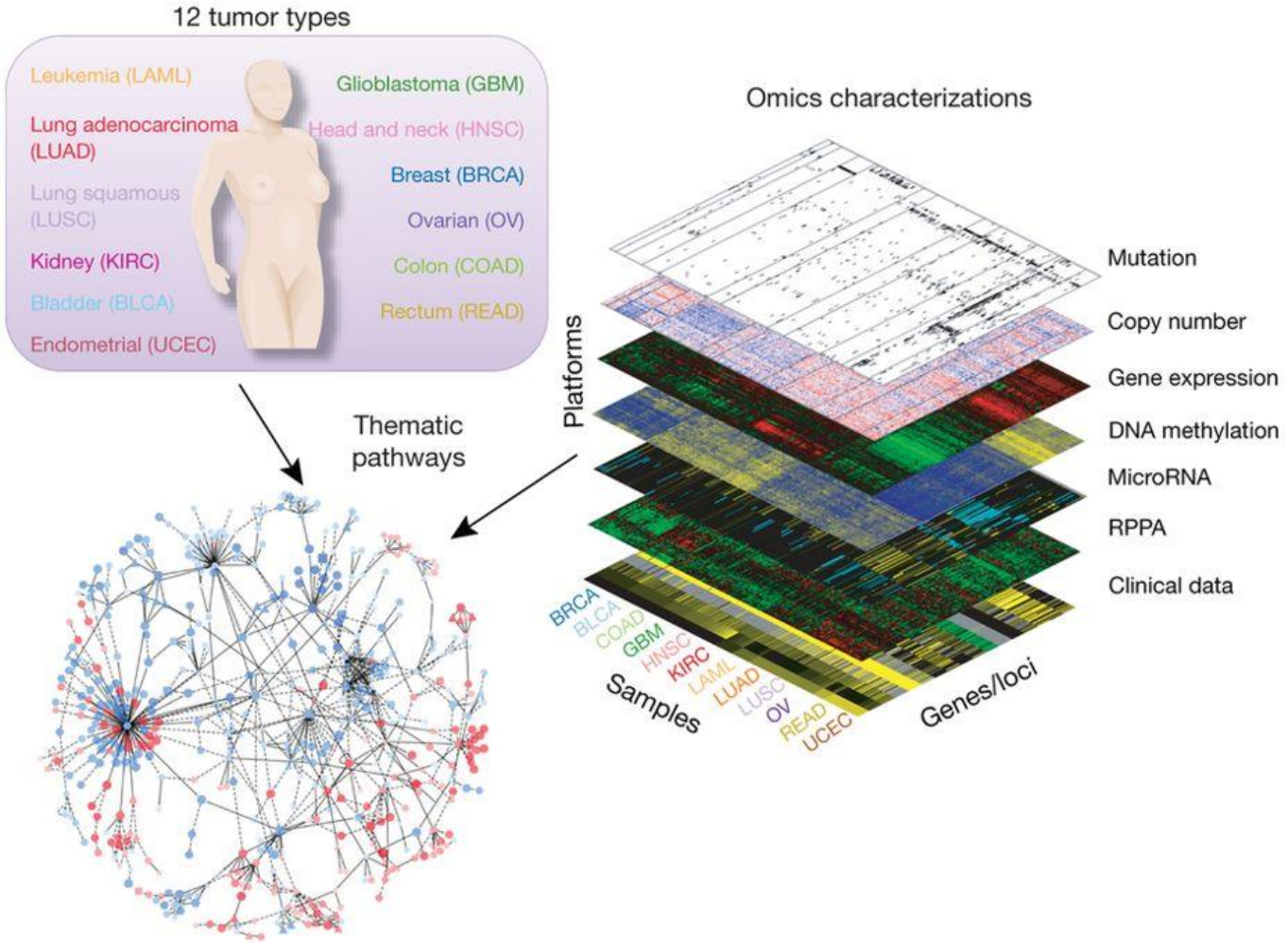
The Cancer Genome Atlas Pan-Cancer analysis project

