

# *Global –Omics data: What do we have and what can we do with it?*

Arcadi Navarro

VIII Conferencia Anual de las Plataformas Tecnológicas

Madrid, 24-03-2015

# Outline

- 10' on global -Omics data

- Origins & amount
- The issue of sharing



- 10' on a project done so far with very little data

- The pleiotropic theory of senescence?
- Genome-Phenome data
- Genome-wide evidence for a theory of senescence



# Outline

## ○10' on global -Omics data

- Current status
- Recent and incoming novelties
  - Unified brand
  - EGA 2.0

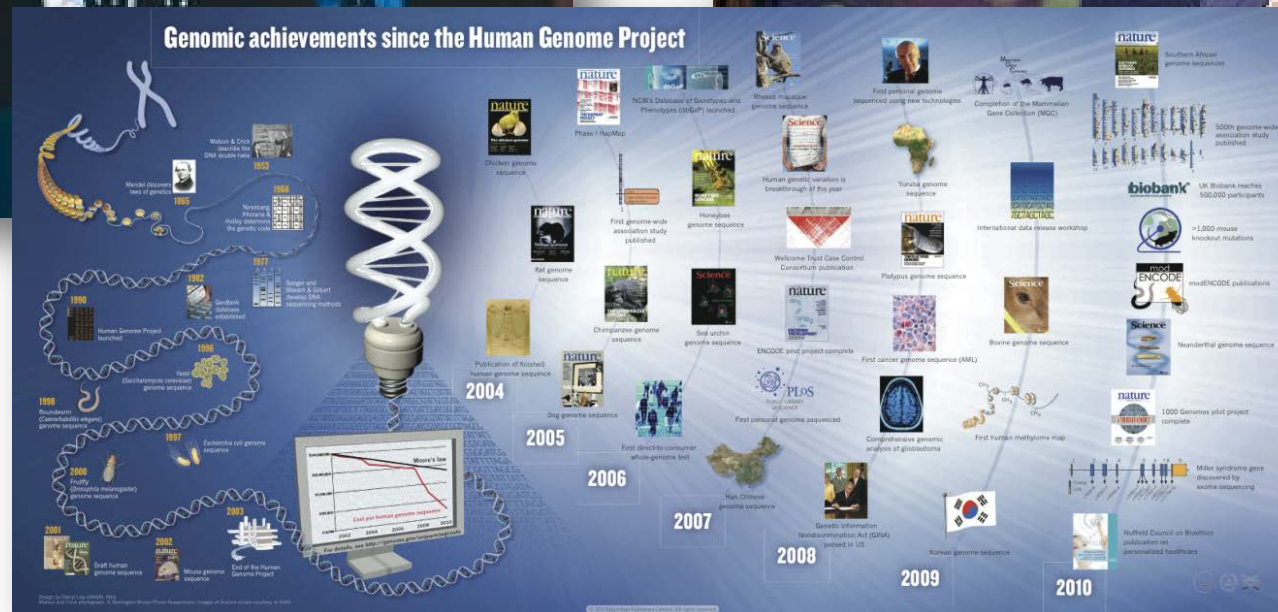


## ○10' on a project done so far with very little data

- The pleiotropic theory of senescence?
- Genome-Phenome data
- Genome-wide evidence for a theory of senescence

