The Cancer Genome Atlas (TCGA) &
International Cancer Genome Consortium (ICGC)

Session 4 – Rebecca Poulos

Introductory bioinformatics for human genomics workshop, UNSW
28th – 29th January 2016
Facts on cancer

➢ An estimated 126,800 new cases of cancer will be diagnosed in Australia this year, with that number set to rise to 150,000 by 2020

➢ Cancer is a leading cause of death in Australia. In 2012, > 43,000 people died from cancer, accounting for about 3 in every 10 deaths.

Source: Cancer Council Australia (2016)
Cancer is a disease of the genome

- Challenges in treating cancer:
  - Every patient is different
  - Every tumour is different, even in the same patient
  - Tumours can be highly heterogeneous
  - High rate of genomic abnormalities (few drivers, many passenger)
What can go wrong in cancer genomes?

<table>
<thead>
<tr>
<th>Types of changes</th>
<th>Some common technologies used to study these changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA mutations</td>
<td>WGS; WXS</td>
</tr>
<tr>
<td>- Point mutations</td>
<td></td>
</tr>
<tr>
<td>- Insertions &amp; deletions</td>
<td></td>
</tr>
</tbody>
</table>

WGS = whole genome sequencing, WXS = whole exome sequencing
RRBS = reduced representation bisulfite sequencing, WGBS = whole genome bisulfite sequencing
Goal of cancer genomics

• Identify changes in the genomes of tumors that drive cancer progression

• Identify new targets for therapy

• Select drugs based on the genomics of the tumour – i.e. personalised therapy
Cancer Sequencing Projects

The Cancer Genome Atlas (TCGA)

- Led by NIH
- Initiated in 2006 (as a pilot program) and expanded in 2009

- Aim:
  To make the genomes of 20 cancers publically available

- Update today:
  33 cancer types & subtypes analysed (11,000 samples)
TCGA pipeline

Publically available for researchers
Types of Cancers

- **Breast**
  - Ductal carcinoma
  - Lobular carcinoma

- **Central nervous system**
  - Glioblastoma multiforme
  - Lower grade glioma

- **Endocrine**
  - Adrenocortical carcinoma
  - Papillary thyroid carcinoma
  - Paraganglioma and pheochromocytoma

- **Gastrointestinal**
  - Cholangiocarcinoma
  - Colorectal Adenocarcinoma
  - Liver Hepatocellular Carcinoma
  - Pancreatic Ductal Adenocarcinoma
  - Stomach-Esophageal Cancer

- **Gynecological**
  - Cervical Cancer
  - Ovarian Serous Cystadenocarcinoma
  - Uterine Carcinosarcoma
  - Uterine Corpus Endometrial Carcinoma

- **Head and neck**
  - Squamous cell carcinoma
  - Uveal melanoma

- **Hematologic**
  - Acute myeloid leukemia
  - Thymoma

- **Skin**
  - Cutaneous melanoma

- **Soft tissue**
  - Sarcoma

- **Thoracic**
  - Lung Adenocarcinoma
  - Lung Squamous Cell Carcinoma
  - Mesothelioma

- **Urologic**
  - Chromophobe Renal Cell Carcinoma
  - Clear Cell Kidney Carcinoma
  - Papillary Kidney Carcinoma
  - Prostate Adenocarcinoma
  - Testicular Germ Cell Cancer
  - Urothelial Bladder Carcinoma
Datasets

**Data types**
- Clinical data
- Images
- Microsatellite instability
- DNA sequencing
- miRNA sequencing
- Protein expression
- mRNA & RNA sequencing
- Array-based expression
- DNA methylation
- Copy number

**Data access tiers**
- Open access
  - De-identified
  - Requires no certification
- Controlled access
  - No direct identifiers
  - Must complete Data Access Request (DAR) form
Click on “Cases with Data” for tumour of interest.
### Data Matrix Datasets

The Data Matrix only provides the latest revision of each archive; older revisions are available through bulk download or HTTP access. Also, it does not allow for querying across multiple disease studies.

