Cancer Informatics Lecture
Mayo-UIUC Computational Genomics Course
June 14, 2019

Krishna Rani Kalari Ph.D.
Associate Professor
Outline

• The Cancer Genome Atlas (TCGA)
• Genomic Data Commons (GDC)
• COSMIC database (mutations database)
• cBioPortal for cancer genomics
• GTEx
• Precision Medicine in Cancer
• Single-cell RNA-Seq
• Design of experiments (GEO)
The Cancer Genome Atlas (TCGA)

TCGA: A Community Resource Looking for a Broader Community

Kenna Shaw, Ph.D.
Director
The Cancer Genome Atlas

Reusing the slides from Kenna Shaw's presentation
TCGA core objectives

Launched in 2006 as a pilot and expanded in 2009, the goals of TCGA are to:

• Establish infrastructure for effective team science

• Develop a scalable “pipeline” beginning with highest quality samples

• Determine the feasibility of a large-scale, high throughput approach to identifying the molecular ‘parts-list’

• Evaluate using statistically-robust sample sets

• Make the data publicly and broadly available to the cancer community while protecting patient privacy
TCGA multiple data types

25* forms of cancer

- glioblastoma multiforme (brain)
- squamous carcinoma (lung)
- serous cystadenocarcinoma (ovarian)

Biospecimen Core Resource with more than 150 Tissue Source Sites

- 6 Cancer Genomic Characterization Centers
- 3 Genome Sequencing Centers
- 7 Genome Data Analysis Centers
- Data Coordinating Center

Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec
Rare tumors projects initiated in 2012

- Adrenocortical Carcinoma
- Adult ALL (B-cell and T-Cell)
- Anaplastic Thyroid
- Cholangiocarcinoma
- Chromophobe kidney
- High Risk MDS (del 5q- cases)
- Mesothelioma
- Paraganglioma/Pheochromocytoma
- Testicular Germ Cell
- Thymoma
- Uterine Carcinosarcoma
- Sarcomas
- Others??
Genomic Data Commons - GDC
Genomic Data Commons

- A NCI repository for The Cancer Genome Atlas and Genomics data.
- It consists of data from 68 primary sites
- >33,605 cases
- >Three million mutations
- 374,699 files
## Projects

**NATIONAL CANCER INSTITUTE**

**GDC Data Portal**

**Projects**

**Start searching by selecting a facet**

### Top Mutated Cancer Genes in Selected Projects

- **% of Cases Affected**
- **# of Cases Affected**

![Graph showing top mutated cancer genes](image)

**Case Distribution per Project**

33,696 Cases across 47 Projects

### Table

<table>
<thead>
<tr>
<th>Project</th>
<th>Disease Type</th>
<th>Primary Site</th>
<th>Program</th>
<th>Cases</th>
<th>Seq</th>
<th>Exp</th>
<th>SNV</th>
<th>CNV</th>
<th>Metll</th>
<th>Clinical</th>
<th>Bio</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILAD</td>
<td>23 Disease Types</td>
<td>42 Primary Sites</td>
<td>FM</td>
<td>16,008</td>
<td>0</td>
<td>0</td>
<td>16,008</td>
<td>0</td>
<td>16,008</td>
<td>16,008</td>
<td>38,134</td>
<td></td>
</tr>
<tr>
<td>TARGET-NBL</td>
<td>Neuroblastoma</td>
<td>Nervous System</td>
<td>TARGET</td>
<td>1,120</td>
<td>277</td>
<td>152</td>
<td>216</td>
<td>0</td>
<td>1,120</td>
<td>1,120</td>
<td>2,807</td>
<td></td>
</tr>
<tr>
<td>TCGA-IRCA</td>
<td>9 Disease Types</td>
<td>Breast</td>
<td>TCGA</td>
<td>1,080</td>
<td>1,080</td>
<td>1,044</td>
<td>1,056</td>
<td>1,056</td>
<td>1,056</td>
<td>1,056</td>
<td>31,598</td>
<td></td>
</tr>
<tr>
<td>TARGET-AML</td>
<td>Myeloid Leukemias</td>
<td>Blood</td>
<td>TARGET</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>2,452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARGET-WT</td>
<td>High-Risk Wilms Tumor</td>
<td>Kidney</td>
<td>TARGET</td>
<td>652</td>
<td>652</td>
<td>652</td>
<td>652</td>
<td>652</td>
<td>652</td>
<td>1,410</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA-GBM</td>
<td>2 Disease Types</td>
<td>Brain</td>
<td>TCGA</td>
<td>617</td>
<td>617</td>
<td>166</td>
<td>326</td>
<td>599</td>
<td>423</td>
<td>617</td>
<td>617</td>
<td>12,060</td>
</tr>
<tr>
<td>TCGA-OV</td>
<td>2 Disease Types</td>
<td>Ovary</td>
<td>TCGA</td>
<td>600</td>
<td>575</td>
<td>492</td>
<td>443</td>
<td>597</td>
<td>602</td>
<td>602</td>
<td>15,156</td>
<td></td>
</tr>
<tr>
<td>TCGA-LUAD</td>
<td>3 Disease Types</td>
<td>Bronchus and lung</td>
<td>TCGA</td>
<td>585</td>
<td>582</td>
<td>519</td>
<td>569</td>
<td>518</td>
<td>570</td>
<td>585</td>
<td>585</td>
<td>17,074</td>
</tr>
<tr>
<td>TCGA-LIHC</td>
<td>4 Disease Types</td>
<td>2 Primary Sites</td>
<td>TCGA</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>18,187</td>
<td></td>
</tr>
<tr>
<td>TCGA-KIRC</td>
<td>Adenomas and Adenocarcinomas</td>
<td>Kidney</td>
<td>TCGA</td>
<td>537</td>
<td>537</td>
<td>537</td>
<td>537</td>
<td>537</td>
<td>537</td>
<td>15,157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA-HNSC</td>
<td>Squamous Cell Neoplasms</td>
<td>13 Primary Sites</td>
<td>TCGA</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>528</td>
<td>15,228</td>
</tr>
<tr>
<td>TCGA-LGG</td>
<td>Gliomas</td>
<td>Brain</td>
<td>TCGA</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>14,741</td>
<td></td>
</tr>
</tbody>
</table>

