NATIONAL CANCER INSTITUTE

Center for Cancer Genomics | Genomic Data Commons

portal.gdc.cancer.gov | @NCIGDC_Updates | support@nci-gdc.datacommons.io

The Genomic Data Commons (GDC) provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.

Access Data & Tools

- Harmonized genomic data uniformly processed and aligned to latest reference genome.
- Mutation calls, expression levels, and other high-level data generated via best-in-class pipelines.
- Web-based tools to search, download and analyze data from over 33,000 cancer cases.
- Visualize biological & clinical relationships in real-time, then download publication-ready figures.
- Consume data efficiently with GDC's API or Data Transfer Tool

Projects Available

Clinical Proteomic Tumor Analysis Consortium (**CPTAC**), Foundation Medicine (**FMI**), The Cancer Genome Atlas (**TCGA**), Therapeutically Applicable Research to Generate Effective Treatments (**TARGET**), and other cancer studies

Examples of Supported Data Types

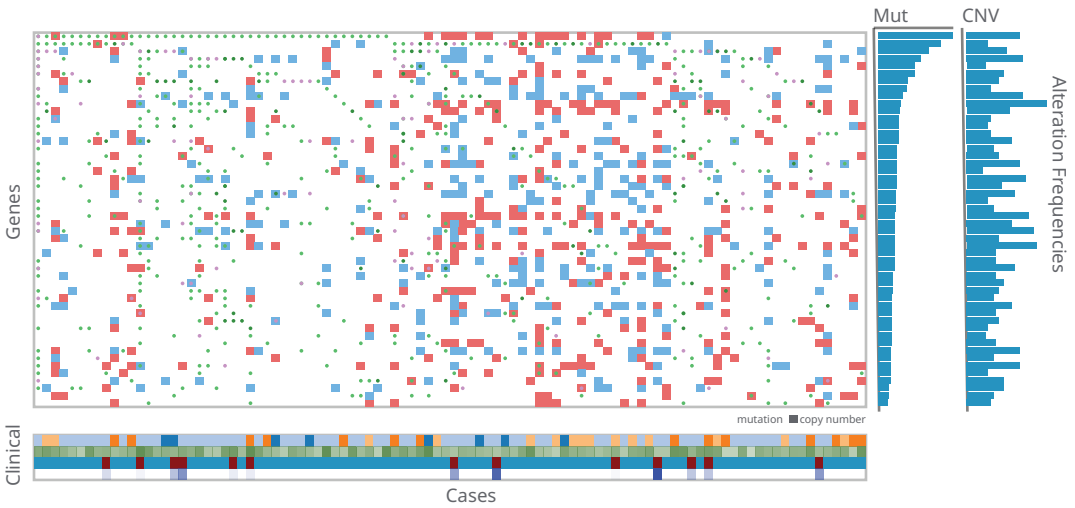
Data type	File Format
Clinical & Biospecimen	TSV, XML, JSON
Sequencing (e.g., WGS, WXS, RNA)	BAM, FASTQ
Array (e.g., SNP, Methylation)	TXT, IDAT
High-level Data	TSV (expression, splice junctions), TXT (CNV, methylation), MAF & VCF (mutations)

Data Analysis, Visualization & Exploration (DAVE) tools are web-based and open-access, helping make genomic data accessible for anyone.

Analyze with DAVE

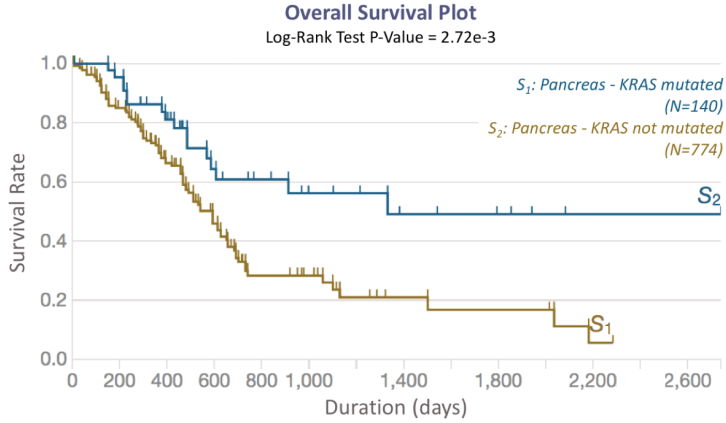
OncoGrid

Visualize combinations of gene mutations & copy number variants for a project or custom cohort



