



Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική
Σχολή Θετικών Επιστημών
Πανεπιστήμιο Θεσσαλίας

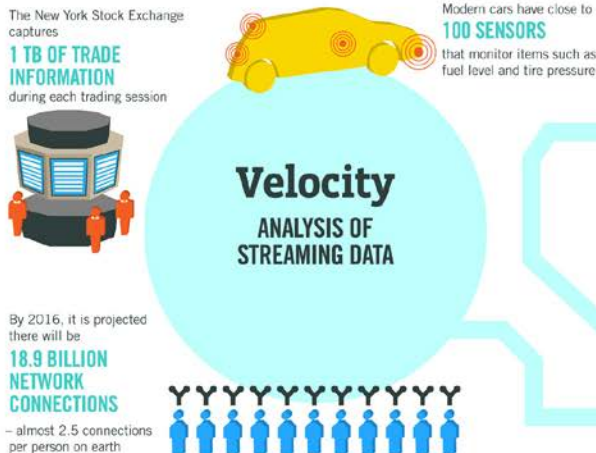
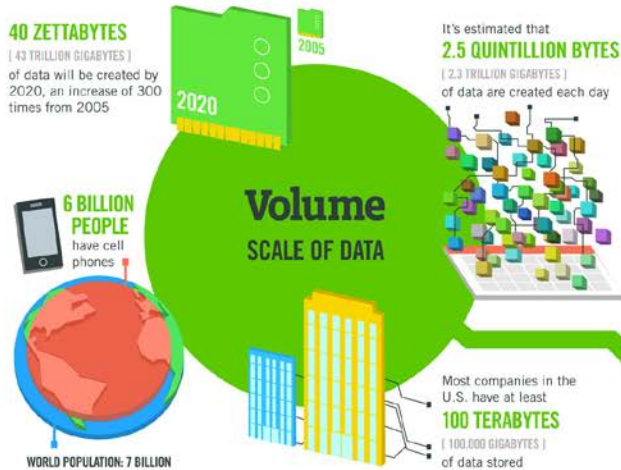


ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Εισαγωγικό Μάθημα

Αριστείδης Γ. Βραχάτης, Dipl-Ing, M.Sc, PhD

Big Data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

Source: IBM Business Analytics

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT are shared on Facebook every month



By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month



400 MILLION TWEETS are sent per day by about 200 million monthly active users

Variety
DIFFERENT FORMS OF DATA



1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



27% OF RESPONDENTS

in one survey were unsure of how much of their data was inaccurate

Veracity
UNCERTAINTY OF DATA

2019 *This Is What Happens In An Internet Minute*



Created By:
@LoriLewis
@OfficiallyChadd

What to do with these data?



- Aggregation and Statistics
- Indexing, Searching, and Querying
 - Keyword based search
 - Pattern matching (XML/RDF)
- Knowledge discovery
 - Data Mining
 - Statistical Modeling

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.

