Opportunities and Lessons Learned from AACR’s Project GENIE

Presented By:
Mia Levy, MD, PhD
Vanderbilt University representative to the GENIE Steering Committee and Chair of the GENIE Operations Sub-committee
10/29/2019
Disclosures

- **Consulting/Advisory:** GenomOncology, Personalis, Roche
- **Research funding:** Pfizer, BMS, GenomOncology
- **Equity:** GenomOncology, Personalis
GENIE Overview

- International pancancer registry built through data sharing
  - Driven by openness, transparency, and inclusion

- GOAL: improve clinical decision making
  - Linking clinical genotype to clinical outcomes

- Eight founding participants, now 19
  - North America & Europe
  - Plans for future expansion

- Sponsored research
- Collaborative projects
- Grants and philanthropy
Getting Started

- Initial Concept for Project GENIE
- Think tank of potential participants
- Project GENIE approved by AACR Board of Directors
- Budget and business plan developed
- Public Announcement
- Summer Summit
- Winter Summit
- Summer Summit
- Data Freeze Meeting
- Winter Summit
- First Public Data Release
- Sponsored research projects start

January 2014

January 2015

January 2016

January 2017

June 2017

DOI: 10.1200/CCI.17.00083 JCO Clinical Cancer Informatics - published online February 16, 2018
Continued Progress

- 1st public release: 1/05/2017
  - 18,860 total sequenced tumors

- 2nd release: 11/22/2017
  - 31,673 total sequenced tumors

- 3rd release: 01/09/2018
  - 39,600 total sequenced tumors

- 4th release: 07/16/2018
  - 48,500 sequenced tumors

- 5th release: 01/08/2019
  - ~60,000 sequenced tumors

GENIE will be three November 6, 2018!
Steering Committee makes decisions by majority vote
GENIE Today

GENOMICS
✓ Somatic Tumor DNA

PHENOMICS
Tumor type
Histology
Demographics
Vital status

47,500 Tumors
8 Cancer Centers
Data made publicly available 12 months after date of sequencing

Sponsored Research

PHENOMICS
Tumor type
Histology
Demographics
Vital status
Detailed Clinicopathology
Prior Tx
Outcomes

Specific Cohorts
Variable # of Centers
Data made public at time of publication

www.aacr.org/genie
How the Registry Operates: Baseline Data

- Data mapped to common ontology and harmonized
- Limited PHI removed
- Data governance, provenance, and versioning in a secure, HIPAA-compliant environment.

www.aacr.org/genie/data
Define Virtual Cohorts of Interest

Cancer Type Gene(s) of interest

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>#</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Small Cell Lung</td>
<td>7682</td>
<td>15.86%</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>5506</td>
<td>11.36%</td>
</tr>
<tr>
<td>Colorectal Cancer</td>
<td>5193</td>
<td>10.72%</td>
</tr>
<tr>
<td>Glioma</td>
<td>2651</td>
<td>5.47%</td>
</tr>
<tr>
<td>Melanoma</td>
<td>2163</td>
<td>4.46%</td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td>1827</td>
<td>3.77%</td>
</tr>
<tr>
<td>Ovarian Cancer</td>
<td>1733</td>
<td>3.58%</td>
</tr>
<tr>
<td>Leukemia</td>
<td>1681</td>
<td>3.47%</td>
</tr>
<tr>
<td>Pancreatic Cancer</td>
<td>1670</td>
<td>3.45%</td>
</tr>
<tr>
<td>Soft Tissue Sarcoma</td>
<td>1459</td>
<td>3.01%</td>
</tr>
<tr>
<td>Endometrial Cancer</td>
<td>1363</td>
<td>2.81%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cancer Type Detailed</th>
<th>#</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Adenocarcinoma</td>
<td>6057</td>
<td>12.50%</td>
</tr>
<tr>
<td>Breast Invasive Ductal</td>
<td>3702</td>
<td>7.64%</td>
</tr>
<tr>
<td>Colon Adenocarcinoma</td>
<td>2876</td>
<td>5.94%</td>
</tr>
<tr>
<td>Prostate Adenocarcinoma</td>
<td>1791</td>
<td>3.70%</td>
</tr>
<tr>
<td>Pancreatic Adenocarcinoma</td>
<td>1322</td>
<td>2.73%</td>
</tr>
<tr>
<td>Colorectal Adenocarcinoma</td>
<td>1314</td>
<td>2.71%</td>
</tr>
<tr>
<td>Acute Myeloid Leukemia</td>
<td>1140</td>
<td>2.35%</td>
</tr>
<tr>
<td>Cutaneous Melanoma</td>
<td>946</td>
<td>1.95%</td>
</tr>
<tr>
<td>Glioblastoma Multiforme</td>
<td>941</td>
<td>1.94%</td>
</tr>
<tr>
<td>Bladder Urothelial Carcinoma</td>
<td>927</td>
<td>1.91%</td>
</tr>
<tr>
<td>High-Grade Serous Ovarian Malignan</td>
<td>838</td>
<td>1.73%</td>
</tr>
</tbody>
</table>