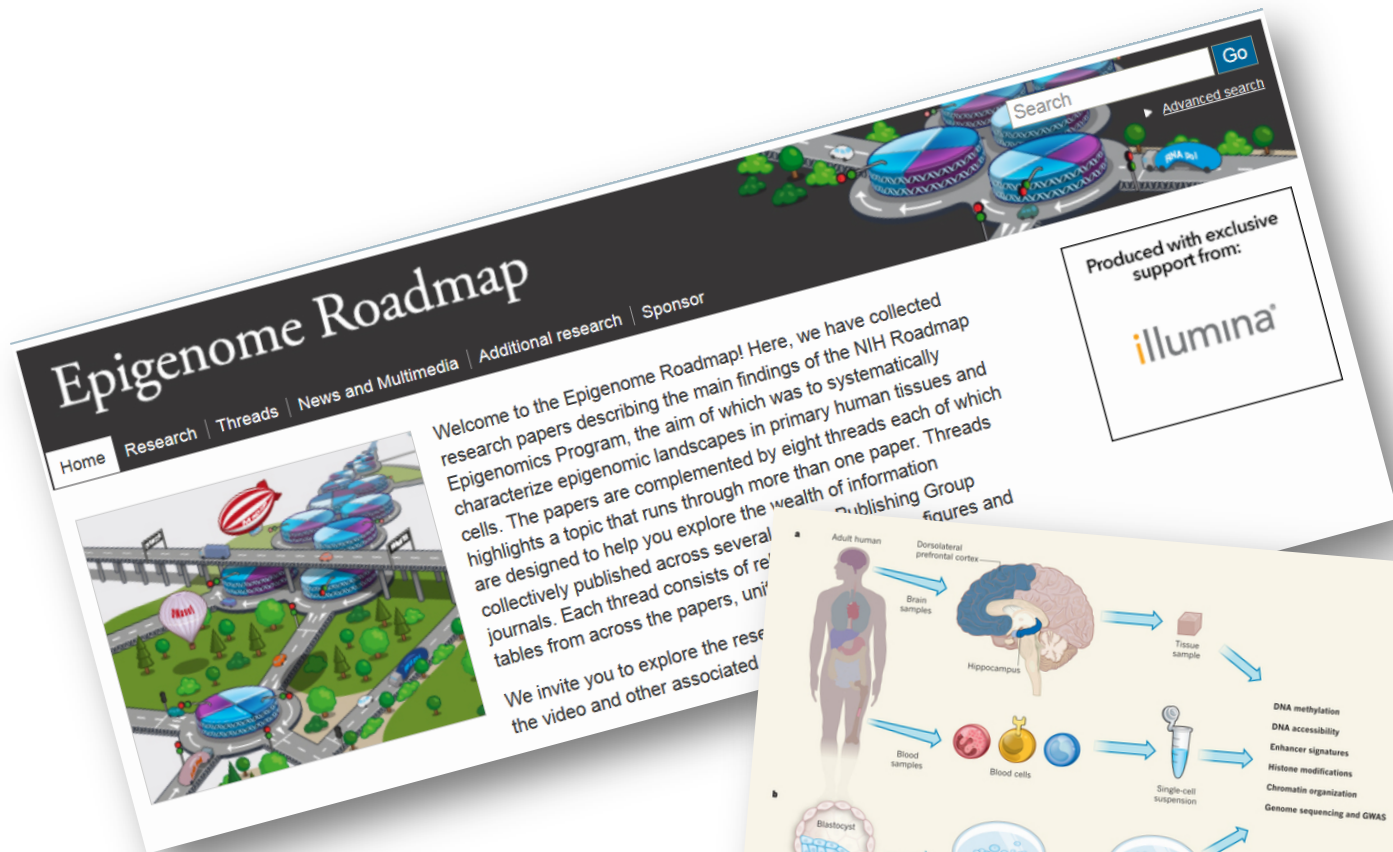


The promise in the 90s: “In 10 years we will unravel the genetic bases of complex diseases!!” ... It was the reason under the Human Genome Project





# There's plenty of additional information



**Epigenome Roadmap**

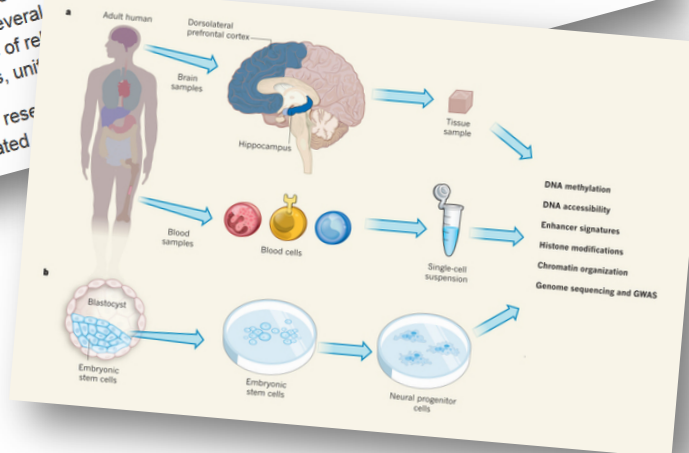
Home | Research | Threads | News and Multimedia | Additional research | Sponsor

Search  Go

Advanced search

Produced with exclusive support from: **illumina**

Welcome to the Epigenome Roadmap! Here, we have collected research papers describing the main findings of the NIH Roadmap Epigenomics Program, the aim of which was to systematically characterize epigenomic landscapes in primary human tissues and cells. The papers are complemented by eight threads each of which highlights a topic that runs through the wealth of information published in the papers. Each thread consists of several papers, collectively published across several journals. Each thread consists of several papers, unified by a common theme. We invite you to explore the results of the program and other associated information.