©2017 MFMER | 3702274-10
Exploration

View Files in Repository

Cases (33,605)  Genes (22,372)  Mutations (3,142,246)  OncoGrid

OncoGrid

200 Most Mutated Cases and Top 50 Mutated Genes By SSM

Mutations

- Show Mutations

- Missense
- Start Lost
- Stop Gained

- Frame Shift
- Stop Lost

CNV Changes

- Show Copy Number Variations

- Loss
- Gain
Analysis

Set Operations
Display Venn diagram and find intersection or union, etc. of your sets of the same type.

Cohort Comparison
Display the survival analysis of your case sets and compare characteristics such as gender, vital status and age at diagnosis.
COSMIC

“COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.”
Expert curated database

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

- **COSMIC**
  The core of COSMIC, an expert-curated database of somatic mutations

- **Cell Lines Project**
  Mutation profiles of over 1,000 cell lines used in cancer research

- **COSMIC-3D**
  An interactive view of cancer mutations in the context of 3D structures

- **Cancer Gene Census**
  A catalogue of genes with mutations that are causally implicated in cancer
COSMIC v85, released 08-MAY-18

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world’s largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

- COSMIC
  The core of COSMIC, an expert-curated database of somatic mutations

- Cell Lines Project
  Mutation profiles of over 1,000 cell lines used in cancer research

COSMIC News

A step by step guide to account revalidation

We have introduced a new way of managing the COSMIC accounts by asking you to annually revalidate your account. More...

COSMIC Release v85

The May COSMIC release (v85) is now live. There are 3 new fully curated genes, substantial curation updates for TERT, 1 new fusion pair, and 162 samples from 8 new systematic screen papers. We have also added 25 new hallmark genes to the Cancer... More...

New COSMIC file download service

The SFTP server will be deprecated Aug 2018 - We have put in place new web endpoints to make it easier to access our data download files from scripts. Please change to the new endpoints as soon as possible. More...
The gene view histogram is a graphical view of mutations across KRAS. These mutations are displayed at the amino acid level across the full length of the gene by default. Restrict the view to a region of the gene by dragging across the histogram to highlight the region of interest, or by using the sliders in the filters panel to the left. Show more.
This section gives an overview of KRAS, along with links to any related data and resources.

**Gene view**

**Overview**

**External links**

**Drug resistance**

**Tissue distribution**

**Mutation distribution**

**Variants**

**References**

---

### COSMIC gene

**KRAS (COSG4)**

### Genomic coordinates

12:25209795..25250931 (negative strand)

### Synonyms

KRAS2, CCDS8702.1, P01116, ENSG00000133703

### COSMIC-3D

There are 88 structures for KRAS. View them in COSMIC-3D.

---

### Number of samples

242082 unique samples
43548 unique samples with mutations

### Alternative transcripts

**KRAS, ENST00000256078**

### Sequences

You can see various sequences for this gene:
- cDNA (ENST00000311936)
- Protein (KRAS)
- Transcript and protein aligned (ENST00000311936+KRAS)

### Gene fusions

KRAS is involved in 1 fusion, with the following gene:
- **UBE2L3** (1 mutation in 1 sample)

### Drug sensitivity data

Mutations in KRAS are associated with altered sensitivity to the following 7 drugs:
- RDEA119
- PD-0325901
- Trametinib

See all drug sensitivity data for KRAS.
Drug Resistance

Cancer Feature: KRAS_mut

Select a tissue specific analysis: Pan-Cancer

Volcano Plot

Schatten plots

IC50 Effect

Drug | Drug Target | Effect size | P-value | FDR | No. of altered cell lines | Tissue analysis
--- | --- | --- | --- | --- | --- | ---
Refametinib | MEK1, MEK2 | -0.722 | 1.67e-10 | 9.04e-06 | 116 | PANCANCER
Selumetinib | MEK1, MEK2 | -0.682 | 7.25e-10 | 4.28e-05 | 121 | PANCANCER
PD0325901 | MEK1, MEK2 | -0.707 | 1.07e-08 | 0.000352 | 107 | PANCANCER
Refametinib | MEK1, MEK2 | -0.696 | 6.65e-08 | 6.00e-06 | 100 | PANCANCER
CI-1040 | MEK1, MEK2 | -0.535 | 3.46e-06 | 0.159 | 109 | PANCANCER
Trametinib | MEK1, MEK2 | -0.671 | 5.33e-06 | 0.144 | 112 | PANCANCER
Nutlin-3a | MDH2 | 0.00355 | 6.00e-07 | 0.539 | 106 | PANCANCER
Bleomycin (50 uM) | dsDNA break induction | -0.084 | 0.00001 | 1.74 | 123 | PANCANCER
Cell Lines Project

