

META DATA FOR DATASETDD.CSV

1. "Case UUID"- UUIDs are generally used for identifying information that needs to be unique within a system or network thereof
2. "Case ID" - ID of all the subjects
3. "Project" -The organization which undertook this project
4. "Primary Site" - The primary site of tumor growth
5. "Gender" -Gender is defined as a set of characteristics that distinguish persons based on their social roles.
6. "Seq"- Number of the sequencing reads (an inferred sequence of base pairs)
7. "Exp" – The transcriptome profiling
8. "SNV" – simple nucleotide variation
9. "CNV"-Copy number variation, the number of copies of a particular gene varies from one individual to the next
- 10."Meth" -Number of DNA methylation that took place
- 11."Clinical" – The number of clinical alterations
- 12."Bio" -biospecimen
- 13."Mutations" -The number of mutations the subject's tumor site had
- 14."Genes" - The genes altered in the subjects.
- 15."Program" – The type of program the subject is enrolled in "The Cancer Genome Atlas" or the "count me in"
- 16."Disease Type" – This describes if it is Adenomas and Adenocarcinomas, the location where the cancer is formed.
- 17."Age at diagnosis" -Age at the time of diagnosis calculated from their day of birth to the day they were diagnosed of cancer.
- 18."Days to death" -The date from when they were diagnosed, and they died
- 19."Vital Status" – the survival rate of the person
- 20."Primary Diagnosis"- Their initial pathological diagnoses of cancer.
- 21."Ethnicity" -An individual's self-identified social and cultural category, particularly whether they identify as Hispanic or Latino. The provided figures are based on the US Office of Management and Business's categories, which are used by the US Census Bureau.
- 22."Race" -A taxonomic group that is a division of a species that is classified arbitrarily. It is characterized by shared heredity, physical attributes, and behavior, and in the case of humans, by common history, nationality, or geographic distribution. It usually arises because of geographical isolation

within a species and is characterized by shared heredity, physical attributes, and behavior. The provided figures are based on the US Office of Management and Business's categories, which are used by the US Census Bureau.