Functionalizing the Cancer Genome

Lynda Chin
Harvard Medicine School
Belfer Institute for Applied Cancer Science
Dana-Farber Cancer Institute
Broad Institute

Disclosure

AVEO Pharmaceuticals: co-founder and advisor
Metamark Genetics: founder, director and advisor
Eden; Epizyme; Agios: Consultants;
GSK: Sponsored Research;
Merck; sanofi-aventis: Corporate alliance partnerships
Major Goals in Cancer Medicine

- Prevention
- Detection
- Intervention

Genome Science → Personalized Medicine
Potential of Cancer Genomics

- Enable prevention
  - Understanding the underlying etiology → strategy

- Facilitate early detection
  - Identify risk alleles / genomic events for screening
  - Early events may be detectable in serum or by imaging

- Guide evidence-based intervention
  - Stratify high vs low risk patients to treat or not
  - Identify new therapeutic targets for drug discovery
  - Inform selection of the right patient for the right drug
  - Define combination / co-extinction strategies

⇒ Personalized Cancer Medicine
TCGA Phase II

Pilot
2007 | 2008 | 2009

GBM Report

Funding
2010 | 2011

Ovarian Report

Continuation
2012 - 2014

Complete Data on 3000 New Cases

Number of Cases


Kenna Shaw, NCI
# TCGA Phase II Projects

<table>
<thead>
<tr>
<th>Tissue</th>
<th>TCGA Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>GBM and low-grade gliomas</td>
</tr>
<tr>
<td>Breast</td>
<td>Ductal &amp; lobular breast adenocarcinomas</td>
</tr>
<tr>
<td>Stomach</td>
<td>Intestinal-type gastric adenocarcinoma</td>
</tr>
<tr>
<td>Liver</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>Intestine</td>
<td>Colon and rectal adenocarcinomas</td>
</tr>
<tr>
<td>Gynecologic</td>
<td>Serous ovarian adenocarcinoma; endometrial and cervical squamous carcinomas</td>
</tr>
<tr>
<td>Prostate</td>
<td>Prostate adenocarcinoma</td>
</tr>
<tr>
<td>Bladder</td>
<td>Non-papillary bladder cancer</td>
</tr>
<tr>
<td>Head and Neck</td>
<td>Squamous cell and thyroid papillary carcinomas</td>
</tr>
<tr>
<td>Hematopoietic</td>
<td>Acute myeloid leukemia</td>
</tr>
<tr>
<td>Skin</td>
<td>Metastatic cutaneous melanoma</td>
</tr>
<tr>
<td>Lung</td>
<td>Non-small cell lung cancer, adenocarcinoma and squamous subtypes</td>
</tr>
<tr>
<td>Kidney</td>
<td>Renal clear cell and renal papillary carcinomas</td>
</tr>
<tr>
<td>Pancreas</td>
<td>Pancreatic adenocarcinoma</td>
</tr>
</tbody>
</table>
Active Tumor Projects

Timeline to Completion of Comprehensive Analysis for Each Tumor Project

- Colon (and rectal)
- Lung Squamous
- AML
- Kidney (Clear Cell)
- Breast (Ductal + Lobular)
- Lung Adeno
- Endometrial
- Bladder Non Papillary
- Stomach
- Prostate
- Melanoma
- Thyroid
- LGG
- Liver
- Cervical
- Kidney Papillary
- Lymphoma

Timeline:
- Mar-11
- Jun-11
- Sep-11
- Dec-11
- Mar-12
- Jun-12
- Sep-12
- Dec-12
- Mar-13
- Jun-13
- Sep-13
- Dec-13
- Mar-14
- Jun-14
Massively Parallel Sequencing

OLD

NEW

Point mutation
Indel
Homozygous deletion
Hemizygous deletion
Gain
Translocation Breakpoint
Pathogen

Copy-number alterations

chr1 reference sequence
chr5 Non-human sequence

Gad Getz
Scale of Growth is unprecedented

- 240 tumor cases/month = 480 exome files (38 – 52 Mbases)
- 48 whole genomes (3,000 Mbases)

Exome (16 Gb)  Genome (200 Gb)

