



Introducción al *big data* biomédico

Grupo de Neurocomputación Biológica
Escuela Politécnica Superior
Universidad Autónoma de Madrid

<http://www.ii.uam.es/~gnb>

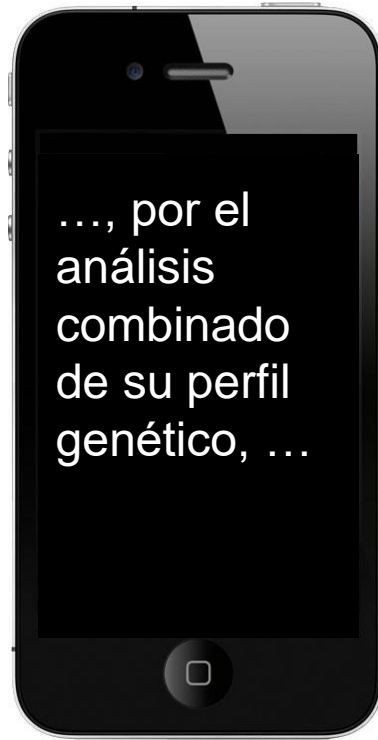
pablo.varona@uam.es

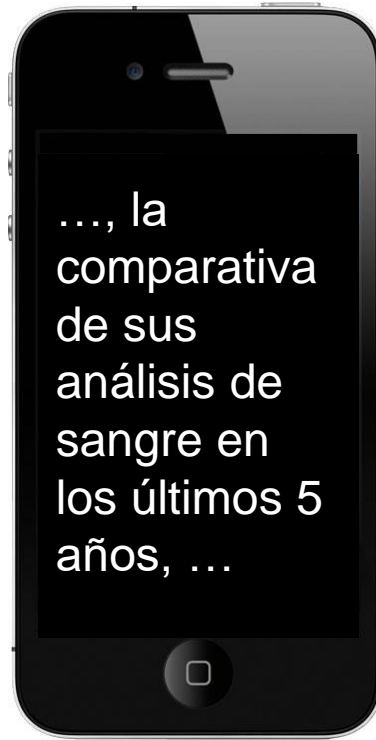
En un futuro cercano una persona recibe un mensaje en el móvil







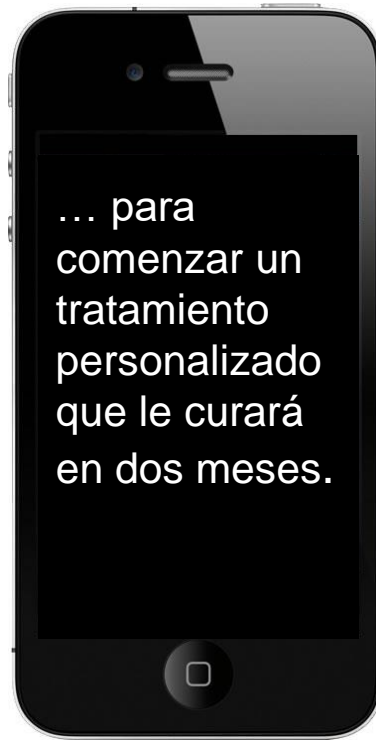
















Esta aplicación es posible con la tecnología actual

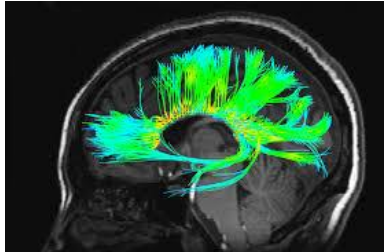
- Parte de esa tecnología se ha descrito en este máster.
- Sin embargo, no existe todavía la infraestructura para implementarla.
- Ni tampoco un marco legal bien definido para que exista.

Incremento de la información biomédica

- Se ha estimado que en 2012 se produjeron unos 500 petabytes de datos digitales de salud y que en 2020 serán entorno a 25,000 petabytes (1 PB = 10^{15} bytes).

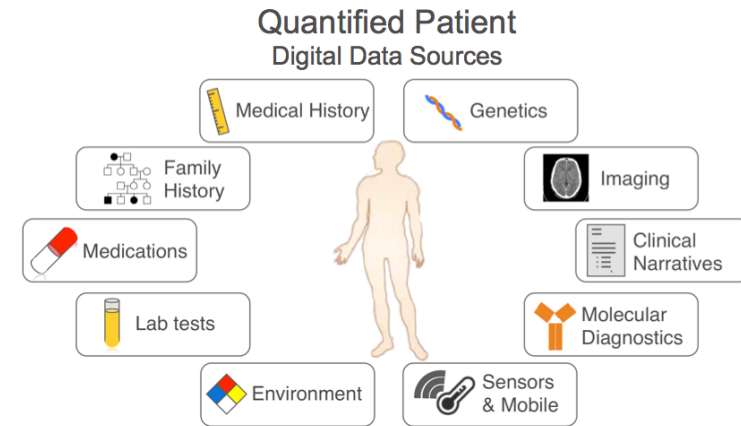
Laura B. Madsen. Data-Driven Healthcare: How Analytics and BI are Transforming the Industry. 2014. Wiley.

Filmar la vida de una persona (100 años) en alta definición (10 megapíxels, 50 fotogramas por segundo) ocuparía 0,5 petabytes.



Incremento de la información biomédica

- El conocimiento médico se va acumulando a partir de la investigación básica y clínica.
- En los últimos años se han desarrollado nuevas tecnologías médicas de coste cada vez más reducido (dispositivos, sensores, software, etc.) y los datos asociados son almacenados rutinariamente.
- Los dispositivos móviles disponen de sensores capaces de recoger, almacenar, tratar y comunicar información relacionada con la salud en tiempo real.



Beneficios de la integración de datos biomédicos

- Nuevos tratamientos, diagnósticos, mecanismos de alarmas, protocolos de seguimiento más eficaces y con un mayor grado de automatización, aplicaciones de telemedicina, etc.
- Medicina personalizada con tratamientos específicos para la singularidad de cada paciente en el instante adecuado.
- Descubrimiento de nuevas enfermedades.
- Identificación precisa del origen y control automático de epidemias.
- Mejoras en la administración de la gestión médica.
- Abaratamiento de los costes sanitarios.
- Desarrollo de tecnología bio-inspirada...

