

Genomic Mutations Analysis in GDC 2.0

22 April 2024

Bill Wysocki, Ph.D. – GDC User Services Lead
Center for Translational Data Science
University of Chicago

Agenda

1. *Overview of Genomic Mutations in the GDC*
2. *Custom Cohorts in GDC 2.0*
3. *Mutation Frequency Tool*
4. *OncoMatrix Tool*
5. *ProteinPaint Tool*
6. *Downloading Genomic Mutation Files*
7. *Questions*



Overview of Genomic Mutations in the GDC

Somatic Mutations

Simple Somatic Mutations in the GDC – Pipeline

1. WXS and Targeted Seq. tumor and normal reads are aligned to the GRCh38 reference genome (BAM x 2)
2. Variant calls are generated with four different pipelines (VCF x 4)
3. Variant calls are annotated (VCF/MAF x 4)
4. Variant calls are aggregated and filtered (MAF x 2)
 - Four single-pipeline files → One file
 - See the September 2020 Webinar: *GDC DNA-Seq Data Processing*
5. Variants are indexed in the GDC API from the open-access MAF

Copy Number Variation Mutations in the GDC – Two Pipelines

WGS Pipeline:

1. WGS tumor and normal reads are aligned to the GRCh38 reference genome (BAM)
2. Gene-level CNV calls are generated with ASCAT-NGS

OR

SNP6 Array Pipeline:

1. SNP6 array tumor – normal pairs are used to generate CNV calls with ASCAT2

-
- Gene-level CNV calls are indexed in the GDC API

Accessing Genomic Mutations - GDC Data Portal Tools

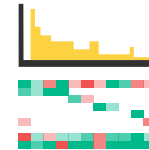
1. Mutation Frequency

- Summarize mutations in your cohort
- Visualize survival in mutated vs. non-mutated
- Filter and create sets of genes/mutations



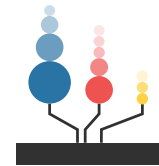
2. OncoMatrix

- Visualize mutations X cases in a matrix
- Visualize mutated cases alongside clinical attributes



3. ProteinPaint

- Visualize mutations on a polypeptide chain
- Display mutations on full genome



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 ▾

GDC 2.0 Workflow - Step 1

Build Cohort



Cohort Builder

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



Step 1: Build a cohort based on clinical or biospecimen attributes

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


 SAM Slicing Download +
120 Cases


 Clinical Data Analysis +
120 Cases


 Cohort Comparison +
120 Cases


 Gene Expression Clustering +
120 Cases

 Mutation Frequency +
120 Cases

 OncoMatrix +
120 Cases

 ProteinPaint +
120 Cases

 Sequence Reads +
120 Cases

 Set Operations +
120 Cases

GDC 2.0 Workflow - Step 2

Step 2: Use the cohort with tools in the analysis center.

Tools will be automatically applied to the cohort.

Download Cohort Data

View Projects

Repository

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

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** ▶ Demo
1,310 Cases


 **Cohort Comparison** ▶ Demo
1,310 Cases


 **Gene Expression Clustering** ▶ Demo
1,039 Cases

 **Mutation Frequency** ▶ Demo
1,039 Cases

 **OncoMatrix** ▶ Demo
1,039 Cases

 **ProteinPaint** ▶ Demo
1,039 Cases

 **Sequence Reads** ▶
1,121 Cases

 **Set Operations** ▶ Demo

Demo: Building a Cohort

Demo: Creating a Cohort for Mutation Analysis



Building a Cohort



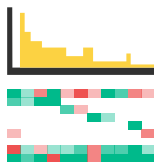
- **Goal:** Create two cohorts for mutation analysis
 - Program: TCGA
 - **Cohort 1** - Tissue or Organ of Origin: **upper lobe, lung**
 - **Cohort 2** - Tissue or Organ of Origin: **lower lobe, lung**

Demo: GDC Mutation Analysis Tools

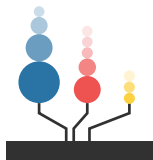
1. Mutation Frequency



2. OncoMatrix



3. ProteinPaint



Mutation Frequency



- Tool Layout
 - Displays mutations and genes from active cohort as well as:
 - Most frequently mutated genes
 - Survival plot
- Toggle between gene and mutation view
- Narrow down genes and mutations using:
 - Filters on the left side of the portal
 - Custom gene/mutation sets

Mutation Frequency - Goal 1



Goal #1:

- Create a smaller cohort of patients who have the most frequently occurring mutation in the original cohort.

