Eric D Perakslis PhD
Chief Science Officer, Datavant

Lecturer, Department of Biomedical Informatics, Harvard Medical School

Strategic Advisor on Innovation
Manson Unit
Médecins Sans Frontières
Health Science Today…and Tomorrow

Medicine today is built on hundreds of years of individual observations…

What if we could integrate and model them instead?
Disease Knowledge Networks: multi-level mechanistic hypotheses

- Syndrome
- Symptom/Traits
- Cellular Systems/Signaling Pathways
- Proteome/Metabolome
- Genome
- Human Interactome
- Microbiome
- Envirome
- Internal Mind State
- Interventions

Modified from http://www.phenomics.ucla.edu/phenomics/cognitive_phenomics.htm
Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

Report from National academy of science, USA, 2011
Challenges in Hypothesis Testing:

From “Big Data” to narrow questions

Total Population → Sampled Population → On Treatment → Reporting Treatment → Reporting Treatment X → Switched Treatment X

Harder questions to answer

The highest priority group we aim to understand
Deep human data phenotyping and molecular profiling strategies are becoming ubiquitous.

Wearable Devices in Clinical Trials: Hype and Hypothesis.
Izmailova ES¹, Wagner JA¹, Perakslis ED¹

Effective knowledge management in translational medicine.
Szalma S¹, Koka V, Khasanova T, Perakslis ED.
IT and Informatics enabling drug review & approval
How Integrative Informatics was Enabling Translational Science in 2007

- Disease definition
- Patient stratification
- Drug target identification
- Drug indication selection
- Epidemiology
Disease Identification

Cancer classically identified solely by effected organ system

We are learning that the basic biology of cancer often transcends organ system definitions

We are finding ‘subtypes within subtypes within subtypes’ it is essential to classify diseases, treatments and responses at the molecular level

We are no longer looking for blockbusters…
Patient Stratification

Need to understand best treatment for any given molecular subtype of cancer

Which patient will respond best to which drug and why? Predictors of response.

What is the risk of metastasis for the different subtypes? Any predictors of recurrence?

When is it OK not to treat, if ever? Signature of indolence?

What is the best endpoint to measure?
Drug Target Identification

It can be very challenging to associate molecular events with disease states

If I inhibit the production of a given protein, will a tumor stop Growing? Will it shrink and die?

Can I change the regulatory cascade and produce a desired Outcome.

Remember: DNA => RNA => Proteins => pretty much controls everything…?? (probably not!)
The provision of access to clinical trial results that include patient-level data is generating much debate. A growing chorus of transparency advocates is pushing for open access to these data...
Improving Patient Safety through Transparency

Allen Kachalia, M.D., J.D.

Transparency — especially when things go wrong — is increasingly considered necessary to improving the quality of health care. By being candid with both patients and clinicians, health care organizations can promote their leaders’ accountability for safer systems, better engage clinicians in improvement efforts, and engender greater patient trust.

Last year, less than two thirds of staff members reported having a favorable perception of their hospital’s openness in communication, and less than half reported that their hospital reports respond to errors in a non-punitive way.
Objectives:

1. **Integration** of clinical, biological and ‘omics data in one place – hypothesis free –

2. Generation of **hypothesis** by Clinicians / Researchers
Analysis of WES\Demographics\Population for subsets:

**Subset 1**

- CEU: 1
- CHB: 3
- YRI: 5

**Subset 2**

- CEU: 2
- CHB: 4
- YRI: 11

<table>
<thead>
<tr>
<th>Category</th>
<th>Subset 1 (n)</th>
<th>Subset 1 (%)n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU</td>
<td>3</td>
<td>33.3%</td>
</tr>
<tr>
<td>CHB</td>
<td>1</td>
<td>11.1%</td>
</tr>
<tr>
<td>YRI</td>
<td>5</td>
<td>55.6%</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Subset 2 (n)</th>
<th>Subset 2 (%)n</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU</td>
<td>11</td>
<td>64.7%</td>
</tr>
<tr>
<td>CHB</td>
<td>2</td>
<td>11.8%</td>
</tr>
<tr>
<td>YRI</td>
<td>4</td>
<td>23.5%</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100%</td>
</tr>
</tbody>
</table>