Retos por superar:



- Existe una separación de datos biomédicos por la división tradicional de disciplinas (p.ej. fisiología, genética).
- Muchos médicos y gerentes son reacios a compartir datos e historiales clínicos por temor a futuras demandas, sobre todo en los países más desarrollados tecnológicamente.
- Hay pocas iniciativas públicas para crear las infraestructuras y el marco legal necesarios.
- Riesgos de malos usos e invasión no consentida de privacidad.

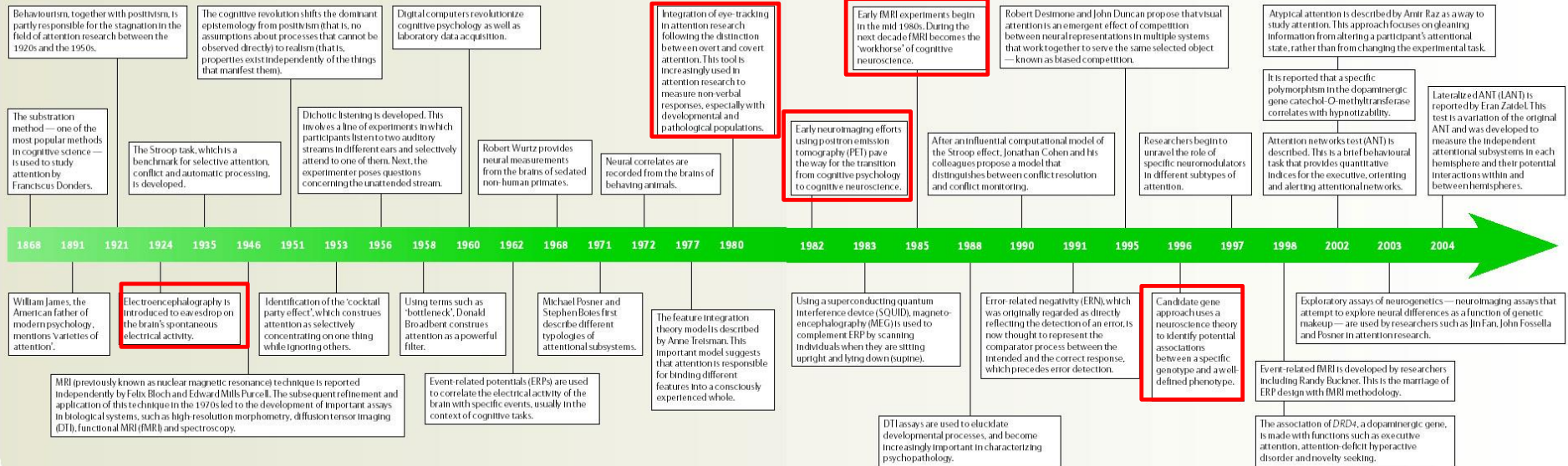
Big data en investigación biomédica:

- Disciplina relativamente reciente
- Pocas bases de datos con información parcial
- Enorme potencial de avance en el conocimiento biomédico.

Fisiología (EEG+fMRI)+Genética+Comportamiento

Evolución científica e integración de conocimiento

Timeline | A sketch of selected landmarks in the study of attention



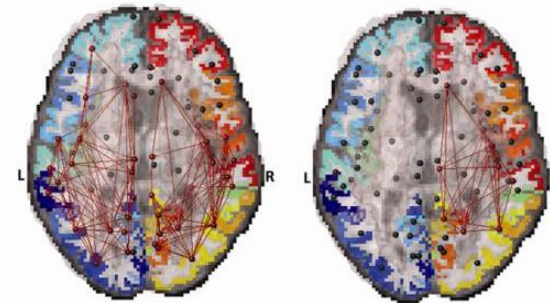
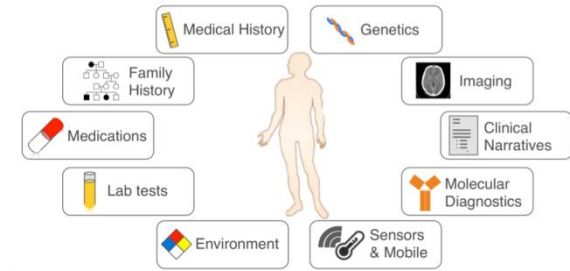
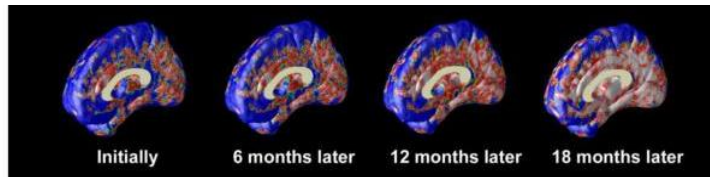
Amir Raz & Jason Buhle. [Typologies of attentional networks](#). *Nature Reviews Neuroscience* 7, 367-379 (2006)

Ejemplo de integración de datos

➤ Pacientes de Alzheimer:

➤ Se integran datos de un número elevado de pacientes y su familia (abuel@, padre o madre, niet@s):

- Imagen de resonancia magnética
- EEG
- Perfil genético
- Medicación aplicada



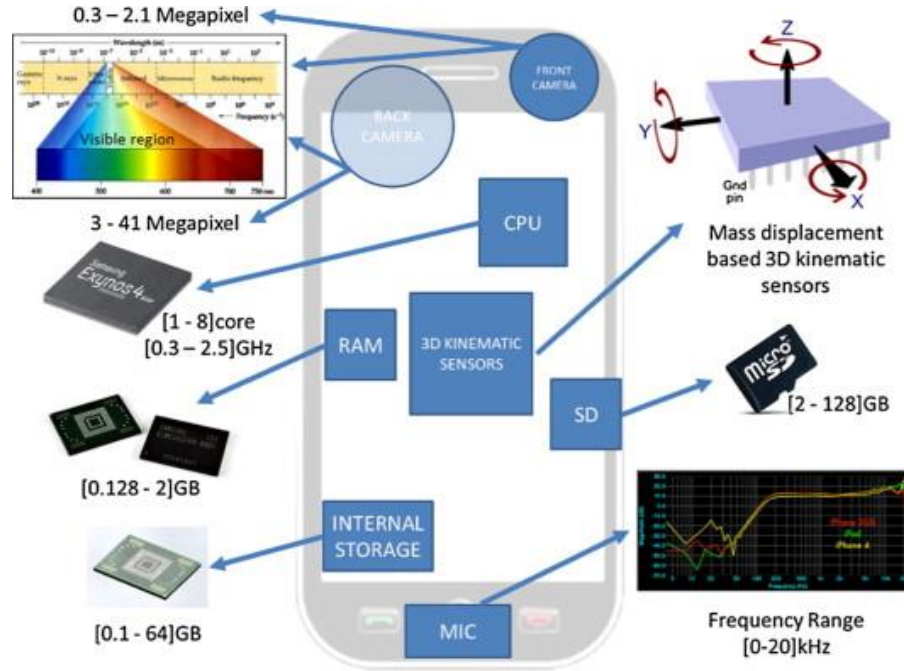
Roussotte et al. 2014. Neuroimaging and Genetic Risk for Alzheimer's Disease and Addiction-Related Degenerative Brain Disorders. *Brain Imaging Behav.* 2014 Jun;8(2):217-33. doi: 10.1007/s11682-013-9263-y

