

开发和利用 PathoChip 微阵列芯片以检测 FFPE 癌症组织中不同的致病微生物特征

作者

Erle S. Robertson
宾夕法尼亚大学
微生物学系
美国宾夕法尼亚州费城

Josh Zhiyong Wang
安捷伦科技公司
美国加利福尼亚州圣克拉拉市

摘要

人们日益意识到有益微生物与有害微生物之间的平衡对于人类健康的重要性。这一平衡会影响人体代谢和免疫反应，而传染性致病因子（包括 HPV 和幽门螺旋杆菌）是致癌风险最高的因素之一。基于这些发现，筛查人肿瘤组织中的数千种病毒、致病菌/真菌和寄生虫有助于揭示它们在癌症进展及治疗反应中的作用。本文将介绍 PathoChip 微阵列芯片的开发，它可以检测福尔马林固定石蜡包埋 (FFPE) 肿瘤组织中来自数千种病毒和致病微生物的 DNA 和 RNA。该方法包括专门针对病原体检测的多个上游样品前处理流程和下游数据分析流程。此外，我们将展示该方法可以成功检测多种癌症组织（包括三阴性乳腺癌和口腔/口咽鳞状细胞癌）中不同的微生物特征。这些癌症通过适合使用独立 PCR 或捕获测序方法验证的特征病毒和微生物病原体进行鉴定。结果表明，使用 PathoChip 可以成功获得复杂 FFPE 样品中全面的病原体信息。

前言

正常的人类微生物组包含数千种微生物。大量的研究集中于已知具有常驻微生物组的组织系统，包括胃肠道、皮肤、气道和免疫系统。病毒、细菌和寄生虫等传染性病原体可能是导致肝、胃、宫颈和血液等各种组织癌变的主要因素。

有许多方法可以检测微生物，包括传统的培养法。但越来越需要基于微生物遗传信息的宏基因组学工具来高效识别与特定疾病相关的传染性病原体。PCR 扩增结合 16s rRNA 测序是鉴定细菌的常用方法，但这种方法并不适用于病毒和真核微生物。对样品总 DNA 进行鸟枪法测序可实现无偏差检测，但宿主 DNA 的高背景会严重影响检测效率。

与此同时，DNA 微阵列芯片因能够快速、经济地对大量样品进行广泛的微生物筛选而崭露头角。市面上已有一些解决方案，但它们涵盖微生物组的不同部分（在某些情况下有所重叠），都无法提供全面的信息。

本应用简报将介绍基于 Agilent SurePrint 微阵列芯片技术的 PathoChip 微阵列芯片（图 1）的开发。PathoChip 微阵列芯片包含针对所有已知的公开病毒序列以及数百种致病菌、真菌和寄生虫的探针，以更经济的方式实现对病原体的广泛覆盖。在可能的情况下，使用多个靶向各基因组不同区域的探针来提高检测性能。此外，虽然 PathoChip 探针内容是根据已知靶标的序列进行开发的，但是由于还设计了针对病毒家族内部和家族之间保守序列的探针，除了识别已知的病原体，仍能够识别新的菌株、毒株或生物体。本文还介绍了用于分析 FFPE 肿瘤样品的支持工作流程，包括同时检测 DNA 和 RNA 靶标。

本研究使用 PathoChip 微阵列芯片成功分析了来自三阴性乳腺癌和口腔/口咽鳞状细胞癌的数百个 FFPE 样品，以检测适合使用独立 PCR 或捕获测序方法验证的微生物特征。

材料与方法

微阵列芯片设计

通过美国国家生物技术信息中心 (NCBI) 基因组、基因和核苷酸编号数据库 (<http://www.ncbi.nlm.nih.gov/pubmed>) 查询所有分类病毒注释，以及通过文献检索和网络资源 (<http://www.niaid.nih.gov>: 新发和再发传染病, A、B 和 C 类优先病原体) 汇编的来自原核和真核人类病原体列表获得目标编号。将得到的编号序列组装为一个宏基因组，并分成 58 条虚拟“染色体”，每条染色体包含大约 500–1000 万个核苷酸 (nt)。如文献所述，识别并包含了该宏基因组中的独特区域和保守区域^[6]。

使用安捷伦微阵列芯片比较基因组杂交 (aCGH) 设计算法，设计了针对宏基因组中独特区域和保守区域（最多 60 个 nt）的探针序列。然后对这些序列进行过滤，以降低与人类基因组序列交叉杂交的可能性，从而降低杂交背景噪音。

在微阵列芯片设计中，如果可用于源编号的探针少于 10 个，则默认将映射到病原体基因组，或任何原核或真核病原体编号的独特或保守区域的探针添加到设计中。当创建了超过 10 个探针/编号时，使用以下标准对探针进行过滤：最多 20 个探针/编号，最小探针间距 100 bp，均匀分布并覆盖整个序列。致癌病毒不受探针数量的限制，创建一个饱和平铺探针组来覆盖所有致癌病毒^[6]。

使用 Agilent SurePrint 8 × 60 K 微阵列芯片，每张芯片可处理 8 个样品。最初创建了两个微阵列芯片 (PathoChip v2a 和 v2b) 来评估探针性能，微阵列芯片包含 60000 个探针，分别靶向独特区域和保守区域。然后将 PathoChip v2a 和 v2b 的高性能探针组合成 PathoChip v3，其中包含针对独特靶标的 37704 个探针，以及针对保守靶标的 23627 个探针。这样可以覆盖所有致癌和致病病毒。使用 PathoChip v3 分析本应用简报中所述的各种 FFPE 癌症组织。

