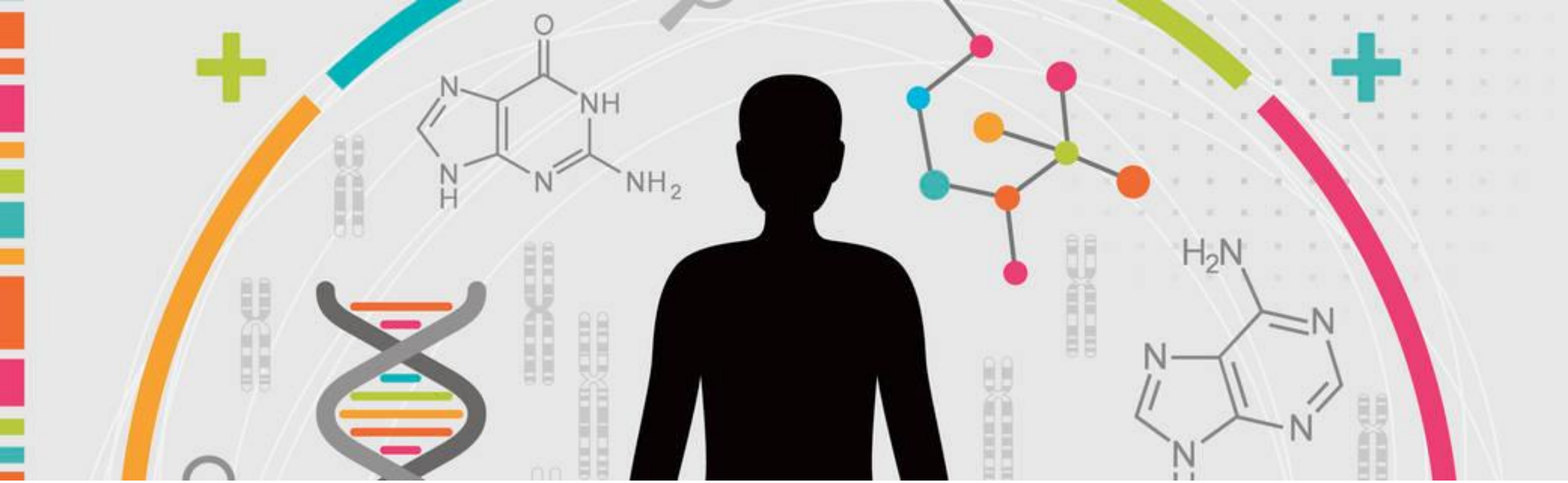


Precision Medicine & Artificial Intelligence

Agenda

1. Precision medicine
2. Artificial intelligence
3. Future scope

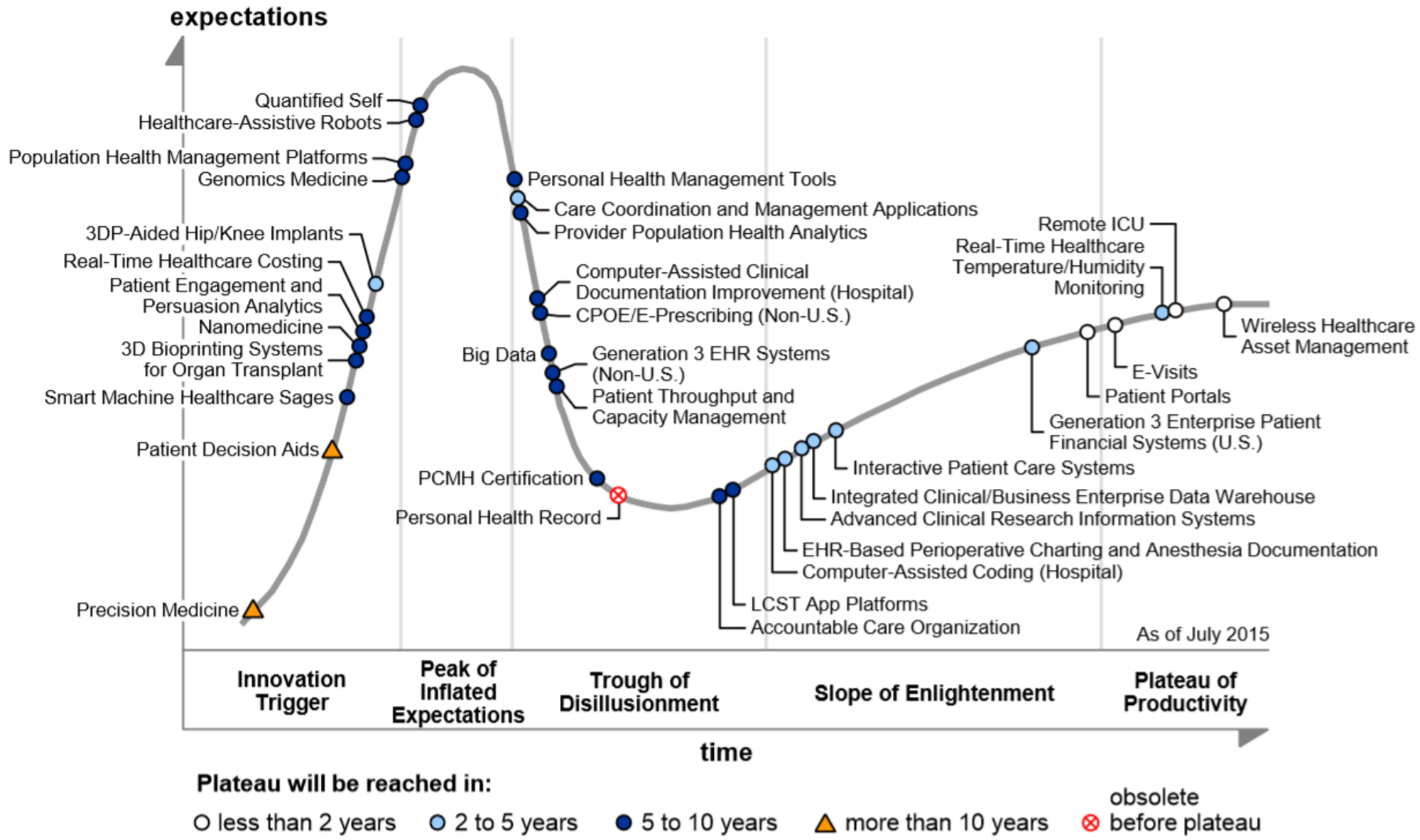
1. Precision medicine

As is : prediction of diseases

- **Examples of diseases of high population burden that currently do not have specific predictive biomarkers include**
  - **Alzheimer's disease**
  - **Type II diabetes mellitus.**

Healthcare Provider Applications, Analytics and Systems, 2015

Figure 1. Hype Cycle for Healthcare Provider Applications, Analytics and Systems, 2015



Source: Gartner (July 2015)


THE PRECISION MEDICINE INITIATIVE

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

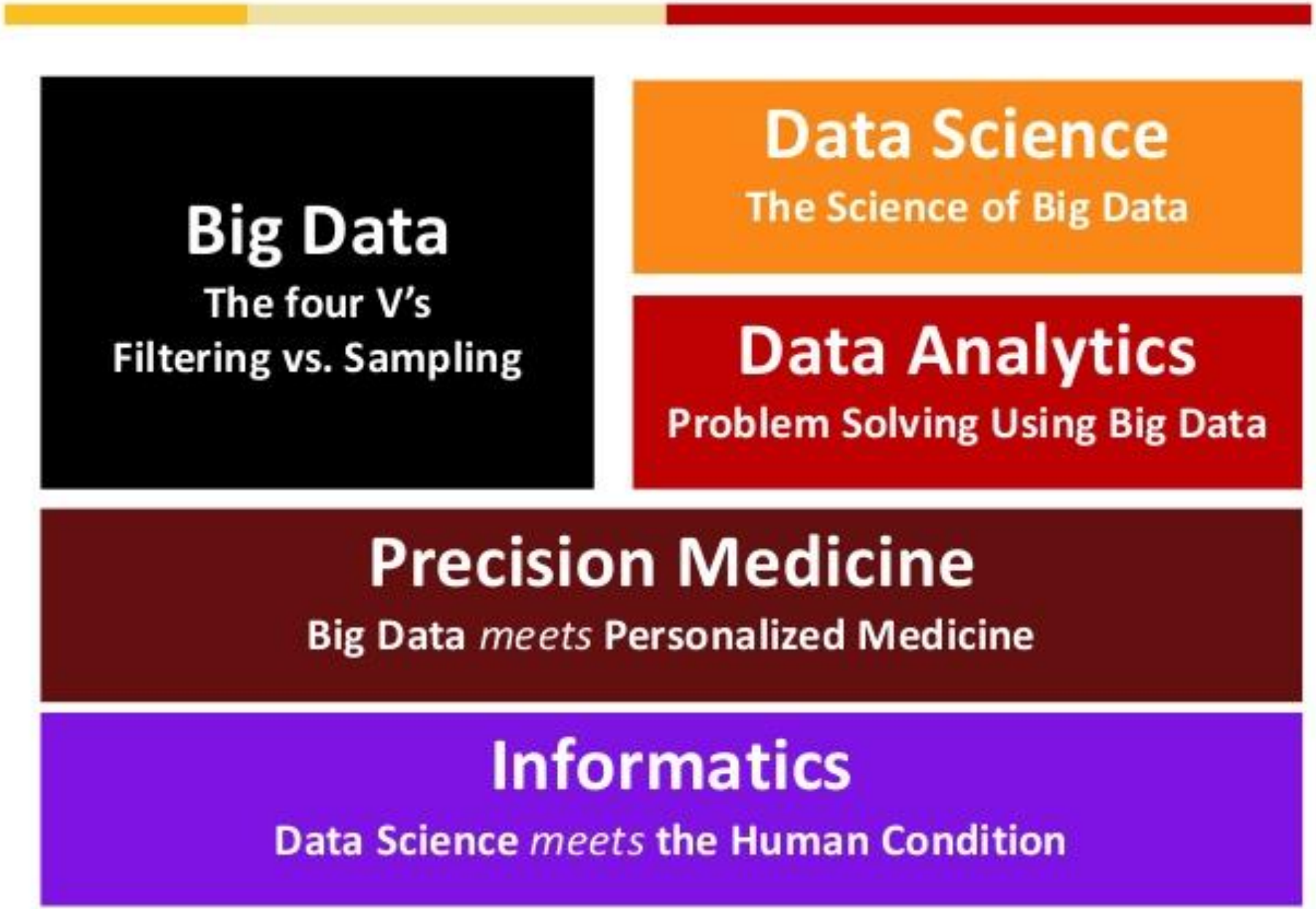
— President Barack Obama, State of the Union Address, January 20, 2015