Mutated Genes (48447 profiled samples)

<table>
<thead>
<tr>
<th>Gene</th>
<th># Mut</th>
<th># Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>18923</td>
<td>35.60%</td>
</tr>
<tr>
<td>KRAS</td>
<td>6898</td>
<td>14.19%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>5813</td>
<td>10.97%</td>
</tr>
<tr>
<td>APC</td>
<td>6416</td>
<td>10.06%</td>
</tr>
<tr>
<td>KMT2D</td>
<td>4529</td>
<td>9.23%</td>
</tr>
<tr>
<td>ARID1A</td>
<td>3519</td>
<td>8.01%</td>
</tr>
<tr>
<td>TERT</td>
<td>3080</td>
<td>8.46%</td>
</tr>
<tr>
<td>BRAF</td>
<td>2810</td>
<td>5.57%</td>
</tr>
<tr>
<td>PTEN</td>
<td>3085</td>
<td>5.30%</td>
</tr>
<tr>
<td>EGFR</td>
<td>3133</td>
<td>5.35%</td>
</tr>
<tr>
<td>ATM</td>
<td>2804</td>
<td>4.98%</td>
</tr>
</tbody>
</table>

www.aacr.org/genie
How the Registry Operates: Detailed Clinical Data

Clinical queries are posed based on registry content.

Clinical data required to answer the question are manually abstracted.

Genomic and clinical data linked.

Consortium/sponsor-only access to time of publication.

www.aacr.org/genie
A Cancer Patient’s Journey

- Demographics
- Vital status

• Pancancer
• Cancer-specific

Treatments

OUTCOMES

n
GENIE of Tomorrow (2.0)

GENOMICS
- Somatic Tumor DNA
  - Germline DNA
  - cfDNA
  - RNA Seq
  - Epigenetics

PHENOMICS
- Tumor type
- Histology
- Demographics
- Vital status
- Medications
- Treatment Outcomes

100,000 Tumors
19+ Cancer Centers

Data to Drive Discoveries
Diverse Gene Panels

Panel Sizes

<table>
<thead>
<tr>
<th>Panel</th>
<th>Genes Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>VICC-01-T5a</td>
<td></td>
</tr>
<tr>
<td>VICC-01-myeloid</td>
<td></td>
</tr>
<tr>
<td>UHN-54-V1</td>
<td></td>
</tr>
<tr>
<td>UHN-48-V1</td>
<td></td>
</tr>
<tr>
<td>MSK-IMPACT468</td>
<td></td>
</tr>
<tr>
<td>MSK-IMPACT341</td>
<td></td>
</tr>
<tr>
<td>MDA-46-V1</td>
<td></td>
</tr>
<tr>
<td>JHU-50GP-V2</td>
<td></td>
</tr>
<tr>
<td>GRCC-MOSC3</td>
<td></td>
</tr>
<tr>
<td>DFCI-…</td>
<td></td>
</tr>
<tr>
<td>DFCI-…</td>
<td></td>
</tr>
</tbody>
</table>

- Total Genes Covered: 1324
- Shared by All Panels: 8
- Shared by All Large Panels: 143

Genes Shared by All Panels
- BRAF
- HRAS
- IDH1
- KIT
- KRAS
- NRAS
- PTEN
- TP53

Shared by All Large and Solid Tumor Panels
- BRAF
- AKT1
- HRAS
- ALK
- IDH1
- CTNNB1
- KIT
- EGFR
- KRAS
- ERBB2
- NRAS
- FGFR1
- PTEN
- FGFR2
- TP53
- FGFR3
- PDGFRA
- MET
- PIK3CA
- RET
Distribution of Samples Across Centers

GENIE Contributions by Center

- DFCI: 31.68%
- GRCC: 2.06%
- JHU: 7.15%
- MDA: 8.47%
- MSK: 39.46%
- NKI: 1.87%
- UHN: 4.63%
- VICC: 4.67%
Racial and Gender Distribution

**GENIE Patient Demographics**

<table>
<thead>
<tr>
<th>Race</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>41.83%</td>
<td>36.40%</td>
</tr>
<tr>
<td>Black</td>
<td>3.35%</td>
<td>2.08%</td>
</tr>
<tr>
<td>Asian</td>
<td>2.60%</td>
<td>1.83%</td>
</tr>
<tr>
<td>Native American</td>
<td>0.07%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Other</td>
<td>1.05%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Undefined</td>
<td>0.04%</td>
<td>0.78%</td>
</tr>
<tr>
<td>Unknown</td>
<td>4.86%</td>
<td>4.99%</td>
</tr>
</tbody>
</table>

**Gender Breakdown:**
- **Female:** White (41.83%), Black (3.35%), Asian (2.60%), Native American (0.07%), Other (1.05%), Undefined (0.04%), Unknown (4.86%)
- **Male:** White (36.40%), Black (2.08%), Asian (1.83%), Native American (0.06%), Other (0.06%), Undefined (0.78%), Unknown (4.99%)
Distribution by Patient Age at Sequencing

