5th Annual

Cloudera Government Forum

cloudera

Cloudera Government Forum

Precision Medicine in the Big Data World

Shawn Dolley, Industry Leader - Health & Life Science

cloudera

Agenda

- Precision medicine overview
- Impact of big data on precision medicine
- Precision medicine analytics and Cloudera Omics overview
- Use cases toward precision medicine
- Summary

cloudera

Cloudera
Government
Forum

Precision Medicine Overview

cloudera

WHAT IS IT?

Precision medicine is an emerging approach for disease prevention and treatment that takes into account people's individual variations in genes, environment, and lifestyle.

The Precision Medicine Initiative® will generate the scientific evidence needed to **move the concept of precision medicine into clinical practice.**

http://www.nih.gov/precisionmedicine/infographic-printable.pdf

WHY NOW?

The **time is right** because of:

Sequencing of the human genome

Improved technologies for biomedical analysis

New tools for using large datasets

http://www.nih.gov/precisionmedicine/infographic-printable.pdf

LONGER-TERM GOALS

Create a research cohort of **> 1 million American volunteers** who will share genetic data, biological samples, and diet/lifestyle information, all linked to their electronic health records if they choose.

http://www.nih.gov/precisionmedicine/infographic-printable.pdf

Precision Medicine - Use of an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis and treatment of disease.

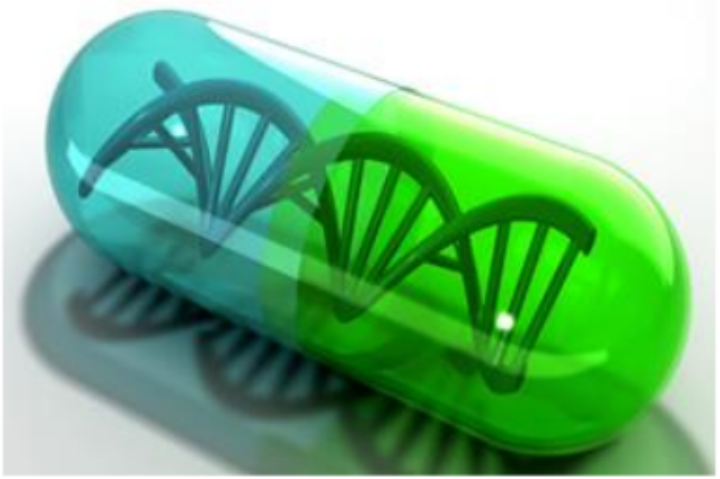


Table 1. Examples of Conditions in Which Precision Medicine Has Been Used.*

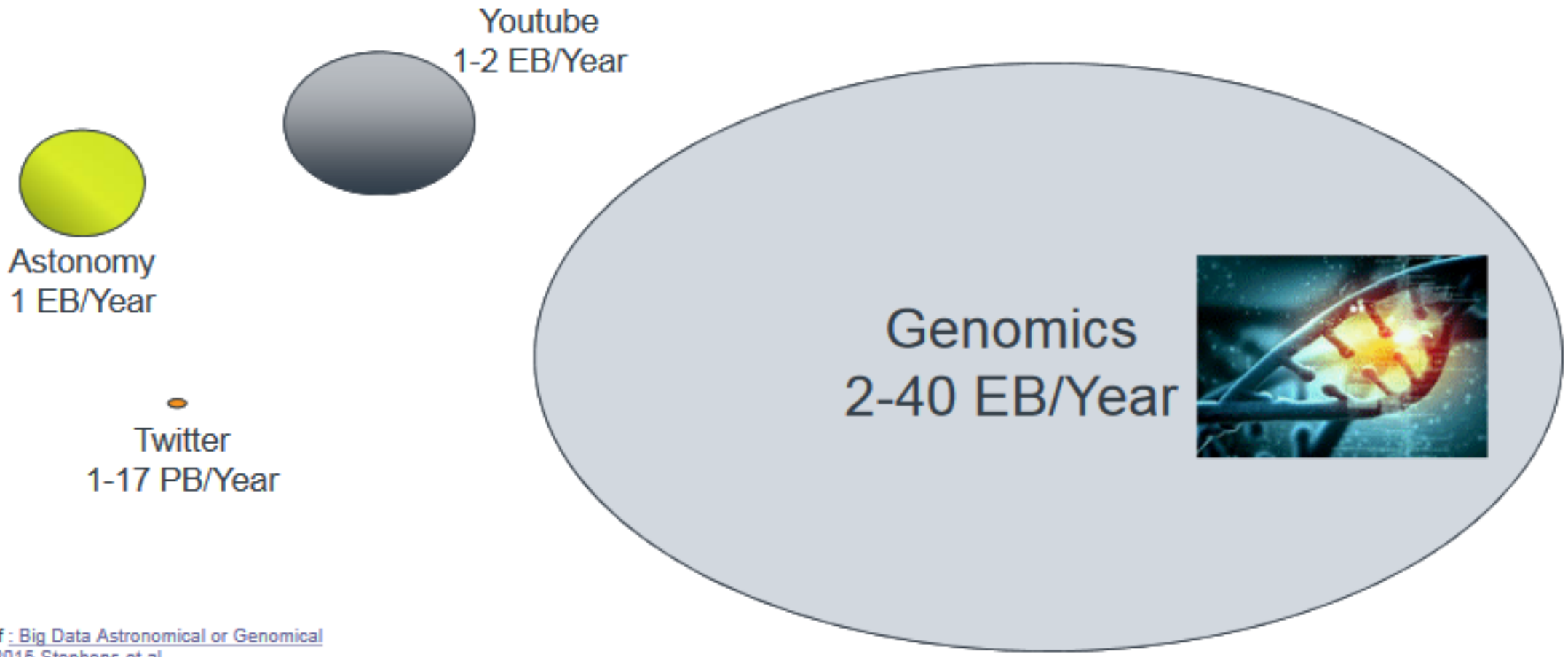
Medical Field	Disease	Biomarker	Intervention
Cancer	Chronic myeloid leukemia	BCR-ABL	Imatinib[4]
	Lung cancer	EML4-ALK	Crizotinib[3]
Hematology	Thrombosis	Factor V Leiden	Avoid prothrombotic drugs[5]
Infectious disease	HIV/AIDS	CD4+ T cells, HIV viral load	Highly active antiretroviral therapy[6]
Cardiovascular disease	Coronary artery disease	*CYP2C19*	Clopidogrel[7]
Pulmonary disease	Cystic fibrosis	*G551D*	Ivacaftor[8]
Renal disease	Transplant rejection	Urinary gene signature	Antirejection drugs[9]
Hepatology	Hepatitis C	Hepatitis C viral load	Direct-acting antiviral agents[10]
Endocrine disease	Multiple endocrine neoplasia type 2	*RET*	Prophylactic thyroidectomy[11]
Metabolic disease	Hyperlipidemia	LDL cholesterol	Statins[12]
Neurology	Autoimmune encephalitis	CXCL13	Immunotherapy[13]
Psychiatry	Alcohol-use disorder	*GRIK1*	Topiramate[14]
Pharmacogenomics	Smoking cessation	*CYP2A6*	Varenicline[15]
Ophthalmology	Leber's congenital amaurosis	*RPE65*	Gene therapy[16]

* In the biomarker column, proteins or genes that are probed to find the specific variants of interest are shown. AIDS denotes acquired immunodeficiency syndrome, HIV human immunodeficiency virus, and LDL low-density lipoprotein.