**1000 Genomes**  
A Deep Catalog of Human Genetic Variation

Home | About | Data | Analysis | Participants | Contact | Browser | Wiki | FTP search

**LATEST ANNOUNCEMENTS**

WEDNESDAY OCTOBER 31, 2012  
**An integrated map of genetic variation from 1092 human genomes**  
The Phase 1 publication: An integrated map of genetic variation from 1092 human genomes is now available from Nature and can be downloaded directly from the 1000 Genomes Project website. The paper is distributed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 license. Please share our paper appropriately.  
All the data files associated with this paper can be found in our [phase1 analysis results directory](#).

**Recent project announcements**

FRIDAY NOVEMBER 9, 2012  
**1000 Genomes Tutorial and Poster Slides #ASHG2012**  
The slides from our tutorial session which was held on Wednesday 7th and the Data Access Poster are both available from our website.

**NAVIGATION**

- Frequently Asked Questions

**LINKS**

- All Project Announcements
- Sample and Project Information
- Media Archive
- Download the 1000 Genomes



Genome Wide Association Studies (GWAS) have unveiled hundreds of SNPs associated to complex disease

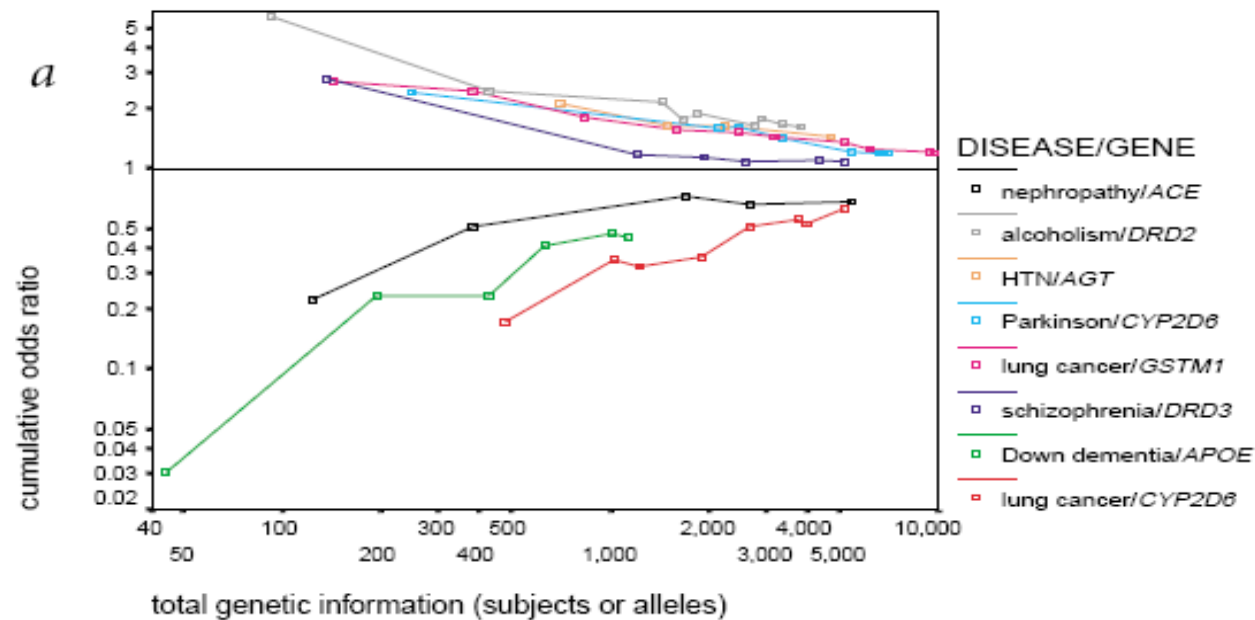
–*Omics* data has gone BIG

Plenty of risk and protective **marker alleles** have been identified in more than 2,000 GWAS