Precision Medicine

- **an approach for more accurate diagnoses, more rational disease prevention strategies, better treatment selection, and the development of novel therapies.**

- **redefine understanding of disease onset and progression, treatment response, and health outcomes through the more precise measurement of molecular, environmental, and behavioral factors**

- **engages individuals as active partners – not just as patients or research subjects.**


Big Data
The four V's
Filtering vs. Sampling
Data Science
The Science of Big Data
Data Analytics
Problem Solving Using Big Data
Precision Medicine
Big Data meets Personalized Medicine
Informatics
Data Science meets the Human Condition

Examples of early success of PM

- **Development of targeted treatments for cancer and cystic fibrosis that are effective in patients who share an underlying causal genotype.**
- **The economic impact of investments in large-scale biomedical research has proven true in the past**
  - **$4 billion investment in the Human Genome Project has spurred an estimated $965 billion in economic growth – a 178-fold return on investment**

Personalized medicine vs. Precision medicine

- **Individual based,**
- **Genomics solutions,**

Big data + AI technique involved..

- **➔ Individual + Population based,**
- **➔ Genomics + Informatics solutions,**

Many of the ideas aren't new, but the "scale" and "scope" are.

GENOMICS

Our genes can suggest what diseases we *might* be predisposed to, but it's an incomplete picture of human health.

PHENOTYPE

A snapshot of the current state of health that can be used to prevent, diagnose and treat disease or improve health.

LIFESTYLE/ENVIRONMENT

External factors like diet, exercise, medications, microbiota and even where we live influence our metabolic state.

Why now?

- **Thanks to advances in**
  - **genomic technologies**
  - **data collection and storage**
  - **computational analysis**
  - **mobile health applications**
  - **→the creation of a large-scale precision medicine cohort is now possible in a way that it was not before.**
- **The PMI cohort is being launched at a time of explosive growth in the number, size, and complexity of potentially relevant data resources.**
  - **The "big data" of human biology (full genomes and high resolution digital images)**
  - **weather patterns, environmental monitoring, and streaming physiologic sensor data from study participants.**

A perfect convergence of technological leaps underlies the PMI:

- **Genome sequencing costs are down 10-million fold from 1998, with time collapsed from a decade to a day.**
- **Improved mass spectroscopy can analyze more metabolites that can serve as biomarkers, enabling us to track more diseases and our immune systems' responses.**
- **Electronic medical records store clinical data over time.**
- **Mobile health ("mHealth") apps and technologies enable anyone to track almost anything, Many of us already monitor our daily food intake, number of steps taken, blood pressure, heart rate, sleep patterns, even blood sugar levels.**
- **Social media is connecting patients, creating an incredibly science-savvy constellation of online communities.**

Dramatically reduced barriers

- **IT revolution has provided remarkable reductions in the cost of data storage, and comparable increases in analytic capabilities.**
- **Cost of DNA sequencing has been reduced nearly 10 million-fold from the time the sequencing phase of the Human Genome Project began in 1998, with complete human genomes now being routinely sequenced and analyzed for less than $2000 in several days.**
- **Wide implementation of EHR in 95% of U.S. hospitals**
  - **complete longitudinal health care records at extremely low cost.**
- **Personal mobile technologies have been adopted at an exponential rate**
  - **medical technologies are becoming mobile, home-based, and/or consumer-operated (mHealth)**
  - **data from sensors and software applications can enrich self-reported data on lifestyle and environment**

Table 1.1: Proposed PMI Budget Allocations for FY 2016 Department of Health and Human Services		
Investment	Agency	Purpose
$130 million	National Institutes of Health	To develop a voluntary national research cohort to propel our understanding of health and disease and set the foundation for a new way of doing research.
$70 million	NIH National Cancer Institute	To scale up efforts to identify genomic drivers in cancer and develop more effective approaches to cancer treatment.
$10 million	Food and Drug Administration	To acquire additional expertise and advance the development of high quality, curated databases to support the regulatory structure needed to advance innovation in precision medicine.
$5 million	Office of the National Coordinator	To support the development of interoperability standards and requirements that address privacy and enable secure exchange of data across systems.

Why cohort?

- **Prospective cohort studies have the ability to identify biomarkers and causative factors contributing to future disease.**
- **Example**
  - ➢ **Framingham Heart Study**
    - ✓ **smoking, high LDL cholesterol, and high blood pressure as major independent risk factors for heart attack and stroke**
      - → **leading to dramatic reductions in morbidity and mortality from these diseases by mitigation of these risk factors**

Table 3.1: Estimated number of prevalent diseases in a population of one million.
Each number represents the number of unique diseases expected at each threshold.

Participants	**Case count threshold:**			
	>2,500	**>5,000**	**>10,000†**	**>20,000‡**
Diseases among participants, any age	420	261	136	59
Diseases among participants, aged 40-69	413	267	151	68

†Threshold for study of genetics on particular phenotypes
‡Threshold for phenotype×gene×environment interaction (e.g., pharmacogenomics) studies.

Utility of the PMI cohort

- Quantitative estimates of risk for a range of diseases by integrating environmental exposures, genetic factors, and gene-environment interactions
- Determinants of safety and efficacy for commonly used therapeutics
- Biomarkers that identify individuals with an increased risk of developing common diseases
- mHealth technologies to correlate body measurements and environmental exposures with health outcomes.
- Cinical impact of loss-of-function mutations.
- New disease classifications and relationships.
- Empowering participants with data to improve their own health
- Platform to enable trials of targeted therapy.

Key six areas for development of PMI-CP

- **Cohort assembly**
- **Participant engagement**
- **Data**
- **Biobanking**
- **Policy**
- **Governance.**

Cohort Assembly.

- **2 distinct methods to recruit participants**
  - **Direct volunteers**
    - ✓**consent to be part of the PMI cohort**
    - ✓**agree to be recontacted**
    - ✓**undertake a PMI baseline health exam**
    - ✓**provide a biospecimen**
    - ✓**share available health data by either directing their EHR data to the PMI-CP and/or by undergoing an initial exam with a health care provider.**
  - **Collaborate with healthcare provider organizations (HPOs) to recruit participants**

Data Considerations

- **The Working Group recommends that the initial core data set acquired from all PMI cohort participants be collected and stored centrally (5.13, 5.23).**
  - **data from EHRs**
  - **health insurance organizations**
  - **participant surveys**
  - **mHealth technologies**
  - **biologic investigations**

Standards

- **Working Group recommends**
  - **Common data model (CDM) to organize data similarly across HPOs and from direct volunteers**

    (many data types useful from clinical investigation may not easily be transformed to an standard)
  - **Leverage existing data standards and CDM**

    (while standards do not exist for many emerging modalities such as sensor)
  - **Early selection of commonly used mHealth technologies to gain experience in use and integration of these new modalities.**

- **The best approach will balance**
  - **normalization of only the highest value data initially for all participants**
  - **followed by on-demand data curation of other data as driven by scientific demand.**

Hub-and-spoke model

- **Data access, data normalization, and participant engagement**
- **Coordinating Center providing a single point of contact for coordinating data, biospecimens, participant communication and engagement, and research studies**
- **Consider novel collaborations with not-for-profit and commercial organizations**
- **For data storage and access**
  - ➢ **pursue a hybrid data and analytics architecture**
    - ✓ **leverages centralized data storage of core data**
    - ✓ **federated access to additional data at the nodes across the network.**