Chi-Squared: 2.8215
p-value: 0.24397

The results are not significant at a 95% confidence level.
Implemented Disruptive Technology

• Open source
• Cloud computing
• Pre-competitive sharing

What was/is/are the real goal(s)?
Collaborative Sharing of IMI Consortium Data – U-BIOPRED Respiratory Repository

- ECLIPSE
- TranSMART
- KM Repository
- IC: CLOUD
- AstraZeneca
- 30 Patient Study
- Novartis
- 20 Patient Study
- Pfizer
- 20 Patient Study
• Integrated platform to support translational research
• Initiated by Johnson & Johnson et Recombinant 6 years ago
  • PI: Eric Perakslis

• Open-source since January 24th, 2012
• Installed at HEGP Hospital, Paris since May, 2012
• Today, driven and maintained by the tranSMART

http://transmartfoundation.org
tranSMART Adopters

- **eTRIKS (IMI), $€24MM, 5 years, 16 partners**
  - [http://www.imi.europa.eu/content/etriks](http://www.imi.europa.eu/content/etriks)
  - 10 EFPIA, 4 Academic / Non-profits, 2 Other Partners

- **EMIF (IMI), $€24MM, 5 years, 55 partners**
  - [http://www.imi.europa.eu/content/emif](http://www.imi.europa.eu/content/emif)
  - 9 EFPIA, 36 Academic / Non-profit, 3 Patient Organizations, 7 Other Partners

- **TraIT (CTMM), $€16MM, 4 years, 26 partners**
  - [http://www.ctmm.nl/pro1/general/start.asp?i=6&j=1&k=0&p=0&itemid=330](http://www.ctmm.nl/pro1/general/start.asp?i=6&j=1&k=0&p=0&itemid=330)
  - The TraIT project is a joint initiative between CTMM, the Dutch Cancer Society, the Dutch Heart Foundation, the Netherlands Federation of University Medical Centers (NFU), the Netherlands Bioinformatics Centre (NBIC), the String of Pearls Initiative (PSI) and the Netherlands eScience Center (NLeSC).

- **Pompidou University Hospital in Paris (APHP - HEGP)**
tranSMART Adopters

- TBIG (Janssen, Millennium, Sanofi & Pfizer)
  Coordination of tranSMART enhancements on pre-competitive basis

- University of Michigan
  Metabolomics, Glomerular Disease, NCIBI Tool Integration,
  CORECT Study (http://epi.grants.cancer.gov/gameon/)
  PostgreSQL Conversion (Janssen, U Michigan, Recombinant)

- One Mind for Research
  Traumatic Brain Injury
  Neuroscience Portal (Citizens, Researchers, Clinicians & Care Givers)

- Food and Drug Administration
  Several Use Case Evaluations; Drug Safety

- University of Minnesota Supercomputing Center and CTSI

- St Jude Children’s Hospital and Research Foundation
### tranSMART main installations

- **International Research Initiatives**
  - IMI – eTRIKS, EMIF
  - CTMM – TraIT
- **Pharma & Biotech**
  - Sanofi, Millennium, Pfizer, JNJ, Roche
- **Government aligned Institutions**
  - FDA
- **Non-Profits**
  - 1Mind4Research, Orion Bionetworks
- **Hospitals / Academics**
  - U Michigan, John’s Hopkins, St. Jude, HEGP, Harvard/Boston Children Hospital
- **Service Providers**
  - Thomson Reuters, Recombinant(Deloitte), theHyve, Rancho Biosciences, BTGS