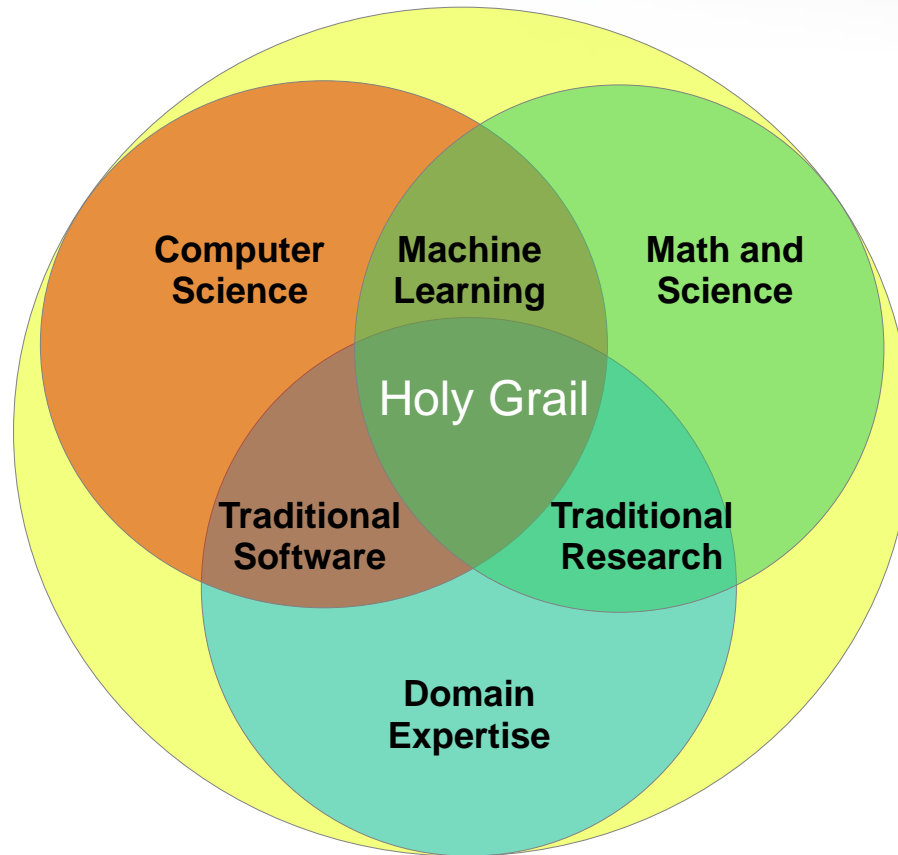


1950's 1960's 1970's 1980's 1990's 2000's 2010's

An Example: Self Driving Cars



Data Science



Data is the New Oil!

Forbes / Tech

APR 2, 2012 @ 11:09 AM 25,812 VIEWS

Is Data The New Oil?



Perry Rotella, CONTRIBUTOR

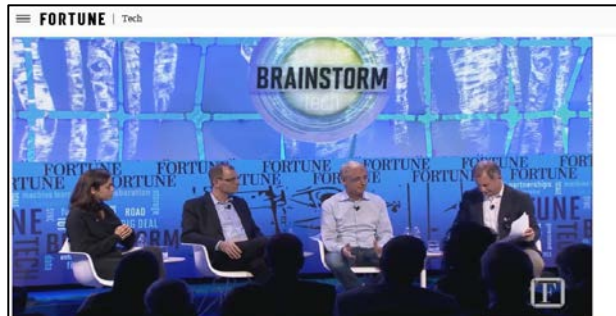
[FULL BIO](#)

Opinions expressed by Forbes Contributors are their own.

Recently, on a CNBC Squawk Box segment, “[The Pulse of Silicon Valley](#),” host Joe Kernan posed the question, “What is the next really big thing?” to [Ann Winblad](#), the legendary investor and senior partner at Hummer-Winblad. Her response: “Data is the new oil.”

Winblad talked about predictive analytics as the new hot spot for venture investing and discussed the growth of companies that can derive value from the huge amounts of data being stored.

This was not the first time we heard the phrase “data is the new oil,” and it certainly will not be the last. For example, marketing commentator [Michael Palmer](#) blogged (http://ana.blogs.com/maestros/2006/11/data_is_the_new.html) back in 2006: “Data is just like crude. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc., to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”



FORTUNE | Tech

BRAINSTORM

Why Data Is The New Oil

Jonathan Vanian
Jul 11, 2016

The field of artificial intelligence is red hot thanks in part to big companies like Google, Facebook, and Microsoft using AI-related techniques to train computers to recognize objects in photos and understand human language.

WIRED Data Is the New Oil of the Digital Economy

DATA IS THE NEW OIL OF THE DIGITAL ECONOMY



Image: verifex/Flickr

DATA IN THE 21st Century is like Oil in the 18th Century: an immensely, untapped valuable asset. Like oil, for those who see Data's fundamental value and learn to extract and use it there will be huge rewards.

The Fourth Industrial Revolution



1st Industrial Revolution WATER & STEAM

Steam and water power replace human and animal power with machines.



2nd Industrial Revolution ELECTRICITY

Electricity, internal combustion engines, airplanes, telephones, cars, radio, and mass production.



3rd Industrial Revolution AUTOMATION

Electronics, the internet and IT used to further the automation of mass production.



