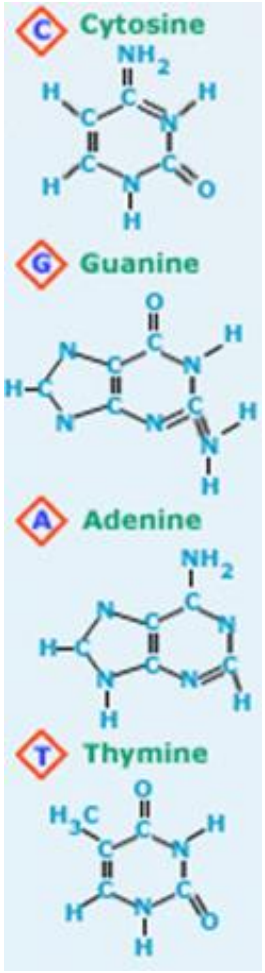
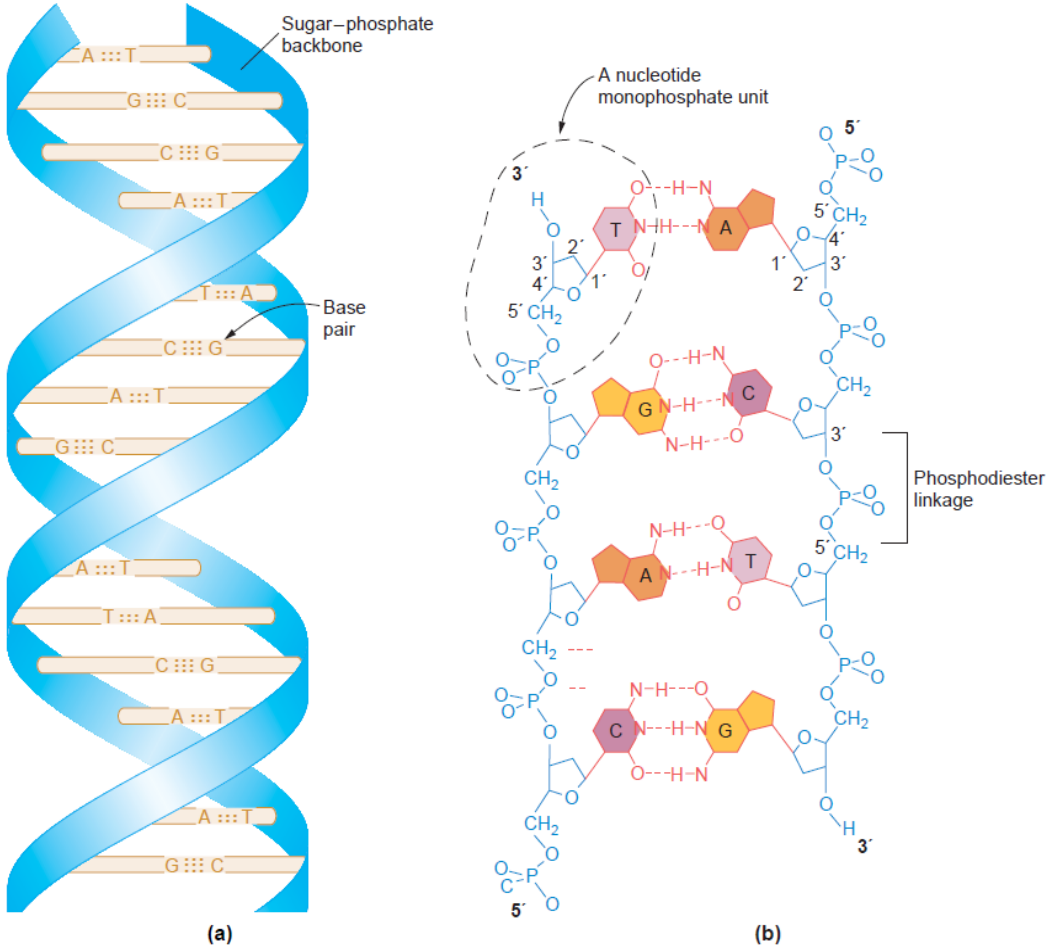
Project : TCGA data analysis

