

**DNA sequencing**  
is the technology that experienced  
the ***most dramatic advances***  
in the human history

the last ten years

2008

2018

**50,000 b**

day  
1 equipment

**18,000,000,000,000 b**

day  
1 equipment

**0.000016**

human genomes  
day

100,000,000 USD

**600**

human genomes  
day

1,000 USD



illumina®

# Projected annual storage in 2025

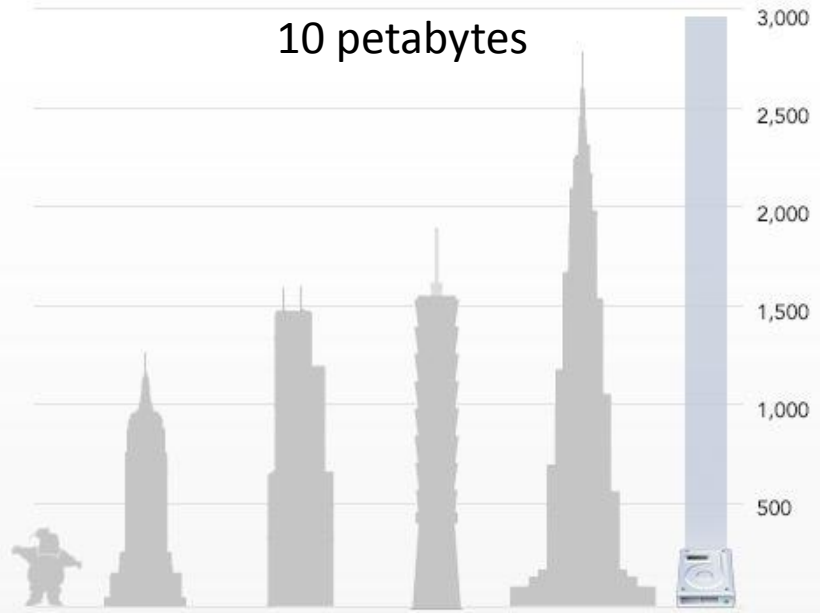
**Twitter:** 1–17 petabytes per year

**Astronomy:**  
1,000 PB/year

**YouTube:**  
1,000–2,000 PB/year

**Genomics:**  
2,000-40,000  
PB/year

## BACKBLAZE DRIVES STACKED



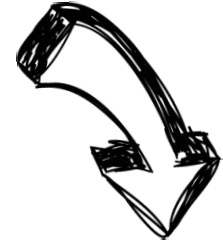
Structure	Height (feet)
Empire State Building	1,250
Willis (Sears) Tower	1,451
Taipei 101	1,671
Burj Khalifa	2,717
Drives	2,968

\* 6,195 drives x 5.75 inches of drive height = 35,621 inches or 2,968 feet

*Biological interpretation  
(social, legal issues)  
Lagging Behind*



*Technology  
at the forefront*





# THE DNA OF A NATION

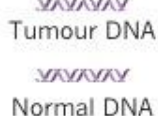
*The United Kingdom aims to sequence 100,000 human genomes by 2017. But screening them for disease-causing variants will require innovative software.*

**~50,000**  
people with rare  
diseases and  
their parents



DNA

**~25,000**  
people with  
cancer



Tumour DNA

Normal DNA

## THE CLINICAL GENOME

Genomics England plans to sequence 100,000 genomes by 2017. The genomic data will be crucial for diagnosing and treating disease, but its interpretation will require automated, specialized software.



### RECRUITMENT OF 75,000 PEOPLE

The 100,000 Genomes Project is recruiting people with cancer and rare diseases. The genomes of both normal and tumour cells will be sequenced in people with cancer.

### NEXT-GENERATION SEQUENCING

The Californian company Illumina will use UK-based high-throughput sequencing machines to produce whole-genome sequences and identify genetic variants.

### AUTOMATED INTERPRETATION

Four UK and US companies will use specialized software to automatically analyse the genetic variants that may be linked to disease.

