

# 采用 Agilent Avida DNA 工作流程和 SeraSeq ctDNA Mutation Mix v4 参考品， 全面评估对 cfDNA 的检测性能

## 作者

Margherita Corioni 博士，  
Xiaomu Chem 硕士，  
Heng Wang 硕士和  
Grace Zhao 博士，  
研发科学家，  
安捷伦科技公司

## 来自 LGC Clinical Genomics 的合作伙伴

Andrew Anfora，  
产品管理副总监

David Merriam，  
商务总监

Yves Konigshofer，  
技术开发总监

## 前言

使用二代测序 (NGS) 技术分析游离 DNA (cfDNA) 极大地提高了我们对癌症生物学的理解，同时推动了无创分子诊断技术的发展。然而，传统的 NGS 工作流程往往较为复杂耗时。此外，每个步骤（尤其是纯化环节）都容易导致 DNA 损失，因此难以分析起始量较少的样品。

Agilent Avida DNA 工作流程采用新型杂交技术，无需在杂交捕获之前的文库制备过程中进行扩增和纯化。这种简化的工作流程减少了 PCR 引起的偏倚，并将杂交时间缩短至一小时，也不会影响回收率。这对于需要高灵敏度和高特异性的低起始量样品分析至关重要。此外，尽可能减少与 PCR 扩增相关的偏倚，有利于更准确地检测拷贝数变异 (CNV)。

在评估 cfDNA NGS 的分析性能以及在产品开发过程中进行质控时，使用稳定、全面且变异内容准确的参比物质至关重要。在本应用简报中，我们使用了 SeraSeq ctDNA Mutation Mix v4 参比物质，因为它具有以下优势：全面覆盖癌症相关变异，与之前的 ctDNA 参比物质相比，背景噪音更低，且在片段分子量分布和连接效率方面更接近真实 cfDNA 样本<sup>[1]</sup>。

Avida 技术可捕获目标 DNA 的正链 (+) 和负链 (-)，并使用分子标识符 (UMIs) 进行错误校正和准确的分子计数。我们使用 UMIs 报告了

- 不同起始量下，三个肿瘤基因 Panel (26 kb、345 kb 和 2.7 Mb) 的 DNA 分子回收情况，每个 Panel 均在不同的测序深度下进行测序
- 三个 Panel 的变异检测结果，并对中等规模的 Panel (345 kb) 进行了深入的评估，该 Panel 包含肿瘤学研究相关的重要变异，同时测序成本经济可接受

本应用简报中报告的测序指标可以推广应用于其他 Avida 目录 Panel 或定制 Panel，用于预测在给定的样品起始量和测序深度下可检测到的变异等位基因频率 (VAF) 以及分子回收情况。

## 方法

### 样品与样品前处理

直接使用制造商提供的 Sereq ctDNA Mutation Mix v4 (货号：0710-3101、0710-3099 和 0710-3100, SeraCare)，VAF 分别为 0% (野生型)、0.5% 和 5%。该参比物质包含的变异的完整列表可在 [SeraCare 网页](#) 上找到。使用 0% 野生型储备液稀释 0.5% 和 5% VAF 储备液，得到 0.2%、1% 和 2% VAF 样品。使用 Agilent 4200 TapeStation 系统 (货号 G2991A) 和 Agilent D1000 DNA ScreenTape (货号 5067-5582) 以及相应试剂 (货号 5067-5583) 确认并定量了 Sereq ctDNA Mutation Mix v4 的浓度。

使用安捷伦的三种 Avida 目录 Panel 对各种起始量的参比物质进行靶向序列捕获：

- Avida DNA Focused 癌症基因 Panel (货号 5280-0050)
- Avida DNA Expanded 癌症基因 Panel (货号 5280-0047)
- Avida DNA Discovery 癌症基因 Panel (货号 5280-0044)

每种实验条件下至少进行了两次技术重复测试。

### Panel 详细信息

- Avida DNA Expanded 癌症基因 Panel：涵盖 105 种相关癌症基因，其中包括 Sereq ctDNA v4 参比物质中的 60 种单核苷酸变异 (SNVs) 和插入/缺失 (indels)、7 种易位和 7 种 CNVs。由于该 Panel 规模合适且与参比物质有较高的重叠，因此本次研究优先选择该 Panel

