Is big data ready to improve patient outcomes or is it a new generation of garbage in/garbage out?

**PRO**

Rodrigo Dienstmann, MD
No conflicts of interest.
BIG DATA – LOST IN TRANSLATION?

Questionable value in our attempt to implement Precision Medicine.
WHAT IS BIG DATA?
THE 5 Vs OF BIG DATA

1. Volume: The size of the data
2. Velocity: The speed at which the data is generated
3. Variety: The different types of data
4. Veracity: The trustworthiness of the data in terms of accuracy
5. Value: Just having Big Data is of no use unless we can turn it into value
Biomedical research has and will continue to generate large amounts of data in many formats and at all levels.
The generation of ever-larger data sets is not restricted to large genome centers or pharmaceutical companies.

Technological advances now allow any academic life scientist or molecular diagnostic lab to compile terabytes of data.
Drug screening

Microbiome

Genome

Epigenome

Transcriptome

Metabolome

Lipidome

Proteome

Database integration engine

Hematological cancers (AML, CML, MM, T-PLL), solid tumors (OvCa, mCRPC)
Pemovska et al, Cancer Discov 2016
# Personalized OncoGenomics at BCC

## Analysis and Reporting

### Pathology
- Genetic abnormalities that affect the treatment plan.
- Mutations with an associated available clinical trial, such as in PIK3K.

### Genomics & Transcriptomics
- Drug Target Analysis

## Potential Clinical Impact
- Clinical Trial
- Approved Therapy
- Off-Label Therapy
- Potential Trial but Not Available Locally or at This Time
- Diagnostic Clarification
- Patient Factors Unable to Treat
- Biological Interest
- Data of Uncertain Significance

## Terms and Definitions

<table>
<thead>
<tr>
<th>TERM</th>
<th>DEFINITION</th>
<th>EXAMPLE</th>
</tr>
</thead>
</table>
| Actionable           | Identification of a potential target or risk factor that affects the treatment plan | - Established target-drug relationships such as EGFR activating mutations in lung cancer or BRAF V600E mutations in melanoma.  
                        |                                                                             | - Established germline risks such as detection of a BRCA mutation.  
                        |                                                                             | - Identification of a primary or primary unknown cancers or amendment of a diagnosis.  
                        |                                                                             | - Mutations with an associated available clinical trial, such as in PIK3K. |
| Informative          | Identification of an interesting feature that may not have prognostic or therapeutic relevance at this time | - Pathway abnormalities that are potentially targetable but no clinical trials are open or available.  
                        |                                                                             | - Detection of a somatic p53 mutation. |
| Data of Uncertain Significance | Identification of biological feature or abnormality of unknown significance; may have potential clinical or biological relevance but is not established | - Germline variants of uncertain or unknown significance.  
                        |                                                                             | - Mutation or pathways that seem important but are unclear.  
                        |                                                                             | - For example: MYC / BAX / MET / MUC1 / PTEN   
                        |                                                                             | - Abnormality not defined as cancer-related in literature or databases (such as in COSMIC). |

Laskin et al, Mol Case Study 2015
Personalized OncoGenomics at BCC

- Biopsy
  - 100 cases
  - 78 Sequenced/ Analyzed
  - 65 Informative
  - 55 Actionable
    - 25 Approved therapy
    - 8 Off-label therapy
    - 1 Clinical trial
    - 34 Treated with POG ‘informed’ systemic therapy
    - 14 Clinical / radiographic improvement
    - 13 Too ill to receive further treatment
Lessons from the Cancer Genome

Levi A. Garraway$^{1,2,4}$ and Eric S. Lander$^{3,4,5,*}$

New cancer genes (in new cancer types)
Mutational mechanisms
Long tail
Tumor heterogeneity
Heredity
Saturation analysis of cancer genes

Lawrence et al, Nature 2014
TCGA therapeutic landscape

Supporting evidence in the tested tumor

- **clinical guidelines/recommendations**
- **late trials**
- **early trials**
- **case reports**
- **preclinical data**

