

Sanger测序和片段分析

# Applied Biosystems SeqStudio Flex 系统 — 最新一代中通量基因分析仪



## 在本应用指南中，我们证明：

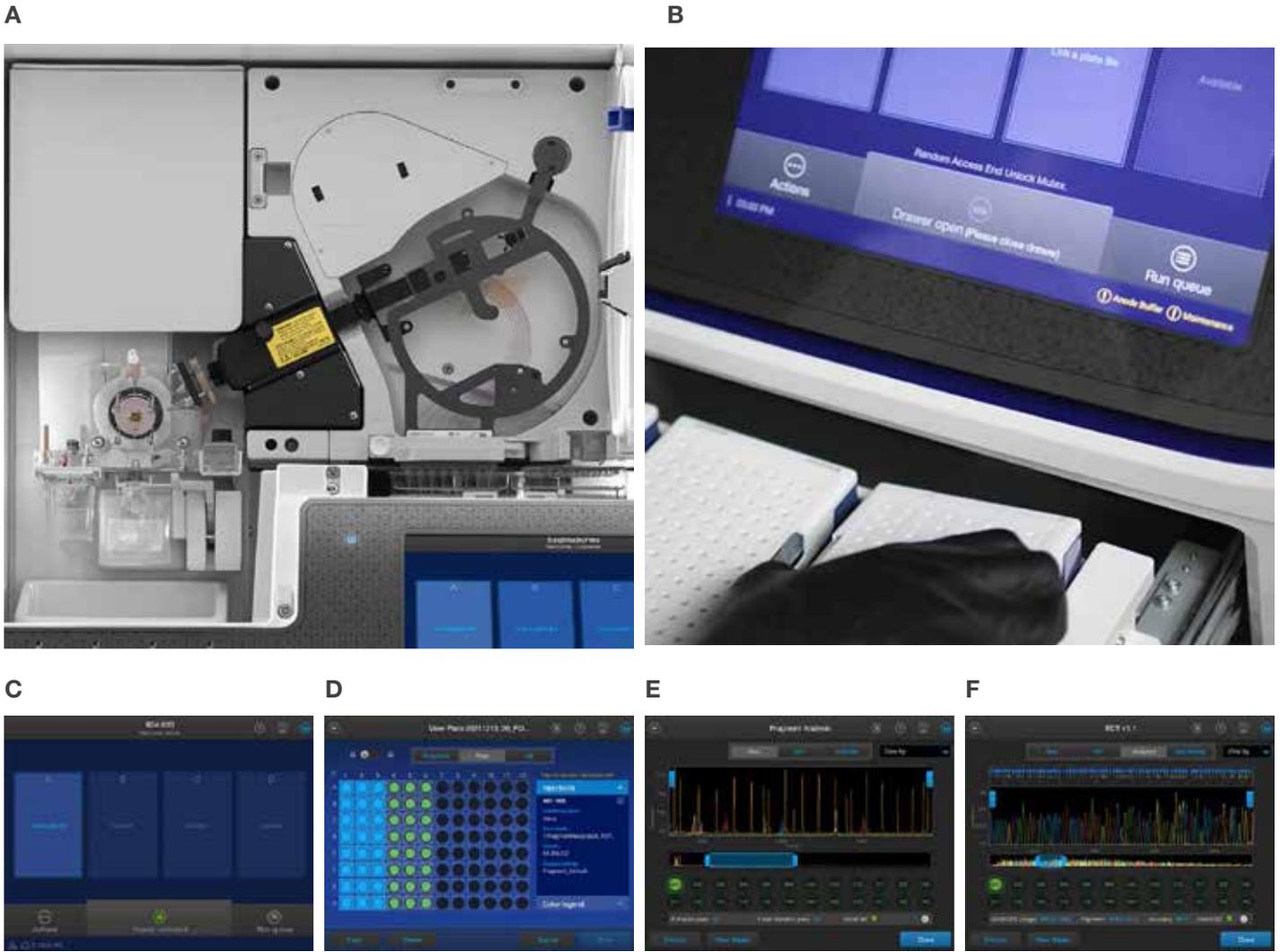
- 高质量数据以及与其他Applied Biosystems™基因分析仪的可比性
- 简单运行设置和二次分析软件
- Sanger测序和多重荧光片段分析应用的亮点，包括：
  - 细胞系鉴定
  - 微卫星不稳定性
  - dsDNA质控
  - 多重PCR分析
  - 基因编辑确认
  - SARS-CoV-2测序
  - 次要等位基因突变测序
  - 二代测序确认
  - 质粒测序
  - 困难模板测序

## 引言

荧光毛细管电泳(CE)是根据片段大小分离荧光标记DNA片段的基因组分析方法，是Sanger测序和片段分析的基础。简单的工作流程、单碱基分辨率、快速分析时间、较小样本量和灵活性使其在各种基础和转化应用、临床研究以及药物研发、新型疗法和疫苗研发中被广泛采用。Applied Biosystems™作为CE仪器、试剂和软件方面的领先品牌，不断提供并满足现代研究的创新解决方案。

Applied Biosystems™ SeqStudio™ Flex是加入我们CE产品组合的最新仪器，它是最先进的中通量基因分析仪，可产生高质量的Sanger测序和多重荧光片段分析结果。SeqStudio Flex基于十分成熟的荧光CE技术(图1A)，通过优化设计和先进的通信提高灵活性、易用性、连通性和可维护性。新的亮点包括连续反应板载入和4个反应板容量(图1B)、改进的机载计算机直观用户界面(图1C)、远程运行和数据监控(图1D-F)、远程故障排除和动态光谱校准。这些在使用性能和高级通信方面的创新简化了CE工作流程，使Applied Biosystems CE仪器具备灵活性，对新的、激动人心的基因分析应用产生影响。

在本指南中，我们介绍能够在SeqStudio Flex仪器上运行的宽谱基因分析应用的选择，并表明所选基因分析应用产生的数据质量与4毛细管SeqStudio™和24毛细管3500xL系统等其他Applied Biosystems™基因分析仪的数据质量相同。



**图1. 介绍SeqStudio Flex基因分析仪。** (A) SeqStudio Flex系统组建在相同的基础平台上，使用与3500系列基因分析仪相同的缓冲液和聚合物耗材。(B) 多达4个反应板可载入该仪器，软件允许用户在持续运行期间定义和载入反应板。(C) 直观用户界面的机载触屏计算机。(D) 简化的反应板添加或变更。实时或完成运行后即可看到片段分析(E)或测序数据(F)。数据能够从该仪器通过USB驱动器、局域网(LAN)或Wi-Fi远程访问进行传输。

# 结果

## 片段分析应用

片段分析是一种高度灵活的方法，通过CE分离不同片段大小和不同标记的DNA片段。通过聚合酶链反应(PCR)获得DNA片段；由于PCR引物选择的灵活性，易于生成特定大小的DNA片段。可用多达四种的不同荧光团标记DNA片段，研究人员在实验设计上有高度灵活性。

### 细胞系鉴定

疾病研究依靠于体外获取和操作的样本分析。这些样本包括人类细胞系、诱导多能干细胞(iPSC)和嵌合抗原受体(CAR) T细胞。污染和人为误差可能发生，影响研究结果或危及生命。

例如，国际细胞系鉴定委员会(ICLAC)发现，研究人员使用的至少451个细胞系被错误识别[1]。此外，2019年的一项研究发现32755篇文章的报告结果通过错误识别的细胞获得，并发现这些结果被引用到大约50万篇文章中[2]。

最后，当研究进展到体外细胞疗法，即细胞从供体取出，在实验室中操作，并返回到宿主，所以确认供体细胞有预期基因型并与预期受体匹配非常重要。因此，鉴定人类细胞的种源非常关键。

Applied Biosystems™确认人类样本身份解决方案基于短串联重复序列(STR)的片段分析。STR是高突变性微卫星序列，为人类样本提供唯一的分子指纹。Applied Biosystems™ GlobalFiler™细胞系鉴定(CLA)试剂盒生成24个不同STR基因座的分子指纹，而Applied Biosystems™ Identifiler™ Plus CLA试剂盒分析16个STR基因座。这些简单且方便的试剂盒用于Applied Biosystems系列的CE基因分析仪。有关更多信息，见参考文献3。

为了说明SeqStudio Flex仪器相对于其他Applied Biosystems基因分析仪的性能，对来自六种常用人类细胞系的DNA(A549、Jurkat、U2OS、HEK293，M4A4GFP和HeLa)使用Invitrogen™ RecoverAll™试剂盒进行纯化。对于STR分析，根据提供的方案，1ng的基因组DNA与GlobalFiler CLA或Identifiler CLA试剂盒一起使用。每份样本重复运行三次。样本在SeqStudio Flex、3500xL和SeqStudio基因分析仪上运行。使用Applied Biosystems™ GeneMapper™ 6.1软件进行等位基因指认。通过检查常见细胞系STR数据库的结果确认细胞系的真实性[4]。

比较所有仪器上五个细胞系所有基因座的STR峰形(图2A)。使用GlobalFiler CLA试剂盒，我们发现与使用SeqStudio Flex和3500xL基因分析仪的所有细胞预期STR等位基因图谱100%匹配，与SeqStudio基因分析仪98.7%匹配(表1和表2)。同样，使用Identifiler CLA试剂盒，我们得到与SeqStudio Flex系统的100%匹配，与3500xL系统的99.6%匹配，与SeqStudio系统的98.8%匹配。

为了检测细胞系污染的能力，在50%、25%、10%和5% HeLa DNA比例下，我们将M4A4GFP和HeLa gDNA混合。每种gDNA混合物各取1ng，使用Identifiler CLA试剂盒进行分析。HeLa细胞有16个等位基因未在M4A4GFP图谱中发现。我们发现，即使污染的HeLa细胞组成5%的gDNA样本，基因分析仪也能够检测9%到23%之间的独特HeLa等位基因(图2B)。这些结果证明，与其他基因分析仪系列相比，SeqStudio Flex仪器生成质量相似或更好的细胞系鉴定结果。