- Avida DNA Focused 癌症基因 Panel：一款非常小型的癌症基因 Panel (26 kb)，经过优化，能够高效捕获特定的癌症靶点，主要包括热点突变和外显子突变，覆盖了 14 个肿瘤学关键基因
- Avida DNA Discovery 癌症基因 Panel：Avida 癌症基因 Panel 系列中最大的基因 Panel (2.7 Mb)，涵盖超过 680 个关键基因和生物标志物，用于肿瘤学研究中的发现和评估

Focused Panel 和 Discovery Panel 分别包含 Sereq ctDNA v4 参比物质中的 23 种和 66 种 SNVs 及插入缺失变异。我们将展示使用这些 Panel 进行等位基因频率 (AF) 检测。

### 文库制备与靶向序列捕获

ctDNA 参比物质按照 Avida DNA 试剂盒 (G9418A, 安捷伦) 制造商的方案，使用 Focused、Expanded 或 Discovery 癌症基因 Panel 进行处理。使用 D1000 ScreenTape 对加标签的 PCR 产物进行定量分析，然后再进行混合和测序。

### 数据分析

Avida 接头中包含 UMIs，UMIs 是随机寡核苷酸短序列，用于在 PCR 扩增之前对样本文库中的每个分子进行标记。使用 UMI 条形码进行测序比对可以降低变异检测的假阳性率，并提高对起始 DNA 样品中存在的真实突变的检测能力。该方法利用 PCR 重复序列，根据唯一 UMIs 对测序 reads 进行分组，并根据 reads 的频率和质量评分确定准确的共有序列。该过程利用重复序列进行序列错误校正，从而提高数据质量。“结果”部分讨论了测序预算、重复率和变异检测之间的关系。

使用起始/终止位点的位置信息和内嵌的 UMIs，在单链和/或双链模式下对测序 reads 进行去重，并使用 bwa-mem 与基因组比对。

使用 VarDict (1.5.0 版) 鉴定 Sereq ctDNA Mutation Mix v4 参比物质中的 SNV 和插入缺失。使用 CNVkit (0.9.8 版) 分析 CNVs，并使用 GeneFuse (0.6 版) 检测融合。

在本研究中，检出率 (灵敏度) 定义为分析 Panel 涵盖的 Sereq ctDNA Mutation Mix v4 参比物质中的变异，成功被检测到的比例。假阳性率 (特异性) 定义为在野生型 (0%) 参比物质中检测到的变异比例。

## 结果

### 测序 reads 数量选择和基于 UMI 的去重与错误校正

本研究在 Avida DNA 工作流程中使用了 10–20 ng 的 Seraseq ctDNA Mutation Mix v4 参比物质以及以下 Avida 癌症目录 Panel: Focused (26 kb)、Expanded (345 kb) 和 Discovery (2.7 Mb)。根据 Panel 大小和 DNA 起始量调整测序量。表 1 为本研究中测序 read 对的数量与设计内容的平均原始中位覆盖度之间的相关性示例。

测序 read 对预算是变异检测分析的一个重要参数，必须选择合适的测序量预算，才能确保对特定变异达到足够的覆盖度，满足检测阈值要求。通常，起始 DNA 量越高、需要检测的 AF 越低，需要的 read 预算也就越高。如“方法”部分所述，测序 reads 的数量也直接关系到 UMI 去重过程的准确性，而 UMI 去重过程会为成功识别的每个唯一 UMI 选择最准确的共有序列。

考虑到起始量和捕获步骤中 Panel 的大小，所选的样本测序深度应使去重步骤中的重复率接近或超过 80%（针对 UMI 单链模式去重）。这样可以对样本中的唯一 UMIs 进行全面筛查，并在去重过程中为每个 UMI 分配唯一共有序列。

在利用内嵌 UMI 的高精度去重来检测低频变异时发现，当重复率显著低于 80% 时，有相当一部分分子未被测序，因此无法为变异检测提供有效信息。样本测序不足会导致只能获得文库的部分信息，无法全面反应文库的复杂性（样本中独特分子的数量），从而更有可能漏检罕见变异。然而，增加测序深度，使重复率远高于 80%，在识别更多独特分子以及了解文库组成方面的帮助有限。但是，如果条件允许，增加测序量仍是可取的，因为它有助于增加每个 UMI 家族的测序深度，从而支持错误校正。