7,700 Gbytes  9,600 Gbytes

Per month 17.3 Terabytes TOTAL

NHGRI Center Production

After NextGen

- FY2008
- FY2009
- FY2010*

Brad Ozenberger, NHGRI
# Summary of TCGA Tumor Data

**Ingested into Broad GDAC Pipeline**

01/14/2011 Run

<table>
<thead>
<tr>
<th>TumorType</th>
<th>Biospecimen</th>
<th>Any Level 1</th>
<th>Clinical</th>
<th>CNA</th>
<th>Methylation</th>
<th>mRNA</th>
<th>miR</th>
<th>MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA</td>
<td>346</td>
<td>186</td>
<td>244</td>
<td>265</td>
<td>186</td>
<td>280</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>COAD</td>
<td>203</td>
<td>151</td>
<td>130</td>
<td>137</td>
<td>167</td>
<td>155</td>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>GBM</td>
<td>508</td>
<td>448</td>
<td>490</td>
<td>466</td>
<td>288</td>
<td>444</td>
<td>415</td>
<td>169</td>
</tr>
<tr>
<td>HNSC</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KIRC</td>
<td>355</td>
<td>39</td>
<td>19</td>
<td>254</td>
<td>219</td>
<td>41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KIRP</td>
<td>48</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LAML</td>
<td>202</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>188</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LGG</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LUAD</td>
<td>128</td>
<td>21</td>
<td>11</td>
<td>56</td>
<td>128</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LUSC</td>
<td>160</td>
<td>116</td>
<td>42</td>
<td>117</td>
<td>133</td>
<td>116</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OV</td>
<td>584</td>
<td>570</td>
<td>532</td>
<td>519</td>
<td>425</td>
<td>519</td>
<td>566</td>
<td>384</td>
</tr>
<tr>
<td>READ</td>
<td>79</td>
<td>52</td>
<td>72</td>
<td>51</td>
<td>69</td>
<td>69</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>STAD</td>
<td>82</td>
<td>35</td>
<td>0</td>
<td>81</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UCEC</td>
<td>145</td>
<td>24</td>
<td>0</td>
<td>114</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| Totals    | 2909        | 1681        | 1540     | 2076 | 1991        | 1698 | 981 | 629 |
Status of TCGA Analysis Pipeline (Jan 14, 2010 Run)

Mike Noble; Doug Voet

- Pre-defined analyses
- Automated and Fast, producing standard human-readable summary reports
- Reproducible
  - Associated input, algorithms and parameters are tracked
  - Uniform intermediate data files for higher level analyses

Firehose Workflow
Complete catalogues will be generated

Complete Compendia

What does it take?
Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA, IDH1, EGFR, and NF1*

Roel G.W. Verhaak,1,2,17 Katherine A. Hoadley,3,4,17 Elizabeth Purdom,7 Victoria Wang,8 Yuan Qi,4,5 Matthew D. Wilkerson,4,5 C. Ryan Miller,4,6 Li Ding,9 Todd Golub,1,10 Jill P. Mesirov,7 Gabriele Alexe,1 Michael Lawrence,1,2 Michael O’Kelly,1,2 Pablo Tamayo,1 Barbara A. Weir,1,2 Stacey Gabriel,1 Wendy Winckler,1,2 Supriya Gupta,1 Lakshmi Jakkula,11 Heidi S. Feiler,11 J. Graeme Hodgson,12 C. David James,12 Jarr N. Sarkaria,13 Cameron Brennan,14 Ari Kahn,15 Paul T. Spellman,11 Richard K. Wilson,9 Terence P. Speed,7,16 Joe W. Gray,11 Matthew Meyerson,1,2 Gad Getz,1 Charles M. Perou,3,4,8 D. Neil Hayes,4,5,1 and The Cancer Genome Atlas Research Network
What is driving the molecular difference among subtypes?

The most significant difference is observed between PN and MS subtypes.
CLR - Context Likelihood of Relatedness

- Extension of relevance networks
- Based on Mutual Information scores

Ayla Ergun; Jim Collins
Define miRNA-mRNA regulatory network

microRNA expression

mRNA expression

Å TCGA Glioblastoma data

534 miRNA, 19692 mRNAs

29610 edges: 252 miRNA and 7373 mRNAs

194 matched samples
A subset of the miRNAs show strong correlations with subtype signature genes.

17 miRNA with strong correlations with the proneural and mesenchymal signature genes.
miR34a as a candidate determinant of PN molecular subtype

- Integration with copy number reveals miR34a resides in region of loss
- miR34a is low in PN subtype GBM
- PN signature is enriched for miR34a edges defined by CLR
miR34a is tumor suppressive in human GBM models \textit{in vivo}

**LOF studies in Human GBM Cells**

- LN319 scr n=5
- LN319 mir-34a n=5
- A1207 scr n=5
- A1207 mir-34a n=5

**GOF studies in E6/E7 Astrocytes**

- Decoy
- Ctrl

A1207 pHAGE ct

A1207 pHAGE mir-34a
How does miR34a regulate the PN/MS transcriptomic signatures?

- 3'UTR Luciferase reporter: Direct regulation of PDGFRa and DLL1 by miR34a
- Modulation of miR34a regulates PDGFRa and DLL1 (as well as Notch downstream targets) in human astrocytes and GBM cells

⇒ Relevance?
Chromosomal and genetic aberrations involved in the genesis of glioblastoma.