COSMIC
The core of COSMIC, an expert-curated database of somatic mutations

Cell Lines Project
Mutation profiles of over 1,000 cell lines used in cancer research

COSMIC-3D
An interactive view of cancer mutations in the context of 3D structures

Cancer Gene Census
A catalogue of genes with mutations that are causally implicated in cancer
Cell lines project

- Mutation profiles of over 1,000 cell lines used in cancer research
- MCF7
COSMIC-3D

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

- **COSMIC**: The core of COSMIC, an expert-curated database of somatic mutations
- **Cell Lines Project**: Mutation profiles of over 1,000 cell lines used in cancer research
- **COSMIC-3D**: An interactive view of cancer mutations in the context of 3D structures
- **Cancer Gene Census**: A catalogue of genes with mutations that are causally implicated in cancer
COSMIC-3D

- A platform for understanding cancer mutations in the context of 3D protein structure.

EGFR [P00533] In Census 178 structures

Receptor tyrosine kinase binding ligands of the EGF family and activating several signaling cascades to convert extracellular cues into appropriate cellular responses.

Information

Click on a mutation in the Sequence Feature Viewer below to view more information.
COSMIC

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:

- **COSMIC**
  The core of COSMIC, an expert-curated database of somatic mutations

- **Cell Lines Project**
  Mutation profiles of over 1,000 cell lines used in cancer research

- **COSMIC-3D**
  An interactive view of cancer mutations in the context of 3D structures

- **Cancer Gene Census**
  A catalogue of genes with mutations that are causally implicated in cancer
Gene Tiers in Cancer Gene Census

- **Census tiers – 719 genes**
- **Tier 1** – A gene must possess a documented activity relevant to cancer, along with evidence of mutations in cancer which change the activity of the gene product in a way that promotes oncogenic transformation.
- **Tier 2** - Consists of genes with strong indications of a role in cancer but with less extensive available evidence.
# Breakdown of Genes/mutations

The gene list has been annotated with information concerning chromosomal location, tumour types in which mutations are oncogenic, and other genetic properties. We have sorted the data in a number of ways to list subsets of cancer genes we recommend that those wishing to scrutinise the list in detail should download it in its entirety from the table in the 'Cancer Gene Census'.

<table>
<thead>
<tr>
<th>Sorted By</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplifications</td>
<td>24</td>
</tr>
<tr>
<td>Chromosome</td>
<td>574</td>
</tr>
<tr>
<td>Frameshift Mutations</td>
<td>155</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>574</td>
</tr>
<tr>
<td>Germline Mutations</td>
<td>102</td>
</tr>
<tr>
<td>Large Deletions</td>
<td>40</td>
</tr>
<tr>
<td>Missense Mutations</td>
<td>251</td>
</tr>
<tr>
<td>Nonsense Mutations</td>
<td>153</td>
</tr>
<tr>
<td>Other Mutations</td>
<td>37</td>
</tr>
<tr>
<td>Somatic Mutations</td>
<td>534</td>
</tr>
<tr>
<td>Splicing Mutations</td>
<td>73</td>
</tr>
<tr>
<td>Translocations</td>
<td>314</td>
</tr>
</tbody>
</table>

Showing 1 to 12 of 12 entries
cBioPortal
Public cancer genomics data for mining

- Cbioportal
- Barrowed slides from cbioportal website
- Walkthrough an example using BRCA1, BRCA2 genes

http://www.cbioportal.org/index.do
Overview of Tabs in a Single Study Query

Note that depending on the data available for a particular study, not all of these will be present (e.g. a study without outcome data will not have a Survival tab)

- **OncoPrint**: Overview of genetic alterations per sample in each query gene
- **Cancer Types Summary**: Frequency of alteration in each query gene in the detailed cancer types included in this study
- **Mutual Exclusivity**: Statistical analysis to determine if query genes are mutually exclusively altered
- **Plots**: Explore the relationships among genetic alterations, gene expression, protein levels, DNA methylation and available clinical features
- **Mutations**: Details about mutations called in each query gene
- **Co-Expression**: Explore which genes have mRNA/protein levels correlated with query genes
- **Enrichments**: Explore which genes are altered in the set of samples with query gene alterations or in the set of samples without query gene alterations
- **Survival**: Compare survival of patients with alterations in query genes to the rest of the cohort
- **Network**: Explore gene networks centered on the query genes
- **CN Segments**: Explore copy number changes with the Integrated Genomics Viewer (IGV)
- **Download**: Download data or copy sample lists
- **Bookmark**: Link to save the query

http://www.cbioportal.org/index.do
Query overview

Browse available datasets and initiate queries.

Download data.

Search studies.

Number of studies for each organ system (click to filter).

List of all studies, organized by organ system.

Click here for a drop-down menu with some common searches and examples of advanced search features.

http://www.cbioportal.org/index.do

Download data.
Selecting a study: from Query

1. Start typing tumor type of interest...

2. Click on “View study summary” button

3. Or click on “Data Sets” to browse Data Sets page
Single study query

1. Start typing tumor type of interest to filter the list of studies.

2. Check the box for study of interest.

3. This section will update to include all data types available for the selected study. Select data types to query.

4. Select sample set. For some studies, an appropriate sample set will be automatically selected given the data types selected in Step 3.