<table>
<thead>
<tr>
<th>Start</th>
<th>Organization</th>
<th>Type</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Johnson &amp; Johnson</td>
<td>Pharma</td>
<td>Production</td>
</tr>
<tr>
<td>2008</td>
<td>Recombinant by Deloitte</td>
<td>Services</td>
<td>Multiple</td>
</tr>
<tr>
<td>2010</td>
<td>Sage Bionetworks</td>
<td>Non profit</td>
<td>Production</td>
</tr>
<tr>
<td>2010</td>
<td>Thomson Reuters</td>
<td>Services</td>
<td>Support</td>
</tr>
<tr>
<td>2010</td>
<td>U-BIOPRED</td>
<td>Consortium</td>
<td>Production</td>
</tr>
<tr>
<td>2011</td>
<td>SAFE-T</td>
<td>Consortium</td>
<td>Pilot</td>
</tr>
<tr>
<td>2011</td>
<td>University of Michigan, Comprehensive Cancer Center (UMCCC)</td>
<td>Academic</td>
<td>Production</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Organization</th>
<th>Type</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>APHP-HEGP Paris France</td>
<td>Academic</td>
<td>Production</td>
</tr>
<tr>
<td>2012</td>
<td>BT Cure</td>
<td>Consortium</td>
<td>Pilot</td>
</tr>
<tr>
<td>2012</td>
<td>CTMM/TraIT</td>
<td>Consortium</td>
<td>Development</td>
</tr>
<tr>
<td>2012</td>
<td>FDA</td>
<td>Government</td>
<td>Development</td>
</tr>
<tr>
<td>2012</td>
<td>IMI/eTRIKS</td>
<td>Consortium</td>
<td>Development</td>
</tr>
<tr>
<td>2012</td>
<td>Merck</td>
<td>Pharma</td>
<td>Pilot</td>
</tr>
<tr>
<td>2012</td>
<td>Millennium Pharmaceuticals</td>
<td>Pharma</td>
<td>Production</td>
</tr>
<tr>
<td>2012</td>
<td>One Mind for Research (1M4R)</td>
<td>Non profit</td>
<td>Production</td>
</tr>
<tr>
<td>2012</td>
<td>Pfizer</td>
<td>Pharma</td>
<td>Production</td>
</tr>
<tr>
<td>2012</td>
<td>Roche</td>
<td>Pharma</td>
<td>Evaluation</td>
</tr>
<tr>
<td>2012</td>
<td>Sanofi-Aventis</td>
<td>Pharma</td>
<td>Development</td>
</tr>
<tr>
<td>2012</td>
<td>St. Jude</td>
<td>Non profit medical center</td>
<td>Development</td>
</tr>
<tr>
<td>2012</td>
<td>University of Michigan, Department of Computational Medicine &amp; Bioinformatics (DCM&amp;B)</td>
<td>Academic</td>
<td>Development, Pilot</td>
</tr>
<tr>
<td>2013</td>
<td>Agios</td>
<td>Biotechnology</td>
<td>Evaluation</td>
</tr>
<tr>
<td>2013</td>
<td>CARPEM – Cancer personalized medicine</td>
<td>Academic French grant</td>
<td>Development</td>
</tr>
<tr>
<td>2013</td>
<td>Harvard Medical School / Boston Children Hospital</td>
<td>Academic</td>
<td>Autism Pilot</td>
</tr>
<tr>
<td>2013</td>
<td>Boehringer Ingelheim</td>
<td>Pharma</td>
<td>Pilot</td>
</tr>
<tr>
<td>2013</td>
<td>Bristol Myers Squibb</td>
<td>Pharma</td>
<td>Evaluation</td>
</tr>
<tr>
<td>2013</td>
<td>BT Global Services</td>
<td>Services</td>
<td>Pilot</td>
</tr>
<tr>
<td>2013</td>
<td>Accelerated Cure Project for MS</td>
<td>Non profit</td>
<td>Development</td>
</tr>
<tr>
<td>2014</td>
<td>Personalized medicine and colorectal cancers</td>
<td>Academic French grant</td>
<td>Development</td>
</tr>
<tr>
<td>2014</td>
<td>PCORI PRRN Phelan-McDermid Syndrome Data Network</td>
<td>Academic US grant</td>
<td>Development</td>
</tr>
</tbody>
</table>
tranSMART and i2b2

