Predicting outcome from cancer data

Jeffrey Chuang

The Jackson Laboratory for Genomic Medicine
Drowning in data, thirsting for knowledge

What is AACR Project GENIE

NCI-MATCH Trial (Molecular Analysis for Therapy Choice)

TARGET: Therapeutically Applicable Research To Generate Effective Treatments

PDX DEVELOPMENT AND TRIAL CENTERS RESEARCH NETWORK
The median 5-year survival rate for the best treatment across cancers is 55%.
Diverse cancer data can be used to predict outcome

Tumor Genomics and Heterogeneity

Gene Expression

Tumor Microenvironment

http://compbio.cs.brown.edu/projects/btp/

https://www.nature.com/articles/nature12626
Mutation rates vary within and across cancer types


Cancer typically have 10-1000 coding mutations in their genome
Recurrent mutations are driver candidates

Driver prediction requires data processing and algorithmic standardization

Driver genes predicted from:

- Prevalence of mutations in gene relative to background
- Ratio of different types of mutations in gene (e.g. protein-changing vs. silent)

Computational driver prediction methods vary in their estimates of the number of driver genes.

Challenges of Testing Driver Candidates

Clinical trials take time and resources.

Difficult to compare alternative treatments on same patient.

Distinguishing effects of different mutations requires many patients.
Too Many Trials, Not Enough Participants

A Cancer Conundrum: Too Many Drug Trials, Too Few Patients

By GINA KOLATA  AUG. 12, 2017
Patient-Derived Xenografts (PDXs) can be used to test drug sensitivity of tumors.
Primary tumors and xenografts have the same mutations
The National Cancer Institute launched the PDX Development and Trial Centers Research Network (PDXNet) in September 2017 to accelerate translational research as part of the Cancer Moonshot Initiative.
PDXnet Consortium

- **PDTCs – U54**

<table>
<thead>
<tr>
<th>Code</th>
<th>Institution</th>
<th>Tumor Type</th>
<th>Number</th>
<th>PI3K/MAPK</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA224083</td>
<td>Govindan Washington U.</td>
<td>Breast/PDAC</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>CA224070</td>
<td>Herlyn Wistar Institute</td>
<td>Melanoma</td>
<td>450</td>
<td>PI3K/MAPK in mBRAF</td>
</tr>
<tr>
<td>CA224065</td>
<td>Roth MDAnderson</td>
<td>Lung/PDAC/CRC</td>
<td>500</td>
<td>MAPK and DDR</td>
</tr>
<tr>
<td>CA224076</td>
<td>Welm Huntsman/Baylor</td>
<td>Breast/Gyn</td>
<td>162</td>
<td>Empiric screening/TICs</td>
</tr>
</tbody>
</table>

- **PDC&CC – U24**

<table>
<thead>
<tr>
<th>Code</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA224067</td>
<td>Chuang Jackson/Seven Bridges</td>
</tr>
</tbody>
</table>

The PDXNet comprises four PDX Development and Trial Centers, the PDX Data Commons and Coordinating Center, Frederick National Laboratory for Cancer Research, and the National Cancer Institute that will jointly perform hundreds of PDX treatment studies and develop clinical trials.
PDXNet Integrates with the NCI Cancer Genomics Cloud to Standardize Data Analysis

- Search capabilities across broad cancer datasets including PDXNet, TCGA, and TARGET
- Rapid data and workflow visualization
The Cancer Genomics Cloud Includes >350 Quality-Controlled Analysis Workflows
Current Pilot Projects in PDXnet

- Comparative testing of drug efficacy with alternate treatment protocols
- Comparative testing of mutation calling and driver identification pipelines on shared data.
- Multi-cancer studies of evolution and resistance processes within tumors
PDXnet summary

• PDXnet will yield datasets with controlled phenotyping and vocabularies following the PDX minimal information standards, allowing systematic data mining.

• These data will be used to develop a jointly planned clinical trial.

• PDXnet data and analysis pipelines will be shared with the general research community.