TCGA

- The Cancer Genome Atlas
  - Systematic cancer genomic project

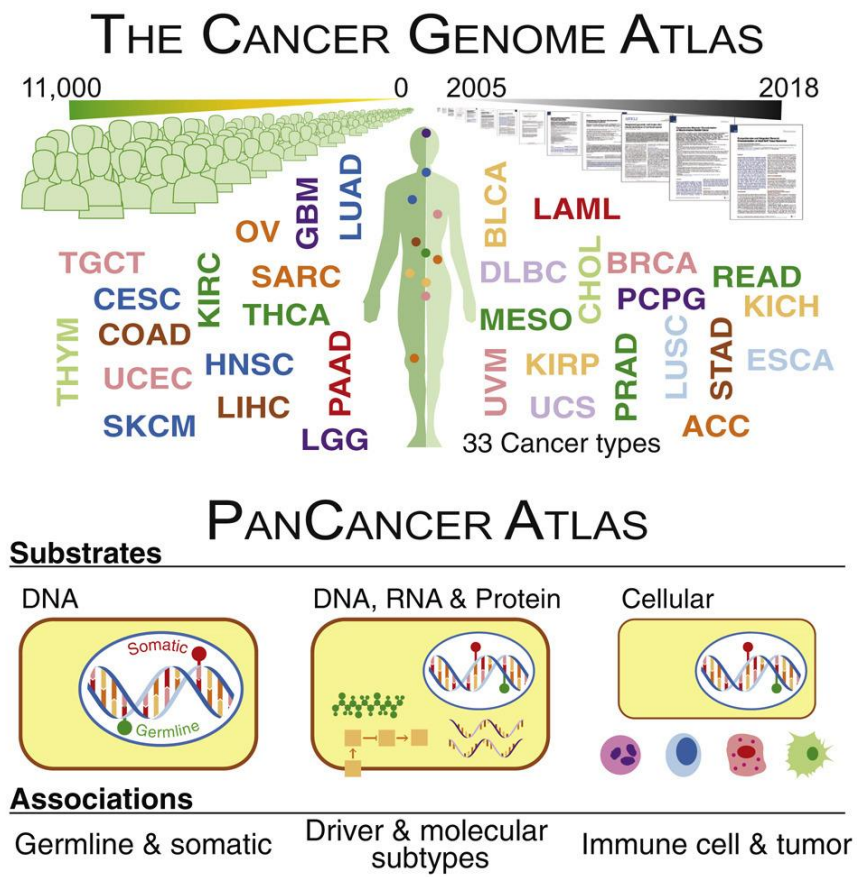
- Emerging technologies
  - Analysis of specific tumor types
  - Identification of novel oncogenic drivers & establishing of subtypes
  - Identification of new biomarkers on the basis of genomic, transcriptomic, proteomic and epigenomic alterations