Jameson JL, Longo DL. N Engl J Med 2015. DOI: 10.1056/NEJMsb1503104

Public domain slide, by Brian Wells, Penn Medicine, http://tinyurl.com/zsayqld

Big Data in 2025 Per Year Growth Estimates – Astronomy, Genomics, Twitter, Youtube

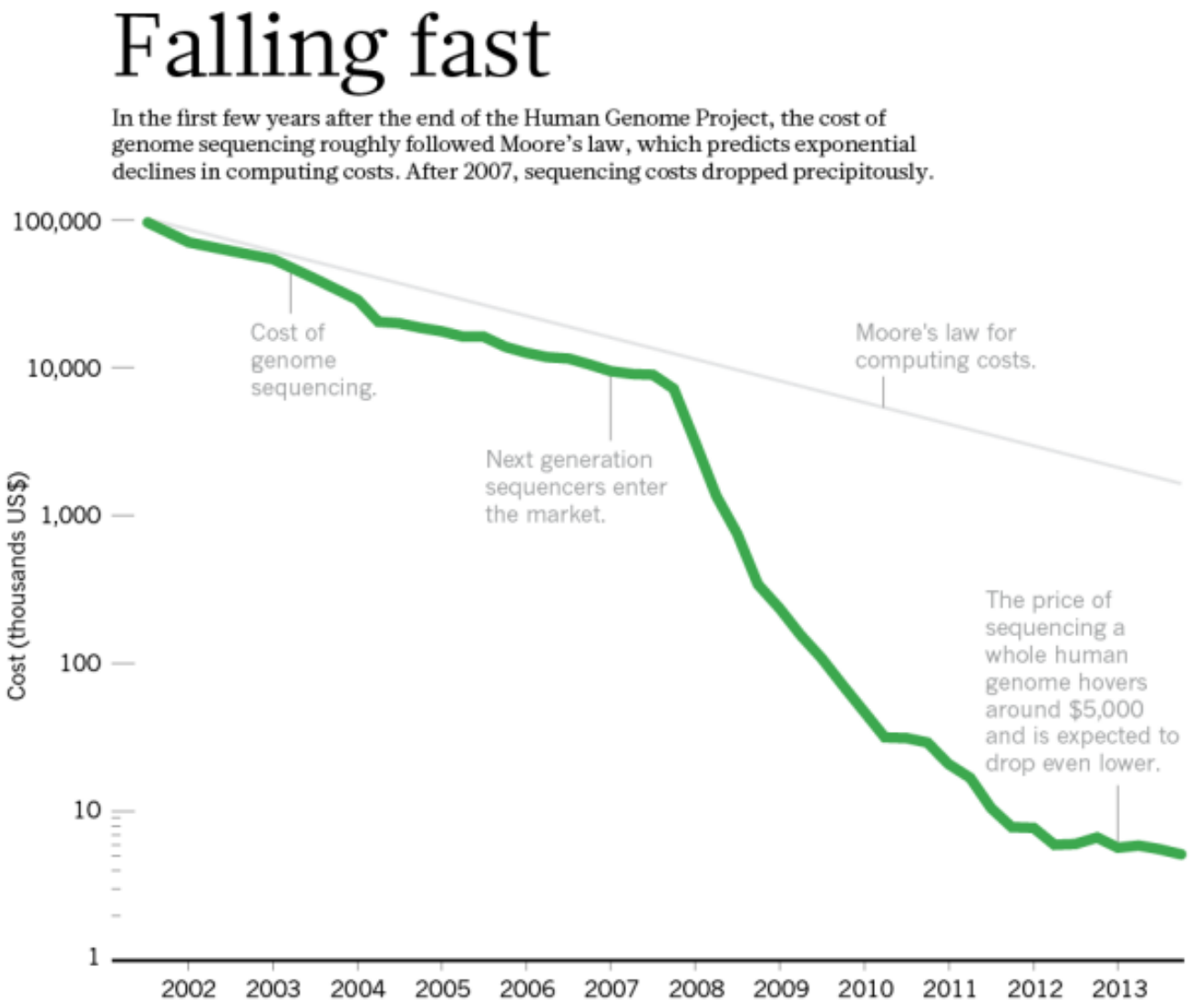


Ref : Big Data Astronomical or Genomical @2015 Stephens et al.
Image Shutterstock

cloudera

Courtesy Cloudian

One Driver: Price of Sequencing Plummets



"In 10 years we've come from a $300M genome to one that's realistically available at around $3000. That's a 100,000 fold drop!"