5. Type gene(s) or select from pre-defined gene lists. cBioPortal will confirm that all entries are valid gene symbols.

6. Submit query

http://www.cbioportal.org/index.do
Select Studies:

- **Breast**
  - **Breast**
    - Breast Cancer (MSK, Cancer Cell 2018)
    - Breast Fibroepithelial Neoplasms
      - Breast Fibroepithelial Tumors (Duke-NUS, Nat Genet 2015)
    - Invasive Breast Carcinoma
      - Breast Cancer (Xenografts (British Columbia, Nature 2015)
      - Breast Invasive Carcinoma (British Columbia, Nature 2012)
      - Breast Invasive Carcinoma (Gimel, Nature 2012)
      - Breast Invasive Carcinoma (Ganger, Nature 2012)
      - Breast Invasive Carcinoma (TOGA, Cell 2015)
      - Breast Invasive Carcinoma (TOGA, Nature 2012)
      - Breast Invasive Carcinoma (TOGA, PanCancer Atlas)
  - **Pancancer**
    - PanCancer Carcinoma
      - PanCancer Carcinoma (TOGA, PanCancer Atlas)

Enter Genes:

Hint: Learn Omic Query Language (OQL) to write more powerful queries.

User-defined List

Enter HUGO Gene Symbols, Gene Names, or OQL

Submit Query

Please select one or more cancer studies.
Summary of alterations per sample. Each sample is a column. Each gene is a row. Different kinds of genetic alterations are highlighted with different colors.

The percentage of samples with an alteration in each query gene

The number (percentage) of samples with an alteration in any of the query genes

http://www.cbioportal.org/index.do
OncoPrint: Advanced Features

Add clinical tracks (options will vary depending on the data available for each study)

Add a heatmap with RNA or protein levels

Change the sample sorting order

Customize visualization

Download figure as PNG, PDF or SVG. Download patient/sample IDs in same order as OncoPrint.

When your mouse hovers, this toolbar appears with options to customize OncoPrint.

http://www.cbioportal.org/index.do
To change the order, click on a gene name and drag, or click on the . Samples will re-sort based on this new order.
Breast Invasive Carcinoma (TCGA, PanCancer Atlas)
All Complete Tumors (993 samples) / 2 Genes

Case Set: All Complete Tumors (993 patients / 993 samples)

Altered in 210 (21%) of 993 sequenced cases/patients (993 total)

BRCA1  13%

BRCA2  11%

Genetic Alteration
- Inframe Mutation (unknown significance)
- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Fusion
- Amplification
- Deep Deletion
- mRNA Upregulation
- No alterations
Mutual Exclusivity with Glioblastoma example

All pairwise combinations of query genes analyzed for mutual exclusivity or co-occurrence in the queried samples.

On the OncoPrint tab we could see visually that alterations in these three query genes tended to be mutually exclusive. Here we can address that same question with a statistical analysis.

The query contains 3 gene pairs with mutually exclusive alterations (2 significant), and no gene

- **Mutual exclusivity**
- **Co-occurrence**
- **Significant only**

<table>
<thead>
<tr>
<th>Gene A</th>
<th>Gene B</th>
<th>Neither</th>
<th>A Not B</th>
<th>B Not A</th>
<th>Both</th>
<th>Log Odds Ratio</th>
<th>p-Value</th>
<th>Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>IDH1</td>
<td>40</td>
<td>24</td>
<td>217</td>
<td>2</td>
<td>&lt;-3</td>
<td>&lt;0.001</td>
<td>Significant</td>
</tr>
<tr>
<td>IDH1</td>
<td>IDH2</td>
<td>52</td>
<td>218</td>
<td>12</td>
<td>1</td>
<td>&lt;-3</td>
<td>&lt;0.001</td>
<td>Significant</td>
</tr>
<tr>
<td>EGFR</td>
<td>IDH2</td>
<td>244</td>
<td>26</td>
<td>13</td>
<td>0</td>
<td>0.278</td>
<td>0.278</td>
<td>Mutually exclusive</td>
</tr>
</tbody>
</table>

Click on any column header to sort. Hover over the column names for more details about how values are calculated.

A positive value here suggests that alterations in these genes co-occur in the same samples, while a negative value suggests that alterations in these genes are mutually exclusive and occur in different samples.

p-Value comes from Fisher Exact Test. Note that this is an unadjusted p-value and may need to be corrected for multiple hypothesis testing.

\[ \log_2 \left( \frac{\text{odds of alteration in B given alteration in A}}{\text{odds of alteration in B given lack of alteration in A}} \right) \]
Plots

Depending on available data types for a given study, this tab allows for plots comparing copy number, gene expression, protein levels and DNA methylation of query genes, along with any available clinical attributes.

Choose genetic or clinical
Select a query gene
Select data type and processing
Swap horizontal & vertical axis
If checked, vertical axis will automatically show the same gene as horizontal axis.

Each dot is a sample, color-coded by mutation status.

http://www.cbioportal.org/index.do
Q: What are the hotspots for EGFR mutation in glioma?
A: Look at the lollipop diagram: G598V is the most common alteration. The Furin-like domain also appears to be frequently mutated.
Mutations are drawn as lollipops along the domain structure of the gene. The height of the lollipop reflects how many times that mutation was detected. This plot will update based on any filters applied to the table below. Hover over any lollipop for additional details.
This mutation is in OncoKB as a Level 3 variant. Hover over this symbol to see additional information, including that this is a known oncogenic mutation.

This mutation is in My Cancer Genome.

This mutation is annotated in CIViC. Hover over this symbol for additional information.