DNA



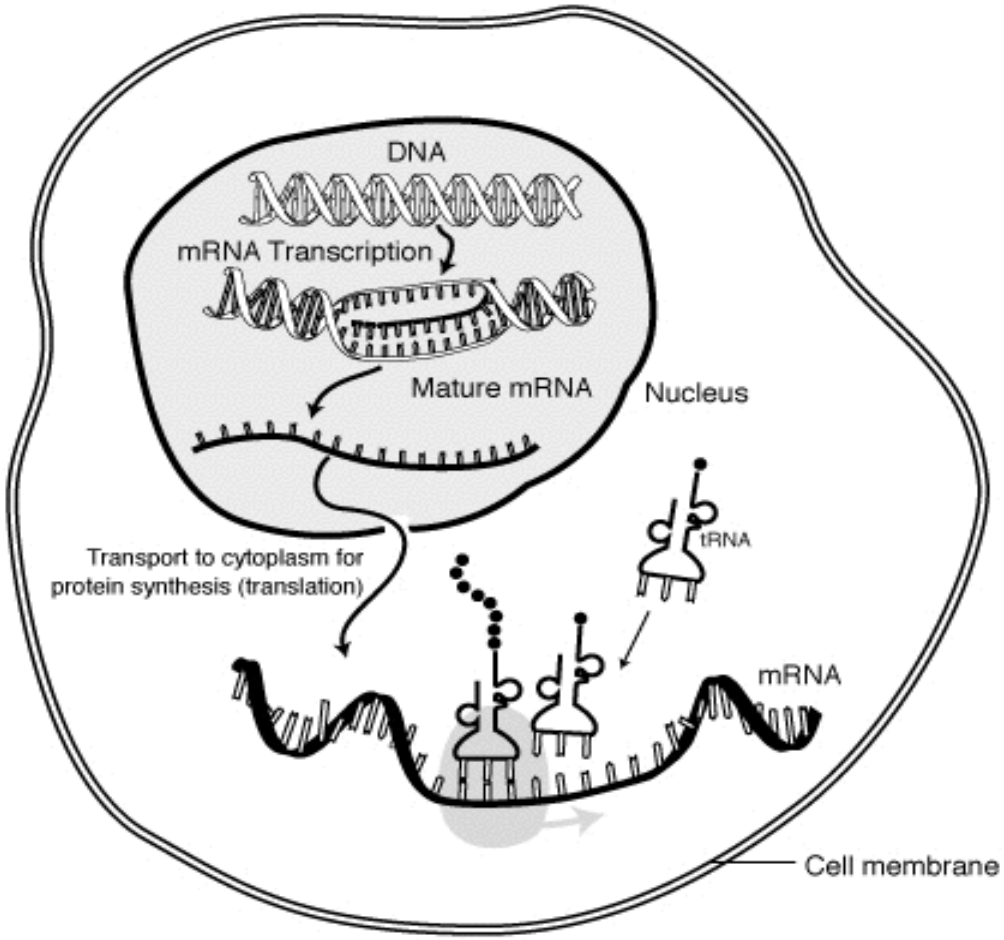
http://esheets.biz/dexterity-dna-and-rna-worksheet/step-nucleic-acids-dna-and-rna/



Source: Introduction to Genetic Analysis, Griths et al., 2008.

DNA to Protein

- DNA
  (map of human body)
- mRNA
  (transcription)
- Amino acid
  (translation)
- Protein
- miRNA