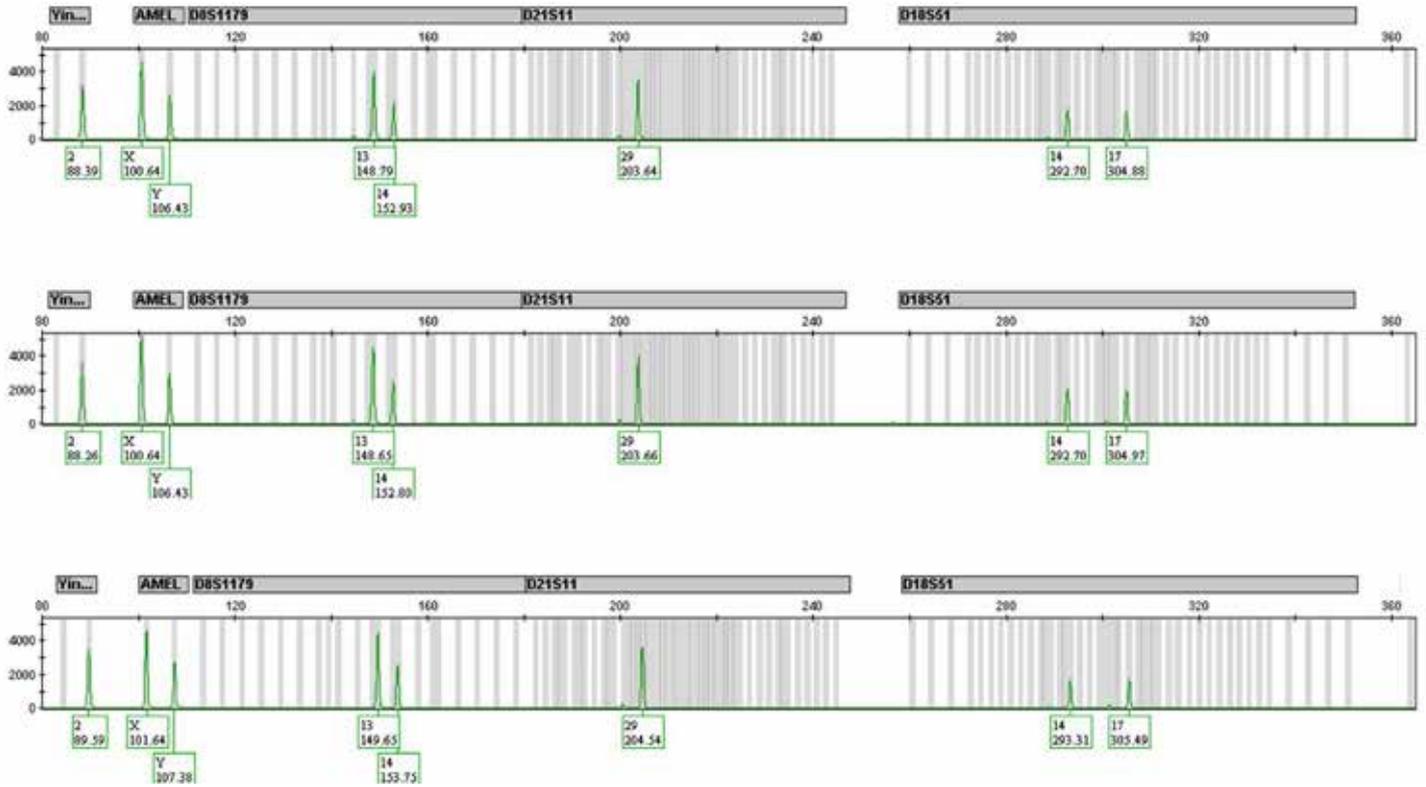


图2A. 三种基因分析仪平台的CLA峰形态。使用GlobalFiler CLA试剂盒，SeqStudio Flex (顶部)、3500xL (中间)和SeqStudio (底部)基因分析仪产生相似的峰形态和等位基因。显示了单通道；在五色通道和其他样本中看到相似结果。

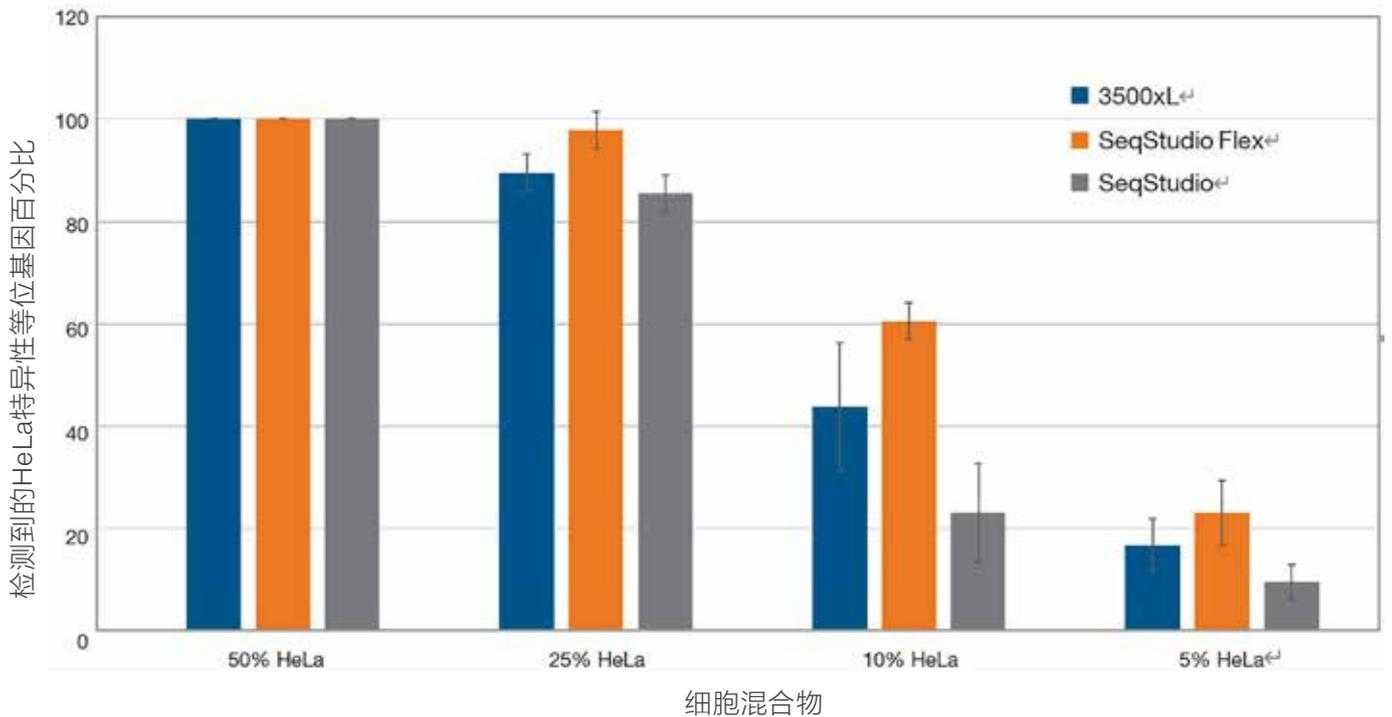


图2B. 在低浓度下检测污染的等位基因。减少的HeLa gDNA量添加到M4A4GFP gDNA中。DNA总量保持为1ng。使用Identifiler Plus CLA试剂盒对混合物重复分析三次。确定在每次滴定与预期数量中检测到的HeLa特异性等位基因数量。注意，在三种仪器上，检测5%样本的HeLa特异性等位基因。

表1. 在SeqStudio Flex、3500xL和SeqStudio基因分析仪上收集的等位基因图谱。

所示示例来自使用GlobalFiler CLA试剂盒分析的人类A549细胞。

基因座	通道	SeqStudio Flex		3500xL		SeqStudio	
AMEL	G	X	Y	X	Y	X	Y
CSF1PO	B	10	12	10	12	10	12
D10S1248	P	13	16	13	16	13	16
D12S391	P	18		18		18	
D13S317	R	11		11		11	
D16S539	B	11	12	11	12	11	12
D18S51	G	14		14		14	
D198433	Y	13		13		13	
D1S1656	P	17	18.3	17	18.3	17	18.3
D21S11	G	29		29		29	
D22S1045	R	15		15		15	
D2S1338	P	24		24		21	
D25441	Y	10	13	10	13	10	13
D3S1358	B	16		16		16	
D5S818	R	11		11		11	
D7S820	R	8	11	8	11	8	11
D8S1179	G	13	14	13	14	13	14
DYS391	G	10		10		10	
FGA	Y	23		23		23	
SE33	R	19	25.2	19	25.2	19	25.2
TH01	Y	8		8		8	
TPOX	B	8	11	8	11	8	11
VWA	B	14		14		14	
Yindel	G	2		2		2	

表2. 对于六个细胞系的三次复制，合计使用GlobalFiler CLA试剂盒准确指认的等位基因的总数。每个细胞系包括略微不同的纯合子和杂合子等位基因数量组合，因此细胞系之间的等位基因总数不同。

	SeqStudio Flex	3500xL	SeqStudio
A549	102/102	102/102	102/102
Jurkat	153/153	153/153	153/153
U2OS	105/105	105/105	105/105
HEK293	111/111	111/111	104/111*
M4A4GEP	96/96	96/96	96/96
HeLa	117/117	117/117	115/117*

\* 三次进样中一次进样期间的技术问题导致这些样本中的一些异常指认。

## 微卫星不稳定性

多种类型的癌症显示DNA错配修复(MMR)的缺陷，产生基因组总体较高的突变率[5]。较高的突变率通常意味着新抗原产生率较高，为免疫治疗研究提供机会[6]。

存在至少11个不同的基因座，均与MMR有关[7]。寻找这些基因座序列的灭活事件会很复杂、耗时和昂贵。通过微卫星不稳定性(MSI)分析检验MMR通路是一种实用的替代方法。由于微卫星序列的高度重复性质，在NGS系统上进行MSI分析会非常困难。因此，使用CE进行微卫星基因座长度的片段分析是一种检测MSI的广泛应用方法。

Applied Biosystems™ TrueMark™ MSI分析系统分析一组13个微卫星基因座，包括Bethesda标志物组[8]，作为微卫星不稳定的一个指标。TrueMark MSI检测还提供了两个高突变基因座的STR信息，从而进行样本身份确认。

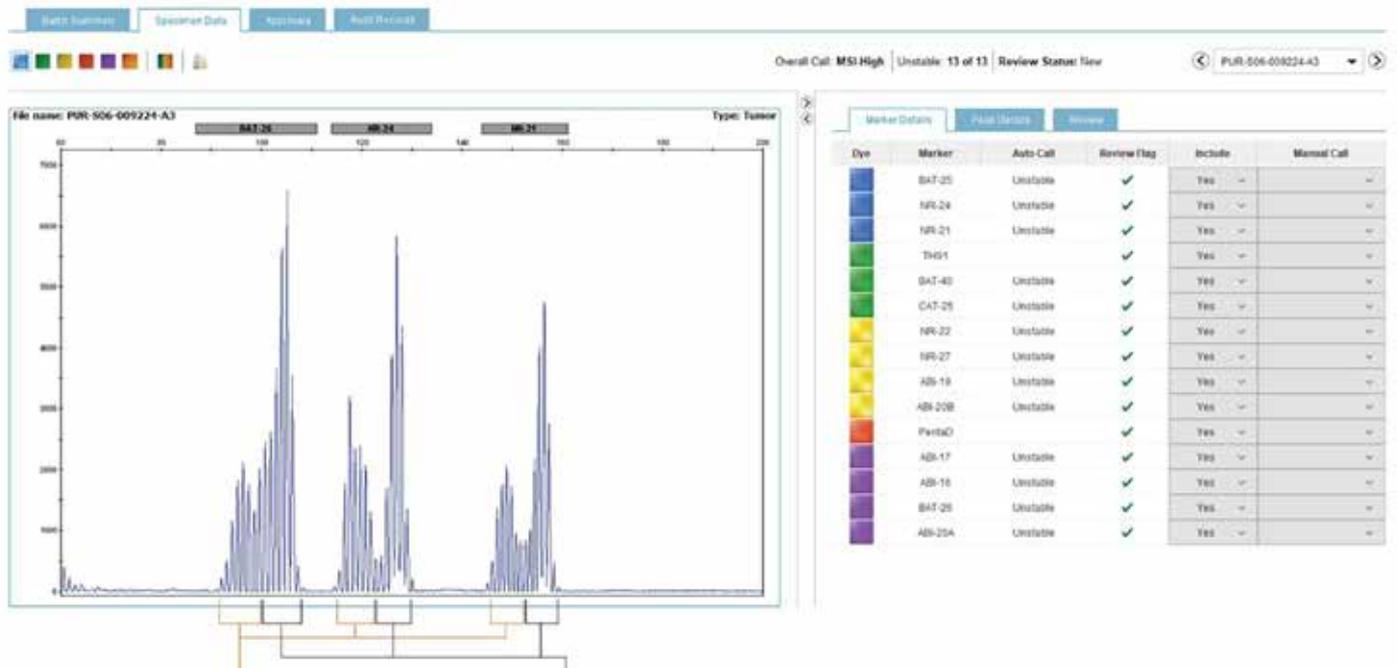
为了便于分析MSI原始数据，我们开发了Applied Biosystems™ TrueMark™ MSI分析软件，简化了鉴定各基因座位置(图3A)。此外，我们不需要并排分析正常非肿瘤组织，减少了所需的样本数量和MSI成本。有关更多信息，见参考文献9。