This mutation is a recurrent hotspot based on a statistical analysis of mutation frequency.

You may also see this symbol which means the mutation is a recurrent hotspot based on a statistical analysis of 3D protein conformation.

This mutation is in My Cancer Genome.
Co-Expression

Select from available data types

Each gene appears on a separate tab

Click on a gene name to see correlation plot

Compares mRNA/protein level expression of your query genes against all other genes. Only genes with Pearson and Spearman correlations >0.3 or <-0.3 are shown.

Check boxes to color-code sample dots by mutation status or change x- or y-axis to log scale

http://www.cbioportal.org/index.do
<table>
<thead>
<tr>
<th>Correlated Gene</th>
<th>Cytoband</th>
<th>Pearson's Correlation</th>
<th>Spearman's Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP2A</td>
<td>17q21.2</td>
<td>0.42</td>
<td>0.65</td>
</tr>
<tr>
<td>PSME3</td>
<td>17q21.31</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>ATAD5</td>
<td>17q11.2</td>
<td>0.58</td>
<td>0.63</td>
</tr>
<tr>
<td>KNTC1</td>
<td>12q24.31</td>
<td>0.58</td>
<td>0.61</td>
</tr>
<tr>
<td>KPNB1</td>
<td>17q21.32</td>
<td>0.57</td>
<td>0.61</td>
</tr>
<tr>
<td>SUZ12</td>
<td>17q11.2</td>
<td>0.56</td>
<td>0.60</td>
</tr>
<tr>
<td>PRIM1</td>
<td>12q13.3</td>
<td>0.50</td>
<td>0.59</td>
</tr>
<tr>
<td>DTL</td>
<td>1q32.3</td>
<td>0.48</td>
<td>0.58</td>
</tr>
<tr>
<td>RACGAP1</td>
<td>12q13.12</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>CCDC43</td>
<td>17q21.31</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>KIF20B</td>
<td>10q23.31</td>
<td>0.55</td>
<td>0.58</td>
</tr>
<tr>
<td>TIMELESS</td>
<td>12q13.3</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>WDR76</td>
<td>15q15.3</td>
<td>0.52</td>
<td>0.58</td>
</tr>
<tr>
<td>WDHD1</td>
<td>14q22.2-q22.3</td>
<td>0.50</td>
<td>0.57</td>
</tr>
<tr>
<td>NEMP1</td>
<td>12q13.3</td>
<td>0.49</td>
<td>0.57</td>
</tr>
<tr>
<td>CENPE</td>
<td>4q24</td>
<td>0.49</td>
<td>0.56</td>
</tr>
<tr>
<td>TOPBP1</td>
<td>3q22.1</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>CDC6</td>
<td>17q21.2</td>
<td>0.25</td>
<td>0.55</td>
</tr>
<tr>
<td>KNL1</td>
<td>15q15.1</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>POLQ</td>
<td>3q13.33</td>
<td>0.46</td>
<td>0.55</td>
</tr>
<tr>
<td>CKAP2L</td>
<td>2q14.1</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>RBL1</td>
<td>20q11.23</td>
<td>0.42</td>
<td>0.55</td>
</tr>
</tbody>
</table>

mRNA Expression Batch Normalized/Merged from Illumina HiSeq_RNASeqV2 syn4976369: BRCA1 vs.

Pearson: 0.42
Spearman: 0.65

- BRCA1 mutated
- TOP2A mutated
- Neither mutated

http://www.cbioportal.org/index.do
Enrichments

This tab takes samples with alterations in any query gene as a set and looks to see whether other genes are frequently altered in the same set of samples (co-occurring) or in the set of samples without query gene alterations (mutually exclusive).

Select type of data to examine

Hover over a dot to see the gene name

Filter table with these options

Click on any column header to sort. Hover over the for more details about how values are calculated.

http://www.cbioportal.org/index.do
Enrichments Tab

### Mutations Tab

#### Add checked genes to query: CDH1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Cytoband</th>
<th>Samples with alteration in altered group</th>
<th>Samples with alteration in unaltered group</th>
<th>Log Ratio</th>
<th>p-Value</th>
<th>q-Value</th>
<th>Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH1</td>
<td>15q22.1</td>
<td>13 (5.19%)</td>
<td>117 (14.94%)</td>
<td>-1.27</td>
<td>2.952e-4</td>
<td>0.230</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3q26.32</td>
<td>57 (27.14%)</td>
<td>289 (36.91%)</td>
<td>-0.44</td>
<td>4.829e-3</td>
<td>0.289</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>5q11.2</td>
<td>11 (5.24%)</td>
<td>79 (10.09%)</td>
<td>-0.95</td>
<td>0.0158</td>
<td>0.398</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>UVRAG</td>
<td>11q13.5</td>
<td>0 (0.00%)</td>
<td>17 (2.17%)</td>
<td>&lt;-10</td>
<td>0.0170</td>
<td>0.398</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>ASXL1</td>
<td>20q11.21</td>
<td>0 (0.00%)</td>
<td>12 (1.53%)</td>
<td>&lt;-10</td>
<td>0.0567</td>
<td>0.455</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>TBX3</td>
<td>12q24.21</td>
<td>3 (1.43%)</td>
<td>30 (3.83%)</td>
<td>-1.42</td>
<td>0.0572</td>
<td>0.455</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>ATP8B3</td>
<td>19p13.3</td>
<td>0 (0.00%)</td>
<td>11 (1.40%)</td>
<td>&lt;-10</td>
<td>0.0722</td>
<td>0.491</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>ABCC9</td>
<td>12p12.1</td>
<td>1 (0.48%)</td>
<td>17 (2.17%)</td>
<td>-2.19</td>
<td>0.0790</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>LRP12</td>
<td>8q22.3</td>
<td>1 (0.48%)</td>
<td>17 (2.17%)</td>
<td>-2.19</td>
<td>0.0790</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>SCN5A</td>
<td>3p22.2</td>
<td>1 (0.48%)</td>
<td>17 (2.17%)</td>
<td>-2.19</td>
<td>0.0790</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>ARRB1</td>
<td>11q13.4</td>
<td>0 (0.00%)</td>
<td>10 (1.28%)</td>
<td>&lt;-10</td>
<td>0.0918</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>ITGB2</td>
<td>21q22.3</td>
<td>0 (0.00%)</td>
<td>10 (1.28%)</td>
<td>&lt;-10</td>
<td>0.0918</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>UBA6</td>
<td>4q13.2</td>
<td>0 (0.00%)</td>
<td>10 (1.28%)</td>
<td>&lt;-10</td>
<td>0.0918</td>
<td>0.529</td>
<td>Mutual exclusivity</td>
</tr>
<tr>
<td>RGS9</td>
<td>17q24.1</td>
<td>1 (0.48%)</td>
<td>15 (1.92%)</td>
<td>-2.01</td>
<td>0.116</td>
<td>0.534</td>
<td>Mutual exclusivity</td>
</tr>
</tbody>
</table>