Table 5.1: Categories, Sources, and Uses of Data				
Category	**Examples**	**Source(s)**	**Example Uses**	**Core/ Subgroup**
Individual demographics and contact information	Date and place of birth, sex and gender, detailed and multiple races/ethnicities (e.g., Asian of Indian descent, Asian of Chinese descent), name, mailing address, phone number, cell phone number, email address, marital status, educational status, occupation/income	Study participant, healthcare provider organizations	Participant-specific communications, analytics, risk stratification, assessment of covariates and confounds, study appointment reminders, invitations to participate in sub-studies	C
Terms of consent and personal preferences for participation in the project	Fine-grained consent for options to participate e.g., receive research results	Study participant	"Precision Participant Engagement"	C
Self-reported measures	Pain scales, disease-specific symptoms, functional capabilities, quality of life and well-being, gender identity, structured family health history	Study participant	Many	C/S
Behavioral and lifestyle measures	Diet, physical activity, alternative therapies, smoking, alcohol, assessment of known risk factors (e.g., guns, Illicit drug use)	Study participant (retrospective and prospective) and healthcare provider organizations	Correlation with clinical events, drug response, and health outcomes	C/S
Sensor-based observations through phones, wearables, home-based devices	Location, activity monitors, cardiac rate and rhythm monitoring, respiratory rate	Smartphone sensors, commercial and research-grade physiologic monitors	Functional ability and impairment assessment	C/S
Structured clinical data derived from Electronic Health Records (EHRs)	ICD/CPT billing codes, clinical lab values, medications, problem lists	Multiple provider organizations per study participant, via institutionally managed channels or direct from	Correlation of clinical events with other categories of data	C

		participant via personal download/upload		
Unstructured and specialized types of clinical data derived from EHRs	Narrative documents, images, EKG and EEG waveform data	Multiple providers, via federated queries rather than inclusion in core dataset	Correlation of clinical events with other categories of data	S
PMI baseline health exam	Vital signs, medication assessment, past medical history	Study participant interacting with healthcare provider organization	Provides baseline measures on all participants	C
Healthcare claims data	Periods of coverage, charges and associated billing codes as received by public and private payers, outpatient pharmacy dispensing (product, dose, amount)	CMS and other federal sources, private insurers, pharmacy benefits management organizations	Assessments requiring complete longitudinal record of exposures/outcomes during specific periods, e.g., within X years of a diagnosis or medication exposure; health services research, exposure and outcomes assessment	C
Research specific observations	Research questionnaires, ecological momentary assessments, performance measures (six minute walk test), disease specific monitors (e.g. glucometers, spirometers)	Study participants, research organizations	Many	S
Biospecimen-derived laboratory data	Genomics, proteomics, metabolites, cell-free DNA, single cell studies, infectious exposures, standard clinical chemistries, histopathology	Study participants, provider organizations, outsourced laboratories	Correlation of tissue findings and high throughput biomolecular data with other categories of data	C
Geospatial and environmental data	Weather, air quality, environmental pollutant levels, food deserts, walkability, population density, climate change	Public and private sources not directly part of PMI	Epidemiology, epidemic surveillance	C/S
Other data	Social networking e.g., Twitter feeds, social contacts from cell phone text and voice, OTC medication purchases	Public and private sources not directly part of PMI	Predictive analytics	S

initial PMI core data set

- A set of standard self-report measures via direct patient assessment (all participants).
- A brief, standardized baseline health exam at enrollment for all PMI cohort participants.
- Structured clinical data
  - Data expected from HPOs
    - ✓ a. All ICD codes with dates.
    - ✓ b. All CPT codes with dates.
    - ✓ c. Select, high-value clinical laboratory results in a structured form
    - ✓ d. All available lifetime medication data
    - ✓ E. Vital measurements, including all weights, heights, heart rate, blood pressure, and pain score values
    - ✓ A record of all encounters (e.g., dates of clinic visits, inpatient visits, ER visits)
    - ✓ health plan data
  - Clinical Data from Direct Volunteers:
    - ✓ they could transmit to the PMI cohort through data transfer protocols
    - ✓ Current Blue Button functionality at many locations (e.g., from CMS) provides for ICD, CPT, some problem list data, and/or medication data. These data would be aggregated at the Coordinating Center.
- Biospecimen-derived data (all participants).
- mHealth data (many participants)

Biobanking

- **Each PMI cohort participant should provide a new blood specimen**

- **Samples collected for the PMI cohort will be sent to a central biorepository**

Policy considerations

- **The success and longevity of the PMI-CP will be heavily influenced by the laws, regulations, and policies surrounding research, data security and privacy, and access and interoperability of EHRs**
- **An internal framework of PMI-CP policies will need to be developed to address**
  - **participant inclusion**
  - **IRB review and consent**
  - **privacy**
  - **misuse of information, and security**
  - **sharing of data and specimens with researchers**
  - **sharing of data and research results with participants**
- **the use of a single IRB to reduce administrative burden and associated costs of the cohort, review time, and to harmonize inconsistent or conflicting policies between PMI cohort nodes**

Governance

- **Governance structures for the PMI-CP should be coordinated with other federal agencies**
  - **Centers for Medicare & Medicaid Services**
  - **Health Resources and Services Administration**
  - **FDA**
  - **ONC (Office of the National Coordinator for Health Information Technology)**
  - **Department of Veterans Affairs (including the Million Veteran Program)**
  - **Department of Defense**

mHealth devices

- **Smartphone**
- **Select wireless sensors**
  - **research grade or commercial grade wrist-worn accelerometers**
  - **wireless weight scales**
  - **movement sensors in the home**
  - **continuous heart rate and pulse oxygen monitors**
  - **respiration monitors**
  - **glucometers**
  - **spirometers**
  - **other FDA approved wireless medical monitoring devices used in the home**

Too Much Information: The Doctor's Data Dilemma

By 2020, doctors will face

200x

the amount of medical data and facts that a human could possibly process.[1]

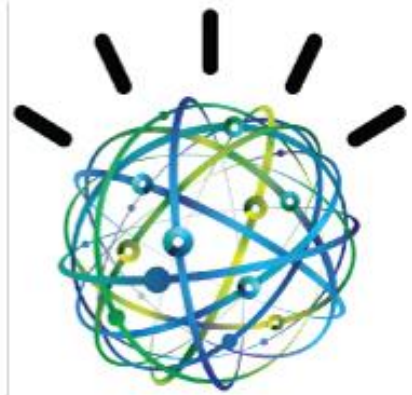
And it will get worse...
The volume of medical data **doubles** every five years.[2]

81% of physicians can't even spare **5 hours per month** to keep up.[3]

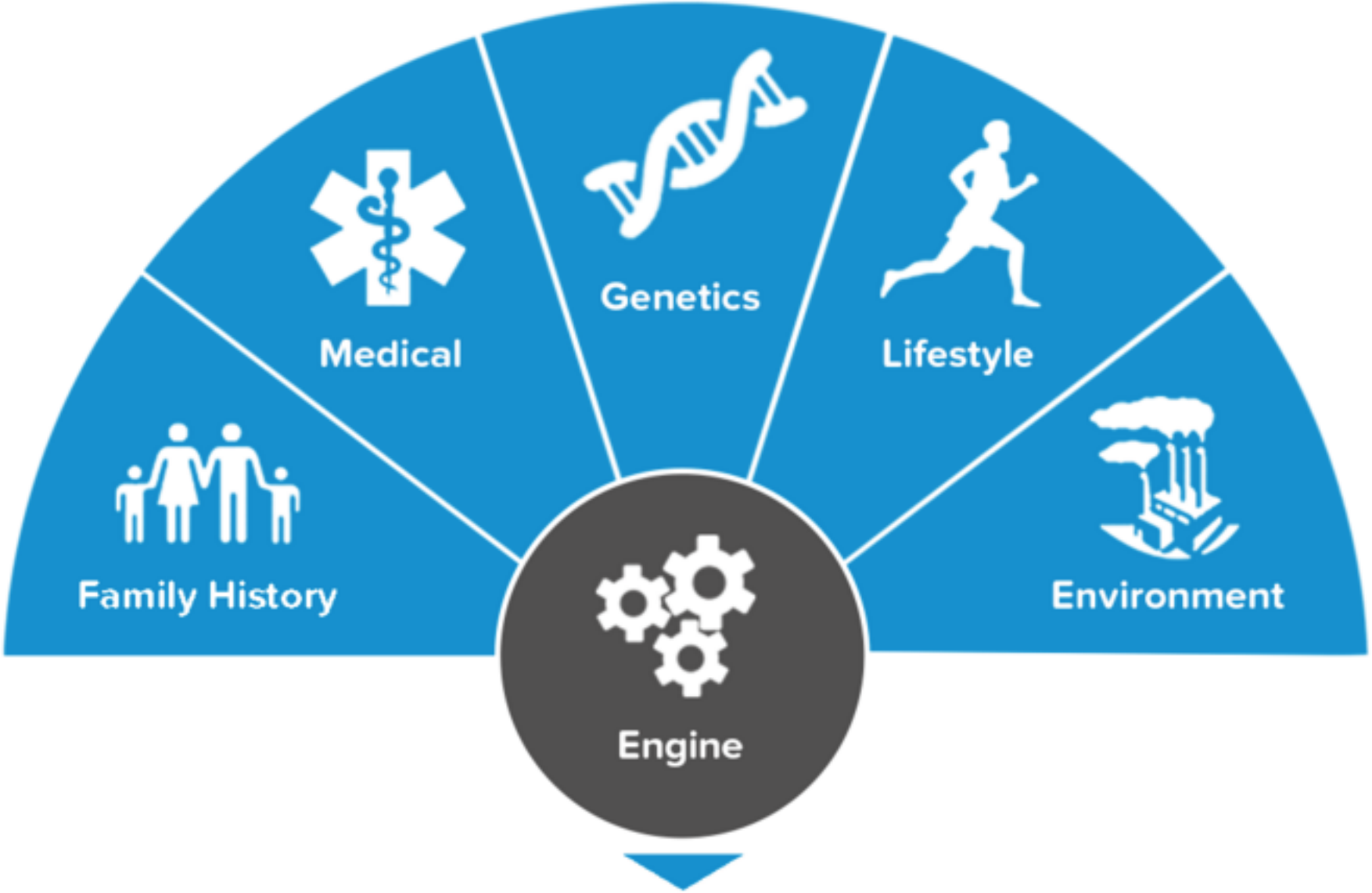
The answer? IBM Research and the Cleveland Clinic are bringing IBM® Watson™ to medical school to create a learning application for students.