©2007 by Cold Spring Harbor Laboratory Press
TCGA GBM cohort shows enrichment of NOTCH in Classical and PN subtypes
p53 and Pten loss in neural progenitor cells results in malignant gliomas

- 42/57 (73%) p53L/L Pten L/+ mice:
  - Acute neurological symptoms
  - 28 Grade III; 14 Grade IV
  - Astrocytic morphology (95%)
  - Diffuse & proliferative
  - Necrosis
  - Glioma markers

Zheng (DePinho), Nature 2008
P53/Pten GBM model is PN-like
**Pdgfra overexpression is a hallmark of mouse PN GBMs**

- miR34a modulates Pdgfra expression
- Pdgfra is functionally epistatic to miR34a

(Zheng et al., Nature 2008)
miR34a-low GEMM tumors show Notch activation in vivo

- miR34a modulates Notch activity and its downstream target gene expression
miR34a is a determinant of PN molecular phenotype in GBM

- Integrative genomic data set enables
  - network modeling to generate testable hypothesis
  - development of framework for understanding complex cancer genomics data
- miR34a defines a subset of GBM with concurrent PDGFRα and Notch activation
Promise of Cancer Genomics

- Enable prevention
  - Understanding the underlying etiology → strategy
- Facilitate early detection é
  - Identify risk alleles / genomic events for screening
  - Early events may be detectable in serum or by imaging
- Guide evidence-based interventioné
  - Stratify high vs low risk patients to treat or not
  - Identify new therapeutic targets for drug discovery
  - Inform selection of the right patient for the right drug
  - Define combination / co-extinction strategies
Early-staged patients make up the majority of US cancer diagnoses.

60% of the diagnoses
In Physician’s Health Cohort (n=405)

- 4-marker outperforms Gleason in predicting lethal disease
  - Gleason-only C Index = 0.774;
  - 4-marker only C Index = 0.829

- Carries molecular information not captured by clinical parameters
  - 4-marker + Gleason C Index = 0.882, p = 0.015 for improvement
Metastatic potential of a primary tumor can be determined early on in evolution.

- Deregulated early in development of primary tumors
- Actively “drive” metastasis
- Are oncogenic

Normal → Premalignant → Early Cancer → Advanced Cancer

No progression

Patient 1: Metastasis

Patient 2: No Metastasis
- Deregulated in early-staged primary tumors
- Active drivers of core processes of metastasis
- Bona fide oncogenes
- May serve as
  - Prognostic biomarkers
  - Therapeutic targets
  - Predictive biomarkers

**Early Cancer Detected**

**Pathology analysis and clinical staging**

**Molecular Characteristics**

**Molecular Characteristics**

**High**
- Rx
- Aggressive treatment reserved for hi-risk patients

**Low**
- No Rx
- Cost savings
  - Improved quality of life

10-15% carry inherently poor prognosis

Predictive Info From MOA

Adapted from: Rethinking Screening for Breast and Prostate Cancer. JAMA, 2009
Cancer Genomics → Genomic Medicine

- Enable prevention
  - Understanding the underlying etiology → strategy

- Facilitate early detection
  - Identify risk alleles / genomic events for screening
  - Early events may be detectable in serum or by imaging

- Guide evidence-based intervention
  - Stratify high vs low risk patients to treat or not
  - Identify new therapeutic targets for drug discovery
  - Inform selection of the right patient for the right drug
  - Define combination / co-extinction strategies
Acknowledgement

At the Bench:
Denise Spring
Nate Goldstein
Steven Quayle
Larry Kwong
Papia Ghosh
Gianni Genovese
Chengyin Min
Sharmistha Sarka
Kunal Rai
Nina Seitzer
Erik Uhlmann
Terrence Wu
Ian Watson
Yonathan Lissanu Deribe
Benito Campos
Robert Dewan
Huiyu Liu
Ruprecht Wiedemeyer
Kenneth Scott
Omar Kabbarah
Cristina Nogueira
Tim Heffernan

Metastasis oncogenes
Jason Hanna
David Rimm (Yale)

GOLPH3:
Kwok-kin Wong (DFCI)
Mei-Chih Liang

CLR Network Modeling:
Ayla Ergun
Jim Collins (BU)

RNAi / Human ORFeome
Jesse Bohem (Broad)
Bill Hahn

Molecular Pathology
Gerry Chu
Perry Samuels

Cancer Genomics
Alexei Protopopov
Elena Ivanova
Maria Alimova
Ilana Perna Otey
Georgia Ren

Bioinformatics
Yonghong Xiao
Juhiua Zhang
Spring Liu
Sachet Shukla
Hailei Zhang
Terrence Wu

Molecular Pathology
Gerry Chu
Perry Samuels
Acknowledgement

**Broad GDAC**
- Gaddy Getz
- Mike Noble
- Doug Voet
- Gordon Saksena
- Mike Lawrence
- Lihua Zou
- Rui Jing
- Juihua Zhang
- Spring Liu
- Sachet Shukla
- Hailei Zhang
- Terrence Wu

**Harvard GCC**
- Raju Kucherlapati
- Jon Seidman
- Peter Park
- Alexei Protopopov
- Ilana Perna
- Georgia Ren
- Juinhua Zhang
- Sachet Shukla
- Juihua Zhang
- Hailei Zhang
- N Sathiamoorthy
- Oleg Iartchouk