**Query Genes**

- CDH1

---

[http://www.cbioportal.org/index.do](http://www.cbioportal.org/index.do)

©2017 MFMER | 3702274-48
Case Set: All Complete Tumors (993 patients / 993 samples)

Altered in 361 (36%) of 993 sequenced cases/patients (993 total)

- **BRCA1**: 13%
- **BRCA2**: 11%
- **CDH1**: 17%

Genetic Alteration:
- Inframe Mutation (unknown significance)
- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Fusion
- Amplification
- Deep Deletion
- mRNA Upregulation
- No alterations
CNV

Breast Invasive Carcinoma (TCGA, PanCancer Atlas)
All Complete Tumors (993 samples) / 4 Genes

Gene Set / Pathway is altered in 420 (42.3%) of queried samples

Case Set: All Complete Tumors (993 patients / 993 samples)

Altered in 420 (42%) of 993 sequenced cases/patients (993 total)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>13%</td>
</tr>
<tr>
<td>BRCA2</td>
<td>11%</td>
</tr>
<tr>
<td>CDH1</td>
<td>17%</td>
</tr>
<tr>
<td>KDM3B</td>
<td>10%</td>
</tr>
</tbody>
</table>

Genetic Alteration:
- Inframe Mutation (unknown significance)
- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Truncating Mutation (unknown significance)
- Fusion
- Amplification
- Deep Deletion
- mRNA Upregulation
- mRNA Downregulation
- No alterations
mRNA overexpressed

Case Set: All Complete Tumors (993 patients / 993 samples)

Altered in 455 (46%) of 993 sequenced cases/patients (993 total)

BRCA1: 13%
BRCA2: 11%
CDH1: 17%
KDM3B: 10%
CENPH: 8%

Genetic Alteration:
- Inframe Mutation (unknown significance)
- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Truncating Mutation (unknown significance)
- Fusion
- Amplification
- Deep Deletion
- mRNA Upregulation
- mRNA Downregulation
- No alterations
Network

Visualize biological interaction networks centered on your query genes, with color-coding and filter options based on the frequency of genomic alterations in each gene. Click on the “Help” tab for a more detailed explanation.

Change zoom and move around network

View or modify the nodes included in the network (e.g. add drugs, filter genes by alteration frequency)

View or modify the types of interactions (edges) utilized in the plot

Click on any node to see detailed information about the gene here
BRCA1/2 string PPI network
CN Segments

View copy number for each sample at each query gene via the Integrated Genomics Viewer (IGV).

Plots for each gene appear on a separate tab.

Toggle track labels, a vertical line marking the center of the viewing screen, and a vertical line that moves with your cursor. Use to zoom in or out.

Click 🌡️ for track settings, including expanding the height of each sample (see below).

Each row is a single sample

Gene structures

Click on a read for sample ID and copy number value.

Download a copy number segment file for the selected samples.
Download data or copy lists of samples.

Download mutations and copy number

Frequency of gene alteration for each gene in the query

List of all samples with status of each query gene (blank = no alteration)

List of samples that have an alteration in one or more query genes

List of all samples with summary classification:
- 0 = no alteration in any query gene
- 1 = alteration in one or more query genes

Advanced feature: use this list as a custom sample list to run a new query in only the subset of samples with a particular genetic alteration.
Bookmark

Save a link to the current session. Useful for sharing with others or returning to a query at a later date.