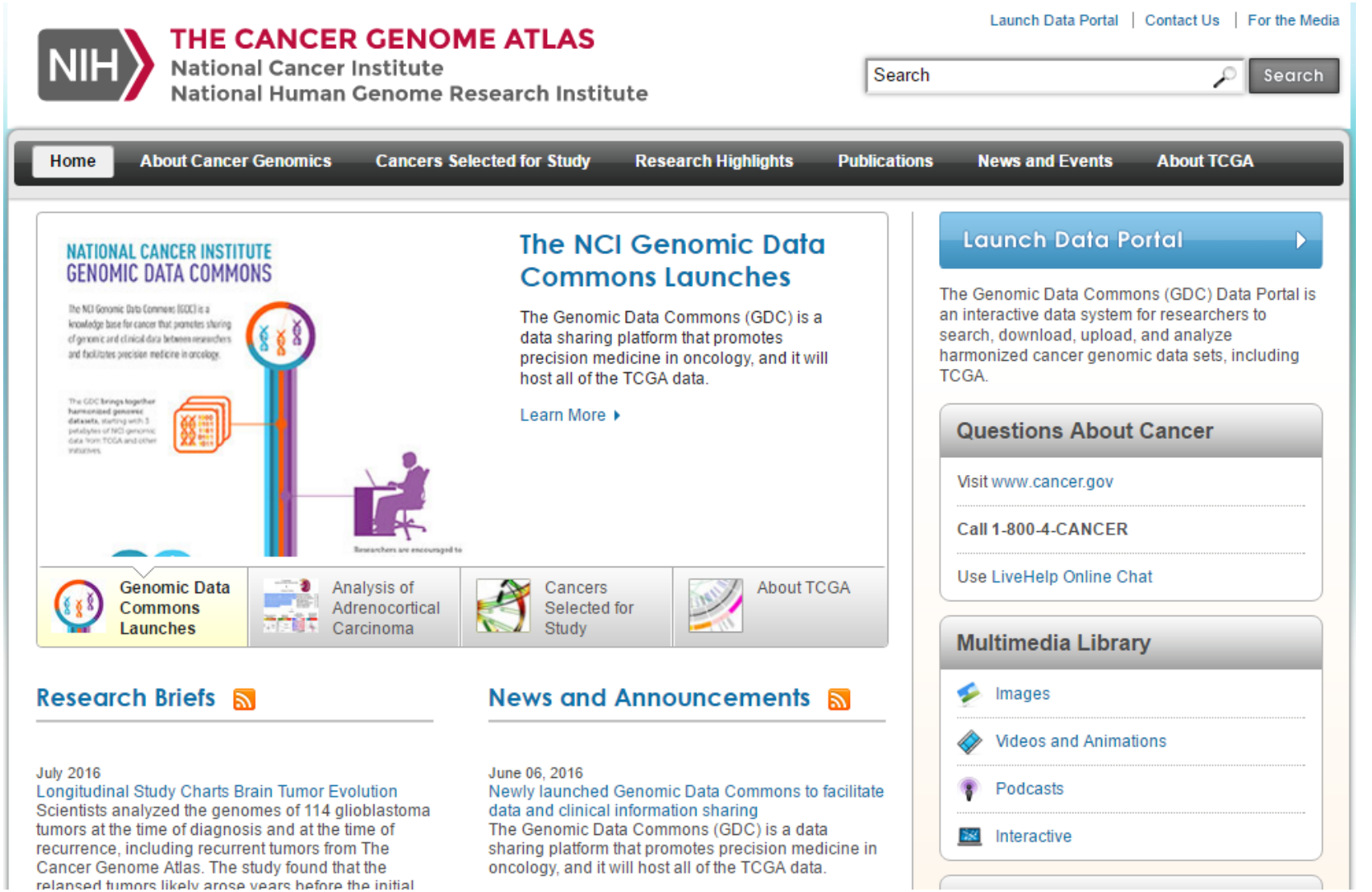


Source: http://en.wikipedia.org/wiki/File:MRNA-interaction.png
(genome.gov, 2004)