Survival Analysis

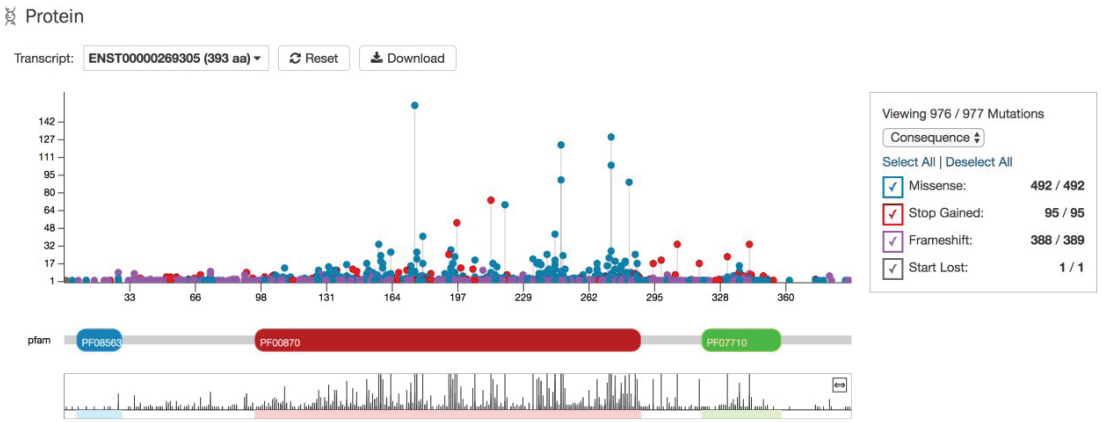
Compare overall survival of any two cohorts, such as patients with & without a mutated gene of interest



Analyze with DAVE

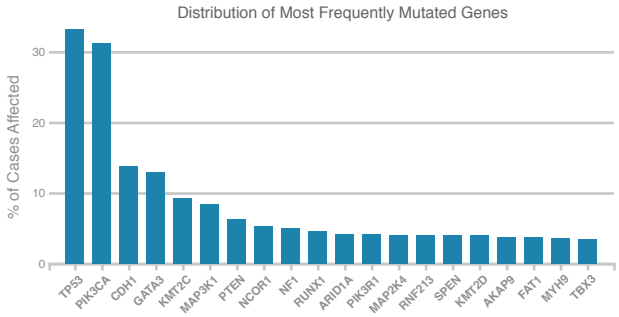
Protein Viewer

Visualize gene mutations mapped to their protein functional domains



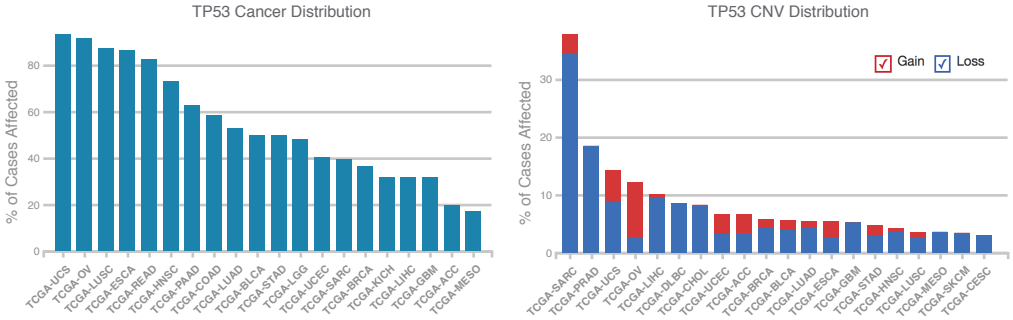
Visualize Frequent Alterations

View the most frequently mutated genes for any cohort



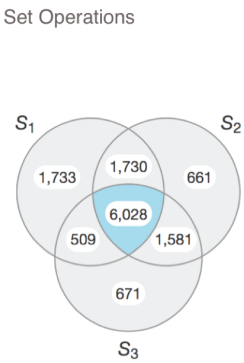
Symbol	Name	# SSM Affected Cases in Cohort	# SSM Affected Cases Across the GDC	# CNV Gain	# CNV Loss	# Mutations	Annotations	Survival
TP53	tumor protein p53	360 / 981 (36.70%)	4,008 / 10,202	16 / 1,062 (1.51%)	47 / 1,062 (4.43%)	240		
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	339 / 981 (34.56%)	1,403 / 10,202	151 / 1,062 (14.22%)	15 / 1,062 (1.41%)	82		
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	149 / 981 (15.19%)	385 / 10,202	36 / 1,062 (3.39%)	46 / 1,062 (4.33%)	133		
GATA3	GATA binding protein 3	141 / 981 (14.37%)	362 / 10,202	113 / 1,062 (10.64%)	10 / 1,062 (0.94%)	107		
KMT2C	lysine (K)-specific methyltransferase 2C	101 / 981 (10.30%)	1,003 / 10,202	51 / 1,062 (4.80%)	97 / 1,062 (9.13%)	128		

Plot frequencies of cases with mutations and copy number variants for a selected gene



Analyze Custom Sets

Create custom sets of cases, genes, or mutations
Compare survival and clinical features of case sets



Alias	Item Type	Name	# Items
S_1	Mutations	Bladder, High impact, Mutect2	10,000
S_2	Mutations	Bladder, High impact, Varscan	10,000
S_3	Mutations	Bladder, High impact, Muse	8,789

Select	Set Operation	# Items
☑	(S1 ∩ S2 ∩ S3)	6,028
☐	(S1 ∩ S2) - (S3)	1,730
☐	(S2 ∩ S3) - (S1)	1,581
☐	(S1 ∩ S3) - (S2)	509
☐	(S1) - (S2 ∪ S3)	1,733
☐	(S2) - (S1 ∪ S3)	661
☐	(S3) - (S1 ∪ S2)	671
Union of selected sets		6,028