4th Industrial Revolution CYBER-PHYSICAL SYSTEMS

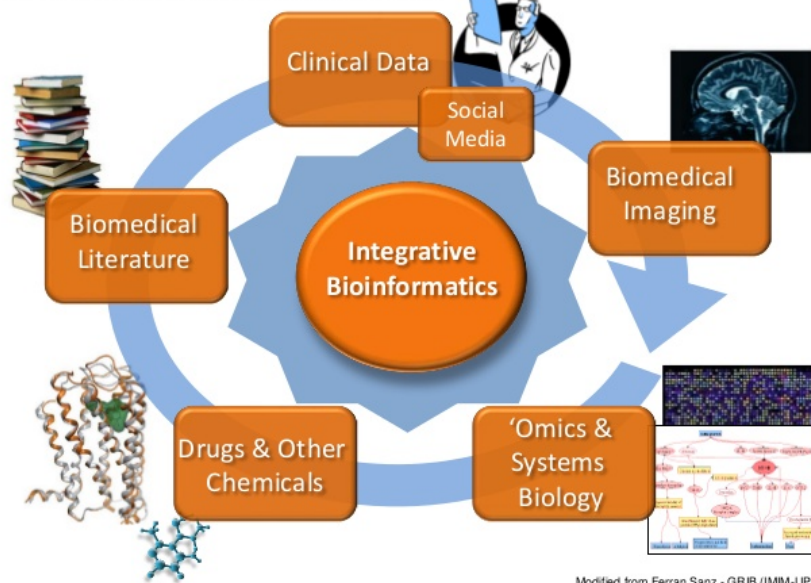
Driverless cars, smart robotics, materials that are lighter and tougher, and a manufacturing process built around 3D printing.

Big Data in BioMedicine



- Big data analytics is transforming biology and medicine by allowing researchers to extract more information and knowledge from rapidly growing diverse types of data.

Integration of heterogeneous biomedical information
to gain a more complete and powerful view on diseases and therapeutics



Modified from Ferran Sanz - GRIB (IMM-UPF)

