arXiv:2611.29410v1 [cs.AR] 22 Nov 2026

Motivo：一种基于基序的算法-架构协同设计，用于快速、节能的基于图的基因组分析

Eirini Maria Tzermpou[1] Konstantina Koliogeorgi[1] Konstantinos Kanellopoulos[1] AtaberkOlgun[1]Nisa Bostancı[1] Dimitrios Soudris[2] Nika Mansouri Ghiasi[1] Onur Mutlu[1]1苏黎世联邦理工学院 2雅典国立技术大学

基因组序列分析通过研究生物体的基因组信息，在许多领域发挥着关键作用。读段比对是基因组序列分析的关键步骤，它将测序基因组的每个片段（称为读段）与已知参考基因组中的位置进行匹配。基因组图 是一种基于图的参考基因组，它以替代路径的形式表示一个种群的遗传多样性，从而提高读段比对准确率。然而，基因组图的规模和复杂性使序列到图（*S2G*）比对（即将读段比对到基因组图）成为基因组序列分析的主要瓶颈。我们发现，先前用于序列到图比对的软件和硬件方法会执行冗余的高成本操作（例如图遍历和动态规划），而这些操作并不会带来有效的比对。

本研究旨在通过减少冗余计算，并以轻量级操作替代开销高昂的计算，为序列到图映射提供一种高性能、节能的设计。我们的关键发现是，基因组图由一些小型子图构成，这些子图共享有限种反复出现的拓扑结构，我们称之为基序。基于这一发现，我们提出*Motivo*，一种面向序列到图映射的拓扑感知算法-架构协同设计。*Motivo* 的核心是一种新的基于基序的序列到图映射算法，它以内存查找和算术运算取代对每条候选路径进行的图遍历和动态规划。该算法通过两种方式利用基序。首先，它将每个节点的预计算模式存储在紧凑索引中，以一次内存查找取代图遍历。其次，它利用基序的拓扑结构及其节点属性（即序列长度和碱基组成），通过算术运算找出最合适的少数几条映射路径。*Motivo*的算法支持：*(i)* 在两种常规硬件平台（*Motivo-CPU* 和*Motivo-GPU*）上实现高性能执行；*(ii)* 在*Motivo-ASIC*上实现高性能、低能耗执行。*Motivo-ASIC*是一种我们与该算法协同设计的新型基于 *ASIC* 的加速器。由于该算法的路径选择仅使用内存查找和算术运算，*Motivo-ASIC*仅由简单的硬件单元构成。

*Motivo-CPU*相较于 *GraphAligner*、*vg* 和 *minigraph*，分别实现了 *2.1-18.4*×、*2.8-343*× 和 *1.4-10.1*× 的加速比；*Motivo-GPU* 相较于 *vg-giraffe* 实现了 *1.5-1.6*× 的加速比。*Motivo-ASIC* 相较于最先进的基于中央处理器的工具，在短读段（长读段）上实现了 *7.4-11.7*×（*138-274*×）的加速比。与两种最先进的专用集成电路加速器 *SeGraM*（短读段）和 *HARP*（长读段）相比，*Motivo-ASIC* 的每瓦吞吐量分别高出 *5.3*× 和 *4.0-4.3*×，每平方毫米吞吐量分别高出 *1.7*× 和 *2.7-3.0*×，并且吞吐量高出 *1.9-2.7*× 倍，高于 *SeGraM*。*Motivo* 在两种基因组图上对短读段和长读段的比对准确率与其他工具相当或更高（*96.8%-99.7%*）。

1. 引言

基因组序列分析通过研究生物体的基因组信息，在许多领域发挥着关键作用，例如个性化医疗 [1–28], 癌症研究 [29–75],疾病监测 [76–101], 进化生物学 [102–119],和农业 [120–135]。测序仪会产生许多生物体 DNA 的片段，称为读段 [136]。读段分为短读段 （最长可达数百个碱基）或长读段（数千至数百万个碱基）。为了跟上不断增长的测序数据量，许多既有研究加速了基因组序列分析（例如， [76, 132, 137–242]）。读段比对是基因组序列分析的关键步骤，它将每个读段与其在已知参考基因组 中的位置进行匹配（即代表某个物种的序列）。当参考基因组是一条单一序列时，读段比对属于序列到序列（*S2S*）映射 问题。

单个参考基因组无法代表一个物种各个个体之间的遗传变异，这会降低与参考基因组不同的读段的比对准确率 [243–245]。基因组图是一种基于图的参考基因组，它将群体的遗传多样性表示为替代路径，从而提高读段比对准确率 [244, 246–249]。如图 1 所示，基因组图中的每个节点都存储一个序列，每条边连接两个序列可以彼此相接的节点，而图中的每条路径都表示一种可能的基因组序列。变异，即两个基因组之间的差异，会在图中形成替代路径。基因组图为应用开辟了新的方向

例如群体规模 [244, 246, 250, 251] 和个性化基因组学 [1–28]。



图 1：由三个基因组（#1-#3）构建的基因组图。每个节点存储一段序列，每种颜色标示一个基因组经过图的路径通过图中的一个基因组。

将读段比对到基因组图是一个序列到图映射（*S2G* 比对）问题：S2G 比对工具会在基因组图中寻找一条路径，其序列与读段最相似 [252–255]。比对前，该工具会为基因组图构建图索引，以便快速查找图中的序列 [255, 256]。S2G 比对通常采用种子-

扩展策略，分为三个步骤：种子生成、过滤和扩展。首先，种子生成步骤从读段中提取较短的子序列（种子），并查询图索引，找出图中与每个种子匹配的位置，即种子命中 [255]。其次，过滤步骤会丢弃、聚类或链接种子命中，以减少扩展步骤的候选项数量（例如，[253, 257, 258]）。第三，扩展步骤将每个剩余的种子命中扩展为局部种子命中子图，并将读段与其中的每条路径进行比对，从而为每条路径生成一个候选比对。比对会逐碱基比较两个序列，并找出将一个序列转换为另一个序列所需的最少改动（即编辑操作）[259–261]。如果编辑操作的数量不超过阈值，则该比对有效。比对是 S2S 和 S2G 映射中的计算瓶颈，因为它依赖耗时的动态规划（DP），以找出读段与参考序列之间的编辑操作 [151, 152, 178, 262–264]。

尽管序列到图映射具有准确率优势，但它也面临两个关键挑战。首先，基因组图包含许多替代路径，因此产生的候选比对数量多于参考基因组的单一序列 [245, 252, 265]。候选比对越多，原本开销就很高的比对步骤成本也越高 [266–269]。其次，基因组图规模庞大，包含数亿个节点 [248, 258]。随着测序技术以更低的成本产生更多数据，基因组图所表示的基因组数量不断增加，图的规模也随之扩大。在更大的图中，一个读段会有更多种子命中，每个种子命中又对应许多替代路径，这进一步增加了序列到图映射的成本 [255, 268]。

为缓解这些挑战，软件工具会优化序列到图映射的各个步骤，例如索引构建 [270–277]、种子生成 [278–282]、过滤 [245, 253, 257, 258, 283–287] 和扩展 [259, 260, 266, 267, 288–291]。其他研究则借助图形处理器（GPU）[218, 292] 或基于专用集成电路（ASIC）的硬件加速器 [210, 211] 来加速序列到图映射。

关键局限。 尽管软件工具和硬件加速器都能为 S2G 比对带来优势，但它们都会执行冗余操作，即对无法产生有效比对的候选项执行代价高昂的操作（例如图遍历和动态规划）。我们对最先进的 S2G 比对工具（§3）的分析表明，仅对被丢弃候选项执行的操作就占

短读段端到端映射时间的 31–99.3%，以及长读段端到端映射时间的 83–99.9%。现有过滤步骤无法消除这些冗余操作。

我们的目标是通过减少冗余计算，并以轻量级操作替代开销较大的计算，为序列到图映射提供一种高性能、低能耗的设计。我们的**关键观察**是，基因组图由许多小型子图组成，这些子图共享有限种反复出现的拓扑结构，我们将其称为基序。大多数遗传变异形成小跳跃，即一条短分支，在几个节点内离开图的其余部分并重新汇入其中（§3），因此相同的小型拓扑结构会在基因组图中反复出现（图 2）。

基于这一观察，我们提出了 Motivo，这是一种面向拓扑结构的算法-架构协同设计，用于序列到图（S2G）比对。Motivo 的核心是一种新型的基于基序的序列到图映射算法，它用内存查找和算术运算（例如加法和比较）取代了对每条候选路径进行的图遍历和动态规划。该算法的**关键思想**是利用基序在比对之前识别出种子命中周围少数可能产生有效比对的路径。该算法仅将读段与这些路径进行比对。Motivo的算法支持：（*i*）在两种常规硬件平台（即中央处理器和图形处理器）上实现高性能运行（Motivo-CPU和Motivo-GPU）；（*ii*）在Motivo-ASIC上实现高性能、低能耗运行。Motivo-ASIC 是一种新型的基于 ASIC 的加速器，我们与该算法协同设计了这一加速器。

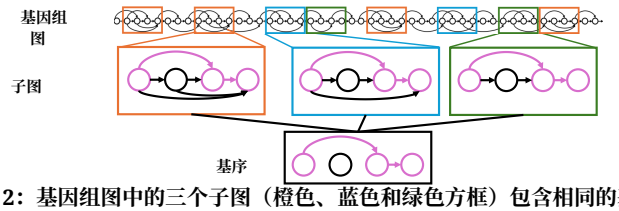


图 2：基因组图中的三个子图（橙色、蓝色和绿色方框）包含相同的基序（粉色）：一条经过连续四个节点、跳过第二个节点的路径。

关键机制。Motivo包含三个关键组件（§4）。首先，MotivoIndex是一种紧凑的图索引，为基因组图中的每个节点存储其预计算模式及其属性（即序列长度和碱基组成）。其次，MotivoExt是一种新的扩展机制，通过内存查找和算术运算，为每个种子命中找到最合适的比对路径。MotivoExt从MotivoIndex中获取种子命中的基序，丢弃无法导向有效比对的候选路径，并为其余路径排序，仅保留排名前-K 的路径。因此，Motivo将序列到图映射简化为每个种子命中只需进行少量 S2S 比对。第三，我们在三种硬件平台上实现了Motivo算法。Motivo-CPU和Motivo-GPU 分别是在中央处理器和图形处理器上对该算法的直接软件实现。它们的性能优于高度优化的最先进的、基于中央处理器和图形处理器的工具，这表明Motivo 的优势源自算法本身。Motivo-ASIC 则在硬件中端到端地执行该算法。由于该算法的路径选择仅使用内存查找和算术运算，Motivo-ASIC由简单的硬件单元构成，能够以较小的面积和功耗提供高吞吐量。

关键结果。我们使用短读段和长读段，在两类基因组图上评估 Motivo：以往工作采用的 Genome in a Bottle (GIAB) 图 [210]，以及更稠密的人类泛基因组参考联盟 (HPRC) 图 [248]。我们将 Motivo 与最先进的基于中央处理器的系统（GraphAligner [253]、vg [245]、minigraph [258]）、基于图形处理器的系统（vg-giraffe [293]）和基于专用集成电路的系统（SeGraM [210]、HARP [211]）进行比较。我们得出四项主要发现。第一，仅Motivo的算法就能提升性能。与 GraphAligner、vg 和 minigraph 相比，Motivo-CPU的加速比分别为 2.1–18.4×、2.8–343× 和 1.4–10.1×；与 vg-giraffe 相比，Motivo-GPU的加速比为 1.5–1.6×。第二，对于短读段，Motivo-ASIC 相比最先进的基于中央处理器和图形处理器的基线，分别实现了 7.4–11.7× 和 7.7–8.2× 的加速比，

对于长读段，相比最先进的基于中央处理器的基线，加速比为138–274× 。采用相同算法时，Motivo-ASIC 处理短读段（长读段）的速度比Motivo-CPU 快 4.2–6.9× （68–113×），比Motivo-GPU快 5.2–5.3× 。第三，Motivo-ASIC 优于两款最先进的基于专用集成电路的加速器。对于 GIAB 图上的短读段，Motivo-ASIC 的吞吐量高出 1.9–2.7× 倍，每瓦吞吐量高出 5.3× 倍，每平方毫米吞吐量高出 1.7× 倍。对于 HPRC 图上的长读段，Motivo-ASIC的每瓦吞吐量比 HARP 高出 4.0–4.3×倍，每平方毫米吞吐量高出 2.7–3.0× 倍。第四，在 GIAB 图上，Motivo 对短读段和长读段的比对准确率分别达到 99.1% 和 99.7%；在 HPRC 图上则分别达到 98.0% 和 96.8%，甚至高于部分最先进的基线。

本研究的主要贡献如下：

•我们提出了**Motivo**。据我们所知，Motivo 是首个利用重复出现的子图拓扑进行序列到图（S2G）比对的算法-架构协同设计。Motivo 利用基序在比对之前 识别基因组图中少数可能产生读段有效比对的路径。

• 我们介绍Motivo的两种关键机制：(*i*) MotivoIndex，一种基于基序的图索引，用于存储基序及基因组图中每个节点的属性；以及 (*ii*) MotivoExt，一种利用MotivoIndex 通过轻量级算术运算，为读段识别出最合适的少数比对路径，而非对每条候选路径执行动态规划的机制。

• 我们展示了Motivo 算法能够提升中央处理器和图形处理器上的序列到图比对性能，并将该算法与Motivo-ASIC协同设计；Motivo-ASIC是一种基于专用集成电路（ASIC）的全新硬件加速器。

• 我们表明，对于短读长和长读长，Motivo的吞吐量和能效均高于最先进的基于中央处理器、图形处理器和专用集成电路的序列到图映射系统。与两种最先进的基于专用集成电路的加速器相比，Motivo-ASIC 的吞吐量最高可提升 2.7× 倍，每瓦吞吐量最高可提升 5.3× 倍，每平方毫米吞吐量最高可提升 3.0× 倍2，比对准确率为 96.8%–99.7%。

2. 背景

2.1. 基因组序列分析

基因组样本经过三个主要步骤处理：(*i*) 测序、(*ii*) 碱基识别和 (*iii*) 分析。第一步测序会将样本的基因组信息（例如，DNA 或 RNA 中编码的信息；通常会在测序前将 RNA 转换为 DNA [294, 295]）转换为数字信号。测序仪无法提供完整 DNA 分子的端到端序列，而是将其拆分为许多相互重叠的短片段，称为读段。随后，碱基识别会将每个读段的原始信号转换为由四种核苷酸碱基 A、C、G、T 组成的字符串 [296]。读段长度（即读段所含的碱基数）以碱基对 (bp) 为单位。基因组分析 会处理由此得到的字符串集合，即读段集合，并将其与参考基因组（即代表某个物种的已知序列）或包含多个物种的参考基因组及先前测序样本的大型数据库进行比对。基因组分析的一项重要步骤是读段比对：将测序样本中的每个读段与该生物所属物种的参考基因组进行匹配，以确定读段的原始位置以及二者之间的差异 [261, 286]。比对 通过逐碱基比较两条序列来识别这些差异，并计算它们的编辑距离，即将一条序列转换为另一条序列所需的最少碱基替换、插入和缺失操作数 [297]。

2.2. 基因组图表示

基因组的线性参考表示，即基于单个或少数几个个体的基因组表示，无法体现群体遗传多样性 [243, 298]。个体基因组与参考基因组之间存在多种类型的变异。最常见的是单核苷酸多态性（SNP）：在单个位点以一个碱基替换另一个碱基（例如，参考基因组在该位点为 G，而实际为 A）。插入和缺失（插入缺失）会增加或移除一个或多个碱基，而结构变异（例如，大型插入缺失、重复和倒位）则涉及较长的序列片段 [254]。由于线性参考基因组在每个位点只能表示群体中众多序列变异之一，研究人员已转向采用图表示 [210, 251, 299–305]，将多个基因组的信息整合起来，并准确捕捉群体的遗传多样性。在基因组图中，节点对应基因组序列，边表示序列之间的邻接关系，而个体路径对应单个基因组 [245, 258]。在已提出的图模型中，最常用的类型之一是**变异图**，它将所有变异信息纳入单一结构中。

不同基因组之间共享的序列会合并到共同节点中，而变异则表现为分叉结构：单核苷酸多态性会形成一条短分支，替代的单碱基路径在一个节点后重新汇合；插入缺失和结构变异则会形成更长的路径，这些路径从主序列分出，之后再重新汇合 [254]。这样既能紧凑地编码共享骨架，也能表达复杂的个体特异性变异 [277]。

2.3. S2G比对工作流程概述

在传统的线性参考比对（S2S）中，目标是在参考字符串中找到最合适的位置，使其与查询序列匹配，同时尽可能减小编辑距离 [286, 297]。相比之下，序列到图（S2G）比对将这一问题推广到基于图的参考，目标转变为在图中找到最可能的比对路径 [245, 252, 253]。这要求同时优化路径选择和比对，从而大幅增加复杂度，但能在高变异区域实现更准确的比对 [257]。大多数S2G比对算法包括三个在线步骤：种子生成、聚类/过滤和扩展

扩展 [245, 253, 257, 258, 265, 280, 284, 287, 289],，同时，它们对所有查询使用离线预构建参考图索引。

离线参考图索引构建。 为了在比对过程中快速查找短匹配，比对前（即离线）会针对图的序列构建索引。常见的设计包括哈希表、压缩全文索引（如基于 FM/BWT 的结构） [306], 以及轻量级草图索引。草图化会根据固定规则，从序列的*k-mer*（即长度为 k 的子序列）中选取一小部分，因此只存储选出的 k-mer，而非全部 k-mer [256, 274, 307]。一种广泛使用且节省内存的草图化规则会生成最小化器：给定 k-mer 长度 k 和窗口大小 w，窗口方案沿序列滑动，并在每个由连续 k-mer 构成的窗口中选出哈希值最小的 k-mer 作为最小化器 [307]。索引会存储每个最小化器的哈希值及其在图中的位置（即节点 ID 和该节点序列中的起始偏移量）。**种子生成**。 序列到图映射的第一阶段（即种子生成）会在读段与图的节点序列之间查找短的精确或近似匹配 [255]。按照索引构建时使用的相同草图化规则（例如最小化器），从读段中提取短子序列，然后在索引中查询这些子序列，以获取它们在图中的位置 [245, 252, 253, 256, 258, 281, 286, 307–313]。

聚类/过滤/链式连接。 单条读段可能在图中产生许多种子命中，因为短子序列会出现在许多位置，但其中只有少数能够产生有效比对。此步骤会减少比对器需要检查的种子命中数量 [255]。过滤 会丢弃不太可能产生有效比对的种子命中，例如在图中孤立的种子命中。聚类 会将图中彼此靠近的种子命中归为一组，因为它们很可能指向参考基因组的同一区域，从而使比对器能够将它们作为一个候选项进行检查。链式连接 会连接在读段和图中顺序一致、距离相容的种子命中，并使用动态规划为每条链评分，从而只扩展得分最高的链 [280, 287]。**扩展**。 最后一步会将未丢弃的种子命中扩展为候选种子命中子图，收集足够多的相邻节点以覆盖读段，然后在读段与子图路径之间执行碱基级比对，以产生最终比对 [255, 297]。比对分为两个阶段。首先，距离计算 （DC）计算读段与候选参考路径之间的编辑距离，即能够解释二者差异的最少编辑操作数。其次，回溯 （TB）恢复实际比对，也就是读段相对于候选路径的各个匹配、替换、插入和缺失的序列及位置，从而达到这一最少编辑操作数 [181, 314]。如果 DC 计算出的编辑距离超过预先设定的编辑距离阈值，则该候选项无法产生有效比对，因此会在执行 TB 之前被拒绝。在序列到图映射中，此过程既确定实际比对，也确定种子命中子图中产生该比对的路径，即有效比对路径。我们将读段的每个比对称为读段与种子命中子图中一条路径的比对称为候选比对。编辑距离未超过阈值的候选项是有效比对；否则，就会被丢弃。读段的最佳有效比对就是其最终比对，而包含该比对的子图则是最终比对子图。由于图比对的计算开销很大，人们已提出多种优化算法和硬件设计，以加速端到端映射流程或其中的单个步骤（例如， [210–223, 253, 258, 265, 280, 315–322]）。

3. 动机分析

我们对最先进的S2G比对工具进行性能分析，以找出它们的主要低效之处。我们发现，(*i*) 它们大部分开销巨大的比对操作都针对无法产生有效比对的候选比对执行；(*ii*) 这些冗余操作占据了端到端映射时间的大部分。我们分析了最先进的软件比对工具 vg [245] 和 GraphAligner [253]，以及最先进的S2G映射加速器 SeGraM [210]（详细的性能分析与评估方法见 §6）。

3.1. 观察：冗余比对工作

图 3(a) 展示了短读段和长读段中每个读段的种子命中数、候选比对数、过滤后的候选比对数以及有效比对数。对于短读段，vg、GraphAligner 和 SeGraM 找到的种子命中数分别比有效比对数多出 38×、 66× 和 185× 。对于长读段，这一差距分别扩大到 3601×、 4038× 和 9407×。长读段会扩大差距，原因有二。首先，一个长读段包含

种子更多，因此产生的种子命中也多得多，但有效比对仍然只有一个或几个。其次，长读段跨越的变异位点多得多，因此其种子命中子图中包含的替代路径也多得多。链式连接和比对步骤需要检查这些路径，进一步导致可能存在大量冗余的高开销操作。

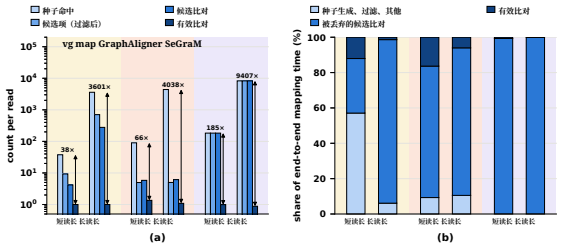


图3：每个读段的（a）种子命中数与有效比对数，以及（b）S2G工具端到端映射时间的细分。

每个通过过滤的种子命中（即图3中过滤后的候选项）都会触发高开销的基于动态规划的操作，例如链式连接和比对。因此，这些操作几乎全都耗费在最终被丢弃的候选项上。值得注意的是，vg 和 GraphAligner 已经对种子命中进行过滤、聚类和链式连接，但仍需检查比所需数量多出几个数量级的候选项。因此，现有的过滤启发式方法无法消除冗余比对。

对端到端映射时间的影响。 比对占据了这三种工具的大部分运行时间：它分别占 vg 和 GraphAligner 映射时间的 43–96% 和 89–91%，并占 SeGraM 全部映射时间。

如图3(b)所示，对最终被舍弃的候选项进行的操作，单独就占这些工具处理短读段端到端映射时间的 31–99.3%，以及处理长读段端到端映射时间的 83–99.9%。因此，在进行比对之前识别哪些候选项能够产生有效比对，可以大幅缩短端到端映射时间。

3.2. 机遇：反复出现的拓扑结构

基因组图中的大多数遗传变异由单核苷酸多态性和小型插入缺失构成 [248], , 它们会形成小跳跃，即较短的分支，并在几个节点内重新汇入共享骨架 [243, 245]。因此，许多子图具有相同的拓扑结构。这带来了一个机遇：如果少数几种拓扑结构就能描述图中的大部分内容，那么无需在运行时遍历图来寻找所有候选比对路径，也能确定种子命中周围的路径。图4显示，对于短读段，SeGraM 中 98.5% 的有效比对子图和 97.4% 的种子命中子图都少于 8 个节点。对于长读段，96.8% 的种子命中子图可以分解为由仅 1–2 个节点隔开的此类小型子图链。因此，最多包含 7 个节点的子图几乎涵盖了序列到图映射所处理的所有子图，或完全涵盖，或涵盖其中一部分。

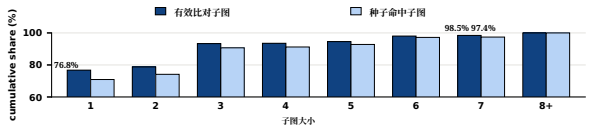


图4：种子命中子图和有效比对子图大小的累积分布百分比。

我们将这些反复出现的小型拓扑结构称为基序。一个基序跨越 n 个连续节点，从第一个节点开始，到最后一个节点结束，并指定它经过哪些中间节点。我们将每种基序（$3 \le n \le 7$）的出现频率定义为以其为起点的图节点所占的百分比。图5显示，基序会频繁重复出现。因此，只需少量基序便足以描述不同密度的基因组图的拓扑结构。

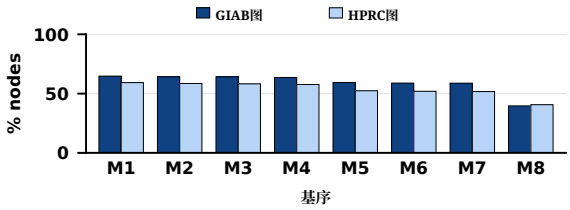


图5：GIAB 和 HPRC 图中最常见基序的出现频率（每种基序起始节点所占的百分比）。

我们的目标是通过消除冗余的比对操作，提高端到端序列到图映射的性能和能效。为此，我们利用基序在比对之前识别每个种子命中周围少数能够产生有效比对的路径。

4. Motivo：高层概述

我们提出Motivo，这是一种具备拓扑感知能力的序列到图（S2G）比对算法-架构协同设计。其核心思想是

[1]SeGraM流水线利用其比对单元进行种子生成，从而完全隐藏种子查找延迟。

Motivo 的目标是利用基序，在比对之前识别出种子命中附近少数几条可能产生有效比对的路径。A
基序是基因组图上 n 连续节点组成的路径（$3 \le n \le 7$，§3.2）。图 6 展示了 Motivo 的三个关键组件（一个索引、一个算法以及该算法的三种硬件实现）和 Motivo运行时所经历的六个步骤（ - ）。

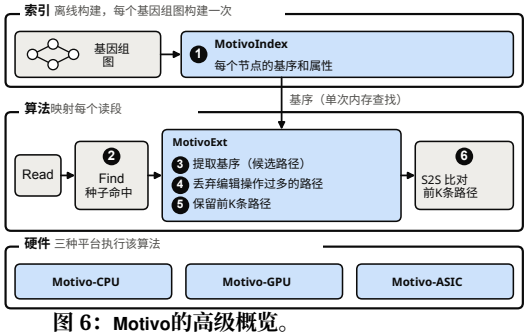


图 6：Motivo的高级概览。

构建索引。Motivo 根据基因组图一次性离线构建 MotivoIndex（ ，§5.1）。MotivoIndex 是一种紧凑的图索引，为基因组图中的每个节点存储在该节点开始或结束的预计算模式，以及节点的属性（即其序列长度和碱基组成，也就是其序列中 A、C、G 和 T 碱基的数量）。