**Legend:**
- **A**: Available
- **P**: Pending
- **N**: Not Available
- **N**: Not Applicable

- **Projected data**

#### Data levels explained:

https://tcga-data.nci.nih.gov/tcga/tcgaDataType.jsp

- The green A indicates that the relevant data is available for that sample.
## Data Levels and Data Types

The table below shows the relationship of TCGA data types to data levels as well as information on important metadata.

Please see the TCGA Data Primer for a detailed guide on TCGA data types.

*“Restricted” indicates data that are in the restricted-access data file. This phrase “restricted-access” is also used to denote such data in TCGA in order to access them. Please see the Access Tumor page for more information.

### Relationship of Data Levels to Data Types

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Data Subtypes</th>
<th>Cancer Types Applicable</th>
<th>Data Type Name</th>
<th>Data Type Name</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Important Metadata</th>
<th>How to Retrieve Data Files</th>
</tr>
</thead>
</table>
| Clinical Data   | 1. Clinical date       | All                     | Clinical       | Clinical       | Available clinical information for each participant (may include demographics, information, treatment information, survival data, etc) | n/a     | info                                                                  | The BCR data dictionary (dtd) describes the clinical and biospecimen data elements in TCGA | Data Matrix: Select “Clinical” for Data Type  
Bulk Download: Select “Complete Clinical Set” for Data Type |
|                 |                        |                         |                |                |                                                                        |        |                                                                        |                                                                                  |                                                                                   |
| Biopsy specimen data | All                     | Clinical                 | Biopsy specimen | Biopsy specimen | Information on how samples from each participant were processed by the Biopsy specimen Core Resource Center (CRF) (dtd) is defined in the table. | n/a     | info                                                                  | The BCR data dictionary (dtd) describes the clinical and biospecimen data elements in TCGA | Data Matrix: Select “Clinical” for Data Type  
Bulk Download: Select “Complete Clinical Set” for Data Type |
| Images          | 1. Diagnostic image    | All                     | Diagnostic     | Diagnostic     | Tissue images used to diagnose participant                           | n/a     | info                                                                  | Available images are listed in the biospecimen biotab and annotations | Bulk Download: Select “Diagnostic Images” for Platform. Images cannot be retrieved via the Data Matrix. |
|                 |                        |                         |                |                |                                                                        |        |                                                                        |                                                                                  |                                                                                   |
| Pathology Reports | All                     | Pathology Reports       | Pathology      | Pathology      | Pathology reports for a subset of participants                      | n/a     | info                                                                  |                                                                                  | Bulk Download: Select “Pathology Reports” for Platform. Pathology reports cannot be retrieved via the Data Matrix. |
|                 |                        |                         |                |                |                                                                        |        |                                                                        |                                                                                  |                                                                                   |
| Microsatellite Instability (MSI) | All                     | MSI                     | MSI            | MSI            | Markers indicating presence or absence of MSI shift, whole homozygous in heterozygosity, and loss of heterozygosity (LOH) observed in the tumor sample for each participant. (dtd) is defined in the table. | n/a     | info                                                                  | Level 1 data are submitted as part of a standard TCGA archive  
Level 1: Data Matrix & Level 2: Data Matrix are contained in the BCR clinical data archives. | Level 1: Data Matrix &  
Bulk Download: Select “Tumor Biology Results” for Data Type  
Level 2: Data Matrix: Select “Tumor Biology Results” for Data Type |
|                 |                        |                         |                |                |                                                                        |        |                                                                        |                                                                                  |                                                                                   |
| DNA Sequencing  | 1. Whole exome sequence | All                     | Whole exome     | Whole exome     | Whole exome sequence for both tumor and normal sample for each participant. (dtd) is defined in the table. | n/a     | info                                                                  | Experimental protocol, including primer information, is contained in the multidata xml file | See CGHub site  
Bulk Download: Select “Whole Exome Sequencing” for Data Type |

**Generally:**
- **Level 1** = Raw data
- **Level 2** = A little processed
- **Level 3** = Normalised and processed

Raw data is controlled access data and requires application to dbGAP

Raw sequence data is held at CGHub (https://cghub.ucsc.edu/)
Click on RNA-seq to select all RNA-seq samples.
After filling out email and hitting download, a link to the achieve with the files below will be sent to you...
TCGA data portal is difficult to use...