HUMAN GENOME, IN NUMBERS



6 billion **22,000**
DNA LETTERS **GENES**

A T
C G
T A

46 CHROMOSOMES

\$9,500 **500GB** **SIZE**
COST TO SEQUENCE **ON**
DISK

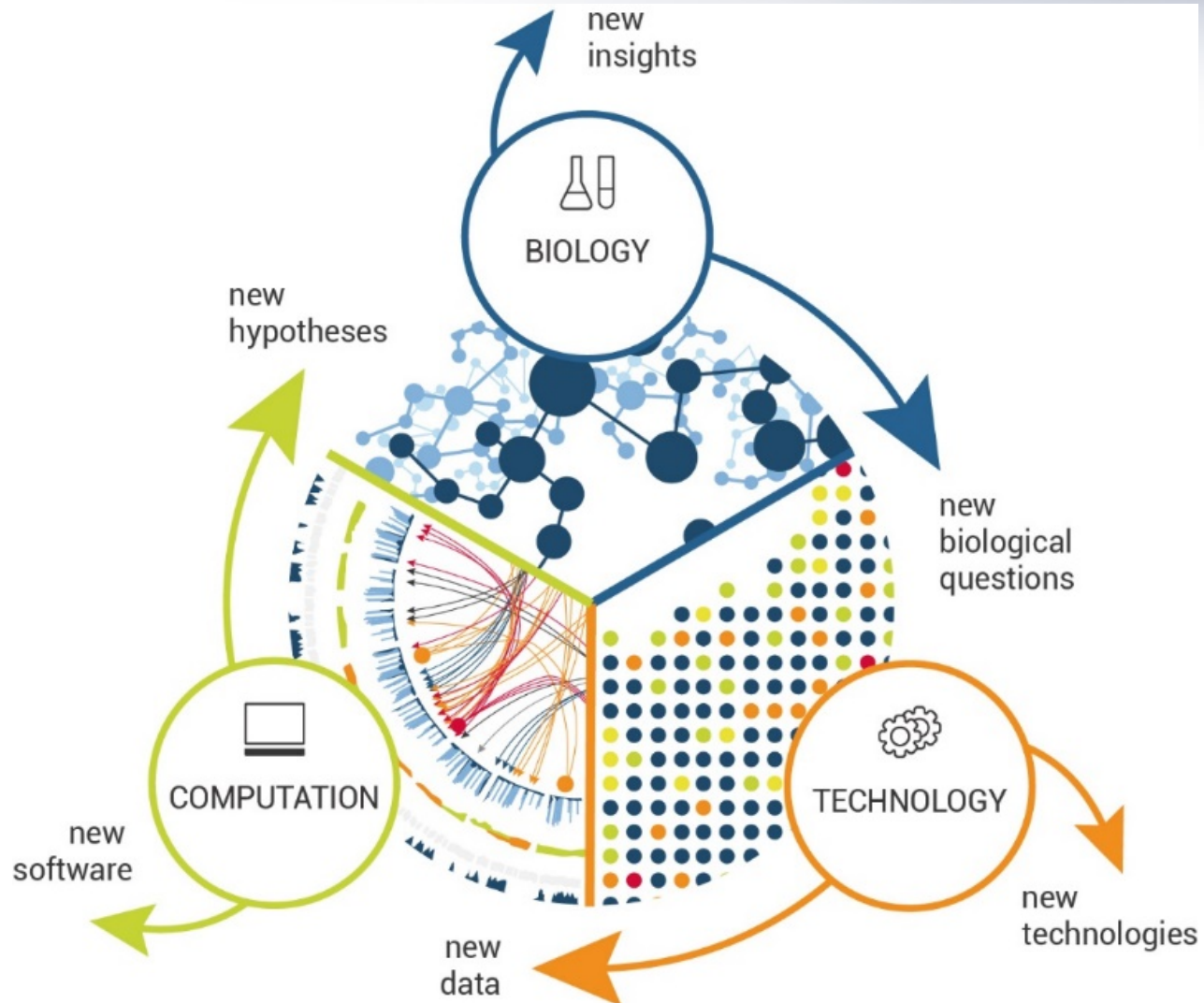
Sources: NIH, Illumina



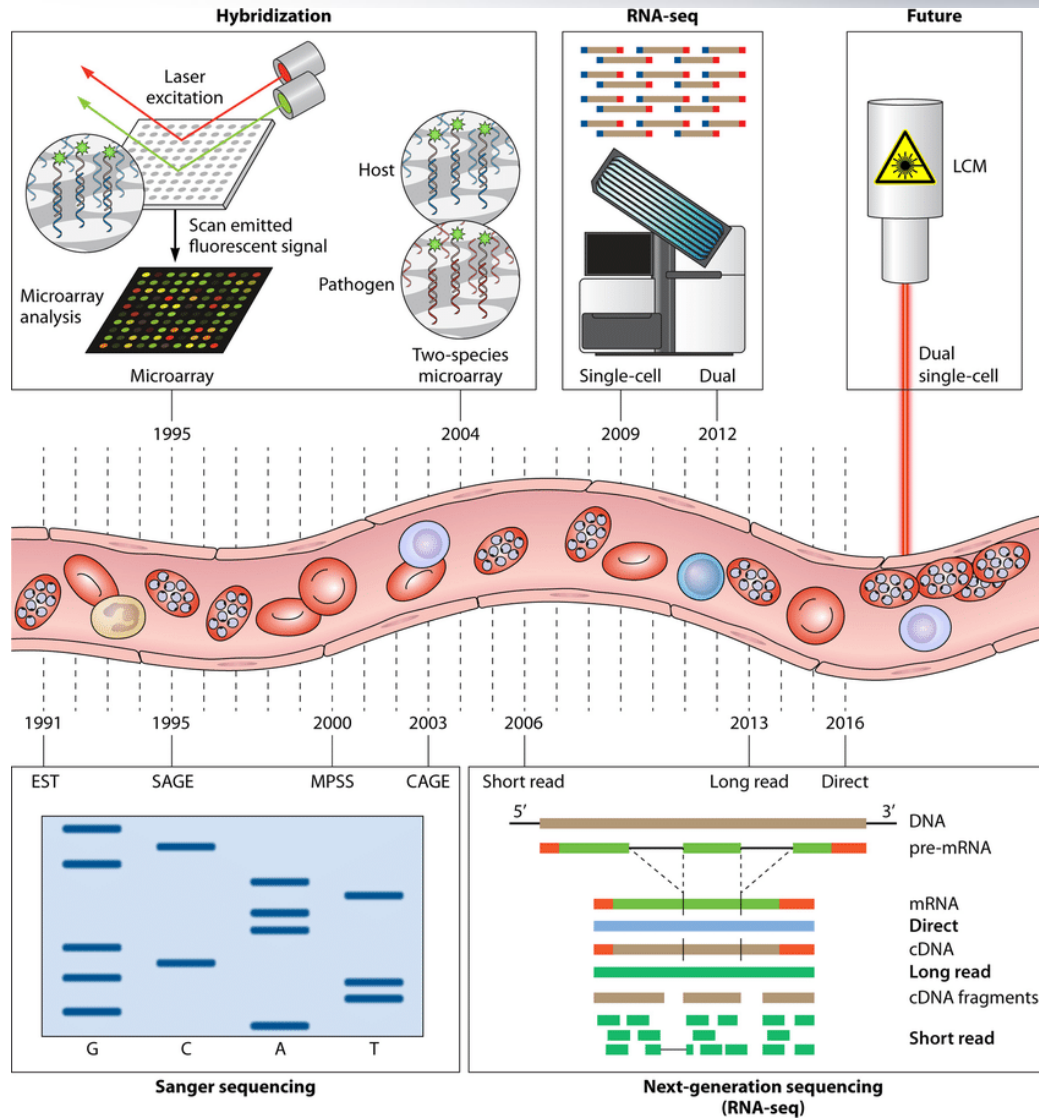
106 STARTUPS TRANSFORMING HEALTHCARE WITH AI



