (最佳 2~6 kb)。引物的位点特异配对部分 (18~22 nt, T<sub>m</sub> 值为 60°C~63°C), 需使用 NCBI 中的 Primer-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'- atcgcctggctccacgctccgagttNNNNNNNNNNNNNNNNNNNNNNNN-3' (NNN...为 18~22 nt 位点特异性序列, T<sub>m</sub> 值为 58°C~63°C; 小写字母为与 pos<sup>N</sup>-F 引物特异性配对的桥接序列)。

特异引物 Target-R: 5'- tacgctggctccacgtccctagaNNNNNNNNNNNNNNNNNNNNNNNNNN-3' (NNN...为为 18~22 nt 位点特异性序列, Tm 值为 58°C~63°C; 小写字母为与 plate<sup>N</sup>-R 引物特异性配对序列)。

注：参考 STI PCR 方法，在 Target-F 和 Target-R 的 5' 端设置相同序列（下划线），在扩增产物变性单链时形成反向互补关系。对于引物二聚体和较短的非特异扩增产物，两末端距离近较容易配对形成茎环结构，从而抑制此类产物的扩增，但不影响较长的特异产物的扩增，最终提高扩增特异性和效率（参考 Zhao et al., 2022, Molecular Plant））

ii) 通用引物 pos<sup>N</sup>-F/col<sup>N</sup>-F 和 plate<sup>N</sup>-R/row<sup>N</sup>-R 的设计

对于检测样本数大于 96 的文库，使用 **pos<sup>N</sup>-F** 和 **plate<sup>N</sup>-R** 在每个样本两端加入特异性 barcode，LaDecode 中内置的 **pos<sup>N</sup>-F** 和 **plate<sup>N</sup>-R** 的序列结构如下：

pos<sup>N</sup>-F: 5'- cttgNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAtcgctggctccacgctccgagtt -3'  
(NNN...为 24-nt 的孔位置 barcode 序列, 用于指示各样品对应 96 孔 PCR 板上的孔位置)

plate<sup>N</sup>-R: 5'-cgacNNNNNNNNNNNNNNtacgctggctccacgctccctaga-3' (NNN...为 12-nt 板 barcode 序列, 用于指示不同的 PCR 板编号, 一个 PCR 板对应 1 条 plate-N 引物)

将 96 条 pos<sup>N</sup>-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。

对于检测样本数小于 96 的文库构建，可选择使用 col<sup>N</sup>-F 和 row<sup>N</sup>-R 在每个样本两端加入特异性 barcode，LaDecode 中内置的 col<sup>N</sup>-F 和 row<sup>N</sup>-R 的序列结构如下：

col<sup>N</sup>-F: 5'- cttgNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNtgcgctggctccacgctccgagtt -3'  
(NNN...为 24-nt 的列位置 barcode 序列, 用于指示各样品对应 96 孔 PCR 板上的列位置, 单个文库构建最多使用 12 条 col<sup>N</sup>-F)

row<sup>N</sup>-R: 5'-cgacNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNtaccgctggctccacgtccctaga-3' (NNN...为 24-nt 的行位置 barcode 序列，用于指示各样品对应 96 孔 PCR 板上的行位置，单个文库构建最多使用 8 条 col<sup>N</sup>-F)

将 12 条 col<sup>N</sup>-F (6.0 μM) 和 8 条 row<sup>N</sup>-R (6.0 μM) 各吸 10 μL 至 96 孔 PCR 板对应的位置中并混匀, 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。

LaDecode 中内置的 pos<sup>N</sup>-F/col<sup>N</sup>-F 和 plate<sup>N</sup>-R/row<sup>N</sup>-R 具体序列见附表 2-LaDecode 所有引物序列.xlsx

