Information Exchange and Data Transformation (INFORMED) Initiative and the Role of Real World Evidence in Regulatory Science

Sean Khozin, MD, MPH
Senior Medical Officer
Office of Hematology and Oncology Products
Food and Drug Administration

Disclosures: None
The views in this presentation are my own and do not necessarily reflect the policies of FDA
About 3 quintillion bytes of data per day
Big data pipeline in biomedicine

Growth of DNA Sequencing

- Recorded growth
- Double every 7 months (Historical growth rate)
- Double every 12 months (Illumina Estimate)
- Double every 18 months (Moore's Law)

Cumulative Number of Human Genomes

- 1e+00
- 1e+03
- 1e+06
- 1e+09

Year


Worldwide Annual Sequencing Capacity

- 1 Zbp
- 1 Ebp
- 1 Pbp
- 1 Tbp

1st Sanger
IHGSC et al. Venter et al.

1st Personal Genome
Levy et al.

1st 454
Wheeler et al.

1st Illumina
Bentley et al. Wang et al. Ley et al.

1000 Genomes

ExAC

Current Capacity

TCGA

1st PacBio
Chaisson et al.

Large amount of diversity

Multiomics

- Epigenetics
- Genomics

DNA

RNA
- Transcriptomics

Protein
- Proteomics

Metabolites
- Metabolomics

Microbiome
The expanding universe of big data in the biomedical enterprise

- **Velocity**:
  - Real time
  - Near-real time
  - Periodic

- **Volume**:
  - GB
  - TB
  - PB

- **Variety**:
  - EHR
  - Databases
  - Clinical trials
  - Biometrics
  - Omics

- **Current capacity**:
  - (generalization)
Must focus on breaking siloes

- Unstructured content
- Omics
- Biometric/sensor technologies
- EHR/PHR
- Clinical trials
Why?
We are nearing the limits of siloed approaches

“Driver” mutations in non-small cell lung cancer (NSCLC):

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Adenocarcinoma</th>
<th>Squamous-cell carcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRAS</td>
<td>5-15%*</td>
<td>&lt;5%†</td>
</tr>
<tr>
<td>EGFR</td>
<td>11.3%</td>
<td>8.3%</td>
</tr>
<tr>
<td>MET exon 14 skipping</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MET</td>
<td>&lt;5%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>&lt;5%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>RAF</td>
<td>&lt;5%</td>
<td>0</td>
</tr>
<tr>
<td>AKT1</td>
<td>1.7%</td>
<td>1.7%</td>
</tr>
<tr>
<td>MAP2K1</td>
<td>1.7%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Amplifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>2.2%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>


**NSCLC: 2016**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Frequency in NSCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>Mutation</td>
<td>1%</td>
</tr>
<tr>
<td>ALK</td>
<td>Rearrangement</td>
<td>3–7%</td>
</tr>
<tr>
<td>BRAF</td>
<td>Mutation</td>
<td>1–3%</td>
</tr>
<tr>
<td>DDR2</td>
<td>Mutation</td>
<td>~4%</td>
</tr>
<tr>
<td>EGFR</td>
<td>Mutation</td>
<td>10–35%</td>
</tr>
<tr>
<td>FGFR1</td>
<td>Amplification</td>
<td>20%</td>
</tr>
<tr>
<td>HER2</td>
<td>Mutation</td>
<td>2–4%</td>
</tr>
<tr>
<td>KRAS</td>
<td>Mutation</td>
<td>15–25%</td>
</tr>
<tr>
<td>MEK1</td>
<td>Mutation</td>
<td>1%</td>
</tr>
<tr>
<td>MET</td>
<td>Amplification</td>
<td>2–4%</td>
</tr>
<tr>
<td>NRAS</td>
<td>Mutation</td>
<td>1%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Mutation</td>
<td>1–3%</td>
</tr>
<tr>
<td>PTEN</td>
<td>Mutation</td>
<td>4–8%</td>
</tr>
<tr>
<td>RET</td>
<td>Rearrangement</td>
<td>1%</td>
</tr>
<tr>
<td>ROS1</td>
<td>Rearrangement</td>
<td>1%</td>
</tr>
</tbody>
</table>
Siloed data $\rightarrow$ Big Data $\rightarrow$ Smart Data

**Reductionist**
- One-gene one-drug
- Trials with strict eligibility criteria
- Leap of faith clinical development

**Holistic**
- Multiomics
- Pragmatic trials
- Systems biology, predictive analytics
Information Exchange and Data Transformation (INFORMED)

A holistic approach to oncology regulatory science and big data analytics

- Sponsor
- Regulatory
  - Formal submission
- Research
  - Direct
- CDISC
- Clinical trials
- Real world data working group
  - Real world, biometrics, omics, social media
- Data exchange/visualization/analytics
  - New data pipelines
  - Data exported for further analysis if needed

*Technology and software development
Uncertainty management

• Using novel pipelines of high quality data in regulatory decision making can reduce uncertainty
  – RWE
  – Patient reported
  – Biometrics (wearables, implantable, etc)
Substantial evidence from adequate and well controlled investigations

Safety

Efficacy

Risk

Benefit

Marketing approval

Uncertainty

Reduced data quality

data quantity \( f(\text{time}) \)
Challenges
More than just technology

Technical
- Computation
- Accuracy

Sociopolitical
- Data linkage and sharing
- Data standards
- Education
- Blurred boundaries

Adapted with modification from Kohane IS. Science. Vol. 349 no. 6243 pp. 37-38
Avoid data standards proliferation
Big picture: today

Preclinical
- Dose, delivery, and manufacturing process optimization

Clinical trials
- PK/PD
- Patients
- Omics
- Biosensors
- Outcomes
- Validation

Hypothesis

Real world
- Patients
- Omics
- Outcomes
- Biosensors
- Unstructured content

Predictive analytics
Bigger picture: tomorrow

The learning health system (IOM)

A system where science, informatics, incentives, and culture are aligned for continuous improvement and innovation

Discovery as a product of the healthcare delivery experience