比对读段。算法首先找到读段的种子命中（ ）。对于每个种子命中，MotivoExt 随后分三个步骤选择少量最合适的比对路径。首先，MotivoExt 通过一次内存查找（MotivoIndex， ）检索从种子命中的节点开始（向右扩展）或结束（向左扩展）的基序。每个基序都是种子命中附近的一条候选路径，因此 MotivoExt 无需遍历图即可获得候选路径。其次，MotivoExt丢弃无法导向有效比对的候选路径（ ）。MotivoExt利用节点的序列长度确定读段覆盖每条路径的部分，累加该部分节点预计算的碱基组成，并将结果与读段的碱基组成进行比较。四种碱基计数之差的总和，即*ACGT* 距离，给出了该路径编辑操作数量的下界，因为一次编辑操作至多使 ACGT 距离变化二。MotivoExt 会丢弃下界超过编辑距离阈值的路径，因此此步骤不会丢弃任何可能导向有效比对的路径。第三，MotivoExt保留排名前-K的路径，即下界最小的 K 条路径（ ；我们使用 $K = 2$）。此步骤是一种启发式方法；§7.3 量化了它对比对准确率的影响。最后，Motivo 仅构建排名前-K 路径的序列，并将读段分别与每条序列进行比对（ ）。因此，MotivoExt 将序列到图映射简化为每个种子命中的少量序列到序列（S2S）比对，而动态规划（DP）仅保留在 S2S 比对器中。

执行该算法。Motivo 的算法能够实现：（*i*）在两种传统硬件平台（Motivo-CPU 和Motivo-GPU）上高性能运行；以及（*ii*）在一种全新的算法

架构协同设计的基于 ASIC 的加速器（Motivo-ASIC，§5.3）。Motivo-CPU 和Motivo-GPU 分别是在中央处理器和图形处理器上对该算法进行的直接软件实现。它们的性能超过了经过高度优化、最先进的基于中央处理器和图形处理器的工具（§7.1.1），这表明Motivo 的优势源自算法本身。Motivo-ASIC 在硬件中端到端执行该算法。该算法通过两种方式实现Motivo-ASIC 的轻量化设计。首先，由于该算法的路径选择（步骤 - ）仅使用内存查找和算术运算，Motivo-ASIC 由简单的硬件单元构成，并组织为完全流水化的五级数据流。其次，MotivoIndex 将每个基序编码为位掩码，所有基序的集合都能容纳在一小块片上存储器中。因此，Motivo-ASIC 可实现每平方毫米^2 及每瓦特的高吞吐量（§7.1.3 和 §7.2）。

5. Motivo：详细设计

5.1. 数据结构

本节介绍我们新颖的基于基序的索引方案 MotivoIndex。MotivoIndex 的最终目标是在比对时快速检索种子命中周围的预计算路径，从而避免代价高昂的图遍历。为此，MotivoIndex 对每个节点编码其参与的基序，使得围绕该节点的所有候选路径都能通过一次查找获得。MotivoIndex由两部分组成：(*i*) 基于基序的索引，用于存储每个节点发起（终止）的基序；以及 (*ii*) 紧凑结构，用于存储节点属性（例如序列长度、碱基组成）。MotivoIndex通过将发起（或终止）同一组基序的节点分组，充分利用了基序重复出现的特性。这样可以仅存储节点所属组的标识符（ID），从而实现紧凑、资源高效且适合硬件的存储。

索引构建前：我们使用 Kahn 算法对图进行拓扑排序，这与最先进的 S2G 加速器和比对器类似 [210, 211, 323]。Motivo 支持各种规模和密度的图（例如 GIAB 和 HPRC图），且不要求图为无环图。对于包含环的泛基因组图（如 HPRC），我们采用 vg 生成的排序图中的节点顺序，并将少量反向边存储为特殊边。除 MotivoIndex 外，我们还构建了基于最小化器的索引（$k = 15$，$w = 20$），如第 §2 节所述，该索引在种子生成步骤中是必需的。

基于基元的索引。 MotivoIndex (*i*) 对基序拓扑结构进行编码作为基序位掩码，(*ii*)记录每个节点在基序映射中发起（终止）的基序，以及 (*iii*)将基序映射相同的节点分组，从而通过一次查找即可获取节点的基序。我们介绍这些结构，并以一个四节点图为例（图 7）。

模体位掩码。 大小为 n 的基序是从一个 n节点子图的起始节点到末尾节点的一条路径。根据我们的性能分析（§3），我们枚举了所有包含 $3 \le n \le 7$的基序。每个基序都编码为一个 n位基序位掩码 （图7），如果第 i个节点位于路径上，则设置位 i。位0和 $n-1$（起始节点和末尾节点）始终置位。总共有 $N = \sum_{i=3}^{n} 2^{i-2} = 2^{n-1} - 2$个基序。在图7中，对于大小为 $n = 4$（I）的图，其起始节点为节点 A 和终止节点 B，我们展示了它可能包含的所有基序 M_0–M_5 (II)。表格 (III) 显示了它们编码为 4 位位掩码的逐位表示。对于 $n = 7$，基序的数量为 62。

基序映射。 每个节点都有一个起始基序映射和一个终止基序映射，分别编码该节点发起的基序（起始基序）和终止的基序（终止基序）。基序映射 的长度等于基序的数量，即图 7 中为 6 个，完整的MotivoIndex中为 62 个。如果节点发起（终止）基序 M_j，则基序映射的第 j 位会被置位。图 7 展示了节点 A 和 B 的起始与终止基序映射（IV）。

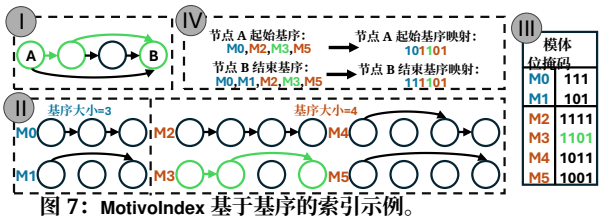


图 7：MotivoIndex 基于基序的索引示例。

基于模体的节点分组。 为图中的所有节点创建起始和终止基序映射后，我们利用基序的重复出现特性（见 §3.2）来降低内存容量需求。具体而言，如图 8 所示，在两种图类型中，少数几种基序映射即可描述大多数节点：仅出现频率最高的起始基序映射就涵盖了全部节点的 23–25%，而出现频率最高的 256 种则涵盖了 GIAB图（HPRC图）中 99.85%（98.1%）的节点。因此，我们可以根据基序映射对节点进行分组，并为每个节点存储其所属组的组 ID。我们也根据终止基序映射进行同样的分组。

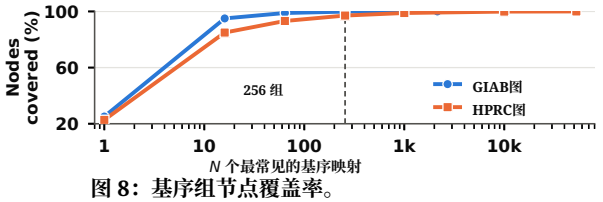


图 8：基序组节点覆盖率。

节点属性。 MotivoIndex 的这一部分存储节点属性，供 MotivoExt 和比对步骤使用。对于图中的每个节点 i，我们存储其实际序列、序列长度以及 ACGT 含量（即节点序列中每种核苷酸碱基的数量）。

特殊情况。 对于包含结构变异和环的复杂高阶图，有些节点不属于任何频繁组。为了处理这些节点，同时无需 在比对过程中访问图的邻接表，我们为每个节点引入两个位：（*i*）稀有拓扑 位和（*ii*）特殊边 位。如果某个节点的稀有拓扑 位置为 1，则说明该节点具有罕见的基序组合（即它所属的组并非最常见的组），此时保留其基序映射，而不是组 ID。我们的 GIAB（HPRC）人类基因组图中，只有 0.0027%（0.0138%）的节点具有稀有拓扑。由于 amotif 只能表示排序顺序中相邻节点之间的边，我们引入特殊边位 ，以紧凑地存储其余少量边，即指向远处节点的边（例如由结构变异造成的边）和反向边。我们使用一个列表来记录具有特殊边位的节点的远距离邻居或后向邻居

标记。我们的 GIAB（HPRC）人类基因组图中，只有 0.18%（3.17%）的节点有远距离邻居，而 HPRC 图中只有 1.23% 的节点有后向邻居。在全部 24 条染色体上，GIAB（HPRC）图的完整索引（MotivoIndex 和最小值索引）大小为 8.15 GB（14.25 GB）。

5.2. S2G映射算法

逐步流程。图 9 展示了Motivo的端到端 S2G 在线读段比对算法，该算法包含五个步骤：（*1*）最小值提取，（*2*）种子生成，（*3*）扩展情况选择，（*4*）扩展，以及（*5*）S2S 比对。首先，*最小值提取* 从读序列中提取最小值子序列。其次，*种子生成* 使用这些最小化器查询基于最小化器的索引，并在图中查找种子命中位置。第三，*扩展情况选择* 确定所需的扩展方向，以便获得足够的核苷酸与读段进行比对。随后，*MotivoExt* 查询 MotivoIndex，根据所选的扩展方向，检索种子命中节点发起或终止的基序。MotivoExt 利用基序拓扑结构和节点的碱基组成，确定哪个基序所代表的路径最有可能产生有效比对。最后，*S2S 比对* 将这些路径的序列与读段进行比对，并生成最终比对及其比对质量（MAPQ）。**最小值提取**。该

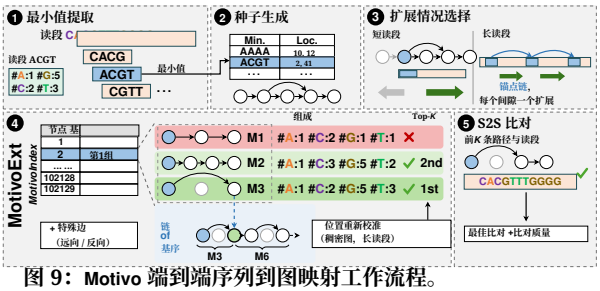


图 9：Motivo 端到端序列到图映射工作流程。

第一步提取读段的最小值，使用与基于最小化器的索引相同的哈希函数和参数（k, w），并统计其碱基，以获得读段的 ACGT含量，供后续阶段使用。长读段被划分为不重叠的窗口；我们计算每个窗口的 ACGT含量，并以密度 D 采样最小化器，以限制种子生成工作量。**种子生成**。种子生成步骤会使用读段的最小值查询基于最小化器的索引。每个种子命中都会保留其图中的位置（节点 ID 及其在节点序列中的偏移量）和其在读段中的偏移量。**扩展情况选择**。此步骤利用种子命中的位置信息，确定是否需要扩展种子命中子图，以及需要沿哪个方向扩展。目标是根据种子命中在参考序列和读序列中的偏移量，使子图覆盖足够多的字符，以完整比对读段。扩展方向有四种可能情况：(*i*) 无延伸（NE），即种子命中子图仅包含种子命中节点；(*ii*) 左扩展（LE），即种子命中子图需要包含种子命中节点左侧的节点；(*iii*) 右扩展（RE），即种子命中子图需要包含种子命中节点右侧的节点；以及(*iv*) 双向扩展（BE），即同时进行左扩展和右扩展。对于长读段，此步骤会在扩展之前执行链式连接。链式连接将读段和图中彼此接近的种子命中连接成锚点链。**扩展**。扩展步骤将种子命中扩展为种子命中子图，并过滤掉无法形成有效比对的子图。它使用碱基计数（BC）过滤器处理情况NE，并使用、RE、BE与MotivoExt处理情况。所有情况均使用MotivoIndex以及读段与候选比对序列之间的 ACGT 距离，$\sum_{i\in\{A,\ C,\ G,\ T\}} |SeqA_content_i - SeqB_content_i|$。两个序列之间的最少编辑次数为$\lceil ACGT_distance/2 \rceil$。无需扩展时（即参考子序列完全位于种子命中节点内），我们从MotivoIndex中提取该子序列；若其与读段之间的最少编辑次数超过编辑距离阈值，BC 过滤器便会将其丢弃。对于多节点种子命中子图，MotivoExt 首先从MotivoIndex中提取种子命中节点的起始（结束）基序，用于右（左）扩展。其次，它利用节点的序列长度，找到每个基序的结束（起始）节点中的精确末端（起始）偏移量。第三，它将该节点片段的 ACGT含量加到基序中间节点预先计算的 ACGT含量上，并丢弃最少编辑次数超过编辑距离阈值的基序。第四，对于足够长、能够完全位于读段范围内的节点，MotivoExt会进行轻量的位置重新校准，即在预期位置附近，对节点序列和读段进行短片段精确匹配比较。这能使读段沿路径保持正确位置，因此仅对这些位置之间的短读段片段计算 ACGT含量；候选路径也只有在这些片段上才会实际产生差异。对于长读段，由于变异位点的许多替代路径通常具有相同的 ACGT含量，MotivoExt 还会使用位运算，将两者分别直接与读段进行比较。这些特性使 MotivoExt 仅通过比较和加法，就能适应更稠密的图和噪声更大的读段。最后，MotivoExt保留排名前K 的候选路径，并按最少编辑次数排序；对于并列或编辑次数相差不超过一次的候选项，则使用 S2S 比对器的一部分，根据其与短读段片段之间的精确编辑距离重新排序。这些路径构成排名前K 的种子命中子图，即有效比对基序。如果种子命中的节点没有基序覆盖所需数量的碱基，MotivoExt 就会从每条保留路径的最后一个节点继续扩展，将路径构建为基序链。对于基序中的稀有节点，它会从其远端邻居或后向邻居继续扩展。对于长读段，同一过程会扩展每个锚点，直至到达链中的下一个锚点。

比对。每条候选路径都可能与读序列产生有效比对。我们提取其节点的序列，将它们拼接成一条线性序列，再使用基于位运算的比对算法 GenASM [181] 将其与读段进行比对。所有候选路径中最佳的比对结果即为比对结果，也就是路径坐标和 CIGAR 字符串；CIGAR 字符串以紧凑形式编码了该比对中的匹配、错配、插入和缺失。由于Motivo 会保留排名靠前的 K 条路径，因此也会保留读段排名靠前的基因座的最佳比对结果

，并像传统比对器一样，根据最佳和次佳比对结果之间的差异计算比对质量 [286,310]。

5.3. 序列到图映射硬件

Motivo的序列到图映射硬件设计采用完全流水线化的五级数据流架构，各级之间使用双缓冲。图 10 展示了该设计的高层概览，其中包含五个模块：最小值提取模块 、种子生成模块 、扩展情况选择模块 、*MotivoExt* 模块 和*S2S* 比对与位点模块 。

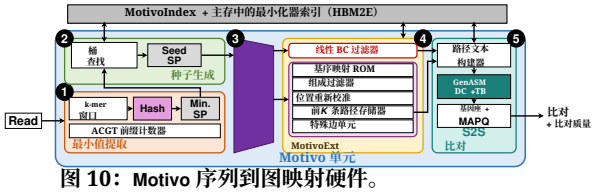


图 10：Motivo 序列到图映射硬件。

5.3.1. 最小值提取。最小化器通过滚动窗口单次遍历提取，使用一个 w项循环缓冲区来存放当前窗口中的 k-mer，并由组合逻辑最小化器查找器每个周期选出一个最小化器。每个最小化器及其读段偏移量都会写入片上暂存区（最小化器暂存区），因此在种子生成模块查询前一个最小化器的同时，预取下一个最小化器。如果找到种子命中，则将位置偏移量转发至模块 。

5.3.2. 种子生成。 基于最小值的索引是由固定大小桶组成的哈希表，因此每次查找只需进行一次主存访问。种子命中位置会写入片上暂存存储器（种子暂存区）。

5.3.3. 扩展情况选择。 模块 是一个简单的组合电路，用于比较种子命中在读段中的偏移量、在种子命中节点中的偏移量与该节点的序列长度，并将结果解码为 $enable$ 相应扩展模块的信号。长读段的链式连接由一个小型控制器实现，该控制器通过简单比较节点 ID 和位置距离来构建锚点链。

5.3.4. 扩展。 该模块包含 MotivoExt 模块和简单线性碱基计数过滤器，后者用于无扩展情况。碱基计数过滤器执行简单的碱基计数、用于计算 ACGT 距离的算术运算，以及根据给定的编辑距离阈值比较并舍弃读段/参考序列对。

***MotivoExt*模块。**扩展模块的核心是MotivoExt 模块。图 11 展示了MotivoExt 模块的硬件架构，该模块负责对种子命中进行基于基序的延伸。 基序检索 。得益于基因组变异图固有的属性，包括已知基序的存在及其可重复性，我们可以

高效地将其位掩码和组基序映射保存在片上只读存储器中，以减少片外存储器访问。为尽量减小该只读存储器的面积，我们利用以下观察：最常见的 256 个组覆盖了我们的图中 99.85%（98.1%）的节点

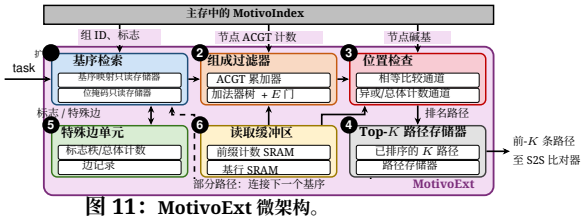


图 11：MotivoExt 微架构。

（GIAB（HPRC）图）；因此，我们只需 4 KB 片上存储器即可存储它们（用于起始组和结束组），并根据扩展方向进行选择。由于我们只有 62 个 7 位基序位掩码，因此将其存储在一个小型只读存储器中。每个节点的组 ID、稀有位和跳转位、序列长度及 ACGT含量都被打包到一条 16 字节记录中，因此一次存储器访问便可同时获取其拓扑结构和属性。由于这些记录按拓扑序存储，一个基序的节点会落在一到两个相邻的 64 字节缓存行中，小型行缓存可利用这一点。只有对于稀有节点，才会从主存中读取该节点自身的基序映射。以位运算方式表示基序映射和基序位掩码，只需简单地移位并提取最低有效位，即可检测当前基序及其节点。 组成过滤器 。ACGT 累加器会累加基序中各节点预先计算的 ACGT含量；由于边界节点可能只需使用其序列的一部分，因此通过对 2 位编码序列进行掩码群体计数来统计所需部分。 读取缓冲区 存储读段的 ACGT 前缀计数，因此只需进行一次减法，即可得到任意读取片段的 ACGT含量。加法器树计算最少编辑次数，比较器则丢弃超过编辑阈值的基序。位置重新校准 使用并行相等比较器在读段中定位长节点，并重新锚定读段。对于长读段，异或和群体计数通道将替代等位基因与读段进行比较。 *Top-K* 路径存储器 将 Top-K 条路径保存在一个小型排序缓冲区中，并将其中的部分路径反馈给块 ，以串联它们的下一个基序。 跳转单元 会检索跳转 位（§5.1）被置位的节点的远端和后向邻居。由于只有 Top-K 条路径会传递到 S2S 比对步骤，因此每次扩展只需调用比对器少数几次，而不是针对每个候选项都调用一次，从而将开销高昂的比对集中在最有希望的子图上。

5.3.5. 比对。为了构建排名靠前的-K路径序列以供比对步骤使用，路径文本构建器模块从内存中获取每条候选路径的节点序列并将其拼接起来，然后将结果传递给基于 GenASM-DC/TB [181] 的 S2S 比对模块。GenASM-DC 使用位运算（如与/或/异或/移位）执行位向量动态规划，GenASM-TB 则遍历存储的动态规划位向量以恢复比对结果。一个小型暂存存储器保存每个排名靠前位点的最佳比对，而查找表则将它们的编辑距离差转换为 MAPQ 分数。

6. 评估方法

数据集。 我们评估两组图。(*i*)**GIAB图**：与先前工作类似 [210],，我们使用 GRCh38 参考基因组和 GIAB项目（v3.3.2）中 HG001-HG007 的变异，通过 vg construct 构建24个图（染色体1-22,

表 1：评测系统和数据集（√：由我们测量；○：引自系统论文；–：不支持）。

平台	系统	GIAB图 短读长	GIAB图 长读长	HPRC图 短读长	HPRC图 Long
CPU	GraphAligner [253]	√	√	√	√
2×AMDEPYC 9554 vg [245]		√	√	√	√
	minigraph [258]	–	√	–	√
	Motivo-CPU	√	√	√	√
	vg-giraffe	√	–	√	–
GPU	vg-giraffe [293]	√	–	√	–
英伟达 A100 80 GB	Motivo-GPU	√	–	√	–
ASIC	SeGraM [210]	○	○	–	–
28纳米	HARP [211]	–	–	–	○
	Motivo-ASIC	√	√	√	√

X和Y）。(*ii*) **HPRC图**：为评估更稠密的泛基因组图，我们使用人类泛基因组参考联盟（HPRC） [248, 326]按染色体构建的Minigraph-Cactus 图。对于每条染色体和每个图，我们使用Mason [327],模拟（*i*）三组短读长数据集，读段长度分别为100、150和250碱基对，错误率为1%（Illumina特征）；以及（*ii*）两组长读长数据集，每组包含10千碱基对的 PacBio连续长读段，并使用 PBSIM3 [328]进行模拟。我们采用 PacBio错误模型 ERRHMM-RSII，平均错误率分别为5%和10%。每组包含100万条读段。

评测系统。我们在三个平台上评估Motivo 算法：Motivo-CPU和Motivo-GPU 分别在中央处理器和图形处理器上以软件方式运行该算法，而Motivo-ASIC 则通过硬件端到端运行该算法。我们将每种实现与其所在平台上最先进的S2G映射系统进行比较，结果汇总于表1。我们将Motivo-ASIC实现为专用集成电路，以确保与最先进的基于ASIC的S2G加速器进行公平比较 [210, 211]。我们在各系统支持的所有数据集类型上进行评估。基于 CPU 的系统均在同一台高端服务器上运行，基于图形处理器的系统则均在同一块 NVIDIA A100 GPU 上运行，并采用各自性能最佳的配置。[2] 对于 SeGraM [210] 和 HARP [211], , 我们采用其论文报告的面积和功耗。SeGraM 仅在 GIAB图上进行评估，因为对于更稠密的图，尤其是长读长（<50%），其准确率会大幅下降，这是由于其子图大小的上限较低。这样的设置将Motivo 算法带来的优势与硬件带来的优势区分开来。首先，由于Motivo-CPU 和 Motivo-GPU 与各自的基线运行在相同的平台和数据集上，将 Motivo-CPU （Motivo-GPU）与基于中央处理器（图形处理器）的基线进行比较，即可单独评估算法带来的优势。其次，由于 Motivo-GPU 生成的映射与Motivo-CPU相同，且我们通过 Motivo-CPU验证Motivo-ASIC ，因此比较这三种实现即可单独评估硬件带来的优势。对于所有系统，我们仅测量映射步骤，不包括离线图构建和索引构建；后两者每个参考基因组只需执行一次，并不处于基因组分析的关键路径上 [329]。

Motivo 性能。 我们实现了全部 Motivo 可综合的 Verilog RTL，并针对 Motivo-CPU 验证了该 RTL，使用 Verilator [330]。我们使用根据 RTL 周期数校准的周期级模拟器评估吞吐量，该模拟器通过 Ramulator 2.1 [331, 332]（基于早期版本的 Ramulator [333–336]）对片外存储器进行建模。我们使用该模拟器，以便在大型读数据集和各种图上快速评估 Motivo 。默认配置采用 32 个 Motivo 单位和一个 HBM2E堆栈（8 个通道、16 个伪通道、16 GB）。参考索引（对于较大的图，大小最高达 14.25GB）仅加载到堆栈中一次，供所有单位共享（详情见 §7.1.5）。我们选择 HBM2E，即 SeGraM [210], 所采用的存储技术，以便在相同的存储技术下比较这两种加速器。

面积、功耗和能耗。 我们使用 Yosys v0.33 [337] 对Motivo 进行综合（采用 Nangate45标准单元库，目标时钟频率为 1GHz），并使用 OpenROAD 开源 RTL到GDS流程 [338] 进行物理设计，以获取布局后的面积和功耗。对于所有专用集成电路系统，我们仅比较硬件计算单元的面积、功耗和能耗。该比较对基线有利：HARP 在主机CPU上执行种子生成，但其面积和功耗中不包含主机CPU；SeGraM 使用四个 HBM2E堆栈和复制的索引，而Motivo-ASIC 仅使用一个堆栈。我们使用 Ayna [339]，, 根据 Ramulator 仿真的 DRAM命令跟踪对 HBM 功耗进行单独建模，并将其从 HBM2 缩放至 HBM2E。对于 CPU系统，我们测量处理器封装的能耗计数器。对于 GPU系统，我们使用 NVML [340] 测量 GPU板卡的功耗。对于所有系统，我们仅测量映射步骤期间的性能、功耗和能耗，不包括索引加载。

7. 评估分析

7.1. 性能

图 12 展示了基于中央处理器和图形处理器的基线，以及Motivo的三种实现：Motivo-CPU、Motivo-GPU和Motivo-ASIC的吞吐量。我们从两个角度分析这些结果。首先，我们将Motivo-CPU和Motivo-GPU 与相同硬件上的最先进的基线进行比较，这体现了Motivo 算法的优势（§7.1.1）。其次，我们比较Motivo的三种实现方式，这体现了硬件的优势（§7.1.2）。

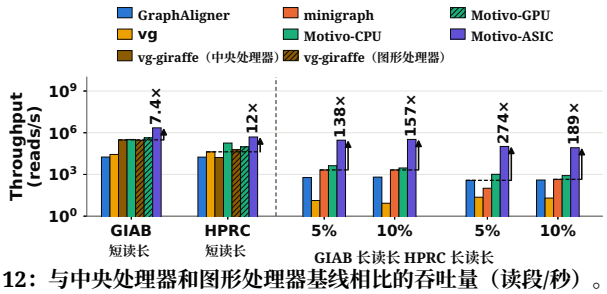


图 12：与中央处理器和图形处理器基线相比的吞吐量（读段/秒）。

7.1.1. 与基于 CPU 和 GPU 的系统比较。我们得出两个关键观察结果。首先，在相同硬件上，Motivo的算法优于最先进的系统。在中央处理器上，Motivo-CPU 相比 GraphAligner、vg、vg-giraffe 和 minigraph 的加速比分别为 2.1-18.4×、 2.8-343×、 1.1-7.1×、

[2]由于 vg-giraffe 仅支持短读段，因此我们仅使用短读段对 GPU 进行评估。

以及 1.4-10.1× 。在图形处理器上，Motivo-GPU 相比 vg-giraffe 的加速比为 1.5- 1.6×，且每个读段的能耗低 1.7-1.9×（图 17）。其次，随着图密度增加，算法的优势也会扩大。更稠密的 HPRC图 在每个种子命中周围具有更多替代路径。因此，基线会逐节点遍历更稠密图中的额外路径，而Motivo则通过基序编码获取预计算路径。例如，在 GIAB图 上，对于短读段，vg-giraffe 的性能接近Motivo-CPU ；但在 HPRC图 上，Motivo-CPU 比 vg-giraffe 快 7.1× 倍。