- James Hadfield, Next Generation Sequencing, 2014

Source: Nature, 2014

cloudera

The Race to *n*

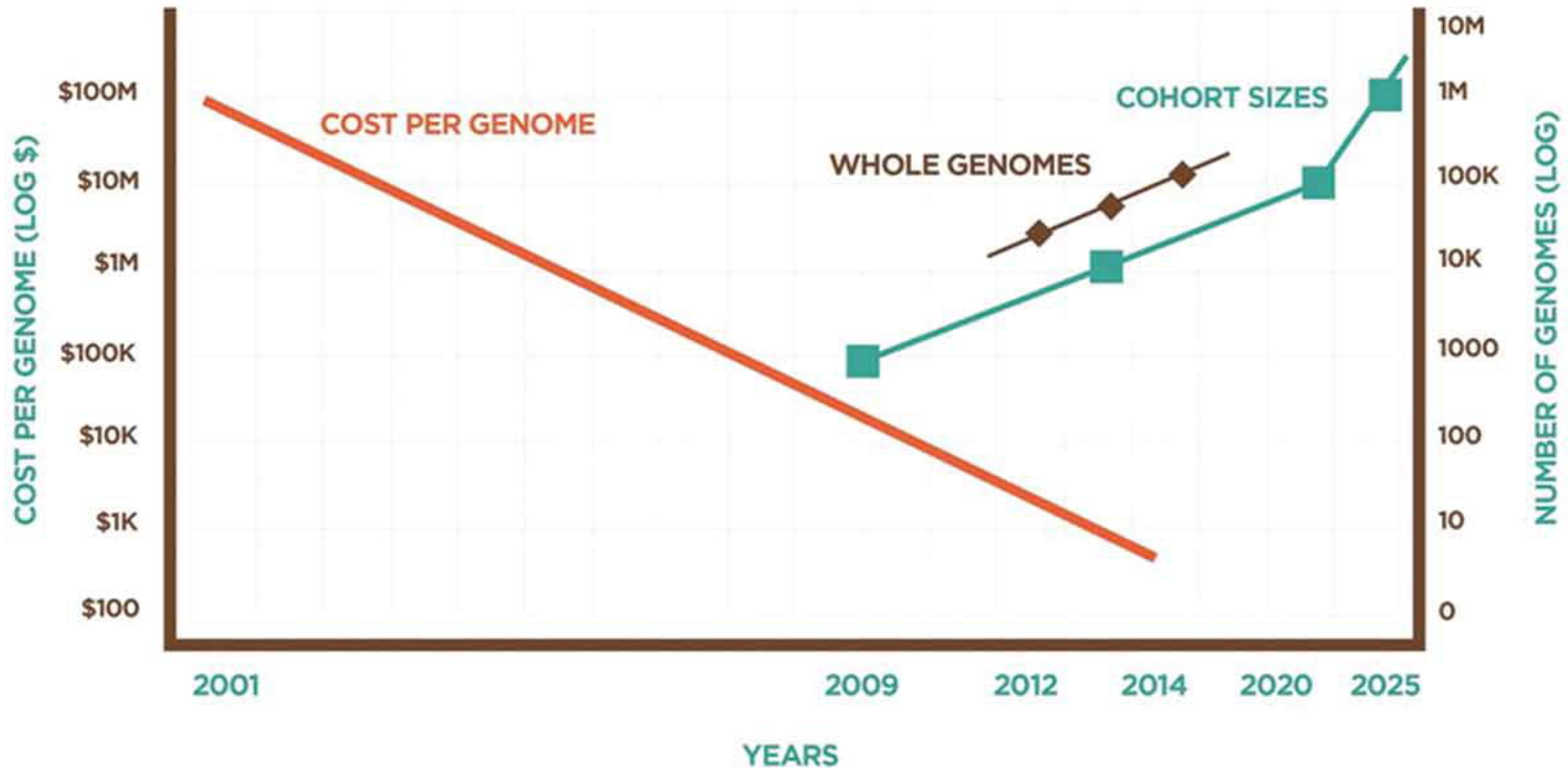


Figure 2. Harnessing the network effect of genome sequencing.

Published in: Susan K. Delaney; Michael L. Hultner; Howard J. Jacob; David H. Ledbetter; Jeanette J. McCarthy; Michael Ball; Kenneth B. Beckman; John W. Belmont; Cinnamon S. Bloss; Michael F. Christman; Andy Cosgrove; Stephen A. Damiani; Timothy Danis; Massimo Delledonne; Michael J. Dougherty; Joel T. Dudley; W. Andrew Faucett; Jennifer R. Friedman; David H. Haase; Tom S. Hays; Stu Heilsberg; Jeff Huber; Leah Kaminsky; Nikki Ledbetter; Warren H. Lee; Elissa Levin; Ondrej Libiger; Michael Linderman; Richard L. Love; David C. Magnus; AnneMarie Martland; Susan L. McClure; Scott E. Megill; Helen Messier; Robert L. Nussbaum; Latha Palaniappan; Bradley A. Patay; Bradley W. Popovich; John Quackenbush; Mark J. Savant; Michael M. Su; Sharon F. Terry; Steven Tucker; William T. Wong; Robert C. Green; *Expert Review of Molecular Diagnostics* Ahead of Print
DOI: 10.1586/14737159.2016.1146593