Starting with cohorts in mutation freq. tool:

1. Switch to “Mutations”
2. Choose top row cohort toggle
3. ”Save As” new cohort

Mutation Frequency - Goal 2



Goal #2:

- Which CGC genes are most commonly mutated, with high VEP impact mutations, in each cohort
- Compare survival curves in the second-most mutated gene

Starting with cohorts in mutation freq. tool:


1. Filters:
 - CGC = Active
 - VEP Impact = High
2. Choose survival icon in second row
3. See survival plot

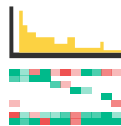
Mutation Frequency - Goal 3



Goal #3: How many cases have mutated CGC genes that are unique to their cohort?

Starting with cohorts in mutation freq. tool:

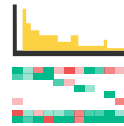
1. Find the CGC genes mutated in each cohort
2. Create a set for each
3. Set Operations 
4. Use newly created sets with the Mutation Frequency tool and corresponding cohort



Tool Layout

- Columns are cases
- Rows are genes
- Colored cells represent mutation occurrences
 - Each color represents a mutation consequence
- Appearance is customizable at the top
 - Gene sets are customizable
 - Cases are based on the active cohort

OncoMatrix - Goal



Goal: Visualize these cohorts in one matrix with Lung Cancer Survival genes along with the cases' gender and primary diagnosis

Starting with cohort in OncoMatrix tool:

- Customize genes
 - “Genes” button
 - Edit Group
- Add clinical attributes
 - Variables button
 - Choose attributes

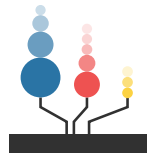
ProteinPaint – Layout 1



Tool Layout (1/2)

- Prompted to enter gene of interest
- Each mutation from your cohort is represented as a lollipop
 - Number represents mutated cases in the active cohort
 - Color represents mutation consequence
 - Clicking the mutation will cause the lollipop to move to a more prominent position
- Protein domains are shaded on the polypeptide chain

ProteinPaint – Layout 2



Tool Layout (2/2)

- Clicking a mutation also shows a list of cases (and case info) associated with it
- Disco plot: representing the full set of mutations in the genome – can be visualized and exported
- Default view: Polypeptide
 - Select alternative views

ProteinPaint Goal



Goal: Visualize missense mutations on the SSX1 gene and the full genome of one case with mutations

Starting with cohort in ProteinPaint tool:

1. Enter SSX1 into field
2. Choose “Missense” and “Show Only”
3. Choose mutation and click on it
4. Choose ”Disco Plot”

Demo: Downloading the Mutation Files



Mutation File Download



Goal: Query the files that are used to populate the analysis tools.

1. Simple Somatic Mutations

- Masked Somatic Mutations
- Targeted Seq. or WXS

2. Copy Number Variation

- ASCAT-NGS (WGS)
- ASCAT2 (SNP6)

GDC Support



- GDC Video Guides
- GDC User's Guides: <https://docs.gdc.cancer.gov>
- GDC Website: <https://gdc.cancer.gov>
- GDC Help Desk

Email: support@nci-gdc.datacommons.io

GDC Data Portal Features: Coming Soon!

- **Cohort MAF Tool:** Build an aggregated MAF from your cohort
- Enhanced OncoMatrix Features
- New Gene Expression Clustering Filter
- Bringing the **Annotation Browser** and **Sample Property Filtering** to the new GDC Data Portal
- ... and More!

Questions?

U.S. Department of Health & Human Services
National Institutes of Health | National Cancer Institute

<https://www.cancer.gov/>

1-800-4-CANCER

Produced April 2024