Watson will help students navigate medical information and make the best decisions for **improving patient care.**

Students will also be able to **teach and train Watson** to advance its knowledge.



WATSON



Bigdata analysis + Cloud engine

Key design principle for building big data analytic systems

- **Data systems with an expectation that future growth that may be dramatic.**

- **Data systems' functional requirements should incorporate at least a small number of extreme use cases of large data sources.**
  - **A useful method to doing so is to implement one or two key use cases in the initial phase of the PMI-CP**

Cloud computing

- require highly elastic data storage and computing resources
- “tools go to the big data” rather than “data being downloaded to the tools”
- e.g., National Cancer Institutional Genomics Cloud Pilot projects125)
- state-of-the-art cloud computing environments may involve novel public-private and academic-commercial partnerships.

Coordinating Center

- **Centralized resources for data acquisition and management**

Fig. 5.1: Organizational relationships

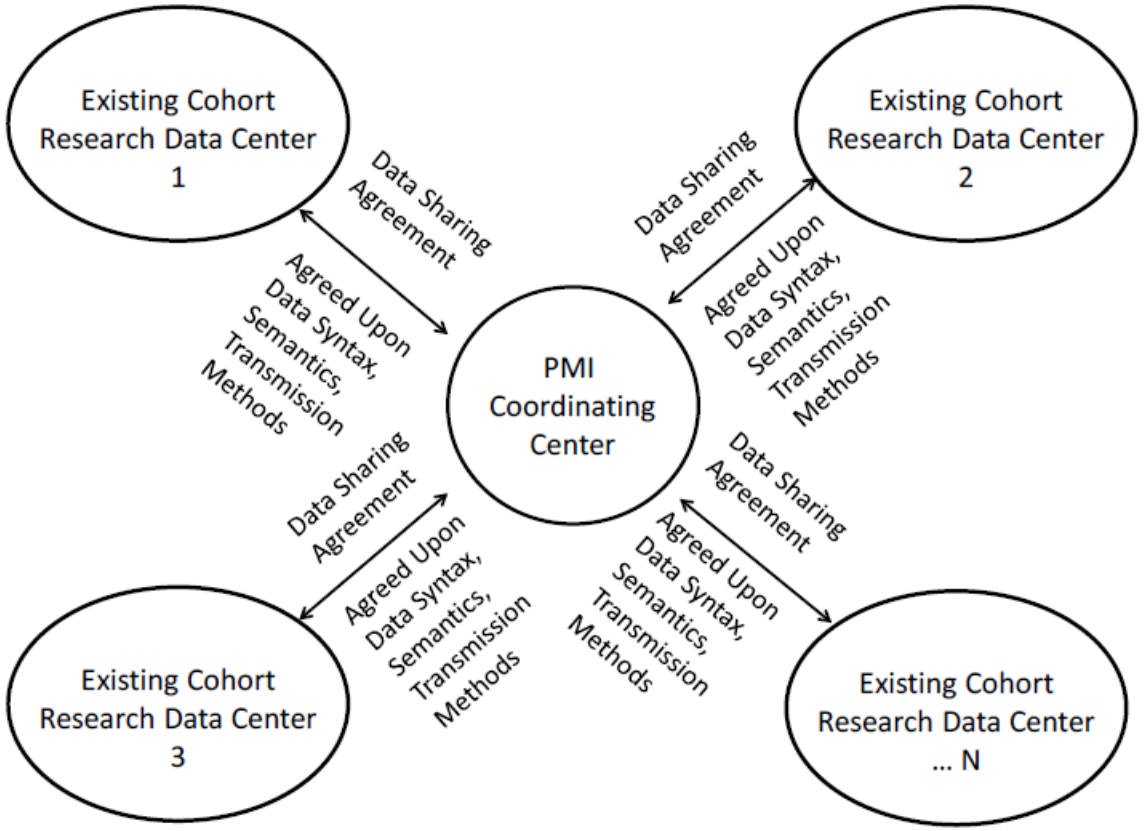


Table 3.5: Select existing biobanks with healthcare provider data. Participants in all biobanks listed below are recontactable.				
Biobank	**HPO system size**	**Current Biobank Size**	**Recruitment method**	**Time to achieve size (during active enrollment)**
Million Veteran Project	6 million	400,000	Mailed veterans info about MVP and enrolled at visit	4 years in 54 sites
Kaiser Permanente	10.1 million	245,000 (goal of 500,000)	Mailed consent and mailed saliva sample (N = 189,500); electronic or in-person consent and blood samples (N= 50,000)	3.5 years using direct mail to 2 million
Partners Healthcare Biobank	6 million	>30,000	In-person at outpatient visits and inpatient floors; Electronic consent via emails using patient portal	5 years since launch: 2 year pilot study; 3 years via in person recruitment; eConsent for past 1 year; current rate is 1100/month
Geisinger MyCode	1.3 million with an EHR	>86,000	In-person during routine outpatient	10 years; however, current rate is 1000/

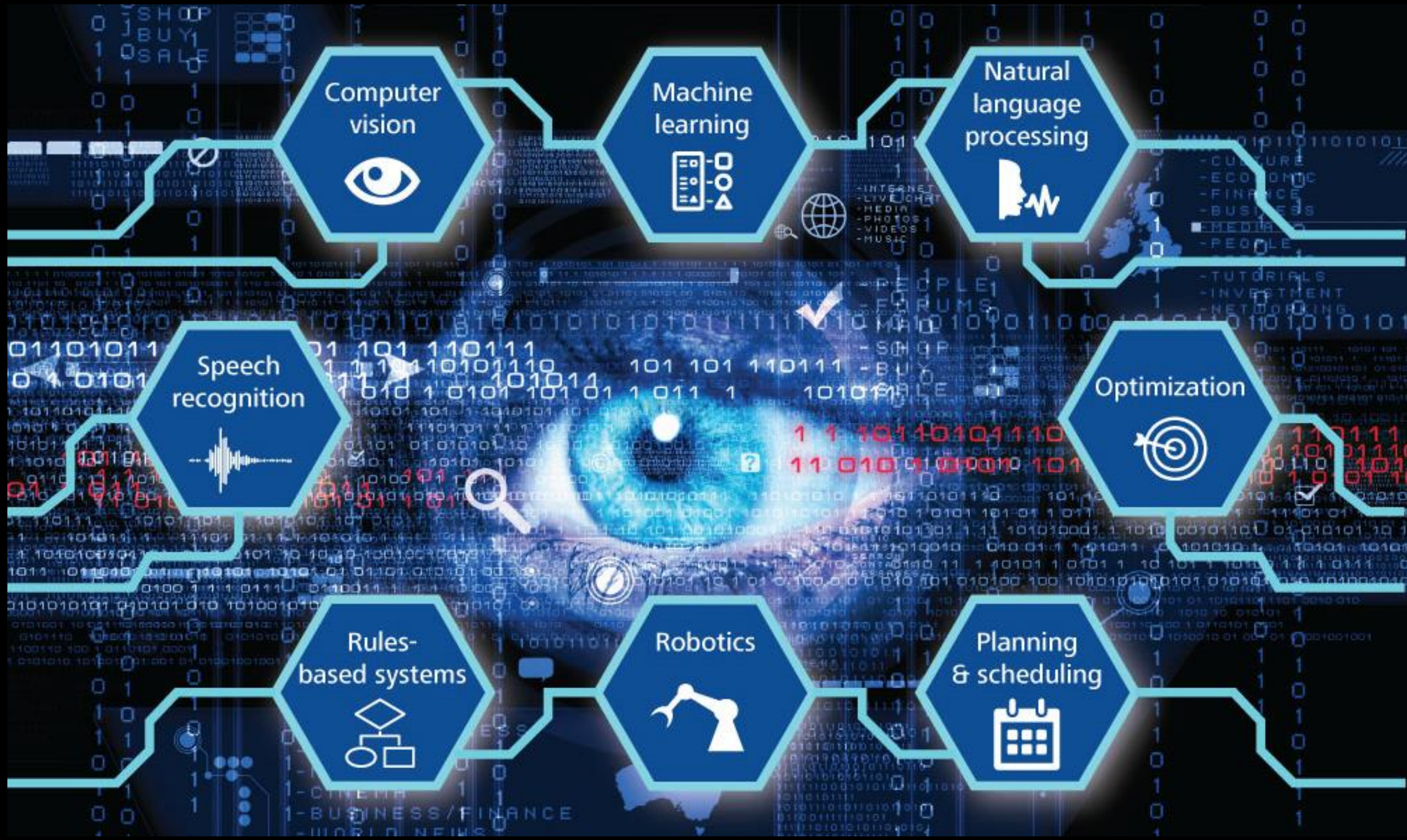
	encounter in last 10 years		visit; Electronic consenting pending	week
Marshfield Clinic Personalized Medicine Research Program	>2 million	20,000	In-person, recruited via phone and mailers	16 months at 4 sites of Marshfield clinic
Mayo Clinic	2 million	>60,000	In-person consent at clinic	7 years, current rate about 8000-9000/year
Children's Hospital of Philadelphia	2.5 million	110,000	In-person consent at clinics	9 years
Cincinnati Children's Hospital Medical Center	670 thousand	>56,000	Hospital-wide consent by registrars at registration	4 years