表 1. 使用 Agilent Avida 癌症基因 Panel，通过 UMI 单链模式去重进行 UMI 校正时，测序深度对设计内容的中位覆盖度和重复率的影响。测序预算以 read 对来表示。总 reads 预算是 read 对数量的两倍

Panel	分子量 [kb]	起始量 [ng]	测序 Read 对 [M]	原始中位覆盖度 [k]	重复率 [%]
Avida DNA Focused 癌症基因 Panel	26	20	15	40	83
		20	10	27	76
		20	5	13	60
Avida DNA Expanded 癌症基因 Panel	345	10	50	14	88
		10	40	10.6	78
		10	20	5.3	62
	345	20	80	21.6	81
		20	50	14	72
		20	40	10.8	66
		20	20	5.4	47
Avida DNA Discovery 癌症基因 Panel	2700	20	400	19	80
		20	200	9	64
		20	100	4.7	45

### 三种癌症基因 Panel 的总体性能（以 20 ng 参比物质作为起始样品）

使用 20 ng 起始量，通过 Avida DNA Expanded 癌症基因 Panel 捕获 Seraseq ctDNA Mutation Mix v4 参比文库，在 80 M 测序 read 对（总计 160 M reads）预算下，该 Panel 的平均覆盖深度达到了 4000x（表 2）。除了高覆盖深度之外，覆盖也很均匀，其中超过 95% 的目标碱基在去重后，其覆盖深度达到或超过预期平均覆盖深度的一半。这意味着 Panel 上超过 95% 的区域（包括外显子区域和内含子区域）的覆盖深度超过 2000x。

因此，当 AF 为 0.2% 且起始量为 20 ng 时，预计在 95% 的基因组区域内，每个位点的平均拷贝数至少为 4 拷贝，而变异的平均拷贝数为 8 拷贝。基于上述条件，在使用 Expanded Panel 分析其涵盖的 Seraseq ctDNA Mutation Mix v4 参比中的 60 种 SNV 和插入/缺失时，在 AF 为 0.2% 且起始量为 20 ng 时，检出率达到 94%（表 2）。

使用 20 ng 0.2% VAF 参比 DNA 作为 Avida DNA Focused Panel (26 kb) 的起始样品（涵盖 23 种变异），测序预算为 10 M read 对（总 reads 20 M），SNV 的检出率为 100%（表 2）。在平均 UMI 覆盖深度为 6000x 的情况下（表 2），预计可以检测到 12 个变异拷贝，因此 100% 的检出率与预期一致。

对于大型 Avida DNA Discovery Panel (2.7 Mb)，使用 20 ng 起始量 (0.5% VAF)，测序得到约 170 M read 对（总计 340 M reads），SNVs 和插入/缺失（涵盖 66 种变异）的检出率为 94%（表 2）。根据 Expanded 和 Focused Panel 检测的相同逻辑，在给定起始量、Panel 大小和测序预算的情况，实现了 2600x 的平均 UMI 覆盖深度，预计每个变异平均有 13 个拷贝。值得注意的是，对于这个相对大型的 Panel，给定测序深度下的重复率仅为 68%（表 2）。增加测序深度将能捕获更多 UMI 并可能提高检测灵敏度。

使用相同的起始量和相似的测序深度分析野生型 (0%) 参比物质，检测结果显示假阳性率很低。在 20 ng 起始量下，其中一个样本使用 Focused Panel 以及 10 M read 对（相当于总计 20 M reads）得到一个假阳性结果；另一个样本使用 Expanded Panel 以及 80 M read 对（相当于总计 160 M reads）得到四个假阳性结果。其他的 Panel、起始量和测序深度组合下均未检测到假阳性。

表 2. 使用 20 ng Seraseq ctDNA Mutation Mix v4 作为起始样品时，Agilent Avida DNA Focused 癌症基因 Panel、Avida DNA Expanded 癌症基因 Panel 和 Avida DNA Discovery 癌症基因 Panel 的一般性能指标：分子回收率（UMI 去重后的覆盖度）和变异检测的灵敏度