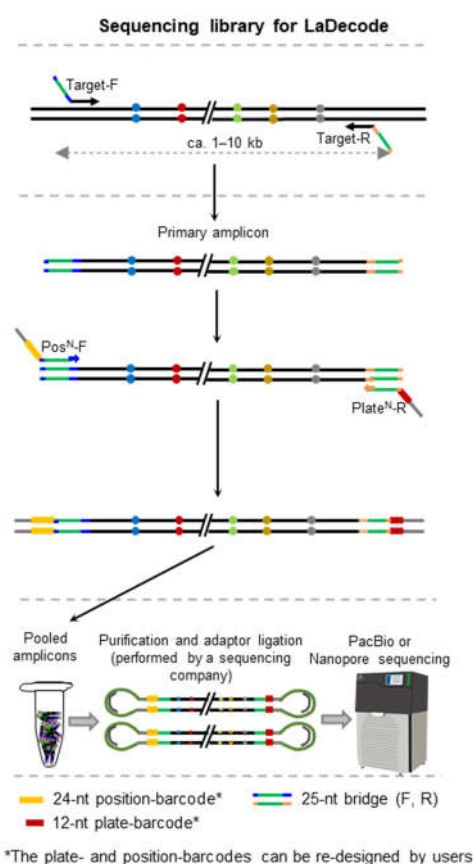


图 3-1 LaDecode 长片段扩增子文库的构建流程

## (2) 扩增步骤

对于使用 pos<sup>N</sup>-F 和 plate<sup>N</sup>-R 引物构建的文库，将每种 LaDecode 的 pos<sup>N</sup>-F 中 1 管配制成 3 μM 溶液，各取 20 μL 加入到一块 96 孔 PCR 板对应的孔即为 3 μM pos<sup>N</sup>-F 工作液，冷冻保存。

对于使用 col<sup>N</sup>-F 和 row<sup>N</sup>-R 引物构建的文库，将每种 LaDecode 的 col<sup>N</sup>-F 和 row<sup>N</sup>-R 配制成 6 μM 溶液，各取 10 μL 加入到一块 96 孔 PCR 板对应的孔混匀，即为 3 μM col<sup>N</sup>-F/row<sup>N</sup>-R 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（每反应 15 μL）

2 × 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 Mix (10 mM)	0.3 μL × $n$
Phanta Max Super-Fidelity DNA Polymerase (1U/μL)	0.3 μL × $n$
ddH <sub>2</sub> O	补足到 15.5 μL × $n$

注：扩增 5kb 以下片段使用较便宜的 Phanta；扩增 5 kb 以上的片段建议使用高保真度高性能酶，如艾科瑞 ApexHF HS DNA Polymerase CL 或东洋纺 KOD FX Neo 等。

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°C 2 min	30~32 super-cycles
变性	96~98°C 30 s	
退火	60°C 20 s	

延伸	$(65^{\circ}\text{C } 10 \text{ s}, 68^{\circ}\text{C } 10 \text{ s}, 72^{\circ}\text{C } 15 \text{ s}) \times n$ (根据片段长度设置 <u>内循环数</u> $n$ , $\sim 1 \text{ kb}/n$ )	
终延伸	$70^{\circ}\text{C } 2 \text{ min}$	

注：使用 STI PCR 方法 (Zhao et al., 2022 Mol Plant) 的链延伸变温内循环可提高扩增特异性和效率；序列 GC 含量较高时变性温度  $97\sim 98^{\circ}\text{C}$ 。

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。

#### ii) 第二轮 barcoding PCR 扩增

将第一轮 PCR 扩增产物稀释 3 倍 (加入  $30 \mu\text{L ddH}_2\text{O}$ )。根据样本数配制第二轮 PCR 总混合液 (表 3-3) (样本数为  $n$ )，将 PCR 反应总混合液分装  $15 \mu\text{L}$  至 96 孔 PCR 板中。

表 3-3 LaDecode 第二轮 PCR 反应总混合液体系

$2 \times \text{Phanta Max Buffer}$	$7.5 \mu\text{L} \times n$
$\text{plate}^{\text{N-R}} (10 \mu\text{M})$ (注: 对于检测样本数 小于 96 且使用 $\text{col}^{\text{N-F}}/\text{row}^{\text{N-R}}$ 引物构建的 文库, 体系中不需加入 $\text{plate}^{\text{N-R}}$ )	$0.3 \mu\text{L} \times n$ (终浓度 $0.2 \mu\text{M}$ )
dNTPs Mix (10 mM)	$0.3 \mu\text{L} \times n$
Phanta Max Super- Fidelity DNA Polymerase (1U/ $\mu\text{L}$ )	$0.3 \mu\text{L} \times n$
ddH <sub>2</sub> O	补足到 $15.5 \mu\text{L} \times n$

对于使用  $\text{pos}^{\text{N-F}}/\text{plate}^{\text{N-R}}$  构建的文库, 使用 96 孔针取样器或者多通道移液器从分装在 96 孔板的  $3 \mu\text{M pos}^{\text{N-F}}$  取  $\sim 1 \mu\text{L}$  加入到 PCR 反应液中 (终浓度  $\sim 0.2 \mu\text{M}$ )。