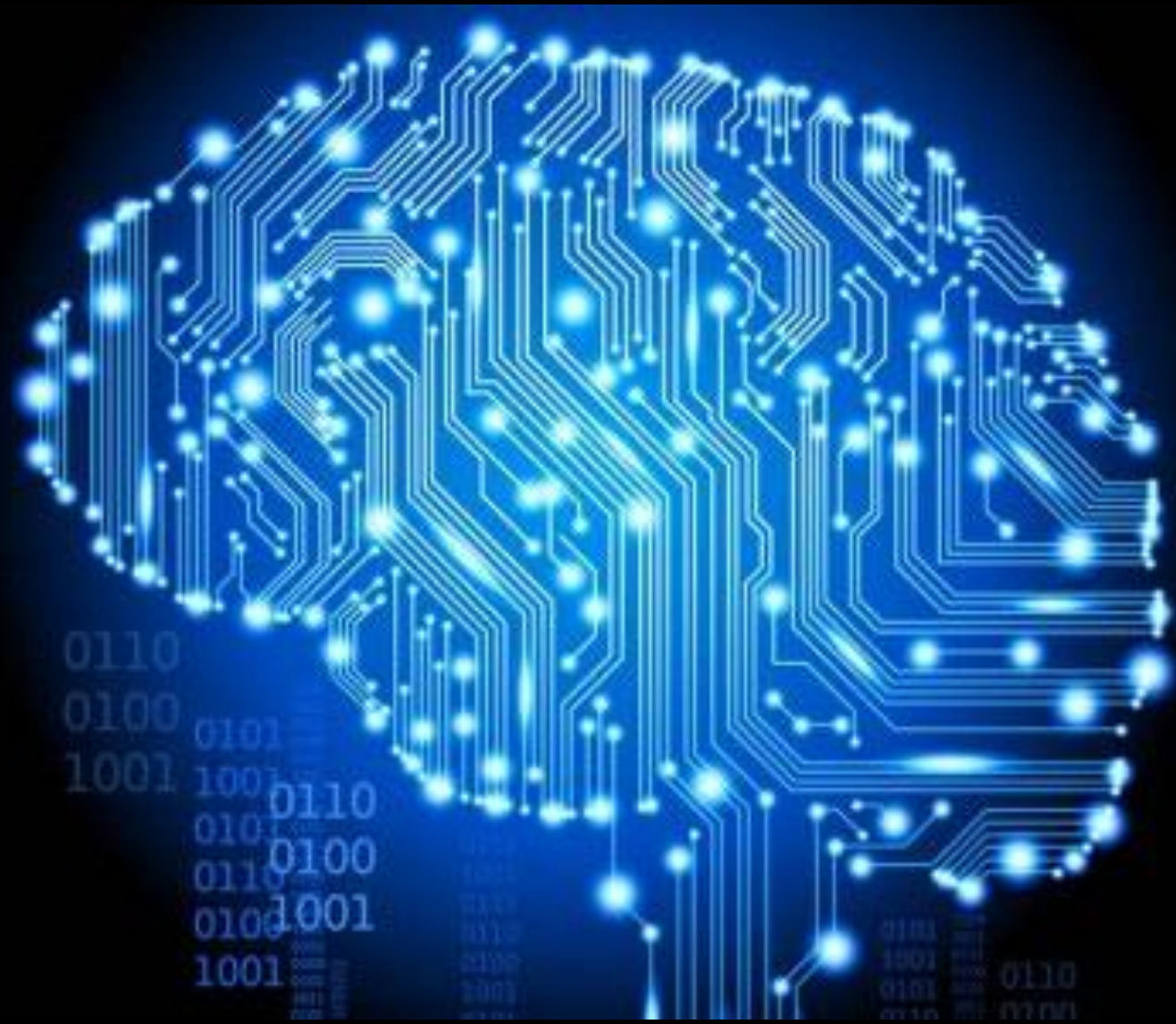
Table 5.3: Types of Access. RAS=Resource Access Subcommittee					
	Open Access (no login required)	Anyone with a login (no application necessary)	After approval of brief project review	With RAS and IRB review	With RAS and IRB review and additional participant consent
Newsletters, ongoing PMI studies and general updates	X				
Aggregate counts of individuals	X				
Graphical query to assess study feasibility using counts		X			
Query interface exact counts of rare events			X		
Access to de-identified individual-level data				X	
Access to identified data				X	
Recontact individuals				X	X
Clinical trials					X

Workshops

- **On April 28-29, 2015, *Unique Scientific Opportunities for the National Research Cohort*, at the NIH in Bethesda, Maryland**
  - **use cases**
- **On May 28-29, 2015, *Digital Health Data in a Million-Person Precision Medicine Initiative Cohort* Workshop at Vanderbilt University in Nashville, Tennessee**
  - **inclusion of specific participant groups**
  - **most appropriate model for PMI cohort data aggregation and sharing.**
- **On July 1-2, 2015, *Participant Engagement and Health Equity* Workshop at the NIH in Bethesda, Maryland**
- **On July 27-28, 2015, *Mobile and Personal Technologies in Precision Medicine*, at the Intel Corporation in Santa Clara, California**
  - **incorporation of sensor and mobile technology data, their data standards**

2. Artificial intelligence


Computer vision
Machine learning
Natural language processing
Speech recognition
Optimization
Rules-based systems
Robotics
Planning & scheduling