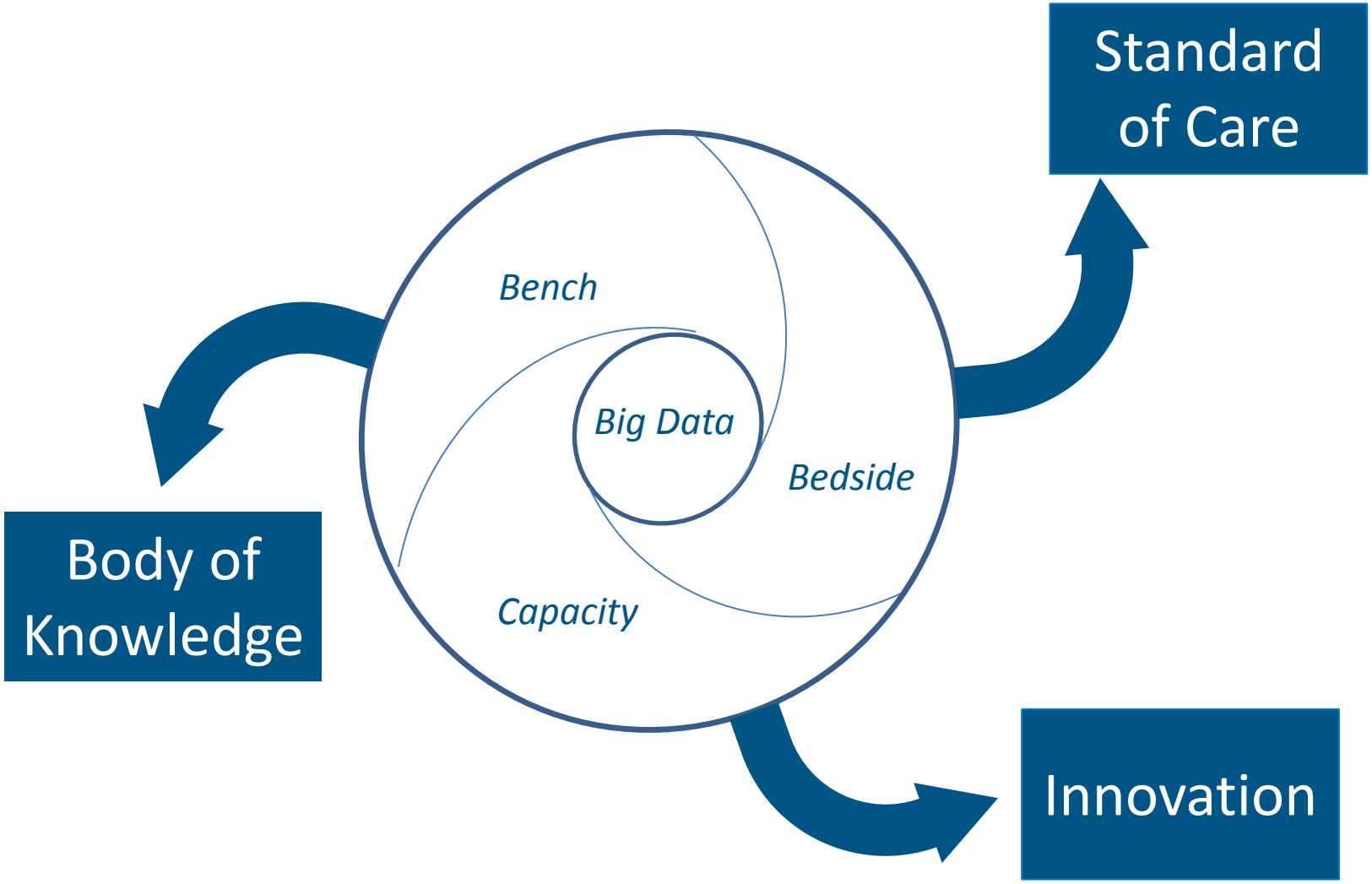
cloudera

Cloudera Government Forum

Impact of Big Data on Precision Medicine

cloudera

Big Data Bringing Clinicians & Researchers Together

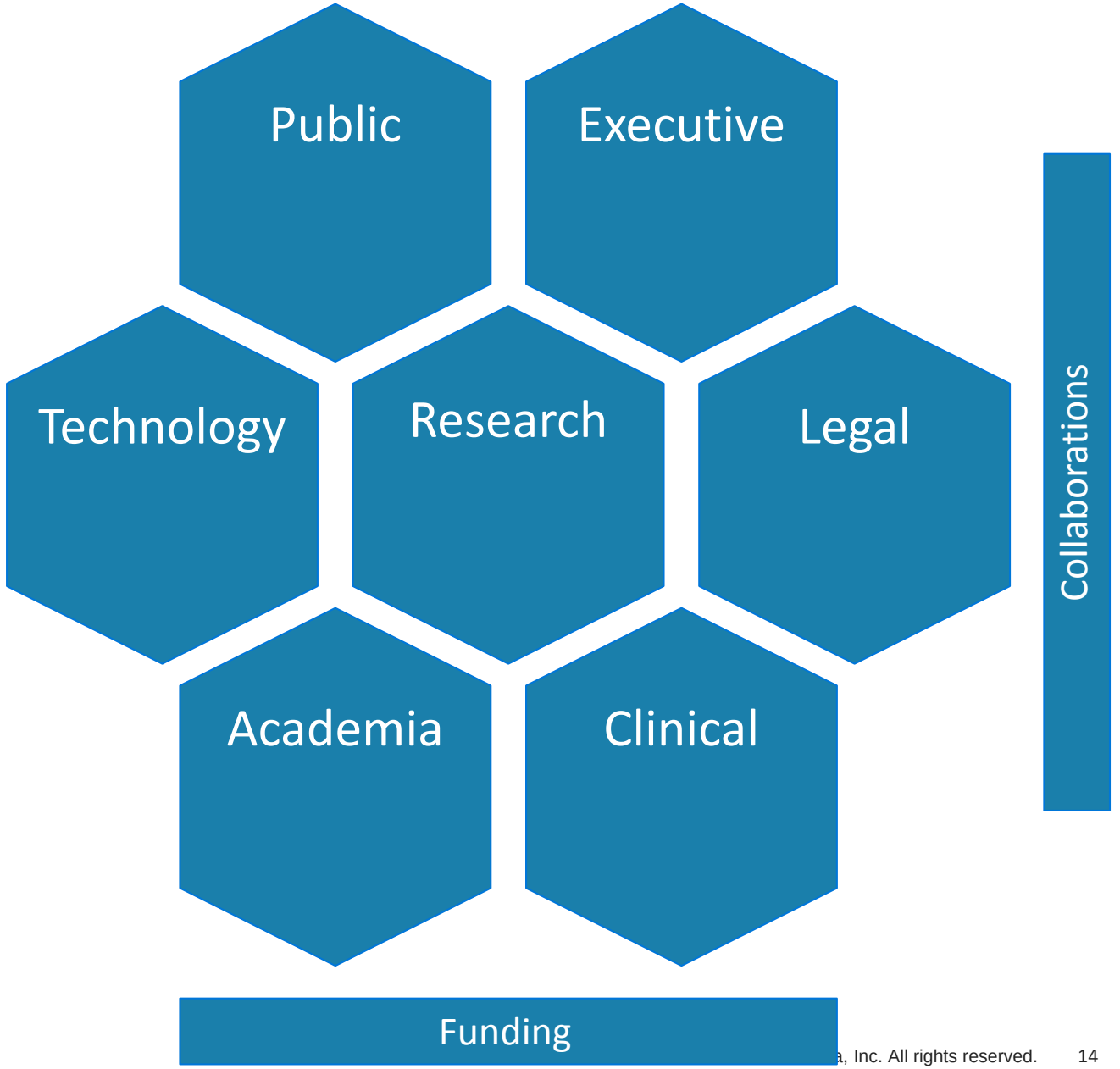


Courtesy: Todd Sharpe, Children's Healthcare of Atlanta/Emory

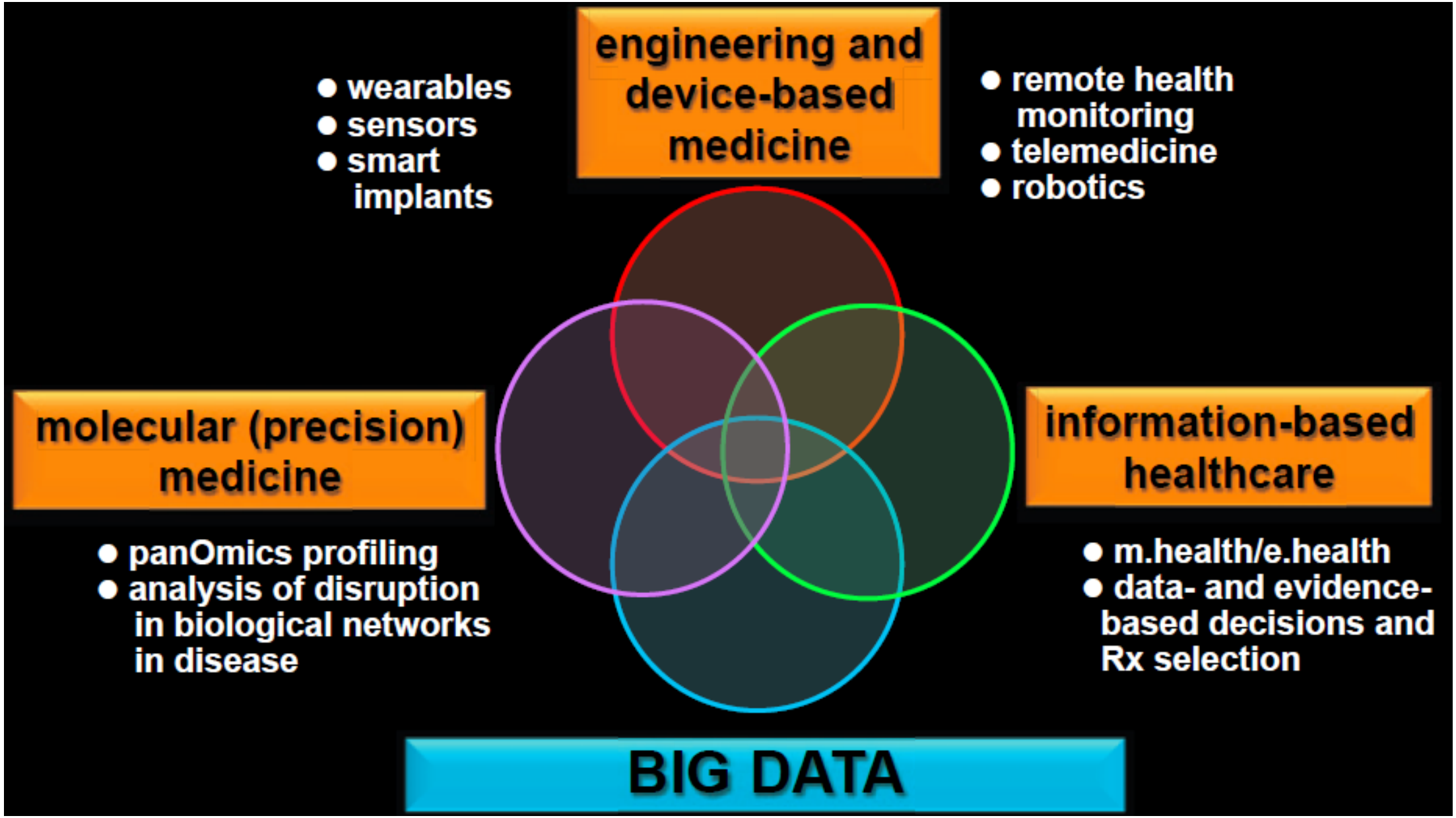
Big Data Needs a New ‘core’