为了说明SeqStudio Flex仪器相对于其他基因分析仪的性能，获取了九个肿瘤/正常相邻对和一个成块仅肿瘤FFPE样本。切片(10μm)从FFPE块上切下，gDNA从切片上提取，根据已发布的方案，将1–2ng的基因组DNA用于 TrueMark检测。在SeqStudio Flex和3500xL基因分析仪上分析片段，并使用TrueMark MSI分析软件解读结果(图 3A)。

为了比较基因分析仪之间的性能，我们记录可称为稳定、不稳定或无指认的基因座的数量(图3B)。各类基因座的总数高度一致。一个样本(S07-001886-A5)的回收gDNA很少，因此不可能使用2ng的推荐输入量。然而，两个仪器能够进行相似的可检测基因座指认。

来自相同FFPE块的样本用于MMR蛋白的标准免疫染色，使用SeqStudio Flex和3500xL系统的总体MSI检测结果与免疫染色结果完全一致。这些结果证明，SeqStudio Flex系统生成的MSI分析数据与那些使用3500xL系统得到的数据相同。

A



微卫星不稳定性出现的片段大小

每个基因座片段大小的正常范围

B



**图3. SeqStudio Flex和3500xL仪器产生相似的MSI分析数据。(A)** TrueMark MSI检测分析了13个微卫星基因座的不稳定性, 包括广泛应用的Bethesda标准。确定为不稳定的基因座是自动认定; 然后软件会使用指认总数进行样本的总体认定。检测包括两个高突变性短串联重复(STR)序列(THO1和PentaD), 这两个序列可用于确认样本身份。软件采用的专有算法不需要同时分析正常非肿瘤组织, 以进行稳定/不稳定认定。**(B)** 使用TrueMark MSI检测, 对九个肿瘤/正常相邻对和一个仅肿瘤样本进行了分析。使用两个仪器的数据, 通过软件指认的基因座的数量高度相似。样本S07-001886-A5的回收gDNA并不理想; 并非所有基因座都等量扩增, 因此两个仪器的结果略微不同。

## dsDNA质控

许多基因组分析方案从基因组DNA开始。开始进一步研究之前，通常确认混合的DNA片段大小和数量。例如，从FFPE保存组织提取的DNA可能导致片段化DNA太小而无法提供有用信息。因此，进行实验前，确定整体片段大小分布非常有用。

在二代测序(NGS)实验中，混合片段文库通常在加载到序列读取芯片前生成。知道混合片段文库中DNA片段的大小和丰度对预计NGS结果成功非常重要。

最后，确认PCR反应是否产生特异性扩增子非常有益。琼脂糖凝胶通常对此有用，然而，虽然它们便宜，但并非精确无比或高通量。

双链DNA片段的大小和相对丰度可在Applied Biosystems基因分析仪上进行分析。对于该应用，使用样本孵育Invitrogen™ TOTO™-1 DNA插入荧光染料。随后，在毛细管加热器关闭的情况下，使用Applied Biosystems™ POP-7™聚合物运行该样本。这有助于保持DNA主要在双链和非变性状态下。通过比较所得检测片段的迁移与已知dsDNA标准品的迁移(例如，经HindIII酶切的λ噬菌体DNA、经BsuRI酶切的phiX174噬菌体DNA)，可确定检测片段的大小。

为了评价SeqStudio Flex系统相对于3500xL系统分析dsDNA的性能，我们获得了从细胞系提取的基因组DNA样本、从FFPE玻片提取的基因组DNA、用于循环测序的*BRAF*扩增子或λ HindIII和phiX174 BsuRI酶切DNA样本。将dsDNA样本与50 nM的TOTO-1染料和单链(ss) Applied Biosystems™ LIZ™片段标准品溶液混合于水中。将样本在37°C条件下孵育10分钟，并在基因分析仪上分离，使用E5染料组校准，毛细管加热关闭。片段使用GeneMapper 6.1软件进行了分析。

当使用该方法分析λ HindIII标准品时，可以看到八个片段(23130 bp-125 bp)熟悉的图谱(图4A)。在相同进样的不同毛细管中，我们分析了从细胞系中纯化的50ng基因组DNA。通过比较gDNA迁移与λ标准品表明，大多数完整DNA范围为~9 kb-20 kb。

同样，我们在另一个毛细管中运行λ HindIII和phiX174 BsuRI标准品的混合物。除八个λ片段外，可看到范围为1353-72 bp的10多个片段(图4B)。从FFPE玻片提取的DNA运行，片段范围为~<70bp-~200bp，加上一些范围为4kb-10kb的较大DNA片段。

最后，循环测序前，我们检查*BRAF*扩增子的质量(图4C)。根据λ和phiX174标准品，该扩增子在~200-230bp的范围内迁移。该扩增子的预期片段大小为208bp，与观察到的片段大小相匹配。注意，在每个示例中，在SeqStudio Flex和3500xL分析仪上得到的图谱非常相似，证明SeqStudio Flex和3500xL系统生成相同的dsDNA片段数据。

该方法提供了dsDNA样本快速简单质控的方法。一些改进会增加从这些运行中得到的信息。例如，单链LIZ内标可以重新校准到λ/phiX标志物组。一旦重新确定它们，当与未知样本运行在相同的毛细管中时，LIZ标准品可作为内部片段标准品。因此，如果同时运行标准曲线，则可能确定峰值的量。最后，我们发现该应用还在4通道毛细管SeqStudio仪器上有效(数据未给出)。



**图4. 通过CE分析双链DNA。** (A) 在降低温度下，不同的DNA样本通过含POP-7聚合物的毛细管进行电泳。单链LIZ标记片段包含在毛细管中。在不同毛细管中，对  $\lambda$  HindIII酶切片段和来自A549细胞的基因组DNA进行了分析。注意  $\lambda$  片段的一般图谱(上组)。这些可用于估计大约9–23 kb的gDNA群体的片段大小(下组)。观察到两个仪器具备相同性能。(B) 在不同毛细管中，对  $\lambda$  HindIII和phiX174 BsuRI酶切混合物以及从FFPE肿瘤样本提取的DNA进行了分析。注意噬菌体片段的一般图谱(上组)；这些可用于估计片段化DNA (75–190 bp)和较大基因组DNA的片段大小(下组)。观察到两个仪器性能相同。(C) 循环测序前，在不同毛细管中，对含BRAF V600E区的扩增子与  $\lambda$  HindIII和phiX174 BsuRI酶切混合物进行了分析。注意噬菌体片段的一般图谱(上组)；这些可用于估计BRAF扩增子(194–230 bp)的片段大小。这与预期片段大小208 bp非常相似。观察到两个仪器结果相同。

## 多重PCR分析

通过PCR进行DNA序列扩增是分子生物学和基因组研究的主要工具。由于知识和科学问题的深度变得更加复杂，因此分析单样本许多靶标的需求越来越多。然而，单重PCR较为冗长，涉及设置各预期靶标的个体反应。

能够分析单样本多个PCR扩增子的方法会精简工作流程并节省宝贵样本。解决它的一个方法是，使用多重能力的多个荧光染料，通过CE进行片段分析。此时，由于能够根据单个毛细管中的片段大小和荧光染料分离特异性PCR产物，因此可分析单样本的大量靶标。实际上，上述CLA和MSI分析试剂盒是多重PCR分析片段的示例。

我们证明如何使用自主设计通过多重PCR进行片段分析，以检测呼吸道病毒性病原体。我们设计一组PCR引物，它们能够检测12种不同的呼吸道RNA病毒。设计引物对，使每个生物体产生不同片段大小的扩增子。Oligo包含在5'端的6-FAM™标记，将这组引物合并，使其与样本即可一步混合。使用任意CE基因分析仪，能够分离和检查所得扩增子。

为了说明SeqStudio Flex仪器相比其他基因分析仪的性能，我们获取了常见呼吸道病毒的Invitrogen™ GeneArt™合成DNA靶标或纯化基因组RNA靶标(ATCC)。使用Applied Biosystems™ TaqMan®多重预混液和上述引物对，对这些单一或混合的靶标进行了扩增。

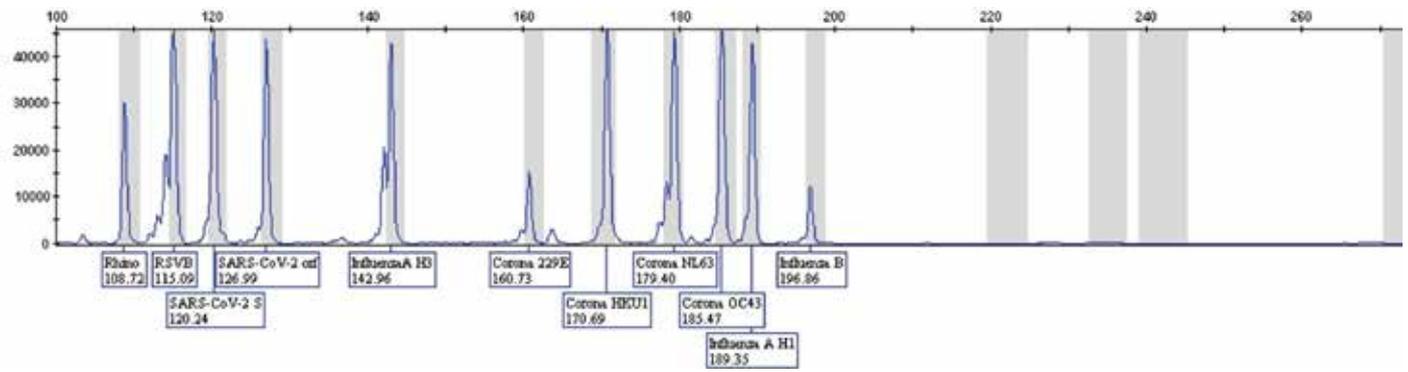
进行PCR后，扩增子与Applied Biosystems™ Hi-Di™甲酰胺和LIZ标准品混合，并在SeqStudio Flex和3500xL基因分析仪上进行分离。使用GeneMapper v6.1软件，对所得数据进行了分析。

通过针对其同源靶标(数据未给出)和所有DNA靶标与合并引物组混合物来分别检测每个引物对，确认引物功能(图5A)。在单重或多重检测中，每个引物对能够识别其靶标(如存在)。我们确认，通过合成单一病毒基因组或所有病毒基因组混合物的cDNA，然后使用单个引物或合并引物组进行PCR扩增，该方法对RNA基因组有效。如果相应靶标存在于样本中，则可检测到(图5B)。