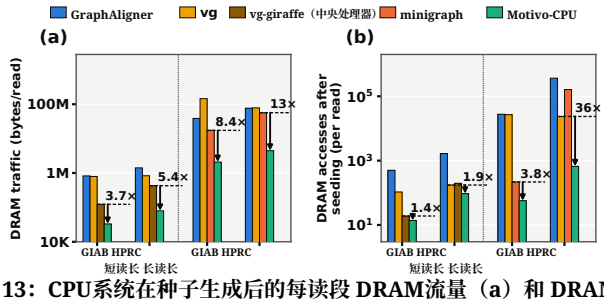


图 13：CPU系统在种子生成后的每读段 DRAM流量（a）和 DRAM访问（b），即候选选择和比对期间未命中所有缓存的加载操作。

性能提升的来源。图 13 展示了基于 CPU 的系统每读段的 DRAM流量，以及种子生成后的 DRAM访问。Motivo-CPU 在所有读段集合中的 DRAM流量最低，比表现最佳的基线低 3.7-13×（图 13a），这是因为MotivoIndex 用一次内存查找取代了每次图遍历（图 13b）。图越密集，这种减少就越明显。对于长读段，Motivo-CPU 在 GIAB图上的种子生成后 DRAM访问次数比 minigraph 少 3.8×，在更密集的 HPRC 图上则比表现最佳的基线 vg 最多少 36×。原因在于基线执行的图遍历次数更多。基线会逐个节点遍历更密集图中的额外路径，而Motivo的基序则能紧凑地对其进行编码。

7.1.2. Motivo 实现方案的比较。 由于Motivo-CPU、Motivo-GPU和 Motivo-ASIC 运行相同的算法，因此相互比较可以单独考察硬件带来的收益。我们有两个关键观察。首先，通用硬件限制了算法的收益。在 GIAB图上，Motivo-GPU 仅比Motivo-CPU 快 1.3×，而在 HPRC图上则更慢。在MotivoIndex 消除了大部分存储器访问（§7.1.1）后，剩余的工作就是路径选择。路径选择由轻量级查找和算术运算组成，但这些操作形成了由多个短小依赖步骤构成的链，其长度在不同种子命中之间差异很大（图 15b）。在 CPU 上，这些依赖步骤会因存储器延迟和分支而串行化；而在 GPU 上，不同的链长度使线程束中始终只有少数线程处于活动状态。因此，通用硬件将其资源耗费在控制流、存储器

延迟和分化，而非这些轻量级操作本身。其次，Motivo-ASIC消除了这些限制，从而带来了显著优势。由于路径选择只涉及查找和算术运算，Motivo-ASIC 通过MotivoExt执行路径选择。MotivoExt 是一种简单的专用单元，可处理

流水线中的依赖步骤，且不会产生指令开销或分化；此外，它还配备了小型片上存储器，以位掩码形式存储所有基序，因此很少需要进行片外访问。在采用相同算法的情况下，对于短读段（长读段），Motivo-ASIC的速度比Motivo-CPU 快 4.2–6.9× 倍（68–113× 倍），比Motivo-GPU快 5.2–5.3× 倍。长读段的加速比最大，因为其扩展过程会形成最长的依赖步骤链。因此，对于短读段，Motivo-ASIC 相较于最先进的基于CPU和基于GPU的基线，分别实现了 7.4–11.7× 倍和 7.7–8.2× 倍的加速；对于长读段，相较于最先进的基于CPU的基线，则实现了 138–274× 倍的加速。归根结底，在这三种实现方案中进行选择是一项设计决策。Motivo-CPU 和Motivo-GPU可利用现有硬件实现近期部署，而Motivo-ASIC 则（*i*）性能最高，（*ii*）能效最高（§7.2），并且（*iii*）仅需 36.7mm^2 和 10.6–12.0W，功耗低于我们基线系统中的单个 CPU 插槽或图形处理器，因此可用于紧凑型系统，进行现场基因组分析 [138, 139, 341–346]。

7.1.3. 与 S2G 加速器的比较。为进行公平比较，我们按照 [347] 中的方法，将 Motivo-ASIC 的面积和功耗缩放至 SeGraM 和 HARP 所采用的 28 纳米工艺技术节点。图 14 展示了吞吐量

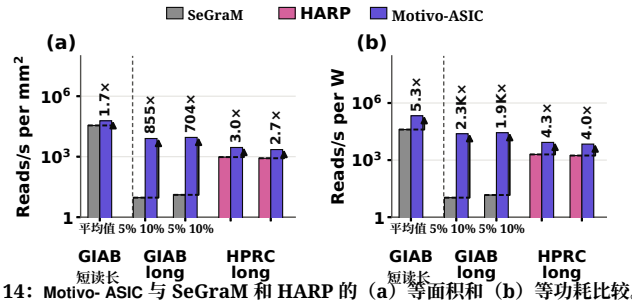


图 14：Motivo- ASIC 与 SeGraM 和 HARP 的（a）等面积和（b）等功耗比较。

（等面积）每平方毫米^2 和每瓦（等功耗）的吞吐量。我们得出两个关键结论。首先，在 GIAB图上，Motivo-ASIC 的性能优于 SeGraM。对于短读段，Motivo-ASIC的每平方毫米^2 （每瓦）吞吐量高出 1.7× （5.3×）倍；对于长读段，则高出 704-855×（1883-2286×）倍；其短读段（长读段）原始吞吐量则高出 1.9-2.7×（932-1132×）倍。其次，在 HPRC图上，Motivo-ASIC 的性能优于 HARP。对于长读段，Motivo-ASIC 的每平方毫米^2 吞吐量比 HARP 高出 2.7-3.0×倍，每瓦吞吐量高出 4.0-4.3×倍。除此之外，Motivo-ASIC 的功能也更加全面：与 HARP 不同，Motivo-ASIC 还可以比对短读段，并在芯片上执行种子生成。

改进来源。SeGraM 会比对每个种子命中的子图，而Motivo-ASIC 则会在比对前，利用基序选出少数最合适的路径。因此，在 GIAB 图上，与 SeGraM 相比，Motivo-ASIC 对短读段（长读段）减少了 1.5-4.5× (243-521×) 个候选项，评估的 DP 单元减少了 1.2-9.7× (310-710×) 个，且在种子生成后发出的 HBM 访问减少了 1.2-1.9× (36-85×) 次。HARP 对每条长读段比对的候选项数量相近（2.6-2.9 对比 4.0-4.2），但每个Motivo-ASIC 候选项的计算开销低 2.0-3.0× 倍，因此减少了 1.4-2.0× 个

DP 单元。此外，HARP 每条染色体的索引需要 1MB 片上存储器（占其面积的 42%）。

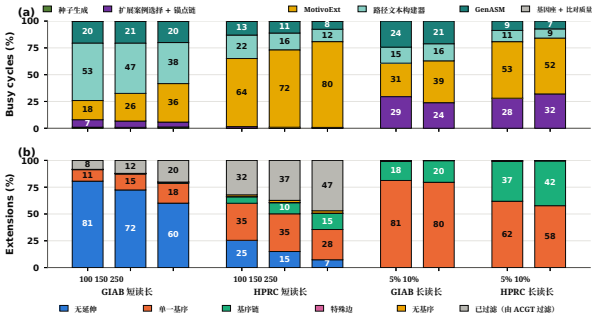
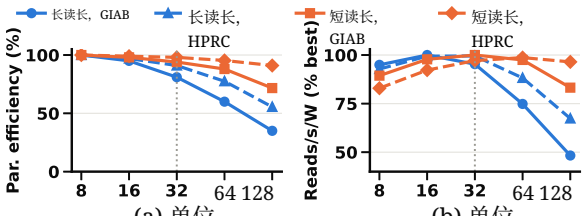


图 15：（a）Motivo 单元中各模块繁忙环数的占比。（b）MotivoExt 如何处理扩展。

7.1.4. Motivo-ASIC的执行细分。图 15 展示了（a）Motivo-ASIC单元中各模块繁忙环数的占比，以及（b）MotivoExt 如何处理扩展。

短读长。 在 GIAB图上，60-81% 的种子命中无需扩展，即其子图仅包含一个节点（图 15b），因此 MotivoExt 占用的繁忙周期少于 36%（图 15a）。最大的模块是路径文本构建器（38-53%），它获取候选路径的节点序列，并构建用于比对的线性参考序列。对于这些种子命中，Motivo-ASIC 和 SeGraM 使用相同的算法 GenASM 执行相同的 S2S 比对，这解释了在短读长上 Motivo-ASIC 相比 SeGraM 的适度性能提升（§7.1.3）。在更稠密的 HPRC图上，41-46% 的种子命中需要单一基序或基序链，而MotivoExt 会过滤掉 32-47% 的候选子图（图 15b），因此 MotivoExt 占据了繁忙周期的主要部分（64-80%）。由于这种过滤，比对不再是瓶颈，仅占繁忙周期的 8-13%。长读长。 锚点链会将每个长读段划分为短间隙，而单一基序或较短的基序链即可解析这些间隙。因此，Motivo-ASIC 通过少量基序查找来解析每个间隙，而不是在该间隙的子图上执行基于 DP 的比对。GenASM 至多占用繁忙周期的 24%，而 SeGraM 的映射时间几乎全部由其占用（§3）。Motivo-ASIC因此将瓶颈从基于 DP 的比对转移到了轻量级的基于基序的延伸，并且其优势会随着图密度的增加而扩大。



(a) 单位　　(b) 单位

图16：在使用共享索引的单个 HBM2E 堆栈上，Motivo 从8个扩展到128个单元时的可扩展性。（a）相对于8个单元的并行效率；（b）相对于每个数据集最佳单元数量的每瓦读段数（计算单元和 HBM）。

7.1.5. 可扩展性。图 16a 和图 16b 分别展示了在单个 HBM2E 堆栈上将 Motivo 的单位数从 8 个扩展到 128 个时的吞吐量和能效。在这种情况下，所有通道共享一个索引。对于 8 个单位，我们为每个通道分配一个单位；对于 16、32、64 和 128 个单位，我们分别为每个伪通道分配 1、2、4 和 8 个单位。能效测量包含这些单位和 HBM2E 堆栈的功耗。我们有两个观察结果。第一，对于短读段（并行效率为 93-99%）以及 HPRC图上的长读段（90-92%），吞吐量在扩展到 32 个单位之前几乎呈线性增长，超过 32 个单位后则趋于饱和。例如，对于 HPRC图上的长读段，从 32 个单位扩展到 64 个单位，会使并行效率从 90-92% 降至 74-81%。对于 GIAB图上的长读段，并行效率在 32 个单位时就已降至 79-83%。第二，32 个单位在吞吐量和能效之间实现了最佳平衡。例如，在所有数据集上，32 个单位均可达到最佳每瓦读段数的 93-100%；而对于长读段，128 个单位只能达到 43-72%。因此，我们采用 32 个单位，以充分发挥单个 HBM2E 堆栈的性能和能效。

7.2 面积、功耗与能耗

面积和功耗。 表2展示了采用28纳米工艺的Motivo单元的面积构成。单个单元占用1.146 mm^2,，因此32单元加速器占用36.7 mm^2,；相比之下，SeGraM占用27.7 mm^2，HARP占用7.93 mm^2。因此，Motivo-ASIC 的面积大于这两种加速器，功耗也高于 HARP（见下文），但其每平方毫米和每瓦的吞吐量均高于二者（§7.1.3）。我们有两点观察。首先，该单元面积中占比最大的是实现Motivo基于基序的机制。MotivoExt （23.2%）以及结合锚点链的扩展情况选择（21.0%）合计占据该单元的44.2%。扩展情况选择模块占用较大面积，是因为对于长读段，它需要存储每个读段的种子命中并构建锚点链。其次，基序处理的开销很低：MotivoExt 占用的面积不到该单元的四分之一。

表2：Motivo 单元的面积分解（28纳米）。

模块	mm^2 (%)
最小值提取 + 种子生成	0.053 (4.5)
扩展情况选择 +锚点链	0.241 (21.0)
MotivoExt	0.265 (23.2)
S2S 比对和基因座	0.490 (42.8)
包括 GenASM (DC +TB) [181]	0.334 (29.2)
共享读取缓冲区 + 流水线控制	0.097 (8.5)
总计	**1.146 (100)**

Motivo 单元的功耗在处理短读段时为 0.33 W，在处理长读段时为 0.376W。32 个单元组成的加速器功耗分别为 10.6W 和 12.0W；相比之下，SeGraMand 为 24.3W，HARP 为 3.82 W。HBM2E 内存的功耗为 1.2-7.5 W，具体取决于读段长度和图类型。

能耗。 图 17 展示了所有系统每条读段的能耗。Motivo 在所有数据集和各类图上的每条读段能耗均最低。与 S2G 加速器相比，在 GIAB图上处理短读段时，Motivo的每条读段能耗比 SeGraM 低 5.3×；在 HPRC图上处理长读段时，其能耗比 HARP 低 4.0-4.3×。图 17 展示了各系统的计算能耗。如果再计入Motivo的 HBM2E 内存能耗，整个Motivo系统每条读段的能耗仍低于其他所有系统单是计算部分的能耗：在 GIAB图上处理短读段时，比 SeGraM 低 4.8×；在 HPRC图上处理长读段时，比 HARP 低 3.1-3.2×。

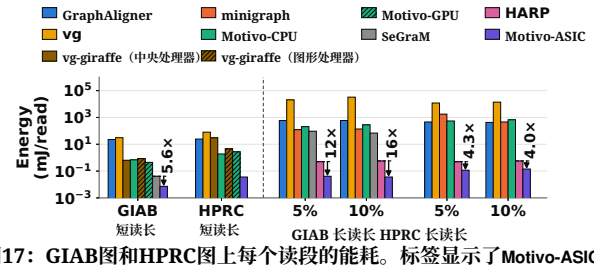


图17：GIAB图和HPRC图上每个读段的能耗。标签显示了Motivo-ASIC相较于SeGraM（短读段）和HARP（长读段）的能耗降低幅度。对于短读段，各柱状条表示三种读段长度下每个读段能耗的平均值（5.6× 相较于SeGraM）；若改为像图14那样对每瓦读段数取平均，则结果为5.3×。

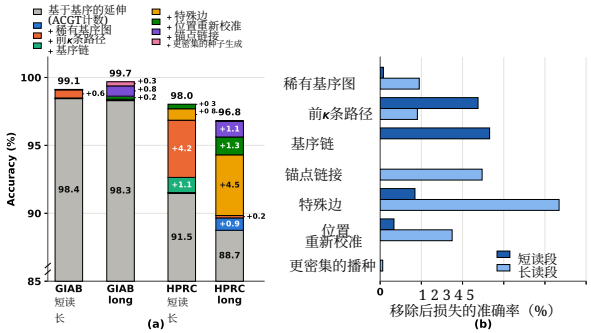


图19：（a）各机制对准确率的贡献。（b）移除单个机制导致的准确率损失（HPRC）。

7.3. 准确率

我们使用模拟读段确定其真实映射，并据此衡量Motivo的准确率。如果读段报告的图上起始和结束位置与其真实来源相同，且报告的比对（CIGAR）最多包含 E 个编辑操作（§2.3），则认为该读段比对正确。总体而言，Motivo在 GIAB图上的短读段和长读段准确率分别达到 99.1% 和 99.7%，在更稠密的 HPRC图上则分别达到 98.0% 和 96.8%。最先进的 S2G 映射器 GraphAligner 和 vg 报告的准确率分别为 96.6% 和 93.8% [253, 257]。

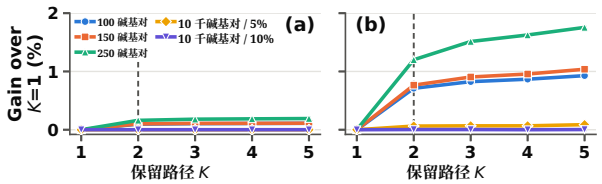


图18：对于（a）GIAB图和（b）HPRC图，使用 K 条路径而非单条路径进行比对时的准确率提升。

顶部K 路径分析。图18展示了扩展路径数量增加时准确率的提升。在GIAB图上，单条路径（$K = 1$）已经能够达到约 99%的准确率，增加第二条路径仅能提升0.1–0.2%。在更稠密的HPRC图上，第二条路径可使短读段准确率提升0.7–1.2%，且提升幅度随读段长度增加而增加；而长读段的提升最多为 0.1%。因此，我们保留 $K = 2$ 条路径，因为对$K > 2$ 条路径进行比对会使短读段（长读段）的比对操作最多增加2.5×（4.3×）。

各机制的贡献。我们通过两种互补的方式评估各机制的贡献（图 19）。首先，图 19a 展示了将各机制逐步加入原始设计后，准确率如何提升（即原始设计只保留一条路径、仅存储 256 个片上基序组，且既不使用特殊边，也不进行位置重新校准）。在 GIAB图上，原始设计已实现近乎完美的准确率（98.3%–98.4%），这证实了对于复杂度较低的图，仅凭基序拓扑和碱基组成就能达到很高的准确率。在 HPRC图上，这些机制使短读长（长读长）的准确率分别提高了 6.6%（8.1%）。其次，为了单独评估各机制对准确率的影响，图 19b 每次从最终设计中移除一种机制，并展示在更稠密的 HPRC图上造成的准确率损失。若不进行位置重新校准或不使用特殊边，设计的准确率最多会损失 4.3%。在更稠密的图上，在 GIAB图上，移除任一MotivoExt 机制，对短读长（长读长）的准确率造成的损失最多为 0.3%（0.2%）。因此，对于复杂度较低的图，即使使用 $K = 1$ 且不采用额外机制，Motivo也能以很高的准确率映射读段。

8. 相关工作

Motivo 是一种新颖的算法-架构协同设计序列到图比对系统，利用了基因组图中观察到的拓扑重复性。

加速S2S映射。 许多研究通过针对映射流水线各步骤进行算法和硬件优化，加速序列到序列（S2S）映射，即将读段映射到单个线性参考基因组 [138, 139, 190, 261, 348], 。**种子生成**通过哈希、最小值、syncmer、strobemer 和基于FM索引的种子，以及优化的种子选择 [271, 281, 286, 310, 312, 313, 349–360]（也适用于原始纳米孔信号 [361–363]）得到加速；此外，还利用FPGA、专用集成电路和存内/近存处理（PIM/PNM）系统上的FM索引搜索与种子查找硬件实现加速 [144,146,148–150, 152, 173, 364, 365]。**比对前过滤** 可在比对前快速拒绝不相似的读段-参考序列对；这一过程既通过软件实现 [366–369] ，也通过FPGA [262, 369–371],、图形处理器 [369, 372], 、PIM/PNM和存内系统 [184,186, 264, 373–376]上的硬件实现。种子命中**链式连接** 依赖稀疏动态规划算法 [286, 377–382] ，并通过中央处理器上的单指令多数据 [383], 、图形处理器 [384–386], 、FPGA [384, 387], 、定制 RISC-V指令 [388], 以及可编程动态规划加速器 [389]实现加速。**比对**在算法层面通过带状、位并行、差分递推、波前和基于 A*的经典动态规划算法实现加速 [181,260, 266, 297, 390–401], ；在硬件层面则通过图形处理器 [157–164, 402–407], 、FPGA [170, 172, 175–177, 408–410], 、专用集成电路和指令集架构扩展 [137, 151–153, 181, 189, 192, 389, 411], 以及存内处理 [145, 147, 182, 185, 412–415]实现加速。其他研究加速了完整的 S2S映射器 [154–156, 165, 166, 168, 169, 171, 178, 179, 387, 416–418] ，以及读段在不同参考基因组之间的重新映射 [419, 420]。所有这些工作都以线性参考基因组为目标，无法处理基因组图的复杂性。

用于序列到图映射的软件工具。S2G 映射工具 [245, 253, 257, 258, 284, 285, 291] 改进了**种子生成** [278–282, 290],**索引构建** [272–277], **过滤** [283, 286, 287], 和**扩展** [260, 288–290], 但仍会执行冗余且开销巨大的

基于 DP 的比对（§3）。Motivo通过基于 amotif 的索引和拓扑感知扩展机制避免了这一问题。

基于图的基因组分析的硬件加速。以往的工作利用 SIMD [265,320],图形处理器 [218, 292], 专用集成电路 [210, 211], 存内处理（PIM） [221], 和存储内处理来加速 S2G 映射。然而，SeGraM [210], 这一用于 S2G 映射的最先进的硬件加速器能够同时映射短读段和长读段，却缺乏冗余计算过滤功能，也不支持新兴的稠密泛基因组图。HARP [211]利用图中的长连续区域来加速索引构建和比对，但仅支持长读段，并且在主机上执行种子生成。MeG2 [221] 将 SeGraM 的种子生成和按位比对映射到存内处理。GRAINS [421]面向超出主存容量的基因组图，通过存储内处理和闪存内处理减少其存储 I/O，并将其识别出的候选图区域转发给 SeGraM。因此，MeG2 和 GRAINS 通过将计算移近数据来加速处理 [371, 422–425]。这些工作都没有利用基因组图中反复出现的局部拓扑结构。Motivo 利用了这一点，将 S2G 映射简化为每个读段仅需进行少量 S2S 比对；它能够准确映射短读段和长读段，并在芯片上运行从种子生成到比对质量评估的完整流水线。这种算法层面的简化与计算运行的位置正交，如第 §7 节所示。Motivo 算法并不局限于某一种平台，而 MotivoExt只需要查找、位掩码操作和加法，这些操作适用于存内逻辑和存储内逻辑。因此，Motivo可以部署在类似 MeG2 的存内处理系统中，也可以替代 SeGraM，充当 GRAINS 的比对引擎，将 GRAINS 的 I/O 节省与每个读段大幅减少的比对次数结合起来。

其他应用的图加速器。 基于图的近似最近邻搜索（ANNS）加速器 [426,427]也会离线建立索引并过滤掉不太可能的候选项，但它们返回的是近似的、最近邻存储的定长向量：无法针对插入缺失进行偏移，也不存储候选路径。将其用于序列到图映射需要将图展平为定长窗口，这要么忽略等位基因组合，要么将其全部枚举并构建规模庞大的索引，而且仍然需要进行比对。

9. 结论

我们提出Motivo，这是首个用于序列到图映射、利用重复出现的子图拓扑的算法-架构协同设计。其核心观察是，基因组图由共享有限种基序的小型子图构成。Motivo的基序算法会在比对前识别出少数能够导向有效比对的路径，以内存查找和算术运算取代图遍历和动态规划。它可在CPU和GPU上实现高性能执行，并在协同设计的Motivo-ASIC上实现高性能、低能耗执行。

参考文献

[1] Can Alkan、Jeffrey M Kidd、Tomas Marques-Bonet、Gozde Aksay、Francesca An- tonacci、Fereydoun Hormozdiari、Jacob O Kitzman、Carl Baker、Maika Malig、Onur Mutlu、S Cenk Sahinalp、Richard A Gibbs 和 Evan E Eichler。利用新一代测序绘制个性化拷贝数和节段性重复图谱。《自然·遗传学》，2009年。

[2] Gaye Lightbody、Valeriia Haberland、Fiona Browne、Laura Taggart、Huiru Zheng、Eileen Parkes 和 Jaine K Blayney。个性化医疗中高通量测序应用综述：研究与临床应用未来发展的障碍与促进因素。《生物信息学简报》，2019年。[3] Stefania Morganti、Paolo Tarantino、Emanuela Ferraro、Paolo D'Amico、Bruno Achutti Duso 和 Giuseppe Curigliano。下一代测序（NGS）：药物基因组学与癌症个性化医疗领域的一项革命性技术。载于癌症个人基因组学时代的转化研究与肿瘤组学应用，2019年。[4] Iuliia Branco 和 Altino Choupina。生物信息学：生命科学与个性化医疗中的新工具和应用。《应用微生物学与生物技术》，2021年。[5] Sameer Quazi。精准医学与基因组医学中的人工智能和机器学习。《医学肿瘤学》，2022年。[6] Samuel J. Aronson 和 Heidi L. Rehm。为精准医学中的基因组学奠定基础。《自然》，2015年。[7] Hannah F. Löchel 和 Dominik Heider。精准医学中下一代测序错误处理策略的比较分析。《科学报告》，2020年。[8] Eirini Papadopoulou、Dimitra Bouzarelou、George Tsaousis、Athanasios Papathana- siou、Georgia Vogiatzi、Charalambos Vlachopoulos、Antigoni Miliou、Panagiota Papachristou、Efstathia Prappa、Georgios Servos、Konstantinos Ritsatos、Aristeidis Seretis、Alexandra Frogoudaki 和 George Nasioulas。下一代测序在心脏病学中的应用：当前与未来的精准医疗意义。《心血管医学前沿》，2023年。[9] Alireza Tafazoli、Henk-Jan Guchelaar、Wojciech Miltyk、Adam J. Kretowski 和 Jesse J. Swen。在临床实践中应用下一代测序平台进行药物基因组学检测。《药理学前沿》，2021年。[10] Valentina Gambardella、Noelia Tarazona、Juan M. Cejalvo、Pasquale Lo mbardi、Marisol Huerta、Susana Roselló、Tania Fleitas、Desamparados Roda 和 Andres Cervantes。个性化医疗：癌症治疗的最新进展。《癌症》，2020年。[11] Rebecca J. Leary、Isaac Kinde、Frank Diehl、Kerstin Schmidt、Chris Clouser、Cisilya Duncan、Alena Antipova、Clarence Lee、Kevin McKernan、Francisco M. De La Vega、Kenneth W. Kinzler、Bert Vogelstein、Luis A. Diaz 和 Victor E. Velculescu。利用大规模并行测序开发个性化肿瘤生物标志物。《科学转化医学》，2010年。[12] Hamburg Margaret A. 和 Collins Francis S. 个性化医疗之路。《新英格兰医学杂志》，2010年。[13] Maaike van der Lee、Marjolein Kriek、Henk-Jan Guchelaar 和 Jesse J. Swen。药物基因组学技术综述。《基因》，2020年。[14] Dabin Moon、Hye W. Park、Dongheon Surl、Dongju Won、Seung-Tae Lee、Saeam Shin、Jong R. Choi 和 Jinu Han。韩国队列遗传性眼病的下一代测序精准医疗。《基因》，2022年。[15] Sumitra Mohan、Victoria Foy、Mahmood Ayub、Hui Sun Leong、Pieta Schofield、Sudhakar Sahoo、Tine Descamps、Bedirhan Kilerci、Nigel K. Smith、Mathew Carter、Lynsey Priest、Cong Zhou、T. Hedley Carr、Crispin Miller、Corinne Faivre-Finn、Fiona Blackhall、Dominic G. Rothwell、Caroline Dive 和 Gerard Brady。对小细胞肺癌患者的循环游离DNA进行靶向和全基因组测序分析。《胸部肿瘤学杂志》，2020年。[16] Claudia C.Y. Chung、Gordon K.C. Leung、Christopher C.Y. Mak、Jasmine L.F. Fung、Mianne Lee、Steven L.C. Pei、Mullin H.C. Yu、Vivian C.C. Hui、Joshua C.K. Chan、Jef- frey F.T. Chau、Marcus C.Y. Chan、Mandy H.Y. Tsang、Wilfred H.S. Wong、Joanna Y.L. Tung、Kin Shing Lun、Yiu Ki Ng、Cheuk Wing Fung、Mabel S.C. Wong、Rosanna M.S. Wong、Yu Lung Lau、Godfrey C.F. Chan、So Lun Lee、Kit San Yeung 和 Brian H.Y. Chung。快速全外显子组测序促进儿科罕见病患者的精准医疗并降低医疗成本。《柳叶刀区域健康—西太平洋》，2020年。[17] Suzette J. Bielinski、Janet E. Olson、Jyotishman Pathak、Richard M. Wein- shilboum、Liewei Wang、Kelly J. Lyke、Euijung Ryu、Paul V. Targonski、Michael D. Van Norstrand、Matthew A. Hathcock、Paul Y. Takahashi、Jennifer B. McCormick、Kiley J. Johnson、Karen J. Maschke、Carolyn R. Rohrer Vitek、Marissa S. Ellingson、Eric D. Wieben、Gianrico Farrugia、Jody A. Morrisette、Keri J. Kruckeberg、Jamie K. Bruflat、Lisa M. Peterson、Joseph H. Blommel、Jennifer M. Skierka、Matthew J. Ferber、John L. Black、Linnea M. Baudhuin、Eric W. Klee、Jason L. Ross、Tamra L. Veldhuizen、Cloann G. Schultz、Pedro J. Caraballo、Robert R. Freimuth、Christopher G. Chute 和 Iftikhar J. Kullo。个性化医疗的预先基因分型：设计合适的药物、剂量和时机——利用基因组数据制定个体化治疗方案。《梅奥诊所学报》，2014年。[18] Dean Ho、Stephen R. Quake、Edward R.B. McCabe、Wee Joo Chng、Edward K. Chow、Xianting Ding、Bruce D. Gelb、Geoffrey S. Ginsburg、Jason Hassenstab、Chih-Ming Ho、William C. Mobley、Garry P. Nolan、Steven T. Rosen、Patrick Tan、Yun Yen 和 Ali Zarrinpar。个性化医疗与精准医学的赋能技术。《生物技术趋势》，2020年。[19] Bashdar Mahmud Hussen、Sara Tharwat Abdullah、Abbas Salihi、Dana Khdr Sabir、Karzan R. Sidiq、Mohammed Fatih Rasul、Hazha Jamal Hidayat、So udeh Ghafouri- Fard、Mohammad Taheri 和 Elena Jamali。下一代测序在临床肿瘤学与个性化医疗中的新兴作用。《病理学：研究与实践》，2022年。[20] Laura E. Russell、Yitian Zhou、Ahmed A. Almousa、Jasleen K. Sodhi、Chuk- wunonso K. Nwabufo 和 Volker M. Lauschke。下一代测序时代的药物基因组学——从字节到临床。《药物代谢评论》，2021年。[21] Renu Verma、Kesia Esther da Silva、Neesha Rockwood、Roeland E. Wasmann、Nombuso Yende、Taeksun Song、Eugene Kim、Paolo Denti、Robert J. Wilkinson,