0110
0100
1001
0101
1001
0110
0100
1001
0110
0100
1001


10 IBM
WATSON-
POWERED
Apps That Are
Changing Our World

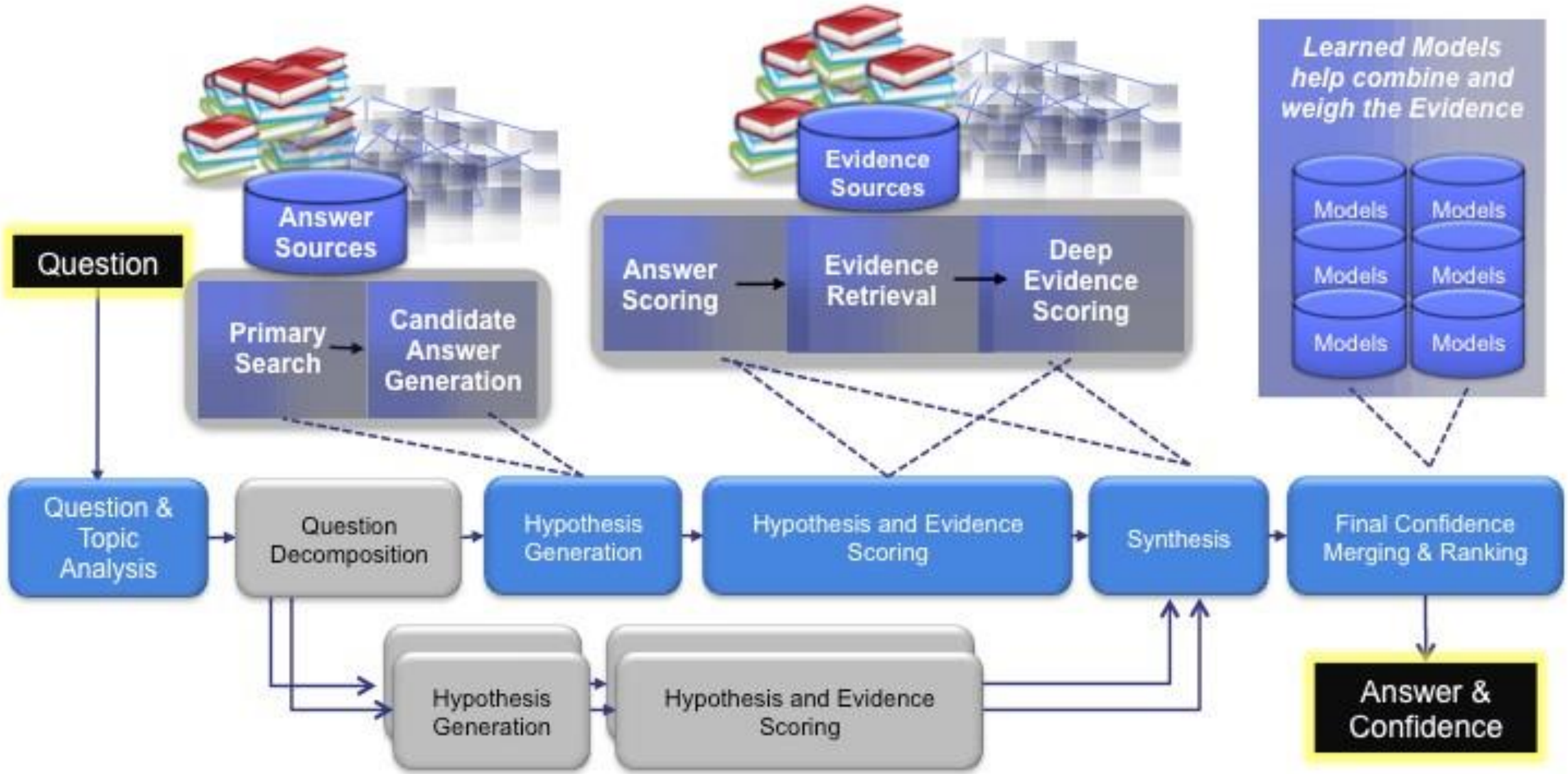
세계적 동향

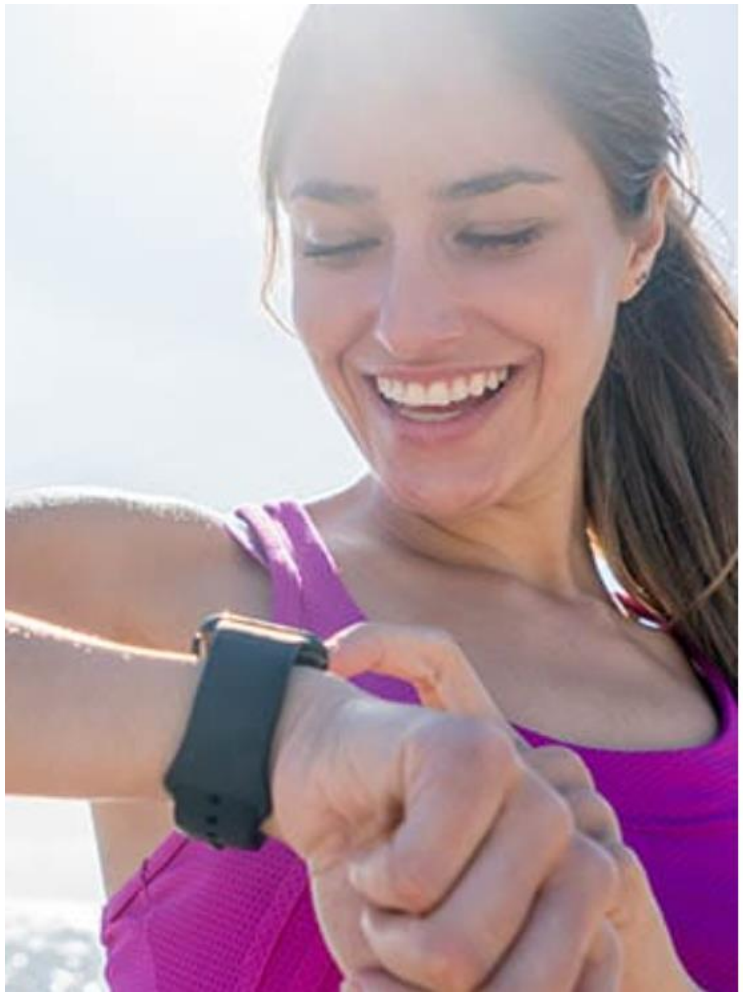
- 구글은 딥러닝의 대가인 제프리 힌튼 교수를 영입하고, 딥러닝 전문 회사인 딥마인드 및 사진 인식 번역 기술을 보유한 워드 렌즈를 인수
- IBM은 B2B 기업 컨설팅 지능 서비스 제공을 위하여 자사의 왓슨(Watson) 시스템을 강화하는데 주력
- 페이스북 딥러닝을 통한 얼굴인식 프로그램 딥 페이스 등 많은 연구를 진행
- 마이크로소프트는 음성 인식을 활용한 지능형 비서 코타나, 스카이프에서 활용 가능한 동시 통역 기술 개발

Cognitive computing in Healthcare

- IBM Watson Health: innovations in care management
- How It Works: IBM Watson Health

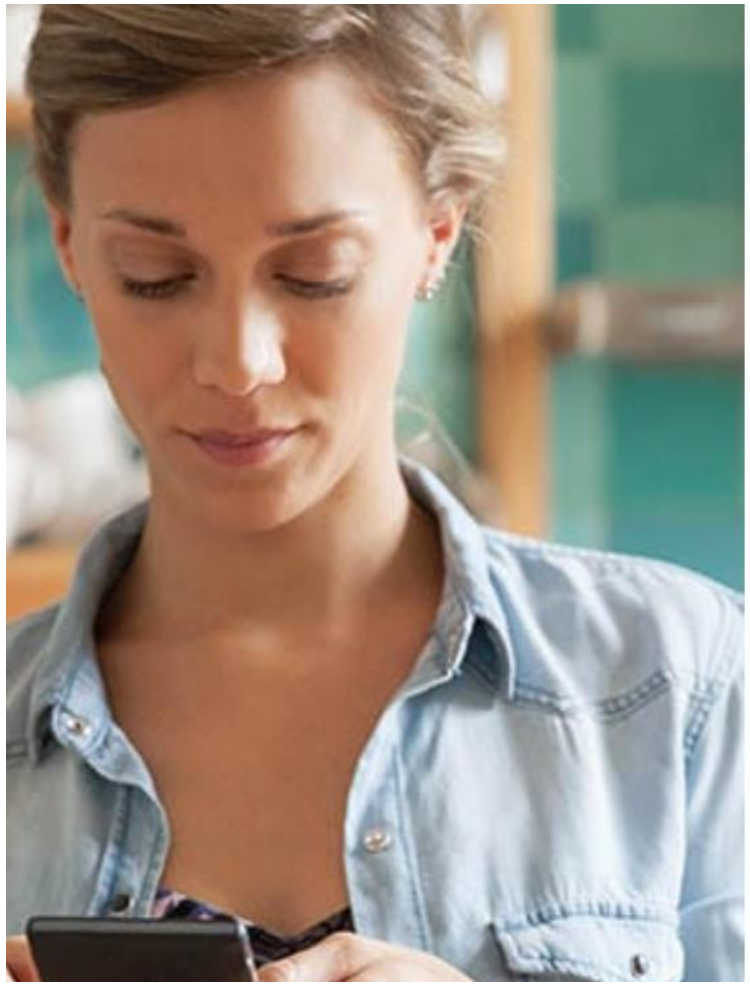
- **Memorial Sloan Kettering collaborates with IBM to teach Watson in the Fight Against Cancer.**


Learned Models help combine and weigh the Evidence
Models
Models
Models
Models
Models
Models
Evidence Sources
Answer Sources
Question
Answer Scoring
Evidence Retrieval
Deep Evidence Scoring
Primary Search
Candidate Answer Generation
Question & Topic Analysis
Question Decomposition
Hypothesis Generation
Hypothesis and Evidence Scoring
Synthesis
Final Confidence Merging & Ranking
Hypothesis Generation
Hypothesis and Evidence Scoring
Answer & Confidence



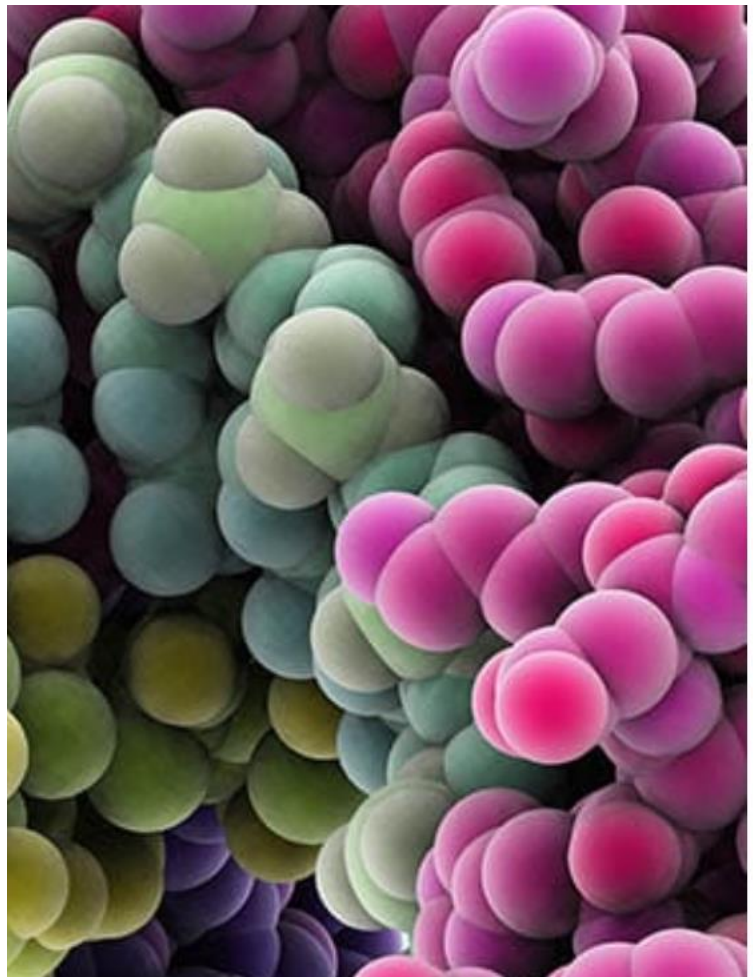
Under Armour®

UA Record is an IBM Watson Health–powered dashboard for your body. Log any workout, track the data and then follow personalized health plans to help break through to the next level of fitness.



Medtronic

Medtronic's insulin pumps, combined with Watson, will help predict dangerous spikes or drops in blood sugar hours in advance and then notify people living with diabetes so they can take action before it's too late.



Baylor College of Medicine

Watson’s voracious reading habit helped researchers analyze 70,000 articles and narrow their focus to six potential proteins that could help in their efforts to win the battle against cancer. Best of all, the potentially life-changing research took weeks instead of years.