The Pan-Cancer project

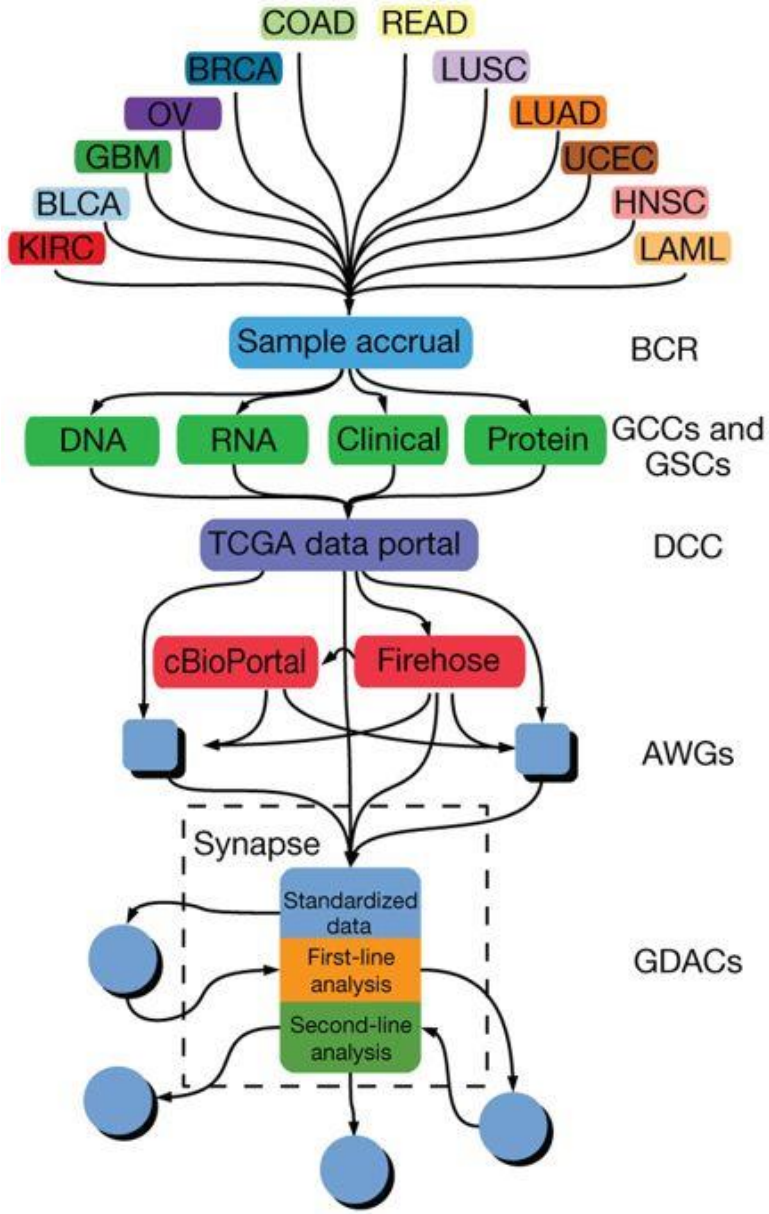


Stuart et al.

Pan-Project Workflow



Mariamidze et al.



Stuart et al.

Pan-Cancer Project

- Goals
  - Assemble of coherent, consistent TCGA data sets across tumor types
- Conclusions so far
  - Sample have distinct alterations not shared with other samples of their cohort
  - Tumors in the same tissue may not resemble each other, similar features are found across different tissue-specific tumors.
- Challenges
  - Integration of Data (generated of different platforms/ versions)
  - Bias caused by unwanted batch-effects

Outlook

- In April 2018, TCGA Research Network marked the end of the TCGA
  - Publishing the Pan-Cancer Atlas
- 11 000 tumors from 33 forms of cancer
- Reclassification of human tumors based on molecular similarity

NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

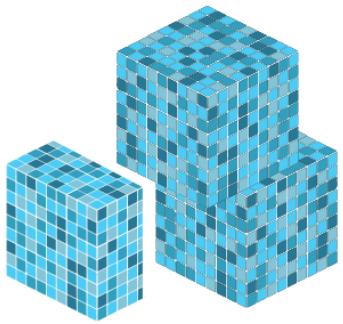
TCGA BY THE NUMBERS

TCGA produced over

2.5

PETABYTES

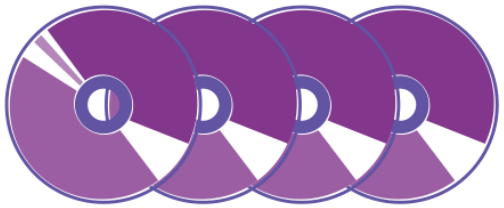
of data



To put this into perspective, **1 petabyte** of data is equal to

212,000

DVDs



TCGA data describes

33

DIFFERENT TUMOR TYPES

...including

10

RARE CANCERS

...based on paired tumor and normal tissue sets collected from

11,000

PATIENTS

...using

7

DIFFERENT DATA TYPES

WHAT'S NEXT?

The Genomic Data Commons (GDC) houses TCGA and other NCI-generated data sets for scientists to access from anywhere. The GDC also has many expanded capabilities that will allow researchers to answer more clinically relevant questions with increased ease.

www.cancer.gov/ccg