Importancia de los datos generados por el paciente

Mode	Elements	Attributes/units
Sensor	Pedometry/accelerometry	Steps, activity intensity
	Sleep	Sleep duration, latency, interruption
	Weight	Pounds/kilograms
	Blood Pressure	mmHg
	Heart Rate	Beats per minute
	Temperature	Celsius/Fahrenheit
	Environmental exposure	Exposure-dependent
	Blood levels	Glucose, medication levels
	Falls	Times fallen
	Geolocation	Coordinates
Data entry	Exercise testing	Self-administered 6 min walk distance, others
	Diet	Calories, composition
	Mood/stress levels	Type, severity, frequency, interference
	Symptoms	Type, severity, frequency, interference
	Health-related quality of life	Scale, instrument-dependent
	Functional status	Scale, instrument-dependent
	Social support	Scale, instrument-dependent
	Medications (including opiates)	Type, frequency
	Tobacco use	Type, frequency
	Alcohol use	Type, frequency
Other	Social connectedness	Activity (e.g. Facebook, Twitter, others)
	Financial data	Medication and health care expenditure co-pays

Wood WA, Bennett AV, Basch E. [Emerging uses of patient generated health data in clinical research](#), *Molecular Oncology*, 9(5):1018-1024 (2015).

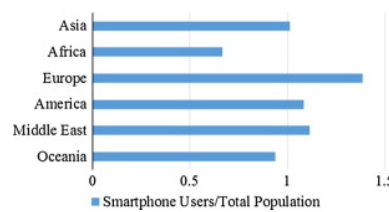
¿Más médicos o más móviles?



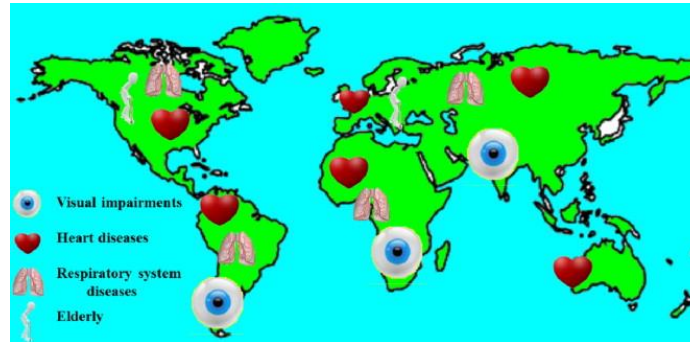
Lamonaca et al. [Health parameters monitoring by smartphone for quality of life improvement](#). *Measurement* 73: 82–94 (2015).

¿Más médicos y/o más móviles?

Distribución de móviles en el mundo:

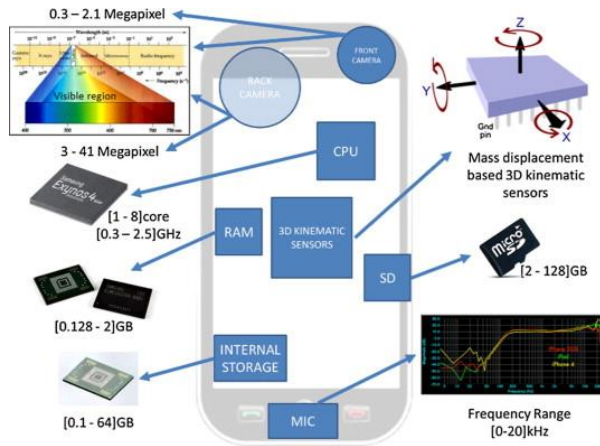


Distribución de enfermedades más comunes



Lamonaca et al. [Health parameters monitoring by smartphone for quality of life improvement](#). *Measurement* 73: 82–94 (2015).