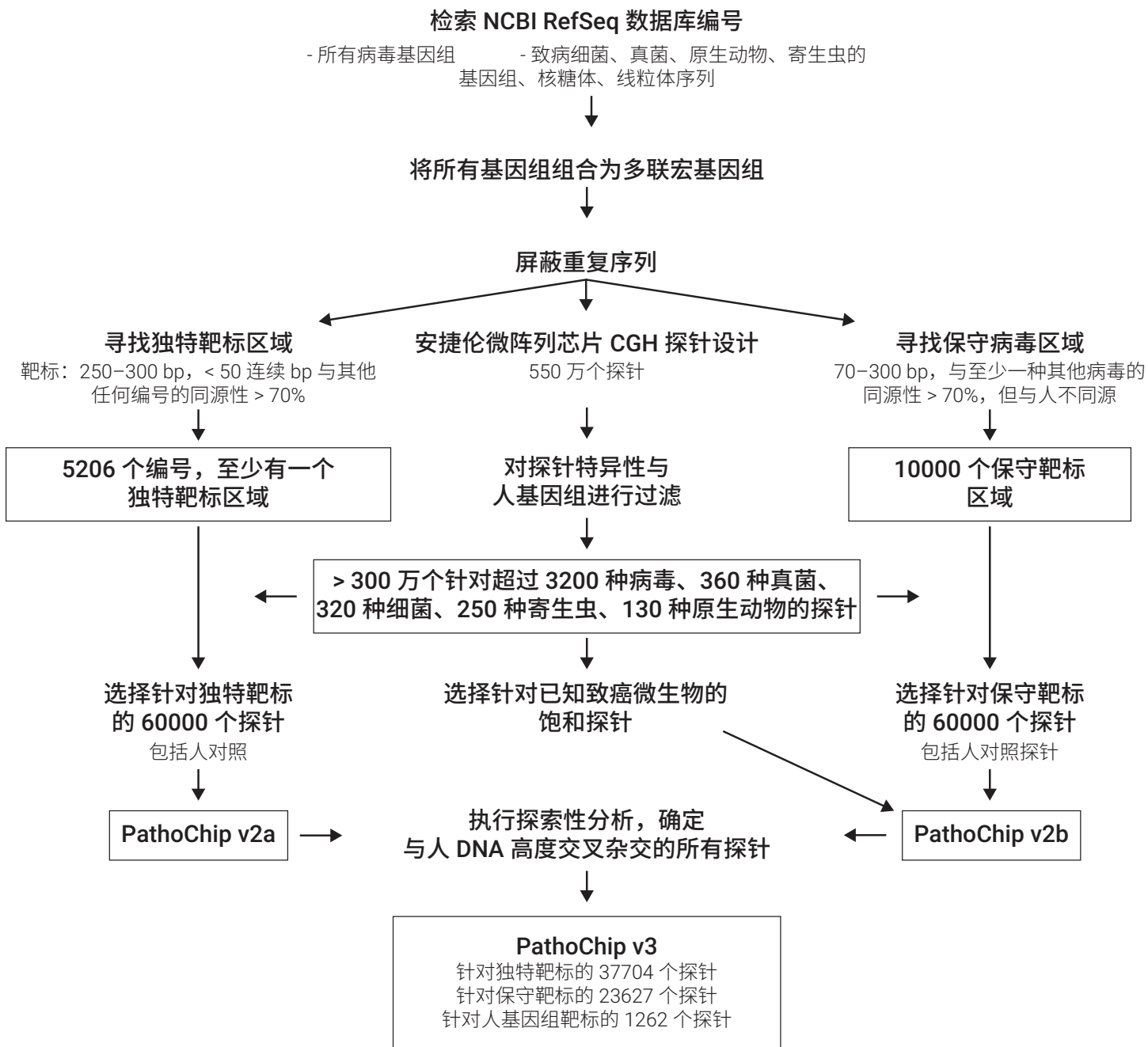


图 1. PathoChip 设计方案和设计迭代

样品

所有样品均来自艾布拉姆森癌症中心的肿瘤组织和生物样本库。由一位病理学住院医师回顾病例病史，并确认了肿瘤类型和癌细胞的划分。如果存在明显的相邻正常组织，将切片封固在不带电的玻片上，使用包含苏木素和伊红 (H&E) 染色切片的模板玻片对肿瘤组织进行研究，并明确划分癌症区域。

收到 100 份去标识的 FFPE 口腔和口咽鳞状细胞癌样品（统称为 OCSCC），均以 10 μm 切片的形式置于不带电玻片上。收到匹配和非匹配对照样品各 20 份^[7]。将来自本研究中 20 名癌症患者的癌旁临床正常样品称为“匹配对照”。而非匹配对照为来自其他健康个体的口腔组织（悬雍垂）。

收到 100 份去标识的 FFPE 三阴性乳腺癌样品，均以 10 μm 切片的形式置于不带电玻片上。17 份匹配对照样品和 20 份非匹配对照样品以石蜡卷的形式提供^[8]。匹配对照与癌症组织来自相同的患者，并采集自相邻的非癌乳腺组织。非匹配对照为来自健康个体的乳腺组织。

样品 DNA/RNA 提取、扩增、标记和杂交

标准安捷伦 CGH 检测方法使用基因组 DNA 作为起始材料，但由于病毒可以以 DNA 或 RNA 作为遗传物质，因此我们试图对流程进行修改。具体而言，我们使用 AllPrep DNA/RNA FFPE 试剂盒 (QIAGEN) 对 FFPE 肿瘤依次进行 DNA 和 RNA 提取。核酸质量控制评估包括 A260/280 比值、产率及通过琼脂糖凝胶电泳获得的大小分布。虽然一些 FFPE RNA 样品显示部分降解，但大部分样品的 RNA 片段大小和提取率足以使用总 RNA 直接生成 cDNA。

将 50 ng 测试样品基因组 DNA 和 50 ng 测试样品总 RNA 一起作为 TransPlex WTA-2 试剂盒 (Sigma-Aldrich) 的起始样品。该试剂盒能够同时扩增基因组 DNA 和 cDNA。使用 QIAquick PCR 纯化试剂盒 (QIAGEN) 纯化扩增产物，然后取 2 μg 纯化

后的扩增产物，使用 SureTag 标记试剂盒（安捷伦）对测试样品进行 Cy3 染料标记。对于参比对照样品，使用 TransPlex WTA-2 试剂盒对 50 ng 病毒阴性参比基因组 DNA (BJAB 细胞系) 进行扩增，然后取 2 μg 纯化后的扩增产物，使用 Agilent SureTag 标记试剂盒进行 Cy5 染料标记。将该参比样品作为对照，以报告探针与人 DNA 的交叉杂交情况。

使用 SureTag 标记试剂盒中的离心柱纯化 Cy3 和 Cy5 标记的 DNA。测量标记 DNA 的比活性，然后将其混合用于杂交。在安捷伦杂交炉中，按照安捷伦建议的 65 °C 杂交温度和 20 rpm 转速，将标记 DNA 与 PathoChip v3 8 × 60 K 微阵列芯片杂交 40 小时。使用标准清洗程序处理微阵列芯片，然后在 Agilent SureScan 微阵列芯片扫描仪 C 或 D (货号 G2565AA 和 G4900DA) 上进行扫描。

微阵列芯片数据分析

使用安捷伦特征提取软件对扫描的微阵列芯片图像进行分析，计算平均像素强度并扣除每个特征的局部背景。作为 QC 流程的一部分，手动检查图像，记录任何受高背景、划痕或其他超出质量控制 (QC) 阈值的技术性干扰影响的阵列。

将 Cy3 和 Cy5 通道的特征强度导入 Partek Genomics Suite (Partek Inc)。计算共杂交测试和人参比 DNA 样品中人基因间对照探针的平均强度。由此确定了一个比例因子，以根据 Cy3 信号平均值对 Cy5 人参比 DNA 信号平均值进行归一化处理。然后将所有 PathoChip 探针的 Cy5 强度乘以比例因子，对染料性能的差异进行归一化处理。计算每个探针的 Cy3/Cy5 比值和 Cy3-Cy5 差值，分别作为双通道或单通道分析流程的输入。将编号平均值 (AccAvg) 定义为一个编号所有探针的平均 Cy3 或 Cy5 强度，将编号信号 (AccSig) 定义为 $\text{AccAvg}(\text{Cy3}) - \text{AccAvg}(\text{Cy5})$ 。

