Computational Challenges and Opportunities in the Era of the Pre-Cancer Genome Atlas

Jennifer Beane, Ph.D.
Boston University School of Medicine
Section of Computational Biomedicine
# The Cancer Genome Atlas (TCGA)

## Tumor Specimens
- 11,000 end-stage tumors
- 33 tumor types
  - Leukemia (LAML)
  - Lung adenocarcinoma (LUAD)
  - Lung squamous (LUSC)
  - Kidney (KIRC)
  - Bladder (BLCA)
  - Endometrial (UCEC)
  - Glioblastoma (GBM)
  - Head and neck (HNSC)
  - Breast (BRCA)
  - Ovarian (OV)
  - Colon (COAD)
  - Rectum (READ)

## Mulit-omic Profiling
- Platforms
- Genes/loci
- Clinical data
- Samples
- Gene expression
- DNA methylation
- MicroRNA
- RPPA
- Mutation
- Copy number

## Broad Goals
- Identify driver mutations
- Identify altered pathways
- Molecular classification
- Pan-Cancer analysis
- Improve cancer treatment
- Functional testing

TCGA, Nature Genetics., 2013
TCGA results for Lung Squamous Cell Carcinoma

Molecular classification

- **Classic (36%)**
  (xenobiotic metabolism)
  KEAP1, NFE2L2, SOX2, TP63 (deltaN), PIK3CA hypermethylation, CIN

- **Primitive (15%)**
  (proliferation/worst survival)
  RB1, PTEN

- **Basal (25%)**
  (cell adhesion)
  NF1

- **Secretory (24%)**
  (immune response)

Identification of Significantly Mutated Genes

Somatically Altered Pathways

- Squamous differentiation
  - SOX2: 21%
  - TP63: 16%
  - NOTCH1: 8%
  - NOTCH2: 5%
  - ASCL4: 3%
  - FOXp1: 4%

- 69% of samples with a potentially targetable alteration

The Unmet Need in Early Detection and Prevention of Cancer

Kensler et al, Cancer Prevention Research, 2016

Campbell et al, Cancer Prevention Research, 2016

TCGA focus

Esophageal AD
Melanoma
Lung AD
Case Study: Bronchial premalignant lesions

Squamous Lung Carcinogenesis

Activation by carcinogens → DNA adduct formation → Mutation, epigenetic gene inactivation → Acquisition of hallmarks of cancer

Normal → Squamous metaplasia → Mild dysplasia → Moderate dysplasia → Severe dysplasia → Carcinoma in situ

Many lesions will regress and only a subset will progress to CIS

Key question:
Can molecular profiling give additional insights into classification and what drives progression or regression of premalignant lesions?

Early Detection Biomarkers and Intervention Therapies

Rationale for extending TCGA-like Studies to Longitudinal Sampled Premalignant Tissues (PCGA)

1. Identification of driver genes

2. Identify the sequence of events

3. Molecular classification

4. Characterize cell population changes

Example (#2) Shared mutational reveal clonal populations in two adjacent CIS. An earlier time point reveals a potential evolutionary ancestor.
Example (#3) Molecular classification and gene/miRNA regulation can identify lesions that will progress and can be used as biomarkers or to inform early intervention strategies.

Gene expression alterations associated with biopsy histological grade

**Mostly Low Grade**

- Subtype 1
- Subtype 2
- Subtype 3

**Mostly High Grade**

- Severe Dysplasia/CIS
- Mild/Moderate Dysplasia
- Metaplasia
- Normal/Hyperplasia

- Associated with increased lesion progression

- miR-182
- FOXF2

- miRNA
- mRNA
- Transcription Factor
- Positive Correlation
- Negative Correlation

- cell migration
- cell adhesion
- EMT
Summary

- Multi-omic profiling of longitudinally collected samples provides unique computational opportunities and challenges
- A PCGA initiative will allow us to advance cancer early detection and prevention
  - Identification of molecular alterations that drive disease progression
  - Serve as biomarkers for early detection of as surrogate endpoints to inform chemoprevention trials
  - Identify novel and personalized early interventions
## Acknowledgements

<table>
<thead>
<tr>
<th>BUMC</th>
<th>DFCL/Broad</th>
<th>Roswell Park Cancer Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avrum Spira</td>
<td>Matthew Meyerson</td>
<td>Mary Reid</td>
</tr>
<tr>
<td>Joshua Campbell</td>
<td>Angela Brooks</td>
<td>Samjot Dhillon</td>
</tr>
<tr>
<td>Marc Lenburg</td>
<td>Alice Berger</td>
<td>Mary Beth Pine</td>
</tr>
<tr>
<td>Gang Liu</td>
<td>Guangwu Guo</td>
<td></td>
</tr>
<tr>
<td>Sherry Zhang</td>
<td>Brad Murray</td>
<td></td>
</tr>
<tr>
<td>Hanqiao Liu</td>
<td>Andrew Cherniack</td>
<td></td>
</tr>
<tr>
<td>Sarah Mazzilli</td>
<td>Marcin Imielinski</td>
<td></td>
</tr>
<tr>
<td>Yaron Gesthalter</td>
<td>Chandra Pedamallu</td>
<td></td>
</tr>
<tr>
<td>Liye Zhang</td>
<td>Josh Francis</td>
<td></td>
</tr>
<tr>
<td>David Jenkins</td>
<td>ChengZhong Zhang</td>
<td></td>
</tr>
<tr>
<td>Ania Tassinari</td>
<td>Xiaoyang Zhang</td>
<td></td>
</tr>
<tr>
<td>Hanqiao Liu</td>
<td>Peter Choi</td>
<td></td>
</tr>
<tr>
<td>Jessica Vick</td>
<td>Alison Taylor</td>
<td></td>
</tr>
<tr>
<td>Stefano Monti</td>
<td>Diana Cai</td>
<td></td>
</tr>
<tr>
<td>Evan Johnson</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>