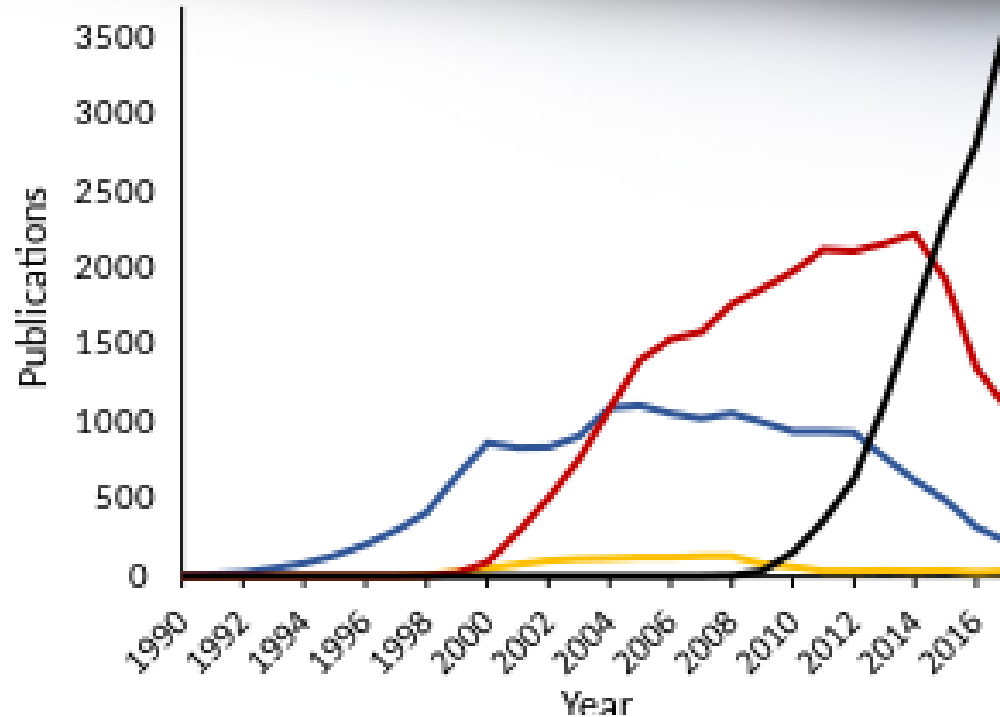
Big Data drives Medicine



Transcriptomics



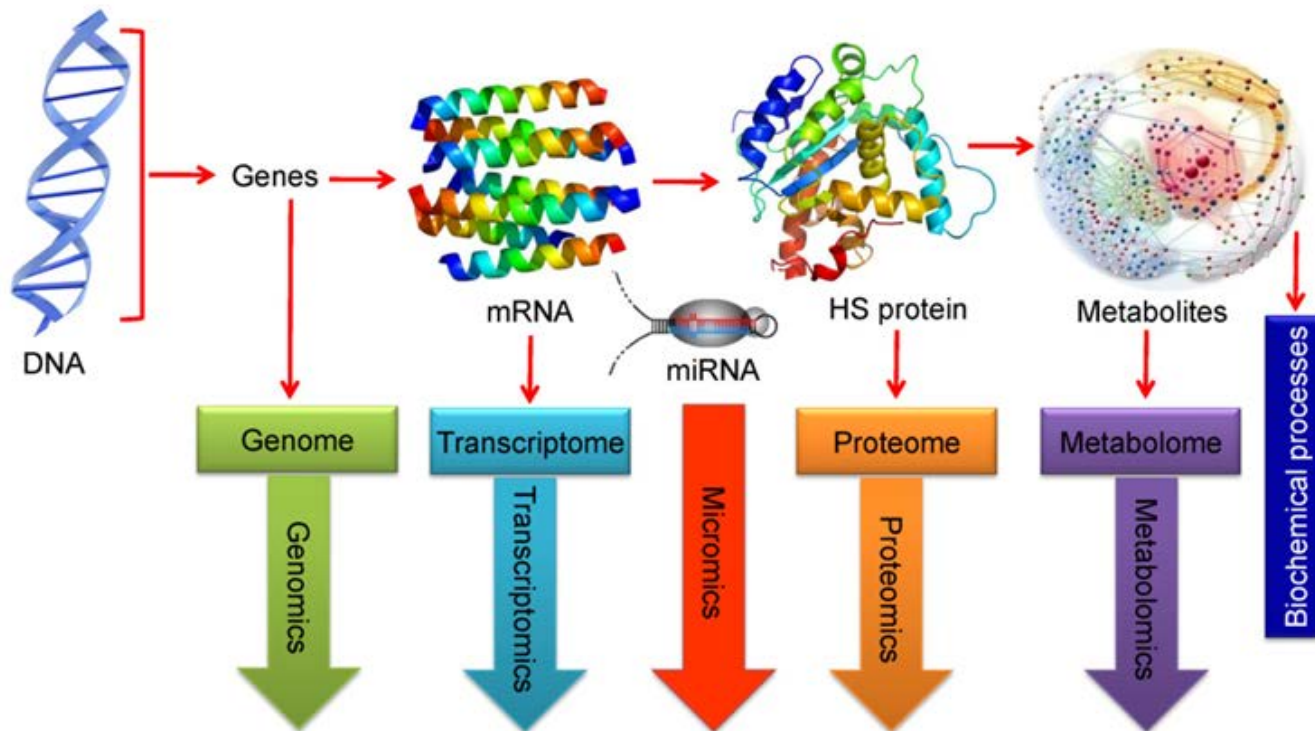
Transcriptomics



- Transcriptomics method use over time.
- Published papers referring to
 - **RNA-Seq (black),**
 - **RNA microarray (red),**
 - **expressed sequence tag (blue)** and
 - **serial/cap analysis of gene expression (yellow)** since 1990.

Omics Data

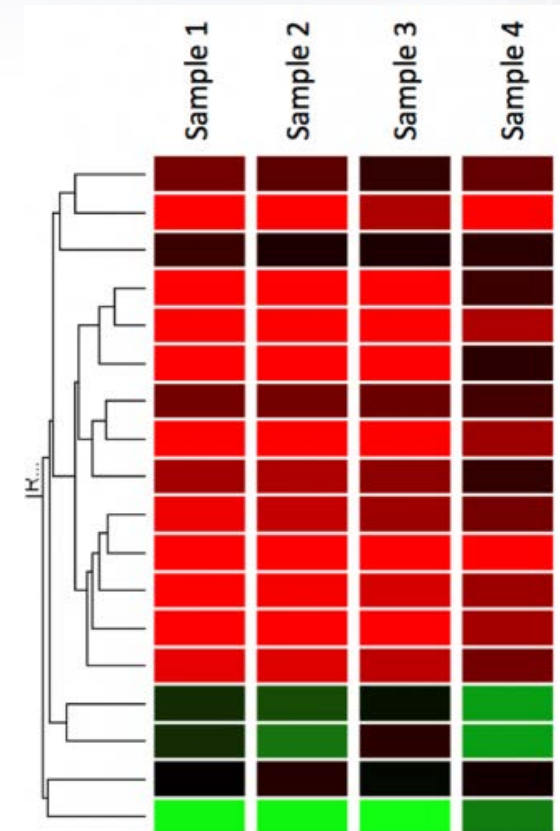
- A field of study in biology ending in -omics, such as genomics, proteomics or metabolomics etc.
- Omics Data Explosion Challenges → Heterogeneous Data Integration, Big Data Analysis etc.