在 Partek 中实施的基于模型的平铺阵列分析 (MAT) 用于每个肿瘤样品探针信号 (Cy3 减 Cy5) 的滑动窗口分析。MAT 参数为 p 值截止限 0.99, 窗口 5000 bp, 最小阳性探针数 5, 舍弃值 0%。按照 MAT 得分 30-300、300-3000 以及 > 3000 对候选区域进行分类。使用 Partek 方差分析 (ANOVA) 工具, 将所有肿瘤样品作为测试条件的平行样品, 并将共杂交的人参比 DNA 平行样品作为对照条件, 进行具有多重检验校正的配对 t 检验。在编号水平上使用 AccAvg(Cy3) 与 AccAvg(Cy5) 进行比较, 在单个探针水平上使用 Cy3 与 Cy5 强度值进行比较。将显著性阈值设定为递增假阳性率 < 0.05, 倍数差异 > 2。此外, 通过计算所有肿瘤 AccSig 或探针信号的标准偏差, 并过滤任何高于总体平均值 2 倍及以上标准偏差的值, 还进行了编号和探针水平的异常值分析^[6]。

在一些数据分析中, R 程序也被用于归一化和数据分析^[7]。使用人探针的绿色和红色通道信号计算比例因子, 比例因子为人探针的绿色通道信号之和与红色通道信号之和的比值。然后使用这些比例因子获得所有其他探针的归一化信号。对于除人探针外的所有探针, 归一化信号为绿色通道信号与比例因子修正后红色通道信号的比值的 log₂ 转换值 (log₂ g-log₂ 比例因子 * r)。对于归一化信号, 通过比较癌症样品与对照 (非匹配和匹配对照), 应用 t 检验选择癌症样品中显示阳性信号的探针, 并选择相对于癌症样品而言, 非匹配或匹配对照中明显存在的探针。显著性截止限为 log₂ 倍数变化 > 0.5, 调整后 p 值 < 0.05。使用 Benjamini-Hochberg 程序对这些调整后 p 值进行多次校正, 在此调整后 p 值截止限下, 在对照样品中检测到的任何探针都认为存在显著性。将对照中标称 p 值 < 0.05 且未进行多次比较校正的高显著性探针与癌症样品中存在的显著性探针进行比较。根据癌症与对照样品中特征的检出率计算阳性率。

结果与讨论

微阵列芯片设计优化

整合后的宏基因组包含 5206 个编号, 涵盖 4200 多种病毒、细菌、真菌和寄生虫, 分布在 58 条人工染色体中, 总长度为 4.489 亿 bp。通过为 CGH 应用构建的安捷伦定制探针设计算法, 从该宏基因组中识别到大约 550 万个探针。预测其中超过 300 万个探针与人基因组序列的交叉杂交风险较低。

最初, 在 PathoChip v2a 微阵列芯片上合成这些探针中映射到所选病原体的独特靶标区域的探针, 在 PathoChip v2b 微阵列芯片上合成覆盖至少两种或更多种病毒之间保守区域的探针^[6]。PathoChip v2b 的一个增强功能是包含覆盖已知与人类癌症高度相关的病原体的 22 个编号的 2085 个探针。使用安捷伦参比人 DNA 进行的探索性分析表明, 针对人序列的探针的强度中值超过 750 个荧光单位 (RFU), PathoChip v2a 上非人类特异性探针的强度中值约为 17 RFU, PathoChip v2b 上非人类保守探针的强度中值约为 120 RFU^[1]。总体而言, 这些分析识别出了 6360 个荧光值 > 150 的探针, 它们可能与人 DNA 杂交。由于此类杂交会导致高背景, 因此在考虑 PathoChip v3 的设计时将这些探针剔除, 该设计结合了独特和保守探针。

有趣的是, 在安捷伦女性参比 DNA 中, 针对 EB 病毒 (EBV; 人疱疹病毒 4 型) 的探针杂交强度较高。制造商证实, 用于制备男性和女性 SureTag 人参比 DNA 的细胞系使用这种病毒进行了感染以产生细胞系。如果不通过女性参比 Cy5 通道中的 EBV 探针信号对该信号进行归一化, 则会生成 EBV 假阳性结果。然后将参比 DNA 换为来自 B 细胞的无病毒参比人 DNA。经过一系列严格的检测步骤, 该参比未显示人病毒信号。

针对 PathoChip 检测开发改进的标记工作流程

与仅处理基因组 DNA (gDNA) 的标准 CGH 工作流程不同, PathoChip 检测同时使用 gDNA 和总 RNA。许多市售 DNA 和 RNA 纯化试剂盒可以处理 FFPE 样品;但我们发现, QIAGEN AllPrep DNA/RNA FFPE 试剂盒可以从同一 FFPE 样品中更高效地提取 gDNA 和总 RNA。该试剂盒从真菌细胞以及革兰氏阴性菌和革兰氏阳性菌中成功提取了核酸,而这些是样品中最难提取的微生物。该试剂盒还从酿酒酵母、蜡样芽孢杆菌和大肠杆菌培养物中高效回收了 gDNA 和 RNA (数据未显示)。

另一个需要考虑的问题是,由于各种微生物可能出现低拷贝数的情况,因此必须对 DNA 和 RNA 进行扩增,使拷贝数增加到检测阈值之上。PathoChip 筛选技术利用 TransPlex WTA-2 扩增试剂盒执行扩增步骤,可以检测基因组拷贝数较低的微生物和病毒。PathoChip 微阵列芯片的这些功能使得可以快速、灵敏地筛选多个肿瘤样品中是否存在各种微生物。

在最初的测试中,将人腺病毒 5 型、JC 多瘤病毒或 BK 多瘤病毒 DNA 添加到 15 ng 人 DNA 背景中,绝对拷贝数范围为 10000 到 10 个病毒基因组。经 TransPlex 扩增后,使用 PathoChip 探针在所有拷贝数下检测到腺病毒 5 型,使用多瘤病毒探针在低至 100 个基因组拷贝中检测到高于背景的信号(图 2A)。将包含腺病毒 5 型的细胞系 DNA 与包含呼吸道合胞病毒的 RNA 混合,用 TransPlex WTA-2 进行同步扩增。针对这两种病毒的探针产生了较强的特异性检测信号。这表明 TransPlex WTA-2 在基因组 DNA 存在的情况下提供了稳定的逆转录,并且 gDNA 和 cDNA 靶标实现了共扩增^[6]。