对于使用  $\text{col}^{\text{N-F}}/\text{row}^{\text{N-R}}$  构建的文库, 使用 96 孔针取样器或者多通道移液器从分装在 96 孔板的  $\text{col}^{\text{N-F}}/\text{row}^{\text{N-R}}$  工作液取  $\sim 1 \mu\text{L}$  加入到 PCR 反应液中 ( $\text{col}^{\text{N-F}}$  和  $\text{row}^{\text{N-R}}$  终浓度均为  $\sim 0.2 \mu\text{M}$ )。

取稀释的第一轮 PCR 产物  $\sim 1 \mu\text{L}$  加入到 PCR 反应液中 (最终稀释约  $45\times$ ), 并进

行 PCR 扩增，PCR 反应程序如表 3-4。

表 3-4 LaDecode 第二轮 PCR 反应程序

变性	96~98°C 30 s	10~12 super-cycles
退火	62°C 20 s	
延伸	(65°C 10 s, 68°C 10 s, 72°C 15 s) × <i>n</i> (根据片段长度设置内循环数 <i>n</i> , ~1 kb/ <i>n</i> )	
终延伸	70°C 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 super-cycles，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%）电泳较短时间（约 20 min），切胶回收目的片段，以纯化试剂盒产物。多靶点编辑产生的突变类型较多，其中包含大片段删除,为保证回收所有不同长度的片段（主要是去除残留引物），电泳时间不要过长。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，推荐每 96 样品的 1–4 kb、4–7 kb 和 7–10 kb 序列分别测~1 Gb、2 Gb 和 3 Gb 的数据量。

## 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、文件输入和 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。



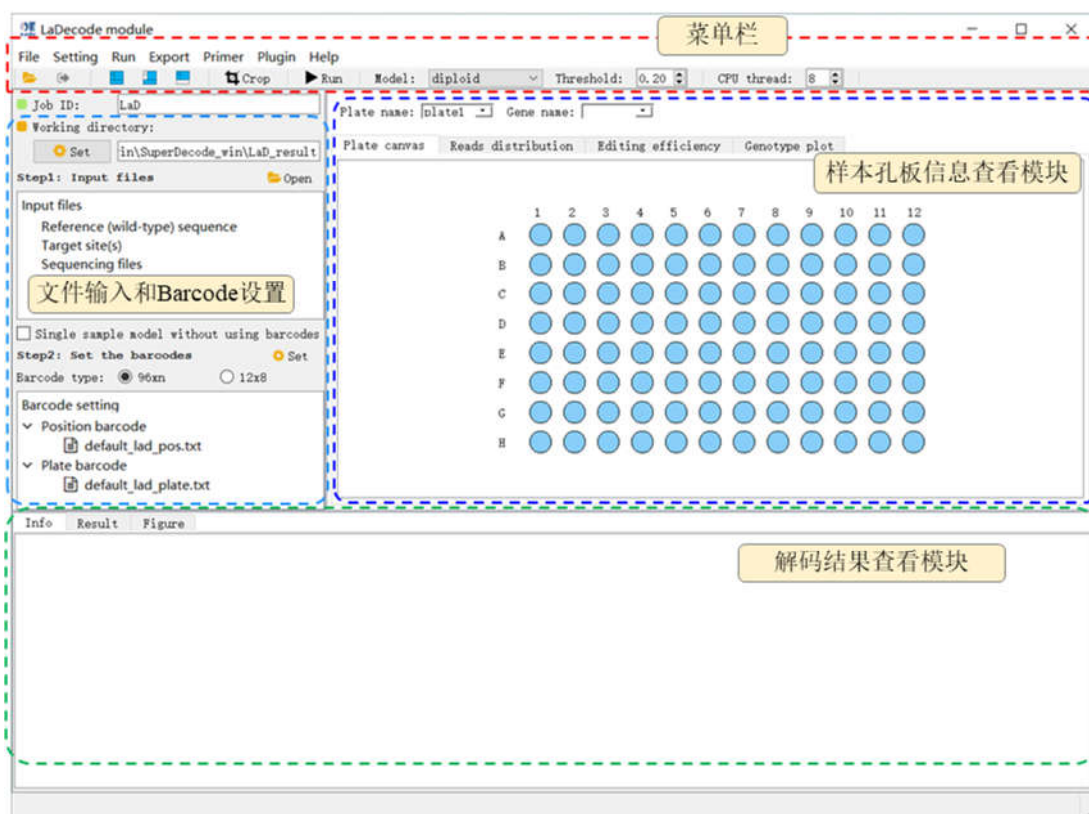


图 3-2 LaDecode 主界面

## (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“File”按钮或文件输入框中的“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。