# So there is plenty of info of the genetic basis of disease, but... is it reliable?

Recall debates on **lack of replicability** and **missing heritability**

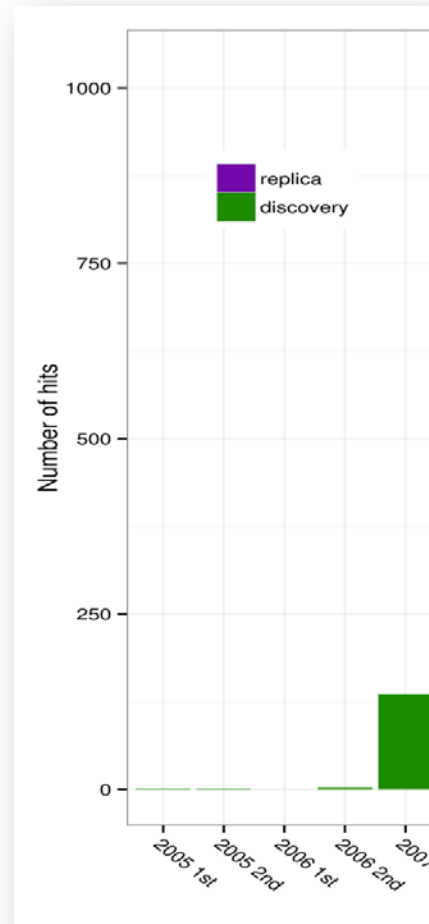


## Replication validity of genetic association studies

John P.A. Ioannidis<sup>1-3</sup>, Evangelia E. Ntzani<sup>1</sup>, Thomas A. Trikalinos<sup>1</sup> & Despina G. Contopoulos-Ioannidis<sup>1,4</sup>  
nature genetics • volume 29 • november 2001