和 Jason R. Andrews。基于纳米孔测序的药物基因组学检测面板：实现结核病药物剂量个性化。《美国呼吸与重症医学杂志》，2024年。

[22] Michelle Clark、Amber Hildreth、Sergey Batalov、Yan Ding、Shimul Chowdhury、Kelly Watkins、Katarzyna Ellsworth、Brandon Camp、Cyrielle I. Kint、Calum Yacoubian、Lauge Farnaes、Matthew N. Bainbridge、Curtis Beebe、Joshua J. A. Braun、Margaret Bray、Jeanne Carroll、Julie A. Cakici、Sara A. Caylor、Christina Clarke、Mitchell P. Creed、Jennifer Friedman、Alison Frith、Richard Gain、Mary Gaughran、Shauna George、Sheldon Gilmer、Joseph Gleeson、Jeremy Gore、Haiying Grunenwald、Raymond L. Hovey、Marie L. Janes、Kejia Lin、Paul D. McDonagh、Kyle McBride、Patrick Mulrooney、Shareef Nahas、Daeheon Oh、Albert Oriol、Laura Puckett、Zia Rady、Martin G. Reese、Julie Ryu、Lisa Salz、Erica Sanford、Lawrence Stewart、Nathaly Sweeney、Mari Tokita、Luca Van Der Kraan、Sarah White、Kristen Wigby、Brett Williams、Terence Wong、Meredith S. Wright、Catherine Yamada、Peter Schols、John Reynders、Kevin Hall、David Dimmock、Narayanan Veeraraghavan、Thomas Defay、and Stephen F. Kingsmore. 重症患儿遗传疾病的诊断：采用快速全基因组测序、自动化表型分析和结果解读。《科学转化医学》, 11(489):eaat6177, 2019年。

[23] Lauge Farnaes、Amber Hildreth、Nathaly M. Sweeney、Melissa C. Clark、Shimul Chowdhury、Shareef Nahas、Yabei Cakici、Wendy Benson、Robert H. Kaplan、Richard Kronick、Matthew N. Bainbridge、Jennifer Friedman、Jeffrey J. Gold、Yan Ding、Narayanan Veeraraghavan、David Dimmock 和 Stephen F. Kingsmore. 快速全基因组测序降低婴儿发病率并减少住院成本化。*npj* 基因组医学，3(1):10，2018年。

[24] Nathaly M. Sweeney、Shareef A. Nahas、Shimul Chowdhury、Sergey Batalov、Michelle Clark、Sara Caylor、Julie Cakici、John J. Nigro、Yan Ding、Narayanan Veeraraghavan、Charlotte Hobbs、David Dimmock 和 Stephen F. Kingsmore。快速全基因组测序对先天性心脏病婴儿的医疗护理和资源利用的影响。《npj 基因组医学》，6(1):29，2021年。[25] Mauricio Flores、Gustavo Glusman、Kristin Brogaard、Nathan D Price 和 Leroy Hood。P4 医学：系统医学将如何改变医疗保健行业和社会。《个性化医疗》，2013年。[26] Geoffrey S Ginsburg 和 Huntington F Willard。基因组学与个性化医学：基础与应用。《转化研究》，2009年。[27] Lynda Chin、Jannik N Andersen 和 P Andrew Futreal。癌症基因组学：从发现科学到个性化医疗。《自然·医学》，2011年。[28] Euan A Ashley。迈向精准医学。《自然综述：遗传学》，2016年。[29] Michael S. Lawrence、Petar Stojanov、Paz Polak、Gregory V. Kryukov、Kristian Cibulskis、Andrey Sivachenko、Scott L. Carter、Chip Stewart、Craig H. Mermel、Steven A. Roberts、Adam Kiezun、Peter S. Hammerman、Aaron McKenna、Yotam Drier、Lihua Zou、Alex H. Ramos、Trevor J. Pugh、Nicolas Stransky、Elena Helman、Jaegil Kim、Carrie Sougnez、Lauren Ambrogio、Elizabeth Nickerson、Erica Shefler、Maria L. Cortés、Daniel Auclair、Gordon Saksena、Douglas Voet、Michael Noble、Daniel DiCara、Pei Lin、Lee Lichtenstein、David I. Heiman、Timothy Fennell、Marcin Imielinski、Bryan Hernandez、Eran Hodis、Sylvan Baca、Austin M. Dulak、Jens Lohr、Dan-Avi Landau、Catherine J. Wu、Jorge Melendez-Zajgla、Alfredo Hidalgo-Miranda、Amnon Koren、Steven A. McCarroll、Jaume Mora、Ryan S. Lee、Brian Crompton、Robert Onofrio、Melissa Parkin、Wendy Winckler、Kristin Ardlie、Stacey B. Gabriel、Charles W. M. Roberts、Jaclyn A. Biegel、Kimberly Stegmaier、Adam J. Bass、Levi A. Garraway、Matthew Meyerson、Todd R. Golub、Dmitry A. Gordenin、Shamil Sunyaev、Eric S. Lander 和 Gad Getz。癌症中的突变异质性及对新型癌症相关基因的探索。《自然》，2013年。[30] Bert Vogelstein、Nickolas Papadopoulos、Victor E. Velculescu、Shibin Zhou、Luis A. Diaz 和 Kenneth W. Kinzler。癌症基因组图谱。《科学》，2013年。[31] Daniel Ramsköld、Shujun Luo、Yu-Chieh Wang、Robin Li、Qiaolin Deng、Omid R Faridani、Gregory A Daniels、Irina Khrebtukova、Jeanne F Loring、Louise C Laurent、Gary P Schroth 和 Rickard Sandberg。来自单细胞水平RNA及单个循环肿瘤细胞的全长mRNA测序。《自然·生物技术》，2012年。[32] Timour Baslan 和 James Hicks。借助单细胞测序揭示癌症生物学并转变研究范式。《自然综述：癌症》，2017年。[33] Ehud Shapiro、Tamir Biezuner 和 Sten Linnarsson。基于单细胞测序的技术将彻底改变全生物体科学。《自然综述：遗传学》，2013年。[34] Yoshitaka Sakamoto、Sarun Sereewattanawoot 和 Ayako Suzuki。癌症基因组学长读长测序的新时代。《人类遗传学杂志》，2020年。[35] Qingzhu Jia、Han Chu、Zheng Jin、Haixia Long 和 Bo Zhu。癌症研究中的高通量单细胞测序。《信号转导与靶向治疗》，2022年。[36] Devon A. Lawson、Kai Kessenbrock、Ryan T. Davis、Nicholas Pervolarakis 和 Zena Werb。单细胞分辨率下的肿瘤异质性与转移。《自然·细胞生物学》，2018年。[37] Chuang Liu、Qiangqiang Shi、Xiangang Huang、Seyoung Koo、Na Kong 和 Wei Tao。基于mRNA的癌症疗法。《自然综述：癌症》，2023年。[38] Bram Van de Sande、Joon Sang Lee、Euphemia Mutasa-Gottgens、Bart Naughton、Wendi Bacon、Jonathan Manning、Yong Wang、Jack Pollard、Melissa Mendez、Jon Hill、Namit Kumar、Xiaohong Cao、Xiao Chen、Mugdha Khaladkar、Ji Wen、Andrew Leach 和 Edgardo Ferran。单细胞RNA测序在药物发现与开发中的应用。《自然综述：药物发现》，2023年。[39] Debyani Chakravarty 和 David B. Solit。临床癌症基因组分析。《自然综述：遗传学》，2021年。[40] Isidro Cortés-Ciriano、Doga C. Gulhan、Jake June-Koo Lee、Giorgio E. M. Melloni 和 Peter J. Park。癌症基因组测序数据的计算分析。《自然综述：遗传学》，2022年。

[41] Ira W. Deveson、Binsheng Gong、Kevin Lai、Jennifer S. LoCoco、Todd A. Richmond、Jeoffrey Schageman、Zhihong Zhang、Natalia Novoradovskaya、James C. Willey、Wendell Jones、Rebecca Kusko、Guangchun Chen、Bindu Swapna Madala、James Blackburn、Igor Stevanovski、Ambica Bhandari、Devin Close、Jeffrey Conroy、Michael Hubank、Narasimha Marella、Piotr A. Mieczkowski、Fujun Qiu、Robert Sebra、Daniel Stetson、Lihyun Sun、Philippe Szankasi、Haowen Tan、Lin-ya Tang、Hanane Arib、Hunter Best、Blake Burgher、Pierre R. Bushel、Fergal Casey、Simon Cawley、Chia-Jung Chang、Jonathan Choi、Jorge Dinis、Daniel Duncan、Agda Karina Eterovic、Liang Feng、Abhisek Ghosal、Kristina Giorda、Sean Glenn、Scott Happe、Nathan Haseley、Kyle Horvath、Li-Yuan Hung、Mirna Jarosz、Garima Kushwaha、Dan Li、Quan-Zhen Li、Zhiguang Li、Liang-Chun Liu、Zhichao Liu、Charles Ma、Christopher E. Mason、Dalila B. Megherbi、Tom Morrison、Carlos Pabón-Peña、Mehdi Pirooznia、Paula Z. Proszek、Amelia Raymond、Paul Rindler、Rebecca Ringler、Andreas Scherer、Rita Shaknovich、Tieliu Shi、Melissa Smith、Ping Song、Maya Strahl、Venkat J. Thodima、Nikola Tom、Suman Verma、Jiashi Wang、Leihong Wu、Wenzhong Xiao、Chang Xu、Mary Yang、Guangliang Zhang、Sa Zhang、Yilin Zhang、Leming Shi、Weida Tong、Donald J. Johann、Timothy R. Mercer、Joshua Xu，以及 SEQC2肿瘤面板测序工作组。评估循环肿瘤DNA测序检测在精准肿瘤学中的分析有效性。《自然·生物技术》，2021年。

[42] Wenming Xiao、Luyao Ren、Zhong Chen、Li Tai Fang、Yongmei Zhao、Justin Lack、Meijian Guan、Bin Zhu、Erich Jaeger、Liz Kerrigan、Thomas M. Blomquist、Tiffany Hung、Marc Sultan、Kenneth Idler、Charles Lu、Andreas Scherer、Rebecca Kusko、Malcolm Moos、Chunlin Xiao、Stephen T. Sherry、Ogan D. Abaan、Wanqiu Chen、Xin Chen、Jessica Nordlund、Ulrika Liljedahl、Roberta Maestro、Maurizio Polano、Jiri Drabek、Petr Vojta、Sulev Kõks、Ene Reimann、Bindu Swapna Madala、Tim-othy Mercer、Chris Miller、Howard Jacob、Tiffany Truong、Ali Moshrefi、Aparna Natarajan、Ana Granat、Gary P. Schroth、Rasika Kalamegham、Eric Peters、Virginie Petitjean、Ashley Walton、Tsai-Wei Shen、Keyur Talsania、Cristobal Juan Vera、Kurt Langenbach、Maryellen de Mars、Jennifer A. Hipp、JamesC. Willey、Jing Wang、Jyoti Shetty、Yuliya Kriga、Arati Raziuddin、Bao Tran、Yuanting Zheng、Ying Yu、Margaret Cam、Parthav Jailwala、Cu Nguyen、Daoud Meerzaman、Qingrong Chen、Chunhua Yan、Ben Ernest、Urvashi Mehra、Roderick V. Jensen、Wendell Jones、Jian-Liang Li、Brian N. Papas、Mehdi Pirooznia、Yun-Ching Chen、Fayaz Seifuddin、Zhipan Li、Xuelu Liu、Wolfgang Resch、Jingya Wang、Leihong Wu、Gokhan Yavas、Corey Miles、Baitang Ning、Weida Tong、Christopher E. Mason、Eric Donaldson、Samir Lababidi、Louis M. Staudt、Zivana Tezak、Huixiao Hong、Charles Wang 和 Leming Shi。《全基因组和全外显子组测序癌症突变检测最佳实践》。《自然·生物技术》，2021年。

[43] Kelly E. Ho、Daoni Ayers、Mirna Jarosz、Prashanth Vedanayagam、Syed M. Mariko Yabe、Catherine C. Coombs、Nicole M. Caltabellotta、Mike Walsh、Kenneth Offit、Zsofia Stadler、Diana Mandelker、Jessica Schulman、Akshar Patel、John Philip、Elsa Bernard、Gunes Gundem、Juan E. Arango Ossa、Max Levine、Juan S. Medina Martinez、Noushin Farnoud、Dominik Glodzik、Sonya Li、Mark E. Robson、Choonsik Lee、Paul D. P. Pharoah、Konrad H. Stopsack、Barbara Spitzer、Simon Mantha、James Fagin、Laura Boucai、Christopher J. Gibson、Benjamin L. Ebert、Andrew L. Young、Todd Druley、Koichi Takahashi、Nancy Gillis、Markus Ball、Eric Padron、David M. Hyman、Jose Baselga、Larry Norton、Stuart Gardos、Virginia M. Klimek、Howard Scher、Dean Bajorin、Eder Paraiso、Ryma Benayed、Maria E. Arcila、Marc Ladanyi、David B.Solit、Michael F. Berger、Martin Tallman、Montserrat Garcia-Closas、Nilanjan Chatterjee、Luis A. Diaz、Ross L. Levine、Lindsay M. Morton、Ahmet Zehir、and Elli Papaem- manuil. 癌症治疗会塑造克隆性造血的适应度景观。《自然·遗传学》，2020年。

[44] Joseph D. Szustakowski、Suganthi Balasubramanian、Erika Kvikstad、Shareef Khalid、Paola G. Bronson、Ariella Sasson、Emily Wong、Daren Liu、J. Wade Davis、Carolina Haefliger、A. Katrina Loomis、Rajesh Mikkilineni、Hyun Ji Noh、Samir Wadhawan、Xiaodong Bai、Alicia Hawes、Olga Krasheninina、Ricardo Ulloa、Alex E. Lopez、Erin N. Smith、Jeffrey F. Waring、Christopher D. Whelan、Ellen A. Tsai、John D. Overton、William J. Salerno、Howard Jacob、Sandor Szalma、Heiko Runz、Gregory Hinkle、Paul Nioi、Slavé Petrovski、Melissa R. Miller、Aris Baras、Lyndon J. Mitnaul、Jeffrey G. Reid、Oleg Moiseyenko、Carlos Rios、Saurabh Saha、Goncalo Abecasis、Nilanjana Banerjee、Christina Beechert、Boris Boutkov、Michael Cantor、Giovanni Coppola、Aris Economides、Gisu Eom、Caitlin Forsythe、Erin D. Fuller、Zhenhua Gu、Lukas Habegger、Marcus B. Jones、Rouel Lanche、Michael Lattari、Michelle LeBlanc、Dadong Li、Luca A. Lotta、Kia Manoochehri、Adam J. Mansfield、Evan K.Maxwell、Ja- son Mighty、Mrunali Nafde、Sean O'Keeffe、Max Orelus、Maria Sotiropoulos Padilla、Razvan Panea、Tommy Polanco、Manasi Pradhan、Ayesha Rasool、Thomas D. Schle- icher、Deepika Sharma、Alan Shuldiner、Jeffrey C. Staples、Cristopher V. Van Hout、Louis Widom、Sarah E. Wolf、Sally John、Chia-Yen Chen、David Sexton、Varant Ku- pelian、Eric Marshall、Timothy Swan、Susan Eaton、Jimmy Z. Liu、Stephanie Loomis、Megan Jensen、Saranya Duraisamy、Jason Tetrault、David Merberg、Sunita Badola、Mark Reppell、Jason Grundstad、Xiuwen Zheng、Aimee M. Deaton、Margaret M. Parker、Lucas D. Ward、Alexander O. Flynn-Carroll、Caroline Austin、Ruth March、Menelas N. Pangalos、Adam Platt、Mike Snowden、Athena Matakidou、Sebastian Wasilewski、Quanli Wang、Sri Deevi、Keren Carss、Katherine Smith、Morten So- gaard、Xinli Hu、Xing Chen、Zhan Ye，以及 UKB-ESC研究团队。通过对英国生物样本库进行外显子组测序，推进人类遗传学研究和药物发现。《自然·遗传学》，2021年。

[45] Nicholas Navin 和 James Hicks. 单细胞测序的未来医学应用-测序在癌症中的应用。*GenomeMedicine*，2011。

[46] Mingye Feng、Shuang Tao、Ling Zhang、Ting Dian、Xuanmei Huang、Shaohui Huang、Shu-Juan Xie、Zhen-Dong Xiao 和 Hua Zhang。RNA测序：癌症研究中的新

技术与应用。*Journal of Hematology &Oncology*，2020年。[47] Yalan Lei、Rong Tang、Jin Xu、Wei Wang、Bo Zhang、Jiang Liu、Xianjun Yu 和 Si Shi。单细胞测序在癌症研究中的应用：进展与展望。*Journal ofHematology&Oncology*，2021年。[48] Yingying Han、Dan Wang、Lushan Peng、Tao Huang、Xiaoyun He、Junpu Wang 和 Chunlin Ou。单细胞测序：揭示肿瘤转移机制的有前景方法。*Journal of Hematology & Oncology*，2022年。[49] Giulia Federici 和 Silvia Soddu。高通量基因组测序时代意义未明变异的解读：乳腺癌和卵巢癌的启示。*Journal of Experimental & Clinical Cancer Research*，2020年。[50] Yijie Zhang、Dan Wang、Miao Peng、Le Tang、Jiawei Ouyang、Fang Xiong、Can Guo、Yanyan Tang、Yujuan Zhou、Qianjin Liao、Xu Wu、Hui Wang、Jianjun Yu、Yong Li、Xiaoling Li、Guiyuan Li、Zhaoyang Zeng、Yixin Tan 和 Wei Xiong。单细胞RNA测序在癌症研究中的应用。*Journal of Experimental & Clinical Cancer Research*，2021年。[51] Xianwen Ren、Boxi Kang 和 Zemin Zhang。利用单细胞测序了解肿瘤生态系统：前景与局限。*Genome Biology*，2018年。[52] Liqing Tian、Yongjin Li、Michael N. Edmonson、Xin Zhou、Scott Newman、Clay McLeod、Andrew Thrasher、Yu Liu、Bo Tang、Michael C. Rusch、John Easton、Jing Ma、Eric Davis、Austyn Trull、J. Robert Michael、Karol Szlachta、Charles Mullighan、Suzanne J. Baker、James R. Downing、David W. Ellison 和 Jinghui Zhang。CICERO：一种利用癌症RNA测序数据检测复杂多样驱动基因融合的通用方法。*Genome Biology*，2020年。[53] Eoghan R. Malone、Marc Oliva、Peter J. B. Sabatini、Tracy L. Stockley 和 Lillian L. Siu。精准癌症治疗的分子谱分析。*Genome Medicine*，2020年。[54] Xiaoning Tang、Yongmei Huang、Jinli Lei、Hui Luo 和 Xiao Zhu。单细胞测序：新进展与医学应用。*Cell & Bioscience*，2019年。[55] Darrell L. Ellsworth、Heather L. Blackburn、Craig D. Shriver、Shahrooz Rabizadeh、Patrick Soon-Shiong 和 Rachel E. Ellsworth。单细胞测序与肿瘤发生：加深对肿瘤演化和转移的理解。*Clinical and Translational Medicine*，2017年。[56] Yiming Zhong、Feng Xu、Jinhua Wu、Jeffrey Schubert 和 Marilyn M. Li。新一代测序在实验室医学中的应用。*AnnalsofLaboratory Medicine*，2021年。[57] Zsofia K. Stadler、Anna Maio、Debyani Chakravarty、Yelena Kemel、Margaret Sheehan、Erin Salo-Mullen、Kaitlyn Tkachuk、Christopher J. Fong、Bastien Nguyen、Amanda Erakky、Karen Cadoo、Ying Liu、Maria I. Carlo、Alicia Latham、Hongxin Zhang、Ritika Kundra、Shaleigh Smith、Jesse Galle、Carol Aghajanian、Nadeem Abu-Rustum、Anna Varghese、Eileen M. O'Reilly、Michael Morris、Wassim Abida、Michael Walsh、Alexander Drilon、Gowtham Jayakumaran、Ahmet Zehir、Marc Ladanyi、Ozge Ceyhan-Birsoy、David B. Solit、Nikolaus Schultz、Michael F. Berger、Diana Mandelker、Luis A. Diaz、Kenneth Offit 和 Mark E. Robson。晚期癌症患者胚系检测的治疗意义。*Journal of Clinical Oncology*，2021年。[58] Aaron C. Tan 和 Daniel S.W. Tan。针对具有致癌驱动基因分子改变的肺癌患者的靶向治疗。*Journal of Clinical Oncology*，2022年。[59] Andrea Degasperi、Xueqing Zou、Tauanne Dias Amarante、Andrea Martinez-Martinez、Gene Ching Chiek Koh、João M. L. Dias、Laura Heskin、Lucia Chmelova、Giuseppe Rinaldi、Valerie Ya Wen Wang、Arjun S. Nanda、Aaron Bernstein、So-phie E. Momen、Jamie Young、Daniel Perez-Gil、Yasin Memari、Cherif Badja、Scott Shooter、Jan Czarnecki、Matthew A. Brown、Helen R. Davies、Genomics England Research Consortium3†、Serena Nik-Zainal、J. C. Ambrose、P. Arumugam、R. Bevers、M. Bleda、F. Boardman-Pretty、C. R. Boustred、H. Brittain、M. J. Caulfield、G. C. Chan、T. Fowler、A. Giess、A. Hamblin、S. Henderson、T. J. P. Hubbard、R. Jackson、L. J. Jones、D. Kasperaviciute、M. Kayikci、A. Kousathanas、L. Lahnstein、S. E. A. Leigh、I. U. S. Leong、F. J. Lopez、F. Maleady-Crowe、M. McEntagart、F. Minneci、L. Moutsianas、M. Mueller、N. Murugaesu、A. C. Need、P. O'Donovan、C. A. Odhams、C. Patch、D. Perez-Gil、M. B. Pereira、J. Pullinger、T. Rahim、A. Rendon、T. Rogers、K. Savage、K. Sawant、R. H. Scott、A. Siddiq、A. Sieghart、S. C. Smith、A. Sosinsky、A. Stuckey、M. Tanguy、A. L. Taylor Tavares、E. R. A. Thomas、S. R. Thompson、A. Tucci、M. J. Welland、E. Williams、K. Witkowska 和 S. M. Wood。英国群体全基因组测序癌症中的碱基替换突变特征。*Science*，2022年。[60] Junfen Xu、Yifeng Fang、Kelie Chen、Sen Li、Sangsang Tang、Yan Ren、Yixuan Cen、Weidong Fei、Bo Zhang、Yuanming Shen 和 Weiguo Lu。单细胞RNA测序揭示人类高级别浆液性卵巢癌的组织结构。*Clinical Cancer Research*，2022年。[61] Peter Horak、Christoph Heining、Simon Kreutzfeldt、Barbara Hutter、Andreas Mock、Jennifer Hüllein、Martina Fröhlich、Sebastian Uhrig、Arne Jahn、Andreas Rump、Laura Gieldon、Lino Möhrmann、Dorothea Hanf、Veronica Teleanu、Christoph E. Heilig、Daniel B. Lipka、Michael Allgäuer、Leo Ruhnke、Andreas Laßmann、Volker Endris、Olaf Neumann、Roland Penzel、Katja Beck、Daniela Richter、Ulrike Winter、Stephan Wolf、Katrin Pfütze、Christina Geörg、Bettina Meißburger、Ivo Buchhalter、Marinela Augustin、Walter E. Aulitzky、Peter Hohenberger、Matthias Kroiss、Peter Schirmacher、Richard F. Schlenk、Ulrich Keilholz、Frederick Klauschen、Gunnar Fol-precht、Sebastian Bauer、Jens Thomas Siveke、Christian H. Brandts、Thomas Kindler、Melanie Boerries、Anna L. Illert、Nikolas von Bubnoff、Philipp J. Jost、Karsten Spiekermann、Michael Bitzer、Klaus Schulze-Osthoff、Christof von Kalle、Barbara Klink、Benedikt Brors、Albrecht Stenzinger、Evelin Schröck、Daniel Hübschmann、Wilko Weichert、Hanno Glimm 和 Stefan Fröhling。全面的基因组与转录组分析，为罕见癌症患者的治疗决策提供指导。*Cancer Discovery*，2021年。[62] Xiaoyan Zhang