### CLINICAL INTERPRETATION

Around 2,000 UK scientists and clinicians will pore over the data to validate or better understand how the variants may cause disease before the information is fed back to patients.





DATA

# Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

# So you wanna be a data scientist? A guide to 2015's hottest profession

## Science

- 1.Math
- 2.Machine learning

## Data

- 1.Statistics
- 2.R programming

## Art

- 1.Visualisation
- 2.Creativity  
(find signal in noise)



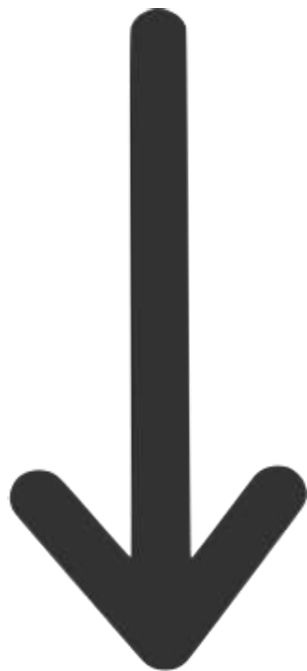


Population power

## HiSeq X Ten

20,000 genomes/year

600 genomes/run



Going backwards

illumina®



## MiSeq

3 exomes/run

Individual power

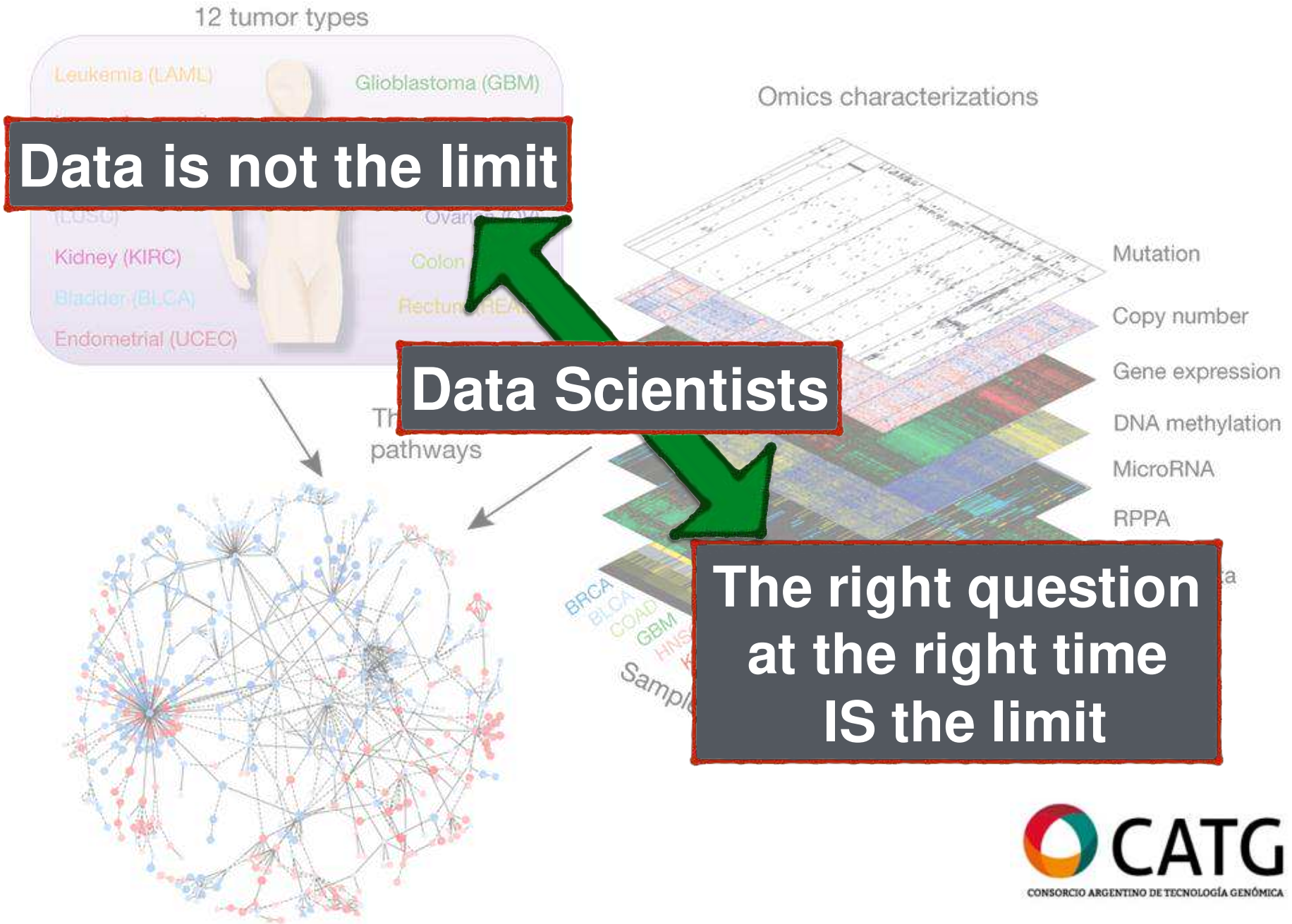


Production power is not a limitation anymore

*Too much information  
is good and bad  
at the same time*

*Most relevant question in the big data era:  
**What do we want from the data???***

# Big Data in genomics



**Data is not the limit**

**Data Scientists**

**The right question at the right time IS the limit**

# Genomics route to the clinic

Research

Clinical Diagnostics



## CAREERS

**TRANSITIONS** From building houses to building molecules p.13

**FUTURE PLANS** Three steps to prepare for the next five years [nature.com/fpdr](http://nature.com/fpdr)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



DPA\_SCHWARTZ/GETTY

GENETICS

## Fluent in DNA

*As genomics migrates to the clinic, job options are emerging for genetic counsellors to explain the meaning in mutations.*