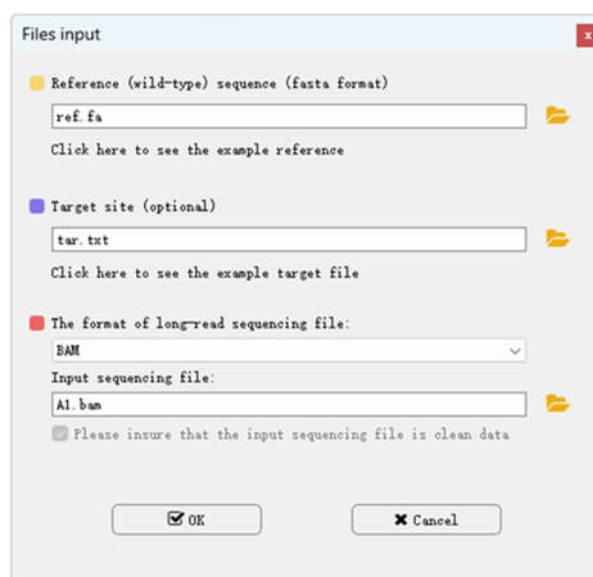


图 3-3 LaDecode 文件输入界面

ii) 根据文库构建方式, 在 barcode 设置区域中选择对应的 barcode 类型。使用  $\text{pos}^{\text{N}}\text{-F/plate}^{\text{N}}\text{-R}$  添加特异性 barcode 的文库, 选择  $96 \times n$  类型 barcode, 使用  $\text{col}^{\text{N}}\text{-F/row}^{\text{N}}\text{-R}$  添加特异性 barcode 的文库, 则选择  $12 \times 8$  类型 barcode。对于  $96 \times n$  类型 barcode, 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位置 barcode 和板编号 barcode。对于  $12 \times 8$  类型 barcode, 通过双击左侧的“default\_col\_UID.txt”和“default\_row\_UID.txt”设置扩增时使用的列位置 barcode 和行位置 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。

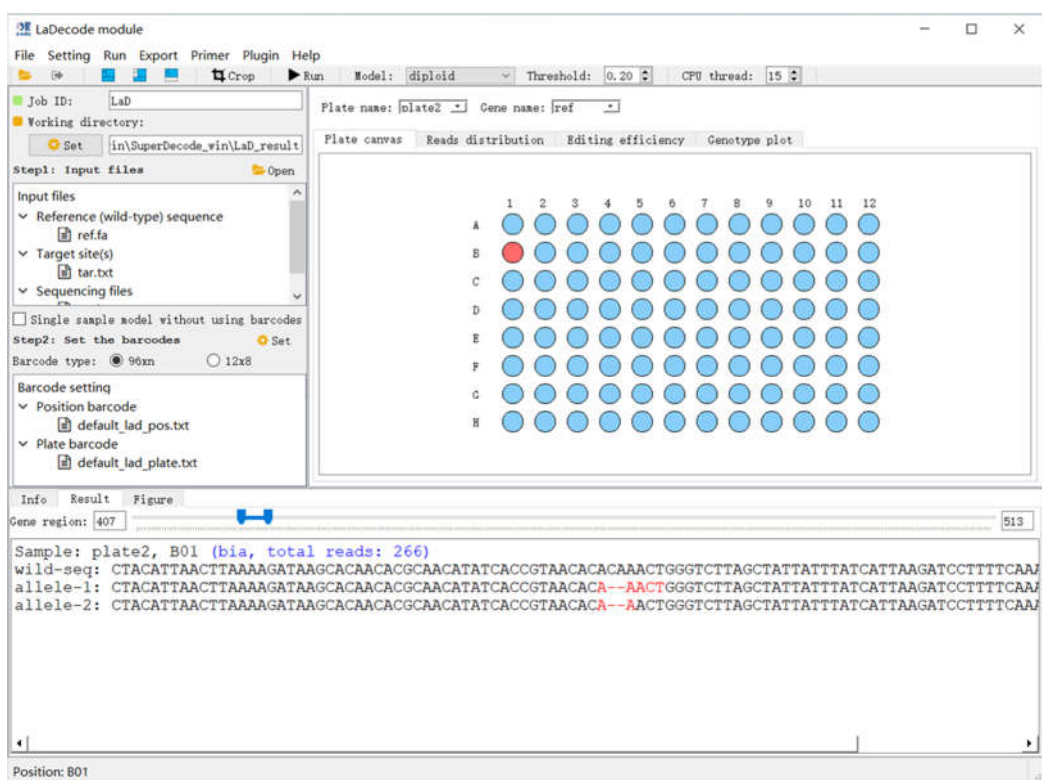


图 3-4 LaDecode 结果分析

LaDecode 可通过交互模式查看总体分析结果。对于所选的板以及对应的基因，点击样本孔板信息模块上方的“Reads distribution”、“Editing efficiency”以及“Genotype plot”可分别显示该板中所有样本的 Reads 分布、突变频率以及基因型分析结果（图 3-5）。

