Overview

- National Challenges
- Using disruptive technologies in care, survivorship and prevention
- Semantics and Data Exchange
- Integration with EHRs
- Predictive modeling
- Building a national learning health system for cancer clinical genomics
National Challenges

• Lowering barriers to data access, analysis and modeling for cancer research

• Integration of data and learning from basic and clinical research with cancer care *that enable prediction and improved outcomes*
We need:

- Open Science (Open Access, Open Data, Open Source) and Data Liquidity for the cancer community
- Semantic interoperability, standards, CDEs and Case Report Forms
- Sustainable models for informatics infrastructure, services, data
Engagement with each other is important

- Informatics: AMIA, BioIT World
- Semantics and ontologies: ICBO, Biocuration
- Computational Biology, Systems Biology and Bioinformatics: ISMB, ECCB, TBC, ISB, AMIA TBI/CRI
- Cancer Informatics: CI4CC
Cancer Informatics for Cancer Centers

CI4CC

Spring Symposium 2014
May 12 - 14

http://ci4cc.org
Cancer Informatics for Cancer Centers

- Fall Symposium
- November 10-12, 2014
- Bay Area
- https://groups.google.com/forum/?fromgroups#!forum/cancer-informatics
- http://ci4cc.org
Where we are

Disruptive technologies
Getting social
Open access to data
Disruptive Technologies

- Printing
- Steam power
- Transportation
- Electricity
- Antibiotics
- Semiconductors design
- http

Systems view - end of reductionism?

- High throughput biology
The era of precision medicine and precision oncology is *predicated* on the integration of research, care, and molecular medicine and the *availability of data* for modeling, risk analysis, and optimal care.

*How do we re-engineer translational research policies that will enable a true learning healthcare system?*
Disruptive Technologies

- Printing
- Steam power
- Transportation
- Electricity
- Antibiotics
- Semiconductors & VLSI design
- http
- High throughput biology
- Ubiquitous computing

World:
- 6.6B active mobile contracts
- 1.9B smart phone contracts
- 1.1B land lines
- World population 7.1B

US:
- 345M active mobile contracts
- 287M smart phone contracts
- US population 313M

Everyone is a data provider
Data immersion

National Cancer Institute
The volume of digital data worldwide is growing rapidly, as the annual IDC Digital Universe study reveals. From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020). From now until 2020, the digital universe will about double every two years.

The majority of information in the digital universe, 68% in 2012, is created and consumed by consumers watching digital TV, interacting with social media, sending camera phone images and videos between devices and around the Internet, and so on.

From the Second Machine Age

FIGURE 3.3 The Many Dimensions of Moore’s Law

- Supercomputer Speed (FLOPS)
- Supercomputer Energy Efficiency (FLOPS/watt)
- Residential Internet Download Speed (kilobytes/second)
- Microprocessor Transistors/Chip
- Hard Drive Cost Efficiency (gigabytes/dollar)

From: The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies by Erik Brynjolfsson & Andrew McAfee
Why should we care about social media?

- Social media may be one avenue for modifying behaviors that result in cancer
- Properly orchestrated, social media can have dramatic impact on quality of life for patients and survivors
- Public education, information, engagement
Public Health

- As a community, we already know how to prevent 50% of the current cancer burden worldwide. Making more effective use of social media, mHealth approaches, virtual communities should enable us to impact vaccination rates (HPV, EBV, mono, hepatitis), and promote healthy lifestyles, including diet, exercise, and smoking cessation.
Public Health

• These three factors - infectious disease, smoking, and poor nutrition and exercise contribute to at least 50% of our current cancer burden. And the cost from loss of quality of life and pain and suffering is incalculable.
Opportunities in prevention

• How do we work together as a community to make our prevention, communication and education researchers more effective and translate this to effect global change. We need to partner with social media and technology-savvy next generation behavioral psychologists!
What about semantics and interoperability?

• Providing powerful, simple resources that enhance data capture, data analysis, and meta analysis is foundational

• A few simple examples focused on current NCI resources
Lowering barriers for the community

- Simplify the creation and distribution of CDE-based forms (caDSR redesigned). Use existing medical terminologies (SNOMED, ICD, LOINC, RxNorm) whenever possible. Link every concept to UMLS as soon as feasible.
Lowering barriers for the community

- Simplify access to EVS, CDEs, NCI Thesaurus (knowledge dissemination too!)
  - Other agency partners: NLM, CDISC, FDA, ONC, PCORI, …

- Creative and appropriate security – we all will need to live in a FISMA moderate world