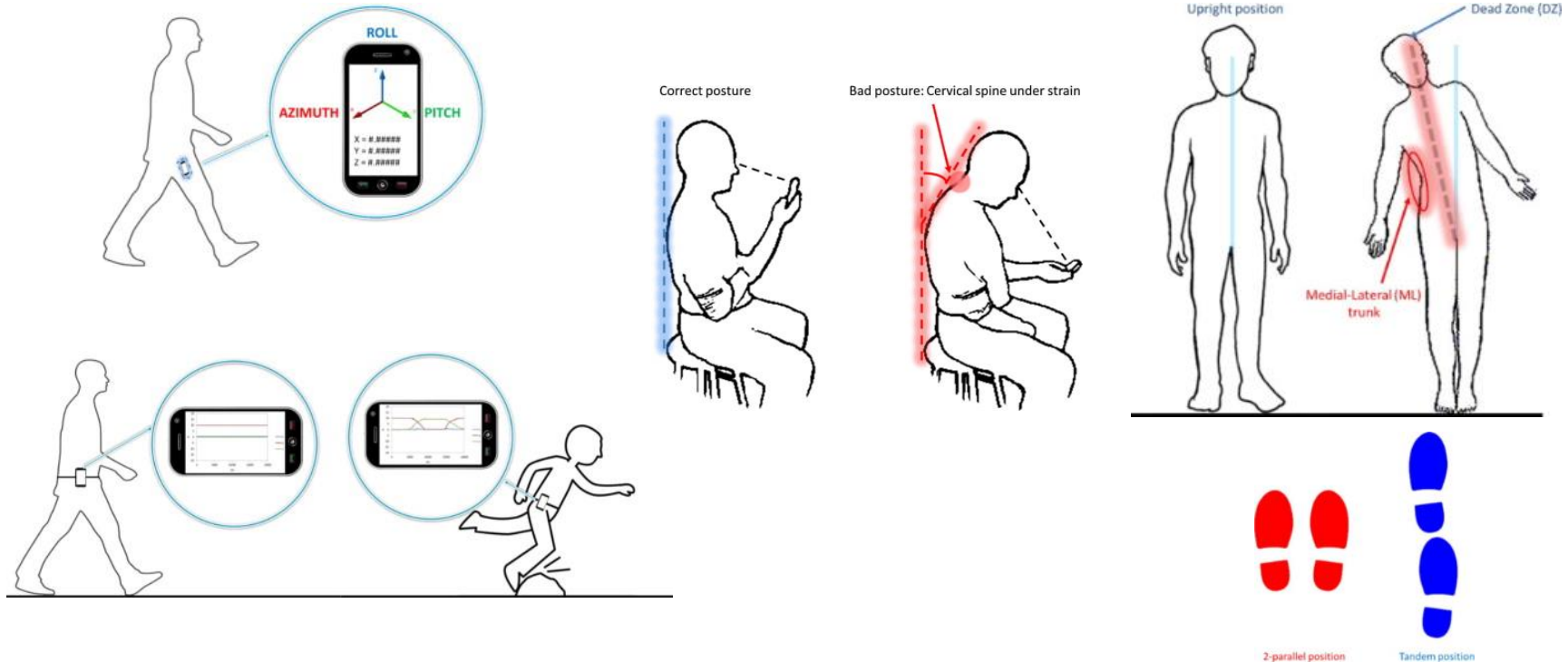
Uso médico de los sensores de un móvil



Tipo de sensor	Aplicación médica
Cinemático: acelerómetros, GPS	Monitorización de actividad
	Detección de accidentes y caídas
	Monitorización postural
	Evolución a la patología (Parkinson, escoliosis, etc.)
Cámara(s)	Estimación de la presión sanguínea
	Ritmo cardíaco
	Evaluación ocular
	Seguimiento ocular para evaluar déficits de atención, dislexia, etc.
Micrófono(s)	Análisis de voz
	Análisis de tos
	Análisis de ronquidos
	Pruebas respiratorias (espirometría)
	Estetoscopio

Lamonaca et al. [Health parameters monitoring by smartphone for quality of life improvement](#). *Measurement* 73: 82–94 (2015).

Uso médico de los sensores de un móvil: sensores cinemáticos

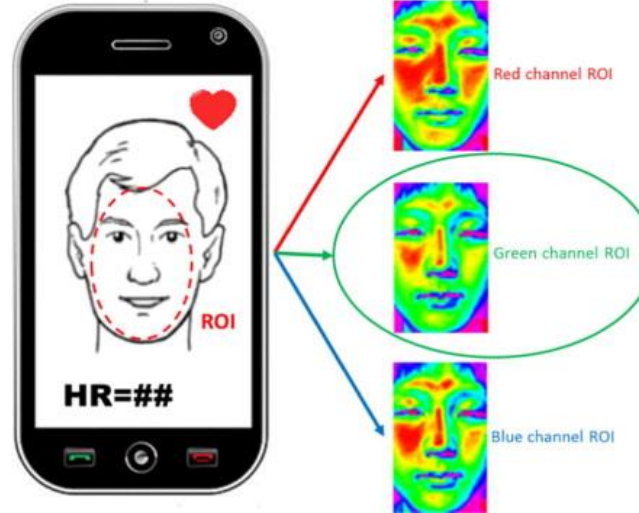


Lamonaca et al. [Health parameters monitoring by smartphone for quality of life improvement](#). *Measurement* 73: 82–94 (2015).

Uso médico de los sensores de un móvil: cámaras



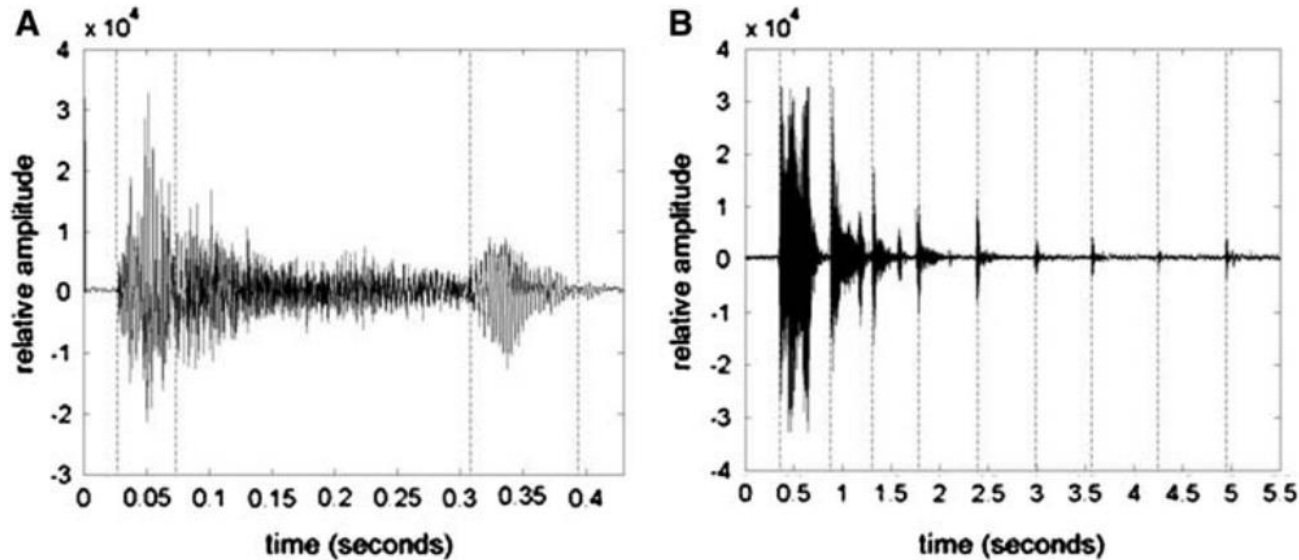
Señal fotopletismográfica: las cámaras pueden detectar la variación volumétrica de la sangre.



