Sequence Reads Analysis with GDC 2.0: Exploring the Sequence Reads and BAM Slicing Tools

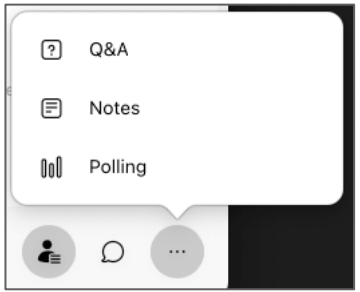
18 November 2024

Bill Wysocki, Ph.D,
Congyu Lu, Ph.D,
Center for Translational Data Science
University of Chicago

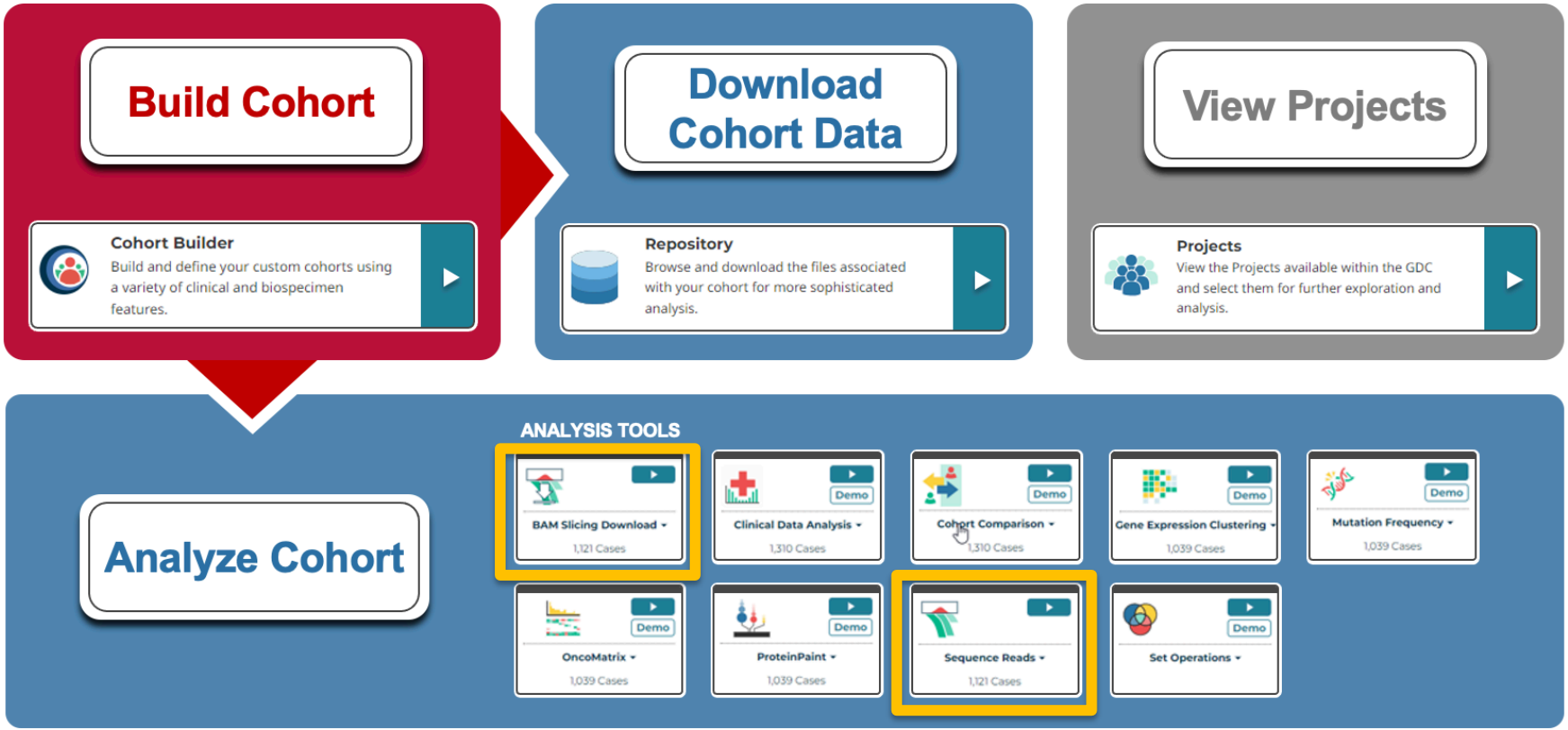
NIH NATIONAL CANCER INSTITUTE

Agenda

1. *Introduction to:*
   - *BAM Slicing Tool*
   - *Sequence Reads Tool*
2. *Demo: BAM Slicing Tool*
3. *Demo: Sequence Reads Tool*
4. *Resources for Support*
5. *Q&A*



GDC 2.0 Workflow

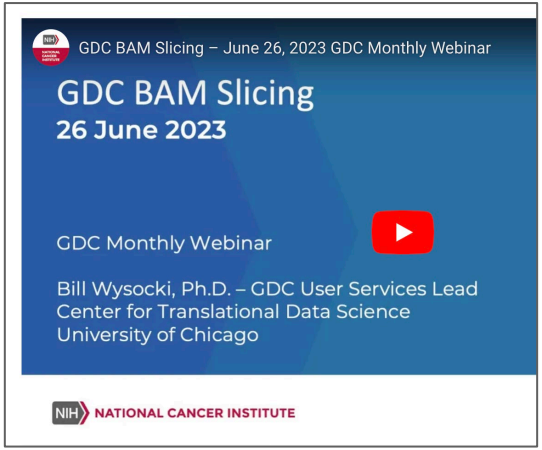


Introduction to the GDC BAM Slicing Tool

Download a Region

GDC BAM Slicing Download Tool

- Allows users to save space by downloading a specific region of any GDC-aligned BAM file
- Can be used to download custom genomic positions, genes, or entire chromosomes
- Also available through the GDC API
  - See the June 2023 Webinar



What do I need to use the BAM Slicing Tool?

- ✓ dbGaP access to the project that has the BAM
- ✓ A specific position in the genome to download
  - ✓ Coordinates / Mutation
  - ✓ Gene / Chromosome
- ✓ An active cohort that includes the case associated with the BAM
- ✓ Cannot be the transcriptome RNA-Seq BAM