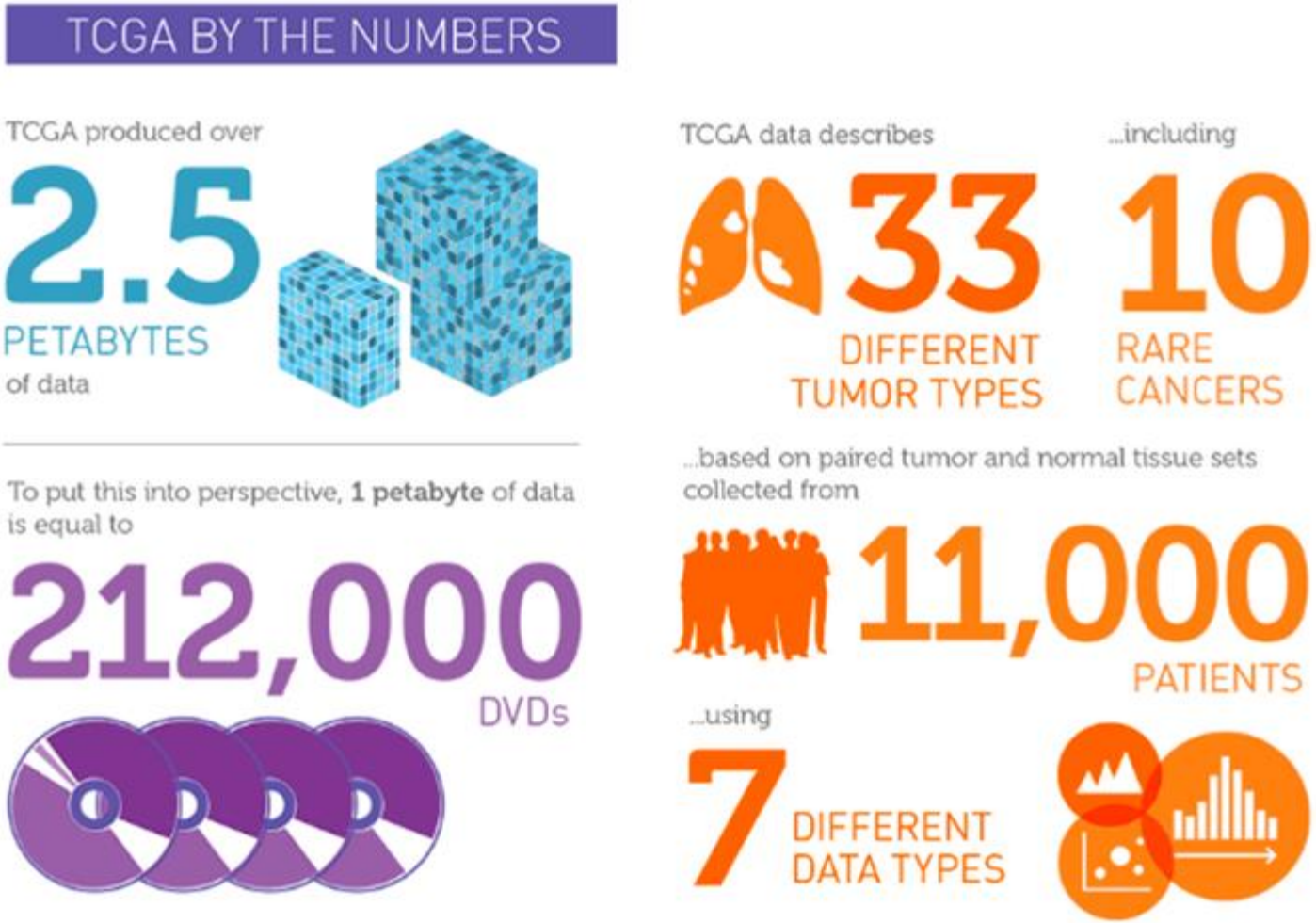
DNA to Protein

https://www.youtube.com/watch?v=gG7uCskUOrA

TCGA data

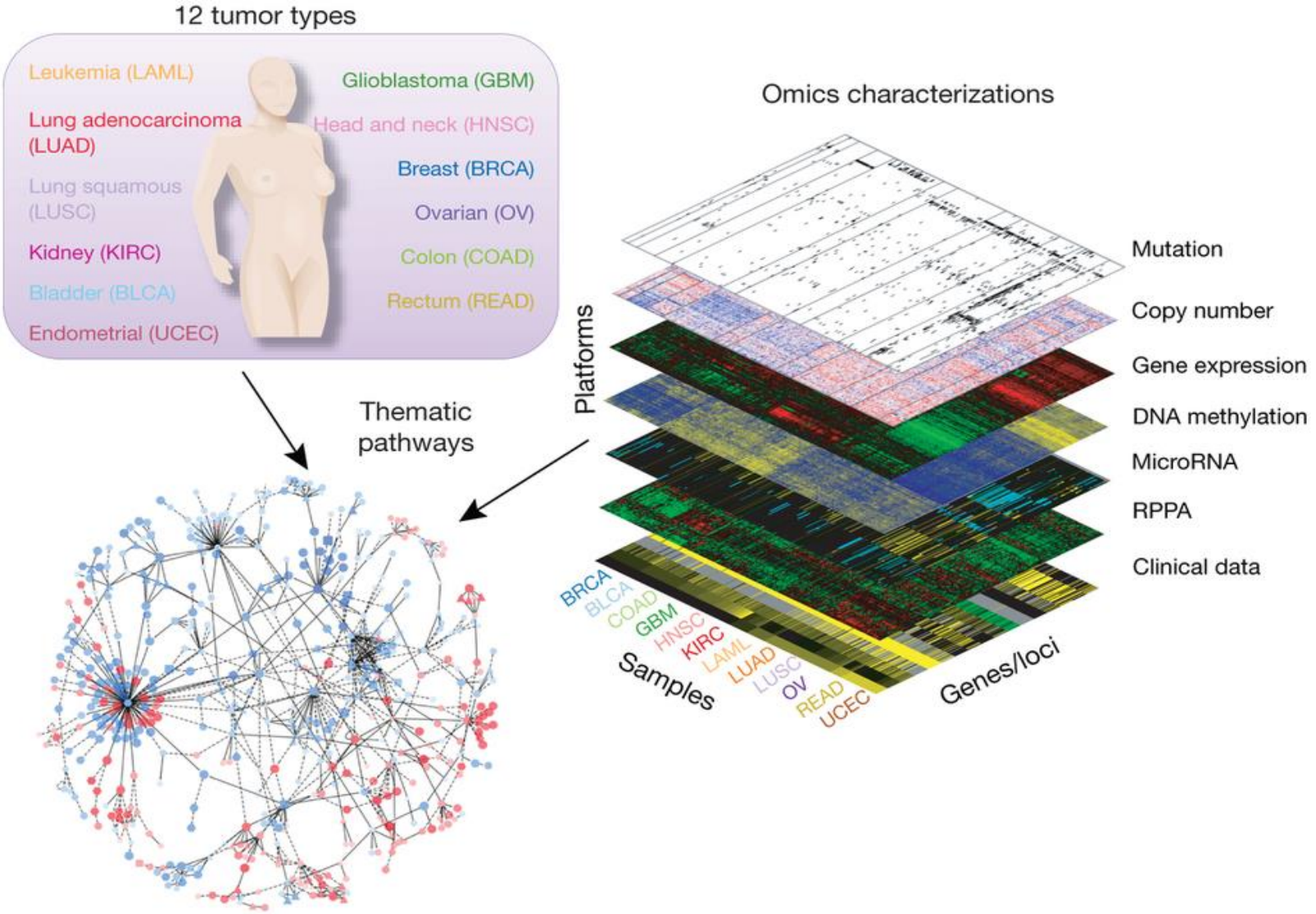


TCGA data



http://cancergenome.nih.gov/abouttcga/overview

TCGA data : the TCGA Pan-Cancer project

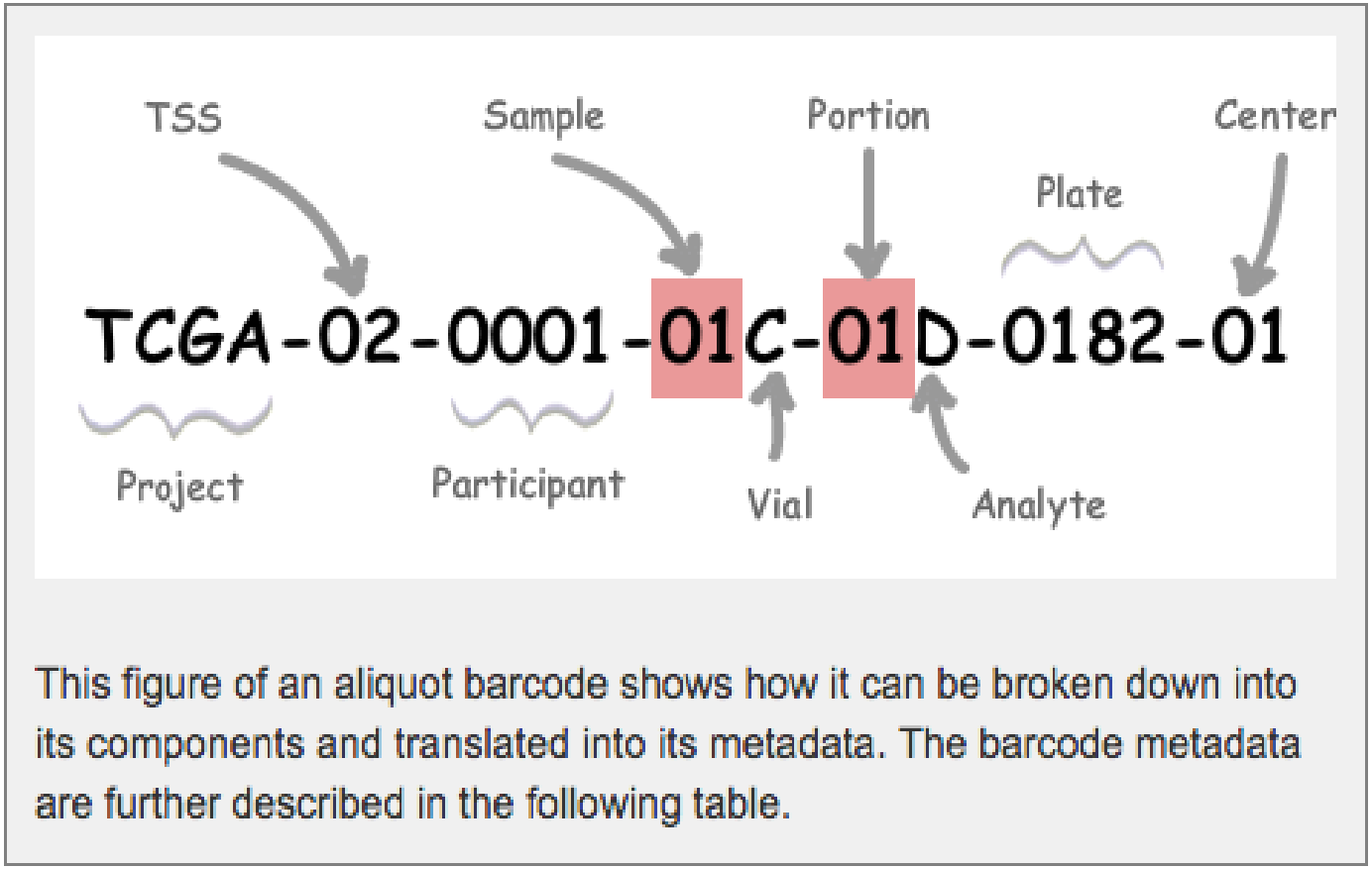


http://www.nature.com/ng/journal/v45/n10/fig_tab/ng.2764_F1.html (Weinstein et al., 2013)

TCGA data : sample ID

sample types 0
1 (solid tumor n
ormal)

11 (solid tissue
normal)



This figure of an aliquot barcode shows how it can be broken down into its components and translated into its metadata. The barcode metadata are further described in the following table.

https://wiki.nci.nih.gov/display/TCGA/TCGA+barcode

TCGA data : sample ID

Label	Identifier for	Value	Value description	Possible values
Project	Project name	TCGA	TCGA project	TCGA
TSS	Tissue source site	02	GBM (brain tumor) sample from M D Anderson	See Code Tables Report
Participant	Study participant	0001	The first participant from MD Ande rson for GBM study	Any alpha-numeric value
Sample	Sample type	01	A solid tumor	Tumor types range from 01 - 09, normal types from 10 – 19, and control samples from 20 - 29. See Code Tables Report for a compl ete list of sample codes
Vial	Order of sample in a sequence of sa mples	C	The third vial	A to Z
Portion	Order of portion in a sequence of 10 0 - 120 mg sample portions	01	The first portion of the sample	01-99
Analyte	Molecular type of analyte for analysis	D	The analyte is a DNA sample	See Code Tables Report
Plate	Order of plate in a sequence of 96-w ell plates	0182	The 182nd plate	4-digit alphanumeric value
Center	Sequencing or characterization cent er that will receive the aliquot for ana lysis	01	The Broad Institute GCC	See Code Tables Report