Impacts

- Focus on genomics drives attention to the research function
- Clinical genetics drive to sequencing drives attention to the hospital
- Big Data’s history in open source is adding weight to ‘open data’



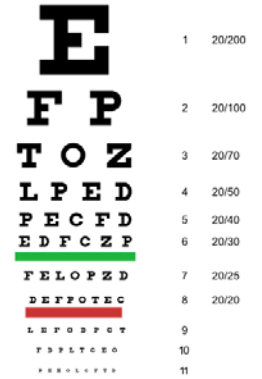
cloudera



Public domain slide, by Dr. George Poste, Univ. of Mich, http://tinyurl.com/j7nr27x

The Big Data

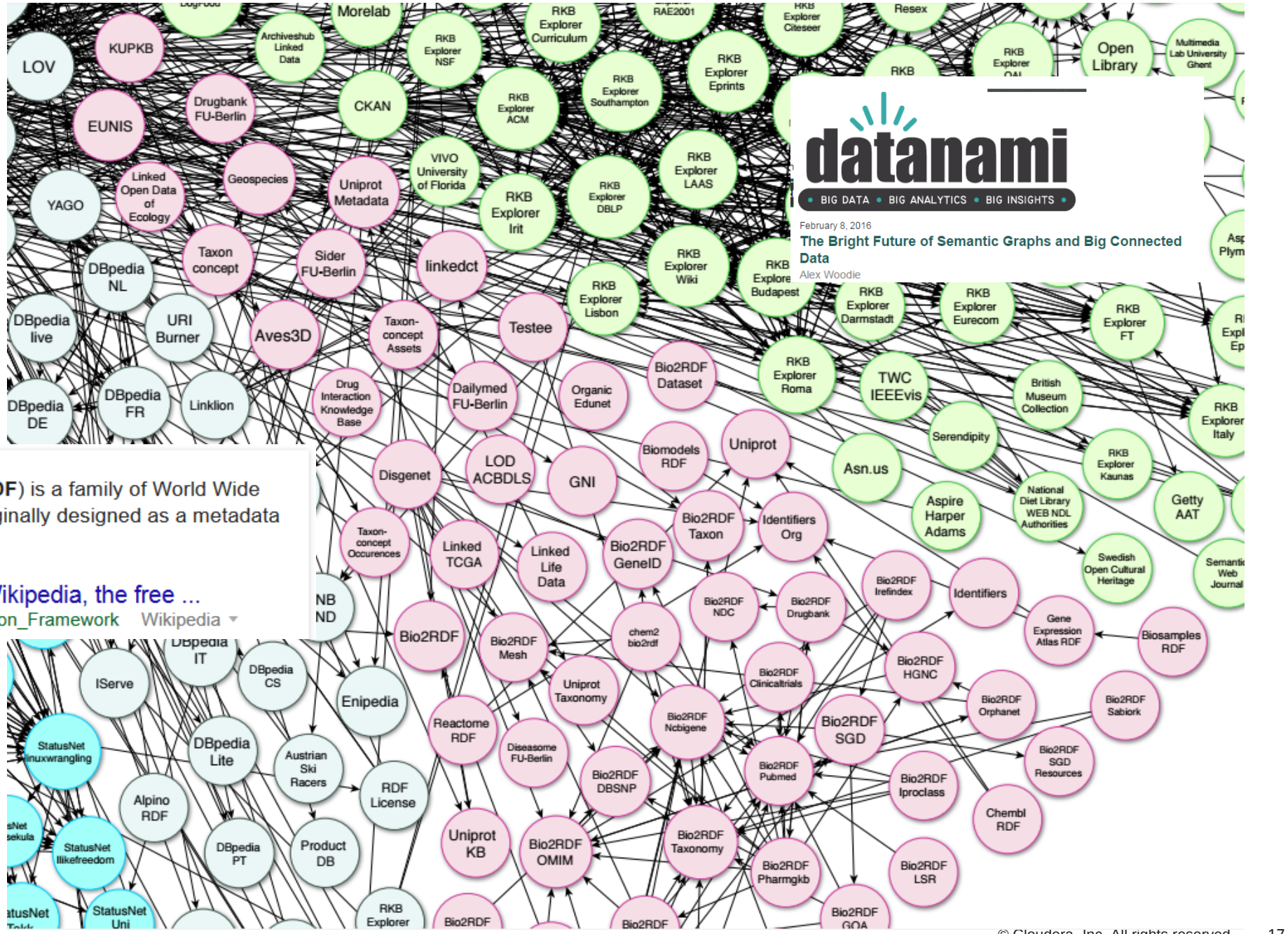
The 'Long Tail' starts with linking full electronic medical record & the whole exome



Electronic Health Record	ADT & Patient Documents	Genomics/RNA/ Microbiome	WHO Signal	PubMed	1000 Genomes, ExAC
Clinical Genetics	Dictionaries / Ontologies	Proteomic/ Metabolomic	AHRQ Datasets	Social Signal	ENCODE
Image Repository	Study Data	Remote Monitor Streams	CMS Datasets	Licensed/POS Signal	Geuvadis
Labs & Screens	Custom Exchanges	Inpatient Bedside Streams	US HHS OpenData	Elsevier Datasets	GTex
Claims	Clinical Trials	Wearable Streams	USAID DHS	GNS REFS	SRA
Real-Time Location	NPDM	NaviNet Provider Portal	NIH Datasets	Premier	COSMIC
Financial & Supply Chain	Measures/HEDIS	Apple HealthKit / PHR	IMS Health, Symphony	Thomson Reuters Cortellis	TCGA
Registries	TeleHealth	Partner Data Exchanges	CERNER HealthFacts	TR Clinical Genetic Toolkit	HGMD

cloudera

Open Web 2014



The Resource Description Framework (**RDF**) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model.