, Sadie L. Marjani , Zhaoyang Hu , Sherman M. Weissman , Xinghua

Pan、Shixiu Wu。《单细胞测序助力精准癌症研究：进展与展望》。癌症研究，2016。[63] Rossella Bruno 和 Gabriella Fontanini。《新一代测序在肺癌基因融合分析中的应用：文献综述》。《*Diagnostics*》，2020。[64] Antonella De Luca、Riziero Esposito Abate、Anna M. Rachiglio、Monica R. Maiello、Claudia Esposito、Clorinda Schettino、Francesco Izzo、Guglielmo Nasti 和 Nicola Normanno。《癌症中的FGFR融合：从诊断方法到治疗干预》。《*International Journal of Molecular Sciences*》，2020。[65] Michael R. Waarts、Aaron J. Stonestrom、Young C. Park 和 Ross L. Levine。《癌症突变靶向治疗》。《*The Journal of Clinical Investigation*》，2022。[66] Bora Lim、Yiyun Lin 和 Nicholas Navin。《以单细胞基因组学推进癌症研究与医学》。《*Cancer Cell*》，2020。[67] Ramon Colomer、Rebeca Mondejar、Nuria Romero-Laorden、Arantzazu Alfranca、Francisco Sanchez-Madrid 和 Miguel Quintela-Fandino。《癌症患者何时应接受新一代测序检测？》。《*eClinicalMedicine*》，2020。[68] Assieh Saadatpour、Shujing Lai、Guoji Guo 和 Guo-Cheng Yuan。《癌症基因组学中的单细胞分析》。《*Trends in Genetics*》，2015。[69] Nazli Dizman、Zeynep E. Arslan、Matthew Feng 和 Sumanta K. Pal。《转移性肾细胞癌的序贯治疗》。《*Urologic Clinics*》，2020。[70] Anton Buzdin、Maxim Sorokin、Andrew Garazha、Alexander Glusker、Alex Aleshin、Elena Poddubskaya、Marina Sekacheva、Ella Kim、Nurshat Gaifullin、Alf Giese、Alexander Seryakov、Pavel Rumiantsev、Sergey Moshkovskii 和 Alexey Moiseev。《RNA测序在临床肿瘤学研究与诊断中的应用》。《*Seminars in Cancer Biology*》，2020。[71] Yi Xiao 和 Dihua Yu。《肿瘤微环境作为癌症治疗靶点》。《*Pharmacology & Therapeutics*》，2021。[72] Arun Nandwani、Shalu Rathore 和 Malabika Datta。《癌症中的长链非编码RNA：调控与治疗意义》。《*Cancer Letters*》，2021。[73] Francesco Marchetti、Renato Cardoso、Connie L. Chen、George R. Douglas、Joanne Elloway、Patricia A. Escobar、Tod Harper、Robert H. Heflich、Darren Kidd、An-thony M. Lynch、Meagan B. Myers、Barbara L. Parsons、Jesse J. Salk、Raja S. Settivari、Stephanie L. Smith-Roe、Kristine L. Witt、Carole L. Yauk、Robert Young、Shaofei Zhang 和 Sheroy Minocherhomji。《误差校正的新一代测序——遗传毒性与癌症风险评估的前景与挑战》。《*Mutation Research/Reviews in Mutation Research*》，2023。[74] Xiao Chen、Han Zhu、Chun Qiao、Sishu Zhao、Lu Liu、Yan Wang、Huimin Jin、Sixuan Qian 和 Yujie Wu。《下一代测序揭示急性髓系白血病患者的基因突变图谱与克隆演化》。《*Hematology*》，2021。[75] Nicholas E. Navin。《单细胞癌症基因组学的前五年及未来》。《基因组研究》，2015。[76] Tim Dunn、Harisankar Sadasivan、Jack Wadden、Kush Goliya、Kuan-Yu Chen、David Blaauw、Reetuparna Das 和 Satish Narayanasamy。《SquiggleFilter：用于便携式病毒检测的加速器》。发表于*MICRO*，2021。[77] Claire Bertelli 和 Gilbert Greub。《快速细菌基因组测序：临床微生物学中的方法与应用》。《*Clinical Microbiology and Infection*》，2013。[78] Armando Arias、Simon J. Watson、Danny Asogun、Ekaete Alice Tobin、Jia Lu、My V. T. Phan、Umaru Jah、Raoul Emeric Guetiya Wadoum、Luke Meredith、Lucy Thorne、Sarah Caddy、Alimamy Tarawalie、Pinky Langat、Gytis Dudas、Nuno R. Faria、Simon Dellicour、Abdul Kamara、Brima Kargbo、Brima Osaio Kamara、Sahr Gevao、Daniel Cooper、Matthew Newport、Peter Horby、Jake Dunning、Foday Sahr、Tim Brooks、Andrew J.H. Simpson、Elisabetta Groppelli、Guoying Liu、Nisha Mu-lakken、Kate Rhodes、James Akpablie、Zabulon Yoti、Margaret Lamunu、Esther Vitto、Patrick Otim、Collins Owilli、Isaac Boateng、Lawrence Okoror、Emmanuel Omo-moh、Jennifer Oyakhilome、Racheal Omiunu、Ighodalo Yemisis、Donatus Adomeh、Solomon Ehikhiametalor、Patience Akhilomen、Chris Aire、Andreas Kurth、Nicola Cook、Jan Baumann、Martin Gabriel、Roman Wölfel、Antonino Di Caro、Miles W. Carroll、Stephan Günther、John Redd、Dhamari Naidoo、Oliver G. Pybus、Andrew Rambaut、Paul Kellam、Ian Goodfellow 和 Matthew Cotten。《在塞拉利昂快速开展埃博拉病毒暴发测序，识别出与散发病例相关的传播链》。《*Virus Evolution*》，2016。[79] Jessica Comin、Armando Chaure、Alberto Cebollada、Daniel Ibarz、Jesús Viñuelas、María Asunción Vitoria、María José Iglesias 和 Sofía Samper。《利用全基因组测序调查快速蔓延的结核病疫情》。《*Infection, Genetics and Evolution*》，2020。[80] Joshua Quick、Nicholas J. Loman、Sophie Duraffour、Jared T. Simpson、Ettore Sev-eri、Lauren Cowley、Joseph Akoi Bore、Raymond Koundouno、Gytis Dudas、Amy Mikhail、Nobila Ouédraogo、Babak Afrough、Amadou Bah、Jonathan H. J. Baum、Beate Becker-Ziaja、Jan Peter Boettcher、Mar Cabeza-Cabrerizo、Álvaro Camino-Sánchez、Lisa L. Carter、Juliane Doerrbecker、Theresa Enkirch、Isabel García Dorival、Nicole Hetzelt、Julia Hinzmann、Tobias Holm、Liana Eleni Kafetzopoulou、Michel Koropogui、Abigael Kosgey、Eeva Kuisma、Christopher H. Logue、Antonio Mazzarelli、Sarah Meisel、Marc Mertens、Janine Michel、Didier Ngabo、Katja Nitzsche、Elisa Pallasch、Livia Victoria Patrono、Jasmine Portmann、Johanna Gabriella Repits、Natasha Y. Rickett、Andreas Sachse、Katrin Singethan、Inês Vitoriano、Rahel L. Yemanaberhan、Elsa G. Zekeng、Trina Racine、Alexander Bello、Amadou Alpha Sall、Ousmane Faye、Oumar Faye、N'Faly Magassouba、Cecelia V. Williams、Victoria Amburgey、Linda Winona、Emily Davis、Jon Gerlach、Frank Washington、Vanessa Monteil、Marine Jourdain、Marion Bererd、Alimou Camara、Hermann Somlare、Abdoulaye Camara、Marianne Gerard、Guillaume Bado、Bernard Baillet、Déborah Delaune、Koumpingnin Yacouba Nebie、Abdoulaye Diarra、Yacouba Savane、Ray-mond Bernard Pallawo、Giovanna Jaramillo Gutierrez、Natacha Milhano、Isabelle Roger、Christopher J. Williams、Facinet Yattara、Kuiama Lewandowski、James Tay-lor、Phillip Rachwal、Daniel J. Turner、Georgios Pollakis、Julian A. Hiscox 和 David A.

Matthews、Matthew K. O' Shea、Andrew McD. Johnston、Duncan Wilson、Emma Hutley、Erasmus Smit、Antonino Di Caro、Roman Wölfel、Kilian Stoecker、Erna Fleischmann、Martin Gabriel、Simon A. Weller、Lamine Koivogui、Boubacar Diallo、Sakoba Keïta、Andrew Rambaut、Pierre Formenty、Stephan Günther 和 Miles W. Carroll。《用于埃博拉监测的实时便携式基因组测序》。《自然》，2016。[81] Esther R. Robinson、Timothy M. Walker 和 Mark J. Pallen。《基因组学与疫情调查：从序列到结论》。*Genome Medicine*，2013。[82] Pierre-Edouard Fournier、Gregory Dubourg 和 Didier Raoult。《基因组学时代细菌病原体的临床检测与表征》。*Genome Medicine*，2014。[83] Claudio U. Köser、Matthew J. Ellington、Edward J. P. Cartwright、Stephen H. Gille-spie、Nicholas M. Brown、Mark Farrington、Matthew T. G. Holden、Gordon Dougan、Stephen D. Bentley、Julian Parkhill 和 Sharon J. Peacock。《微生物全基因组测序在诊断和公共卫生微生物学中的常规应用》。《*PLOS Pathogens*》，2012。[84] Marc Eloit 和 marc lecuit。《利用全基因组新一代测序诊断传染病：新时代即将开启》。《*Frontiers in Cellular and Infection Microbiology*》，2014。[85] Gardy Jennifer L.、Johnston James C.、Sui Shannan J. Ho、Cook Victoria J.、Shah Lena、Brodkin Elizabeth、Rempel Shirley、Moore Richard、Zhao Yongjun、Holt Robert、Varhol Richard、Birol Inanc、Lem Marcus、Sharma Meenu K.、Elwood Kevin、Jones Steven J.M.、Brinkman Fiona S.L.、Brunham Robert C. 和 Tang Patrick。《结核病疫情的全基因组测序与社交网络分析》。《新英格兰医学杂志》，2011。[86] Taylor Angela J.、Lappi Victoria、Wolfgang William J.、Lapierre Pascal、Palumbo Michael J.、Medus Carlota 和 Boxrud David。《利用全基因组测序单核苷酸多态性分析，对肠炎沙门氏菌血清型肠炎亚种的食源性暴发进行表征，以开展监测和暴发检测》。《*Journal of Clinical Microbiology*》，2015。[87] Quainoo Scott、Coolen Jordy P. M.、van Hijum Sacha A. F. T.、Huynen Martijn A.、Melchers Willem J. G.、van Schaik Willem 和 Wertheim Heiman F. L.《细菌病原体的全基因组测序：医院感染暴发分析的未来》。《临床微生物学评论》，2017。[88] Goldberg Brittany、Sichtig Heike、Geyer Chelsie、Ledeboer Nathan 和 Weinstock George M.《从研究实验室迈向临床：下一代测序在传染病诊断中的挑战与机遇》。《*mBio*》，2015。[89] John M. Besser、Heather A. Carleton、Eija Trees、Steven G. Stroika、Kelley Hise、Matthew Wise 和 Peter Gerner-Smidt。《全基因组测序在肠道疾病监测和暴发调查中的结果解读》。《*Foodborne Pathogens and Disease*》，2019。[90] Weiwei Li、Qingpo Cui、Li Bai、Ping Fu、Haihong Han、Jikai Liu 和 Yunchang Guo。《全基因组测序在中国食源性疾病监测国家分子溯源网络中的应用》。《*Foodborne Pathogens and Disease*》，2021。[91] Yinhua Deng、Min Jiang、Patrick S.L. Kwan、Chao Yang、Qiongcheng Chen、Yiman Lin、Yaqun Qiu、Yinghui Li、Xiaolu Shi、Liqiang Li、Yujun Cui、Qun Sun 和 Qinghua Hu。《用于肠炎沙门氏菌血清型肠炎亚种暴发检测与源头追踪的集成式全基因组测序基础设施》。《*Foodborne Pathogens and Disease*》，2021。[92] Jason C. Kwong、Nadine McCallum、Vitali G. Sintchenko 和 Benjamin P. Howden。《全基因组测序在临床与公共卫生微生物学中的应用》。《*Pathology*》，2015。[93] Ruud H. Deurenberg、Erik Bathoorn、Monika A. Chlebowicz、Natacha Couto、Mithila Ferdous、Silvia García-Cobos、Anna M.D. Kooistra-Smid、Erwin C. Raangs、Sigrid Rosema、Alida C.M. Veloo、Kai Zhou、Alexander W. Friedrich 和 John W.A. Rossen。《下一代测序在临床微生物学和感染预防中的应用》。《*Journal of Biotechnology*》，2017。[94] Patrick Tang、Matthew A. Croxen、Mohammad R. Hasan、William W.L. Hsiao 和 Linda M. Hoang。《基因组流行病学新时代的感染控制》。《*American Journal of Infection Control*》，2017。[95] Nicholas J Croucher 和 Xavier Didelot。《基因组学在追踪细菌病原体传播中的应用》。宿主–微生物相互作用：细菌 • 基因组学，2015。[96] Joshua S. Bloom、Laila Sathe、Chetan Munugala、Eric M. Jones、Molly Gasperini、Nathan B. Lubock、Fauna Yarza、Erin M. Thompson、Kyle M. Kovary、Jimin Park、Dawn Marquette、Stephania Kay、Mark Lucas、TreQuan Love、A. Sina Booeshaghi、Oliver F. Brandenberg、Longhua Guo、James Boocock、Myles Hochman、Scott W. Simpkins、Isabella Lin、Nathan LaPierre、Duke Hong、Yi Zhang、Gabriel Oland、Bianca Judy Choe、Sukantha Chandrasekaran、Evann E. Hilt、Manish J. Butte、Robert Damoiseaux、Clifford Kravit、Aaron R. Cooper、Yi Yin、Lior Pachter、Omai B. Garner、Jonathan Flint、Eleazar Eskin、Chongyuan Luo、Sriram Kosuri、Leonid Kruglyak 和 Valerie A. Arboleda。《通过对混合并加条形码的鼻腔和唾液样本进行下一代测序，大规模扩大SARS-CoV-2 RNA检测》。《*Nature Biomedical Engineering*》，5(7):657–665，2021。[97] Ramesh Yelagandula、Aleksandr Bykov、Alexander Vogt、Robert Heinen、Ezgi Özkan、Marcus Martin Strobl、Juliane Christina Baar、Kristina Uzunova、Bence Hajdusits、Darja Kordic、Erna Suljic、Amina Kurtovic-Kozaric、Sebija Izetbegovic、Justine Schae-ffer、Peter Hufnagl、Alexander Zoufaly、Tamara Seitz、Mariam Al-Rawi、Stefan Ameres、Juliane Baar、Benedikt Bauer、Nikolaus Beer、Katharina Bergauer、Wolf-gang Binder、Claudia Blaukopf、Boril Bochev、Julius Brennecke、Selina Brinnich、Aleksandra Bundalo、Meinrad Busslinger、Tim Clausen、Geert de Vries、Marcus Dekens、David Drechsel、Zuzana Dzupinkova、Michaela Eckmann-Mader、Michaela Fellner, Thomas Fellner, Laura Fin, Bianca Valeria Gapp, Gerlinde Grab mann, Irina Grishkovskaya, Astrid Hagelkruys, Dominik Handler, David Haselbach, Louisa Hempel, Louisa Hill, David Hoffmann, Stefanie Horer, Harald Isemann, Robert Kalis, Max Kellner, Juliane Kley, Thomas Köcher, Alwin Köhler, Christian Krauditsch, Sabina Kula, Sonja Lang, Richard Latham, Marie-Christin Leitner, Thomas Leonard, Dominik Lindenhofer, Raphael Arthur Manzenreither, Martin Matl, Karl Mechtler, Anton Meinhart, Stefan Mereiter, Thomas Micheler, Paul Moeseneder, Tobias Neu- mann, Simon Nimpf, Magnus Nordborg, Egon Ogris, Michaela Pagani, Andrea Pauli, Jan-Michael Peters, Petra Pjevac, Clemens Plaschka, Martina Rath, Daniel Reumann, Sarah Rieser, Marianne Rocha-Hasler, Alan Rodriguez, Nathalie Ropek, James Julian Ross, Harald Scheuch, Karina Schindler, Clara Schmidt, Hannes Schmidt, Jakob Schnabl, Stefan Schüchner, Tanja Schwickert, Andreas Sommer, Daniele Soldoroni, Johannes Stadlmann, Peter Steinlein, Marcus Strobl, Simon Strobl, Qiong Sun, Wen Tang, Linda Trübestein, Johanna Trupke, Christian Umkehrer, Sandor Urmosi- Incze, Gijs Versteeg, Vivien Vogt, Michael Wagner, Martina Weissenboeck, Barbara Werner, Johannes Zuber, Manuela Födinger, Franz Allerberger, Alexander Stark, Luisa Cochella, Ulrich Elling 和 VCDI。《高通量SARSeq多重检测SARS-CoV-2及其他呼吸道感染》，《自然·通讯》，12(1):3132，2021年。[98] Vien Thi Minh Le 和 Binh An Diep。《全基因组测序在暴发调查中的应用：精选见解》。《重症医学观点》，2013年。[99] Vlad Nikolayevskyy、Katharina Kranzer、Stefan Niemann 和 Francis Drobniewski。《利用结核分枝杆菌全基因组测序检测近期传播并追踪暴发：系统综述》。《结核病》，2016年。[100] Shaofu Qiu、Peng Li、Hongbo Liu、Yong Wang、Nan Liu、Chengyi Li、Shenlong Li、Ming Li、Zhengjie Jiang、Huandong Sun、Ying Li、Jing Xie、Chaojie Yang、Jian Wang、Hao Li、Shengjie Yi、Zhihao Wu、Leili Jia、Ligui Wang、Rongzhang Hao、Yansong Sun、Liuyu Huang、Hui Ma、Zhengquan Yuan 和 Hongbin Song。《利用全基因组测序追踪中国由新型HAdV 7型血清型变异株引起的两次ARD暴发之间的传播关联》。《科学报告》，5(1):13617，2015年。[101] Carol A Gilchrist、Stephen D Turner、Margaret F Riley、William A Petri 和 Erik L Hewlett。《全基因组测序在暴发分析中的应用》。《临床微生物学评论》，2015年。[102] Minoru Kanehisa。《探寻细胞生物的起源与进化》。《蛋白质科学》，2019年。[103] Jun Qing、Yi-De Meng、Feng He、Qing-Xin Du、Jian Zhong、Hong-Yan Du、Pan-Feng Liu、Lan-Ying Du 和 Lu Wang。《全基因组重测序揭示杜仲的遗传多样性与进化模式》。《分子遗传学与基因组学》，2022年。[104] Patricia J. Wittkopp 和 Gizem Kalay。《顺式调控元件：分化背后的分子机制与进化过程》。《自然综述：遗传学》，2012年。[105] Irene Gallego Romero、Ilya Ruvinsky 和 Yoav Gilad。《基因表达与基因调控进化的比较研究》。《自然综述：遗传学》，2012年。[106] Xinchao Wang、Hu Feng、Yuxiao Chang、Chunlei Ma、Liyuan Wang、Xinyuan Hao、A' lun Li、Hao Cheng、Lu Wang、Peng Cui、Jiqiang Jin、Xiaobo Wang、Kang Wei、Cheng Ai、Sheng Zhao、Zhichao Wu、Youyong Li、Benying Liu、Guo-Dong Wang、Liang Chen、Jue Ruan 和 Yajun Yang。《群体测序增进对茶树进化的认识》。《自然·通讯》，2020年。[107] Mark S. Hill、Pétra Vande Zande 和 Patricia J. Wittkopp。《产生基因表达变异的分子与进化过程》。《自然综述：遗传学》，2021年。[108] Eeshit Dhaval Vaishnav、Carl G. de Boer、Jennifer Molinet、Moran Yassour、Lin Fan、Xian Adiconis、Dawn A. Thompson、Joshua Z. Levin、Francisco A. Cubillos 和 Aviv Regev。《基因调控DNA的进化、可进化性与工程化》。《自然》，2022年。[109] Xingtan Zhang、Shuai Chen、Longqing Shi、Daping Gong、Shengcheng Zhang、Qian Zhao、Dongliang Zhan、Liette Vasseur、Yibin Wang、Jiaxin Yu、Zhenyang Liao、Xindan Xu、Rui Qi、Wenling Wang、Yunran Ma、Pengjie Wang、Naixing Ye、Dongna Ma、Yan Shi、Haifeng Wang、Xiaokai Ma、Xiangrui Kong、Jing Lin、Liufeng Wei、Yaying Ma、Ruoyu Li、Guiping Hu、Haifang He、Lin Zhang、Ray Ming、Gang Wang、Haibao Tang 和 Minsheng You。《单倍型分辨基因组组装揭示茶树（Camellia sinensis）的进化史》。《自然·遗传学》，2021年。[110] Justin C. Fay 和 Patricia J. Wittkopp。《评估自然选择在基因调控进化中的作用》。《遗传》，2008年。[111] Aquillah M. Kanzi、James Emmanuel San、Benjamin Chimukangara、Eduan Wilkin- son、Maryam Fish、Veron Ramsuran 和 Tulio de Oliveira。《新一代测序与生物信息学分析在家族遗传中的应用》。《遗传学前沿》，2020年。[112] Gregory A. Wray、Matthew W. Hahn、Ehab Abouheif、James P. Balhoff、Margaret Pizer、Matthew V. Rockman 和 Laura A. Romano。《真核生物转录调控的进化》。《分子生物学与进化》，2003年。[113] Aiping Wu、Lulan Wang、Hang-Yu Zhou、Cheng-Yang Ji、Shang Zhou Xia、Yang Cao、Jing Meng、Xiao Ding、Sarah Gold、Taijiao Jiang 和 Genhong Cheng。《SARS-CoV-2进化一周年》。《细胞·宿主与微生物》，2021年。[114] Sarah A. Signor 和 Sergey V. Nuzhdin。《顺式与反式基因表达的进化》。《*Trends in Genetics*》，2018年。[115] Andrew Whitehead 和 Douglas L. Crawford。《物种内及物种间的基因表达变异：进化的原材料》。《分子生态学》，2006年。[116] Joseph D. Coolon、C. Joel McManus、Kraig R. Stevenson、Brenton R. Graveley 和 Patricia J. Wittkopp。《果蝇调控进化的速度与模式》。

《基因组研究》，2014年。[117] Hans Ellegren。《非模式生物的基因组测序与群体基因组学》。《生态学与进化趋势》，2014年。[118] Javier Prado-Martinez、Peter H. Sudmant、Jeffrey M. Kidd、Heng Li、Joanna L. Kelley、Belen Lorente-Galdos、Krishna R. Veeramah、August E. Woerner、Timothy D. O'Connor、Gabriel Santpere、Alexander Cagan、Christoph Theunert、Ferran Casals、Hafid Laayouni、Kasper Munch、Asger Hobolth、Anders E. Halager、Maika Malig、Jessica Hernandez-Rodriguez、Irene Hernando-Herraez、Kay Prüfer、Marc Pybus、Laurel Johnstone、Michael Lachmann、Can Alkan、Dorina Twigg、Natalia Petit、Carl Baker、Fereydoun Hormozdiari、Marcos Fernandez-Callejo、Marc Dabad、Michael L. Wilson、Laurie Stevison、Cristina Camprubí、Tiago Carvalho、Aurora Ruiz-Herrera、Laura Vives、Marta Mele、Teresa Abello、Ivanela Kondova、Ronald E. Bontrop、Anne Pusey、Felix Lankester、John A. Kiyang、Richard A. Bergl、Elizabeth Lonsdorf、Simon Myers、Mario Ventura、Pascal Gagneux、David Comas、Hans Siegismund、Julie Blanc、Lidia Agueda-Calpena、Marta Gut、Lucinda Fulton、Sarah A. Tishkoff、James C. Mullikin、Richard K. Wilson、Ivo G. Gut、Mary Katherine Gonder、Oliver A. Ryder、Beatrice H. Hahn、Arcadi Navarro、Joshua M. Akey、Jaume Bertranpetit、David Reich、Thomas Mailund、Mikkel H. Schierup、Christina Hvilsom、Aida M. Andrés、Jeffrey D. Wall、Carlos D. Bustamante、Michael F. Hammer、Evan E. Eichler 和 Tomas Marques-Bonet。《类人猿遗传多样性与群体历史》。《自然》，2013年。[119] Ana Prohaska、Fernando Racimo、Andrew J Schork、Martin Sikora、Aaron J Stern、Melissa Ilardo、Morten Erik Allentoft、Lasse Folkersen、Alfonso Buil、J Víc- tor Moreno-Mayar、Thorfinn Korneliussen、Daniel Geschwind、Andrés Ingason、Thomas Werge、Rasmus Nielsen 和 Eske Willerslev。《群体基因组学视角下的人类疾病变异》。《细胞》，2019年。[120] 拟南芥基因组计划。《开花植物拟南芥基因组序列分析》。《自然》，2000年。[121] Haocheng Zhu、Chao Li 和 Caixia Gao。《CRISPR–Cas在农业与植物生物技术中的应用》。《自然综述：分子细胞生物学》，2020年。[122] Jae Young Choi、Zoe N. Lye、Simon C. Groen、Xiaoguang Dai、Priyesh Rughani、Sophie Zaaijer、Eoghan D. Harrington、Sissel Juul 和 Michael D. Purugganan。《基于纳米孔测序的环巴斯马蒂稻基因组组装与进化基因组学》。《基因组生物学》，2020年。[123] Kristian A Stevens、Jill L Wegrzyn、Aleksey Zimin、Daniela Puiu、Marc Crepeau、Charis Cardeno、Robin Paul、Daniel Gonzalez-Ibeas、Maxim Koriabine、Ann E Holtz- Morris、Pedro J Martínez-García、Uzay U Sezen、Guillaume Marçais、Kathy Jermstad、Patrick E McGuire、Carol A Loopstra、John M Davis、Andrew Eckert、Pieter de Jong、James A Yorke、Steven L Salzberg、David B Neale 和 Charles H Langley。《糖松巨型基因组序列》。《遗传学》，2016年。[124] Maria Doroteia Campos、Maria do Rosário Félix、Mariana Patanita、Patrick Mater- atski 和 Carla Varanda。《高通量测序揭示番茄—病原体互作，助力可持续植物育种》。《园艺研究》，2021年。[125] Caixia Gao。《用于作物改良与未来农业的基因组工程》。《细胞》，2021年。[126] Aalt Dirk Jan van Dijk、Gert Kootstra、Willem Kruijer 和 Dick de Ridder。《机器学习在植物科学与植物育种中的应用》。《iScience》，2021年。[127] Yanqing Sun、Lianguang Shang、Qian-Hao Zhu、Longjiang Fan 和 Longbiao Guo。《植物基因组测序二十年：成就与挑战》。《植物科学趋势》，2022年。[128] Kyung D. Kim、Yuna Kang 和 Changsoo Kim。《基因组大数据在植物育种中的应用：过去、现在与未来》。《植物》，2020年。[129] Mahendar Thudi、Ramesh Palakurthi、James C. Schnable、Annapurna Chitikineni、Susanne Dreisigacker、Emma Mace、Rakesh K. Srivastava、C. Tara Satyavathi、Damaris Odeny、Vijay K. Tiwari、Hon-Ming La m、Yan Bin Hong、Vikas K. Singh、Guowei Li、Yunbi Xu、Xiaoping Chen、Sanjay Kaila、Henry Nguyen、Sobhana Sivasankar、Scott A. Jackson、Timothy J. Close、Wan Shubo 和 Rajeev K. Varshney。《可持续农业植物育种中的基因组资源》。《植物生理学杂志》，2021年。[130] Todd P Michael 和 Robert VanBuren。《构建近完整植物基因组》。《植物生物学当前观点》，2020年。[131] Yanting Shen、Guoan Zhou、Chengzhi Liang 和 Zhixi Tian。《基于组学的跨学科研究正在加速植物育种》。《植物生物学当前观点》，2022年。[132] Taha Shahroodi、Mahdi Zahedi、Can Firtina、Mohammed Alser、Stephan Wong、Onur Mutlu 和 Said Hamdioui。《Demeter：一种利用存内超维计算实现快速节能食品分析器》。《IEEE Access》，10:82493–82510，2022年。[133] Shiv Prasad、Lal Chand Malav、Jairam Choudhary、Sudha Kannojiya、Monika Kundu、Sandeep Kumar 和 Ajar Nath Yadav。《健康养分循环的土壤微生物组》，第1–21页。Springer Singapore，新加坡，2021年。[134] Martin Mascher、Murukarthick Jayakodi、Hyeonah Shim 和 Nils Stein。《作物转化基因组学的前景与挑战》。《自然》，2024年。[135] Mona Schreiber、Murukarthick Jayakodi、Nils Stein 和 Martin Mascher。《用于作物改良、生物多样性与进化的植物泛基因组》。《自然综述：遗传学》，2024年。[136] Frederick Sanger、Steven Nicklen 和 Alan R Coulson。《使用链终止抑制剂进行DNA测序》。《美国国家科学院院刊》，74(12):5463–5467，1977年。[137] Max Doblas、Po Jui S hih、Oscar Lostes-Cazorla、Miquel Moreto、Christopher Batten 和 Santiago Marco-Sola。《SMX：用于通用序列比对加速的异构架构》。载于 MICRO，2025年。[138] Onur Mutlu 和 Can Firtina。《通过算法—架构协同设计加速基因组分析》。载于 DAC，2023年。[139] Mohammed Alser

