



The Cancer Genome Atlas

--TCGA 简介

1、背景介绍

目前有至少 200 种形式的癌症，以及众多亚型。这些都是由于细胞中 DNA 突变引起的细胞生长失控。识别每一个癌症一整套的 DNA 变异情况，以及了解这些变化的相互作用，为提高癌症预防，早期的诊断和治疗提供基础。

The Cancer Genome Atlas (TCGA)是由美国国家人类基因组研究所和美国国家癌症研究所与 2005 年共同斥资建立的项目，对人类目前多种癌症进行了大量的基因组测序工作，已被社会各界广泛引用，发表了 1000 多篇刊物其中不乏 nature 等重量级的文章。

TCGA 创立的使命是通过各方面综合努力，包括大规模的基因组测序分析等手段来探究癌症治病的分子层面的原理。旨在对人类目前的 33 例癌症的突变频谱框架进行研究，希望在精准医疗方向有更大的进展，实际上也确实意义非凡，众多药物公司根据 TCGA 的数据，开发出多种靶向药物。

2、数据产生历程

TCGA 创造了一个集获取，筛选，分析于一体的网络，可以大规模地对人类的癌症组织进行研究。

TCGA 下属的几个部门：

1、**Biospecimen Core Resource (BCR)** – Tissue samples were carefully cataloged, processed, checked for quality and stored, complete with important medical information about the patient.

2、**Genome Characterization Centers (GCCs)** – The Genome Characterization centers used several technologies to analyze genomic changes involved in cancer including gene expression levels and structural rearrangements of the genome.

3、**Genome Sequencing Centers (GSCs)**– High-throughput Genome Sequencing Centers identified the changes in DNA sequences that are associated with specific types of cancer.

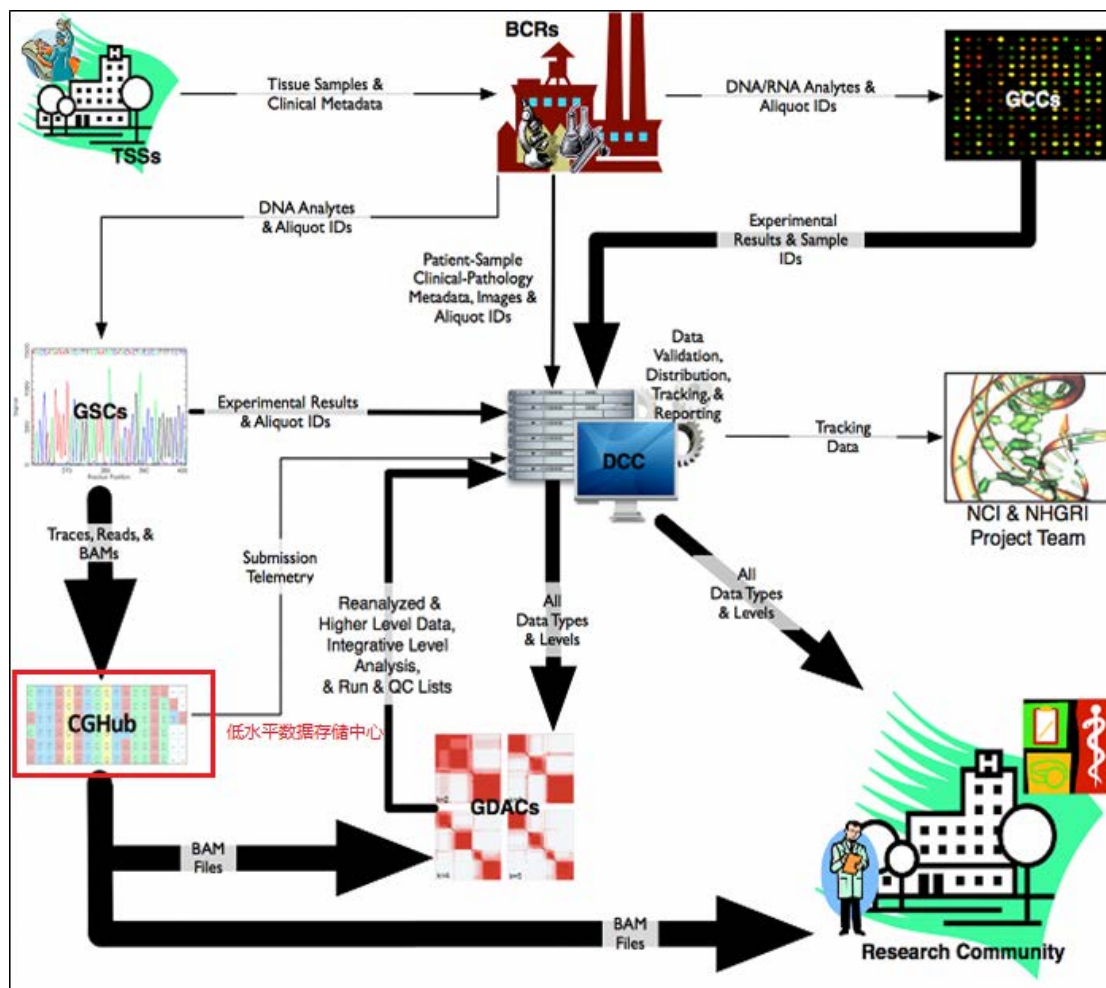
4、**Proteome Characterization Centers (PCCs)**– The centers, a component of NCI's Clinical Proteomic Tumor Analysis Consortium, analyzed the total proteomic content of a subset of TCGA samples.