Resource Description Framework - Wikipedia, the free ...
https://en.wikipedia.org/wiki/Resource_Description_Framework Wikipedia

http://www.datanami.com/2016/02/08/the-bright-future-of-semantic-graphs-and-big-connected-data/

cloudera

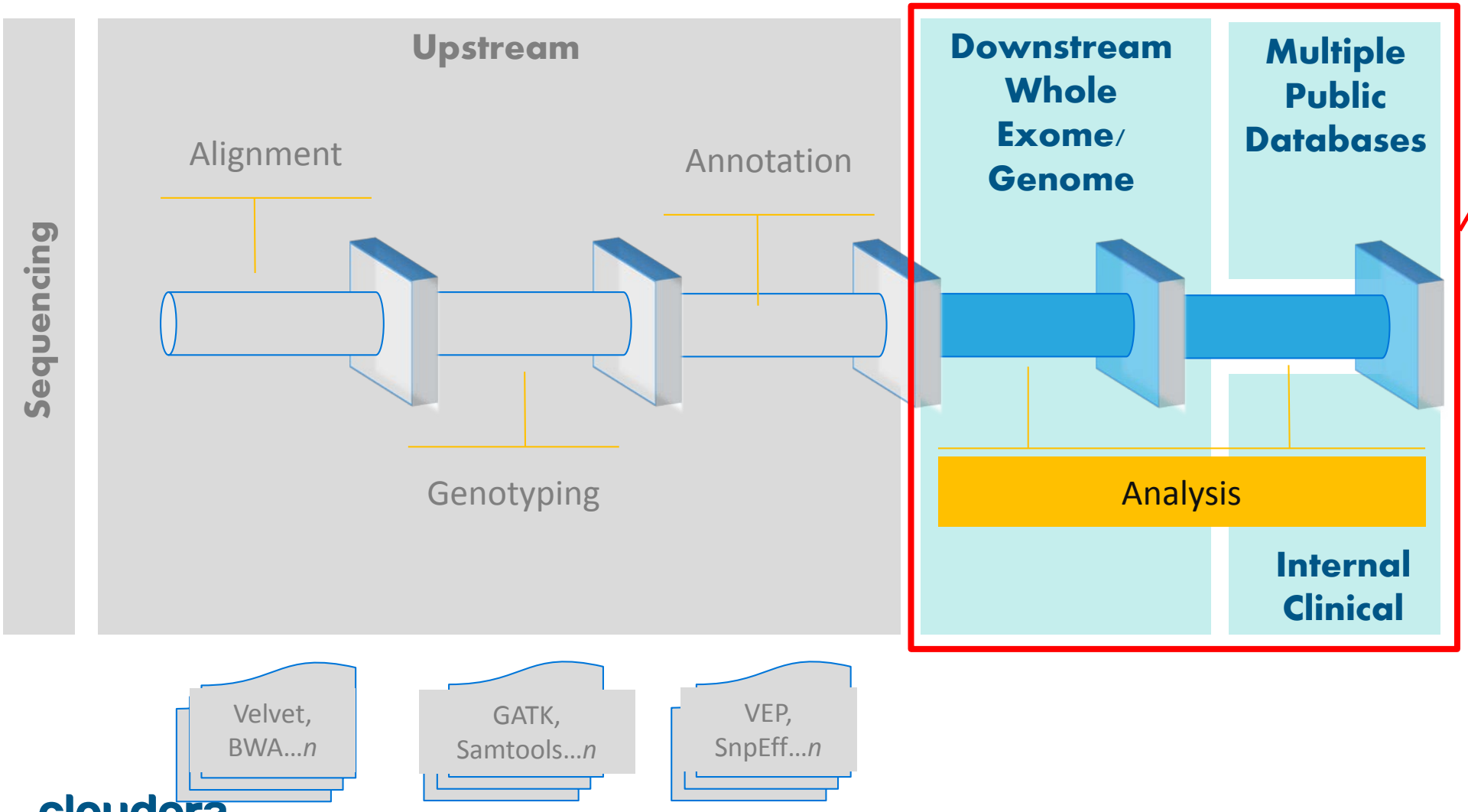
http://lod-cloud.net/versions/2014-08-30/lod-cloud_colored.pdf