，Joel Lindegger ，Can Firtina ，Nour Almadhoun ，Haiyu Mao ，

Gagandeep Singh、Juan Gomez-Luna 和 Onur Mutlu。《从分子到基因组变异：通过智能算法和架构加速基因组分析》。《计算与结构生物技术期刊》，2022年。[140] Qian Lou、Sarath Chandra Janga 和 Lei Jiang。《Helix：用于加速纳米孔基因组碱基识别的算法/架构协同设计》。收录于 PACT，第293–304页，2020年。[141] Qian Lou 和 Lei Jiang。《Brawl：面向纳米孔基因组测序的、基于自旋电子学的便携式存内碱基识别架构》。《计算机体系结构快报》，2018年。[142] Taha Shahroodi、Gagandeep Si ngh、Mahdi Zahedi、Haiyu Mao、Joel Lindegger、Can Firtina、Stephan Wong、Onur Mutlu 和 Said Hamdioui。《Swordfish：利用非理想忆阻器和存内计算评估基于深度神经网络的碱基识别的框架》。收录于 MICRO，第1437–1452页，2023年。[143] Ryan Marcus、Andreas Kipf、Alexander van Renen、Mihail Stoian、Sanchit Misra、Alfons Kemper、Thomas Neumann 和 Tim Kraska。《学习型索引基准测试》。《VLDB基金会会刊》，14(1)：1–13，2020年。[144] Arun Subramaniyan、Jack Wadden、Kush Goliya、Nathan Ozog、Xiao Wu、Satish Narayanasamy、David Blaauw 和 Reetuparna Das。《利用枚举基数树加速基因组序列比对的种子生成》。收录于 ISCA，第388–401页，2021年。[145] Wenqin Huangfu、Shuangchen Li、Xing Hu 和 Yuan Xie。《RADAR：一种基于三维ReRAM的DNA比对加速器架构》。收录于 DAC，2018年。[146] S Karen Khatamifard、Zamshed Chowdhury、Nakul Pande、Meisam Razaviyayn、Chris H Kim 和 Ulya R Karpuzcu。《GeNVoM：非易失性存储器附近的读段比对》。《TCBB》，2021年。[147] Saransh Gupta、Mohsen Imani、Behnam Khaleghi、Venkatesh Kumar 和 Tajana Rosing。《RAPID：用于DNA序列比对的ReRAM存内处理架构》。收录于 ISLPED，2019年。[148] Xue-Qi Li、Guang-Ming Tan 和 Ning-Hui Sun。《PIM-Align：一种FM索引搜索算法的存内处理架构》。《计算机科学与技术学报》，2021年。[149] Shaahin Angizi、Jiao Sun、Wei Zhang 和 Deliang Fan。《AlignS：利用SOT-MRAM的DNA短读段比对存内加速器》。收录于 DAC，2019年。[150] Farzaneh Zokaee、Hamid R Zarandi 和 Lei Jiang。《AligneR：一种用于ReRAM短读段比对的存内处理架构》。《IEEE计算机体系结构快报》，2018年。[151] Yatish Turakhia、Gill Bejerano 和 William J Dally。《Darwin：一种基因组学协处理器，可将长读段组装加速高达15,000倍》。收录于 ASPLOS，2018年。[152] Daichi Fujiki、Arun Subramaniyan、Tianjun Zhang、Yu Zeng、Reetuparna Das、David Blaauw 和 Satish Narayanasamy。《GenAx：基因组测序加速器》。收录于 ISCA，2018年。[153] Advait Madhavan、Timothy Sherwood 和 Dmitri Strukov。《Race Logic：动态规划算法的硬件加速》。ACM SIGARCH《计算机体系结构新闻》，2014年。[154] Haoyu Cheng、Yong Zhang 和 Yun Xu。《BitMapper2：基于稀疏Q-gram索引的GPU加速全映射器》。《IEEE/ACM TCBB》，2018年。[155] Ernst Joachim Houtgast、Vlad-Mihai Sima、Koen Bertels 和 Zaid Al-Ars。《面向更长读段长度的BWA-MEM基因组短读段定位硬件加速》。《计算生物学与化学》，2018年。[156] Ernst Joachim Houtgast、VladMihai Sima、Koen Bertels 和 Zaid AlArs。《一种高效的GPU加速BWA-MEM基因组短读段定位实现》。《ACM SIGARCH计算机体系结构新闻》，2017年。[157] Alberto Zeni、Giulia Guidi、Marquita Ellis、Nan Ding、Marco D Santambrogio、Steven Hofmeyr、Aydın Buluç、Leonid Oliker 和 Katherine Yelick。《LOGAN：高性能GPU长读段X-drop比对》。收录于 IPDPS，2020年。[158] Nauman Ahmed、Jonathan Lévy、Shanshan Ren、Hamid Mushtaq、Koen Bertels 和 Zaid Al-Ars。《GASAL2：面向高通量NGS数据的GPU加速序列比对库》。《BMC Bioinformatics》，2019年。[159] Takahiro Nishimura、Jacir L Bordim、Yasuaki It o 和 Koji Nakano。《在GPU上使用按位并行批量计算技术加速Smith-Waterman算法》。收录于 IPDPSW，2017年。[160] Edans Flavius de Oliveira Sandes、Guillermo Miranda、Xavier Martorell、Eduard Ayguade、George Teodoro 和 Alba Cristina Magalhaes Melo。《CUDAlign 4.0：GPU集群上用于精确全染色体比对的增量式推测回溯》。《IEEE TPDS》，2016年。[161] Yongchao Liu 和 Bertil Schmidt。《GSWABE：通过最优比对检索加速GPU序列比对，适用于短DNA序列》。《Concurrency and Computation: Practice and Experience》，2015年。[162] Yongchao Liu、Adrianto Wirawan 和 Bertil Schmidt。《CUDASW++ 3.0：通过结合中央处理器与GPU的SIMD指令，加速Smith-Waterman蛋白质数据库搜索》。《BMC Bioinformatics》，2013年。[163] Yongchao Liu、Douglas L Maskell 和 Bertil Schmidt。《CUDASW++：针对支持CUDA的图形处理器优化Smith-Waterman序列数据库搜索》。《BMC Research Notes》，2009年。[164] Yongchao Liu、Bertil Schmidt 和 Douglas L Maskell。《CUDASW++ 2.0：基于SIMT和虚拟化SIMD抽象，在支持CUDA的GPU上增强Smith-Waterman蛋白质数据库搜索》。《BMC Research Notes》，2010年。[165] Richard Wilton、Tamas Budavari、Ben Langmead、Sarah J Wheelan、Steven L Salzberg 和 Alexander S Szalay。《Arioc：通过GPU加速探索种子-延伸搜索空间，实现高通量读段比对》。《PeerJ》，2015年。[166] Amit Goyal、Hyuk Jung Kwon、Kichan Lee、Reena Garg、Seon Young Yun、Yoon Hee Kim、Sunghoon Lee 和 Min Seob Lee。《超快速下一代人类基因组测序数据处理：用于精准医学的DRAGENTM生物信息技术处理器》

《开放遗传学杂志》，2017年。[167] Yu-Ting Chen、Jason Cong、Zhenman Fang、Jie Lei 和 Peng Wei。《当Spark遇上FPGA：下一代DNA测序加速案例研究》。收录于 USENIX HotCloud，2016年。[168] Peng Chen、Chao Wang、Xi Li 和 Xuehai Zhou。《利用基于FPGA的系统加速下一代长读段比对》。《TCBB》，2014年。[169] Yen-Lung Chen、Bo-Yi Chang、Chia-Hsiang Yang 和 Tzi-Dar Chiueh。《用于全人类基因组短读段定位的高吞吐量FPGA加速器》。《IEEE TPDS》，2021年。[170] Daichi Fujiki、Shunhao Wu、Nathan Ozog、Kush Goliya、David Blaauw、Satish Narayanasamy 和 Reetuparna Das。《SeedEx：在次最小空间内实现最优比对的基因组测序加速器》。收录于 MICRO，2020年。[171] Subho Sankar Banerjee、Mohamed El-Hadedy、Jong Bin Lim、Zbigniew T Kalbarczyk、Deming Chen、Steven S Lumetta 和 Ravishankar K Iyer。《ASAP：可编程硬件上的加速短读段比对》。《IEEE计算机汇刊》，2019年。[172] Xia Fei、Zou Dan、Lu Lina、Man Xin 和 Zhang Chunlei。《FPGASW：在FPGA线性脉动阵列上通过回溯加速大规模Smith-Waterman序列比对应用》。《跨学科科学：计算生命科学》，2018年。[173] Hasitha Muthumala Waidyasooriya 和 Masanori Hariyama。《基于Burrows-Wheeler变换的短读段比对硬件加速》。《TPDS》，2015年。[174] Yu-Ting Chen、Jason Cong、Jie Lei 和 Peng Wei。《一种新型高吞吐量读段比对加速引擎》。收录于 FCCM，2015年。[175] Enzo Rucci、Carlos Garcia、Guillermo Botella、Armando De Giusti、Marcelo Naiouf 和 Manuel Prieto-Matias。《SWIFOLD：利用OpenCL在FPGA上实现长DNA序列的Smith-Waterman比对》。《BMC系统生物学》，2018年。[176] Abbas Haghi、Santiago Marco-Sola、Lluc Alvarez、Dionysios Diamantopoulos、Christoph Hagleitner 和 Miquel Moreto。《用于基因组学两两比对的波前算法FPGA加速器》。收录于 FPL，2021年。[177] Luyi Li、Jun Lin 和 Zhongfeng Wang。《PipeBSW：FPGA上带状Smith-Waterman算法的两阶段流水线结构》。收录于 ISVLSI，2021年。[178] Tae Jun Ham、David Bruns-Smith、Brendan Sweeney、Yejin Lee、Seong Hoon Seo、U Gyeong Song、Young H Oh、Krste Asanovic、Jae W Lee 和 Lisa Wu Wills。《Genesis：基因组数据分析硬件加速框架》。收录于 ISCA，2020年。[179] Tae Jun Ham、Yejin Lee、Seong Hoon Seo、U Gyeong Song、Jae W Lee、David Bruns-Smith、Brendan Sweeney、Krste Asanovic、Young H Oh 和 Lisa Wu Wills。《利用可组合硬件加速框架加速基因组数据分析》。《IEEE Micro》，2021年。[180] Lisa Wu、David Bruns-Smith、Frank A. Nothaft、Qijing Huang、Sagar Karandikar、Johnny Le、Andrew Lin、Howard Mao、Brendan Sweeney、Krste Asanović、David A. Patterson 和 Anthony D. Joseph。《云端FPGA加速插入缺失重比对》。收录于 HPCA，2019年。[181] Damla Senol Cali、Gurpreet S. Kalsi、Zülal Bingöl、Can Firtina、Lavanya Subramanian、Jeremie S. Kim、Rachata Ausavarungnirun、Mohammed Alser、Juan Gomez-Luna、Amirali Boroumand、Anant Norion、Allison Scibisz、Sreenivas Subramoneyon、Can Alkan、Saugata Ghose 和 Onur Mutlu。《GenASM：用于基因组序列分析的高性能、低功耗近似字符串匹配加速框架》。收录于 MICRO，2020年。[182] Fan Zhang、Shaahin Angizi、Jiao Sun、Wei Zhang 和 Deliang Fan。《Aligner-D：利用DRAM内计算加速DNA短读段比对》。《IEEE电路与系统新兴及重点主题期刊》，2023年。[183] Melina Soysal、Konstantina Koliogeorgi、Can Firtina、Nika Mansouri Ghiasi、Rakesh Nadig、Haiyu Mao、Geraldo Francisco de Oliveira Junior、Yu Liang、Klea Zambaku、Mohammad Sadrosadati 等。《MARS：存储子系统内的原始信号基因组分析存内加速》。收录于 ICS，2025年。[184] Jeremie S Kim、Damla Senol Cali、Hongyi Xin、Donghyuk Lee、Saugata Ghose、Mohammed Alser、Hasan Hassan、Oguz Ergin、Can Alkan 和 Onur Mutlu。《GRIM-Filter：利用存内处理技术加速DNA读段定位中的种子位置过滤》。《BMC Genomics》，2018年。[185] Roman Kaplan、Leonid Yavits 和 Ran Ginosasr。《BioSEAL：面向大规模基因组数据的内存内生物序列比对加速器》。收录于 SYSTOR，2020年。[186] Haiyu Mao、Mohammed Alser、Mohammad Sadrosadati、Can Firtina、Akanksha Baranwal、Damla Senol Cali、Aditya Manglik、Nour Almadhoun Alserr 和 Onur Mutlu。《GenPIP：通过紧密集成碱基识别与读段比对实现基因组分析存内加速》。收录于 MICRO，第710–726页，2022年。[187] Yingqi Cao、Anshu Guta、Jason Liang 和 Yatish Turakhia。《DP-HLS：用于加速生物信息学动态规划算法的高级综合框架》。收录于 HPCA，2026年。[188] Zhehong Wang、Tianjun Zhang、Daichi Fujiki、Arun Subramaniyan、Xiao Wu、Makoto Yasuda、Satoru Miyoshi、Masaru Kawaminami、Reetuparna Das、Satish Narayanasamy 和 David Blaauw。《采用字符串无关PE阵列、面向下一代测序的种子扩展加速器，吞吐量达每秒246万读段》。《IEEE JSSC》，2020年。[189] Sumit Walia、Cheng Ye、Arkid Bera、Dhruvi Lodhavia 和 Yatish Turakhia。《TALCO：利用回溯指针收敛进行平铺基因组序列比对》。收录于 HPCA，第91–107页，2024年。[190] Harisankar Sadasivan、Artur Klauser、Juergen Hench、Yatish Turakhia、Gagandeep Singh、Alberto Zeni、Sarah Beecroft、Satish Narayanasamy、Jeff Nivala、Bob Robey、Onur Mutlu、Kristof Denolf 和 Sriranjani Sitaraman。《基因组计算革命：定义未来数十年的基因组学加速方向》。收录于 HPEC

[191] Yatish Turakhia。《面向基因组序列分析的通用加速器》。《ACM通信》，2025年。[192] Yatish Turakhia、Sneha D. Goenka、Gill Bejerano 和 WIlliam J. Dally。《Darwin-WGA：一种通过协处理器提升全基因组比对灵敏度并实现高倍加速的方法》。载于 HPCA，第359–372页，2019年。[193] William Andrew Simon、Leonid Yavits、Konstantina Koliogeorgi、Yann Falevoz、Yoshihiro Shibuya、Dominique Lavenier、Irem Boybat、Klea Zambaku、Berkan Şahin、Mohammad Sadrosadati、Onur Mutlu、Abu Sebastian、Rayan Chikhi、BioPIM联盟和 Can Alkan。《面向基因组学工作负载的存内处理》。《IEEE Micro》，2026年。[194] Peng Jia、Liming Xuan、Lei Liu 和 Chaochun Wei。MetaBinG：利用GPU加速宏基因组序列分类。《PLOS ONE》，2011。[195] Robin Kobus、André Müller、Daniel Jünger、Christian Hundt 和 Bertil Schmidt。MetaCache-GPU：超快速宏基因组分类。载于 ICPP，2021年。[196] Xuebin Wang、Taifu Wang、Zhihao Xie、Youjin Zhang、Shiqiang Xia、Ruixue Sun、Xinqiu He、Ruizhi Xiang、Qiwen Zheng、Zhencheng Liu、Jin'An Wang、Honglong Wu、Xiangqian Jin、Weijun Chen、Dongfang Li 和 Zengquan He。GPMeta：一种从宏基因组序列中超快速鉴定病原体的GPU加速方法。《生物信息学简报》，24(2):bbad092，2023年。[197] Robin Kobus、Christian Hundt、André Müller 和 Bertil Schmidt。在支持CUDA的GPU上加速宏基因组读段分类。《BMC Bioinformatics》，2017年。[198] Xiaoquan Su、Jian Xu 和 Kang Ning。Parallel-META：基于高性能计算的高效宏基因组数据分析。《BMC系统生物学》，2012年。[199] Xiaoquan Su、Xuetao Wang、Gongchao Jing 和 Kang Ning。GPU-Meta-Storms：使用GPU计算大量微生物群落样本之间的结构相似性。《生物信息学》，30(7):1031–1033，2014年。[200] Masahiro Yano、Hiroshi Mori、Yutaka Akiyama、Takuji Yamada 和 Ken Kurokawa。CLAST：基于CUDA实现的大规模比对搜索工具。《BMC生物信息学》，15:1–13，2014年。[201] Antonio Saavedra、Hans Lehnert、Cecilia Hernández、Gonzalo Carvajal 和 Miguel Figueroa。使用草图和硬件加速，在DNA序列中挖掘判别性k-mer。《IEEE Access》，8:114715–114732，2020年。[202] Tianqi Zhang、Antonio González、Niema Moshiri、Rob Knight 和 Tajana Rosing。GenoMiX：加速人类基因组学、微生物组宏基因组学和病毒序列的同步分析。载于 BioCAS，第1–5页，2023年。[203] Gustavo Henrique Cervi、Cecília Dias Flores 和 Claudia Elizabeth Thompson。宏基因组分析：利用高性能计算提高效率的途径。载于 ICICT，2022年。[204] Lingxi Wu、Rasool Sharifi、Marzieh Lenjani、Kevin Skadron 和 Ashish Venkat。Sieve：面向大规模并行k-mer匹配的可扩展原位DRAM加速器设计。载于 ISCA，第251–264页，2021年。[205] Taha Shahroodi、Mahdi Zahedi、Abhairaj Singh、Stephan Wong 和 Said Hamdioui。KrakenOnMem：用于分类学谱系分析的忆阻器增强型软硬件框架。载于 ICS，第1–14页，2022年。[206] Zuher Jahshan、Itay Merlin、Esteban Garzón 和 Leonid Yavits。DASH-CAM：用于基因组分类的动态近似搜索内容寻址存储器。MICRO，2023年。[207] Robert Hanhan、Esteban Garzón、Zuher Jahshan、Adam Teman、Marco Lanuzza 和 Leonid Yavits。EDAM：可容忍编辑距离的近似匹配内容寻址存储器。载于 ISCA，第495–507页，2022年。[208] Zhuowen Zou、Hanning Chen、Prathyush Poduval、Yeseong Kim、Mahdi Imani、Elaheh Sadredini、Rosario Cammarota 和 Mohsen Imani。BioHD：一种利用超维记忆实现的高效基因组序列搜索平台。载于 ISCA，第656–669页，2022年。[209] Po Jui Shih、Hassaan Saadat、Sri Parameswaran 和 Hasindu Gamaarachchi。在资源受限设备上高效实时地进行选择性基因组测序。GigaScience，12:giad046，2023年。[210] Damla Senol Cali、Konstantinos Kanellopoulos、Joël Lindegger、Zülal Bingöl、Gurpreet S. Kalsi、Ziyi Zuo、Can Firtina、Meryem Banu Cavlak、Jeremie Kim、Nika Mansouri Ghiasi、Gagandeep Singh、Juan Gómez-Luna、Nour Almadhoun Alserr、Mohammed Alser、Sreenivas Subramoney、Can Alkan、Saugata Ghose 和 Onur Mutlu。SeGraM：用于基因组序列到图和序列到序列比对的通用硬件加速器。载于 ISCA，第638–655页，2022年。[211] Yichi Zhang、Dibei Chen、Gang Zeng、Jianfeng Zhu、Zhaoshi Li、Longlong Chen、Shaojun Wei 和 Leibo Liu。Harp：利用准序列特征加速长读段的序列到图比对。载于 ASPLOS，第512–527页，2024年。[212] Gang Zeng、Jianfeng Zhu、Yichi Zhang、Ganhui Chen、Zhenhai Yuan、Shaojun Wei 和 Leibo Liu。一种利用动态规划算法实现精确序列到图对齐的高性能基因组加速器。《IEEE并行与分布式系统汇刊》，35(2):237–249，2024年。[213] Zhe-Wei Shen、Jheng-Syun Huang 和 Yi-Chang Lu。一种面向变异富集区域中128路并行序列到图对齐的节省内存加速器。载于 BioCAS，第1–5页，2024年。[214] Jiajie Li、Jan-Niklas Schmelzle、Yixiao Du、Simon Heumos、Andrea Guarracino、Giulia Guidi、Pjotr Prins、Erik Garrison 和 Zhiru Zhang。基于GPU的快速泛基因组图布局。载于 SC，第1–19页，2024年。[215] Kalikinkar Mandal、Bo Yang、Guang Gong 和 Mark Aagaard。一类复合德布鲁因序列的分析与高效实现。《IEEE计算机汇刊》，69(12):1835–1848，2020年。[216] B. Sharat Chandra Varma、Kolin Paul、M. Balakrishnan 和 Dominique Lavenier。FAssem：基于FPGA加速从头基因组组装。载于 FCCM，第