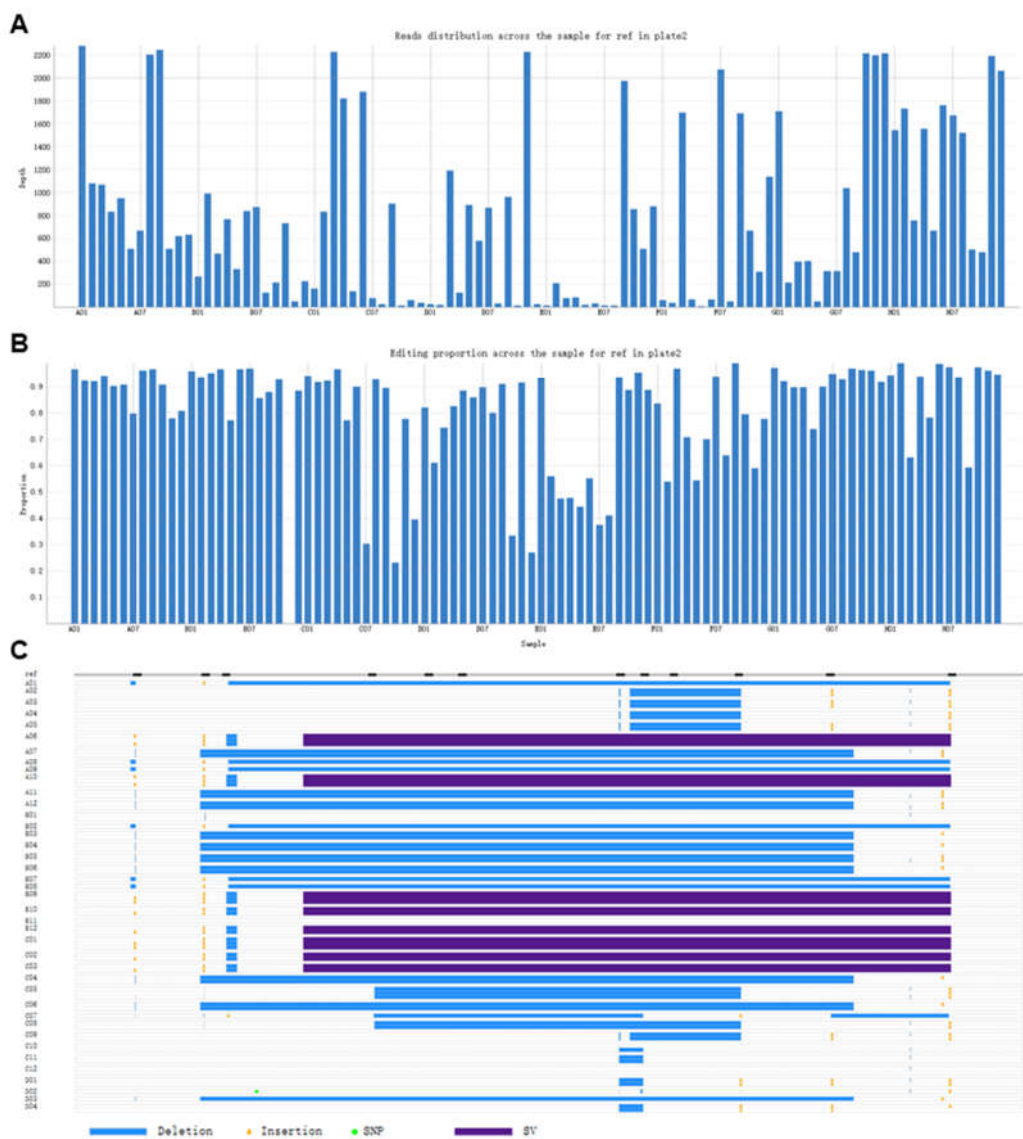


图 3-5 基于 LaDecode 解码的样本 Reads 分布、突变频率以及基因型分析

对于选中的某个样本，点击解码查看模块上方的“Figure”按钮，可查看当前样本的总体突变频率、Indel 片段大小的分布以及所选样本在参考序列中每个位置的突变频率（图 3-6）。