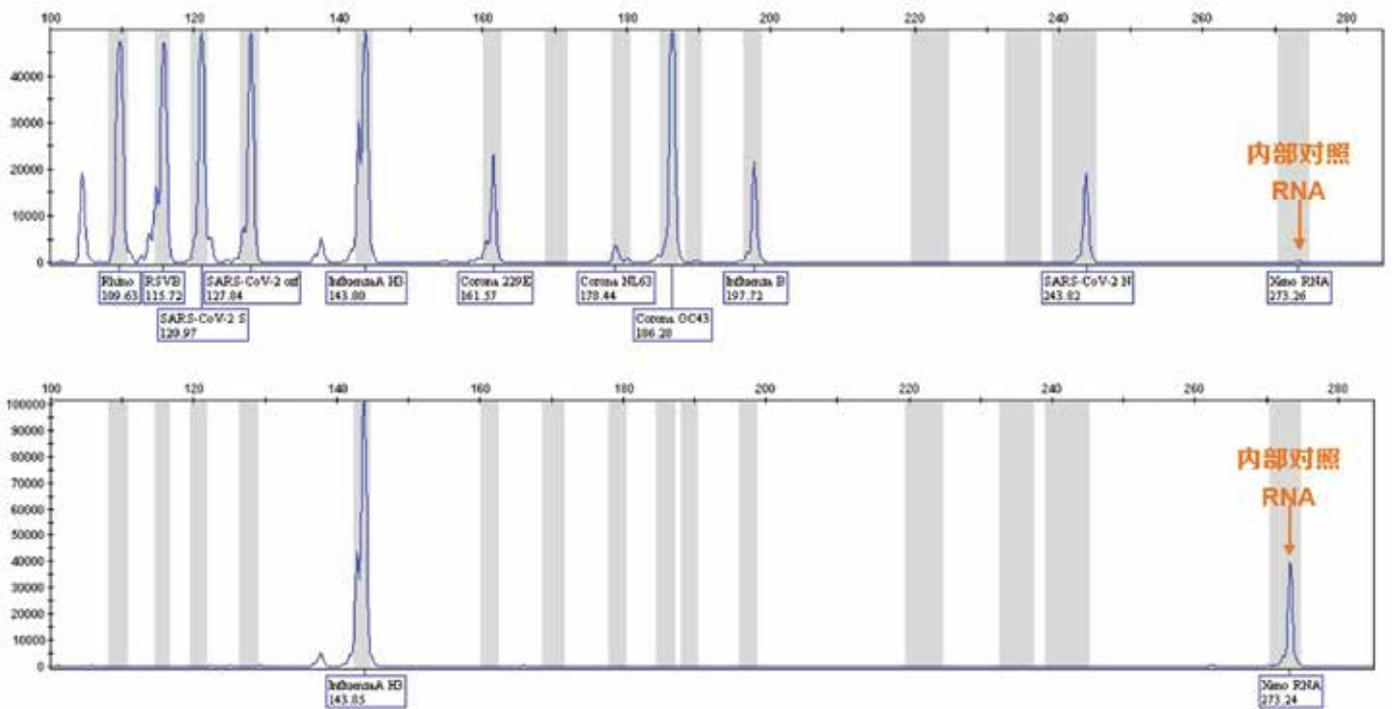
最后，我们比较使用SeqStudio Flex、3500xL和SeqStudio基因分析仪得到的结果。每个仪器产生可比较的结果(图5C)。因此，这些结果证实，SeqStudio Flex系统生成的多重PCR数据与使用其他CE系列仪器得到的数据相同。

结合片段分析的多重PCR提供了绝佳机会，可轻松进行复杂的分析。引物设计仅受有关靶标引物组的能力限制。因此，它能够适应和用于几乎任何场景，这些场景是需要检测单样本许多靶序列。

A

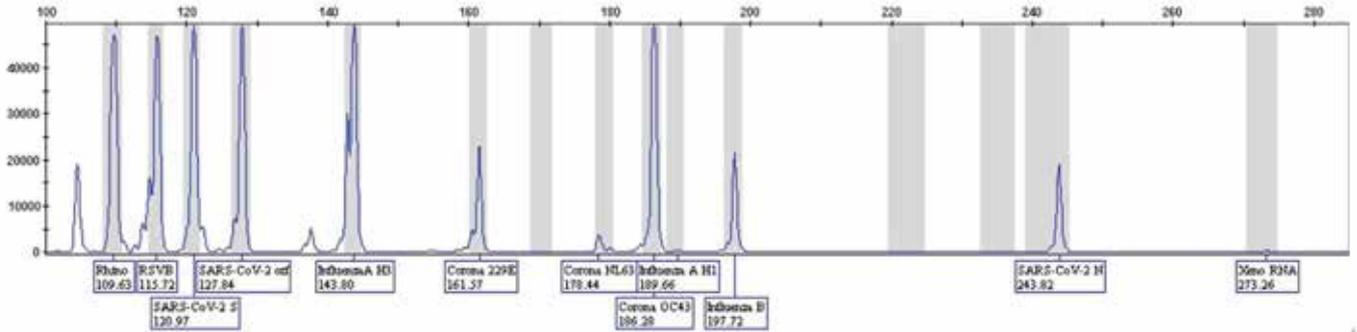


B

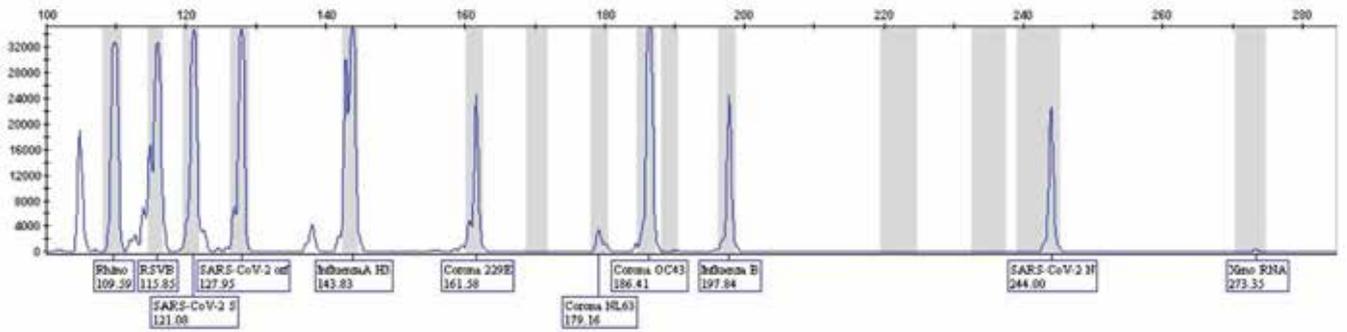


C

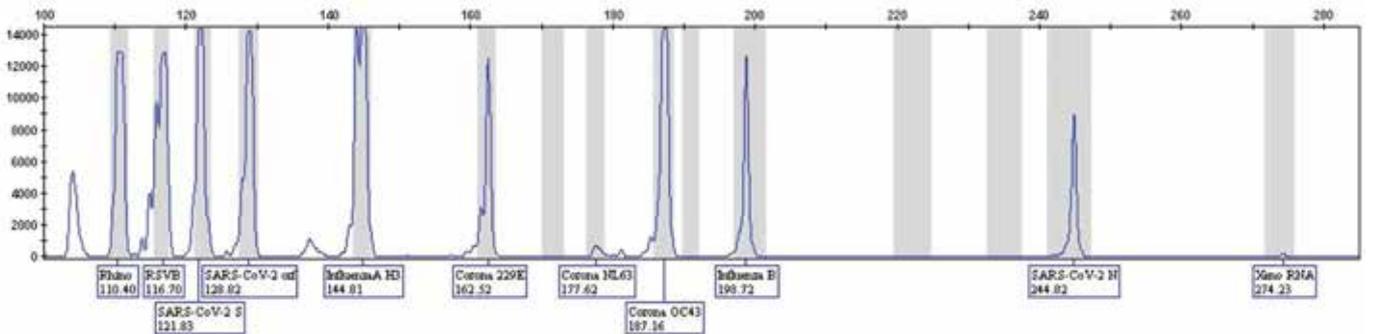
## SeqStudio Flex



## 3500xL



## SeqStudio



**图5. 通过多重PCR和片段分析来分析多个靶标。** (A) 使用6-FAM标记PCR引物对，对上面所示RNA病毒的DNA靶标在单一反应中进行了扩增。每个引物对会产生通过片段分析可以分离和确定的病毒特异性扩增子片段大小。灰色bins被定义为检测的靶标。(B) 上面所示病毒的纯化RNA基因组合并(上图)或单独保留(下图)。使用合并引物扩增后的cDNA合成证实，在单一反应中，能够得到多个或单个靶标。作为cDNA合成和扩增的内部对照，外源RNA (Applied Biosystems™ Xeno™ RNA)添加到RNA池，引物池包含外源RNA的引物。(C) 所有仪器的多重扩增子检测能力相同；在此给出RNA基因组池的示例。

# 测序应用

Sanger测序是获得DNA序列信息的“金标准”。它助力人类基因组计划，研究者继续依靠此方法产生高度准确的靶向测序结果。Applied Biosystems™产品支持快速简单的Sanger测序工作流程，从而提供高度准确性、长读序能力和简单数据分析。从取样到得出结果，整个工作流程能够在不到一个工作日内完成，具有灵活性，支持许多研究领域不同应用。下面描述一些热门应用，它们通常使用Sanger测序方法。

## 质粒测序(长读序长度)

现代基因分析通常需要DNA序列的大量操作。在亚克隆插入至质粒、生成文库、构建基因疗法病毒载体的任务中，要酶切、扩增、连接、克隆和纯化DNA片段。

因为许多步骤需要DNA聚合酶，所以错误掺入碱基和遗传漂变可能导致靶序列发生非预期突变。因此，有必要确认操作的DNA序列，以确保成品中的预期序列正确。

还建议对任何基因工程实验进行序列确认，例如AAV病毒载体构建或基因合成实验。Sanger测序片段简化了化学和分析工作流程需求。

Applied Biosystems™ BigDye™ Terminator v1.1和v3.1循环测序化学是通过CE进行Sanger测序。循环测序后，有各种电泳前纯化选择，包括Centri-Sep™纯化柱和反应板、Applied Biosystems™ ExoSAP-IT™酶混合物。最后，基因分析仪配备有为长测序片段而优化的运行模块。

为了说明SeqStudio Flex系统相对于其他基因分析仪的性能，我们进行了标准测序实验。使用v1.1和v3.1 BigDye Terminator化学，对质粒pGEM™-3Zf(+)进行了正向和反向测序。

在每个方向，进行了8次扩增反应的各方向测序反应。反应产物经过Centri-Sep反应板纯化、干燥并在Hi-Di甲酰胺中重新悬置。

所得反应产物在SeqStudio Flex、3500xL和SeqStudio基因分析仪上运行。使用Applied Biosystems™ Sequence Scanner和Biomatters Geneious™软件，对测序图谱进行分析。使用BigDye Terminator v1.1化学，我们能够在SeqStudio Flex和3500xL系统上得到超过1,000bp的高质量序列，SeqStudio系统上的测序列片段略短(图6A)。

有几项Sequence Scanner软件生成的质控指标。一个测量值是碱基指认质量值(QV)；它代表图谱峰形和峰值信噪比。高质量峰值通常QV大于20。QV是用来评价碱基指认置信度的最客观的指标，并用于计算示踪分数。

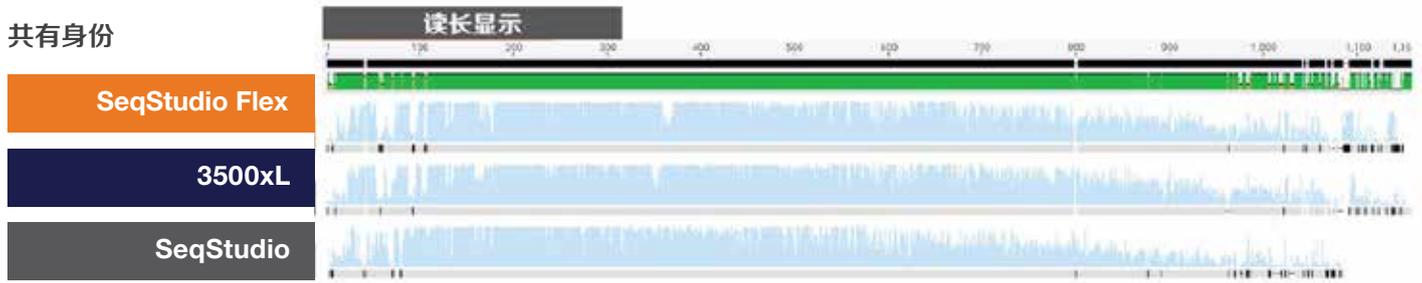
示踪分数是图谱图谱易读范围内碱基的QV平均值。序列连续读长(CRL)是图谱图谱中生成的无低质量中断的最长读序。QV20+值是在整个图谱中质量值大于20(高质量测序片段)的碱基数量。我们使用这些指标来比较性能(图6B)。