总之,我们证明了可以使用 PathoChip 肿瘤提取程序(基于改进的 CGH 工作流程[图 2B])提取和检测来自真核和原核病原体的核酸。

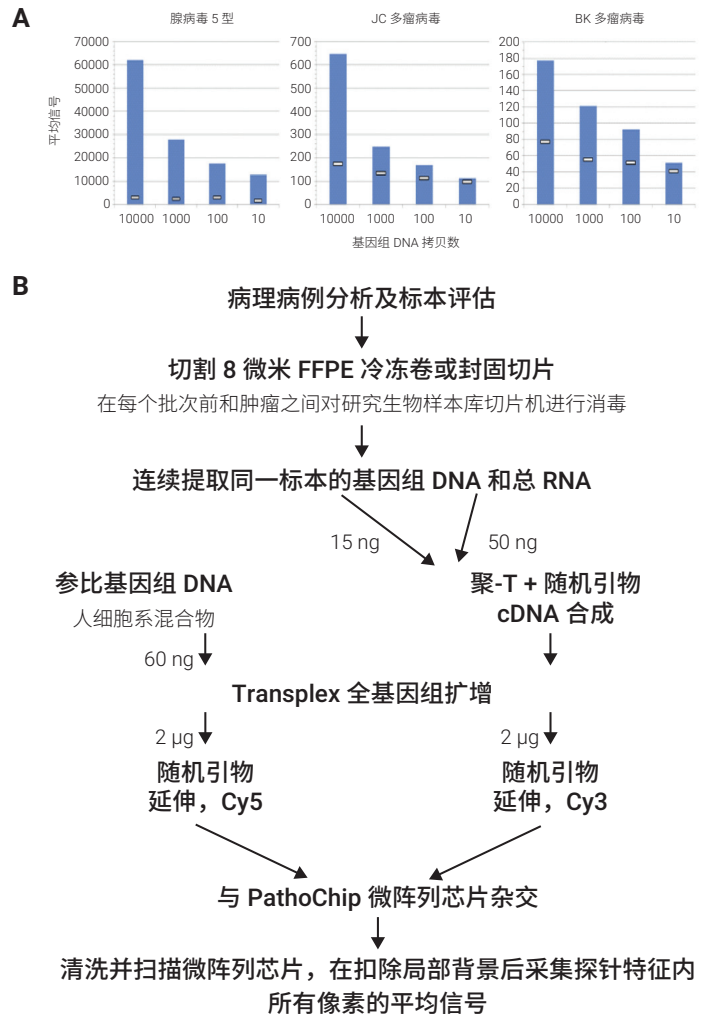


图 2. 用于检测对照 DNA 病原体的 PathoChip 工作流程。在每个样品 10000 到 10 个基因组拷贝的一系列稀释中测量了三种病毒的检测响应。将每种病毒的基因组 DNA 加入参考量的人 DNA 中。蓝色柱状图表示与测试样品杂交的指定病毒所有探针的平均 Cy3 信号, 白线表示对照样品探针的平均 Cy5 信号 (仅人 DNA)

使用 OSCC 样品开发用于 PathoChip 微阵列芯片的数据分析策略

致癌病毒在宿主肿瘤中可能发生明显的基因组重排或缺失。此外，由于病毒株具有广泛的多态性，新病原体的检测可能依赖于来自单个探针的信号。因此，需要通过多个级别的数据分析来检测筛选项目中预期的 3 种主要“匹配”类别（图 3）。编号信号 (AccSig) 是根据人 DNA 交叉杂交调整的编号的所有探针的平均值。计算 AccSig，从而根据编号组中的大部分探针进行一般病原体检测筛选。

通过计算探针滑动窗口的 MAT（基于模型的平铺阵列分析）分数来检测局部的高信号区域，而不考虑编号边界。在单个探针水平上采用具有多重检验校正的 t 检验来识别整个肿瘤群中信号始终高于背景的探针。对高信号探针进行异常值分析，但仅在筛选群体中的一种（或几种）肿瘤中进行。

使用来自包含 100 例 OSCC 肿瘤的筛选项目的数据评估这些分析方法。HPV16 的 AccSig 与 p16 病理报告^[6]一致（图 4），80% 的 p16(+) 肿瘤产生的 AccSig 值超过 100。在 8 例低或无 HPV16 AccSig 的 p16(+) 肿瘤中，有 4 例对部分 HPV16 探针显示出较高的信号，或产生明显的 HPV26 或 HPV92 AccSig 值。滑动窗口分析概述了 AccSig 结果，并强调了 HPV16 全基因组或部分基因组检测事件之间的差异^[6]。

在单个探针水平的分析也证明了在识别候选对象方面的实用性。大多数 HPV16 探针通过了 t 检验，在肿瘤群体中的检测显著性阈值大于背景（表 1）。对于 OSCC 中常见的基因组而言，这是符合预期的。许多 HPV16 探针也通过了异常值测试，这表明，尽管信号始终与背景不同，但群体的强度范围较宽（因此也包含异常值）。相比之下，t 检验结果显著的 HPV18 和 HPV26 探针较少，反映了这些基因组在该肿瘤群体中的表观发生率要低得多。

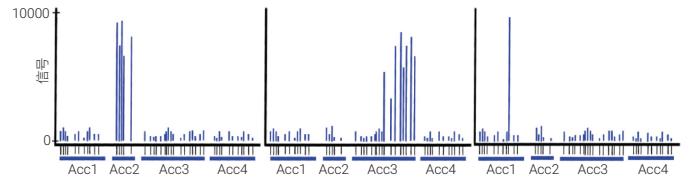


图 3. 模型数据说明了 PathoChip 的三种分析策略。在与 3 个假设的肿瘤样品杂交后，绘制 4 个基因组编号 (Acc) 的单个探针 (x 轴) 的信号。Acc2 的所有探针在肿瘤 1 中均显示出高信号 (左)，因此通过比较测试样品与对照样品的编号的所有探针的平均值，应该可以检测该候选对象。Acc3 探针的一部分在肿瘤 2 中显示出高信号 (中)，可能是由于菌株序列差异或基因组部分缺失，降低了所有探针的编号平均值，并增大了检测难度。在这种情况下，局部探针信号的滑动窗口分析不受编号注释的影响，并且可以实现更灵敏的候选对象鉴定。Acc1 的一个探针在肿瘤 3 中具有高信号 (右)，因此需要进行仅基于单个探针性能的第三级分析，以检测 PathoChip 非特异性靶向但与一个或少数几个探针具有序列同源性的生物体

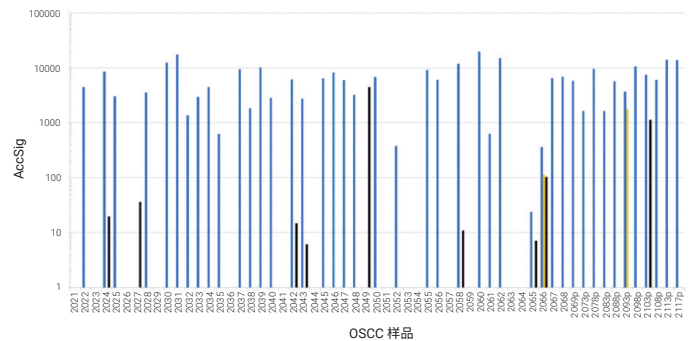


图 4. 肿瘤中 HPV 的编号平均值分析。(A) 通过单独分析 (2021–2068) 或混合分析 (2069p–2117p) 100 个口腔鳞状细胞癌 (OSCC) 样品的 PathoChip 结果计算 HPV16 (蓝色)、HPV18 (橙色) 和 HPV26 (黑色) 的编号信号 (AccSig)

表 1. 人乳头瘤病毒检测的单个探针分析

| 探针 | 探针数量 | | |
|---------|-------|-------|-------|
| | HPV16 | HPV18 | HPV26 |
| 探针总数 | 68 | 85 | 13 |
| 特异性探针 | 67 | 84 | 11 |
| 通过 t 检验 | 64 | 11 | 4 |
| 通过异常值测试 | 65 | 66 | 9 |
| 保守探针 | 1 | 1 | 2 |
| 通过 t 检验 | 1 | 1 | 0 |
| 通过异常值测试 | 1 | 0 | 2 |