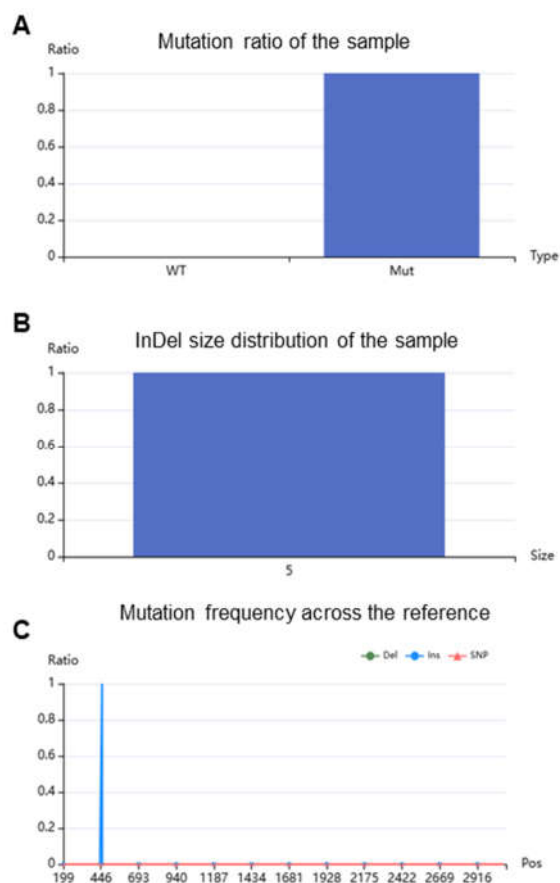


图 3-6 基于 LaDecode 解码的样本突变频率、Indel 片段大小以及突变位置频率统计

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 3-7），“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。



进行分析的插件工具，点击菜单栏“Plugin”中的“Mutation 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 3-9），输出的结果文件自动保存至 reads 文件所在目录。

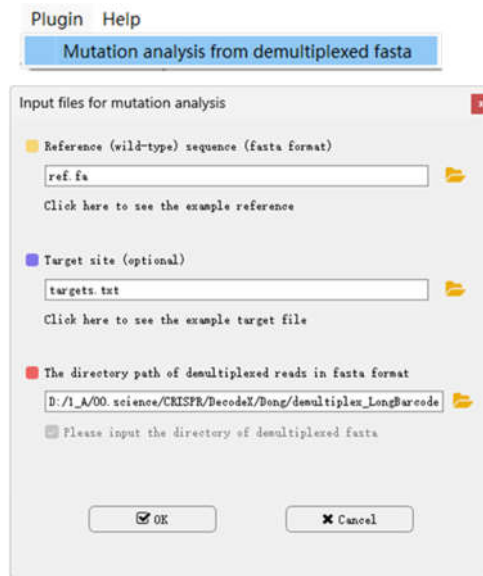


图 3-9 基于 LaDecode 分析已拆分的测序数据

## 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 3-10）。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。



Generate specific target-amplification primers

☒ The bridge sequence Edit

Forward bridge seq (5'-3') atcgctggttcacgctcagtt

Reverse bridge seq (5'-3') tgcgctggttcacgctcagtt

☒ Generate specific primers for target amplification Insert more

	Target name	Forward specific sequence (5'-3')	Reverse specific sequence (5'-3')
1	test-F	TCCCTTCATTCCGATGAG	CTGAGCGTATCTGAGCATG
2			
3			
4			
5			
6			

[Click here to show example](#)

Generate primers Cancel

图 3-10 LaDecode 引物设计输入界面

### iii) 自定义 barcode

对于使用  $\text{pos}^{\text{N}}\text{-F}/\text{plate}^{\text{N}}\text{-R}$  构建的文库，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 3-11），或者点击菜单栏“Setting”下拉框的“Reset the default  $96 \times n$  barcodes (define your own barcodes)”（批量修改，并保存为默认 barcode），可以设置用户自己定义的 barcode。