，2024年。

173–176，2013年。[217] Muaaz Gul Awan、Steven Hofmeyr、Rob Egan、Nan Ding、Aydin Buluc、Jack Deslippe、Leonid Oliker 和 Katherine Yelick。利用GPU加速大规模从头宏基因组组装。载于 SC，第1–11页，2021年。[218] Zonghao Feng 和 Qiong Luo。在异构处理器上加速序列到图对齐。载于 ICPP，第1–10页，2021年。[219] Yichi Zhang、Jianfeng Zhu、Liangwei Li、Gang Zeng、Dibei Chen、Tairan Zhang、Yeyang Deng、Zhicheng Gong、Aoyang Zhang、Yang Liu、Shaojun Wei 和 Leibo Liu。一款28纳米、239碱基对/μJ敏捷泛基因组分析加速器，支持多方案读段比对。《IEEE固态电路期刊》，第1–11页，2025年。[220] Heewoo Kim、Sanjay Sri Vallabh Singapuram、Haojie Ye、Joseph Izraelevitz、Trevor Mudge、Ronald Dreslinski 和 Nishil Talati。NMP-PaK：近存处理加速可扩展的从头基因组组装。载于 ISCA，2025年。[221] Yu Huang、Long Zheng、Haifeng Liu、Zhuoran Zhou、Dan Chen、Pengcheng Yao、Qinggang Wang、Xiaofei Liao 和 Hai Jin。MeG2：用于基因组图分析的存内加速。载于 DAC，第1–6页，2023年。[222] Shaahin Angizi、Naima Ahmed Fahmi、Wei Zhang 和 Deliang Fan。PANDA：基于德布鲁因图的DNA组装的MRAM内处理加速。arXiv:2008.06177，2020年。[223] Shuang Qiu 和 Qiong Luo。在异构处理器上并行构建大型德布鲁因图。载于 ICDCS，第1431–1441页，2017年。[224] Minxuan Zhou、Lingxi Wu、Muzhou Li、Niema Moshiri、Kevin Skadron 和 Tajana Rosing。通过近存计算实现超高效的从头基因组组装加速。载于 PACT，第199–212页，2021年。[225] Aritra Sarkar、Zaid Al-Ars 和 Koen Bertels。QuASeR：量子加速的从头DNA序列重建。《PLOS ONE》，16(4):e0249850，2021年。[226] B. Sharat Chandra Varma、Kolin Paul、M. Balakrishnan 和 Dominique Lavenier。从头基因组组装的硬件加速。《国际嵌入式系统期刊》，9(1):74，2017年。[227] B. Sharat Chandra Varma、Kolin Paul 和 M. Balakrishnan。基于FPGA加速从头基因组组装，第55–79页。Springer Singapore，2016年。[228] Sayan Goswami、Kisung Lee、Shayan Shams 和 Seung-Jong Park。GPU加速的大规模基因组组装。载于 IPDPS，第814–824页，2018年。[229] Georgios Galanos、Pavlos Malakonakis 和 Apostolos Dollas。一种用于加速从头基因组组装的基于FPGA的数据预处理架构。载于 BIBE，第1–6页，2021年。[230] Shaahin Angizi、Naima Ahmed Fahmi、Wei Zhang 和 Deliang Fan。PIM-Assembler：用于基因组组装的存内处理平台。载于 DAC，第1–6页，2020年。[231] Aman Sinha、Huei-Chun Yang、Pei-Yi Liu、Yen-Shi Kuo、Yuhao Fang、Tien-Shuo Chang 和 Ke-Han Li、Bo-Cheng Lai。DSIM：面向基因组组装的近DRAM加速器上的分布式序列匹配。《IEEE电路与系统新兴及重点主题期刊》，12(2):486–499，2022年。[232] Pingfan Meng、Matthew Jacobsen、Motoki Kimura、Vladimir Dergachev、Thomas Anantharaman、Michael Requa 和 Ryan Kastner。面向大规模基因组的新型光学从头组装硬件加速。载于 FPL，第1–8页，2014年。[233] Yuanqi Hu 和 Pantelis Georgiou。一种基于FPGA实现的实时从头DNA测序组装平台。《IEEE/ACM计算生物学与生物信息学汇刊》，13(2):291–300，2016年。[234] Yibo Chen、Jun-Han Huang、Yuhui Sun、Yong Zhang、Yuxiang Li 和 Xun Xu。VRP组装器：利用量子计算对二倍体和多倍体基因组进行单倍型解析的从头组装。Cell Reports Methods，2023年。[235] Santhi Natarajan、N. KrishnaKumar、H. V. Anuchan、Debnath Pal 和 S. K. Nandy。ReneGENE-Novo：用于基因组短读段加速预处理与组装的协同设计算法与架构，第564–577页。施普林格国际出版，2018年。[236] Shanshan Ren、Nauman Ahmed、Koen Bertels 和 Zaid Al-Ars。一种用于微组装的高效GPU版德布鲁因图构建算法。载于 BIBE，第67–72页，2018年。[237] Gagandeep Singh、Mohammed Alser、Kristof Denolf、Can Firtina、Alireza Khodamoradi、Meryem Banu Cavlak、Henk Corporaal 和 Onur Mutlu。RUBICON：用于设计高效的基于深度学习的基因组碱基识别器的框架。《基因组生物学》，25(1):1–29，2024年。[238] Meryem Banu Cavlak、Gagandeep Singh、Mohammed Alser、Can Firtina、Joel Lindegger、Mohammad Sadrosadati、Nika Mansouri Ghiasi、Can Alkan 和 Onur Mutlu。TargetCall：通过碱基识别前过滤消除碱基识别中的无效计算。《遗传学前沿》，2024年。[239] William Andrew Simon、Irem Boybat、Riselda Kodra、Elena Ferro、Gagandeep Singh、Mohammed Alser、Shubham Jain、Hsinyu Tsai、Geoffrey W. Burr、Onur Mutlu 和 Abu Sebastian。CiMBA：通过存内计算实现设备端碱基识别，加速基因组测序。《IEEE并行与分布式系统汇刊》，36(6):1130–1145，2025年。[240] Can Firtina、Kamlesh Pillai、Gurpreet S. Kalsi、Bharathwaj Suresh、Damla Senol Cali、Jeremie S. Kim、Taha Shahroodi、Meryem Banu Cavlak、Joël Lindegger、Mohammed Alser、Juan Gómez Luna、Sreenivas Subramoney 和 Onur Mutlu。ApHMM：加速轮廓隐马尔可夫模型，实现快速、节能的基因组分析。《ACM架构与代码优化汇刊》，21 29，2024 1 1( ): – .[241] Can Firtina、Jeremie S Kim、Mohammed Alser、Damla Senol Cali、A Ercument Cicek、Can Alkan 和 Onur Mutlu。Apollo：一种独立于测序技术、可扩展且准确的组装校正算法。《生物信息学》，2020年。[242] Nika Mansouri Ghiasi、Mohammad Sadrosadati、Harun Mustafa、Arvid Gollwitzer、Can Firtina、Julien Eudine、Haiyu Mao、Joël Lindegger、Meryem Banu Cavlak、Mohammed Alser、Jisung Park 和 Onur Mutlu。MegIS：高性能、节能的

ISCA，2024年。[243] Benedict Paten、Adam M Novak、Jordan M Eizenga 和 Erik Garrison。基因组图与基因组推断的演化。《基因组研究》，27 676，2017年 5 665( ): – 。[244] Rachel M. Sherman 和 Steven L. Salzberg。人类基因组时代的泛基因组学。《自然综述：遗传学》，21(4):243–254，2020年。[245] Erik Garrison、Jouni Sirén、Adam M Novak、Glenn Hickey、Jordan M Eizenga、Eric T Dawson、William Jones、Shilpa Garg、Charles Markello、Michael F Lin 等。通过在参考基因组中表示遗传变异，变异图工具包改进了读段比对。《自然·生物技术》，36(9):875–879，2018年。 [246] Jordan M. Eizenga、Adam M. Novak、Jonas A. Sibbesen、Simon Heumos、Ali Ghaf- faari、Glenn Hickey、Xian Chang、Josiah D. Seaman、Robin Rounthwaite、Jana Ebler、Mikko Rautiainen、Shilpa Garg、Benedict Paten、Tobias Marschall、Jouni Sirén 和 Erik Garrison。泛基因组图。《基因组学与人类遗传学年度评论》，21(第21卷，2020年):139–162，2020年。 [247] Dylan J. Taylor、Jordan M. Eizenga、Qiuhui Li、Arun Das、Katharine M. Jenike、Eimear E. Kenny、Karen H. Miga、Jean Monlong、Rajiv C. McCoy、Benedict Paten 和 Michael C. Schatz。超越人类基因组计划：完整人类基因组序列与泛基因组参考的时代。《基因组学与人类遗传学年度评论》，2024年。 [248] Wen-Wei Liao、Mobin Asri、Jana Ebler、Daniel Doerr、Marina Haukness、Glenn Hickey、Shuangjia Lu、Julian K. Lucas、Jean Monlong、Haley J. Abel、Silvia Buon- aiuto、Xian H. Chang、Haoyu Cheng、Justin Chu、Vincenza Colonna、Jordan M. Eizenga、Xiaowen Feng、Christian Fischer、Robert S. Fulton、Shilpa Garg、Cristian Groza、Andrea Guarracino、William T. Harvey、Simon Heumos、Kerstin Howe、Miten Jain、Tsung-Yu Lu、Charles Markello、Fergal J. Martin、Matthew W. Mitchell、Kather- ine M. Munson、Moses Njagi Mwaniki、Adam M. Novak、Hugh E. Olsen、Trevor Pesout、David Porubsky、Pjotr Prins、Jonas A. Sibbesen、Jouni Sirén、Chad Tomlin- son、Flavia Villani、Mitchell R. Vollger、Lucinda L. Antonacci-Fulton、Gunjan Baid、Carl A. Baker、Anastasiya Belyaeva、Konstantinos Billis、Andrew Carroll、Pi-Chuan Chang、Sarah Cody、Daniel E. Cook、Robert M. Cook-Deegan、Omar E. Cornejo、Mark Diekhans、Peter Ebert、Susan Fairley、Olivier Fedrigo、Adam L. Felsenfeld、Giulio Formenti、Adam Frankish、Yan Gao、Nanibaa' A. Garrison、Carlos Garcia Giron、Richard E. Green、Leanne Haggerty、Kendra Hoekzema、Thibaut Hourlier、Hanlee P. Ji、Eimear E. Kenny、Barbara A. Koenig、Alexey Kolesnikov、Jan O. Korbel、Jennifer Kordosky、Sergey Koren、HoJoon Lee、Alexandra P. Lewis、Hugo Maga- lhães、Santiago Marco-Sola、Pierre Marijon、Ann McCartney、Jennifer McDaniel、Jacquelyn Mountcastle、Maria Nattestad、Sergey Nurk、Nathan D. Olson、Alice B. Popejoy、Daniela Puiu、Mikko Rautiainen、Allison A. Regier、Arang Rhie、Samuel Sacco、Ashley D. Sanders、Valerie A. Schneider、Baergen I. Schultz、Kishwar Shafin、Michael W. Smith、Heidi J. Sofia、Ahmad N. Abou Tayoun、Françoise Thibaud-Nissen、Francesca Floriana Tricomi、Justin Wagner、Brian Walenz、Jonathan M. D. Wood、Aleksey V. Zimin、Guillaume Bourque、Mark J. P. Chaisson、Paul Flicek、Adam M. Phillippy、Justin M. Zook、Evan E. Eichler、David Haussler、Ting Wang、Erich D. Jarvis、Karen H. Miga、Erik Garrison、Tobias Marschall、Ira M. Hall、Heng Li 和 Benedict Paten。人类泛基因组参考草图。《自然》，617(7960):312– 324，2023年。[249] Harun Mustafa、Mikhail Karasikov、Nika Mansouri Ghiasi、Gunnar Rätsch 和 André Kahles。带注释的德布鲁因图上的标签引导的种子-链-延伸比对。《生物信息学》，40(补充_1):i337–i346，2024年。[250] Joel Armstrong、Glenn Hickey、Mark Diekhans、Ian T. Fiddes、Adam M. Novak、Alden Deran、Qi Fang、Duo Xie、Shaohong Feng、Josefin Stiller、Diane Genereux、Jeremy Johnson、Voichita Dana Marinescu、Jessica Alföldi、Robert S. Harris、Kerstin Lindblad-Toh、David Haussler、Elinor Karlsson、Erich D. Jarvis、Guojie Zhang 和 Benedict Paten。Progressive Cactus 是千基因组时代的多基因组比对器。《自然》，587(7833):246–251，2020年。[251] Phelim Bradley、Henk C Den Bakker、Eduardo PC Rocha、Gil McVean 和 Zamin Iqbal。超快速搜索所有已提交的细菌和病毒基因组数据。《自然·生物技术》，2019年。[252] Chirag Jain、Haowen Zhang、Yu Gao 和 Srinivas Aluru。论序列到图比对的复杂性。《计算生物学杂志》，27 654，2020年 4 640( ): – 。[253] Mikko Rautiainen 和 Tobias Marschall。GraphAligner：快速且多用途的序列到图比对。《基因组生物学》，21(1):1–28，2020年。 [254] Glenn Hickey、David Heller、Jean Monlong、Jonas A. Sibbesen、Jouni Sirén、Jordan Eizenga、Eric T. Dawson、Erik Garrison、Adam M. Novak 和 Benedict Paten。使用 vg工具包在泛基因组图中对结构变异进行基因分型。《基因组生物学》，21(1)，2020年。 [255] Yingbo Cui、Chenchen Peng、Zeyu Xia、Canqun Yang 和 Yifei Guo。泛基因组时代序列到图映射算法综述。《基因组生物学》，26(1):138，2025年。[256] Jouni Sirén。变异图的索引构建。载于 *ALENEX*，第13–27页，2017年。[257] Jouni Sirén、Jean Monlong、Xian Chang、Adam M. Novak、Jordan M. Eizenga、Charles Markello、Jonas A. Sibbesen、Glenn Hickey、Pi-Chuan Chang、Andrew Carroll、Namrata Gupta、Stacey Gabriel、Thomas W. Blackwell、Aakrosh Ratan、Kent D. Taylor、Stephen S. Rich、Jerome I. Rotter、David Haussler、Erik Garrison 和 Benedict Paten。泛基因组学使5202个多样化基因组中的已知结构变异得以基因分型。《科学》，374(6574):abg8871，2021年。 [258] Heng Li、Xiaowen Feng 和 Chong Chu。使用minigraph设计和构建参考泛基因组图。《基因组生物学》，21(1)，2020年。 [259] Christopher Lee、Catherine Grasso 和 Mark F Sharlow。使用偏序图进行多序列比对。《生物信息学》，18(3):452–464，2002年。 [260] Santiago Marco-Sola、Juan Carlos Moure、Miquel Moreto 和 Antonio Espinosa。

’与采用存储内处理的低成本宏基因组分析。在

使用波前算法进行快速间隙仿射成对比对。《生物信息学》，2021年。[261] Mohammed Alser、Zülal Bingöl、Damla Senol Cali、Jeremie Kim、Saugata Ghose、Can Alkan 和 Onur Mutlu。加速基因组分析：正在进行的探索之旅简介。IEEE Micro，40(5):65–75，2020年。[262] Mohammed Alser、Hasan Hassan、Hongyi Xin、Oğuz Ergin、Onur Mutlu 和 Can Alkan。GateKeeper：用于加速DNA短读段比对预处理的新型硬件架构。《生物信息学》，2017年。[263] Wenqin Huangfu、Xueqi Li、Shuangchen Li、Xing Hu、Peng Gu 和 Yuan Xie。MEDAL：基于双列直插式内存模块的可扩展近数据处理DNA种子算法加速器。载于MICRO，第587–599页，2019年。[264] Anirban Nag、CN Ramachandra、Rajeev Balasubramonian、Ryan Stutsman、Edouard Giacomin、Hari Kambalasubramanyam 和 Pierre-Emmanuel Gaillardon。Gen- Cache：利用缓存内算子实现高效序列比对。载于MICRO，2019年。[265] Chirag Jain、Sanchit Misra、Haowen Zhang、Alexander Dilthey 和 Srinivas Aluru。加速序列到图的比对。载于IPDPS，第451–461页，2019年。[266] Saul B Needleman 和 Christian D Wunsch。适用于搜索两种蛋白质氨基酸序列相似性的一般方法。《分子生物学杂志》，1970年。[267] Zeyu Xia、Yingbo Cui、Ang Zhang、Tao Tang、Lin Peng、Chun Huang、Canqun Yang 和 Xiangke Liao。Smith-Waterman算法并行实现综述。《跨学科科学：计算生命科学》，14(1):1–14，2022年。[268] Zülal Bingöl、Berkan Şahin、Klea Zambaku、Ricardo Roman-Brenes、Konstantina Koliogeorgi、Can Firtina、Onur Mutlu 和 Can Alkan。泛基因组学的德布鲁因图：基于德布鲁因图工具进行读段比对的深度性能基准测试。《生物信息学简报》，27(4):bbag440，2026年。[269] Noah Kaplan、Jan-Niklas Schmelzle、Yufeng Gu、Erik Garrison、Christopher Batten 和 Reetuparna Das。PangenomicsBench：泛基因组学基准测试套件及其特征分析。载于IISWC，第272–285页，2025年。[270] Stephen F. Altschul、Warren Gish、Webb Miller、Eugene W. Myers 和 David J. Lipman。基本局部比对搜索工具。《分子生物学杂志》，215(3):403– 410，1990年。[271] Ben Langmead、Cole Trapnell、Mihai Pop 和 Steven L Salzberg。将短DNA序列超快速且节省内存地比对到人类基因组。《基因组生物学》，10(3)，2009年。[272] Michael Burrows。块排序无损数据压缩算法。SRS研究报告，124，1994年。[273] Lin Huang、Victoria Popic 和 Serafim Batzoglou。利用基因组群体进行短读段比对。《生物信息学》，29(13):i361–i370，2013年。[274] Jouni Sirén、Niko Välimäki 和 Veli Mäkinen。用于路径查询的图索引及其在基因组研究中的应用。IEEE/ACM计算生物学与生物信息学汇刊，11(2):375–388，2014年。[275] P. Ferragina 和 G. Manzini。具有应用的数据结构。载于FOCS，第390–398页，2000年。[276] Massimiliano Rossi、Marco Oliva、Ben Langmead、Travis Gagie 和 Christina Boucher。MONI：用于查找最大精确匹配的泛基因组索引。《计算生物学杂志》，29(2):169–187，2022年。[277] Jouni Sirén、Erik Garrison、Adam M Novak、Benedict Paten 和 Richard Durbin。单倍型感知图索引。《生物信息学》，36(2):400–407，2020年。[278] Korbinian Schneeberger、Jörg Hagmann、Stephan Ossowski、Norman Warthmann、Sandra Gesing、Oliver Kohlbacher 和 Detlef Weigel。同时将短读段比对到多个基因组。《基因组生物学》，2009年。[279] Mahdi Heydari、Giles Miclotte、Yves Van de Peer 和 Jan Fostier。BrownieAligner：将Illumina测序数据准确比对到德布鲁因图。《BMC Bioinformatics》，19(1):311，2018年。[280] Jun Ma、Manuel Cáceres、Leena Salmela、Veli Mäkinen 和 Alexandru I Tomescu。针对含错误长读段到无环变异图的准确比对进行链式连接。《生物信息学》，39(8):btad460，2023年。[281] Kristoffer Sahlin。Strobealign：灵活的种子大小实现超快速且准确的读段比对。《基因组生物学》，23(1):260，2022年。[282] Heng Li。使用全基因组从头组装探索单样本SNP和插入缺失识别。《生物信息学》，28(14):1838–1844，2012年。[283] Antoine Limasset、Bastien Cazaux、Eric Rivals 和 Pierre Peterlongo。德布鲁因图上的读段比对。《BMC Bioinformatics》，17(1):237，2016年。[284] Tatiana Dvorkina、Dmitry Antipov、Anton Korobeynikov 和 Sergey Nurk。SPAligner：将长而高度分化的分子序列比对到组装图。《BMC Bioinformatics》，21(增刊12):306，2020年。[285] Bo Liu、Hongzhe Guo、Michael Brudno 和 Yadong Wang。deBGA：基于德布鲁因图的种子与延伸读段比对。《生物信息学》，32 3232，2016年 21 3224( ): – 。[286] Heng Li。Minimap2：核苷酸序列的成对比对。《生物信息学》，2018年。[287] Ghanshya m Chandra 和 Chirag Jain。使用对间隙敏感的共线链式连接进行序列到图比对。载于RECOMB，第58–73页，2023年。[288] Haowen Zhang、Shiqi Wu、Srinivas Aluru 和 Heng Li。使用图波前算法进行快速序列到图比对。arXiv预印本arXiv:2206 13574，2022年 . 。[289] Pesho Ivanov、Benjamin Bichsel、Harun Mustafa、André Kahles、Gunnar Rätsch 和 Martin Vechev。AStarix：快速且最优的序列到图比对，第104–119页。施普林格国际出版，2020年。[290] Pesho Ivanov、Benjamin Bichsel 和 Martin Vechev。由种子引导的快速且最优序列到图比对。载于RECOMB，第306–325页，2022年。[291] Davide Cesare Monti

, Brian Riccardi , Jouni Sirén , Paola Bonizzoni , Gianluca

Della Vedova 和 Raffaella Rizzi。更快速、准确且具备重组感知能力的序列到图比对器。《BMC 生物信息学》，2026年。[292] Tong Zhu、Pankaj Vats、Seth Onken、Al Dunstan、Babak Zamirai、Daniel F Puleri、Abhishek Nair、Marco Oliva、Anil Gaihre、Priyanka Sadhnani 等。Parabricks：基于图形处理器加速、适用于所有测序仪的通用基因组学分析软件套件。bioRxiv，第 2025–07 页，2025年。[293] 英伟达公司。giraffe（vg giraffe + GATK）。Clara Parabricks v4.4.0 文档，2025年。访问日期：2026年6月11日。[294] Charlotte J. Houldcroft、Mathew A. Beale 和 Judith Breuer。病毒基因组测序带来的临床与生物学见解。《自然综述：微生物学》，2017年。[295] Natasha Jansz 和 Geoffrey J. Faulkner。病毒基因组测序方法：当前方法的优势与陷阱。《生物化学会学报》，2024年。[296] Peter J. A. Cock、Christopher J. Fields、Naohisa Goto、Michael L. Heuer 和 Peter M. Rice。用于带质量分数序列的 Sanger FASTQ 文件格式及 Solexa/Illumina FASTQ 变体。《核酸研究》，38(6):1767–1771，2009年。[297] Temple F Smith、Michael S Waterman 等。公共分子子序列的识别。《分子生物学杂志》，1981年。[298] Sergey Nurk、Sergey Koren、Arang Rhie、Mikko Rautiainen、Andrey V. Bzikadze、Alla Mikheenko、Mitchell R. Vollger、Nicolas Altemose、Lev Uralsky、Ariel Gersh- man、Sergey Aganezov、Savannah J. Hoyt、Mark Diekhans、Glennis A. Logsdon、Michael Alonge、Stylianos E. Antonarakis、Matthew Borchers、Gerard G. Bouffard、Shelise Y. Brooks、Gina V. Caldas、Nae-Chyun Chen、Haoyu Cheng、Chen-Shan Chin、William Chow、Leonardo G. de Lima、Philip C. Dishuck、Richard Durbin、Tatiana Dvorkina、Ian T. Fiddes、Giulio Formenti、Robert S. Fulton、Arkarachai Fung- tammasan、Erik Garrison、Patrick G. S. Grady、Tina A. Graves-Lindsay、Ira M. Hall、Nancy F. Hansen、Gabrielle A. Hartley、Marina Haukness、Kerstin Howe、Michael W. Hunkapiller、Chirag Jain、Miten Jain、Erich D. Jarvis、Peter Kerpedjiev、Melanie Kirsche、Mikhail Kolmogorov、Jonas Korlach、Milinn Kremitzki、Heng Li、Valerie V. Maduro、Tobias Marschall、Ann M. McCartney、Jennifer McDaniel、Danny E. Miller、James C. Mullikin、Eugene W. Myers、Nathan D. Olson、Benedict Paten、Paul Peluso、Pavel A. Pevzner、David Porubsky、Tamara Potapova、Evgeny I. Rogaev、Jeffrey A. Rosenfeld、Steven L. Salzberg、Valerie A. Schneider、Fritz J. Sedlazeck、Kishwar Shafin、Colin J. Shew、Alaina Shumate、Ying Sims、Arian F. A. Smit、Daniela C. Soto、Ivan Sović、Jessica M. Storer、Aaron Streets、Beth A. Sullivan、Françoise Thibaud- Nissen、James Torrance、Justin Wagner、Brian P. Walenz、Aaron Wenger、Jonathan M. D. Wood、Chunlin Xiao、Stephanie M. Yan、Alice C. Young、Samantha Zarate、Urvashi Surti、Rajiv C. McCoy、Megan Y. Dennis、Ivan A. Alexandrov、Jennifer L. Gerton、Rachel J. O’Neill、Winston Timp、Justin M. Zook、Michael C. Schatz、Evan E. Eichler、Karen H. Miga 和 Adam M. Phillippy。人类基因组完整序列。《科学》，376(6588):44–53，2022年。[299] Camille Marchet、Christina Boucher、Simon J. Puglisi、Paul Medvedev、Mikaël S alson 和 Rayan Chikhi。基于 k-mer 的数据结构：用于查询大型测序数据集。《基因组研究》，31(1):1–12，2021年。[300] Mikhail Karasikov、Harun Mustafa、Daniel Danciu、Oleksandr Kulkov、Marc Zim- mermann、Christopher Barber、Gunnar Rätsch 和 André Kahles。在拍字节规模序列库中进行高效、准确的搜索。《自然》，2025年。[301] Jason Fan、Noor Pratap Singh、Jamshed Khan、Giulio Ermanno Pibiri 和 Rob Patro。Fulgor：用于大规模匹配和颜色查询的快速紧凑 k-mer 索引。载于 WABI，第 273 卷，第 18:1–18:21 页，2023年。[302] Martin D Muggli、Alexander Bowe、Noelle R Noyes、Paul S Morley、Keith E Belk、Robert Raymond、Travis Gagie、Simon J Puglisi 和 Christina Boucher。简洁彩色de Bruijn图。《生物信息学》，33(20):3181–3187，2017年。[303] Zamin Iqbal、Mario Caccamo、Isaac Turner、Paul Flicek 和 Gil McVean。利用彩色de Bruijn图进行从头组装和变异基因分型。《自然·遗传学》，44(2):226–232，2012年。[304] Isaac Turner、Kiran V Garimella、Zamin Iqbal 和 Gil McVean。将长程连通性信息整合到德布鲁因图中。《生物信息学》，2018年。[305] Martin Hunt、Leandro Lima、Daniel Anderson、George Bouras、Michael B Hall、Jane Hawkey、Oliver Schwengers、Wei Shen、John Lees 和 Zamin Iqbal。AllTheBacteria——已组装、可获取且可搜索的所有细菌基因组。bioRxiv，2025年。[306] Heng Li。太碱基规模的 BWT 构建与搜索。arXiv，2024年。[307] Michael Roberts、Wayne Hayes、Brian R Hunt、Stephen M Mount 和 James A Yorke。降低生物序列比较的存储需求。《生物信息学》，2004年。[308] Heng Li。Minimap 和 Miniasm：含噪长序列的快速比对与从头组装。《生物信息学》，2016年。[309] Heng Li 和 Richard Durbin。利用 Burrows–Wheeler 变换实现快速、准确的短读段比对。《生物信息学》，25(14):1754–1760，2009年。[310] Youngmok Jung 和 Dongsu Han。BWA-MEME：采用机器学习方法模拟 BWA-MEM。《生物信息学》，38(9):2404–2413，2022年。[311] Francisco Fernandes 和 Ana T. Freitas。slaMEM：利用采样LCP数组高效检索最大精确匹配。《生物信息学》，30(4):464–471，2013年。[312] Robert Edgar。在生物序列中筛选保守 k-mer 时，Syncmer 比最小化器更灵敏。PeerJ，9:e10805，2021年。[313] Kristoffer Sahlin。利用 strobemer 实现有效的序列相似性检测。《基因组研究》，31(11):2080–2094，2021年。[314] Heng Li、Bob Handsaker、Alec Wysoker、Tim Fennell、Jue Ruan、Nils Homer、Gabor Marth、Goncalo Abecasis 和 Richard Durbin。序列比对/映射格式与 SAMtools。《生物信息学》，2009年。[315] Daehwan Kim、Joseph M Paggi、Chanhee Park、Christopher Bennett 和 Steven L Salzberg。使用 HISAT2 和 HISAT-genotype 进行基于图的基因组比对与基因分型。《自然·生物技术》，37(8):907–915，2019年。[316] Yan Gao、Yongzhuang Liu、Yanmei Ma、Bo Liu、Yadong Wang 和 Yi Xing。abPOA：