Panel	分子量 [kb]	起始量 [ng]	覆盖的 SNV/ 插入缺失	预期 VAF [%]	Read 对 [M]	原始 Reads 覆盖度 [k]	UMI 去重后的覆盖度 [k]	重复率 [%]	实测 VAF [%]	检测灵敏度 [%]
Avida DNA Focused 癌症基因 Panel	26	20	23	0.2	10	27	6	78	0.2	100
Avida DNA Expanded 癌症基因 Panel	345	20	60	0.2	80	21	4.3	81	0.2	94
Avida DNA Discovery 癌症基因 Panel	2700	20	66	0.5	172	7.5	2.6	68	0.5	95

此处所示的测序预算（read 对）经过精心选择，足以实现充分的错误校正和独特分子回收，从而确保样品分析的准确性（重复率%）。测序预算以 read 对来表示。总 reads 预算是 read 对数量的两倍。

## 不同的起始量和测序预算组合下 Avida DNA Expanded 癌症基因 Panel 的检测性能

检测低频变异时（如本研究中），强烈建议准确确定所需的测序预算，应该根据起始量、目标 AF 和捕获 Panel 的大小来确定。理想情况下，分析中的原始中位覆盖度应该足够高，以使样品重复率达到 80% 以上。这表明已经筛选出大量独特分子，并可以准确识别捕获的独特分子，从而能够准确地分析目标样品（见表 2）。

表 3.

A. 使用 Avida DNA Expanded 癌症基因 Panel，并以 20–40 ng Seraseq ctDNA Mutation Mix v4 参比物质作为起始样品，获得的变异等位基因频率 (VAF) 检测结果

起始量 [ng]	测序 Read 对 [M]	重复率 [%]	VAF [%]		
			0.1	0.2	0.5
40	160	82	92	98	
	50	26	86	96	
	25	13	73	91	
	13	6	51	76	
20	80	81		94	100
	50	51		92	100
	25	25		87	98
	13	13		69	96

测序预算对错误校正和回收的分子数量（以重复率%表示）的影响，以及在低 VAF (0.1%–0.5%) 下的变异检测灵敏度。

B. 使用 Avida DNA Expanded 癌症基因 Panel，并以 Seraseq ctDNA Mutation Mix v4 参比物质作为起始样品，在 1–10 ng 超低起始量下获得的变异等位基因频率 (VAF) 检测结果

起始量 [ng]	测序 Read 对 [M]	重复率 [%]	VAF [%]			
			0.5	1	2	5
10	46	81	96			
	32	80		98		
	39	78			99	
5	30	86	82			
	57	93		97		
	68	93				100
3	39	92		98		
1	33	92			98	
	11	94				93

通过测序 read 预算优化分子回收和错误校正（以重复率%表示）的示例，以及在 VAF 0.5%–5% 范围内的检测灵敏度。Expanded Panel 涵盖 Seraseq ctDNA Mutation Mix v4 中的 7 种 CNVs 和 7 种易位变异。两种起始量下的 CNV 检出率均为 100%（表 3C）。0.2% VAF 下的 CNVs 平均倍数变化约为 1.1 倍，0.5% VAF 下约为 1.3 倍。对于易位检测，检出率分别为 93% 和 79%。在使用野生型 (0%) 参比 DNA 进行检测时，在相同的起始量和测序深度条件下，没有检测到假阳性。

C. CNV 和易位

变异类型	VAF [%]	起始量 [ng]	测序 Read 对 [M]	重复率 [%]	检测灵敏度 [%]
CNV	0.2	20	80	90	100
CNV	0.5	10	66.6	81	100
易位	0.2	20	80	90	93
易位	0.5	10	66.6	81	79

鉴于测序深度可能因各种因素受到限制，本研究在减少测序数据量的情况下进行了分析，并给出了相应的重复百分比以及对 AF 检出限的影响。这展示了 reads 数量对变异检出率的影响，并有助于合理规划测序深度以获得尽可能准确的检测结果（见表 1 和 3A）。