三个仪器的每个质控指标值几乎相同。同样，我们使用BigDye Terminator v3.1化学试剂测定pGEM-3Zf(+)克隆载体的序列，并在SeqStudio Flex和3500xL系统上获得非常相似的读序长度(图7A)。此外，示踪分数(图7B)、CRL和QV20+(图7C)值也非常相似。

最后，我们将测序结果与pGEM-3Zf(+)参考序列比对，测序结果包括一些同源但不连续的区域，我们发现所有仪器上得到的全长序列高度同源(图7C)。这些结果证实，SeqStudio Flex系统生成高质量的长测序片段，这些测序片段的质量与使用其他CE仪器得到的测序片段质量相同。

A

共有身份



B

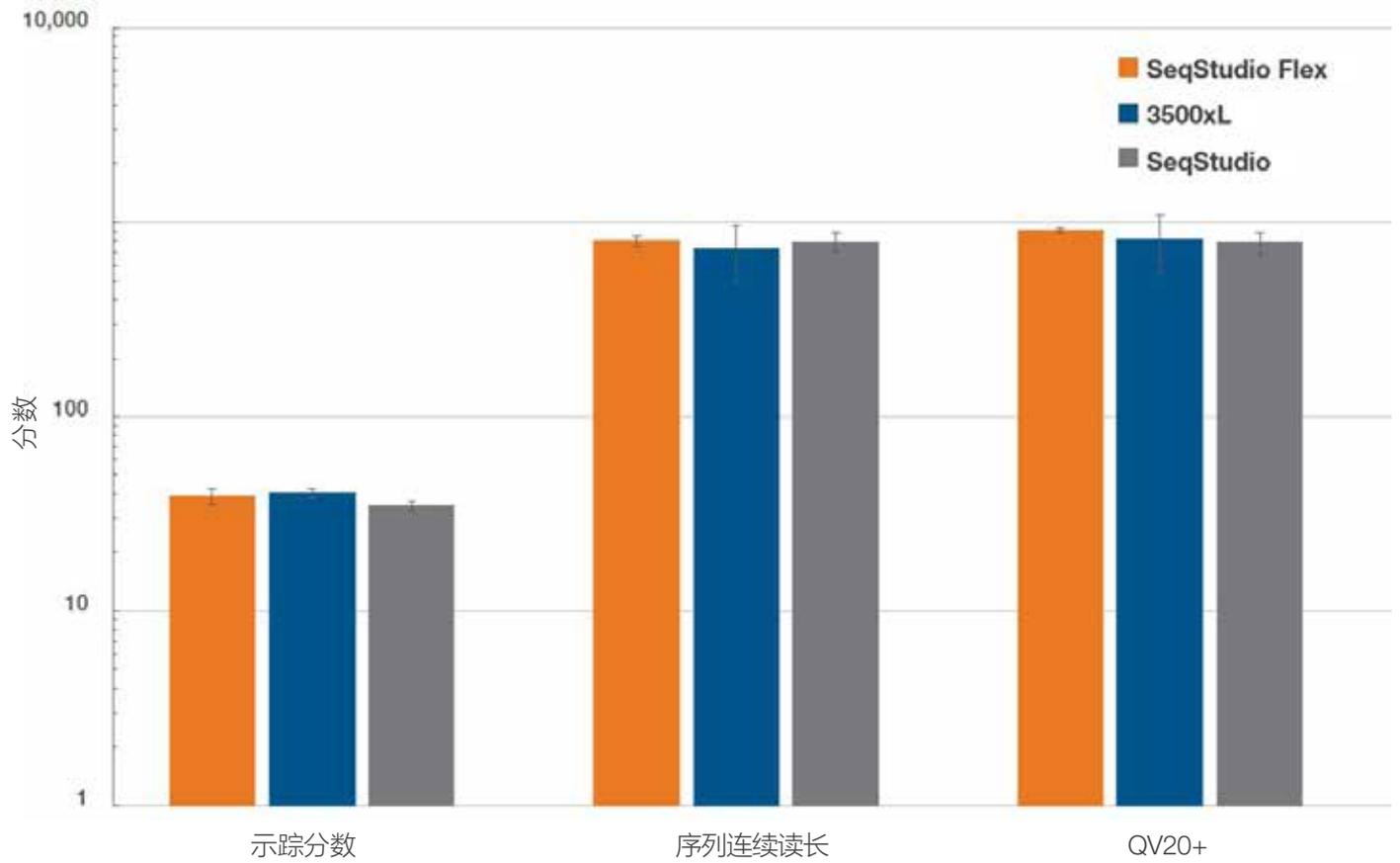
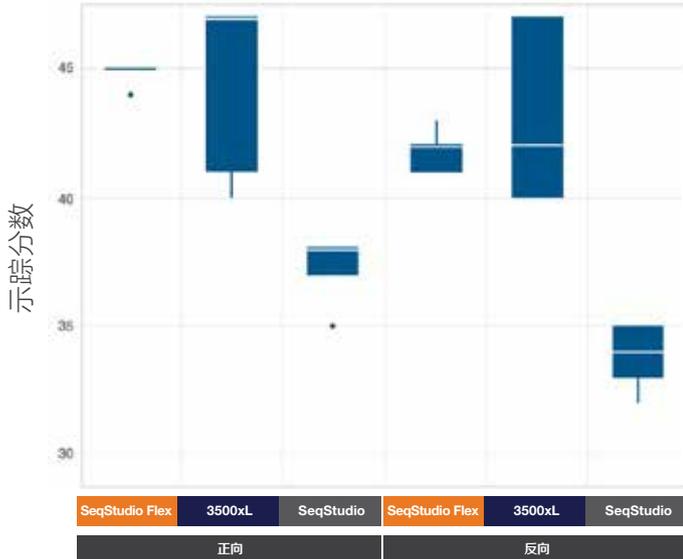


图6. 使用BigDye Terminator v1.1化学分析长测序片段的质粒测序结果。(A) 读序长度见曼哈顿图；与pGEM-3Zf(+)共有序列的对齐见顶部绿色条。淡蓝色条表示该碱基处读序的质量；更高的条是质量更好的测序片段。SeqStudio Flex和3500xL基因分析仪产生几乎相同的结果，能够读取超过1000 bp长度的读序；SeqStudio仪器显示略短的读序长度(大约1000 bp)。(B) 在全部三个仪器上得到了测序片段的质量指标。三个仪器上产生的序列质量指标几乎相同。

A



B



C

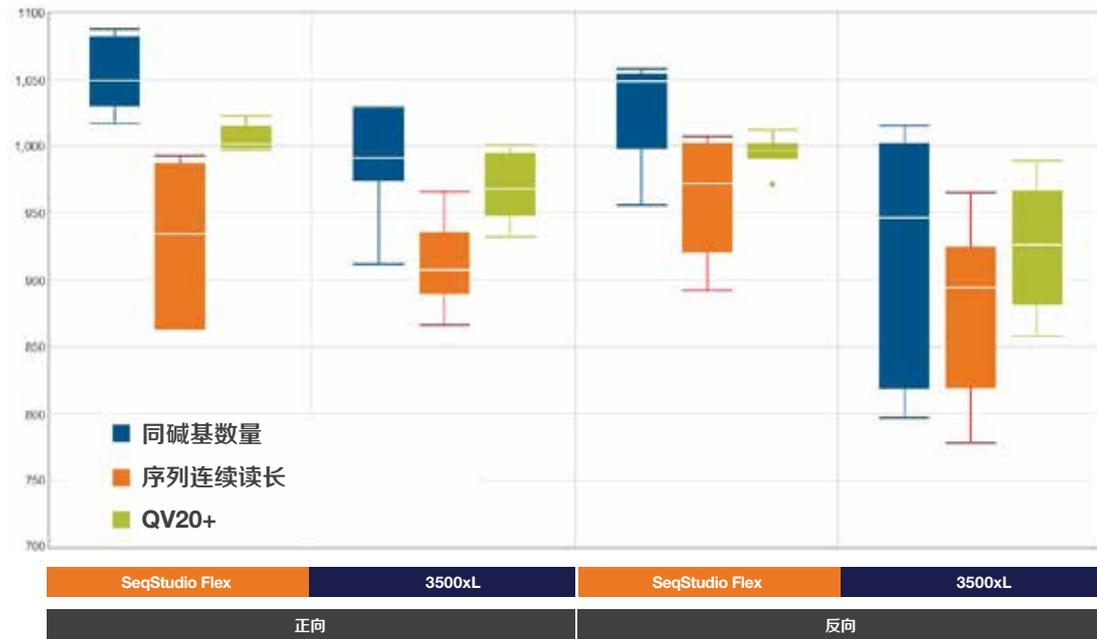


图7. 使用BigDye Terminator v3.1化学分析长测序片段的质粒测序结果。(A) 读序长度见曼哈顿图。淡蓝色条表示每个碱基处读序的质量；更高的条是质量更好的测序片段。SeqStudio Flex和3500xL基因分析仪产生几乎相同的结果，能够读取超过1100 bp长度的读序；SeqStudio仪器显示略短的读序长度(大约 950 bp)。在全部三个仪器上得到了示踪分数。(B) 序列连续读长和QV20+值(C)。此外，还得到了测序片段的同源碱基数量。SeqStudio Flex和3500xL系统提供了相似结果，SeqStudio系统显示略低数值(数据未给出)。

## SARS-CoV-2测序(中读序长度)

一些研究需要基因组特定区域的序列信息。虽然NGS等全基因组提供很大区域的大量数据，但是研究时并不总是需要此类复杂方法。当集中于一个区域时，Sanger测序提供简单、便宜且易于解读的解决方案。