- Simplify data access – move toward a ‘library card’ model?
PROs and the EHR

- Patient reported outcomes measures are here to stay
- PROMIS and NIHtoolbox are important
- Integration with the EHR is critical!
- Integration with REDCap is here!
  - http://nihpromis.org/
  - http://nihtoolbox.org/
  - http://project-redcap.org
Delivery on an iPad
(work at Northwestern)
Results

>4000 ePROs collected in 1 year in 2 clinics

- 482 patients recruited
  - 434 patients completed at least one measure
- Mean age 48yrs
  - 52.5% female, 87.7% white
## Results: Time burden

<table>
<thead>
<tr>
<th>Patient Reported Outcome Measure</th>
<th># of items</th>
<th>Patients (N)</th>
<th>Median time, min (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease Specific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GerdQ</td>
<td>6</td>
<td>413</td>
<td>1.0 (1.6)</td>
</tr>
<tr>
<td>Heartburn Symptom/Experience</td>
<td>13</td>
<td>432</td>
<td>1.3 (1.7)</td>
</tr>
<tr>
<td>Heartburn Vigilance/Awareness</td>
<td>16</td>
<td>424</td>
<td>1.8 (1.8)</td>
</tr>
<tr>
<td>Impaction Dysphagia Questionnaire</td>
<td>6</td>
<td>426</td>
<td>1.3 (1.8)</td>
</tr>
<tr>
<td>Visceral Sensitivity Index</td>
<td>15</td>
<td>432</td>
<td>1.9 (2.0)</td>
</tr>
<tr>
<td><strong>Not Disease Specific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discomfort Tolerance Scale</td>
<td>7</td>
<td>391</td>
<td>1.4 (1.4)</td>
</tr>
<tr>
<td>Anxiety Sensitivity Index</td>
<td>16</td>
<td>432</td>
<td>1.8 (1.7)</td>
</tr>
<tr>
<td>BSI-18</td>
<td>18</td>
<td>430</td>
<td>1.4 (1.4)</td>
</tr>
<tr>
<td>PANAS</td>
<td>20</td>
<td>432</td>
<td>1.8 (1.5)</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td>4</td>
<td>434</td>
<td>0.8 (0.8)</td>
</tr>
</tbody>
</table>

- Most patients required ≤ 2 minutes for each ePRO measure
- Average time to complete all measures: ~ 20 minutes
Observations

• Tablet computing is here to stay
• Patients appreciate direct entry
• Even in palliative care, tablet uptake was 100% over a multiple month pilot
• Patients found the devices helped them ask better questions (requires building educational materials into the experience)
Measuring outcomes

• Incorporating clinical informatics across healthcare will be essential, especially as care will be judged by true outcomes.
Where do we go from here?
Institute of Medicine Report
Sept 10, 2013
Delivering High-Quality Cancer Care: Charting a New Course for System in Crisis

1. Understanding the outcomes of individual cancer patients as well as groups of similar patients

2. Capturing data from real-world settings that researchers can then analyze to generate new knowledge

3. A “Learning” healthcare IT system that learns routinely and iteratively by analyzing captured data, generating evidence, and implementing new insights into subsequent care.
“Learning IT System”

IOM Report on Cancer Care

1. Search Prior Knowledge: Enable clinicians to use previous patients’ experiences to guide future care.

2. Care Team Collaboration: Facilitate a coordinated cancer care workforce & mechanisms for easily sharing information with each other.

3. Cancer Research: Improve the evidence base for quality cancer care by utilizing all of the data captured during real-world clinical encounters and integrating it with data captured from other sources.
What’s next?

1. Searching
2. Mining
3. Predicting
Can searching prior knowledge help future patients?
Can we make a Cinematch for cancer patients?

Netflix’s Cinematch software analyzes each customer’s film-viewing habits and recommends other movies.
Patients like me

• Patients with diagnoses, symptoms and labs like yours are eligible for these trials:
Other predictive models
Where is the weather moving?
Animating the Weather

Dimension of time assists in decision making.
What about the future?

Present

5 Hours into Future
What changed?

1. Algorithms
2. Discoverable data
3. Scalable computation
4. Pervasive computing
If we can forecast the weather, can we forecast cancer?
SOLVING CANCER
YOU CAN’T CURE WHAT YOU DON’T UNDERSTAND

$$ (X + Y = -C) (X + Y = -C) (X + Y = -C) (X + Y = -C) $$

Cancer Institute
Glioblastoma Treatment Outcomes

- Prediction of the outcome of individual patients would be of great significance for monitoring responses to therapy.
- A mathematical model has been developed based on proliferation and invasion.
- Serial medical imaging can be used to track the spatio-temporal behavior of the detectable portion of each lesion.