  - ✓ RNA-Seq BAM aligned to the genome is okay

Introduction to the GDC Sequence Reads Tool

Visualize Aligned Reads

GDC Sequence Reads Tool

- Allows users to visualize reads as they are aligned to the reference genome
- Can be used to investigate specific reads or genomic positions
- Annotations, visualizations, and zoom level are all customizable
- Specific BAM slice you are looking at can be downloaded

What do I need to use the Sequence Reads Tool?

- ✓ dbGaP access to the project that has the BAM
- ✓ A specific position in the genome to look at
  - ✓ Coordinates / Mutation
  - ✓ Gene
- ✓ An active cohort that includes the case associated with the BAM
- ✓ Cannot be the transcriptome RNA-Seq BAM
  - ✓ RNA-Seq BAM aligned to the genome is okay

Tool Demonstration

Open Access Data

Resources for Support

Users Guides, Videos, and More!

Publication

- https://pubmed.ncbi.nlm.nih.gov/37140547/

Bioinformatics. 2023 May 4;39(5):btad300. doi: 10.1093/bioinformatics/btad300.

ppBAM: ProteinPaint BAM track for read alignment visualization and variant genotyping

Robin Paul [1], Jian Wang [1], Colleen Reilly [1], Edgar Sioson [1], Jaimin Patel [1], Gavriel Matt [1], Aleksandar Ačić [1], Xin Zhou [1]

