

Data level

Data level is a method of data categorization used within the TCGA network to facilitate researchers in communicating and locating their data of interest.

Data levels are assigned for each data type, platform and center. There are four data levels: Level 1 (for Raw Data), Level 2 (for Processed Data), Level 3 (for Segmented or Interpreted Data) and Level 4 (for Region of Interest Data).

Contents of this Page

- [Data Level Classification](#)
- [Relationships Between Data Type and Data Level](#)

Data Level Classification

Data level distinguishes raw data from derived data, from higher-level analysis or interpreted results for each data type, platform, and center.

The following table lists and describes each TCGA data level.

Data Level	Level Type	Description	Example
1	Raw	<ul style="list-style-type: none">• Low-level data for single sample• Not normalized	<ul style="list-style-type: none">• Sequence trace file• Affymetrix CEL file¹• BAM file
2	Processed	<ul style="list-style-type: none">• Normalized single sample data• Interpreted for presence or absence of specific molecular abnormalities	<ul style="list-style-type: none">• Putative mutation call for a single sample• Probed locus amplification/deletion/Loss of Heterozygosity (LOH) calls in a sample• Signal of a probe or probe set for a sample
3	Segmented/Interpreted	<ul style="list-style-type: none">• Aggregate of processed data from single sample• Grouped by probed loci to form larger contiguous regions (in some cases)	<ul style="list-style-type: none">• Validated mutation call for a single sample• Amplification/deletion/Loss of Heterozygosity (LOH) calls for a sample region• Expression signal of a gene for a sample• Genomic copy-number data
4	Summary/Regions of Interest (ROI)	<ul style="list-style-type: none">• Quantified association across classes of samples• Associations based on two or more<ul style="list-style-type: none">• Molecular abnormalities• Sample characteristics• Clinical variables	<ul style="list-style-type: none">• Discovery that a genomic region is amplified in 10% of TCGA glioma samples.

Relationships Between Data Type and Data Level

Each platform can produce multiple data types. To understand data categorization, it is important to clarify the relationship between data type and data level.

Each data type is associated with sets of data that span one more data levels. Each center and platform may have a slightly different concept of data level depending on their data types, and the algorithms used for analysis.

For descriptions of data types and corresponding data levels see [Data Types and Data Levels](#).