Processed & Normalized Matrix



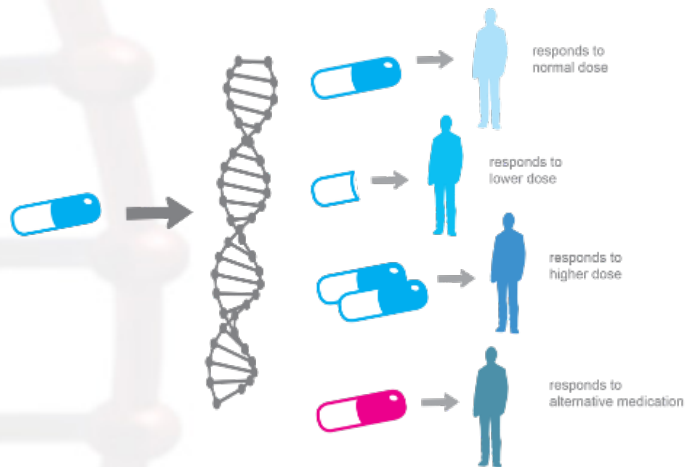
	Sample 1	Sample 2	Sample K
Gene_1	value (1,1)	value (1,2)	value(1,K)
Gene_2	value (2,1)	value (2,2)	value(2,K)
.....
.....
.....
.....
Gene_N	value (1,1)	value (1,1)	value(N,K)



Systems Biology Extensions



— Personalized Medicine



— Precision Medicine



GENOMICS



PHENOTYPE



LIFESTYLE/ENVIRONMENT

— P4 Medicine

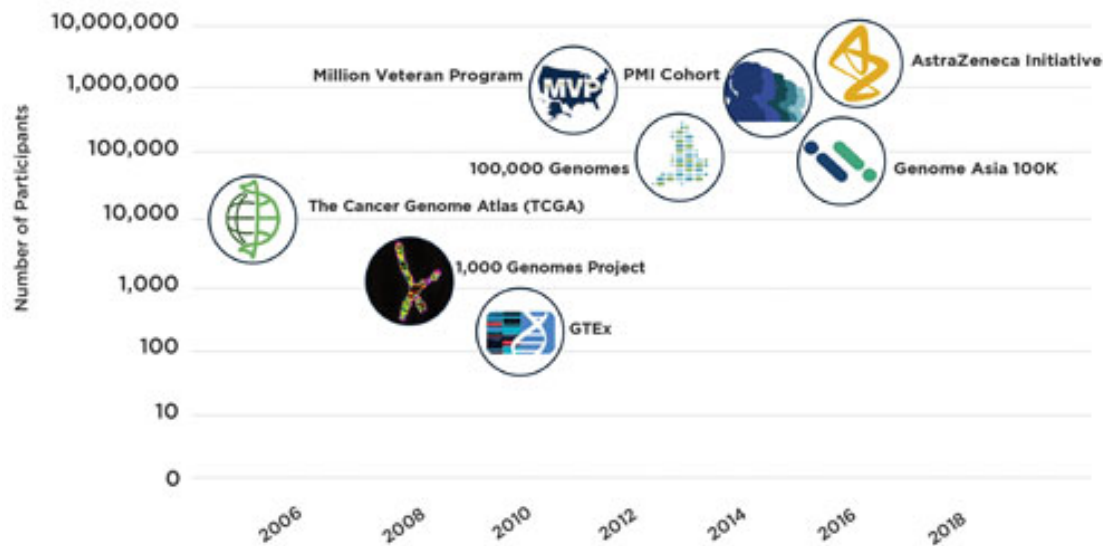
P4 Medicine

● PREDICT ● PREVENT ● PERSONALIZE ● PARTICIPATE



Million-genome era

- We are in the million-genome era
 - several projects now aiming to accrue over 1 million participants.
- By 2025, the annual acquisition of genomic data is anticipated to exceed 2 exabytes (2 million terabytes) and could be considerably higher.
- Getting the most from these data will require robust infrastructure and tools for large-scale analysis of multi-omic datasets.



What will happen in few years ???

- In 10 years a virtual cloud of billions of data points will surround each patient.
- These data will be of many different types and, accordingly, multistage.
- The challenge will be to convert these data into simple hypotheses about health and disease for the individual.