Facebeat app: El ritmo cardiaco se puede detectar de las variaciones del canal verde en una región de interés

Lamonaca et al. [Health parameters monitoring by smartphone for quality of life improvement](#). *Measurement* 73: 82–94 (2015).

Uso médico de los sensores de un móvil: micrófonos



Comparison between the LCM and PCA + RF cough detection algorithms.

Algorithm	Mean true positive rate (MTPR) (%)	Mean false positive rate (MFPR) (%)	Mean false alarms (h)
LCM	91	<1	2.5
PCA + RF	92	0.5	17

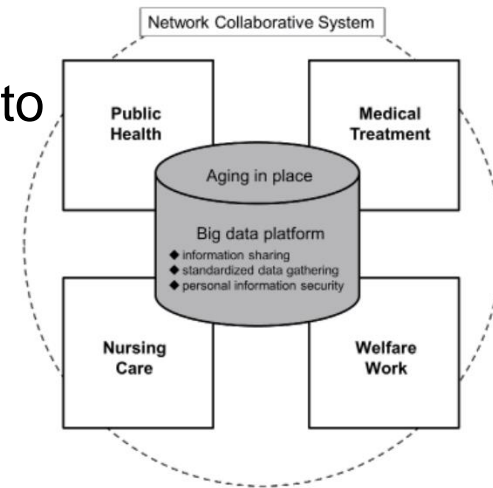
Otros usos: estetoscopia, sonidos respiratorios, etc.

Smith J, Woodcock, A. [New Developments in the Objective Assessment of Cough](#). *Lung* 186(1): 48-54 (2008).

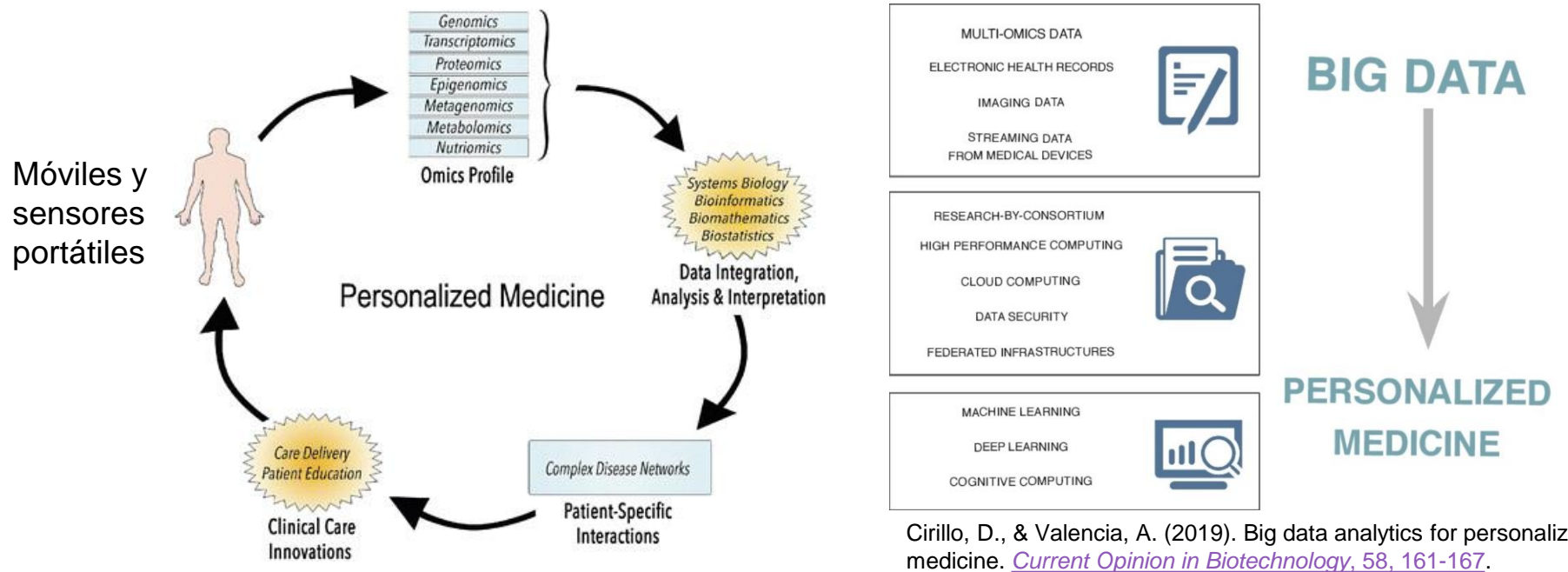
Bigdata para medicina adaptada a la tercera edad

- La nueva generación de sensores en viviendas y dispositivos móviles permite:
 - La comprobación de la evolución de las constantes vitales sin acudir necesariamente a los centros de salud
 - El seguimiento de rutinas diarias de comportamiento
 - La implementación de alarmas automáticas
 - El control de la medicación

Song & Chen. [Public policy response, aging in place, and big data platforms: Creating an effective collaborative system to cope with aging of the population](#). *BioScience Trends* 9(1): 1-6 (2015).