Cloudera
Government
Forum

Precision Medicine Analytics & Cloudera Omics

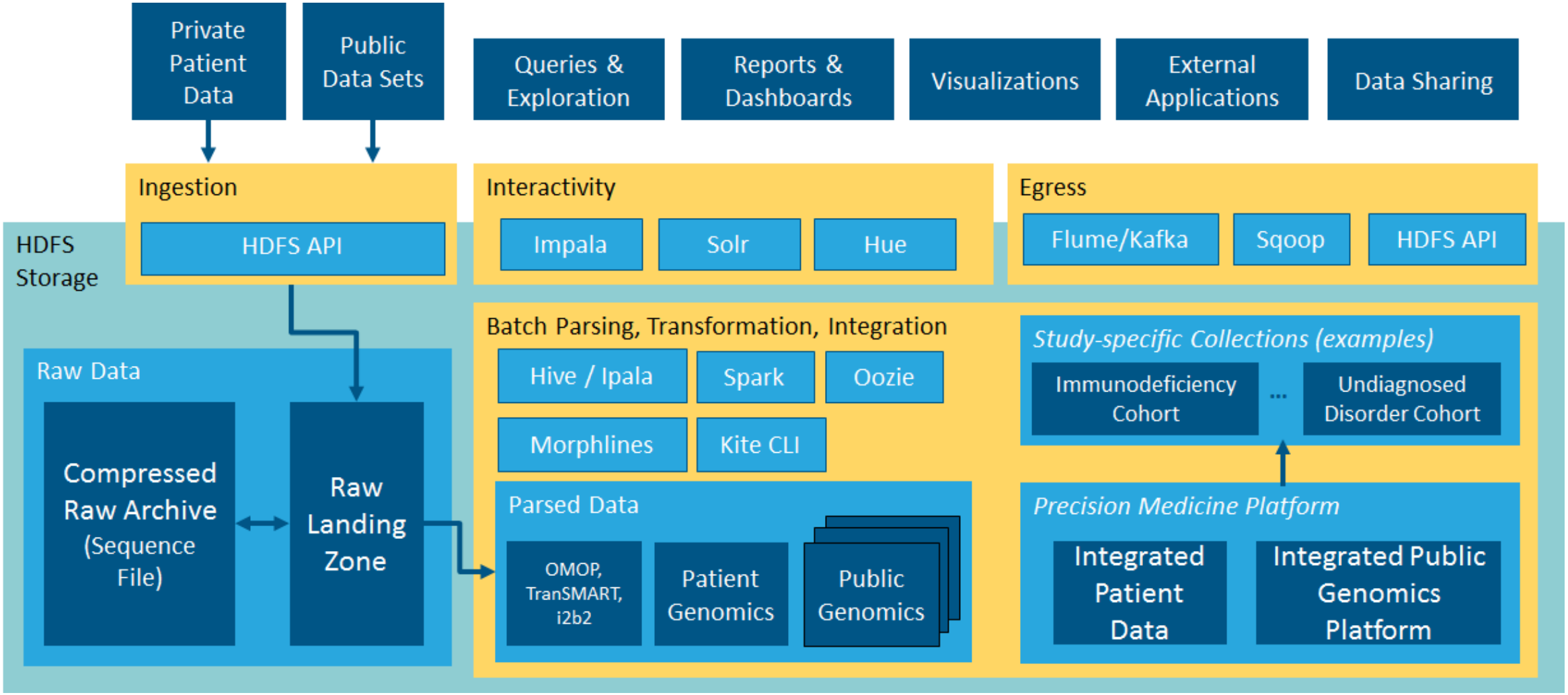
cloudera

The Genomic Analytic Pipeline & Cloudera Omics



cloudera

Cloudera Omics Reference Architecture

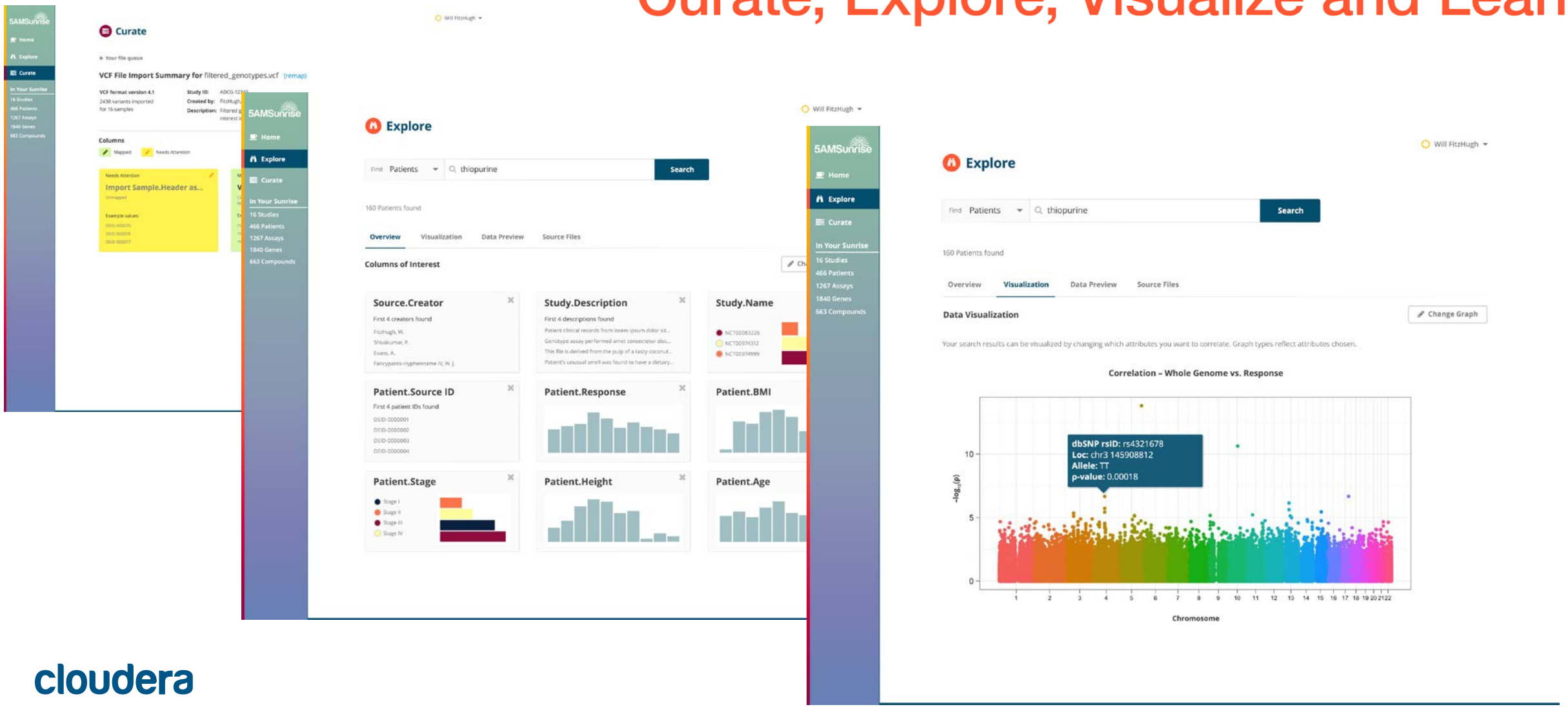


Extended Reference Architecture for Multi-Omics

cloudera

5AMSunrise

Curate, Explore, Visualize and Learn



cloudera

Use Cases that lead us to Precision Medicine

- Signal Detection
- Clinical Prediction
- Real World Evidence

Signal Detection: Epidemiology Treatment Analytics

Top 10 Global Pharma does Epidemiology Research on Cloudera

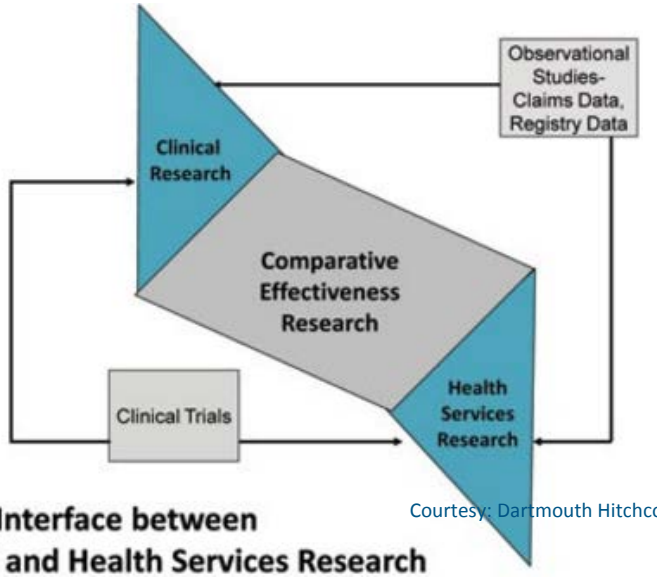
Application: Comparative Effectiveness Research

Data: Claims data from partners consolidated into the Observational Medical Outcomes Partnership data model

Users: 50+ pharmaco-epidemiologists

Challenge: Complex analytics must run in parallel to ever return results; data sizes large; need a platform for multiple analytics, R, SAS, SQL, etc.