一种基于单指令多数据的 C 库，利用自适应带状区域实现快速偏序比对。《生物信息学》，37(15):2209–2211，2020年。[317] Mikko Rautiainen、Veli Mäkinen 和 Tobias Marschall。位并行序列到图对齐。《生物信息学》，35(19):3599–3607，2019年。[318] Ghanshyam Chandra 和 Chirag Jain。使用间隙感知共线链式连接进行序列到图对齐，第 58–73 页。施普林格 · 自然瑞士，2023 年。[319] Pesho Ivanov 、 Benjamin Bichsel 和 Martin Vechev。由种子引导的快速最优序列到图对齐，第 306–325 页。施普林格国际出版，2022年。[320] Charlotte A Darby、Ravi Gaddipati、Michael C Schatz 和 Ben Langmead。VARGAS：一种无启发式比对方法，用于评估线性读段比对器和图读段比对器。《生物信息学》，36(12):3712–3718，2020年。[321] Stephen Hwang、Nathaniel K. Brown、Omar Y. Ahmed、Katharine M. Jenike、Sam Kovaka、Michael C. Schatz 和 Ben Langmead。用于 k-mer 查询的基于 MEM 的泛基因组索引。《分子生物学算法》，20(1)，2025年。[322] Sandra Romain 和 Claire Lemaitre。SVJedi-graph：利用变异图和长读段改进相近及重叠结构变异的基因分型。《生物信息学》，39(Supplement_1):i270–i278，2023年。[323] ParBLiSS。PaSGAL：并行序列到图比对——数据。https:// github.com/ParBLiSS/PaSGAL/tree/master/data，2019年。访问日期：2026-06-18。[324] Valerie A Schneider、Tina Graves-Lindsay、Kerstin Howe、Nathan Bouk、Hsiu- Chuan Chen、Paul A Kitts、Terence D Murphy、Kim D Pruitt、Françoise Thibaud- Nissen、Derek Albracht 等。GRCh38 与从头单倍体基因组组装的评估证实了参考组装经久不衰的质量。《基因组研究》，2017年。[325] Justin M Zook、David Catoe、Jennifer McDaniel、Lindsay Vang、Noah Spies、Arend Sidow、Ziming Weng、Yuling Liu、Christopher E Mason、Noah Alexander 等。对七个人类基因组进行广泛测序，以表征基准参考材料。《科学数据》，2016年。[326] Glenn Hickey、Jean Monlong、Jana Ebler、Adam M. Novak、Jordan M. Eizenga、Yan Gao、人类泛基因组参考联盟、Tobias Marschall、Heng Li 和 Benedict Paten。利用 Minigraph-Cactus 根据基因组比对构建泛基因组图。《自然·生物技术》，42(4):663–673，2024年。[327] M. Holtgrewe。Mason——第二代测序数据读段模拟器。柏林自由大学技术报告，2010年。[328] Yukiteru Ono、Michiaki Hamada 和 Kiyoshi Asai。PBSIM3：适用于所有类型 PacBio 和 ONT 长读段的模拟器。《NAR 基因组学与生物信息学》，4 092，2022年 4( ):lqac 。[329] Nika Mansouri Ghiasi、Talu Güloglu、Harun Mustafa、Can Firtina、Konstantina Koliogeorgi、Konstantinos Kanellopoulos、Haiyu Mao、Rakesh Nadig、Mohammad Sadrosadati、Jisung Park 和 Onur Mutlu。SAGe：一种轻量级算法-体系结构协同设计方案，用于缓解大规模基因组分析中的数据准备瓶颈。arXiv:2504.03732，2025年。[330] Wilson Snyder、Paul Wasson、Duane Galbi、Geza Lore 等。Verilator。https: //github.com/verilator/verilator。[331] Haocong Luo、F. Nisa Bostancı、Ataberk Olgun、Maria Makeenkova、Ziad Malik、Ipek Akdeniz 和 Onur Mutlu。Ramulator 2.1：面向现代 DRAM 系统的可组合内存系统模拟器。载于 ICS 研讨会，第 173–179 页，2026年。[332] Haocong Luo、F. Nisa Bostancı、Ataberk Olgun、Maria Makeenkova、Ziad Malik、Ipek Akdeniz 和 Onur Mutlu。Ramulator 2.1 源代码。https://github.com/CMU-SAFARI/ramulator2/tree/v2.1.0。[333] Yoongu Kim、Weikun Yang 和 Onur Mutlu。Ramulator：快速、可扩展的 DRAM 模拟器。《计算机体系结构快报》，2015年。[334] Yoongu Kim、Weikun Yang 和 Onur Mutlu。Ramulator 1.0 源代码。https: // github.com/CMU-SAFARI/ramulator。[335] Haocong Luo、Yahya Can Tuğrul、F Nisa Bostancı、Ataberk Olgun、A Giray Yağlıkçı 和 Onur Mutlu。Ramulator 2.0：一种现代化、模块化且可扩展的 DRAM 模拟器。载于 IEEE《计算机体系结构快报》，2023年。[336] Haocong Luo、Yahya Can Tuğrul、F Nisa Bostancı、Ataberk Olgun、A Giray Yağlıkçı 和 Onur Mutlu。Ramulator 2.0 源代码。https://github.com/CMU-SAFARI/ ramulator2。[337] Clifford Wolf。Yosys——免费 Verilog 综合套件。载于 Austrochip，2013年。[338] Tolulope Ajayi 等。OpenROAD：迈向自动驾驶式开源数字布局实现工具链。载于 DAC，2019年。[339] Ataberk Olgun、Spiros Galanopoulos、Andreas Kosmas Kakolyris、Haocong Luo、İsmail Emir Yüksel、F. Nisa Bostancı 和 Onur Mutlu。理解现代高带宽内存的功耗：基于真实 HBM2 DRAM 芯片的实验分析与建模。载于 MICRO，2026年。[340] 英伟达。英伟达管理库（NVML）。https://developer.nvidia

.

com/management-library-nvml，2026年。[341] Oxford Nanopore Technologies。MinION Mk1CO。https://nanoporetech.com/document/requirements/minion-mk1c-spec，2024年。[342] Aspyn Palatnick、Bin Zhou、Elodie Ghedin 和 Michael C Schatz。iGenomics：在智能手机上进行全面的 DNA序列分析。*GigaScience*，9 138，2020年 12( ):giaa 。[343] Zachary S. Ballard、Calvin Brown 和 Aydogan Ozcan。用于全球微生物组发现、分析和工程研究的移动技术。*ACS Nano*，12(4):3065–3082，2018年。[344] Josephine B. Oehler、Helen Wright、Zornitza Stark、Andrew J. Mallett 和 Ulf Schmitz。长读长测序在临床环境中的应用。人类基因组学，17(1):73，2023年。[345] Aaron Pomerantz、Nicolás Peñafiel、Alejandro Arteaga、Lucas Bustamante、Frank Pichardo、Luis A Coloma、César L Barrio-Amorós、David Salazar-Valenzuela 和 Stefan Prost。《利用纳米孔测序在雨林中实时进行 DNA 条形码分析：快速生物多样性评估与本地能力建设的机遇》。GigaScience，2018年。[346] Augusto Dulanto Chiang 和 John P Dekker。《从流水线到病床：临床宏基因组学的进展与挑战》。The Journal of Infectious Diseases，221(Supplement_3)：S331–S340，2019年。[347] Aaron Stillmaker 和 Bevan Baas。《180 纳米至 7 纳米制程下准确预测 CMOS 器件性能的缩放方程》。Integration，2017年。[348] Mohammed Alser、Jeremy Rotman、Dhrithi Deshpande、Kodi Taraszka、Huwenbo Shi、Pelin Icer Baykal、Harry Taegyun Yang、Victor Xue、Sergey Knyazev、Ben- jamin D. Singer、Brunilda Balliu、David Koslicki、Pavel Skums、Alex Zelikovsky、Can Alkan、Onur Mutlu 和 Serghei Mangul。《技术决定算法：读段比对的最新进展》。《基因组生物学》，2021年。[349] Ben Langmead 和 Steven L Salzberg。《使用 Bowtie 2 进行快速带间隙读段比对》。《自然·方法》，2012年。[350] Heng Li。《使用 BWA-MEM 比对序列读段、克隆序列和组装重叠群》。arXiv，2013年。[351] Faraz Hach、Fereydoun Hormozdiari、Can Alkan、Farhad Hormozdiari、Inanc Birol、Evan E Eichler 和 S Cenk Sahinalp。《mrsFAST：一种用于短读段比对的缓存无关算法》。《自然·方法》，7(8)：576–577，2010年。[352] David Weese、Anne-Katrin Emde、Tobias Rausch、Andreas Döring 和 Knut Reinert。《RazerS——具有灵敏度控制的快速读段比对》。《基因组研究》，19(9)：1646–1654，2009年。[353] Guillaume Marçais、David Pellow、Daniel Bork、Yaron Orenstein、Ron Shamir 和 Carl Kingsford。《改进最小化器与 Winnowing 方案的性能》。生物信息学，2017年。[354] Chirag Jain、Arang Rhie、Haowen Zhang、Claudia Chu、Brian P Walenz、Sergey Koren 和 Adam M Phillippy。《加权最小化器采样可改善长读段比对》。生物信息学，36(Supplement_1)：i111–i118，2020年。[355] Hongyu Zheng、Carl Kingsford 和 Guillaume Marçais。《实用最小化器的改进设计与分析》。生物信息学，36(Supplement_1)：i119–i127，2020年。[356] Hongyi Xin、Sunny Nahar、Richard Zhu、John Emmons、Gennady Pekhimenko、Carl Kingsford、Can Alkan 和 Onur Mutlu。《Optimal Seed Solver：优化读段比对中的种子选择》。生物信息学，2016年。[357] Hongyi Xin、Mingfu Shao 和 Carl Kingsford。《用于读段比对的上下文感知种子》。Algorithms for Molecular Biology，2020年。[358] Can Firtina、Jisung Park、Mohammed Alser、Jeremie S Kim、Damla Senol Cali、Taha Shahroodi、Nika Mansouri Ghiasi、Gagandeep Singh、Konstantinos Kanellopou- los、Can Alkan 和 Onur Mutlu。《BLEND：一种快速、节省内存且准确的机制，用于在基因组分析中查找模糊种子匹配》。《NAR 基因组学与生物信息学》，5(1)，2023年。[359] Bariş Ekim、Kristoffer Sahlin、Paul Medvedev、Bonnie Berger 和 Rayan Chikhi。《利用 mapquik 在最小化器空间中高效比对准确的长读段》。《基因组研究》，33(7)：1188–1197，2023年。[360] Md Vasimuddin、Sanchit Misra、Heng Li 和 Srinivas Aluru。《面向多核系统的 BWA-MEM 高效架构感知加速》。载于 IPDPS，2019年。[361] Can Firtina、Nika Mansouri Ghiasi、Joel Lindegger、Gagandeep Singh、Meryem Banu Cavlak、Haiyu Mao 和 Onur Mutlu。《RawHash：实现大型基因组原始纳米孔信号的快速、准确实时分析》。生物信息学，39(Supplement_1)：i297–i307，2023年。[362] Can Firtina、Melina Soysal、Joël Lindegger 和 Onur Mutlu。《RawHash2：利用基于哈希的种子生成和自适应量化比对原始纳米孔信号》。生物信息学，2024年。[363] Joël Lindegger、Can Firtina、Nika Mansouri Ghiasi、Mohammad Sadrosadati、Mo- hammed Alser 和 Onur Mutlu。《RawAlign：通过结合种子生成与比对，实现准确、快速且可扩展的原始纳米孔信号比对》。arXiv 预印本 arXiv:2310.05037，2023年。[364] Ann Franchesca Laguna、Hasindu Gamaarachchi、Xunzhao Yin、Michael Niemier、Sri Parameswaran 和 X Sharon Hu。《基于播种与投票的 DNA 读段比对内存内加速器》。载于 ICCAD，2020年。[365] Longlong Chen、Jianfeng Zhu、Guiqiang Peng、Mingxu Liu、Shaojun Wei 和 Leibo Liu。《GEM：通过划分与预测式散射实现读段比对的超高效近存储器可重构加速》。IEEE并行与分布式系统汇刊，34(12)：3059–3072，2023年。[366] Hongyi Xin、Donghyuk Lee、Farhad Hormozdiari、Samihan Yedkar、Onur Mutlu 和 Can Alkan。《使用 FastHASH 加速读段比对》。《BMC 基因组学》，2013年。[367] Hongyi Xin、John Greth、John Emmons、Gennady Pekhimenko、Carl Kingsford、Can Alkan 和 Onur Mutlu。《移位汉明距离：一种快速准确、适用于单指令多数据的过滤器，可加速读段比对中的比对验证》。生物信息学，2015年。[368] Mohammed Alser、Onur Mutlu 和 Can Alkan。《MAGNET：理解并改进基因组预比对过滤的准确性》。arXiv，2017年。[369] Mohammed Alser、Taha Shahroodi、Juan Gómez-Luna、Can Alkan 和 Onur Mutlu。《SneakySnake：一种适用于中央处理器、图形处理器和 FPGA 的快速准确通用基因组预比对过滤器》。生物信息学，2020年。[370] Mohammed Alser、Hasan Hassan、Akash Kumar、Onur Mutlu 和 Can Alkan。《Shouji：一种用于序列比对的快速高效预比对过滤器》。生物信息学，2019年。[371] Gagandeep Singh、Mohammed Alser、Damla Senol Cali、Dionysios Diamantopoulos、Juan Gómez-Luna、Henk Corporaal 和 Onur Mutlu。《基于 FPGA 的近存储器加速：面向现代数据密集型应用》。《IEEE Micro》，2021年。[372] Zülal Bingöl、Mohammed Alser、Onur Mutlu、Ozcan Ozturk 和 Can Alkan。《GateKeeper-GPU：短读段比对中的快速准确预比对过滤

》。载于 IPDPSW，2021年。[373] Roman Kaplan、Leonid Yavits 和 Ran Ginosar。《RASSA：用于近似 DNA 长读段比对的电阻式预对齐加速器》。《IEEE Micro》，2018年。[374] Fazal Hameed、Asif Ali Khan 和 Jeronimo Castrillon。《ALPHA：一种用于加速 DNA 种子定位过滤的新型算法-硬件协同设计》。ITETC，2021年。[375] Peiyi He、Ruibin Mao、Keyi Shan、Yunwei Tong、Zhicheng Xu、Muyuan Peng、Ruibang Luo 和 Can Li。《ShiftCAM：一种利用移位汉明距离进行稳健基因组分析的时域内容寻址存储器》。载于 ICCAD，第 1–9 页，2024年。[376] Nika Mansouri Ghiasi、Jisung Park、Harun Mustafa、Jeremie Kim、Ataberk Ol- gun、Arvid Gollwitzer、Damla Senol Cali、Can Firtina、Haiyu Mao、Nour Almad- houn Alserr、Rachata Ausavarungnirun、Nandita Vijaykumar、Mohammed Alser 和 Onur Mutlu。《GenStore：用于基因组序列分析的高性能存储内处理系统》。载于第 27 届 ACM 编程语言与操作系统架构支持国际会议论文集，第 635–654 页，2022年。[377] Gene Myers 和 Webb Miller。《以次二次时间复杂度链式连接多重比对片段》。载于 SODA，第 38–47 页，1995年。[378] Mohamed Ibrahim Abouelhoda 和 Enno Ohlebusch。《用于多基因组比较的链式连接算法》。《离散算法杂志》，3(2)：321–341，2005年。[379] Christian Otto、Steve Hoffmann、Jan Gorodkin 和 Peter F. Stadler。《使用成对求和间隙代价进行快速局部片段链式连接》。Algorithms for Molecular Biology，6(1)：4，2011年。[380] Veli Mäkinen 和 Kristoffer Sahlin。《重审含重叠的链式连接》。载于 CPM，第 25:1–25:12 页，2020年。[381] Chirag Jain、Daniel Gibney 和 Sharma V. Thankachan。《含重叠和间隙代价的共线链式连接》。载于 RECOMB，第 246–262 页，2022年。[382] Jim Shaw 和 Yun William Yu。《通过对种子-链式连接-延伸启发式进行平均情形分析，证明序列比对器可在几乎 O(m log n) 的时间内保证准确性》。《基因组研究》，33(7)：1175–1187，2023年。[383] Saurabh Kalikar、Chirag Jain、Md Vasimuddin 和 Sanchit Misra。《在现代中央处理器上加速 Minimap2，以用于长读长测序应用》。《自然·计算科学》，2022年。[384] Licheng Guo、Jason Lau、Zhenyuan Ruan、Peng Wei 和 Jason Cong。《基因组测序中长读段两两重叠的硬件加速：FPGA 与 GPU 的较量》。载于 FCCM，2019年。[385] Harisankar Sadasivan、Milos Maric、Eric Dawson、Vishanth Iyer、Johnny Israeli 和 Satish Narayanasamy。《在 GPU 上加速 Minimap2，以实现准确的长读段比对》。Journal of Biotechnology and Biomedicine，6(1)：13，2023年。[386] Juechu Dong、Xueshen Liu、Harisankar Sadasivan、Sriranjani Sitaraman 和 Satish Narayanasamy。《mm2-gb：用于长读段 DNA 比对的 GPU 加速 Minimap2》。载于 BCB，第 1–9 页，2024年。[387] Kisaru Liyanage、Hiruna Samarakoon、Sri Parameswaran 和 Hasindu Gamaarachchi。《使用集成硬件加速链式连接的 minimap2-fpga，实现高效端到端长读段序列比对》。《科学报告》，2023年。[388] Kisaru Liyanage、Hasindu Gamaarachchi、Hassaan Saadat、Tuo Li、Hiruna Sama- rakoon 和 Sri Parameswaran。《使用 RISC-V 自定义指令加速基因组分析中的链式连接》。载于 DATE，2024年。[389] Yufeng Gu、Arun Subramaniyan、Tim Dunn、Alireza Khadem、Kuan-Yu Chen、Somnath Paul、Md Vasimuddin、Sanchit Misra、David Blaauw、Satish Narayanasamy 和 Reetuparna Das。《GenDP：面向基因组测序分析的动态规划加速框架》。载于 ISCA，2023年。[390] Osamu Gotoh。《一种改进的生物序列匹配算法》。《分子生物学杂志》，1982年。[391] Esko Ukkonen。《近似字符串匹配算法》。Information and Control，64(1–3)：100–118，1985年。[392] Gene Myers。《一种基于动态规划的近似字符串匹配快速位向量算法》。JACM，1999年。[393] Martin Šošić 和 Mile Šikić。《Edlib：一种使用编辑距离进行快速、精确序列比对的 C/C++ 库》。生物信息学，2017年。[394] Mengyao Zhao、Wan-Ping Lee、Erik P. Garrison 和 Gabor T. Marth。《SSW 库：用于基因组学应用的单指令多数据 Smith-Waterman C/C++ 库》。PLOS ONE，8(12)：e82138，2013年。[395] Hajime Suzuki 和 Masahiro Kasahara。《引入差分递推关系，以加快长序列的半全局比对》。BMC Bioinformatics，2018年。[396] Santiago Marco-Sola、Jordan M Eizenga、Andrea Guarracino、Benedict Paten、Erik Garrison 和 Miquel Moreto。《在 O(s) 空间内进行最优间隙仿射比对》。生物信息学，39(2)：btad074，2023年。[397] Ragnar Groot Koerkamp 和 Pesho Ivanov。《使用 A*、链式连接种子启发式和匹配剪枝进行精确全局比对》。生物信息学，2024年。[398] Ragnar Groot Koerkamp。《A*PA2：快达 19× 倍的精确全局比对》。载于 WABI，2024年。[399] Joël Lindegger、Damla Senol Cali、Mohammed Alser、Juan Gòmez-Luna 和 Onur Mutlu。《GenASM 算法的算法改进与 GPU 加速》。载于 IPDPSW，2022年。[400] Joël Lindegger、Damla Senol Cali、Mohammed Alser、Juan Gómez-Luna、Nika Man- souri Ghiasi 和 Onur Mutlu。《Scrooge：一种适用于中央处理器、图形处理器和专用集成电路的快速、节省内存的基因组序列比对器》。生物信息学，39(5)，2023年。[401] Maximilian-David Rumpf、Mohammed Alser、Arvid E. Gollwitzer、Joel Lindegger、Nour Almadhoun、Can Firtina、Serghei Mangul 和 Onur Mutlu。《SequenceLab：用于比较基因组序列的计算方法综合基准测试》。arXiv，2023年。[402] Nauman Ahmed、Tong Dong Qiu、Koen Bertels 和 Zaid Al-Ars。《GPU 加速 Darwin 读段重叠器，以对长 DNA 读段进行从头组装》。BMC

《生物信息学》，21(13):388，2020年。[403] Sneha D. Goenka、Yatish Turakhia、Benedict Paten 和 Mark Horowitz。SegAlign：一种可扩展的、基于GPU的全基因组比对工具。载于 SC，2020年。[404] Quim Aguado-Puig、Santiago Marco-Sola、Juan Carlos Moure、Christos Matzoros、David Castells-Rufas、Antonio Espinosa 和 Miquel Moreto。WFA-GPU：使用GPU进行间隙仿射成对比对。bioRxiv，2022年。[405] Seongyeon Park、Hajin Kim、Tanveer Ahmad、Nauman Ahmed、Zaid Al-Ars、H. Peter Hofstee、Youngsok Kim 和 Jinho Lee。SALoBa：通过最大化数据局部性和负载均衡，实现GPU上的快速序列比对。载于 IPDPS，2022年。[406] Seongyeon Park、Junguk Hong、Jaeyong Song、Hajin Kim、Youngsok Kim 和 Jinho Lee。AGAThA：面向长读段比对的引导式序列比对快速高效GPU加速。载于 PPoPP，第431–444页，2024年。[407] Alberto Zeni、Seth Onken、Marco Domenico Santambrogio 和 Mehrzad Samadi。利用差分递推关系实现高性能GPU基因组比对。载于 PACT，第133–143页，2024年。[408] Yi-Lun Liao、Yu-Cheng Li、Nae-Chyun Chen 和 Yi-Chang Lu。适用于长读段的自适应带状Smith-Waterman算法及其硬件加速器。载于 ASAP，第1–9页，2018年。[409] Abbas Haghi、Santiago Marco-Sola、Lluc Alvarez、Dionysios Diamantopoulos、Christoph Hagleitner 和 Miquel Moreto。WFA-FPGA：一种用于短读段和长读段基因组比对的高效波前算法加速器。《未来一代计算机系统》，149:39–58，2023年。[410] Carolina Teng、Renan W. Achjian、Jiang C. Wang 和 Fernando J. Fonseca。在FPGA云实例中调整GACT-X比对工具，以加速Minimap2。《应用科学》，13(7)，2023年。[411] Julian Pavon、Ivan Vargas Valdivieso、Carlos Rojas、Cesar Hernandez、Mehmet Aslan、Roger Figueras、Yichao Yuan、Joël Lindegger、Mohammed Alser、Francesc Moll、Santiago Marco-Sola、Oguz Ergin、Nishil Talati、Onur Mutlu、Osman Unsal、Mateo Valero 和 Adrian Cristal。QUETZAL：面向现代基因组序列分析算法的向量加速框架。载于 ISCA，2024年。[412] Shaahin Angizi、Jiao Sun、Wei Zhang 和 Deliang Fan。PIM-Aligner：一种用于生物序列比对的存内处理MRAM平台。载于 DATE，2020年。[413] Safaa Diab、Amir Nassereldine、Mohammed Alser、Juan Gómez Luna、Onur Mutlu 和 Izzat El Hajj。利用真实存内处理系统实现高吞吐量序列比对的框架。《生物信息学》，39(5)，2023年。[414] Alejandro Alonso-Marín、Ivan Fernandez、Quim Aguado-Puig、Juan Gómez-Luna、Santiago Marco-Sola、Onur Mutlu 和 Miquel Moreto。BIMSA：利用存内处理加速长序列比对。《生物信息学》，40(11)，2024年。[415] Meven Mognol、Dominique Lavenier 和 Julien Legriel。面向长DNA序列比对的带状Needleman & Wunsch算法在UPMEM PiM架构上的并行化。载于 ICPP，第1062–1071页，2024年。[416] Jie Cheng、Lifu Hu、Wei Xu、Hanhua Chen 和 Tian Xia。Minimap2基因组序列比对算法的硬件加速。载于 ICPP，第887–897页，2024年。[417] Jialei Sun、Lingyi Liu、Kunyue Li、Shuaipeng Li、Sai Gao、Zizheng Dong、Jianfei Jiang 和 Fangzhen Wu。一种基于Minimap2、用于第三代基因组测序的高能效CPU-FPGA异构加速系统。《IEEE电路与系统汇刊II：简报》，2025年。[418] Julien Eudine、Chu Li、Zhuo Cheng、Renzo Andri、Can Firtina、Mohammad Sadrosadati、Nika Mansouri Ghiasi、Konstantina Koliogeorgi、Anirban Nag、Arash Tavakkol、Haiyu Mao、Onur Mutlu、Shai Bergman 和 Ji Zhang。GenPairX：一种用于双端读段比对的硬件-算法协同设计加速器。载于 HPCA，2026年。[419] Jeremie S. Kim、Can Firtina、Meryem Banu Cavlak、Damla Senol Cali、Nastaran Hajinazar、Mohammed Alser、Can Alkan 和 Onur Mutlu。AirLift：一种在参考基因组之间重新映射比对结果的快速综合技术。《IEEE/ACM TCBB》，2024年。[420] Jeremie S. Kim、Can Firtina、Meryem Banu Cavlak、Damla Senol Cali、Can Alkan 和 Onur Mutlu。FastRemap：一种快速在基因组组装版本之间重新映射读段的工具。《生物信息学》，2022年。[421] Nika Mansouri Ghiasi 和 Onur Mutlu。通过以存储为中心的系统设计，实现快速、高效、低成本的基因组和宏基因组分析。载于 ICS 研讨会，第97–102页，2026年。[422] Onur Mutlu、Saugata Ghose、Juan Gómez-Luna 和 Rachata Ausavarungnirun。在合适的位置处理数据：实现内存计算。《微处理器与微系统》，2019年。[423] Gagandeep Singh、Lorenzo Chelini、Stefano Corda、Ahsan Javed Awan、Sander Stuijk、Roel Jordans、Henk Corporaal 和 Albert-Jan Boonstra。近存计算架构综述：机遇与挑战。载于 DSD，2018年。[424] Gagandeep Singh、Juan Gomez-Luna、Giovanni Mariani、Geraldo F. Oliveira、Stefano Corda、Sander Stuijk、Onur Mutlu 和 Henk Corporaal。NAPEL：通过集成学习预测近存计算应用的性能。载于 DAC，2019年。[425] Gagandeep Singh、Dionysios Diamantopoulos、Christoph Hagleitner、Juan Gomez-Luna、Sander Stuijk、Onur Mutlu 和 Henk Corporaal。NERO：用于天气预测建模的近高带宽内存模板加速器。载于 FPL，2020年。[426] Yitu Wang、Shiyu Li、Qilin Zheng、Linghao Song、Zongwang Li、Andrew Chang、Hai Li 和 Yiran Chen。NDSEARCH：通过近数据处理加速基于图遍历的近似最近邻搜索。载于 ISCA，2024年。[427] Weihong Xu、Junwei Chen、Po-Kai Hsu、Jaeyoung Kang、Minxuan Zhou、Sumukh Pinge、Shimeng Yu 和 Tajana Rosing。Proxima：面向3D NAND中基于图的近似最近邻搜索的近存储加速。《IEEE计算机汇刊》，2026年。