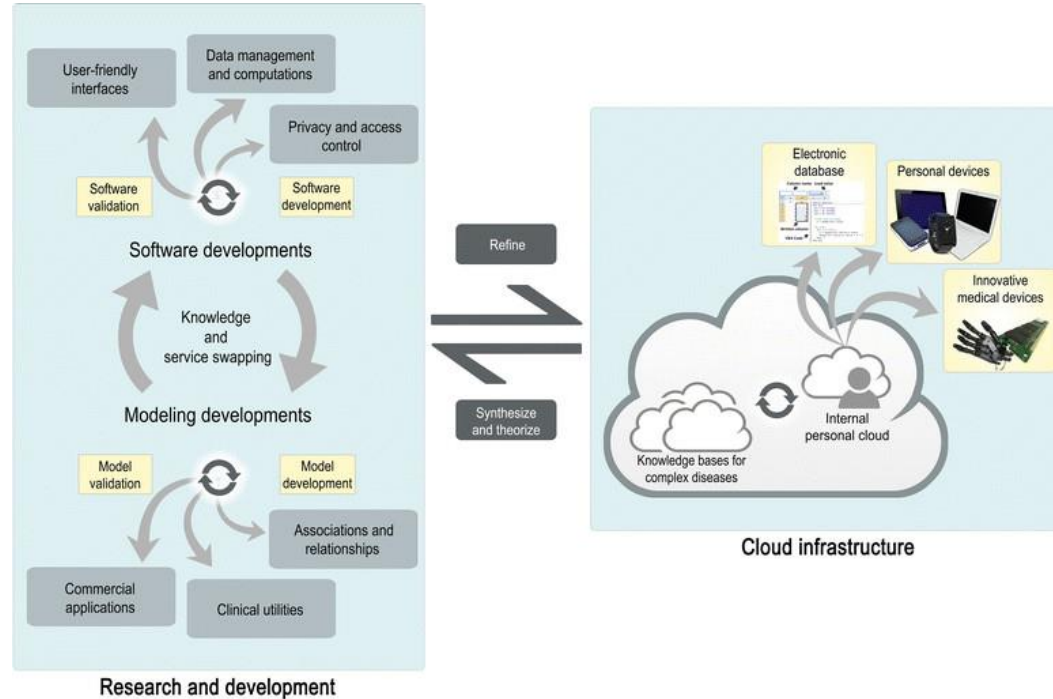
Big data para lograr una medicina personalizada



Esquema básico de medicina personalizada basada en la recogida masiva de datos médicos. La integración de perfiles clínicos, de comportamiento y ambientales permite modelar enfermedades complejas y abre oportunidades para aplicaciones innovadoras.

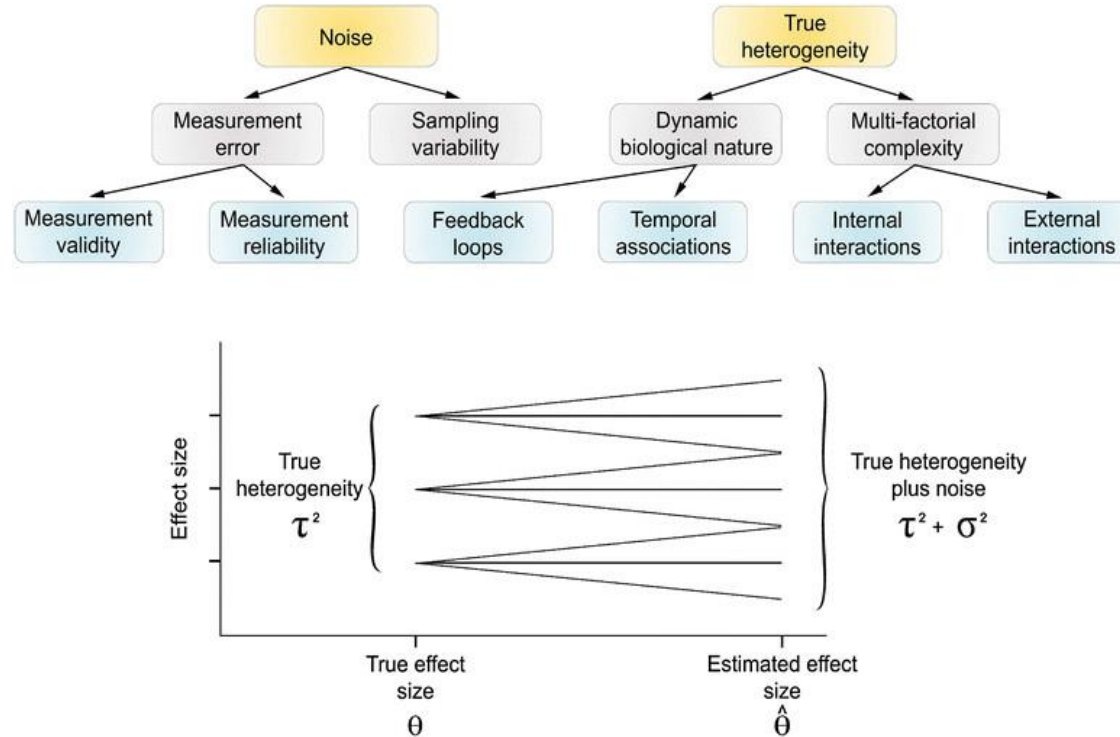
Alyass et al. [From big data analysis to personalized medicine for all: challenges and opportunities](#). *BMC Medical Genomics* 2015; 8:33.

Tecnologías necesarias para la integración y uso de datos



Alyass A, Turcotte M, Meyre D. [From big data analysis to personalized medicine for all: challenges and opportunities.](#) *BMC Medical Genomics* 2015; 8:33.