5、**Data Coordinating Center (DCC) and Cancer Genomics Hub (CGHub)**– The information that was generated by TCGA was centrally managed at the DCC and entered into the TCGA Data Portal and Cancer Genomics Hub as it became available. The data is now stored and distributed by the NCIGenomic Data Commons.

6、**Genome Data Analysis Centers (GDACs)**– GDACs integrated immense

amounts of data from array and sequencing technologies across thousands of samples. These centers provided novel informatics tools to the entire research community to facilitate broader use of TCGA data.

7、**Analysis Working Groups (AWGs)**– AWGs are interdisciplinary, international groups of scientists that perform a global, integrative analysis on each TCGA tumor type. Every AWG studies a particular tumor type using all of the TCGA platforms and publishes an analysis of their findings in a peer-reviewed journal to benefit the cancer research and clinical communities.



TCGA 数据的产生历程:

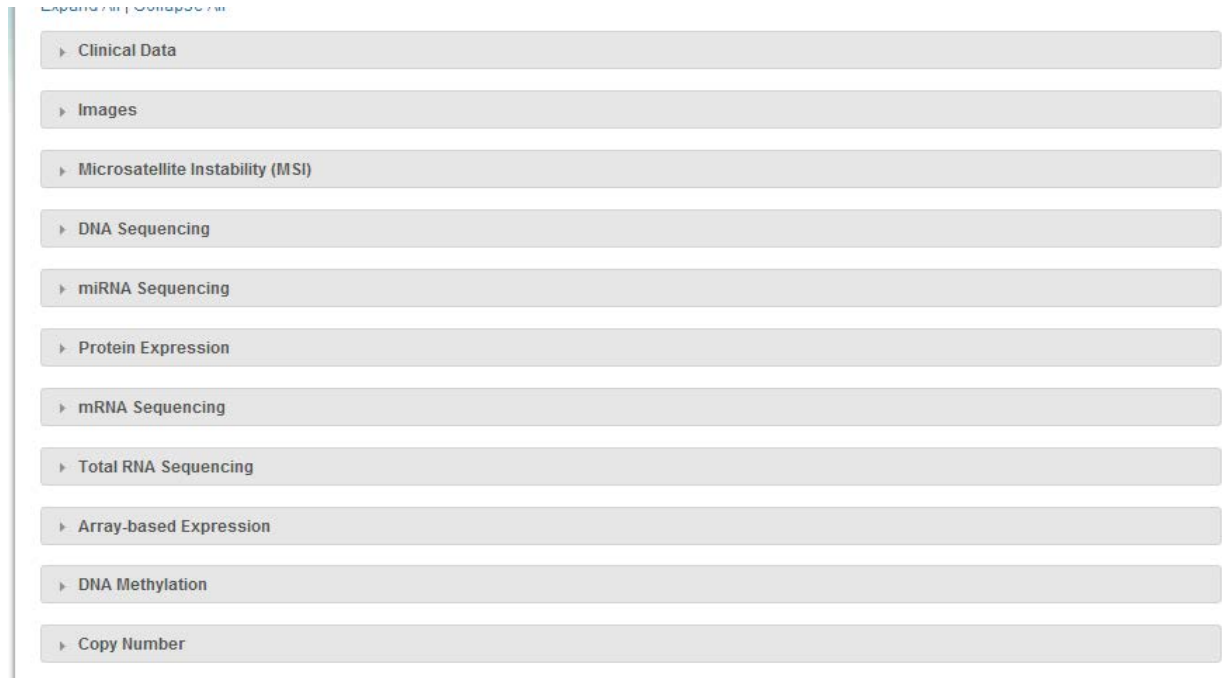
1. 组织样本和临床数据来源网站(TSS)收集的数据发送到 Biospecimen 核心资源 (BCRs)。
2. BCRs 提交临床数据和元数据到数据协调中心(DCC)和测序中心(GSCs),获取组织变异数据,然后提交给 DCC。
3. GSCs 提交跟踪文件,序列比对后文件到癌症基因组学中心(CGHub)。
4. 提交给 DCC 和 CGHub 的数据可供研究团队和基因组数据分析中心(GDACs)使用。
5. 分析 pipeline 以及由 GDACs 产生的数据结果通过 DCC 保存到研究社区 (community) 中。