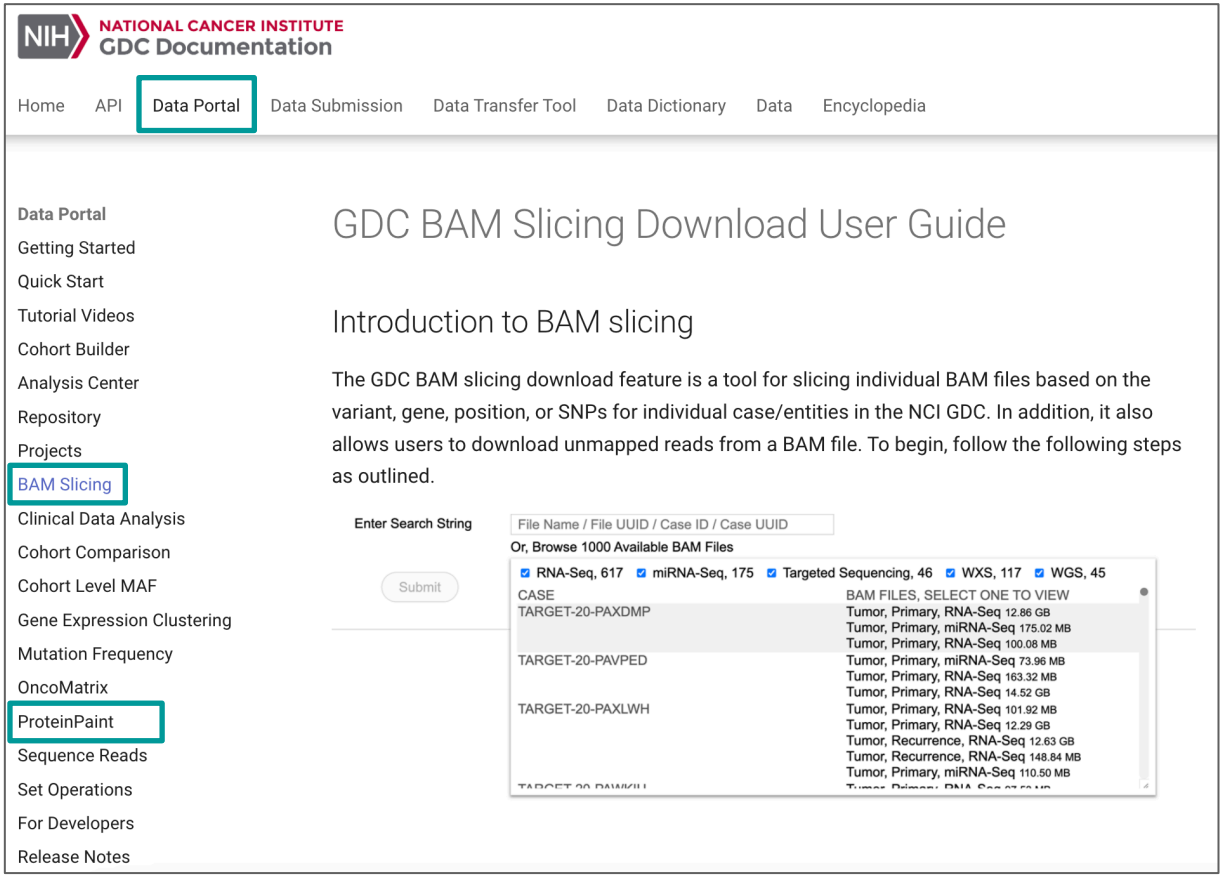
Affiliations + expand

PMID: 37140547 PMCID: PMC10182850 DOI: 10.1093/bioinformatics/btad300

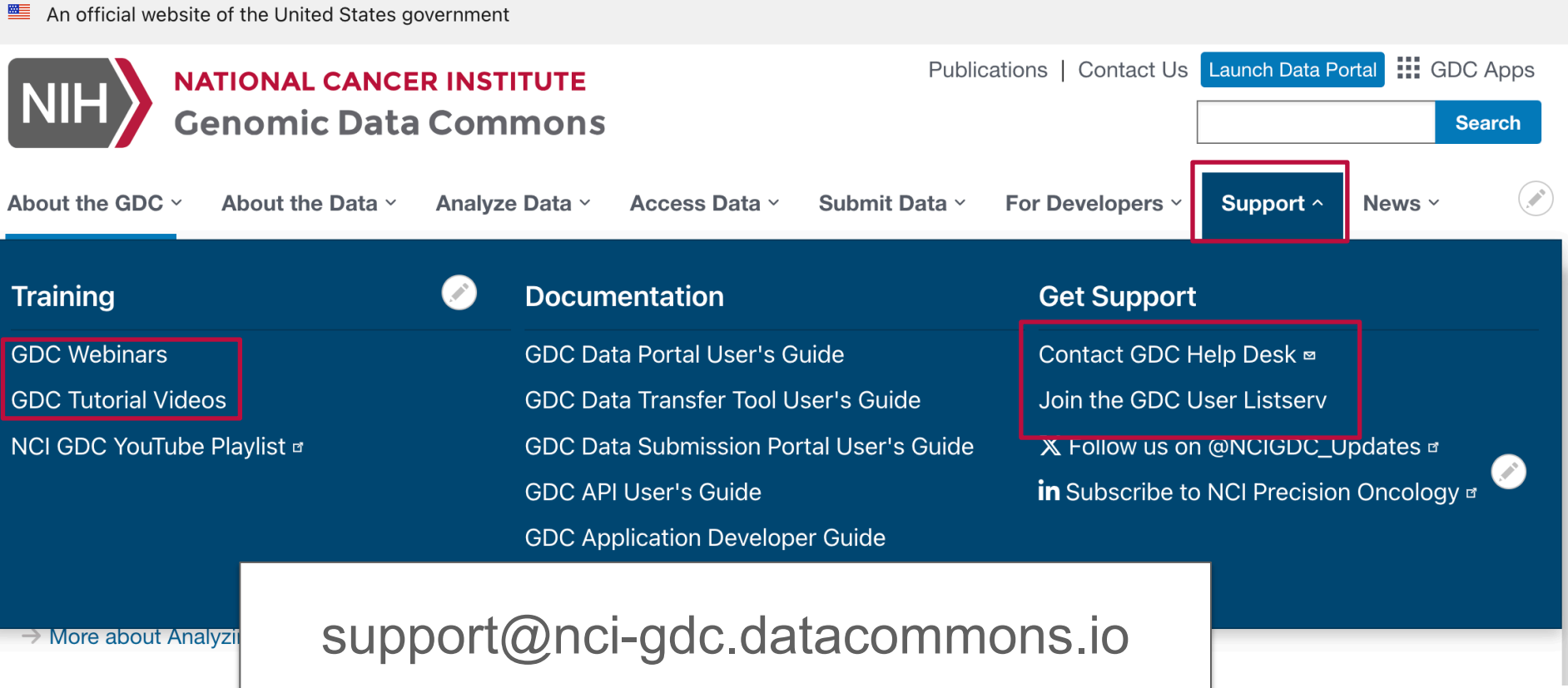
Abstract

Summary: ProteinPaint BAM track (ppBAM) is designed to assist variant review for cancer research and clinical genomics. With performant server-side computing and rendering, ppBAM supports on-the-fly variant genotyping of thousands of reads using Smith-Waterman alignment. To better visualize support for complex variants, reads are realigned against the mutated reference sequence using ClustalO. ppBAM also supports the BAM slicing API of the NCI Genomic Data Commons (GDC) portal, letting researchers conveniently examine genomic details of vast amounts of cancer sequencing data and reinterpret variant calls.

GDC Documentation Website: docs.gdc.cancer.gov



GDC Website: gdc.cancer.gov



Questions?

U.S. Department of Health & Human Services

National Institutes of Health | National Cancer Institute

gdc.cancer.gov

1-800-4-CANCER

Produced November 2024