El reto de la complejidad requiere estandarizaciones y calibraciones

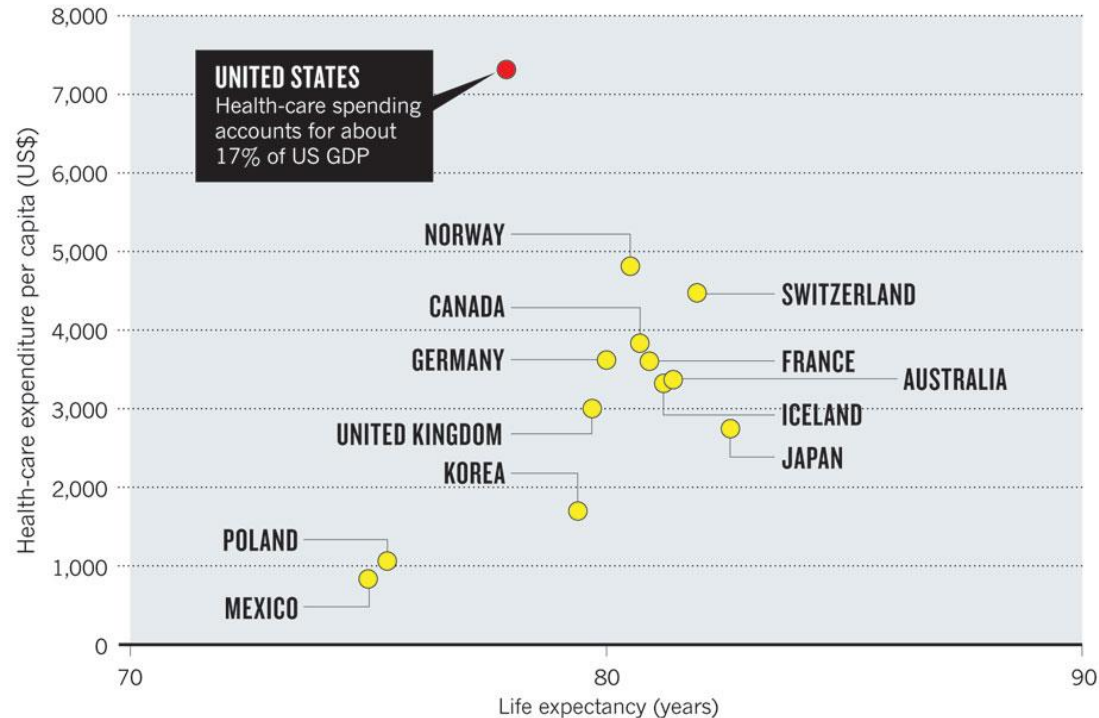


Alyass A, Turcotte M, Meyre D. [From big data analysis to personalized medicine for all: challenges and opportunities.](#) *BMC Medical Genomics* 2015; 8:33.

La inversión en big data biomédico puede ser muy rentable

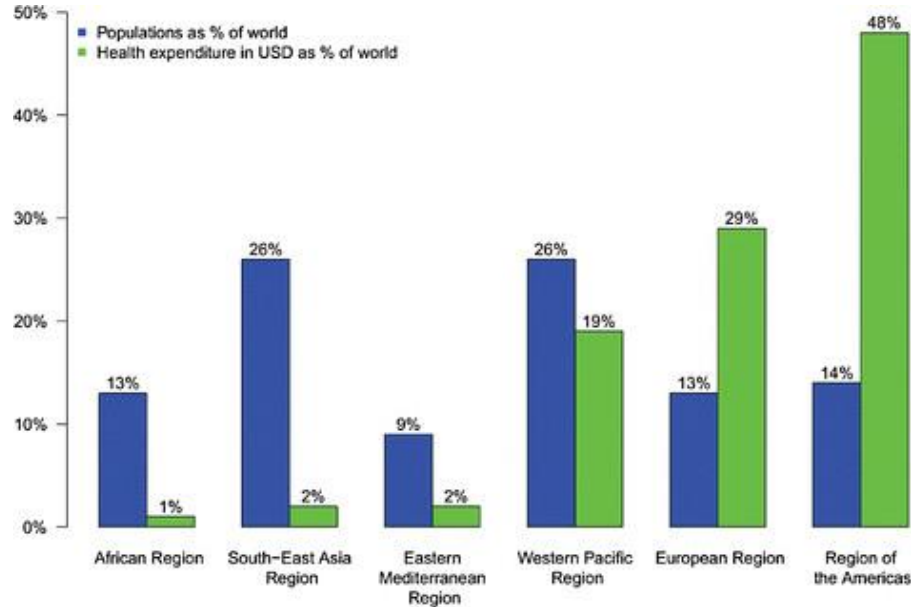
MONEY WELL SPENT?

The United States has not seen an increase in life expectancy to match its huge outlay on health care.



Hearth et al. [Health-care hit or miss?](#) Nature 470, 327–329, 2011.

y de tener un gran impacto en los países menos desarrollados



Alyass et al. [BMC Medical Genomics 2015 8:33](#)

Distribución de la población y gasto global en salud según la Organización Mundial de la Salud 2012

Oportunidades locales (CAM)

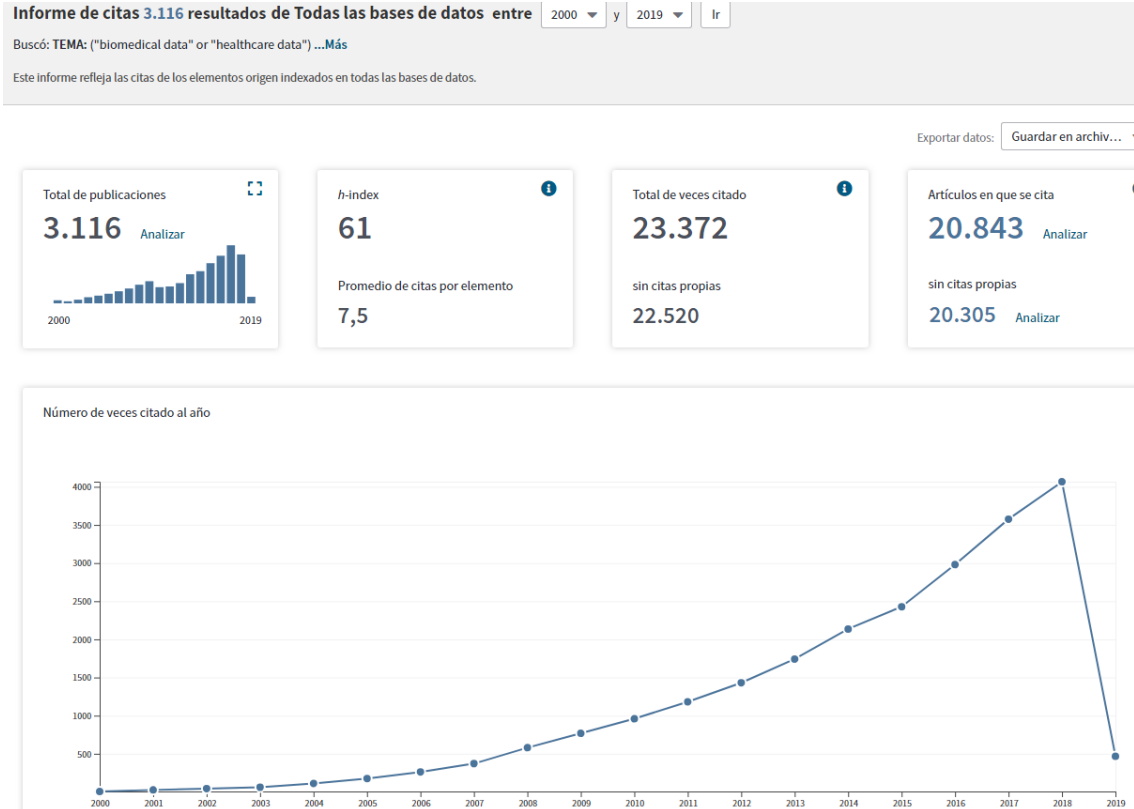
- 400 empresas y 24,000 puestos de trabajo directos relacionados con la tecnología biomédica (farmacéuticas, compañías de dispositivos médicos y biotecnológicos, centros de investigación públicos y privados).
- La CAM es la región líder de España en:
 - Inversión en I+D entre las compañías biotecnológicas privadas (83% del total en España).
 - Número de compañías enteramente dedicadas a tecnología biomédica (32 % del total en España).
 - Número de publicaciones en biomedicina y tecnología biomédicas (31 % del total en España). España ocupa el undécimo lugar del mundo y el séptimo de Europa en esta área científica.
- 75 centros sanitarios con más de 20.000 camas (varios de los más prestigiosos asociados a la UAM).
- Cerca de 27.000 pacientes participan cada año en 1.000 estudios clínicos (las nuevas tecnologías de ingeniería biomédica y big data podrían multiplicar el nº y la eficacia de este tipo de estudios).

