

总体概述：

计算机架构划分：<https://face2ai.com/CUDA-F-1-0-%E5%B9%B6%E8%A1%8C%E8%AE%A1%E7%AE%97%E4%B8%8E%E8%AE%A1%E7%AE%97%E6%9C%BA%E6%9E%B6%E6%9E%84/>

子主题：

计算机体系结构、高性能计算（图像，并行）、硬件技术发展，

研究热点：

计算机硬件顶会：<https://asplos-conference.org/2022/>

<https://arxiv.org/abs/2202.10400>

计算机科学 > 硬件架构

[2022 年 2 月 21 日提交]

GenStore：用于基因组序列分析的高性能、高能效的存储计算系统

Nika Mansouri Ghiasi, Jisung Park, Harun Mustafa, Jeremie Kim, Ataberk Olgun, Arvid Gollwitzer, Damla Senol Cali, Can Firtina, Haiyu Mao, Nour Almadhoun Alser, Rachata Ausavarungnirun, Nandita Vijaykumar, Mohammed Alser, Onur Mutlu

Read mapping 是许多基因组学应用程序中一个基本但计算量大的步骤。它用于识别已测序基因组的片段（称为读数）与已知基因组（称为参考基因组）之间的潜在匹配和差异。为了解决基因组分析中的计算挑战，许多先前的工作提出了各种方法，例如选择必须经过昂贵计算的读数的过滤器、有效的启发式方法和硬件加速。虽然可以有效地减少计算开销，但所有这些方法仍然需要将大量数据从存储设备转移到系统的其余部分，成本高昂，这会显著降低传统和新兴基因组学中读取映射的端到端性能系统。


我们提出 GenStore，第一个为基因组序列分析设计的存储处理系统，通过利用低成本和准确的存储过滤器，大大减少了基因组序列分析的数据移动和计算开销。GenStore 利用硬件/软件协同设计来应对存储处理的挑战，支持具有 1) 不同读取长度和错误率，以及 2) 不同程度的遗传变异的读取。通过对读取映射过程的严格分析，我们精心设计了基于 NAND 闪存 SSD 内的低成本硬件加速器和数据/计算流程。我们使用广泛的真实基因组数据集进行的评估表明，当在三个现代 SSD 中实施 GenStore 时，将最先进的软件（硬件）基线的读取映射性能显著提高了 $2.07\text{--}6.05\times(1.52\text{--}3.32\times)$ 对于与参考基因组高度相似的读取集和 $1.45\text{--}33.63\times(2.70\text{--}19.2\times)$ 对于与参考基因组相似度较低的读取集。

注释： 出现在 ASPLOS 2022

科目： **硬件架构 (cs.AR)** ；分布式、并行和集群计算 (cs.DC)；操作系统 (cs.OS)；基因组学 (q-bio.GN)

引用为： [arXiv:2202.10400](https://arxiv.org/abs/2202.10400) [**cs.AR**]

(或此版本的 [arXiv:2202.10400v1](https://arxiv.org/abs/2202.10400v1) [**cs.AR**])

<https://doi.org/10.48550/arXiv.2202.10400> 

放一个英伟达的学习博客：<https://developer.nvidia.com/zh-cn/blog/evaluating-applications-using-the-nvidia-arm-hpc-development-kit/>

高校研究：

<https://dl.acm.org/doi/full/10.1145/3466795>