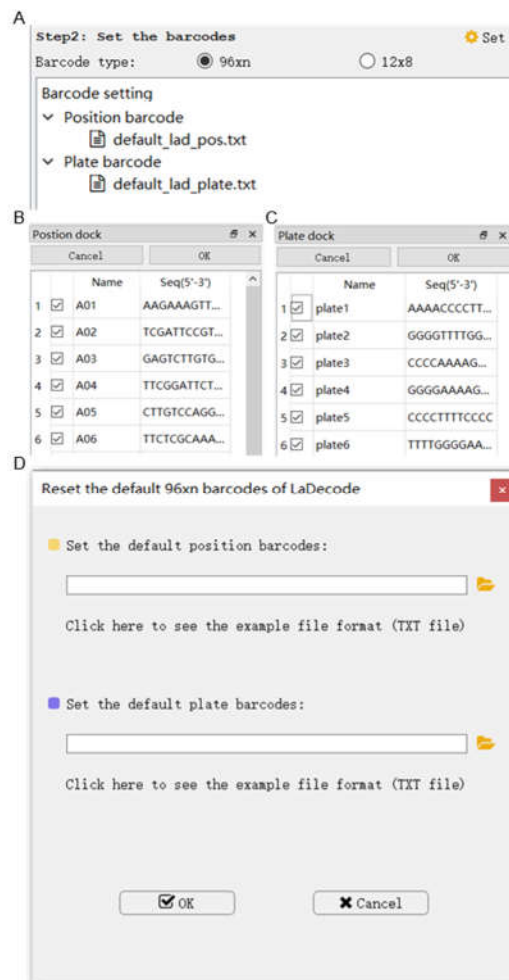


图 3-11 LaDecode 中 96×n 文库的 barcode 定义界面

对于使用  $\text{col}^N\text{-F}/\text{row}^N\text{-R}$  构建的文库，LaDecode 中内置了 12 条 column barcodes 以及 8 条 row barcodes 序列，但是用户同样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 3-12），或者点击菜单栏“Setting”下拉框的“Reset the default 12×8 barcodes (define your own barcodes)”（批量修改，并保存为默认 barcode），可以设置用户自己定义的 barcode。

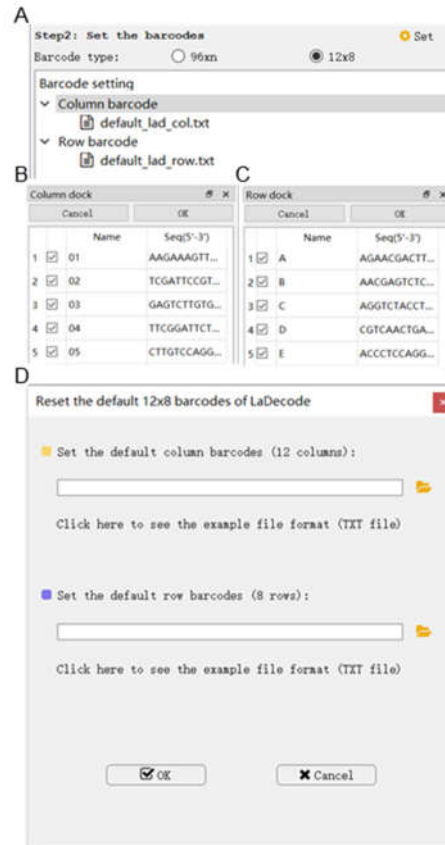


图 3-12 LaDecode 中 12×8 文库的 barcode 定义界面

#### 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，LaDecode 同样提供了从单个样本的文库分析突变的功能，点击文件输入模块下方的“Single sample model without using barcodes”即可打开该功能（图 3-13）。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。