Health care
Health Information System

Clinical Research

1) Analysis tools

2) Structured data from research studies

3) ‘omics data

Browser tools available for Researchers

i2b2

Phenotypic

ETL once a week

i2b2

‘omics

ETL

ETL

ETL

DRG
EHR forms
EHR reports
Biology
Imaging
Pathology
Rx
Organizations are sharing their data and collaborating in open technology, open science and open data

tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform
Elisabeth Scheufele, MD, MS,¹,² Dina Aronzon, MS,¹ Robert Coopersmith, Ph.D,¹ Michael T. McDuffie, MS,¹ Manish Kapoor, MS,¹ Christopher A. Uhrich,¹ Jean E. Avitabile,¹ Jinlei Liu, MS,¹ Dan Housman,¹ and Matvey B. Palchuk, MD, MS¹,²

http://transmartfoundation.org/bibliography/

2017: 18 papers
2016: 18 papers
2015: 27 papers
Translational research platforms integrating clinical and omics data: a review of publicly available solutions

Vincent Canuel*, Bastien Rance*, Paul Avillach, Patrice Degoulet and Anita Burgun

BRISK, iCOD
caTRIP, iDASH
cBio Cancer, tranSMART (i2b2)
Portal
G-DOC
In Stunning Win For Open Science, Johnson & Johnson Decides To Release Its Clinical Trial Data To Researchers

Matthew Herper, Forbes Staff
I cover science and medicine, and believe this is biology's century.
PHARMA & HEALTHCARE | 1/30/2014 @ 7:09AM | 17,254 views
Session presentations and discussion will inform the development of an interim report that will be released for public comment in January 2014.

The interim report will concisely characterize the landscape of clinical trial data and data sharing activities with respect to the Committee’s tentative findings on:

1) Key elements of and approaches to data sharing activities (to clarify definitions and develop a shared nomenclature)

2) Guiding principles that underpin responsible sharing of clinical trial data.
4 Cyber-security Threats: Data Breach, Theft, Device &/or Infrastructure Attack

Study: Data breaches cost healthcare providers $1.6 billion
Author Name Patrick Ouellette | Date February 3, 2014

Wash. Hospital Hit By $1.03 Million Cyberheist
30 APR 13

A New Cyber Concern: Hack Attacks on Medical Devices
The FDA issues guidelines to manufacturers to protect their products

Experts: Cyber Attacks Threaten Hospitals
8.11.2010: By Andy Fell
Experts: Cyber Attacks Threaten Hospitals

8.11.2010: By Andy Fell

- **Hospitals: 'Soft targets'** "Hospitals are a soft target where a cyber attack can cause a lot of damage quite easily,"

- **'Down for a year'** FBI Special Agent Eric Brelsford described a similar attack on a Chicago-area hospital in 2006 that affected everything from cancer treatments to prescriptions.

- **Identity theft** Medical information is also a rich target for thieves.

- **Waves of time attacks** ...By day 16, the nation's hospitals would be in chaos...
  "If a doctor can get to it, I guarantee a bad guy can,"

- **Undermines confidence** A security breach can cause a company to fold..."Trust is key in healthcare, and anything that could break that trust is a big deal,"
Develop deep and pioneering healthcare cyber security and privacy experience


https://www.jmir.org/announcement/view/153
Harvard Medical School to play key coordination role in NIH Undiagnosed Diseases Network

By Raymond MacDougall  Associate Director of Communications, Division of Intramural Research

What’s Wrong With Summer Stiers?
http://www.nytimes.com/2009/02/22/magazine/22Diseases-t.html?_r=3&emc=eta1&

60 Minutes HEALTH & SCIENCE
Hard cases: Investigating rare & tough diseases

NIH Undiagnosed Diseases Program documents two-year pilot as clinic of last resort
Genomic tools prove integral to solving medical mysteries
Cloud computing reduces HIPAA compliance risk in managing genomic data
September 4, 2013

“…a review of data on HIPAA breaches published by the US Department of Health and Human Services (HHS) shows that these concerns are Misplaced.

In fact, by using a cloud-based service with an appropriate security and compliance infrastructure, an organization can significantly reduce its compliance risk…”
Patients are sharing their data and collaborating enthusiastically in open science and open data.