https://gdac.broadinstitute.org/

BROAD INSTITUTE

FIREHOSE Broad GDAC

Liberating scientific knowledge from data, at scale

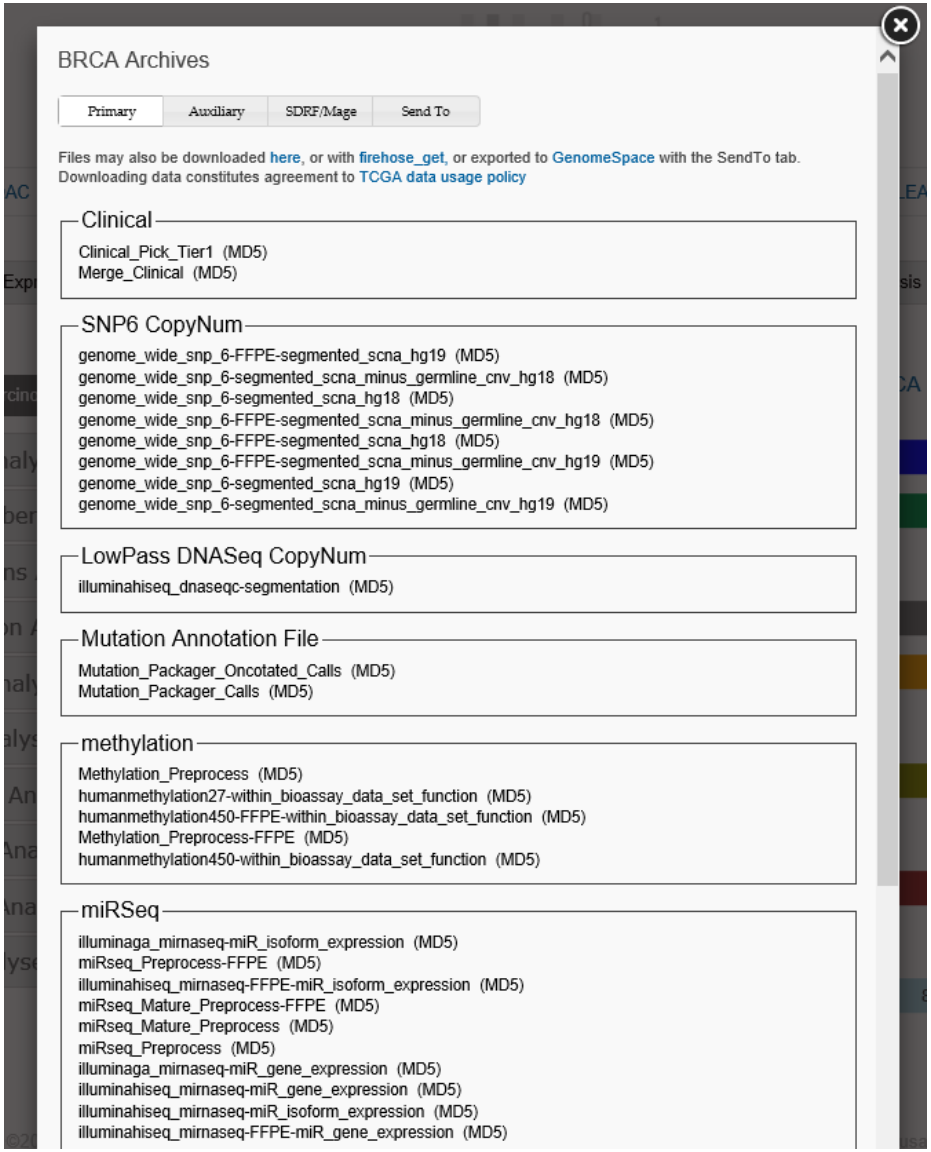
About FireBrowse Dashboards Data Analyses Software Documentation FAQ Download Contact Us What's New?

Disease Name	Cohort	Cases	Analyses	Data
Adrenocortical carcinoma	ACC	92	Browse	Browse
Bladder urothelial carcinoma	BLCA	412	Browse	Browse
Breast invasive carcinoma	BRCA	1098	Browse	Browse
Cervical and endocervical cancers	CESC	307	Browse	Browse
Cholangiocarcinoma	CHOL	51	Browse	Browse
Colon adenocarcinoma	COAD	460	Browse	Browse
Colorectal adenocarcinoma	COADREAD	631	Browse	Browse
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	58	Browse	Browse
Esophageal carcinoma	ESCA	185	Browse	Browse
FFPE Pilot Phase II	FPPP	38	None	Browse
Glioblastoma multiforme	GBM	613	Browse	Browse
Glioma	GBMLGG	1129	Browse	Browse
Head and Neck squamous cell carcinoma	HNSC	528	Browse	Browse
Kidney Chromophobe	KICH	113	Browse	Browse
Pan-kidney cohort (KICH+KIRC+KIRP)	KIPAN	973	Browse	Browse
Kidney renal clear cell carcinoma	KIRC	537	Browse	Browse
Kidney renal papillary cell carcinoma	KIRP	323	Browse	Browse
Acute Myeloid Leukemia	LAML	200	Browse	Browse
Brain Lower Grade Glioma	LGG	516	Browse	Browse
Liver hepatocellular carcinoma	LIHC	377	Browse	Browse
Lung adenocarcinoma	LUAD	585	Browse	Browse
Lung squamous cell carcinoma	LUSC	504	Browse	Browse
Mesothelioma	MESO	87	Browse	Browse
Ovarian serous cystadenocarcinoma	OV	602	Browse	Browse
Pancreatic adenocarcinoma	PAAD	185	Browse	Browse
Pheochromocytoma and Paraganglioma	PCPG	179	Browse	Browse
Prostate adenocarcinoma	PRAD	499	Browse	Browse
Rectum adenocarcinoma	READ	171	Browse	Browse
Sarcoma	SARC	261	Browse	Browse
Skin Cutaneous Melanoma	SKCM	470	Browse	Browse
Stomach adenocarcinoma	STAD	443	Browse	Browse
Stomach and Esophageal carcinoma	STES	628	Browse	Browse
Testicular Germ Cell Tumors	TGCT	150	Browse	Browse
Thyroid carcinoma	THCA	503	Browse	Browse
Thymoma	THYM	124	Browse	Browse
Uterine Corpus Endometrial Carcinoma	UCEC	560	Browse	Browse

