Accessing and using underlying cBioPortal data

Introduction to cBioPortal

Course material

Feedback form

elixir UNITED KINGDOM

4 Recognise the process for accessing and analysing cBioPortal data

Why bother?

"under-the-hood" dataset has more information than displayed publicly

analyse lists of genes quickly

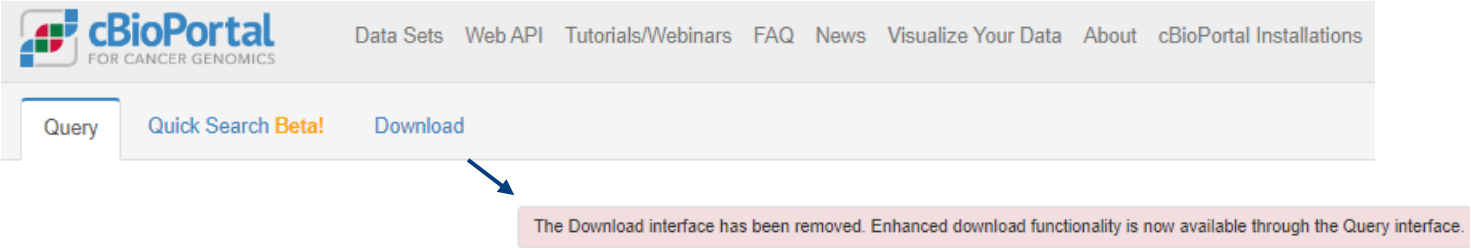
"improve" the plot quality

perform more advanced statistical testing (e.g. DEA, GSEA)

Course material

Feedback form

Downloading data



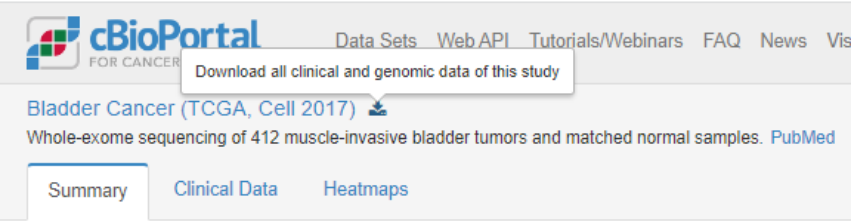
Unhelpful starting point.

Course material

Feedback form

Downloading data

Explore your dataset first, and then download.



Download will start and give a `.tar.gz` file

Course material

Feedback form

Downloading data

blca_tcga_pub_2017 | Search blca_tcga_pub_2017

Name	Date modified	Type	Size
case_lists	25/03/2022 19:07	File folder	
data_clinical_patient.txt	25/03/2022 19:15	TXT File	355 KB
data_clinical_sample.txt	25/03/2022 19:15	TXT File	103 KB
data_cna.txt	25/03/2022 19:15	TXT File	22,499 KB
data_linear_cna.txt	25/03/2022 19:15	TXT File	64,382 KB
data_methylation_hm450.txt	25/03/2022 19:15	TXT File	119,597 KB
data_mrna_seq_v2_rsem.txt	25/03/2022 19:15	TXT File	69,336 KB
data_mrna_seq_v2_rsem_zscores_ref_all_samples.txt	25/03/2022 19:15	TXT File	60,450 KB
data_mrna_seq_v2_rsem_zscores_ref_diploid_samples.txt	25/03/2022 19:15	TXT File	59,941 KB
data_mutations.txt	25/03/2022 19:15	TXT File	264,752 KB
data_mutsig.txt	25/03/2022 19:15	TXT File	2,083 KB
data_rppa.txt	25/03/2022 19:15	TXT File	643 KB
data_rppa_zscores.txt	25/03/2022 19:15	TXT File	569 KB
LICENSE	25/03/2022 19:07	File	1 KB
meta_clinical_patient.txt	25/03/2022 19:07	TXT File	1 KB
meta_clinical_sample.txt	25/03/2022 19:07	TXT File	1 KB
meta_cna.txt	25/03/2022 19:07	TXT File	1 KB
meta_linear_cna.txt	25/03/2022 19:07	TXT File	1 KB
meta_methylation_hm450.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem_zscores_ref_all_samples.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem_zscores_ref_diploid_samples.txt	25/03/2022 19:07	TXT File	1 KB
meta_mutations.txt	25/03/2022 19:07	TXT File	1 KB
meta_rppa.txt	25/03/2022 19:07	TXT File	1 KB
meta_rppa_zscores.txt	25/03/2022 19:07	TXT File	1 KB
meta_study.txt	25/03/2022 19:07	TXT File	1 KB