3、目前已包含癌症种类

Available Cancer Types	# Cases Shipped by BCR*	# Cases with Data*	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [LAML]	200	200	04/29/15
Adrenocortical carcinoma [ACC]	80	80	08/31/15
Bladder Urothelial Carcinoma [BLCA]	412	412	08/31/15
Brain Lower Grade Glioma [LGG]	516	516	09/04/15
Breast invasive carcinoma [BRCA]	1100	1098	09/04/15
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	308	308	09/04/15
Cholangiocarcinoma [CHOL]	36	36	09/04/15
Colon adenocarcinoma [COAD]	461	461	08/31/15
Esophageal carcinoma [ESCA]	185	185	08/31/15
FFPE Pilot Phase II [FPPP]	38	38	09/04/15
Glioblastoma multiforme [GBM]	529	528	09/03/15
Head and Neck squamous cell carcinoma [HNSC]	528	528	09/04/15
Kidney Chromophobe [KICH]	66	66	08/31/15
Kidney renal clear cell carcinoma [KIRC]	536	536	09/04/15
Kidney renal papillary cell carcinoma [KIRP]	291	291	09/04/15
Liver hepatocellular carcinoma [LIHC]	377	377	09/04/15
Lung adenocarcinoma [LUAD]	521	521	08/31/15
Lung squamous cell carcinoma [LUSC]	510	504	09/04/15
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBC]	48	48	08/31/15
Mesothelioma [MESO]	87	87	09/02/15
Ovarian serous cystadenocarcinoma [OV]	586	586	09/04/15
Pancreatic adenocarcinoma [PAAD]	185	185	08/31/15
Pheochromocytoma and Paraganglioma [PCPG]	179	179	09/04/15
Prostate adenocarcinoma [PRAD]	498	498	08/31/15
Rectum adenocarcinoma [READ]	172	171	09/04/15
Sarcoma [SARC]	261	261	08/31/15
Skin Cutaneous Melanoma [SKCM]	470	470	08/31/15
Stomach adenocarcinoma [STAD]	445	443	08/31/15
Testicular Germ Cell Tumors [TGCT]	150	150	08/31/15
Thymoma [THYM]	124	124	09/04/15
Thyroid carcinoma [THCA]	507	507	08/31/15
Uterine Carcinosarcoma [UCS]	57	57	09/04/15
Uterine Corpus Endometrial Carcinoma [UCEC]	548	548	09/04/15
Uveal Melanoma [UVM]	80	80	09/04/15

4、 数据种类和等级

数据种类：



为了方便数据的分类及隐私管理，所有种类的数据实行等级控制，只有 open 的数据才对外开放：

这个网址中有对相应数据等级的划分

<https://tcga-data.nci.nih.gov/docs/publications/tcga/datatype.html>