

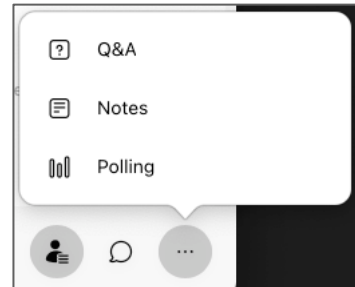
Sequence Reads Analysis with GDC 2.0: Exploring the Sequence Reads and BAM Slicing Tools

18 November 2024

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

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

Build Cohort



Cohort Builder

Build and define your custom cohorts using a variety of clinical and biospecimen features.

Download Cohort Data



Repository

Browse and download the files associated with your cohort for more sophisticated analysis.

View Projects



Projects

View the Projects available within the GDC and select them for further exploration and analysis.

Analyze Cohort

ANALYSIS TOOLS



BAM Slicing Download

1,121 Cases



Clinical Data Analysis

1,310 Cases



Cohort Comparison

1,310 Cases



Gene Expression Clustering

1,039 Cases



Mutation Frequency

1,039 Cases



OncoMatrix

1,039 Cases



ProteinPaint

1,039 Cases



Sequence Reads

1,121 Cases



Set Operations

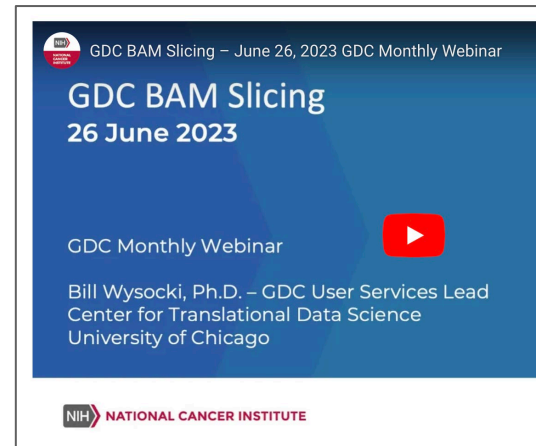


Introduction to the GDC BAM Slicing Tool

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 Acic](#)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

The screenshot shows the GDC Documentation Website interface. At the top left is the NIH logo and the text 'NATIONAL CANCER INSTITUTE GDC Documentation'. Below this is a navigation bar with links: Home, API, Data Portal (highlighted with a red box), Data Submission, Data Transfer Tool, Data Dictionary, Data, and Encyclopedia. On the left side, there is a vertical navigation menu with the following items: Data Portal, Getting Started, Quick Start, Tutorial Videos, Cohort Builder, Analysis Center, Repository, Projects, BAM Slicing (highlighted with a red box), Clinical Data Analysis, Cohort Comparison, Cohort Level MAF, Gene Expression Clustering, Mutation Frequency, OncoMatrix, ProteinPaint (highlighted with a red box), Sequence Reads, Set Operations, For Developers, and Release Notes. The main content area is titled 'GDC BAM Slicing Download User Guide' and contains the following text: 'Introduction to BAM slicing' and 'The GDC BAM slicing download feature is a tool for slicing individual BAM files based on the variant, gene, position, or SNPs for individual case/entities in the NCI GDC. In addition, it also allows users to download unmapped reads from a BAM file. To begin, follow the following steps as outlined.' Below the text is a search interface with a text input field labeled 'Enter Search String' containing the placeholder 'File Name / File UUID / Case ID / Case UUID', a 'Submit' button, and a section titled 'Or, Browse 1000 Available BAM Files'. This section includes a list of filters: RNA-Seq, 617; miRNA-Seq, 175; Targeted Sequencing, 46; WXS, 117; WGS, 45. Below the filters is a table with two columns: 'CASE' and 'BAM FILES, SELECT ONE TO VIEW'. The table lists several cases and their corresponding BAM files with sizes. The cases listed are TARGET-20-PAXDMP, TARGET-20-PAVPED, and TARGET-20-PAXLWH. The BAM files listed include Tumor, Primary, RNA-Seq, Tumor, Primary, miRNA-Seq, and Tumor, Recurrence, RNA-Seq.

NIH NATIONAL CANCER INSTITUTE

GDC Website: gdc.cancer.gov

 An official website of the United States government



Publications | Contact Us [Launch Data Portal](#)  GDC Apps

 [Search](#)

[About the GDC](#) [About the Data](#) [Analyze Data](#) [Access Data](#) [Submit Data](#) [For Developers](#) [Support](#) [News](#)

Training

[GDC Webinars](#)
[GDC Tutorial Videos](#)

[NCI GDC YouTube Playlist](#)



Documentation

[GDC Data Portal User's Guide](#)
[GDC Data Transfer Tool User's Guide](#)
[GDC Data Submission Portal User's Guide](#)
[GDC API User's Guide](#)
[GDC Application Developer Guide](#)

Get Support

[Contact GDC Help Desk](#)
[Join the GDC User Listserv](#)
[Follow us on @NCIGDC_Updates](#)
[Subscribe to NCI Precision Oncology](#)

support@nci-gdc.datacommons.io

[→ More about Analyzi](#)

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