- Data portal is great for downloading data in large tab delimited format – perfect for a bioinformatician

- Data portal files are difficult to use for the average biologist

- Fortunately there are some alternatives:

- A data analysis portal to TCGA data
- Provides functions for visualisation, analysis and download of data.
- Maintained by Memorial Sloan-Kettering Cancer Center
Features of cBioPortal

• Visualising frequency of mutations
• Correlation between occurrence of mutations
• Correlation of expression and CNV or methylation
• Visualisation of mutations
• Survival analysis
• Network analysis

The CBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.

Please adhere to the TCGA publication guidelines when using TCGA data in your publications.

In this query, we are telling cBioPortal to perform an analysis comparing all AML samples with ERG mutation or CNA and those without ERG mutation nor CNA.

Select Cancer Study:
- TCGA
  - Acute Lymphoid Leukemia (1)
  - Infant MLL-Rearranged Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015) 24 samples
  - Acute Myeloid Leukemia (2)
    - Acute Myeloid Leukemia (TCGA, Provisional) 200 samples
    - Acute Myeloid Leukemia (TCGA, NEJM 2013) 200 samples
  - Multiple Myeloma (1)
  - Multiple Myeloma (Broad, Cancer Cell 2014) 205 samples
  - Bone (2)
  - Ewing Sarcoma (2)

Select Genomic Profiles:
- Mutations
- Copy number alterations from GISTIC
- mRNA Expression data
  - mRNA Expression z-Scores (RNA Sec V2 RSEM)
  - mRNA Expression z-Scores (RNA Sec RPKM)

Select Patient/Cases Set:
- Tumor Samples with CAN data (191)
  - To build your own case set, try out our enhanced Study View.

Enter Gene Set:
- Advanced: Onco Query Language (OQL)
  - User-defined list
  - Select Genes from Recurrent CNAs (Gistic)
  - ERG
9 out of 191 samples have alteration in ERG:
- 8 samples have amplifications of ERG
- 1 sample has a deep deletion of ERG
Plots – correlation ERG expression with CNA

Samples with amplification possibly have higher expression
Survival analysis

We know that high ERG expression is associated with poor survival (Marcucci et al JCO 2005).

Seems like ERG amplification is also associated with poor survival.
Network analysis

This network analysis is not that interesting, but it could be more useful with a larger input gene set.
Bookmark

You can make a URL to immediately share analysis with collaborators
**Gene summaries across cancer types**

**Query**

<table>
<thead>
<tr>
<th>Select Cancer Study:</th>
<th>Download Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search...</td>
<td>All (121)</td>
</tr>
<tr>
<td>Adrenal Gland (1)</td>
<td></td>
</tr>
<tr>
<td>Adrenocortical Carcinoma (1)</td>
<td></td>
</tr>
<tr>
<td>Adrenocortical Carcinoma (TCGA, Provisional) 92 samples</td>
<td></td>
</tr>
<tr>
<td>Biliary Tract (5)</td>
<td></td>
</tr>
<tr>
<td>Cholangiocarcinoma (4)</td>
<td></td>
</tr>
<tr>
<td>Cholangiocarcinoma (Johns Hopkins University, Nat Genet 2013) 40 samples</td>
<td></td>
</tr>
<tr>
<td>Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013) 15 samples</td>
<td></td>
</tr>
<tr>
<td>Cholangiocarcinoma (National University of Singapore, Nat Genet 2012) 8 samples</td>
<td></td>
</tr>
</tbody>
</table>

**Select Genomic Profiles:**

**Select Data Type Priority:** Mutation and CNA, Only Mutation, Only CNA

**Select Patient/Case Set:** Select an Option

To build your own case set, try out our enhanced Study View.