Fuente: www.biomadrid.org/es/madrid_bioscience_region.cfm

Oportunidades locales (CAM)

- Aplicación Consult@web
 - Herramienta enfocada al tratamiento de datos masivos, para poder convertirlos en información de utilidad para la práctica clínica diaria.
 - Almacena 120 millones de episodios de Atención Primaria, 120 millones de prescripciones y dispensaciones de fármacos, así como información clínica de 785.000 diagnósticos de alta de Atención Especializada y 60 millones de datos generales del paciente (glucosa, colesterol).
- Unidad de Innovación del Instituto de Investigación Sanitaria del Hospital Clínico San Carlos

Hay mucho trabajo por delante: el número de datos crece más rápido que la investigación en la integración de la información y su desarrollo



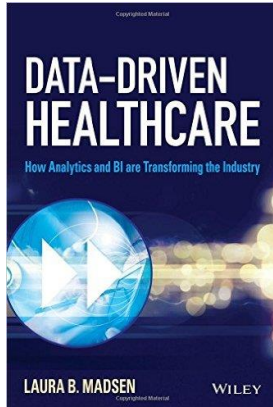
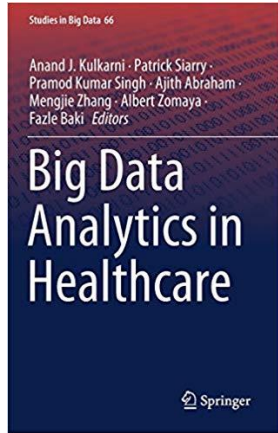
Programas de Big Data del National Institute of Health

NIH "Big Data to Knowledge" (BD2K) centers programs.

BD2K Center of Excellence	Institution	Lead PI(s)
<i>Big Data for Discovery Science (BDDS)</i>	University of Southern California (USC)	Arthur W. Toga
<i>Center for Causal Discovery (CCD)</i>	University of Pittsburgh	Gregory F. Cooper, Ivet Bahar, Jeremy M. Berg
<i>Center for Expanded Data Annotation and Retrieval (CEDAR)</i>	Stanford University	Mark A. Musen
<i>Center for Predictive Computational Phenotyping (CPCP)</i>	University of Wisconsin-Madison	Mark W. Craven
<i>ENIGMA Center for Worldwide Medicine, Imaging, and Genomics</i>	University of Southern California (USC)	Paul M. Thompson
<i>Heart BD2K</i>	University of California Los Angeles (UCLA)	Peipei Ping, Merry Lindsey, Andrew Su, Karol Watson
<i>KnowEnG</i>	University of Illinois at Urbana-Champaign	Jiawei Han, Saurabh Sinha, Jun S. Song, Richard M. Weinshilboum
<i>BD2K-LINCS DCIC</i>	Icahn School of Medicine at Mount Sinai	Avi Ma'ayan, Mario Medvedovic, Stephan C. Schurer
<i>Mobile Sensor Data to Knowledge (MD2K)</i>	University of Memphis	Santosh Kumar
<i>Mobilize</i>	Stanford University	Scott Delp
<i>Center for Big Data in Translational Genomics</i>	University of California Santa Cruz (UCSC)	David Haussler, David Patterson, Laura J. van 't Veer
<i>PIC-SURE</i>	Harvard University	Isaac S. Kohane
<i>Broad Institute LINCS Center for Transcriptomics and Toxicology</i>	Broad Institute	Todd Golub, Aravind Subramanian
<i>bioCADDIE</i>	University of California San Diego (UCSD)	Lucilla Ohno-Machado
BD2K Coordinating Centers		
<i>BD2K Centers Coordinating Center (CCC)</i>	University of California Los Angeles (UCLA)	Peipei Ping, Alex Bui, Wei Wang
<i>BD2K Training Coordinating Center (TCC)</i>	University of Southern California (USC)	John D. Van Horn

For full descriptions of BD2K U01, R25, T32/T15, and other awards, see <https://datascience.nih.gov/bd2k/funded-programs>.

Más información:



- Cirillo, D., & Valencia, A. (2019). Big data analytics for personalized medicine. [*Current Opinion in Biotechnology*, 58, 161-167.](#)
- Bui, A.A.T. Van Horn J.D., [*Envisioning the future of 'big data' biomedicine.* Journal of Biomedical Informatics 2017; 69:115-117.](#)
- Celi LA, Citi L, Ghassemi M, Pollard TJ (2019) [*The PLOS ONE collection on machine learning in health and biomedicine: Towards open code and open data.* PLOS ONE 14\(1\): e0210232.](#)
- Costa F, [*Big Data in Biomedicine.* Drug Discovery Today 2014 19\(4\): 433–440.](#)
- Alyass A, Turcotte M, Meyre D. [*From big data analysis to personalized medicine for all: challenges and opportunities.* BMC Medical Genomics 2015; 8:33.](#)
- Raghupathi W, Raghupathi V. [*Big data analytics in healthcare: promise and potential.* Health Inf Sci Syst. 2014;2\(1\):3.](#)
- Hersh W, Jacko JA, Greenes R, Tan J, Janies D, Embi PJ, et al. [*Health-care hit or miss?* Nature 2011;470\(7334\):327–329.](#)
- Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). [*Health-CPS: Healthcare cyber-physical system assisted by cloud and big data.* IEEE Systems Journal, 11\(1\), 88-95.](#)