有时，用于检测的 DNA 样品量可能有限，或者可能难以估算测序量，所有这些都影响方法的检测能力。表 3A 和 3B 提供了使用 Avida DNA Expanded 癌症基因 Panel 在不同的起始量和测序预算下得出的不同 VAF 的检出率，作为参考。

## Avida DNA Expanded 癌症基因 Panel 的靶标覆盖度统计数据

为了更全面地展示 Avida DNA Expanded 癌症基因 Panel 对起始参比物质的靶向序列捕获能力，我们提供了不同靶点的覆盖度统计数据。图 1 为使用 20 ng 参比物质、Expanded Panel 和测序生成的 80 M read 对（总计 160 M reads）获得的靶标区域的覆盖度数据。所有靶标区域均实现 1000x 覆盖，97% 实现 2000x 覆盖，87% 实现超过 3000x 覆盖。整个 Panel 的平均覆盖度为 4425x。超过 87% 的区域覆盖度超过 3000x（图 1A）。

图 1B 为使用 Expanded Panel 时关键癌症相关基因的平均覆盖度。整个外显子区域的平均覆盖度达到了 3000x 以上。

图 1C 为参比物质中部分变异的检测情况，这些变异与 Expanded Panel 以及图 1B 中显示的基因有重叠。起始量为 20 ng，测序获得了 80 M read 对（总计 160 M reads），检测的预期等位基因频率 (AF) 为 0.5% 和 0.2%。