Benefit: Find differential survival rates for many drugs by population, dose, other factors, find new off-label uses for existing drugs



CER at Interface between Clinical and Health Services Research

Courtesy: Dartmouth Hitchcock

Comparative effectiveness research is designed to inform health-care decisions by providing evidence on the effectiveness, benefits, and harms of different treatment options. The evidence is generated from research studies that compare drugs, medical devices, tests, surgeries, or ways to deliver health care. -US DHHS

cloudera

http://www.statista.com/statistics/272717/top-global-biotech-and-pharmaceutical-companies-based-on-revenue/

Clinical Prediction: Patient Pathways

Health systems adopt predictive analytics for a range of outcome improvement

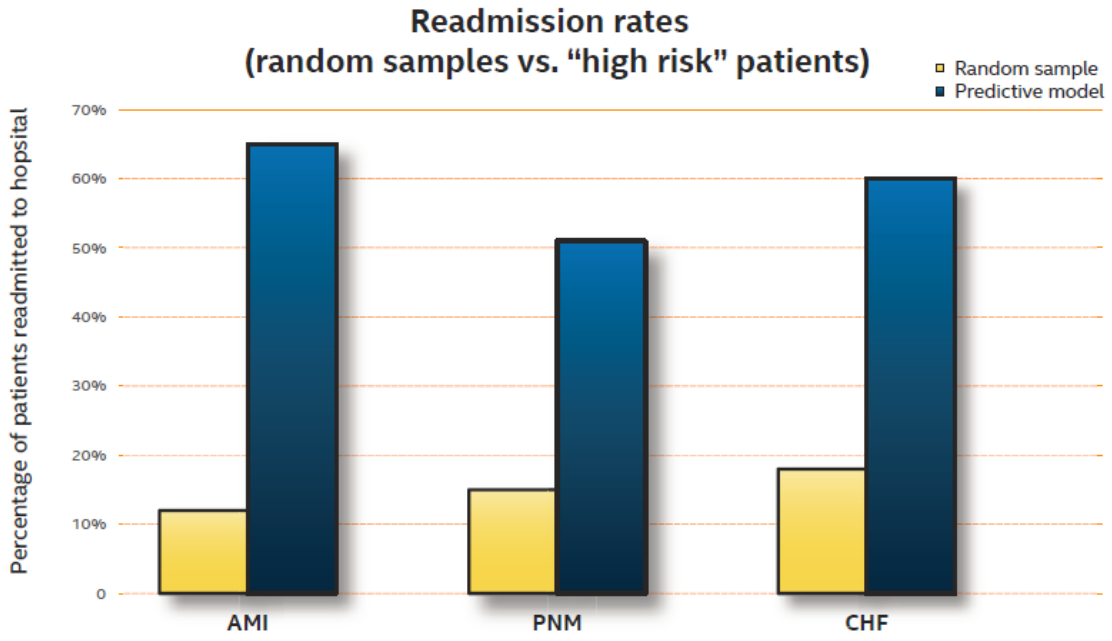
Application: Predict clinical events

Data: Electronic health records per patient, demographics

Users: Care coordinators, CFO, nursing

Challenge: Can't provide care coordination for all patients, and can't achieve hot-spotting

Benefit:



Using the derived predictions from the analysis, the Hospital Group reaps the following annual savings:

- Reduces 6,000 occurrences of patient readmission.
- Avoids $4 million in potential Medicare penalties.
- Saves approximately $72 million in medical service costs.
- Utilizes resources more efficiently by providing extra care to high-risk patients.
- Improves hospital rating based on lower readmission rate and increased patient satisfaction.

cloudera

Real World Evidence

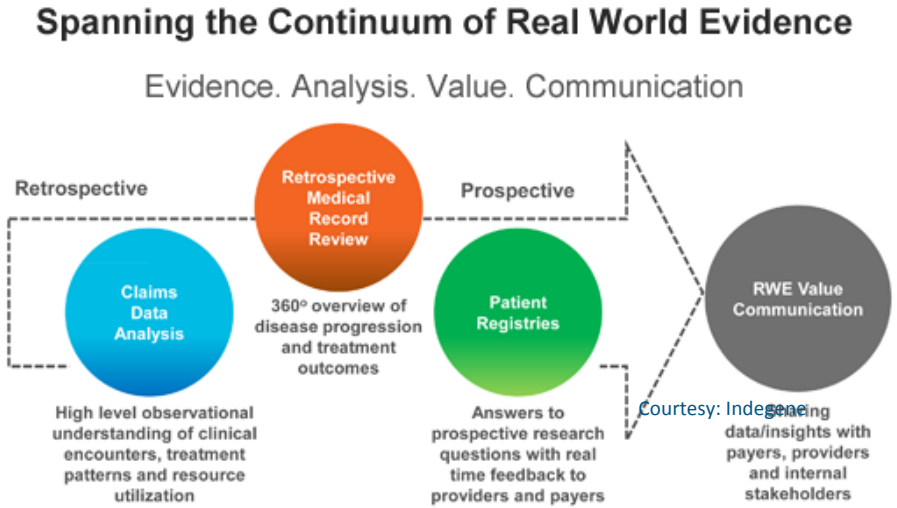
Top 20 Global Pharma does Compiles 'Superset of Lives'

Application: Real World Evidence

Data: Licensed data from Symphony Health, GE, Thomson Reuters, IMS, Optum & others

Challenge: Create a superset of millions-of-patients' health outcomes from the universe of licensed data

Benefit: Create a much larger clinical 'n', that can be used in clinical trials, epi, health economics and outcomes research, pharmacogenomics, pharmacovigilance



Real World Evidence (RWE): Millions of patient experiences in the real world, supplementing randomized clinical trials that typically involve a few thousand patients in a controlled setting. Using RWE, patients, providers of care and those who pay for it can better assess the value of treatments and services based on actual health outcomes and the total cost of care." -IMS

cloudera

http://www.statista.com/statistics/272717/top-global-biotech-and-pharmaceutical-companies-based-on-revenue/

The Best-in-class Organizations Use Cloudera

#1 Largest Payer in the US will be covering 123 million lives and pay out $950B to providers in 2015.

5 out of the top 9 drugs worldwide are made by Cloudera customers.

Over 40 health & life science organizations use Cloudera software.

#1 Largest health data company, with 500M+ anonymous patient records.

#1 Largest Biotech in the world.

#1 commercial hospital chain worldwide.

The maker of the #1 Top Selling Drug worldwide.

#1 most utilized Patient Centered Medical Home program.

this Hospital was one of the first four to receive Stage 7 status from HIMSS, the highest possible distinction in electronic medical records implementation, uses Cloudera to host a variety of data, and was awarded by US DHHS a Gold Medal of Honor.

#1 Largest Health IT company in the World, $3B+ in revenue has 1000's of nodes of Cloudera.

cloudera

Summary

- Genomics is upon us
- Precision medicine is medicine
- It’s all (very) big data
- Data brings us together
- Cloudera can help you get started with technical, subject matter expertise & Omics accelerators

cloudera

Cloudera
Government
Forum

Thank You