Median Age = 60
Distribution of Samples by Cancer Type

- Primary tumor
- Lymph node metastasis
- Distant organ metastasis
- Metastasis site unspecified
Top Mutated Genes

Top 50 Mutated Genes
Genomic Alterations in NSCLC
Landscape of Clinical Actionability

Nikolaus Schultz, MSKCC
658 enrolling trials testing a targeted therapy AND with actionable variants in inclusion criteria resulted in 16,000 matches to GENIE patients.
NCI-MATCH (release 3 data)

<table>
<thead>
<tr>
<th>Arm</th>
<th>Description</th>
<th>Patients in GENIE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>NF1 inactivating mts</td>
<td>1808</td>
</tr>
<tr>
<td>I</td>
<td>PK3CA muts (No KRAS, PTEN mts; No breast carcinoma, lung squamous cell carcinoma)</td>
<td>1528</td>
</tr>
<tr>
<td>W</td>
<td>FGFR1-3 mts or mts</td>
<td>985</td>
</tr>
<tr>
<td>Z1B</td>
<td>CCND1/2/3 mts (No breast carcinoma, mantle cell lymphoma or myeloma)</td>
<td>765</td>
</tr>
<tr>
<td>U*</td>
<td>NF2 inactivating mts</td>
<td>595</td>
</tr>
<tr>
<td>Z1A</td>
<td>NRAS mts (No melanoma)</td>
<td>421</td>
</tr>
<tr>
<td>Q</td>
<td>HER2 amp (No breast carcinoma, gastric/GES adenocarcinoma)</td>
<td>356</td>
</tr>
<tr>
<td>B</td>
<td>HER2 activating mts (No NSCLC)</td>
<td>261</td>
</tr>
<tr>
<td>H</td>
<td>BRAF V600 mts (No melanoma, papillary thyroid cancer, colorectal adenocarcinoma)</td>
<td>247</td>
</tr>
<tr>
<td>C1</td>
<td>MET mts</td>
<td>202</td>
</tr>
<tr>
<td>Y</td>
<td>AKT E17K mts (No KRAS, NRAS, HRAS, or BRAF mts)</td>
<td>201</td>
</tr>
<tr>
<td>V</td>
<td>cKIT mts (No GIST, renal cell carcinoma, PNET)</td>
<td>153</td>
</tr>
<tr>
<td>E</td>
<td>EGFR T790M (No lung adenocarcinoma) and rare EGFR activating mts</td>
<td>136</td>
</tr>
<tr>
<td>S2</td>
<td>GNAQ or GNA11 mts (No uveal melanoma)</td>
<td>32</td>
</tr>
<tr>
<td>Z1D</td>
<td>dMMR (No colorectal cancer)</td>
<td>30</td>
</tr>
<tr>
<td>A</td>
<td>EGFR activating mts (No SCLC or NSCLC)</td>
<td>18</td>
</tr>
<tr>
<td>T</td>
<td>SMO or PTCH1 mts (No basal cell carcinoma)</td>
<td>10</td>
</tr>
<tr>
<td>X</td>
<td>DDR2 5768R, I638F or L239R mts</td>
<td>0</td>
</tr>
</tbody>
</table>

The diagram shows the percentage of patients matched based on the arm and description, with categories for Bowel, Lung, Ovary/Fallopian Tube, Uterus, Bladder/Urinary Tract, Esophagus/Stomach, CNS/Brain, Head and Neck, and Other.
Challenges with Real-World Data

- Missing treatment information pre- and post- treatment at sequencing institution
- Missing diagnostic information, especially biomarker assessments
- Lack of information on responses to therapy
- Lack of information on reasons for discontinuation of therapy
- “Fuzzy” dates
- Resource intensive effort to gather data from outside medical records
- Limited ability of some institutions to release information for patients on clinical trials
- Relatively high proportion of patients lost to follow-up
Many Are Looking at Different Parts of the Same Problem
Patient-driven initiatives

- Count Me In (Broad Institute)
  - Metastatic Breast Cancer Project
  - Angiosarcoma Project
  - Metastatic Prostate Cancer Project
  - ...

- Make an IMPACT (MSKCC)

- Similar efforts emerging

- Thousands of patients joining to share clinical and molecular data to advance research

mpcproject.org
Conclusions

- Patient as aggregator/donor of data
  - Regulation/Policy to make easier for patient to access, download, transfer their data
  - Patient “rights” to use data “What is it you know about me?”
- Funding agencies should require data deposition and require that portion of funding is used for work required to clean and deposit data
- Insufficient infrastructure for public deposition (storage, upload, download, data use)
- More publicly accessible training data for machine learning community
- Data Authorship (credit for secondary use of deposited data)
- Caution FDA to not inadvertently restrict data coming from assays
Thank you