- **only mutation data available**

Resilience project

Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases

Chen et al, Nature Biotechnology 2016
874 genes
13 asymptomatic individuals
8 severe Mendelian conditions
<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Gene</th>
<th>Mutation (cDNA; protein (reference))</th>
<th>Genomic coordinate (hg19)</th>
<th>Mutation severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic fibrosis</td>
<td>CFTR</td>
<td>c.1558G&gt;T; p.V520F (NM_000492.3)</td>
<td>Chr7 117199683</td>
<td>Severe pulmonary disease, childhood-onset</td>
</tr>
<tr>
<td>Smith-Lemli-Opitz syndrome</td>
<td>DHCR7</td>
<td>c.964-1G&gt;C (NM_001360.2)</td>
<td>Chr11: 71146886</td>
<td>Severe developmental disorder, probably embryonic lethal</td>
</tr>
<tr>
<td>Familial dysautonomia</td>
<td>IKBKAP</td>
<td>c.2204+6T&gt;C (NM_003640.3)</td>
<td>Chr9: 111662096</td>
<td>Severe neurological disease, high mortality in early childhood</td>
</tr>
<tr>
<td>Epidermolysis Bullosa simplex</td>
<td>KRT14</td>
<td>c.373C&gt;T; p.R125C (NM_000526.4)</td>
<td>Chr17: 39742714</td>
<td>Severe dermatologic condition, infantile onset</td>
</tr>
<tr>
<td>Pfeiffer syndrome</td>
<td>FGFR1</td>
<td>c.755C&gt;G; p.P252R (NM_023110.2)</td>
<td>Chr8: 38282208</td>
<td>Severe congenital skeletal dysplasia with variable expressivity</td>
</tr>
<tr>
<td>APECED</td>
<td>AIRE</td>
<td>c.769C&gt;T; p.R257* (NM_000383.2)</td>
<td>Chr21: 45709656</td>
<td>Severe childhood-onset autoimmune disease</td>
</tr>
<tr>
<td>Acampomelic campomelic dysplasia</td>
<td>SOX9</td>
<td>c.1320C&gt;G; p.Y440* (NM_000346.3)</td>
<td>Chr17: 70120318</td>
<td>Severe skeletal dysplasia with early childhood death</td>
</tr>
<tr>
<td>Atelosteogenesis</td>
<td>SLC26A2</td>
<td>c.835C&gt;T; p.R279W (NM_000112.3)</td>
<td>Chr5: 149359991</td>
<td>Severe early-onset skeletal dysplasia with variable expressivity</td>
</tr>
</tbody>
</table>

Buffering effects of rare highly penetrant deleterious mutations.

Protective genetic variants?
Therapeutic targets?

Chen et al, Nature Biotechnology 2016
Time to move away from case reports!
Learning from ALL treated patients (not only clinical trials and case reports).
Targeted Agent and Profiling Utilization Registry Study

American Society of Clinical Oncology
IBM Watson Oncology Advisor

- Treatment plan 1: Systemic Chemotherapy, Prognosis: Best, Match: 95%
  - Patient Preferences Match: Acceptable, matches with patient preferences

- Treatment plan 2: Systemic Chemotherapy, Prognosis: Poor, Match: 45%
  - Patient Preferences Match: Unacceptable, does not match with patient preferences

- Treatment plan 3: Systemic Chemotherapy, Prognosis: Good, Match: 8%
  - Patient Preferences Match: Preferred, matches with patient preferences

Radiation and Surgery are unlikely to be appropriate.
Medical Records
mPower study

Parkison’s disease mobile app

Bot et al. Sci Data, 2016
mPower study cohort description

Bot et al. Sci Data, 2016
Assessing drug response with app data

Individual 1

Individual 2

Individual 3

Individual 4

Chaibub Neto et al. Pac Symp Biocomputing, 2016
Big Data in Life Sciences—Example Use Cases

- Drug Repurposing
- Health record-guided drug development
- Next Generation Sequencing
- Personalized Healthcare
- Treatment Adherence (Compliance)
- Adverse Event Detection
- Patient Pre-Profiling
- Influencer Profiling
Deep learning in digital pathology

Günhan Ertuson et al, AMIA Annual Symp Proc 2015
BIG DATA – LOST IN TRANSLATION?

NO.

Direct evidence – individual cases
Indirect evidence – everywhere

BIG HOPES
Tumor evolution = Knowledge evolution
In the Race to the Top you need...

Access to a Variety of Data

Data Subjects who see their Interest in Sharing
Acknowledgements

Stephen Friend
Brian Bot
Justin Guinney
Elias Chaibub Neto
Access to person-level data – ETHICAL considerations

Sensitive data are being re-used, linked and analysed on an unprecedented scale.

Balance private rights with public benefits (and market interests).

In many cases, researchers cannot use completely anonimous data and yet cannot feasibly seek consent.

Exemption from consent for the use of sensitive personal data in medical research in case of legitimate interests.