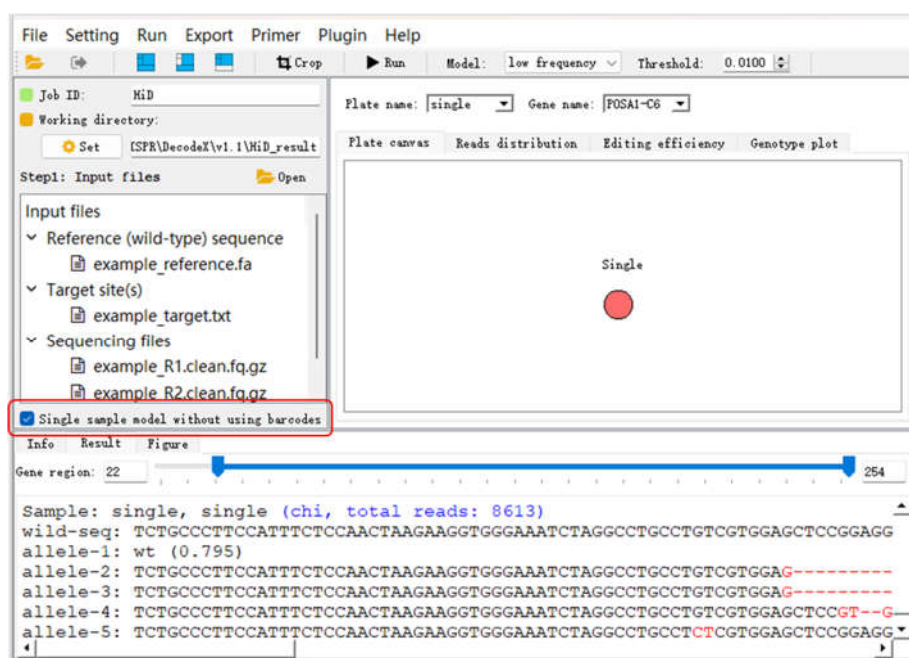


图 3-13 LaDecode 单样本文库突变分析

#### 4、使用网页版在线工具中的 LaDecode 模块分析突变

##### (1) 网页版 LaDecode 界面介绍

LaDecode 网页版主界面包括：文件输入模块和 Barcode 设置模块（图 3-14）。其中，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 设置模块用于设置 Barcode 序列。

The screenshot displays the SuperDecode web interface. At the top, there are navigation links: Home, DSDecode, HiDecode, LaDecode (selected), and Document. Below the navigation bar, a description of LaDecode is provided: "LaDecode: decoding mutations from long-read sequencing data of pooled amplicons. LaDecode is a dedicated tool for decoding multiple and complex mutations induced by multiplex target-site editing within a certain region by long-read TGS of long-range PCR amplicons, including PacBio and Nanopore sequencing. For testing the program, click [download](#) to obtain the test files."

The main configuration area is divided into two sections: "Input files" and "Barcode settings".

**Input files section:**

- Step 1: input reference sequence in fasta format**  
A text input field for the reference sequence is shown with a placeholder ">seq\_name" and "reference sequence". Below it, a button "选择文件" (Select file) is next to the text "未选择任何文件" (No file selected).
- Step 2: input target sequence file (optional):** A button "选择文件" (Select file) is next to the text "未选择任何文件" (No file selected).
- Step 3: input sequencing file (in BAM or FASTA format)**  
A button "选择文件" (Select file) is next to the text "未选择任何文件" (No file selected). Below it, there is a checkbox labeled "Single sample (not a pooled amplicon)".
- Step 4: threshold setting** A text input field shows "0.250" with a range "(0.001-1.00)" next to it.

**Barcode settings section:**

- Using default barcodes of LaDecode** (selected by default)  
Two buttons "Select position barcode:" and "Select plate barcode:" are shown, each with a dropdown arrow.
- Defined your barcode**  
Two buttons "选择文件" (Select file) are shown, each next to the text "未选择任何文件" (No file selected). Below each button is a link to an example file: "Example\_position\_barcode.txt" and "Example\_plate\_barcode.txt".

At the bottom of the form, there is a text input field for "Your email (optional):" with a placeholder "Email address". Below this field is a note: "\*In case the webpage is closed during the program's execution, it is recommended to enter a valid email address to receive the analysis results."

A blue button labeled "Run LaDecode" is located at the bottom left of the form.

图 3-14 网页版 LaDecode 主界面

## (2) 基本分析操作

使用 LaDecode 网络版在线工具从 TGS 扩增子文库分析靶点突变序列的主要步骤包括文件输入、样本分析模式选择、参数设置和解码。详细操作如下：

i) 在文件输入模块中依次输入：**Fasta 格式**的参考序列或包含参考序列的文件（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）。

ii) 在文件输入模块中选择样本分析模式，如果为单个样本的文库分析，则点击“**Single sample (not a pooled amplicon)**”启动单样本文库分析模式。根据样本类型在文件输入模块中设置突变检测阈值 Threshold。

iii) 如果为多样本扩增子混合文库，在上传相关文件以及设置突变检测阈值 Threshold 后，需在参数设置模块中设置 Barcode 序列信息。如果为单样本文库的突变分析则不需设置 Barcode 信息。

iv) 点击“**Run LaDecode**”按钮启动分析任务。（注意：在启动分析前可在 email 输入框中输入接收解码结果的邮箱，防止解码过程中网页关闭导致结果丢失）。

程序运行结束后，解码结果会跳转至新的网页中显示，可以通过鼠标选择相应样本查看对应解码分析结果，包括当前选择样本的序列与参考序列的比对情况、当前样本的总体突变频率、Indel 片段大小的分布以及所选样本在参考序列中每个位置的突变

频率（图 3-15）。解码分析结果可点击网页下面的下载按钮进行保存。



图 3-15 网络版 LaDecode 解码结果

## 五、问题和反馈

欢迎用户在使用本软件包过程中出现的问题和改良建议反馈给我们，联系人：谢先荣（[xiexianrong@scau.edu.cn](mailto:xiexianrong@scau.edu.cn)）。后续我们也将建立软件社区，及时就使用过程中的问题进行沟通。