**Genetic Counsellor**  
Is the next big thing  
in hot professions



## IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

**1. ABILIFY (aripiprazole)**  
Schizophrenia



**2. NEXIUM (esomeprazole)**  
Heartburn



**3. HUMIRA (adalimumab)**  
Arthritis



**4. CRESTOR (rosuvastatin)**  
High cholesterol



**5. CYMBALTA (duloxetine)**  
Depression



**6. ADVAIR DISKUS (fluticasone propionate)**  
Asthma



**7. ENBREL (etanercept)**  
Psoriasis



**8. REMICADE (infliximab)**  
Crohn's disease



**9. COPAXONE (glatiramer acetate)**  
Multiple sclerosis



**10. NEULASTA (pegfilgrastim)**  
Neutropenia



# DEEP MEDICINE

HOW ARTIFICIAL  
INTELLIGENCE  
CAN MAKE  
HEALTHCARE  
HUMAN AGAIN

ERIC TOPOL

With a foreword by  
ABRAHAM VERGHESE,  
*author of Cutting for Stone*



SÍNDROME DE QT LARGO

SÍNDROME DE BRUGADA

SÍNDROME DE QT CORTO

TAQUICARDIA VENTRICULAR

CATECOLAMINÉRGICA POLIMÓRFICA

FIBRILACIÓN ATRIAL FAMILIAR

La fibrilación atrial (FA) familiar es una manifestación asociada a diversos fenotipos electrofisiológicos y/o inclusive cardiopatías estructurales, en los que el desarrollo de FA puede presentarse en forma primaria.

#### GENES COMPROBADOS

GJA5	KCNQ1	SCN5A	
------	-------	-------	--

#### GENES EMERGENTES

ABCC9	DSC2	EMD	HCN4
JPH2	KCNA5	KCND3	KCNE1
KCNE2	KCNE3	KCNJ2	KCNJ8
LMNA	MYH6	NKX2-5	NPPA
SCN3B	SCN4B		



Apple Watch.  
Helping your patients  
identify early warning  
signs.