**Enter Gene Set:** Advanced: Onco Query Language (OQL)

**User-defined List**

ERG
Gene summaries across cancer types

Aberration frequency

Types of aberration

Data types in analysis

Cancer types

- Mutation
- Deletion
- Amplification
- Multiple alterations
Cancer Sequencing Projects

International Cancer Genome Consortium (ICGC)

- Collaboration between 22 countries
- Initiated in 2007
- **Aim:**
  To catalogue genomic abnormalities in tumours from 50 different cancer types & subtypes
- **Update today:**
  88 projects, 17 jurisdictions, >25,000 tumour genomes
- Uses data from TCGA and the Sanger Cancer Genome Project
# Working groups

<table>
<thead>
<tr>
<th>International Cancer Genome Consortium (ICGC) Working Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical and Pathology Issues</strong></td>
</tr>
<tr>
<td>Lynda Chin, Jean-Yves Blay, William Dalton, Tony Green, Stan Hamilton, Timothy Ley, Ed Liu, Paul Mischel, Kenneth Pienta, Rajiv Sarin, Daniel Tan</td>
</tr>
<tr>
<td><strong>Quality Standards of Samples</strong></td>
</tr>
<tr>
<td>Peter Lichter, Carolyn Compton, Andy Futreal, Youyang Lu, Miguel Angel Piris</td>
</tr>
<tr>
<td><strong>Genome Analyses</strong></td>
</tr>
<tr>
<td>Mike Stratton, Olli Kallioniemi, Ed Liu, Marco Marra, John McPherson, Brad Ozenberger, Henk Stunnenberg, Daniel Tan, Brandon Wainwright, Rick Wilson</td>
</tr>
<tr>
<td><strong>Informed Consent and Privacy Protections</strong></td>
</tr>
<tr>
<td>Bartha Knoppers, Martin Bobrow, Wylie Burke, Kazuto Kato, Karen Kennedy, Brad Ozenberger, Daniel Tan, Susan Wallace, Henry Yang</td>
</tr>
<tr>
<td><strong>Sample Size/Study Design</strong></td>
</tr>
<tr>
<td>Eric Lander, Ron DePinho, Doug Easton, Gaddy Getz, Partha P. Majumder</td>
</tr>
<tr>
<td><strong>Data Management/Databases and Coordination</strong></td>
</tr>
<tr>
<td>Lincoln Stein, Cameron Brennan, Arul Chinnaian, Peter Good, Joe Gray, J Gowrishankar, David Haussler, David Housman, Tim Hubbard, Subha Madhavan, Paul Spellman</td>
</tr>
<tr>
<td><strong>Data Release, Data Tiers, Intellectual Property, and Publications</strong></td>
</tr>
<tr>
<td>Mark Guyer, Daniela Gerhard, Karen Kennedy, Brad Ozenberger</td>
</tr>
</tbody>
</table>
ICGC Samples
Data Release 20
November 27th, 2015

Donor Distribution by Primary Site

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer projects</td>
<td>66</td>
</tr>
<tr>
<td>Cancer primary sites</td>
<td>21</td>
</tr>
<tr>
<td>Donors with molecular data in DCC</td>
<td>14,767</td>
</tr>
<tr>
<td>Total Donors</td>
<td>17,867</td>
</tr>
<tr>
<td>Simple somatic mutations</td>
<td>36,985,985</td>
</tr>
<tr>
<td>Mutated genes</td>
<td>57,637</td>
</tr>
</tbody>
</table>
Data types

• Mandatory:
  • Genomic DNA analyses of tumors (and matching control DNA) are core elements of the project.

• Complementary (Recommended):
  • Additional studies of DNA methylation and RNA expression are recommended on the same samples that are used to find somatic mutations.

• Optional:
  • Proteomic analyses
  • Metabolomic analyses
  • Immunohistochemical analyses
## Data access policy

<table>
<thead>
<tr>
<th>ICGC Open Access Datasets</th>
<th>ICGC Controlled Access Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Cancer Pathology</td>
<td>- Detailed Phenotype and Outcome Data</td>
</tr>
<tr>
<td>- Histologic type or subtype</td>
<td>- Patient demography</td>
</tr>
<tr>
<td>- Histologic nuclear grade</td>
<td>- Risk factors</td>
</tr>
<tr>
<td>- Patient/Person</td>
<td>- Examination</td>
</tr>
<tr>
<td>- Gender</td>
<td>- Surgery/Drugs/Radiation</td>
</tr>
<tr>
<td>- Age range</td>
<td>- Sample/Slide</td>
</tr>
<tr>
<td>- Gene Expression (normalized)</td>
<td>- Specific histological features</td>
</tr>
<tr>
<td>- DNA methylation</td>
<td>- Protocol</td>
</tr>
<tr>
<td>- Genotype frequencies</td>
<td>- Analyte/Aliquot</td>
</tr>
<tr>
<td>- Computed Copy Number and Loss of Heterozygosity</td>
<td>- Gene Expression (probe-level data)</td>
</tr>
<tr>
<td>- Newly discovered somatic variants</td>
<td>- Raw genotype calls</td>
</tr>
<tr>
<td></td>
<td>- Gene-sample identifier links</td>
</tr>
<tr>
<td></td>
<td>- Genome sequence files</td>
</tr>
</tbody>
</table>
ICGC data portal (http://dcc.icgc.org/)