然而，异常值分析很容易识别出相对较多的探针，这些探针在少数阳性样品中通过 AccSig 或 MAT 评分检测出 HPV18 或 HPV26。对于这些罕见的候选对象，一些探针的 t 检验结果显著，因为它们在整个群体中产生了较低但一致且超过背景的信号。这可能是存在的基因组拷贝数导致的，不足为奇。这也说明了在考虑后续验证的候选对象时，无论其初始识别使用的是何种方法，都需要检查探针水平的杂交强度，而不仅仅是分析算法输出分数。

使用 PathoChip 微阵列芯片检测与口咽和口腔鳞状细胞癌相关的不同微生物特征

在早期的数据分析中^[6]，为每个样品汇总了 MAT 得分大于 3000 的宏基因组区域。然后按探针信号图中的位置对每个区域内的各个探针进行排序。该分析检测到许多其他生物体，包括口腔致病菌，表明除 HPV 外，OSCC 中还可能涉及广泛的微生物。

我们利用 PathoChip 技术筛选了 100 份 FFPE 病理确诊的口腔鳞状细胞癌 (OSCC) 患者样品，以及 20 份癌旁正常对照 (匹配) 和 20 份来自健康个体的口腔组织 (悬雍垂) (非匹配对照)。该筛选用于分析与肿瘤组织相关的不同病毒和微生物特征^[7]。本研究分析的癌症样品取自舌头、舌根、扁桃体、口底和面颊，但以口咽为主。按照我们改进的工作流程，从样品中提取 DNA 和 RNA，进行全基因组和转录组扩增 (称为 WGTA)、标记，并与 PathoChip 上的探针杂交。

与 OSCC 相关的病毒特征

我们确定了与癌症和对照样品相关的 RNA 和 DNA 病毒 (图 5)。在筛选的 OSCC 样品中，乳头瘤病毒科的病毒序列显示出最高的杂交信号，然后依次为疱疹病毒科、痘病毒科、逆转录病毒科和多瘤病毒科的病毒序列 (图 5A)。在筛选的 100 份 OSCC 样品中，上述所有病毒科的病毒特征阳性率 > 75%。有趣的是，在 98% 的病例中检测到乳头瘤病毒 (图 5A)。

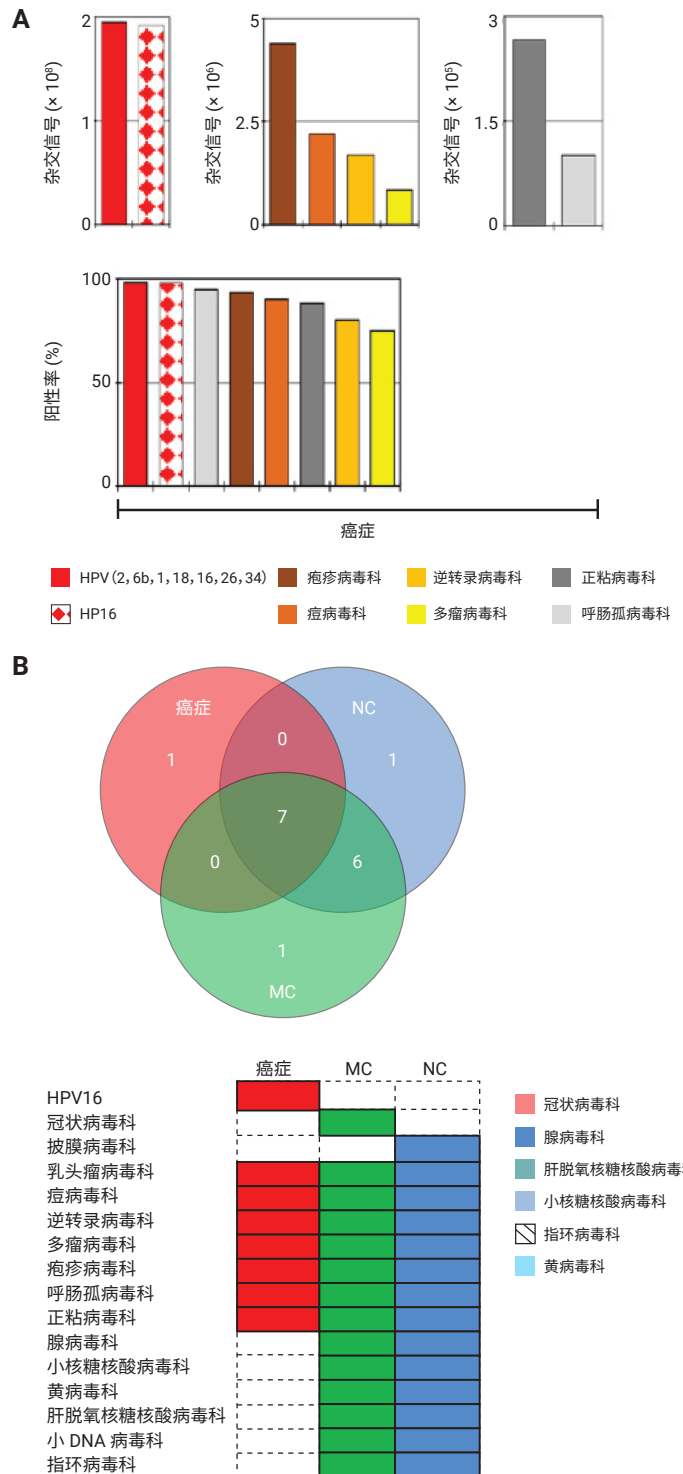


图 5. 在口腔癌和对照样品中检测到的病毒特征。(A) 显示了通过 PathoChip 筛选 100 份口腔癌样品，通过杂交信号 ($g-r > 30$) 检测到的病毒特征，并按杂交信号 (每个编号所有探针的加权得分总和) 和阳性率进行递减排序。(B) 病毒家族的不同分子特征与癌症和对照的相关性，以文氏图和彩色条表示

与匹配和非匹配对照相比，OCSCC 样品中所有乳头瘤病毒的杂交信号要高得多^[9]。重要的是，仅在 OCSCC 样品中检测到 HPV16 的高杂交信号和高阳性率 (98%) (图 5A、5B)。在 OCSCC 样品中检测到呼肠孤病毒科、疱疹病毒科、痘病毒科、正粘病毒科、逆转录病毒科和多瘤病毒科的特征，阳性率高，杂交信号比对照高 2-3 个 log (图 5A)。值得注意的是，在对照中特异性地检测到显著的冠状病毒科、小核糖核酸病毒科、腺病毒科、指环病毒科、肝脱氧核糖核酸病毒科和黄病毒科的病毒特征以及非 HPV16 型乳头瘤病毒科的特征^[7]。这些数据表明，与 OCSCC 组织相比，病毒特征发生了显著变化。