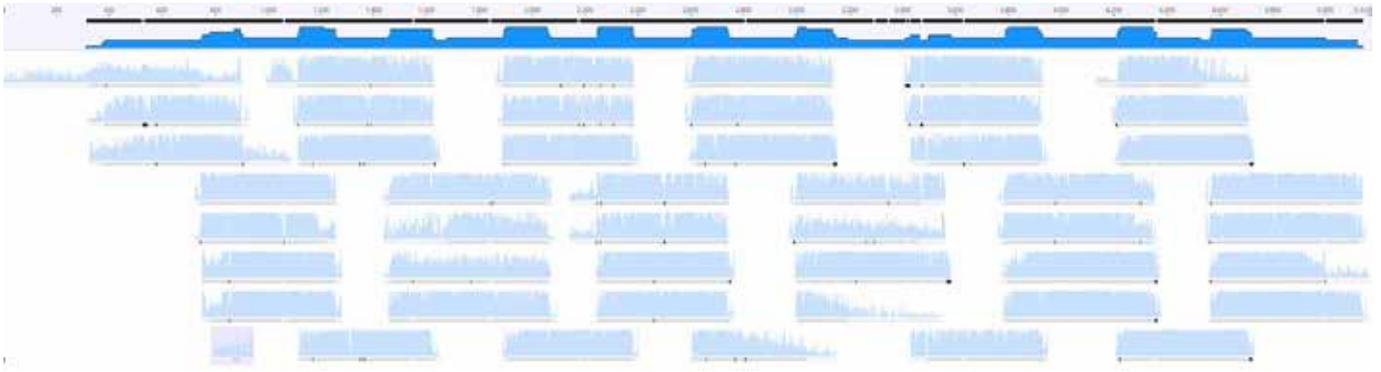
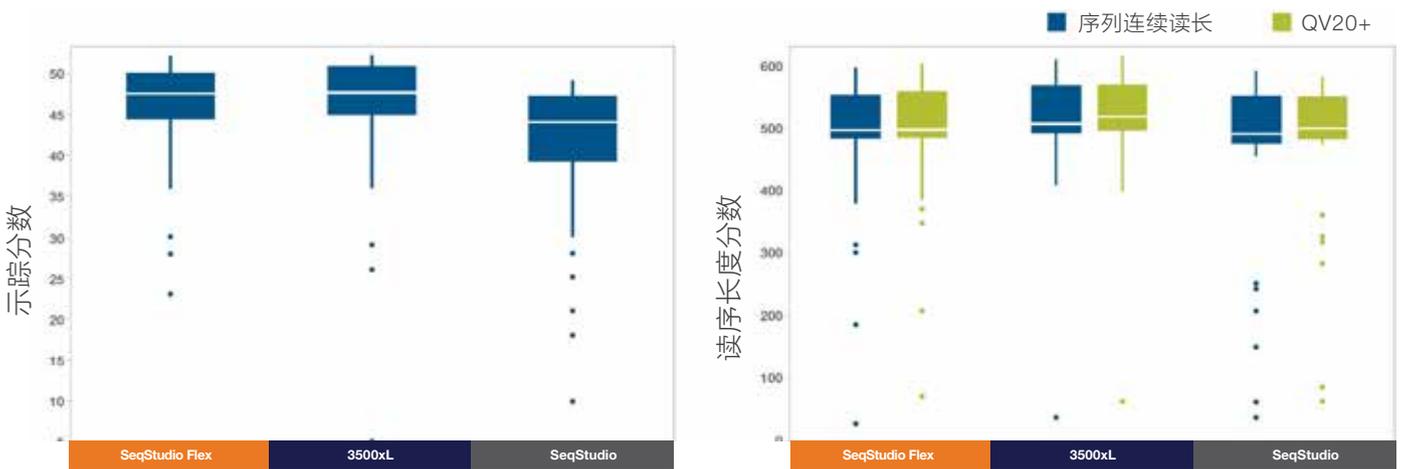
中读序长度测序时，Applied Biosystems™ BigDye™ Terminator v1.1和v3.1化学也被广泛采用。另一个选择是Applied Biosystems™ BigDye™ Direct化学利用通用M13引物来进一步简化循环测序工作流程。

当结合使用Applied Biosystems™ BigDye Xterminator™纯化试剂盒进行净化时，整个循环测序工作流程可在96孔板的单孔中进行。基因分析仪系列包含BigDye Direct和BigDye Terminator测序反应而优化的测序模块。

为了比较靶向测序结果，我们测定了包含Spike基因和相邻序列的SARS-CoV-2基因组区域的序列。

先前已描述全部SARS-CoV-2 Spike基因的重叠M13尾测序引物(12对)[10]。使用Invitrogen™ SuperScript™ IV VILO™预混液，对1000个拷贝的SARS-CoV-2基因组RNA进行逆转录。使用1 μL的cDNA和引物对产生了扩增子(500–600 bp)，使用BigDye Direct试剂盒进行循环测序，使用BigDye Xterminator试剂盒进行纯化。所得反应产物在SeqStudio Flex、3500xL和SeqStudio基因分析仪上运行。使用Sequence Scanner和Geneious软件，对图谱测序图谱进行分析。

SARS-CoV-2的Spike基因区域被涵盖超过5 kb的12个重叠扩增子覆盖。从几乎所有12个扩增子得到正向和反向的高质量Sanger测序结果(图8A)。这些结果可以组装到单一连续读长(重叠群)，该读序通过BLAST比对确认为SARS-CoV-2 Spike基因。三个基因分析仪上的读序质量指标几乎相同(图8B)。正如在三个仪器上得到的数据证明，一致的碱基质量有助于使用Sanger测序进行重叠中长度扩增子组装集中区域的序列，生成覆盖有关靶基因的重叠群。

**A****B**

**图8. 通过重叠扩增子测定SARS-CoV-2 Spike基因序列。** (A) 12个扩增子用于SARS-CoV-2 RNA重复测序两次。扩增子生成的测序片段重组装，获得Spike基因序列。每个扩增子的读序质量通过淡蓝色条的高度表示。在SeqStudio Flex系统上采集了数据；使用3500xL和SeqStudio 系统的数据进行相似对齐，得到高度相似结果。注意一些扩增子存在质量下降；在其他仪器的一些运行中也观察到了这种情况。(B) 比较了三个仪器的示踪分数(左)、序列连续读长和QV20+值(右)。测序质量结果几乎相同。

## 次要等位基因突变测序和NGS确认

二代测序(NGS)等发现式基因组研究通常找出新突变、意外突变或其他序列异常。研究者寻找使用正交法验证这些新发现的方式。Sanger测序是确认NGS结果研究的一种可选方法，原因是其工作流程简单和结果明确。

对于这些确认研究，需要对通常仅覆盖待确认区域的较短扩增子进行测序。此外，存在于不均匀样本的次要等位基因突变能够从包含混合碱基指认的电泳图中通过Sanger测序和软件处理确认。

BigDye Terminator或BigDye Direct循环测序化学和BigDye Xterminator纯化系统常用于这些类型的NGS确认研究。基因分析仪已运行为较短扩增子而优化的模块，因此可以快速获得确认数据。最后，我们开发了专门Applied Biosystems™ Minor Variant Finder软件，它能够检测存在于Sanger测序反应中低至5%的等位基因突变。

为了说明SeqStudio Flex仪器对次要等位基因确认实验的性能，我们关注*BRAF*基因。在50%、10%、5%和2.5%次要等位基因频率下，使用带*BRAF* V600E突变的标准样本作为模板。

使用*BRAF* V600E突变周围覆盖172nt的引物从这些样本中重复生成测序扩增子两次。使用BigDye Direct循环测序和BigDye Xterminator纯化化学，对所得扩增子进行了处理。纯化反应产物在SeqStudio Flex、3500xL和SeqStudio基因分析仪上运行，并使用Sequence Scanner和Minor Variant Finder软件进行了分析。

我们顺利在所检测的各浓度下检测到*BRAF* V600E突变(图9A)。在Minor Variant Finder的结果视图中，在每个样本中，可看到经该软件计算的次要突变的频率。在Minor Variant Finder的图谱视图中，对照和检测样本的正向和反向示踪以及软件处理的示踪可用于验证次要等位基因的软件调用。

在计算的次要等位基因频率下，甚至低于5%的次要等位基因频率下，各基因分析仪产生非常相似的结果。我们还比较了这些反应的测序质量指标，发现三个仪器产生质量高度相似的运行数据(图9B)。这些结果证明，SeqStudio Flex仪器能够通过Sanger测序确认存在次要等位基因突变，其性能与3500xL和SeqStudio基因分析仪相同。

**表3. 在三个基因分析仪平台上次要等位基因检测定量。**注意，虽然测得频率比预期高，但是所有仪器测得的频率相同，这表示样本而非仪器出现异常情况。

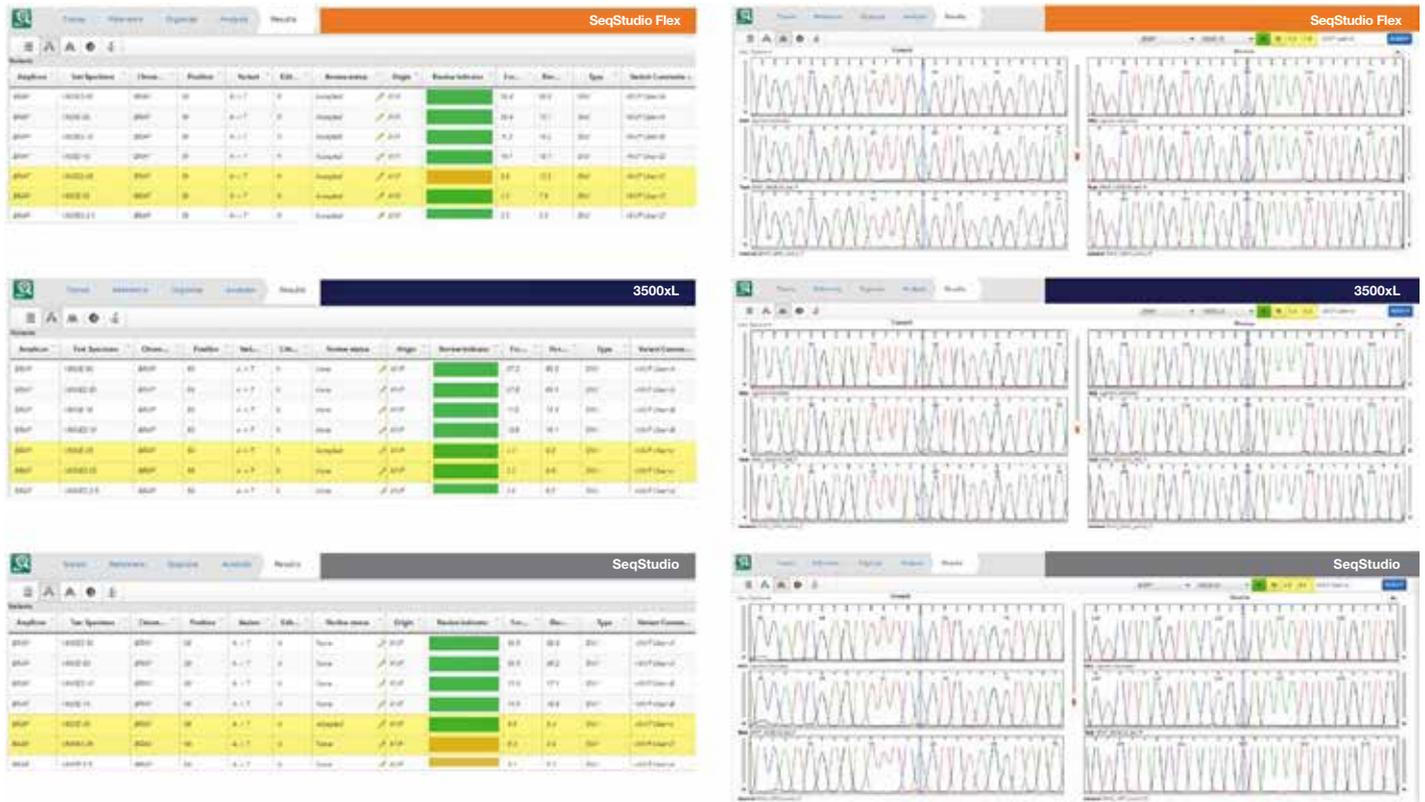
样本	预期值	测得值: SeqStudio Flex仪器	测得值: 3500xL仪器	测得值: SeqStudio仪器
50% <i>BRAF</i> V600E	50	62.4	62.6	62.8
10% <i>BRAF</i> V600E	10	14.8	13.9	15.1
5% <i>BRAF</i> V600E	5	7.8	6.9	7.0
2.5% <i>BRAF</i> V600E	2.5	4.3*	4.8*	4.2*