Date of Report: Feb 01, 2018

Description:

Shortly after birth, the patient was noticed to have decreased tone and difficulty breathing. She struggled to drink from a bottle and would breathe liquids into her airway (aspirate). At 6 months, a G tube was placed. Currently she struggles with chewing and swallowing and is unable to drink out of a cup or straw.

The patient said her first words at 13 months and began walking at 22 months. She gets tired easily (easy fatigability) and makes rhythmic and repetitive movements with her upper and lower limbs (stereotypic movements).
The only 100% common element of digital transformation across all industries is **data**

---

**Harvard Business Review**

*What the Companies on the Right Side of the Digital Business Divide Have in Common*

Robert Bock, Marco Iansiti, Karim R. Lakhani… “differentiate themselves from competitors based on their data platform.”

---

**Forbes**

*The Power Of Digital Transformation In A Data-Driven World*

Peter Bendor-Samuel. “digital transformation journey moves an organization from a process-defined world to a data-driven world”

---

**Information Age**

*Data science needs to be a fundamental component of any digital transformation effort.* Ben Rossi

---

**HIT Consultant**

*Why Data Block Is the Leading Cause of Death for Digital Health Startups.* by Our Thought Leaders

---

**MobiHealthNews**

*Five digital health trends investors are watching in 2017*

By Skip FleshmanFebruary 16, 2017. “#3 Data Integration and Analytics”
Data transformation: what is data solving for?

How is medicine changing?

More Than Half of Kaiser Permanente's Patient Visits Are Done Virtually
Last year, patients and doctors interacted more than 110 million times in total.
By Kia Kokalitcheva Oct 6, 2017

How is biopharmaceutical R&D changing?

FDA issues new guidance to facilitate expanded use of real-world evidence in medical device Development
Deborah Kotz

Digital Health Innovation Action Plan
Digital Health Software Precertification (PreCert) Program
Data transformation: How are patients changing?

They are active, connected, informed and savvy.
Ill Literates or Illiterates? Investigating the eHealth Literacy of Users of Online Health Communities.
Petrič G¹, Atanasova S¹, Kamin T²

The younger & next generation is very different.
Exploring the digital technology preferences of teenagers and young adults (TYA) with cancer and survivors: a cross-sectional service evaluation questionnaire.
Abrol E¹, Groszmann M², Pitman A³, Hough R⁵, Taylor RM⁶, Aref-Adib G⁴,⁷

The reliability of patient reported outcomes is still highly variable.
The Impact of Participation in Online Cancer Communities on Patient Reported Outcomes: Systematic Review.
van Eenbergen MC#¹, van de Poll-Franse LV¹,², Heine P³, Mols F⁴

Direct-to-consumer R&D is becoming mainstream.
The RUDY study: using digital technologies to enable a research partnership
Harriet J A Teare¹, et al

They desire price transparency but do not yet understand how to shop.
Patients' views on price shopping and price transparency.
Semigran HL, Gourevitch R, Sinaiko AD, Cowling D, Mehrotra A¹

The #1 reason patients use pharmacies is information.
Much of the data generated in healthcare remains inaccessible, and unshared – even within a company, let alone across the healthcare ecosystem. As a result, ongoing efforts to leverage medical data are more limited, biased, and inaccurate relative to a solution that employs a linked ecosystem of data sources at the patient level.

Problem: Information about an individual’s health is fragmented and is not analyzed or shared effectively limiting insight into patient outcomes
Datavant vision: To measurably improve healthcare outcomes through a virtuous cycle of streamlining clinical research and drug development

1. Data aggregation
Datavant is collaboratively building the largest medical data network through strategic partnerships

2. Data linking
Increasing data completeness and dimensionality will enable novel analyses

Datavant’s Goal
To double clinical trial success rates through enabling improved research and development efforts

3. Data structuring
Translating natural language and streamlining data pre-processing provides research utility

4. Eroding analytical barriers
Medical data’s utility is limited due to coding, errors, and inaccessibility, preventing analyses on accurate patient subsets and skewing results

5. Improvements in clinical research & patient outcomes

6. Compounding effects
Continuing to aggregate, link, and inform trials will accelerate value realization of Datavant’s data network