与 OCSCC 相关的细菌特征

图 6A 显示了在 OCSCC、匹配对照和非匹配对照样品中发现的各种细菌特征。文氏图对我们的发现进行了汇总，显示代表 13 个属的细菌特征与 OCSCC 样品相关，但与匹配或非匹配对照无关。其中包括变形菌门 11 个属，放线菌门和厚壁菌门各 1 个属^[7]。变形菌门短波单胞菌属和放线菌门动弯杆菌属的阳性率最高 (98%)：癌症病例中变形菌门埃希氏菌属和短波单胞菌属的探针阳性率分别为 88% 和 98%，具有高杂交信号。在 OCSCC 样品中检测到的放线菌门探针也具有高杂交信号，其中最高的是隐秘杆菌属。与病毒的情况一样，与正常的特征相比，细菌微生物特征在 OCSCC 中显示出显著的差异，并且更加稳定。在匹配或非匹配对照样品中，检测到放线菌属、动弯杆菌属和分枝杆菌属的细菌特征。需要重点注意的是，在对照样品中检测到的大多数细菌特征都属于正常的口腔菌群。

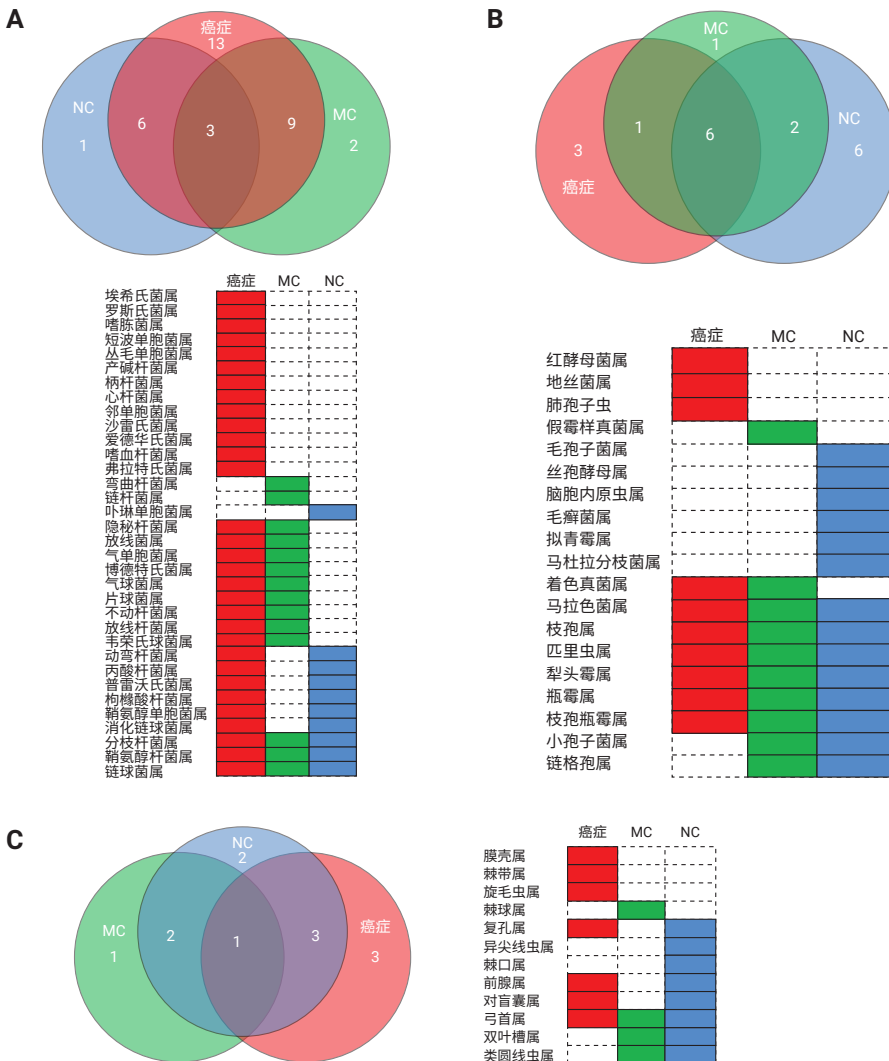


图 6. 在口腔癌样品中检测到的细菌、真菌和寄生虫特征。图 A、B 和 C 显示了不同细菌、真菌和寄生虫属的分子特征与口腔癌和/或对照的相关性，以文氏图和彩色条表示

与 OCSCC 相关的真菌和寄生虫特征

文氏图显示了 OCSCC、匹配和非匹配对照之间共有和特有的真菌特征。值得注意的是与 OCSCC 特异性相关的 3 种真菌特征：红酵母菌属、地丝菌属和肺孢子菌（图 6B）。在筛选的所有 OCSCC 样品中均检测到着色真菌属、马拉色菌属、匹里虫属、红酵母菌属、枝孢瓶霉属和枝孢属的分子特征。在 93% 的癌症样品中检测到肺孢子虫，在 > 75% 的筛选癌症病例中检测到地丝菌属、瓶霉属、犁头霉属和普雷沃氏菌属的特征^[7]。我们注意到，与口腔对照样品相比，OCSCC 的真菌生物群落发生了显著变化。同时，图 6C 中的文氏图汇总了与癌症和对照样品相关的寄生虫特征结果。结果表明，膜壳属、棘带属和旋毛虫属的分子特征仅与 OCSCC 相关。棘球属的特征仅与匹配对照样品相关，异尖线虫属和棘口属的特征仅与非匹配对照样品相关。因此，通过不同的特征可以区分癌症、匹配对照和非匹配对照。

使用 PathoChip 检测与三阴性乳腺癌 (TNBC) 相关的不同微生物特征

乳腺癌是最常见的癌症之一，根据是否存在某些激素和生长受体进行分类。最具侵袭性的乳腺癌是三阴性乳腺癌（缺乏雌激素、孕激素和 HER2 受体），因为不能通过内分泌治疗。遗传、环境和生活方式与乳腺癌的进展有关，但一些关于乳腺癌的研究表明，乳腺癌与疱疹病毒、多瘤病毒、乳头瘤病毒和逆转录病毒相关^[9]。

使用 PathoChip 对 100 份 TNBC 样品，以及 17 份匹配对照和 20 份非匹配对照进行了筛选。所有样品均来自 FFPE 档案样品（见方法部分）。在筛选的 100 份 TNBC 样品中，40 份进行单独筛选，60 份以每次反应 5 个样品（每种 RNA/DNA 10 ng）的形式进行混合筛选。筛选 100 份 TNBC 样品总共使用了 52 个微阵列芯片。将 17 个匹配对照和 20 个非匹配对照进行样品混合，每组使用 4 个微阵列芯片

当 PathoChip 筛选检测到特定探针的 Cy3 (g) 信号高于 Cy5 (r) 信号时，认为该探针为阳性。当癌症样品中特定生物体的可检测杂交信号 ($g-r > 30$) 明显高于匹配或非匹配对照样品时，则认为该特定生物体的探针与癌症样品相关。此外，采用多种检测方法对阳性探针进行了仔细评估。这些方法（表 2）包括编号异常值、编号 t 检验、特异性探针异常值、特异性探针 t 检验、保守探针异常值、保守探针 t 检验和基于模型的平铺阵列分析。