\* 虽然在该实验中检测到2.5%浓度，但是低频率下指认能力在其他样本中会有所不同，并随着其他等位基因而变化。

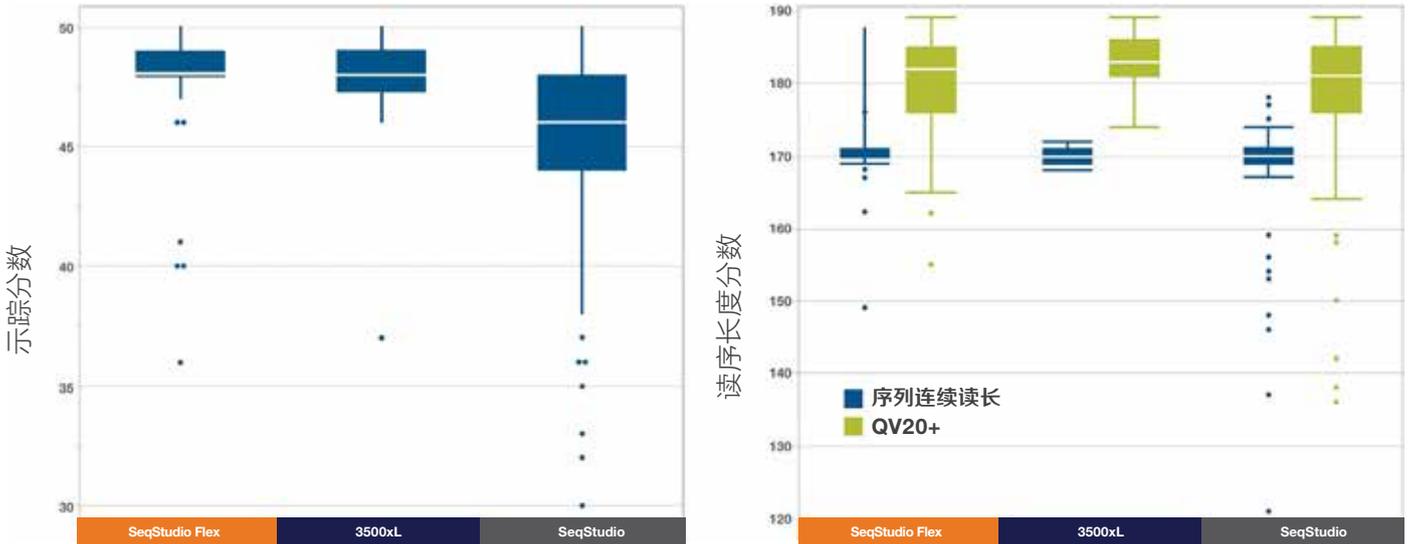
A

结果

图谱图谱



B



**图9. 通过较短扩增子测序检测次要等位基因。(A)** Minor Variant Finder软件可分析Sanger测序示踪, 并检测在5%总数位置出现的序列突变。在50%、10%、5%和2.5%频率下, 在SeqStudio Flex、3500xL和SeqStudio仪器上, 对包含*BRAF* V600E突变的模板进行了分析。在每种情况下, 软件能够重复检测到5%突变(黄色高亮, 左)。在图谱视图(右)中, 可看到对照、检测和软件处理样本的Sanger示踪。三个平台的示踪相同; 在各组中, 通过软件确定的等位基因数量用黄色高亮表示。**(B)** 比较了三个仪器的示踪分数(左)、序列连续读长和QV20+值(右)。这些较小扩增子的测序质量结果几乎相同。

## 困难模板测序

由于较高的GC含量，一些生物体和人类基因组某些区域难以测序。这些区域造成聚合酶滑脱、停止或与测序模板完全分离，妨碍产生高质量结果。此外，基因组的一些区域包含大量高度重复的序列。

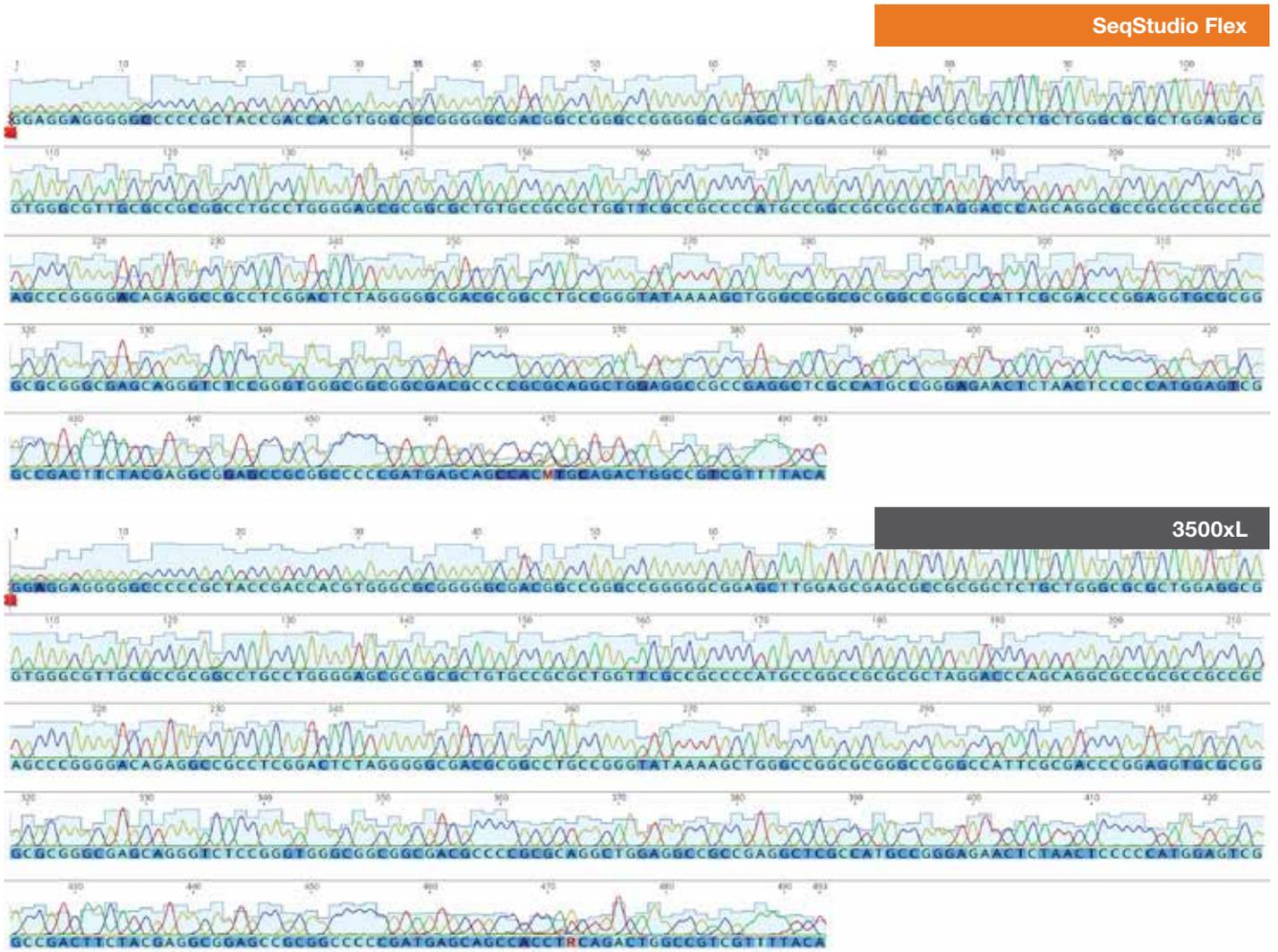
NGS平台难以将此类序列与参考基因组比对。Sanger测序能够克服许多困难，原因是其化学稳健和结果明确，不需要参考基因组。

为了帮助基因组中富含GC序列的PCR扩增，我们研发了GC增强试剂。该添加试剂有助于DNA聚合酶生成可用于循环测序的富含GC序列的模板。在基因分析仪上进行纯化和毛细管电泳后，使用BigDye Terminator或BigDye Direct测序化学。为了分析富含GC或高度重复的序列，基因分析仪需要没有特殊运行模块。

为了确定从困难模板得到的序列的质量，我们分析了一种含平均66% GC区域的人类*CEBPA*基因。对引物进行了选择，以扩增人类*CEBPA*基因中八个不同区域。使用Applied Biosystems™ AmpliTaq Gold™聚合酶和GC增强试剂，对这些区域进行了扩增。使用BigDye Direct循环测序和BigDye Xterminator纯化化学，对扩增子进行了测序。所得反应产物在SeqStudio Flex、3500xL和SeqStudio基因分析仪上运行。

所有仪器能够产生这些富含GC扩增子的高质量结果。比较在SeqStudio Flex和3500xL仪器上得到的电泳图，表明示踪几乎相同(图10A)。虽然仅给出了一个代表性测序图谱，但是看到与其他扩增子和两个测序方向相似的结果。质控分析显示各指标的数值几乎相同，支持了这些结果(图10B)。这些结果表明，即使使用困难模板，SeqStudio Flex仪器生成的数据与其他Applied Biosystems基因分析仪上生成的数据相同。

A



B

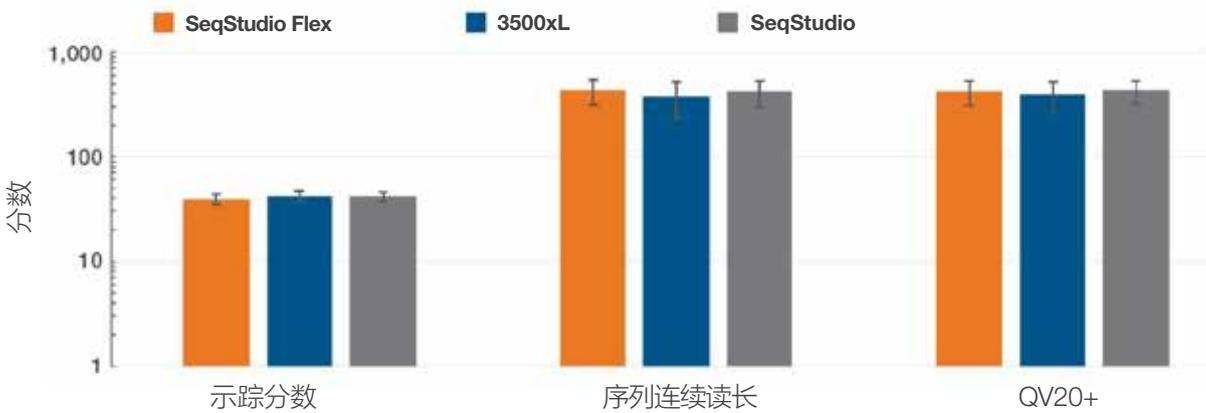


图10. 困难区域测序。测序前，使用GC增强试剂，对平均66% GC含量的人类*CEBPA*基因部分进行扩增。(A) 给出了在SeqStudio Flex和 3500xL基因分析仪上得到的一个扩增子的图谱图谱。特定区域为77.8% GC。注意这些图谱几乎相同。(B) 比较了所有三个仪器所有扩增子的示踪分数(左)、序列连续读长和QV20+值。这些困难扩增子的测序质量结果几乎相同。

## Sanger测序进行基因组编辑验证

基因组编辑彻底改变了进行生物研究的方式。由于其简单和高效，从建立特异性SNP，到敲除基因功能，再到进行大规模基因组重排，研究者可对基因组做出任何类型的改变。

大多数研究人员描述一个连续基因组编辑过程，它从设计指导RNA开始到包含预期编辑的稳定细胞系结束。然而，在整个过程中，对于一些步骤，有必要了解编辑效率。通常，这需要监测细胞培养是否存在和预期编辑的频率。在许多情况下，编辑效率与gRNA和CRISPR相关酶的效率有关。因此，研究人员需要简单的方法来分析他们基因组编辑反应的效率。

基因组编辑验证通常利用使用BigDye Terminator或BigDye Direct化学的循环测序结合基因分析仪的纯化和电泳。为了便于基因组编辑验证，我们开发了Applied Biosystems™ SeqScreener基因编辑验证应用程序。该突破性软件用于分析编辑效率，并显示包含最优编辑的克隆。