Click on cancer projects

Data Release 20
November 27th, 2015
Donor Distribution by Primary Site

Tutorial
EXAMPLE QUERIES
1. BRAF missense mutations in colorectal cancer
2. Most frequently mutated genes by high impact mutations in stage III malignant lymphoma
3. Brain cancer donors with frameshift mutations and having methylation data available
Cancer project view

- Click on BRCA-US

Donor Distribution
17,867 Donors across 66 Projects

Top 20 Mutated Genes with High Functional Impact SSMs
9,155 Unique SSM-Tested Donors

Number of Somatic Mutations in Donor's Exomes Across Cancer Projects
View top mutated genes
Or search by mutation ID/location
Looking at mutations in specific genes

ICGC Data Portal

Type in BRAF here

About Us

The ICGC Data Portal provides tools for visualizing, querying and downloading the data released quarterly by the consortium’s member projects.

To access ICGC controlled tier data, please read these instructions.

New features will be regularly added by the DCC development team. Feedback is welcome.

Data Release 20
November 27th, 2015

Donor Distribution by Primary Site

Tutorial

EXAMPLE QUERIES

1. BRAF missense mutations in colorectal cancer
2. Most frequently mutated genes by high impact mutations in stage III malignant lymphoma
3. Brain cancer donors with frameshift mutations and having methylation data available
## Gene centric view

### General information

<table>
<thead>
<tr>
<th>Summary</th>
<th>External References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>HGNC Gene</td>
</tr>
<tr>
<td>Name</td>
<td>Ensembl (release 75)</td>
</tr>
<tr>
<td>Synonyms</td>
<td>COSMIC</td>
</tr>
<tr>
<td>Type</td>
<td>Entrez Gene</td>
</tr>
<tr>
<td>Location</td>
<td>OMIM</td>
</tr>
<tr>
<td>Strand</td>
<td>UniProtKB/Swiss-Prot</td>
</tr>
</tbody>
</table>

### Annotation

<table>
<thead>
<tr>
<th>Reactome Pathways</th>
<th>Curated Gene Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMS-mediated activation</td>
<td>Cancer Gene Census</td>
</tr>
<tr>
<td>CREB phosphorylation through the activation of Ras</td>
<td></td>
</tr>
<tr>
<td>Frs2-mediated activation</td>
<td></td>
</tr>
<tr>
<td>MAP2K and MAPK activation</td>
<td></td>
</tr>
<tr>
<td>Negative feedback regulation of MAPK pathway</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GC Terms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP binding</td>
<td></td>
</tr>
<tr>
<td>CD4-positive, alpha-beta T cell differentiation</td>
<td></td>
</tr>
<tr>
<td>Fc-epsilon receptor signalling pathway</td>
<td></td>
</tr>
<tr>
<td>MAP kinase kinase kinase activity</td>
<td></td>
</tr>
<tr>
<td>MAPK cascade</td>
<td></td>
</tr>
</tbody>
</table>