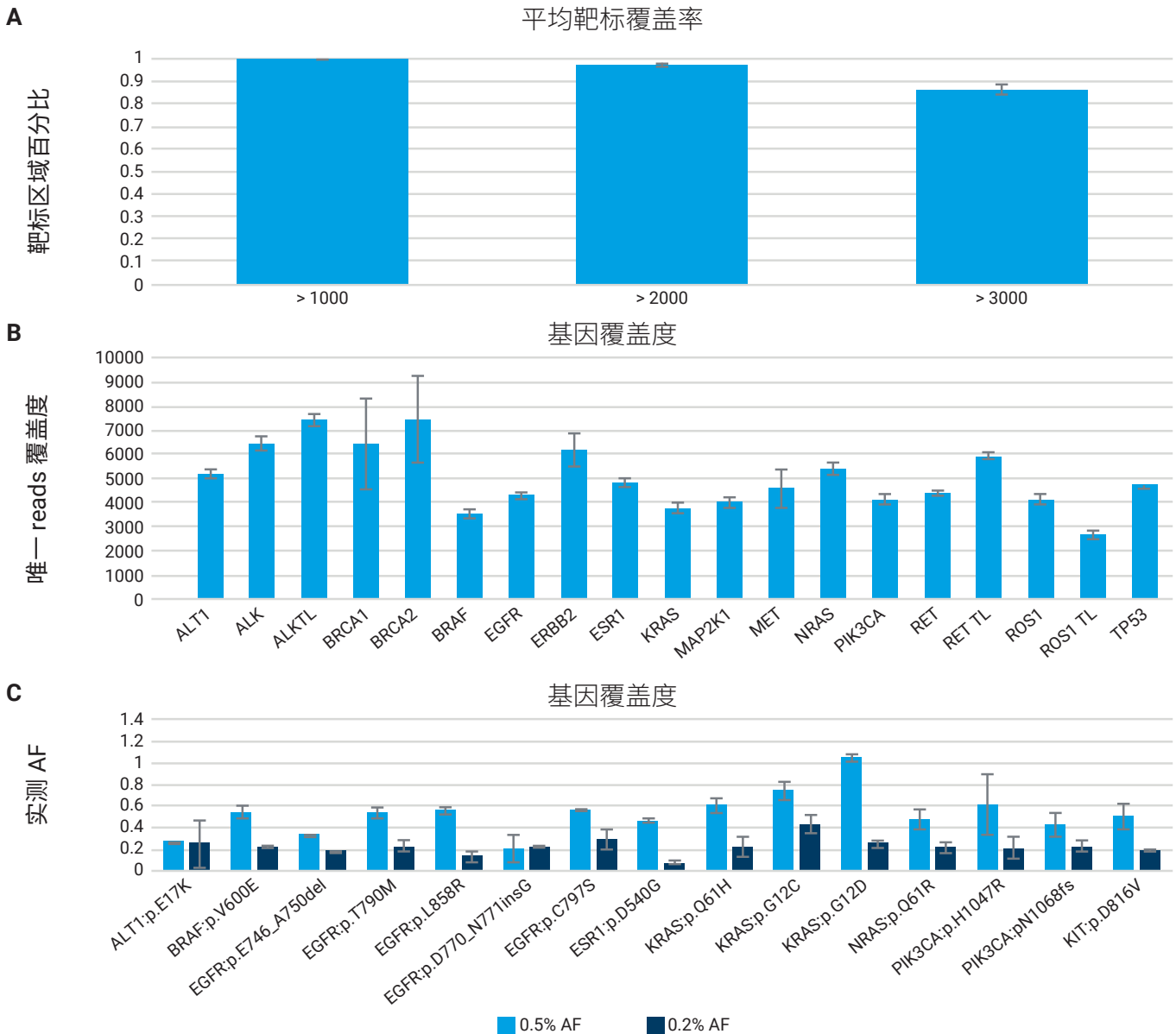


图 1. 使用 20 ng Seraseq ctDNA Mutation Mix v4 参比物质作为起始样品，通过 Agilent Avida DNA Expanded 癌症基因 Panel 进行靶向序列捕获，获得的选定基因的回检情况。A. 根据 80 M 测序 read 对（总计 160 M reads）计算得到的靶标区域的中位覆盖度统计数据（旨在作为 UMI 去重后的唯一 reads 覆盖度）。中位覆盖度来自 6 次重复实验 (N = 6)。B. 癌症相关的关键特征基因的唯一 reads 覆盖度，这些结果由 80 M 测序 read 对（总计 160 M reads）计算得出（“TL”表示特意加入了非外显子靶标区域，以增强对基因易位的检测能力）。C. 从参比物质的 80 M 测序 read 对（总计 160 M reads）中实际检测到的变异与预期变异 (AF 为 0.5% 和 0.2%)。平均 AF 来自 20 ng 起始样品的 2 次重复实验 (N = 2)

## 结论

评估 cfDNA NGS 方法的分析性能非常关键，而使用高质量的参比物质对于该评估过程至关重要。SeraSeq ctDNA Mutation Mix v4 包含多种变异且背景噪音低，经研究证明是一种高效、全面的工具，非常适合用于评估 Avida DNA 工作流程的变异检测性能。

结果表明，以 10–20 ng SeraSeq ctDNA 参比物质作为起始样品，Avida DNA 工作流程能够可靠地捕获足够的独特 DNA 分子，从而能够在 0.5% 和 0.2% VAF 下实现可靠的变异检测。

此外，通过减少测序数据量，使用少于推荐数量的 read 对进行变异检测，Avida 技术仍然能够准确地检测变异，并提供有关样品组成的可靠检测结果。这使得在起始量和测序预算的分配规划方面具有更高的灵活性，有助于对珍贵且有限的样品进行分析。

如本研究所示，Agilent Avida 支持更宽的起始量范围，而无需修改实验方案，也不会影响检测性能（见表 3），非常适合液体活检应用以及其他需要分析单个样品且对分子回收率要求极高的研究。

## 参考文献

1. Ruminski, L.; et al. *Next generation liquid biopsy reference material performance across NGS assays and platforms*. Presented at AACR Annual Meeting, **2024**. 可从以下网址下载: <https://digital.seracare.com/aacr-liquid-biopsy>

查找当地的安捷伦客户中心:

[www.agilent.com/chem/contactus-cn](http://www.agilent.com/chem/contactus-cn)

免费专线:

800-820-3278, 400-820-3278 (手机用户)

联系我们:

[LSCA-China\\_800@agilent.com](mailto:LSCA-China_800@agilent.com)

在线询价:

[www.agilent.com/chem/erfq-cn](http://www.agilent.com/chem/erfq-cn)

[www.agilent.com/chem/avidadna](http://www.agilent.com/chem/avidadna)

仅供科研使用。不用于临床诊断用途。  
PR7001-3473

本文中的信息、说明和指标如有变更，恕不另行通知。

© 安捷伦科技 (中国) 有限公司, 2024  
2024 年 12 月 12 日, 中国出版  
5994-7953ZH-CN