Right click on one of the links below to bookmark your results:

http://www.cbiportal.org/index.do?session_id=59690649498e5df2e292870c

If you would like to use a shorter URL that will not break in email postings, you can use the bitly.com url below:

Genotype Tissue-Expression Project

- Genome-wide association studies (GWAS)
- Cases vs controls
- ~95% of SNPs located in non-coding regions
- 53 tissue sites
Sample and data processing overview

DNA Analysis
- OMNI 2.5M/5M: 450 donors
- WES (100x)
- WGS (30x): HiSeq 2000, HiSeq X

RNA sequencing
- QC: RIN ≥ 5.5
- polyA+ (Illumina TruSeq)
- 2x76bp, ≥ 50M reads

Donor

Blood & skin → LCLs Fibroblasts → RNA DNA

Tissues (PAXgene) → Pathology review → RNA (DNA) → RNA (DNA)

Brain (Liquid N2) → 9-11 sub-regions → RNA (DNA) → Quality control
RNA-seq and eQTL pipeline details

<table>
<thead>
<tr>
<th>Release</th>
<th>V6p</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome build</td>
<td>GRCh37</td>
<td>GRCh37</td>
<td>GRCh38</td>
<td>GRCh38</td>
</tr>
<tr>
<td>GENCODE annotation</td>
<td>v19</td>
<td>v19</td>
<td>v26</td>
<td>v26</td>
</tr>
<tr>
<td>Aligner</td>
<td>TopHat 1.4.1</td>
<td>STAR 2.4.2a</td>
<td>STAR 2.5.3a</td>
<td>STAR 2.5.3a</td>
</tr>
<tr>
<td>Gene expression</td>
<td>RNA-SeQC 1.1.8</td>
<td>RNA-SeQC 1.1.9</td>
<td>RNA-SeQC 1.1.9</td>
<td>RNA-SeQC 1.1.9</td>
</tr>
<tr>
<td>Transcript expression</td>
<td>FluxCapacitor 1.6</td>
<td>RSEM 1.2.22</td>
<td>RSEM 1.3.0</td>
<td>RSEM 1.3.0</td>
</tr>
<tr>
<td>Quality control metrics</td>
<td>RNA-SeQC 1.1.8</td>
<td>RNA-SeQC 1.1.9</td>
<td>RNA-SeQC 1.1.9</td>
<td>RNA-SeQC 1.1.9</td>
</tr>
<tr>
<td>QTL mapper</td>
<td></td>
<td></td>
<td>FastQTL</td>
<td></td>
</tr>
</tbody>
</table>

- Pipeline components selected and updated based on internal and published benchmarks (e.g., Teng et al., Genome Biology, 2016).

Overview of GTEx resources: open-access data

- Expression
  - Gene-level expression (TPM, counts)
  - Transcript-level expression (TPM, counts, isoform proportions)
  - Exon read counts
- QTLs
  - Single-tissue eQTLs (cis- and trans-)
  - Multi-tissue eQTLs
  - Future: splicing QTLs
- Histology images
- De-identified public access sample and subject metadata

All open-access data is available at gtexportal.org
Overview of GTEx resources: protected data

- Sequence data:
  - RNA-seq (2x76 bp, unstranded, >50M reads/sample)
  - WGS (30x coverage) and WES (100x coverage)
  - Illumina Omni2.5/5 microarray genotypes (subset of 450 donors)
- Allele-specific expression (ASE)
- Full sample and subject metadata
- Future: eGTEx sequence data
  - ChIP-seq
  - WGBS-seq

ESR1 query
Exon expression
## Significant Single-Tissue eQTLs for ESR1 (ENSG00000091831.17) in all tissues

Data Source: GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2)

ESR1 Gene eQTL Visualizer

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Variant ID</th>
<th>SNP</th>
<th>P-Value</th>
<th>NES</th>
<th>Tissue</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESR1</td>
<td>6_151998105_G_A_b37</td>
<td>rs1293942</td>
<td>2.2e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151998085_T_G_b37</td>
<td>rs1293943</td>
<td>2.2e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_152346190_TC_T_b37</td>
<td>rs113533024</td>
<td>2.4e-7</td>
<td>0.28</td>
<td>Tests</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_152000028_A_G_b37</td>
<td>rs712220</td>
<td>3.0e-7</td>
<td>-0.20</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151999603_A_G_b37</td>
<td>rs1293938</td>
<td>3.0e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151999507_C_G_b37</td>
<td>rs1293939</td>
<td>3.1e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151998723_A_G_b37</td>
<td>rs980280</td>
<td>3.1e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151990059_G_A_b37</td>
<td>rs1293956</td>
<td>3.7e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151990954_T_C_b37</td>
<td>rs1293955</td>
<td>3.8e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
<tr>
<td>ESR1</td>
<td>6_151990981_A_G_b37</td>
<td>rs1293954</td>
<td>4.3e-7</td>
<td>-0.21</td>
<td>Thyroid</td>
<td>eQTL box plot, IGV eQTL Browser, Multi-tissue eQTL Plot</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 204 entries
No splice QTLs and protein truncating variants found for ESR1
WebQTL software
GeneNetwork WebQTL

Select and Search

Species: Human (hg19)

Group: GTEx v5 All Tissues, RNA-Seq with Genotypes

Type: Breast - Mammary Tissue mRNA

Data Set: GTExv5 Human Breast Mammary Tissue RefSeq (Sep15) RPKM log2

Get Any: esr1

Enter terms, genes, ID numbers in the Get Any field. Use * or ? wildcards (Cypr*?, synap*).
Use Combined for terms such as tyrosine kinase.