Then scroll down...
Gene centric view

Types of cancers with the mutation
Gene centric view

Cancer Distribution
1,144 DONORS AFFECTED BY 1,244 MUTATIONS ACROSS 43 PROJECTS

More detailed information

<table>
<thead>
<tr>
<th>Project</th>
<th>Site</th>
<th>Tumour Type</th>
<th>Tumour Subtype</th>
<th># Donors affected</th>
<th># Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>MELA-AU</td>
<td>Skin</td>
<td>Skin cancer</td>
<td>Melanoma</td>
<td>149 / 183 (81.42%)</td>
<td>501</td>
</tr>
<tr>
<td>SKCA-BR</td>
<td>Skin</td>
<td>Melanoma</td>
<td>48 / 66 (72.73%)</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>THCA-US</td>
<td>Head and neck</td>
<td>Head and Neck cancer</td>
<td>Thyroid carcinoma</td>
<td>238 / 400 (59.50%)</td>
<td>5</td>
</tr>
<tr>
<td>THCA-SA</td>
<td>Head and neck</td>
<td>Thyroid cancer</td>
<td>Papillary thyroid carcinoma</td>
<td>8 / 15 (53.33%)</td>
<td>2</td>
</tr>
<tr>
<td>ESAD-UK</td>
<td>Esophagus</td>
<td>Esophageal cancer</td>
<td>Esophageal adenocarcinoma</td>
<td>58 / 119 (48.74%)</td>
<td>93</td>
</tr>
<tr>
<td>SKCM-US</td>
<td>Skin</td>
<td>Skin cancer</td>
<td>Cutaneous melanoma</td>
<td>148 / 335 (44.18%)</td>
<td>22</td>
</tr>
<tr>
<td>LIRI-JP</td>
<td>Liver</td>
<td>Liver cancer</td>
<td>Hepatocellular carcinoma (Virus associated)</td>
<td>83 / 260 (31.92%)</td>
<td>131</td>
</tr>
<tr>
<td>OV-AU</td>
<td>Ovary</td>
<td>Ovarian cancer</td>
<td>Serous cystadenocarcinoma</td>
<td>29 / 93 (31.18%)</td>
<td>46</td>
</tr>
<tr>
<td>PACA-CA</td>
<td>Pancreas</td>
<td>Pancreatic cancer</td>
<td>Ductal adenocarcinoma</td>
<td>55 / 204 (26.96%)</td>
<td>75</td>
</tr>
<tr>
<td>MALY-DE</td>
<td>Blood</td>
<td>Malignant Lymphoma</td>
<td>Germinal center B-cell derived lymphomas</td>
<td>10 / 44 (22.73%)</td>
<td>19</td>
</tr>
<tr>
<td>PRAD-UK</td>
<td>Prostate</td>
<td>Prostate cancer</td>
<td>Adenocarcinoma</td>
<td>21 / 108 (19.44%)</td>
<td>21</td>
</tr>
<tr>
<td>RECA-EU</td>
<td>Kidney</td>
<td>Renal cancer</td>
<td>Renal cell carcinoma (Focus on but not limited to clear cell subtype)</td>
<td>18 / 95 (18.95%)</td>
<td>23</td>
</tr>
<tr>
<td>COAD-US</td>
<td>Colorectal</td>
<td>Colon cancer</td>
<td>Adenocarcinoma</td>
<td>33 / 216 (15.28%)</td>
<td>11</td>
</tr>
<tr>
<td>LINC-JP</td>
<td>Liver</td>
<td>Liver cancer</td>
<td>Hepatocellular carcinoma (Virus associated)</td>
<td>30 / 244 (12.30%)</td>
<td>34</td>
</tr>
</tbody>
</table>
Gene centric view