Modeling Tumor Growth

Mathematical model: proliferation of cells with the potential for invasion and metastasis

\[
\frac{\partial c}{\partial t} = \nabla \cdot (D(x)\nabla c) + \rho c - R(x, t)c
\]

FIG. 1. Illustration of high proliferation rate ($\rho$) and cell diffusion rate ($D$).

rate of change of glioma cell concentration

\[
\frac{\partial c}{\partial t} = \nabla \cdot (D(x) \nabla c) + \rho c - R(x, t)c
\]
Personalized Tumor Model

Imaging used to seed the model

Example

National Cancer Institute
Personalized Tumor Model

Today

Future

National Cancer Institute
Radiation Treatment Effects

rate of change of glioma cell concentration
\[ \frac{\partial c}{\partial t} = \nabla \cdot (D(x) \nabla c) + \rho c - R(x, t)c, \]

New term defines cell killing

L-Q model used to describe cell killing

\[ R(\alpha, d(x, t)) = \begin{cases} 0, & \text{for } t \notin \text{therapy} \\ 1 - S(\alpha, d(x, t)), & \text{for } t \in \text{therapy.} \end{cases} \]

\[ S(\alpha, d(x, t)) = e^{-\alpha \text{BED}}. \]

\[ \text{BED} = nd \left(1 + \frac{d}{\alpha/\beta}\right), \]
Simulated tumor growth & response to XRT

Does it help make better decisions?

High Diffusion, Low Proliferation

Low Diffusion, High Proliferation
How do we generalize?

- We need to use Rapid Learning Systems to build prediction models
Rapid Learning Systems
Patient-level data are aggregated to achieve population-based change, and results are applied to care of individual patients.

Population Decision Support

Clinical trials, comparative effectiveness research, molecular and biologic data

Information-rich, patient-focused data

Evaluation of outcomes

Transformation of subsequent care delivery

Data aggregation, evidence generation

Predict outcomes
The era of precision medicine and precision oncology is *predicated* on the integration of research, care, and molecular medicine and the *availability of data* for modeling, risk analysis, and optimal care.

*How do we re-engineer translational research policies that will enable a true learning healthcare system?*
Some NCI activities

• TCGA, TARGET and ICGC
  – Cancer Genomics Data Commons
  – NCI Cloud Pilots

• Molecular Clinical Trials:
  – MPACT, MATCH, Exceptional Responders
Cancer Genomics Data Commons

- A data service for the cancer research community
- Increase consistency, QA, calling and annotations
- Cancer genomics (TCGA and TARGET) data housed into a uniform and co-localized database
- Create a foundation for future expanded data access, computational capabilities, and bioinformatics cloud research
NCI Cloud Pilots

• Funding for up to 3 cloud pilots - 24 month pilots that are meant to inform the Cancer Genomics Data Commons
NCI Cloud Pilots

- A way to move computation to the data
- Sustainable models for providing access to data
- Reproducible pipelines for QA, variant calling, knowledge sharing
- Define genomics/phenomics APIs for discovering new variants contributing to cancer, enhancing response, modulating risk
Standard Model of Computational Analysis

Public Data Repositories

Local Data

Network Download

Publicly Available Software

Locally Developed Software

Local storage and compute resources
Growth of TCGA Sequence Data

Gigabytes (GB)

<table>
<thead>
<tr>
<th>Date</th>
<th>Gigabytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/1/09</td>
<td>0</td>
</tr>
<tr>
<td>1/1/10</td>
<td>0</td>
</tr>
<tr>
<td>7/1/10</td>
<td>0</td>
</tr>
<tr>
<td>1/1/11</td>
<td>0</td>
</tr>
<tr>
<td>7/1/11</td>
<td>0</td>
</tr>
<tr>
<td>1/1/12</td>
<td>0</td>
</tr>
<tr>
<td>7/1/12</td>
<td>0</td>
</tr>
<tr>
<td>1/1/13</td>
<td>0</td>
</tr>
<tr>
<td>7/1/13</td>
<td>0</td>
</tr>
<tr>
<td>1/1/14</td>
<td>0</td>
</tr>
<tr>
<td>7/1/14</td>
<td>2,500,000</td>
</tr>
</tbody>
</table>

7/1/14
Multiple orthogonal data types
The future

- Elastic computing ‘clouds’
- Social networks
- Big Data analytics
- Precision medicine
- Measuring health
- Practicing protective medicine

Learning systems that enable learning from every cancer patient

Semantic and synoptic data

Intervening before health is compromised
Thank you

Warren A. Kibbe
Warren.kibbe@nih.gov