For each assay, 1 data file and 1 metadata/information file

Course material

Feedback form

elixir UNITED KINGDOM

Downloading data

blca_tcga_pub_2017 | Search blca_tcga_pub_2017

Name	Date modified	Type	Size
case_lists	25/03/2022 19:07	File folder	
data_clinical_patient.txt	25/03/2022 19:15	TXT File	355 KB
data_clinical_sample.txt	25/03/2022 19:15	TXT File	103 KB
data_cna.txt	25/03/2022 19:15	TXT File	22,499 KB
data_linear_cna.txt	25/03/2022 19:15	TXT File	64,382 KB
data_methylation_hm450.txt	25/03/2022 19:15	TXT File	119,597 KB
data_mrna_seq_v2_rsem.txt	25/03/2022 19:15	TXT File	69,336 KB
data_mrna_seq_v2_rsem_zscores_ref_all_samples.txt	25/03/2022 19:15	TXT File	60,450 KB
data_mrna_seq_v2_rsem_zscores_ref_diploid_samples.txt	25/03/2022 19:15	TXT File	59,941 KB
data_mutations.txt	25/03/2022 19:15	TXT File	264,752 KB
data_mutsig.txt	25/03/2022 19:15	TXT File	2,083 KB
data_rppa.txt	25/03/2022 19:15	TXT File	643 KB
data_rppa_zscores.txt	25/03/2022 19:15	TXT File	569 KB
LICENSE	25/03/2022 19:07	File	1 KB
meta_clinical_patient.txt	25/03/2022 19:07	TXT File	1 KB
meta_clinical_sample.txt	25/03/2022 19:07	TXT File	1 KB
meta_cna.txt	25/03/2022 19:07	TXT File	1 KB
meta_linear_cna.txt	25/03/2022 19:07	TXT File	1 KB
meta_methylation_hm450.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem_zscores_ref_all_samples.txt	25/03/2022 19:07	TXT File	1 KB
meta_mrna_seq_v2_rsem_zscores_ref_diploid_samples.txt	25/03/2022 19:07	TXT File	1 KB
meta_mutations.txt	25/03/2022 19:07	TXT File	1 KB
meta_rppa.txt	25/03/2022 19:07	TXT File	1 KB
meta_rppa_zscores.txt	25/03/2022 19:07	TXT File	1 KB
meta_study.txt	25/03/2022 19:07	TXT File	1 KB

For each assay, 1 data file and 1 metadata/information file

Course material

Feedback form

elixir UNITED KINGDOM

Understanding the data

data_clinical_patient.txt	25/03/2022 19:15	TXT File	355 KB
data_clinical_sample.txt	25/03/2022 19:15	TXT File	103 KB

TSV – feature x patient ID (many missing values, cancer-specific features)

Patient information

Sex, height, weight, race, ethnicity, diagnosis age, survival status

Occupation history, smoking status, family history

Tumour information

Stage, grade, disease codes, metastasis status

Tumour-specific categories (e.g. for bladder, rate of prostate cancer)

Course material

Feedback form

elixir UNITED KINGDOM

Understanding the data

data_cna.txt	25/03/2022 19:15	TXT File	22,499 KB

tumour x gene using GISTIC scale (TSV)

- -2 homozygous "deep" deletion
- -1 shallow deletion (anything that isn't total loss)
- 0 diploid
- 1 gain ("a few" extra copies)
- 2 amplification (often in focal sets)

Course material

Feedback form

Understanding the data

data_mrna_seq_v2_rsem.txt	25/03/2022 19:15	TXT File	69,336 KB
data_mrna_seq_v2_rsem_zscores_ref_all_samples.txt	25/03/2022 19:15	TXT File	60,450 KB
data_mrna_seq_v2_rsem_zscores_ref_diploid_samples.txt	25/03/2022 19:15	TXT File	59,941 KB

tumour x gene, normalised gene expression data (TSV)

- Normalised counts
- Can feed into differential expression pipelines (DESeq2 *etc*), if careful!

- Good for comparisons of <u>one</u> gene across samples
- Harder to compare expression between genes of same sample

Course material

Feedback form

Understanding the data

data_mutations.txt	25/03/2022 19:15	TXT File	264,752 KB

TSV – list of all mutations, sorted by tumour ID

- Includes synonymous mutations as well as non-synonymous
- Data structure is rubbish, requires lots of parsing to find hotspots *etc.*

Course material

Feedback form

Working with the data

Existing UG training and extensive core bioinformatic support

Python support available too – pandas package is versatile

Doable…! But. Memory intensive, and watch delimiters when importing.

Course material

Feedback form

Working with the data… final thoughts

The data is not always complete

- Inconsistent column usage between datasets
- Watch 'whitespace' vs 'tab space' vs comma delimiters

Biological vs Statistical significance

Limited by previous bioinformatic analysis pipelines, genome version *etc.*

- More advanced questions can go back to the raw data

Course material

Feedback form

Introduction to cBioPortal

Course complete!

Feedback form

elixir UNITED KINGDOM