出于报告的目的，我们列出了通过 PathoChip 上的特异性探针检测到的特异性病毒和微生物的名称。但我们注意到，通过特异性探针检测确定的可能是密切相关的家族成员，而不是特异性生物体。这在 TNBC 样品显示一系列杂交信号，或者在针对特定病毒或微生物的探针组中某些探针没有杂交信号的情况下尤为相关。这也可能意味着，在特定的肿瘤中，这些微生物的基因组区域缺失，或者某个菌株出现了变异。

在保守探针中，检测到疱疹病毒科、逆转录病毒科、副痘病毒科、多瘤病毒科和乳头瘤病毒科的病毒特征^[8]。例如，对于疱疹病毒科，人巨细胞病毒 (HCMV)、人疱疹病毒 1 型 (HHV1；单纯疱疹病毒 1 型)、卡波西肉瘤疱疹病毒 (KSHV)、EB 病毒或人疱疹病毒 4 型 (EBV/HHV4) 的探针分别在 92%、65%、96% 和 78% 的乳腺癌样品中检出。对于乳头瘤病毒科，特异性探针分别在 78.8%、75%、84.6% 和 78.8% 的乳腺癌样品中检测到人乳头瘤病毒 (HPV) 6b、HPV18、HPV2 和 HPV16。特异性探针还分别在 82.7%、90.4% 和 86.5% 的癌症样品中检测到肝炎 GB、丙型肝炎和乙型肝炎的信号。有趣的是，并不是上述病毒的所有特异性探针都被检测到^[8]。有几种可能的原因，包括存在探针区域具有相同序列的相似生物体、生物体的片段或生物体 DNA 的整合片段。

将病毒探针按阳性率进行排序（不考虑杂交强度）时，显示为肝脱氧核糖核酸病毒和虫媒病毒的特征 (86.5%)，然后依次为副痘病毒 (83.3%)、疱疹病毒 (83.2%)、逆转录病毒 (79.6%) 和乳头瘤病毒 (79.3%) 的特征。然而，当按杂交信号进行递减排序时，疱疹病毒探针在肿瘤中的杂交信号最高，然后依次为具有高杂交信号的副痘病毒、虫媒病毒、多瘤病毒、逆转录病毒、肝脱氧核糖核酸病毒和乳头瘤病毒的探针（表 2）。细菌、真菌和寄生虫也获得了相似的数据^[8]。

表 2. 在 100 份三阴性乳腺癌样品中检测到的杂交信号（以每个编号所有探针的杂交信号之和计算）和病毒探针的阳性率。还给出了检测到候选对象的方法；AO：编号异常值，AT：编号 t 检验，SO：特异性探针异常值，ST：特异性探针 t 检验，CO：保守探针异常值，CT：保守探针 t 检验，MAT：基于模型的平铺阵列分析

| (a) 相关病毒 | 检测方法 | 检出率 (%) | 探针总和/编号 |
|-----------------|-----------------|---------|----------|
| 人疱疹病毒 5 型/HCMV | AO、SO、ST、CO、MAT | 92 | 14332000 |
| 人疱疹病毒 8 型/KSHV | AO、SO、MAT | 96 | 12119800 |
| 猴病毒 40 | AO、SO、MAT | 75 | 8113970 |
| 丙型肝炎病毒基因 1 型 | SO、CO、MAT | 90 | 7199330 |
| 人嗜 T 淋巴细胞病毒 2 型 | AO、SO、CO、MAT | 88 | 7040500 |
| 羊口疮病毒 | CO、MAT | 75 | 6422460 |
| 伪牛痘病毒 | AO、SO、CO、MAT | 90 | 5037880 |
| 人疱疹病毒 4 型/EBV | AO、SO、CO、MAT | 79 | 5024970 |
| 牛丘疹性口炎病毒 | AO、SO、CO、MAT | 85 | 4214040 |
| 秋葵花叶病毒 | AO、SO、CO、MAT | 75 | 3435060 |
| 人乳头瘤病毒 2 型 | SO、MAT | 85 | 3361460 |
| 人嗜 T 淋巴细胞病毒 1 型 | AO、SO、CO、MAT | 83 | 2745990 |
| 乙型肝炎病毒 | AO、SO、CO、MAT | 87 | 2621640 |
| 人疱疹病毒 1 型 | AO、SO、CO、MAT | 65 | 2319570 |
| 人乳头瘤病毒 16 型 | SO、MAT | 79 | 1651350 |
| 莫洛尼鼠类白血病病毒 | SO、CO、MAT | 58 | 1587600 |
| 默克尔细胞多瘤病毒 | AO、SO、MAT | 90 | 1551830 |
| 小鼠乳腺瘤病毒 | AO、SO、MAT | 79 | 1464980 |
| 人乳头瘤病毒 6b 型 | AO、SO、MAT | 79 | 1271950 |
| 人乳头瘤病毒 18 型 | SO、CT、MAT | 75 | 1184610 |
| JC 多瘤病毒 | AO、CO、SO、MAT | 77 | 755288 |
| 肝炎 GB 病毒 A | SO、MAT | 83 | 749098 |
| 藤浪肉瘤病毒 | SO、CO、MAT | 90 | 691071 |

表 3 和表 4 汇总了在癌症样品中检测到的病毒保守和特异性探针以及细菌、真菌和寄生虫探针^[8]。与分析的非匹配和匹配对照样品相比，在三阴性乳腺癌样品中检测到的病毒、细菌、真菌和寄生虫特征与癌症样品显著相关 ($p < 0.05$)^[8]。

表 3. 使用 PathoChip 筛选三阴性乳腺癌样品检测到的病毒探针特征的数量

| | 逆转录病毒科 | | | | 多瘤病毒科 | | | | 疱疹病毒科 | | | |
|------|--------|------|-------|-------|-------|------|----|------|-------|-----|------|------|
| | MMTV | MMLV | HTLV1 | HTLV2 | FSV | SV40 | JC | MCPV | HCMV | EBV | KSHV | HHV1 |
| 探针总数 | 31 | 24 | 41 | 86 | 8 | 41 | 42 | 62 | 299 | 235 | 259 | 22 |
| 特异性 | 31 | 15 | 37 | 84 | 5 | 41 | 40 | 62 | 275 | 149 | 256 | 15 |
| 异常值 | 1 | 4 | 24 | 43 | 4 | 25 | 12 | 27 | 139 | 67 | 132 | 7 |
| t 检验 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 保守 | 0 | 9 | 4 | 2 | 3 | 0 | 2 | 0 | 24 | 86 | 3 | 7 |
| 异常值 | 0 | 2 | 3 | 2 | 2 | 0 | 1 | 0 | 15 | 2 | 3 | 3 |
| t 检验 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | 乳头瘤病毒科 | | | | 肝脱氧核糖核酸病毒科 | 黄病毒科 | | 痘病毒科 | | |
|------|--------|-------|-------|------|------------|-------|------------|------|-----|-----|
| | HPV16 | HPV18 | HPV6b | HPV2 | HBV | HCV-1 | HepGB 病毒 A | BPSV | PCP | ORF |
| 探针总数 | 68 | 85 | 91 | 92 | 49 | 121 | 14 | 109 | 105 | 111 |
| 特异性 | 67 | 84 | 90 | 92 | 47 | 119 | 14 | 12 | 12 | 13 |
| 异常值 | 19 | 28 | 37 | 49 | 25 | 72 | 7 | 1 | 3 | 1 |
| t 检验 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 保守 | 1 | 1 | 1 | 0 | 2 | 2 | 0 | 97 | 93 | 98 |
| 异常值 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 74 | 80 | 76 |
| t 检验 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