Combined: Enter terms to combine (blood pressure): logical AND

Search  Make Default  Advanced Search

Websites Affiliated with GeneNetwork

- GeneNetwork Time Machine: Full versions from 2009 to 2016 (mm9)
- UTHSC Genome Browser Classic and Newest
- UTHSC Galaxy Service
- UTHSC Bayesian Network Web Server
- GeneNetwork Classic on Amazon Cloud
- GeneNetwork Classic Code on GitHub
- GeneNetwork 2.0 Development Code on GitHub
- GeneNetwork 2.0 Development

Getting Started

1. Select Species (or select All)
2. Select Group (a specific sample)
3. Select Type of data:
   - Phenotype (traits)
   - Genotype (markers)
   - Expression (mRNAs)
4. Select a Database
5. Enter search terms in the Get Any or Combined field: words, genes, ID numbers, probes, advanced search commands
6. Click on the Search button
7. Optional: Use the Make Default button to save your preferences

©2017 MFMER | 370227-69
Precision medicine for cancer patients using clinical and molecular data
Multi-dimensional data to individual patient

1. Multidimensional Data
2. Data Quality Analysis
3. Bioinformatics Pipelines
4. Integrative Modeling
5. Model Testing
6. Individual Patient
PANOPLY – Precision cancer genomics report: single sample inventory
Integration of multi-omics data for precision medicine

• PANOPLY- Precision Cancer Genomic Report: Single Sample Inventory

• PANOPLY: Omics-Guided Drug Prioritization Method Tailored to an Individual Patient.


https://hsr-rsc-prod.mayo.edu:1919/connect/#/apps/51/access

• http://mayo-pgrn-internal.northcentralus.cloudapp.azure.com/panoply_patients/index3.html

• http://www.kalarikrlab.org/Software/Panoply.html
Integration of 17 non-responder PANOPLY reports

Oncomatch – matching best cancer cell line to a patient
OncoMatch

CCLE Database

Matching Algorithms

Best match based on CNA, RNA, DNA combined

Surgery Biopsy

Patient derived Tumor Seq data

Sanger Drug Sensitivity Database

Mining the most sensitive and resistant drug for the cell-line/patient

OncoMatch
Non-coding reads
Unmapped host reads


HGT-ID: An efficient and sensitive workflow to detect human-viral insertion sites using next-generation sequencing data
Single-cell RNA-sequencing
Single Cell vs. Bulk Samples

Conclusions from bulk analysis can be representative of nothing

- Important for answering biological questions where cell-specific changes in transcriptome are important
- New protocols and lower sequencing cost

David Cook, SlideShare, 2017
Two scRNA-Seq Platforms at MGF

Fluidigm C1 (fluidic circuits) (non-UMI)

UMI=Unique Molecular Identifier

Chromium

10X Genomics

Single cell gene expression (Droplet) (UMI)

- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell

MAPRSeq pipeline

Tertiary analysis

10X Genomics Pipeline
## Fluidigm C1 vs. Chromium 10X Genomics

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Fluidigm C1</th>
<th>Chromium 10x Genomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Launched in</td>
<td>2012</td>
<td>10/2016</td>
</tr>
<tr>
<td>Principles (Reference)</td>
<td>Integrated fluidic circuits</td>
<td>Droplet-based</td>
</tr>
<tr>
<td>RNA-Seq solution</td>
<td>Full transcript</td>
<td>3’-tag</td>
</tr>
<tr>
<td>Throughput (# of cells analyzed)</td>
<td>Low-medium (48-800)</td>
<td>High (100-10,000+)</td>
</tr>
<tr>
<td>Visual Inspection</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Cell Selection</td>
<td>Yes (C1 size based)</td>
<td>No</td>
</tr>
<tr>
<td>Starting Amount of Cells</td>
<td>Medium-low</td>
<td>High</td>
</tr>
<tr>
<td>Flexibility (Own Protocols)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Advantage</td>
<td>Allows visual inspection of captured cells</td>
<td>High cell capture efficiency,</td>
</tr>
<tr>
<td></td>
<td>customizable protocols</td>
<td>cell size &lt;50µm, nuclei</td>
</tr>
<tr>
<td></td>
<td></td>
<td>suspensions can be studied,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lower system cost</td>
</tr>
<tr>
<td>Limitation</td>
<td>Size-based cell selection (C1)</td>
<td>High initial cell concentration</td>
</tr>
<tr>
<td></td>
<td>(5-10, 10-17, 17-25 µm)</td>
<td>required, no users modification</td>
</tr>
<tr>
<td></td>
<td></td>
<td>possible</td>
</tr>
</tbody>
</table>
Customized Tertiary Analysis

- **Healthy**
  - Tissues
  - Single-cell RNA-seq
  - Expression profile clustering
  - Cell-type maps

- **Pathological**
  - Disease-associated cells

**Types of analyses**

**Within cell type**
- Stochasticity, variability of transcription
- Regulatory network inference
- Allelic expression patterns
- Scaling laws of transcription

**Between cell types**
- Identify biomarkers
- (Post)-transcriptional differences

**Between tissues**
- Cell-type compositions
- Altered transcription in matched cell types
Unsupervised clustering after gene expression analysis of single blood cells isolated from the whole kidney marrow. Heat map shows high transcript expression in red and low/absent expression in blue. Four major clusters were identified, including the following: erythroid (red), myeloid (green), B cells (light blue), and T cells (dark blue).

http://jem.rupress.org/content/213/6/979
Unsupervised clustering

- Violin plots show the distribution of gene expression of single cells. Cells types were assigned based on hierarchical clustering and assessed for transcript expression of well-known blood cell lineage genes.

http://jem.rupress.org/content/213/6/979
Impact

Prophylactic oophorectomy

- In women who have a known BRCA mutation, prophylactic oophorectomy can decrease breast cancer incidence by 50%.
Questions & Discussion