Mayo Clinic

There are almost 200,000 open clinical trials worldwide and finding the right patients is crucial to their success. Watson helps researchers cross-reference patient data and match patients to open trials.

Memorial Sloan Kettering Cancer Center

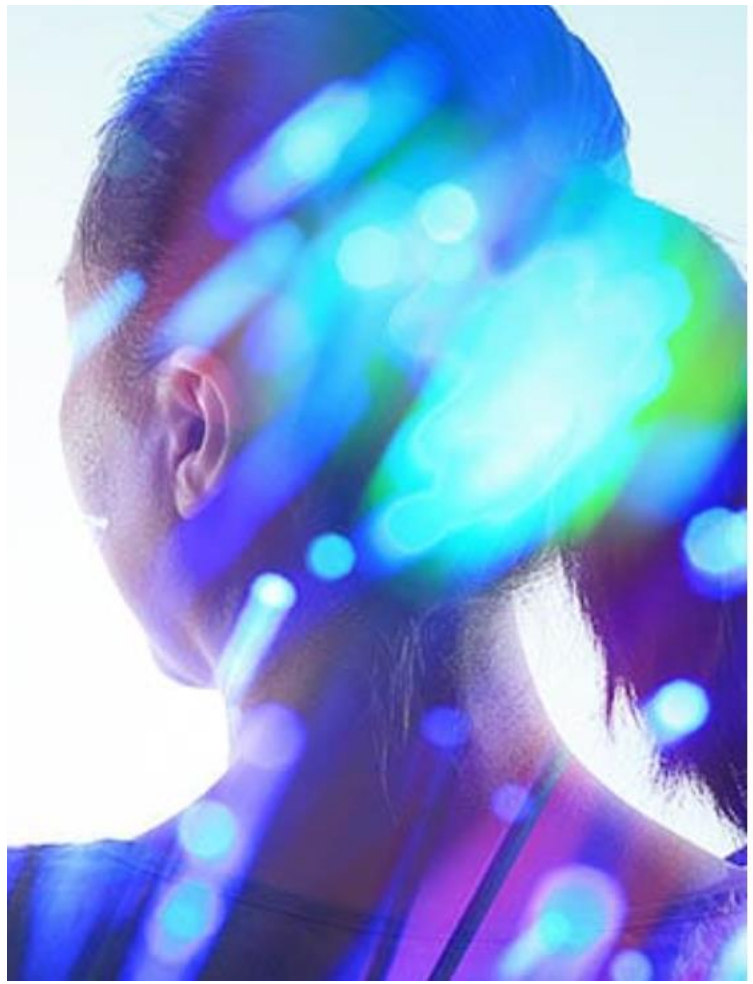
Watson, working in tandem with oncologists from Memorial Sloan Kettering Cancer Center, will help design evidence-based treatment regimens for cancer patients. Since every case of cancer is different, every treatment plan should be too.

iDAvatars

Sophie, iDAvatars' virtual medical assistant, uses Watson natural language processing and sentiment analysis to hear a patient's tone of voice and recognize things that may be troubling them. Since Sophie is cognitive, she gets more familiar with patients every time she "talks" to them. It's virtual support with real results.

Talkspace

Using Watson's Personality Insights API, the online therapy start-up uses a patient's text messages and anonymous chats to match them with a therapist who "gets" them. Online therapy is proven to help remove stigmas attached to seeking counseling—at an 80% lower cost.



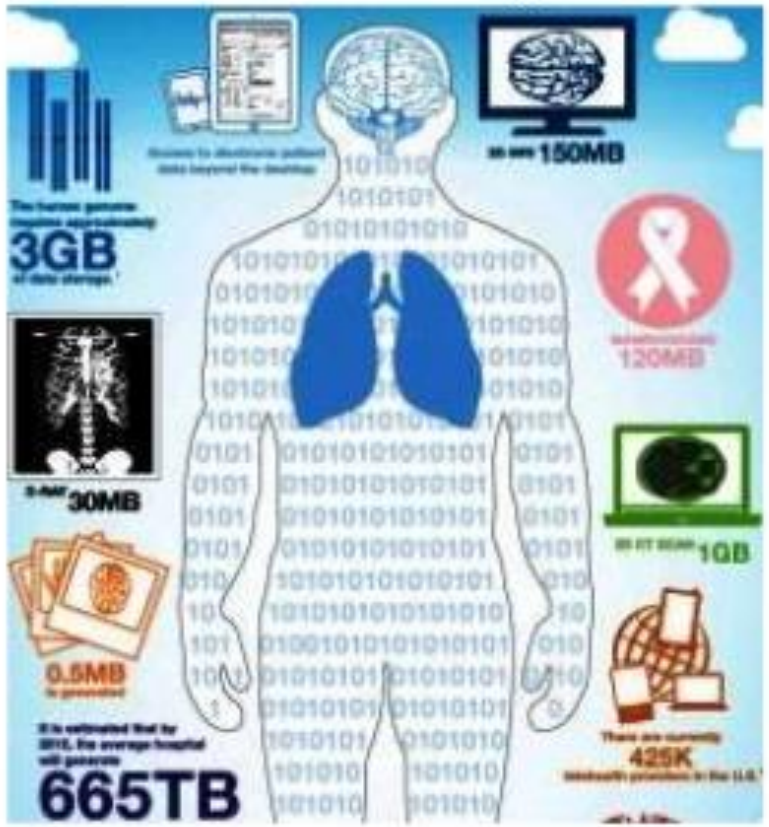
Arch Health Partners

Only 47% of patients with hypertension have it under control, according to the CDC. Arch Health Partners used our population health management platform to develop an eight-week pilot program that automatically detected registrants who were high-risk patients and funneled them towards treatment. As a result, the percentage of patients whose hypertension was under control grew by 14%.

3. Future scope

Why PHR and AI?

Healthcare Big Data



Machine Learning

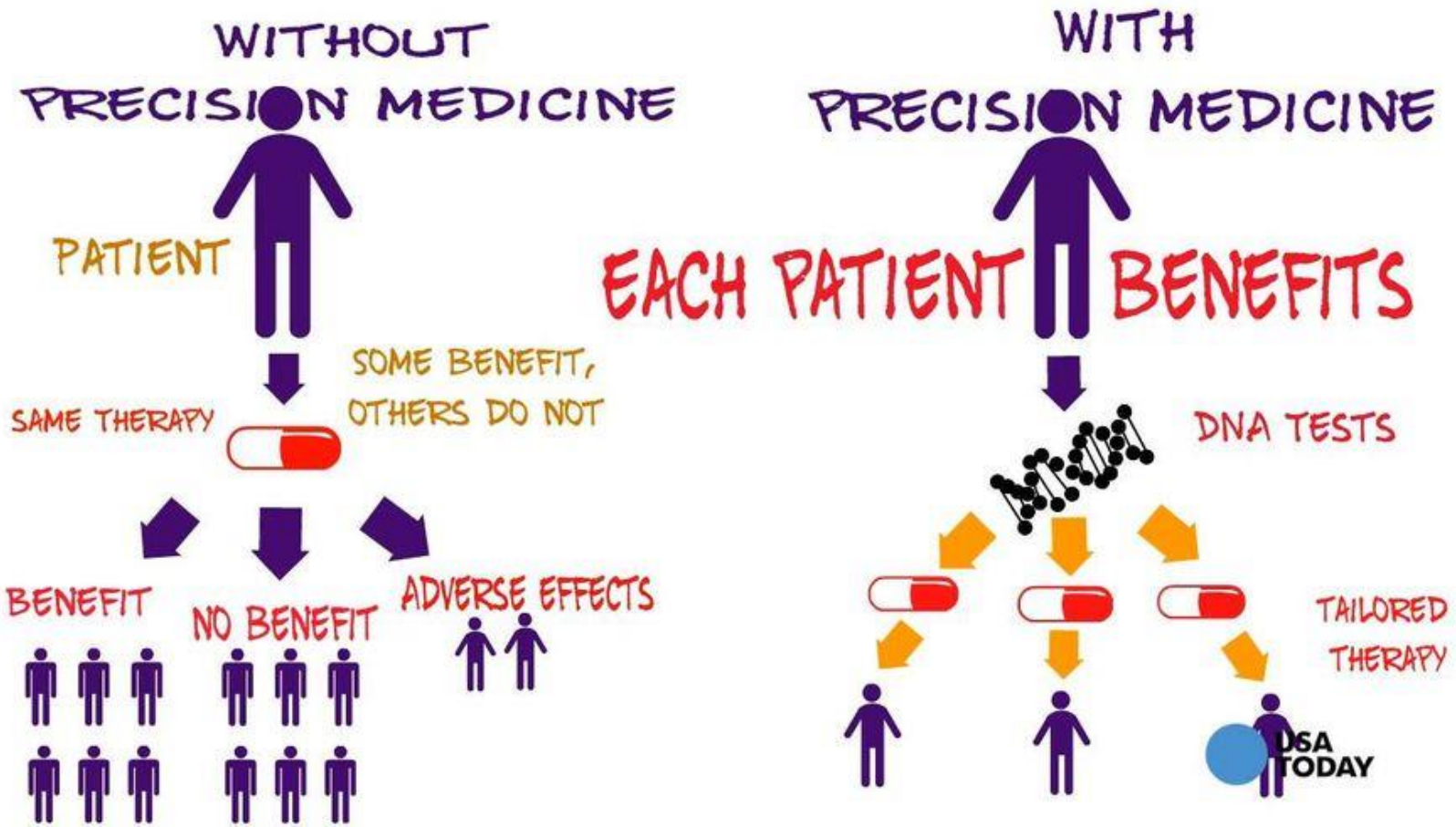
Novel Insights and Applications

Medical AI era: Pros and Cons

- 의료진 역할의 변화
  - 전체 의료진 수가 줄어들지는 않겠으나,
  - 단순한 역할은 AI가 대체하지만 AI를 조정하는 의료진 증가
  - 미래 의사는 환자를 진료하는 수준을 넘어 의료서비스 수준 고도화를 수행하고, 양질의 데이터를 생산·운영하는 역할 수행