https://gdac.broadinstitute.org/

- Colon adenocarcinoma / COAD / 460
- Brain Lower Grade Glioma / LGG / 516
- Liver hepatocellular carcinoma / LIHC / 377
- Stomach adenocarcinoma / STAD / 443
- Lung adenocarcinoma / LUAD / 585
- Thyroid carcinoma / THCA / 503
- Skin Cutaneous Melanoma / SKCM / 470

Download data (Example: BRCA)



Download data : Next Generation Sequencing Technology

MiniSeq System
Power and simplicity for targeted sequencing.

MiSeq Series
Small genome and targeted sequencing.

NextSeq Series
Everyday genome, exome transcriptome sequencing, and more.

HiSeq Series
Production-scale genome, exome, transcriptome sequencing and more.

HiSeq X Series
Population- and production-scale human whole-genome sequencing.

www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Download data

Clinical

Clinical_Pick_Tier1 (MD5)

mRNASeq

illuminahiseq_rnaseqv2-RSEM_genes_normalized (MD5)

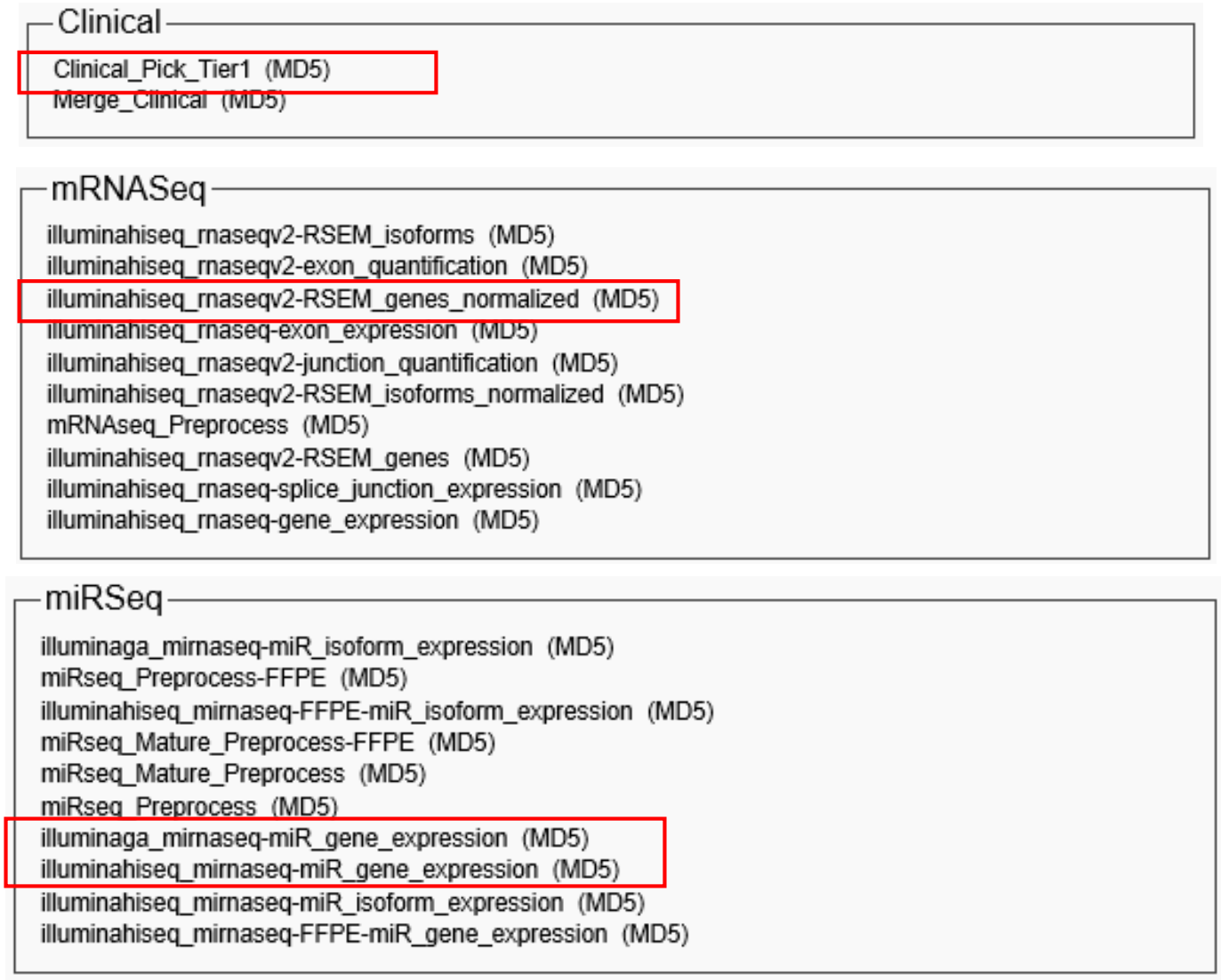
illuminaga_rnaseqv2-RSEM_genes_normalized (MD5)

miRSeq

illuminahiseq_mirnaseq-miR_gene_expression (MD5)