在SeqStudio Flex仪器上进行基因组编辑验证，我们生成了人类早老素基因(PSEN1)中的96次CRISPR重组编辑。对于该实验，我们使用不同的指导RNA、转染方法和同源重组修复模板，得出多种编辑效率。使用PSEN1引物、BigDye Direct循环测序、BigDye Xterminator纯化以及SeqStudio Flex、3500xL和SeqStudio仪器上的电泳图，对转染培养的gDNA进行分离和测序。使用Sequence Scanner和SeqScreener基因编辑验证应用程序软件，对所得测序图谱进行分析。

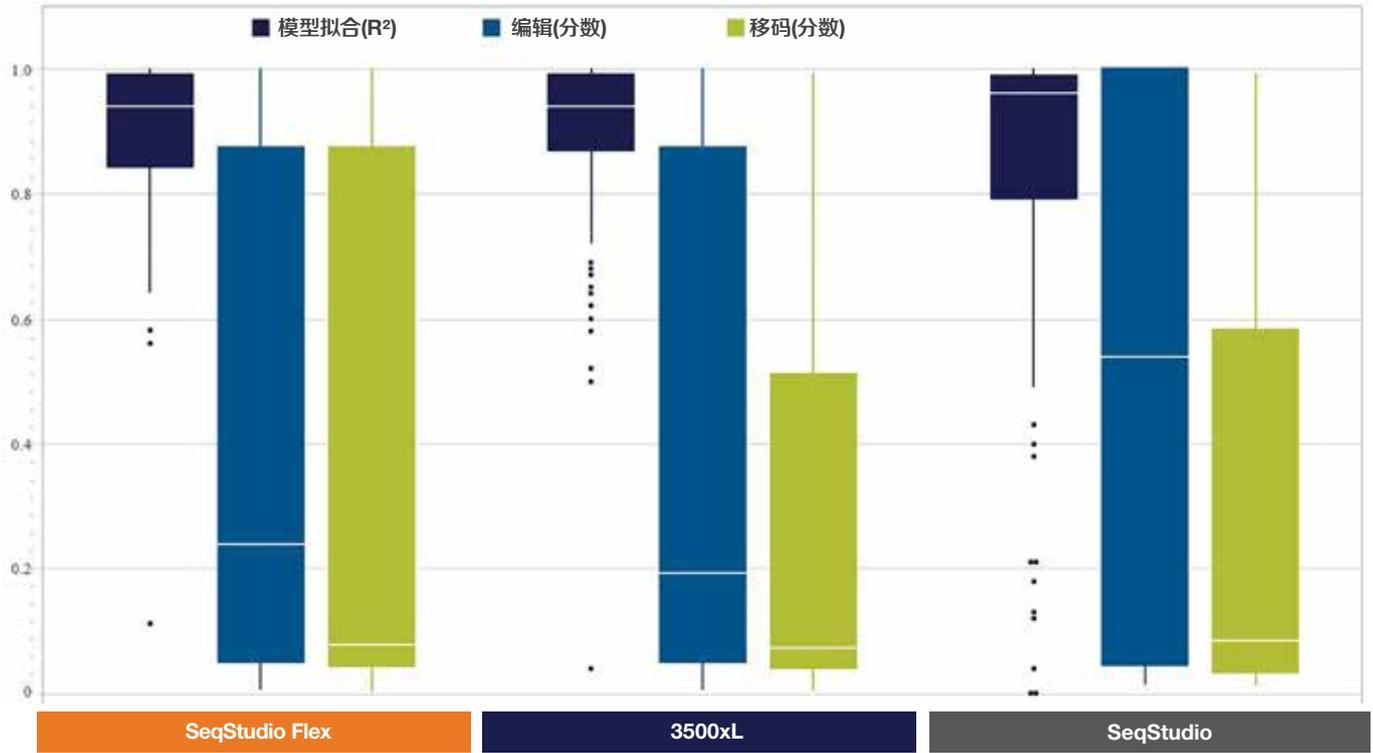
SeqScreener基因编辑验证软件通过几个不同方式计算和显示结果。一种方式是通过产生结果的数值指标(图11A)。模型拟合( $R^2$ )是结果的总体置信度。通常，高质量结果的 $R^2$ 值为0.8或更大；然而，该阈值因样本而变化。编辑值是所测样本中所有成功编辑的累积分数。数值范围广泛表示检测的所有不同编辑条件的效率范围广泛。

移码值是预计消除编码蛋白表达的基因编辑的分数。当进行功能缺失基因敲除筛选时，该信息有用。三个基因分析仪产生非常相似的基因组编辑效率结果。

SeqScreener软件显示结果的另一种方式是在信息图中显示96孔板的孔分析状态，为检测板内包含的结果建模(图11B)。在该图示中，绿孔包含成功编辑，蓝孔表示建议对测序示踪进行目视检查，灰孔表示野生型(W)、对照(C)或问题(!)。标星孔表示包含细胞的孔，是高效、理想化的编辑。点击一个孔可看到关于该孔成功情况的更多信息。

三个仪器产生的孔板信息图非常相似。存在一些差异，与SeqStudio Flex和3500xL系统相比，SeqStudio系统的大多数孔为目视检查与问题分类。这可能是由于序列读序质量下降，因为孔板被多次使用。然而，这些结果证明，对于监测基因组编辑效率，SeqStudio Flex基因分析仪的性能与3500xL和SeqStudio系统的性能相同。

A



B



**图11. Sanger测序进行基因组编辑验证。** (A) 通过Sanger测序和SeqScreener基因组编辑验证应用程序软件，对PSEN1编辑实验的96个样本进行了分析。软件计算了每个孔编辑的效率，并产生三个指标：模型拟合 $R^2$ 值、编辑的孔中DNA分数、移码的孔中编辑分数。全部三个仪器生成的测序数据产生相同的编辑效率指标。(B) SeqScreener基因组编辑验证应用程序软件还生成显示每孔编辑效率的孔板图。绿孔包含成功编辑，蓝孔表示建议对测序示踪进行目视检查，灰孔表示野生型(W)、对照(C)或问题(I)。标星孔表示包含细胞的孔，是高效、理想化的编辑。点击一个孔可看到关于该孔成功情况的更多信息(孔板图右侧)。三个仪器生成的孔板信息图非常相似；大多数差异与是否需要目视检查有关，而不是完全差异。

# 结论

毛细管电泳是一种有效的基因分析方法。无论是片段分析或Sanger测序，CE的灵活性在许多类型的应用可展现。SeqStudio Flex系统是一种先进的基于CE的基因分析仪，有助于发现新的突破性研究。尤其是SeqStudio Flex系统生成的数据质量与其他基因分析仪的数据质量相同，可轻松过渡到新平台。关于该Applied Biosystems SeqStudio Flex基因分析仪的更多信息，访问[thermofisher.cn/seqstudioflex](http://thermofisher.cn/seqstudioflex)或联系本地Thermo Fisher合作伙伴。

## 参考文献

1. <https://iclac.org/databases/cross-contaminations/>
2. Horbach SPJM, Halfman W (2017) The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLoS ONE* 12(10): e0186281. <https://doi.org/10.1371/journal.pone.0186281>
3. Authenticating human cell lines using CLA IdentiFiler and CLA GlobalFiler kits on capillary electrophoresis platforms. <http://assets.thermofisher.cn/TFS-Assets/GSD/Application-Notes/authenticating-human-cell-lines-str-kits-capillary-electrophoresis-application-note.pdf>
4. ATCC STR database. <https://www.atcc.org/en/search-str-database>
5. Gologan A, Sepulveda AR (2005) Microsatellite instability and DNA mismatch repair deficiency testing in hereditary and sporadic gastrointestinal cancers. *Clin Lab Med* 25(1):179–196. doi: 10.1016/j.cl.2004.12.001. PMID: 15749237.
6. Zhao P, Li L, Jiang X et al. (2019) Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. *J Hematol Oncol* 12:54. <https://doi.org/10.1186/s13045-019-0738-1>
7. Reyes GX et al. (2015) New insights into the mechanism of DNA mismatch repair. *Chromosoma* 124:443–462.
8. Boland CR et al. (1998) A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 58:5248–5257.
9. TrueMark MSI Assay For Microsatellite Instability Testing User Guide. [https://assets.thermofisher.cn/TFS-Assets%2FMSG%2Fmanuals%2FMAN0018868\\_MSI%20Assay\\_UG.pdf](https://assets.thermofisher.cn/TFS-Assets%2FMSG%2Fmanuals%2FMAN0018868_MSI%20Assay_UG.pdf)
10. Protocol for Sanger sequencing of the SARS-CoV-2 spike (S) gene. <https://assets.thermofisher.cn/TFS-Assets/GSD/brochures/sequencing-sars-cov-2-spike-gene-protocol.pdf>

## 订购信息

产品	数量	货号
<b>仪器</b>		
<b>SeqStudio 8 Flex基因分析仪, 8通道</b>		A53627
<b>SeqStudio 24 Flex基因分析仪, 24通道</b>		A53630
包括: 带数据采集软件的SeqStudio 24 Flex基因分析仪, 1年保修期、1天现场SmartStart培训以及用于系统性能检查的DNA测序和片段分析试剂盒		
<b>试剂</b>		
CLA Identifiler Plus PCR扩增试剂盒	200次反应	A44660
CLA Identifiler Direct PCR扩增试剂盒	200次反应	A44661
CLA GlobalFiler PCR扩增试剂盒	200次反应	A44662
TrueMark MSI检测	100次反应	A45295
BigDye Terminator v1.1循环测序试剂盒	100次反应	4337450
BigDye Terminator v3.1循环测序试剂盒	100次反应	4337455
BigDye Direct循环测序试剂盒	100次反应	4458687
BigDye XTerminator纯化试剂盒	100次制备	4376486
ExoSAP-IT PCR产物净化试剂	100次反应	78200.200.UL
RecoverAll FFPE总核酸分离试剂盒	40次反应	AM1975

## 赛默飞世尔科技

---

### 上海

上海市浦东新区新金桥路27号3,6,7号楼  
邮编 201206  
电话 021-68654588

### 成都

成都市临江西路1号川投大厦1406室  
邮编 610041  
电话 028-65545388\*5300

### 南京

南京市中央路201号金茂广场南楼1103室  
邮编 210000  
电话 021-68654588\*2901

### 北京

北京市东城区北三环东路36号环球贸易中心C座7层/8层  
邮编 100013  
电话 010-87946888

### 沈阳

沈阳市沈河区惠工街10号卓越大厦3109室  
邮编 110013  
电话 024-31096388\*3901

### 西安

西安市高新区科技路38号林凯国际大厦1006-08单元  
邮编 710075  
电话 029-84500588\*3801

### 广州

广州国际生物岛寰宇三路36、38号合景星辉广场北塔204-206单元  
邮编 510000  
电话 020-82401600

### 武汉

武汉市高新四路22号58众创光谷产业园A座1楼2-5楼  
邮编 430075  
电话 027-59744988\*5401

欲了解更多信息，请扫描二维码关注我们的微信公众账号与官方网站。

赛默飞世尔科技在全国有共14个商业办公室。本资料中的信息，说明和技术指标如有变更，恕不另行通知。

更多信息请见 [thermofisher.cn/seqstudioflex](http://thermofisher.cn/seqstudioflex)



赛默飞  
官方微信



赛默飞  
官方微信

Applied Biosystems

免费 800 820 8982

服务电话 400 820 8982

**ThermoFisher**  
SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures. © 2022 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. TaqMan is a registered trademark of Roche Molecular Systems, Inc., used under permission and license. Geneious is a trademark of Biomatters Limited. pGEM is a trademark of Promega Corporation. **EXT2083 0122**