- 오류(버그)에 대한 대책 필요
  - 의료 AI는 바둑이나 체스 게임과 같이 승패에 관여하지 않고,
  - 환자와 의료진의 “생명”을 다루는 현장에서 발생하기 때문에 오류에 대한 대책이 명확하게 필요함

The first trial target: *Cancer treatment domain*



Future medicine

Box 7.1: The legal, regulatory, and policy landscape affecting PMI-CP

As a publicly funded research initiative involving human research participants, the PMI-CP is subject to a number of laws, regulations, and policies, generally including:

- Healthcare Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule
- Health Information Technology for Economic and Clinical Health (HITECH) Act
- The Privacy Act of 1974
- Clinical Laboratory Improvement Amendments (CLIA)
- HHS regulations for the meaningful use of electronic health record technology
- FDA framework for regulation of genomic technologies
- FDA regulations for medical devices (including *in vitro* diagnostics and mobile health technologies)
- Federal Policy for Protection of Human Research Subjects (the Common Rule)
- NIH Grants Policy Statement (http://grants.nih.gov/grants/policy/nihgps/nihgps.pdf)
- NIH data sharing policies:
    - NIH Data Sharing Policy
    - NIH Genomic Data Sharing Policy
    - (Draft) NIH Policy on Dissemination of NIH-Funded Clinical Trial Information
- State laws as applicable

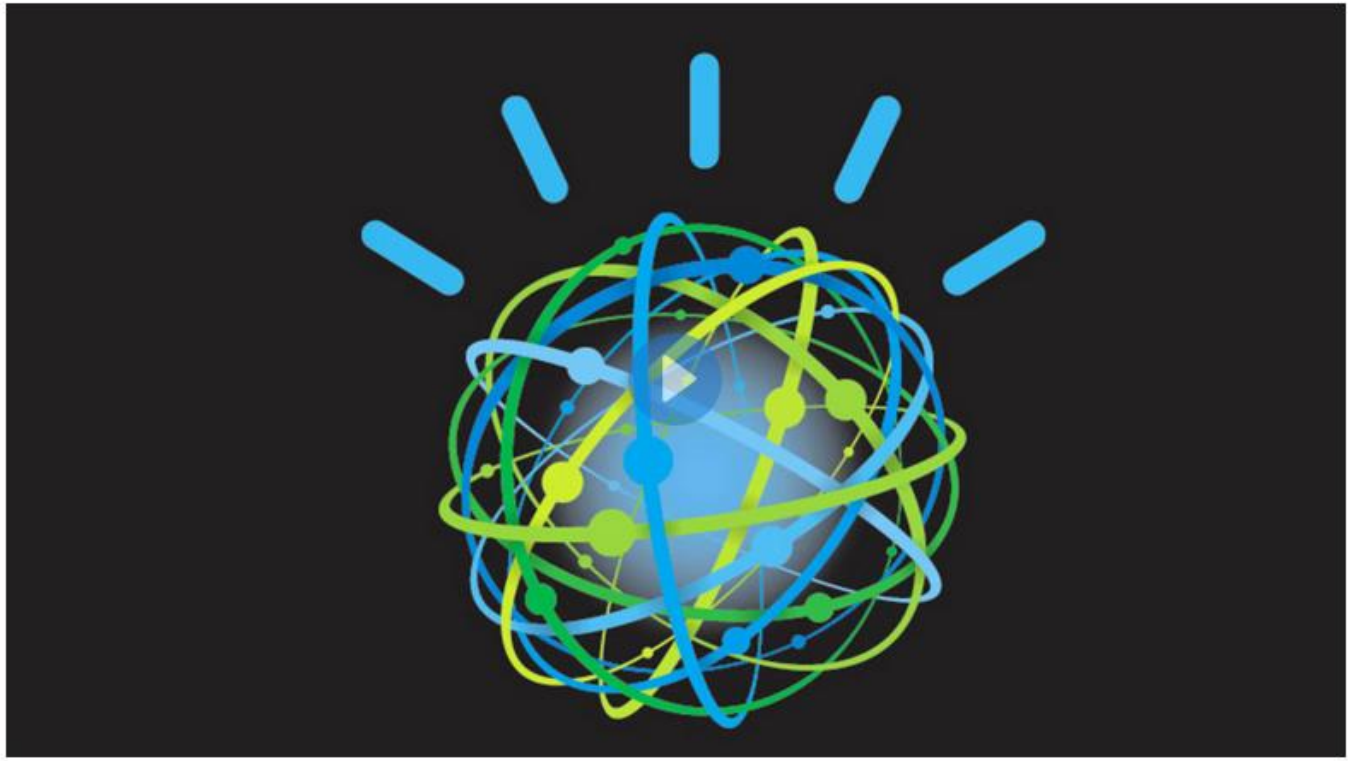
Table 2.1: Timeline when expected PMI cohort capabilities will be realized. The estimated timeline for focused research for each type of investigation is indicated by the number of "+" characters in each cell.

		Time in years			
		0-2	3-5	5-10	>10
Cohort Capabilities	1. Discovery of disease risk factors	+	+++	+++	++++
	2. Pharmacogenomics	+	+++	+++	+++
	3. Discovery of disease biomarkers	+	++	+++	+++
	4. mHealth connections with disease outcomes		+	++	++++
	5. Impact of loss-of-function mutations		+	+++	+++
	6. New classifications of diseases		+	+++	++++
	7. Empowering participants	+++	+++	+++	+++
	8. Clinical trials of targeted therapies		+	+++	+++

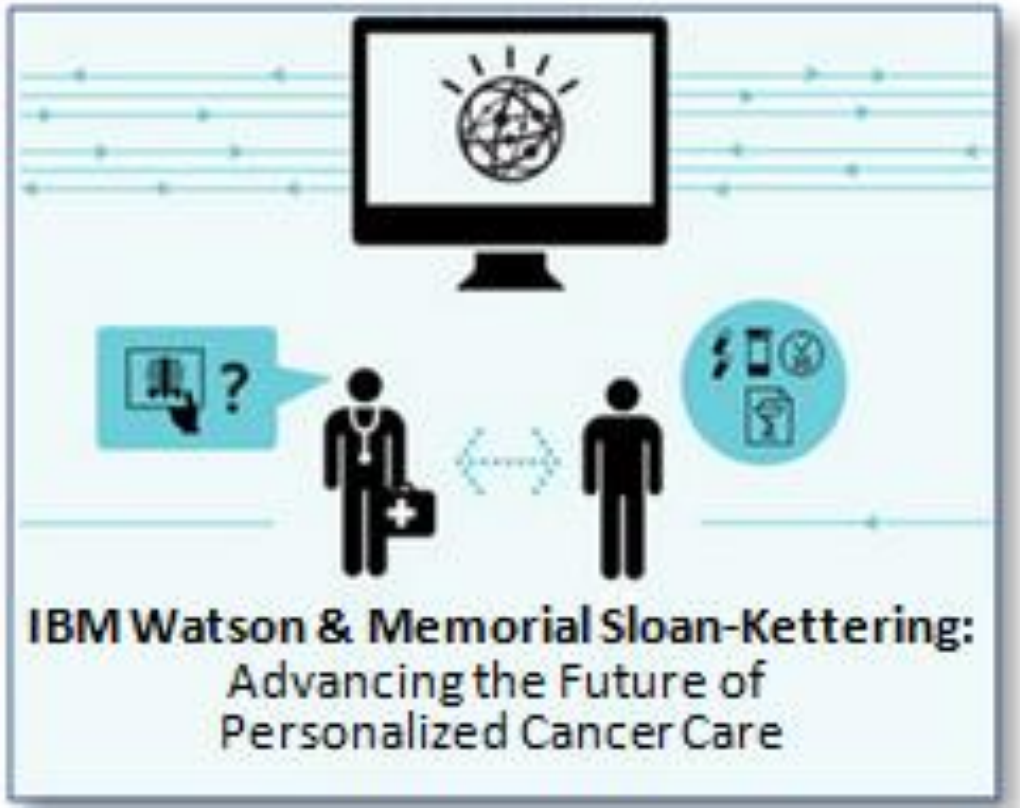
On Cancer

Memorial Sloan Kettering Trains IBM Watson to Help Doctors Make Better Cancer Treatment Choices

on **Friday, April 11, 2014**



A team of physicians and analysts at Memorial Sloan Kettering has been "training" IBM Watson for more than a year to develop a tool that can help medical professionals choose the best treatment plans for individual cancer patients.



Demonstration of Watson Cancer Care Solution

IBM Watson Oncology Advisor

Provide clues for creation, not creation itself

감사합니다