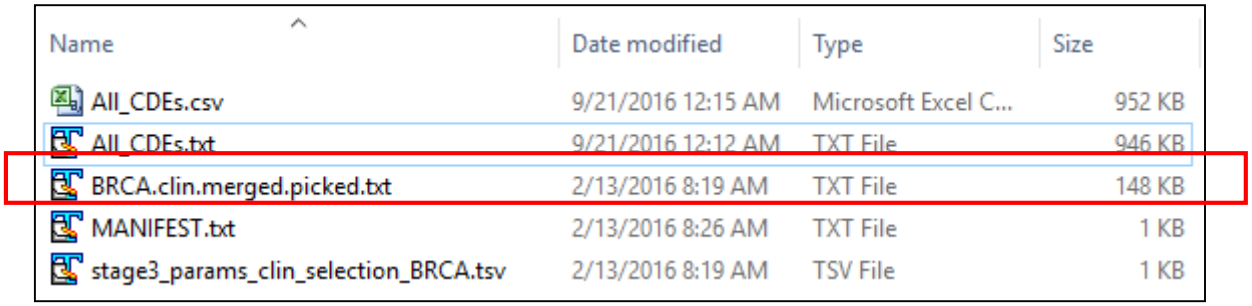
illuminaga_mirnaseq-miR_gene_expression (MD5)

Download data (Example: BRCA)



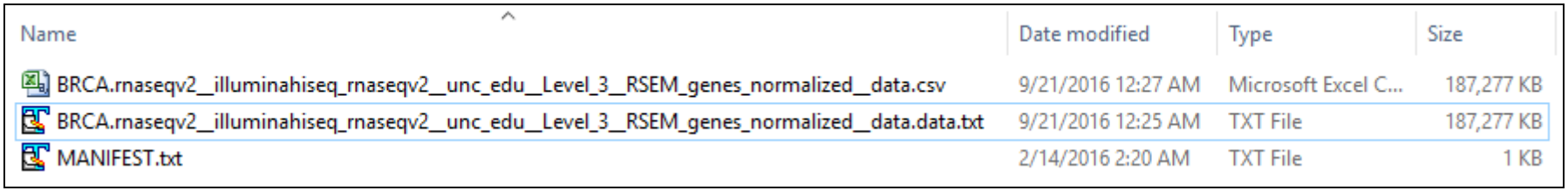
Download data : after extracting the compressed files

Clinical



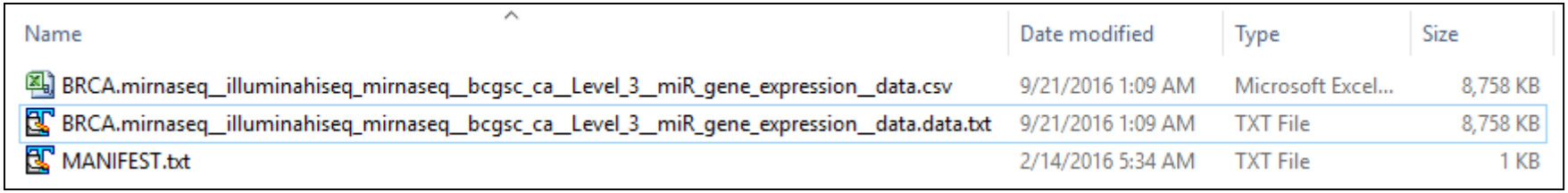
Name	Date modified	Type	Size
All_CDEs.csv	9/21/2016 12:15 AM	Microsoft Excel C...	952 KB
All_CDEs.txt	9/21/2016 12:12 AM	TXT File	946 KB
BRCA.clin.merged.picked.txt	2/13/2016 8:19 AM	TXT File	148 KB
MANIFEST.txt	2/13/2016 8:26 AM	TXT File	1 KB
stage3_params_clin_selection_BRCA.tsv	2/13/2016 8:19 AM	TSV File	1 KB

mRNASeq



Name	Date modified	Type	Size
BRCA.rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__RSEM_genes_normalized__data.csv	9/21/2016 12:27 AM	Microsoft Excel C...	187,277 KB
BRCA.rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__RSEM_genes_normalized__data.data.txt	9/21/2016 12:25 AM	TXT File	187,277 KB
MANIFEST.txt	2/14/2016 2:20 AM	TXT File	1 KB

miRSeq



Name	Date modified	Type	Size
BRCA.mirnaseq__illuminahiseq_mirnaseq__bcgsc_ca__Level_3__miR_gene_expression__data.csv	9/21/2016 1:09 AM	Microsoft Excel...	8,758 KB
BRCA.mirnaseq__illuminahiseq_mirnaseq__bcgsc_ca__Level_3__miR_gene_expression__data.data.txt	9/21/2016 1:09 AM	TXT File	8,758 KB
MANIFEST.txt	2/14/2016 5:34 AM	TXT File	1 KB

Q & A