• PDXnet xenograft models will be available to researchers via the NCI Patient Derived Model Repository.
Research Vignettes
Intratumoral evolution with xenografts
Resistance arises from tumor heterogeneity

Resistant intratumoral populations are selected for by treatment.

Evolution in tumors is a dynamic process of selection, mutation, and drift.

Mutation analysis of triple negative breast cancer PDXs reveals intratumoral evolution

Post-treated tumor samples vary in their cell composition

Hyunsoo Kim
Collaboration with E. Liu
Intratumoral evolution is susceptible to dynamic therapy strategies

One subclone (red) grows faster in the absence of cisplatin but is disfavored in its presence.

Treating to reduce the red cells, but not eradicate them, may slow development of resistance,
Gene expression evolution during metastasis
Using gene expression to predict survival

Primary-metastasis comparisons in triple negative breast cancer
Differentially expressed genes between paired recurrent and primary TNBCs

Clustering of 455 differentially expressed genes among 8 paired metastatic/primary TNBCs
Over-expression of CCNE1 is associated with poor survival

Kaplan-Meier overall survival curves for CCNE1 using METABRIC (Total, N=1981) and TCGA (N=1072) breast cancer databases.
Using immune cell infiltration to predict outcome
Tumor immune infiltration predicts clinical outcome. Example: Astrocytoma

- cytotoxic T cells (CD8; red)
- Th cells (CD4; brown)
- microglia/macrophages (AIF1; brown)

High combined immune cell infiltration of
- cytotoxic T cells
- Th cells
- microglia/macrophages
is associated with better survival in astrocytoma

Donson et al. J. Immunology 2012
Example: Neutrophil-to-lymphocyte ratio predicts survival in pancreatic cancer

Significance of baseline and change in neutrophil-to-lymphocyte ratio in predicting prognosis: a retrospective analysis in advanced pancreatic ductal adenocarcinoma

Yang Chen, Huan Yan, Yanfong Wang, Yan Shi, & Guanghai Dai

Scientific Reports 7, Article number: 753 (2017)
doi:10.1038/s41598-017-00859-5
Received: 06 December 2016
Accepted: 15 March 2017

High neutrophil / lymphocyte ratio (NLR) is associated with worse survival in PDAC

https://www.nature.com/articles/s41598-017-00859-5
Large cancer datasets allow rapid mining of immune effects

High neutrophil infiltration predicts worse survival in UCEC, OV & PRAD. Analysis of TCGA data for survival based on pre-treatment neutrophil infiltration

Javad Noorbakhsh  Zi-Ming Zhao
Classifying histological features in tumor images: connecting phenotype, genotype and clinical outcome

Convolutional neural networks for intratumoral histology classification

Javad Noorbakhsh
Summary

- Large cancer datasets can be used to predict outcome, such as tumor heterogeneity, gene expression, and tumor microenvironment.

- PDX and organoid models can be used to identify features that predict outcome from cancer sequencing data. PDXnet provides a powerful infrastructure to robustly analyze these data with other large datasets and to develop clinical trials.

- Evolution, gene expression, and immune infiltration are critical features that will also be empowered for outcome prediction by large cancer datasets.
Acknowledgements

Chuang Lab
Hyunsoo Kim Javad Noorbakhsh Zi-Ming Zhao Victor Wang
Scott Adamson Ivan Dotu Ada Zhan Minji Kim
Yuliana Tan Sherry Zhang

Edison Liu
Francesca Menghi Pooja Kumar Phung Shreckengast

Charles Lee
CZ Zhang Eliza Cerveira Mallory Ryan Ankit Malhotra
Qihui Zhu

Krish Karuturi
Joshy George Anuj Srivastava Xingyi Woo Vishal Kumar Sarsani Preeti Bais

Carol Bult Peter Robinson Karolina Palucka Olga Anczukow-Camarda
Steve Neuhauser Paul Robson Jim Keck

Young Kwang Chae (Northwestern) Yuan Yuan (City of Hope)

Seven Bridges Genomics Brandi Davis-Dusenbery Anurag Sethi Christian Frech

NCI (P30CA034196, R21CA191848-01A1, U24CA224067-01).