### Cancer Distribution

1,144 donors affected by 1,244 mutations across 43 projects

#### Cancer Site and Mutations

<table>
<thead>
<tr>
<th>Project</th>
<th>Site</th>
<th>Tumour Type</th>
<th>Tumour Subtype</th>
<th># Donors affected</th>
<th># Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>MELA-AU</td>
<td>Skin</td>
<td>Skin cancer</td>
<td>Melanoma</td>
<td>149 / 183 (81.42%)</td>
<td>501</td>
</tr>
<tr>
<td>SKCA-BR</td>
<td>Skin</td>
<td>Melanoma</td>
<td></td>
<td>48 / 66 (72.73%)</td>
<td>81</td>
</tr>
<tr>
<td>THCA-US</td>
<td>Head and neck</td>
<td>Head and Neck cancer</td>
<td>Thyroid carcinoma</td>
<td>238 / 400 (59.50%)</td>
<td>93</td>
</tr>
<tr>
<td>THCA-SA</td>
<td>Head and neck</td>
<td>Thyroid cancer</td>
<td>Papillary thyroid carcinoma</td>
<td>8 / 15 (53.33%)</td>
<td>2</td>
</tr>
<tr>
<td>ESAD-UK</td>
<td>Esophagus</td>
<td>Esophageal cancer</td>
<td>Esophageal adenocarcinoma</td>
<td>58 / 119 (48.74%)</td>
<td>93</td>
</tr>
<tr>
<td>SKCM-US</td>
<td>Skin</td>
<td>Skin cancer</td>
<td>Cutaneous melanoma</td>
<td>148 / 335 (44.18%)</td>
<td>22</td>
</tr>
<tr>
<td>LIRI-JP</td>
<td>Liver</td>
<td>Liver cancer</td>
<td>Hepatocellular carcinoma (Virus associated)</td>
<td>83 / 260 (31.92%)</td>
<td>131</td>
</tr>
<tr>
<td>OV-AU</td>
<td>Ovary</td>
<td>Ovarian cancer</td>
<td>Serous cystadenocarcinoma</td>
<td>29 / 93 (31.18%)</td>
<td>46</td>
</tr>
<tr>
<td>PACA-CA</td>
<td>Pancreas</td>
<td>Pancreatic cancer</td>
<td>Ductal adenocarcinoma</td>
<td>55 / 204 (26.96%)</td>
<td>75</td>
</tr>
<tr>
<td>MALY-DE</td>
<td>Blood</td>
<td>Malignant Lymphoma</td>
<td>Germinal center B-cell derived lymphomas</td>
<td>10 / 44 (22.73%)</td>
<td>19</td>
</tr>
<tr>
<td>PRAD-UK</td>
<td>Prostate</td>
<td>Prostate cancer</td>
<td>Adenocarcinoma</td>
<td>21 / 108 (19.44%)</td>
<td>21</td>
</tr>
<tr>
<td>RECA-EU</td>
<td>Kidney</td>
<td>Renal cancer</td>
<td>Renal cell carcinoma (Focus on but not limited to clear cell subtype)</td>
<td>18 / 95 (18.95%)</td>
<td>23</td>
</tr>
<tr>
<td>COAD-US</td>
<td>Colorectal</td>
<td>Colon cancer</td>
<td>Adenocarcinoma</td>
<td>33 / 216 (15.28%)</td>
<td>11</td>
</tr>
<tr>
<td>LINC-JP</td>
<td>Liver</td>
<td>Liver cancer</td>
<td>Hepatocellular carcinoma (Virus associated)</td>
<td>30 / 244 (12.30%)</td>
<td>34</td>
</tr>
</tbody>
</table>
Hover over the section of the pie chart to see what region it represents.

Each mutation has a unique ID (click for more info).
Advanced Search

Find out which cancers commonly have BRAF missense mutations

Go to the home page and click “advanced search”
Advanced Search

Find out which cancers commonly have BRAF missense mutations

Search for “BRAF”
Advanced Search

Find out which cancers commonly have BRAF missense mutations

Go to the mutations tab

Select “missense”

Hover mouse to see details.

Most common cancers with BRAF missense mutations are thyroid cancer and melanoma.
Limitations of data portal

• The data portal is mutation centric
  – i.e. All queries are related to retrieving tumours/samples with particular mutations in a particular gene

• If you just want expression/methylation data for a particular gene you still have to download the data
Downloading data from ICGC

Go to the home page and click "advanced search"
Downloading data from ICGC

Select cancer type of interest

Click download donor data

Note: go back to Advance search on home page
Downloading data from ICGC

Select the data types of interest

Click “Submit”
Downloading data from ICGC

Or download from the data repository
The advantage of ICGC is that data for all samples is in a single file so it is easier to work with in Excel (if file is small) or Galaxy (if file is big).
Summary

- There are global cancer genome sequencing projects with publically available data

- TCGA data can downloaded or easily viewed through cBioPortal

- ICGC data can be downloaded or viewed from the user interface
Exercises

1. Download patient clinical annotations for AML (TCGA dataset) using TCGA data portal and then using the ICGC data portal.

2. Using the ICGC data portal:
   a. What is the cancer with most frequent RUNX1 mutations?
   b. Which cancer has the most RUNX1 frameshift mutations?

3. Using cBioPortal:
   a. Do kidney renal papillary cell carcinoma patients with BAP1 mutations have worse survival than those without?