表 4. 使用 PathoChip 筛选三阴性乳腺癌样品检测到的细菌、真菌和寄生虫探针特征的数量

| 微生物特征 | Chip 中的探针总数 | 检测到的探针总数 | 三阴性乳腺癌肿瘤中的检测 | 微生物类型 |
|----------|-------------|----------|--------------|-------|
| 成员 | 特异性 | 特异性 | 阳性百分比 | 生物体 |
| 隐秘杆菌属 | 4 | 4 | 75 | 细菌 |
| 短波单胞菌属 | 3 | 3 | 73 | 细菌 |
| 鞘脂杆菌 | 5 | 5 | 67 | 细菌 |
| 普罗威登斯菌属 | 1 | 1 | 67 | 细菌 |
| 普雷沃氏菌属 | 2 | 2 | 67 | 细菌 |
| 布鲁氏菌属 | 10 | 10 | 65 | 细菌 |
| 埃希氏菌属 | 13 | 10 | 64 | 细菌 |
| 放线菌属 | 4 | 4 | 52 | 细菌 |
| 动弯杆菌属 | 4 | 4 | 50 | 细菌 |
| 丙酸杆菌属 | 2 | 2 | 50 | 细菌 |
| 芽孢杆菌属 | 2 | 1 | 44 | 细菌 |
| 罗斯氏菌属 | 3 | 3 | 40 | 细菌 |
| 嗜酸菌属 | 2 | 2 | 39 | 细菌 |
| 二氧化碳嗜纤维菌 | 1 | 1 | 37 | 细菌 |

| 微生物特征 | Chip 中的探针总数 | 检测到的探针总数 | 三阴性乳腺癌肿瘤中的检测 | 微生物类型 |
|--------|-------------|----------|--------------|-------|
| 成员 | 特异性 | 特异性 | 阳性百分比 | 生物体 |
| 匹里虫属 | 8 | 8 | 98 | 真菌 |
| 毛孢子菌属 | 6 | 6 | 90 | 真菌 |
| 产色芽生菌属 | 3 | 3 | 89 | 真菌 |
| 瓶霉属 | 4 | 4 | 87 | 真菌 |
| 拟青霉属 | 4 | 4 | 69 | 真菌 |

| 微生物特征 | Chip 中的探针总数 | 检测到的探针总数 | 三阴性乳腺癌肿瘤中的检测 | 微生物类型 |
|--------|-------------|----------|--------------|-------|
| 成员 | 特异性 | 特异性 | 阳性百分比 | 生物体 |
| 鞭虫属 | 7 | 7 | 96 | 寄生虫 |
| 弓首属 | 1 | 1 | 62 | 寄生虫 |
| 利什曼原虫属 | 6 | 5 | 60 | 寄生虫 |
| 巴倍虫属 | 2 | 2 | 56 | 寄生虫 |
| 吸吮线虫属 | 1 | 1 | 40 | 寄生虫 |
| 并殖吸虫属 | 3 | 2 | 15 | 寄生虫 |

结论

在存在较高的人类基因组序列背景下，高度多重宏基因组分析方法检测小型非人类基因组的能力取决于几个因素，包括核酸提取与回收、靶标大小与拷贝数、扩增效率（如果使用 WTA）及特异性探针性能。在 PathoChip 开发过程中进行了一些修改，以提高检测灵敏度。这些修改包括每个编号包含多个探针，并整合来自不同级别数据分析的候选对象，从而优化筛选项目中的病原体检测。PathoChip 结合了饱和探针组以及 RNA 和 DNA 检测，改善了对已知致癌病原体或其他微生物的筛选。因此，可以使用 PathoChip 检测来全面评估多种生物体共感染的频率，及其与致癌或其他致病事件的相关性。

PathoChip 筛选项目可以根据检测量级、以及基于层次聚类分析和整个样品群体中检出率的多种分析策略进行检测，并进行优先级排序，从而生成候选列表。将这些结果与病毒或致病微生物的注释（例如宿主范围、组织特异性或普通群体中的阳性率）相结合，有助于确定值得进一步关注的微生物。通过 PCR 或捕获探针杂交与二代测序可以进一步跟踪这些结果。这种方法很可能为特定癌症或疾病提供有价值的微生物特征，不同的微生物在其中发挥了不同程度的作用。

参考文献

1. de Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D, Plummer M. "Global burden of cancers attributable to infections in 2008: a review and synthetic analysis." *Lancet Oncol.* 13:607–615 (2012)
2. Brodie EL, Desantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL, Hazen TC, Richardson PM, Herman DJ, Tokunaga TK, Wan JM, Firestone MK. "Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation." *Appl. Environ. Microbiol.* 72:6288–6298, (2006)
3. Wong CW, Heng CL, Wan Yee L, Soh SW, Kartasasmita CB, Simoes EA, Hibberd ML, Sung WK, Miller LD. "Optimization and clinical validation of a pathogen detection microarray." *Genome Biol.* 8:R93, (2007)
4. Chen EC, Miller SA, DeRisi JL, Chiu CY. "Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens." *J. Vis. Exp.* 50:2536 (2011)
5. Tu Q, Yu H, He Z, Deng Y, Wu L, Van Nostard JD, Zhou A, Voodeckers J, Qin Y, Hemme CL, Shi Z, Xue K, Yaun T, Wang A, Zhou J. "GeoChip 4: a functional gene array-based high throughput environmental technology for microbial community analysis." *Mol. Ecol. Resour.* 14:914–928, (2014)
6. Baldwin, D. A., Feldman, M., Alwine, J. C. & Robertson, E. S. "Metagenomic assay for identification of microbial pathogens in tumor tissues." *MBio* 5. e01714–01714, (2014)
7. Banerjee S, Tian T, Wei Z, Peck KN, Shih N, Chalian AA, O'Malley BW, Weinstein GS, Feldman MD, Alwine J, Robertson ES. "Microbial Signatures Associated with Oropharyngeal and Oral Squamous Cell Carcinomas." *Sci Rep.* 7(1):4036 (2017)
8. Banerjee S, Wei Z, Tan F, Peck KN, Shih N, Feldman M, Rebbeck TR, Alwine JC, Robertson ES. "Distinct microbiological signatures associated with triple negative breast cancer." *Sci Rep.* 5:15162, (2015)
9. Shiovitz S, Korde LA. "Genetics of breast cancer: a topic in evolution." *Ann Oncol.* 20, (2015)

查找当地的安捷伦客户中心：

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

在线询价：

www.agilent.com/chem/erfq-cn

www.agilent.com

仅供科研使用。不用于临床诊断用途。
PR7000-2665

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，2020
2020年12月，中国出版
5994-1968ZHCN