# Have GWAS been replicated?

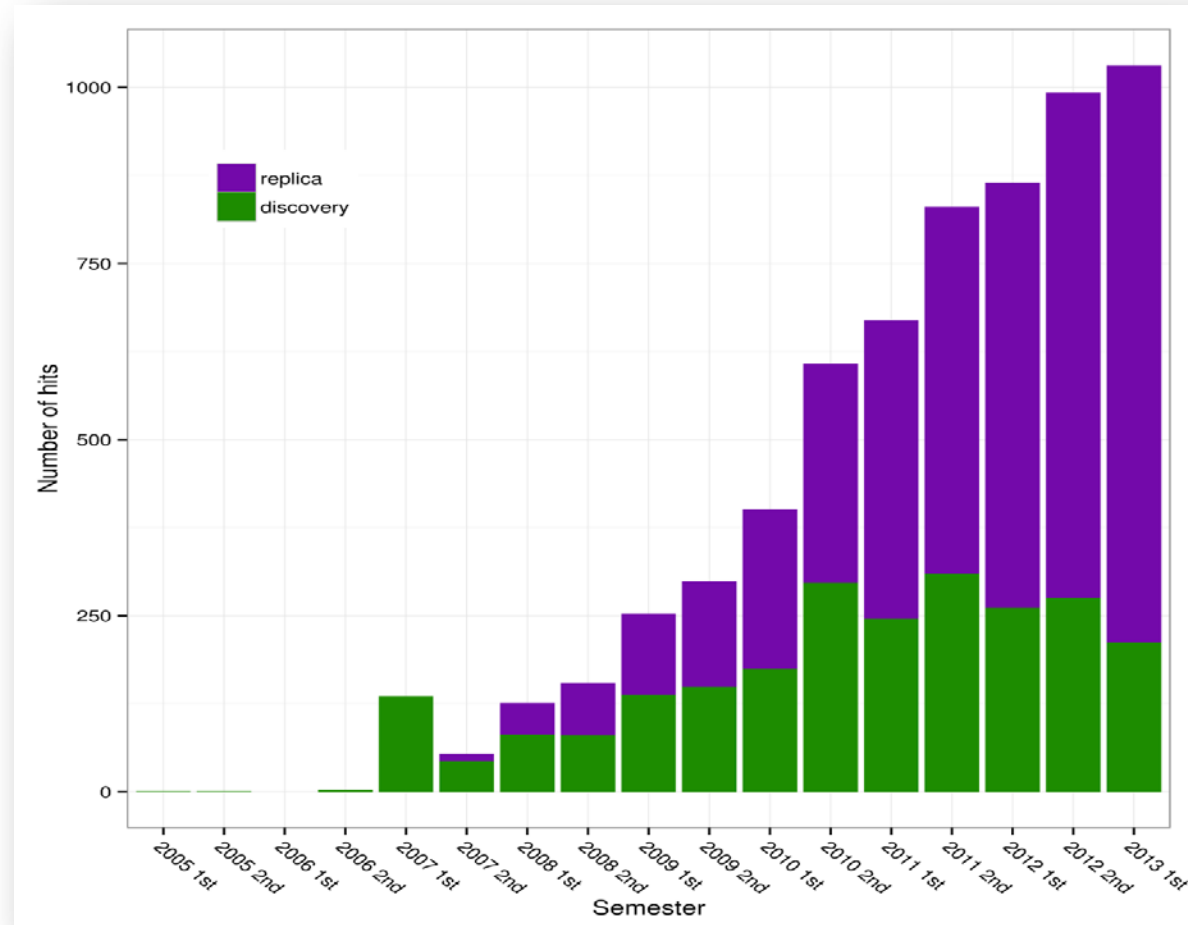
June 2007, publication of the WTCCC





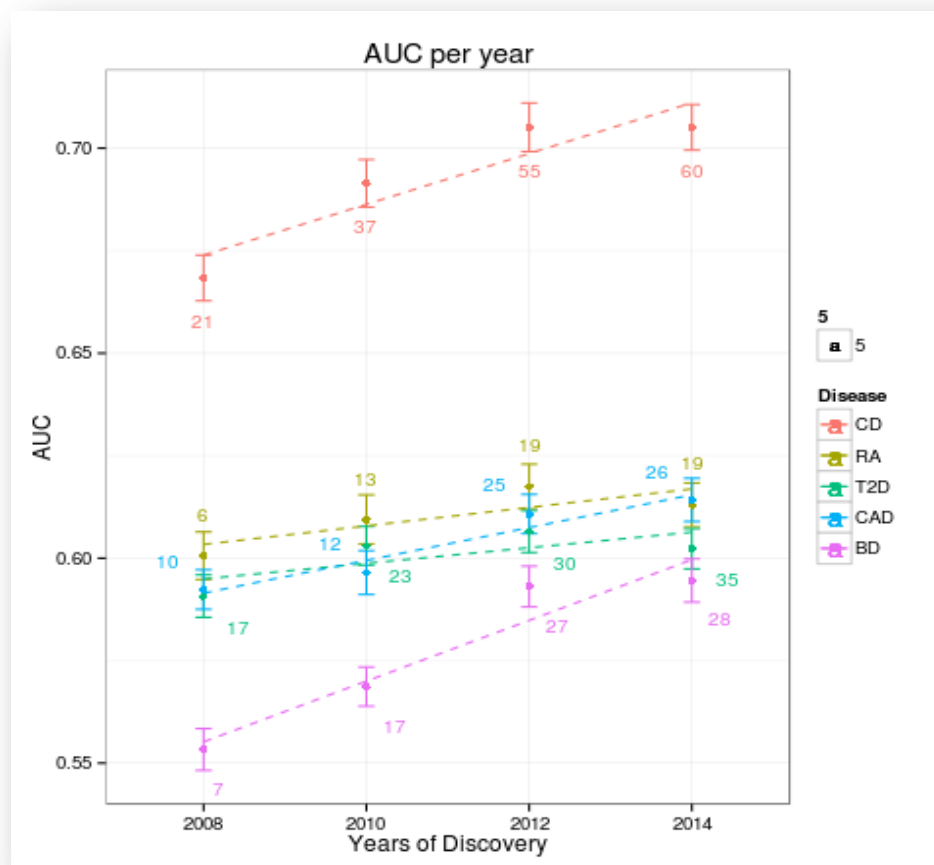
# Have GWAS been replicated?

Contrary to candidate-gene association studies GWAS have produced replicable results

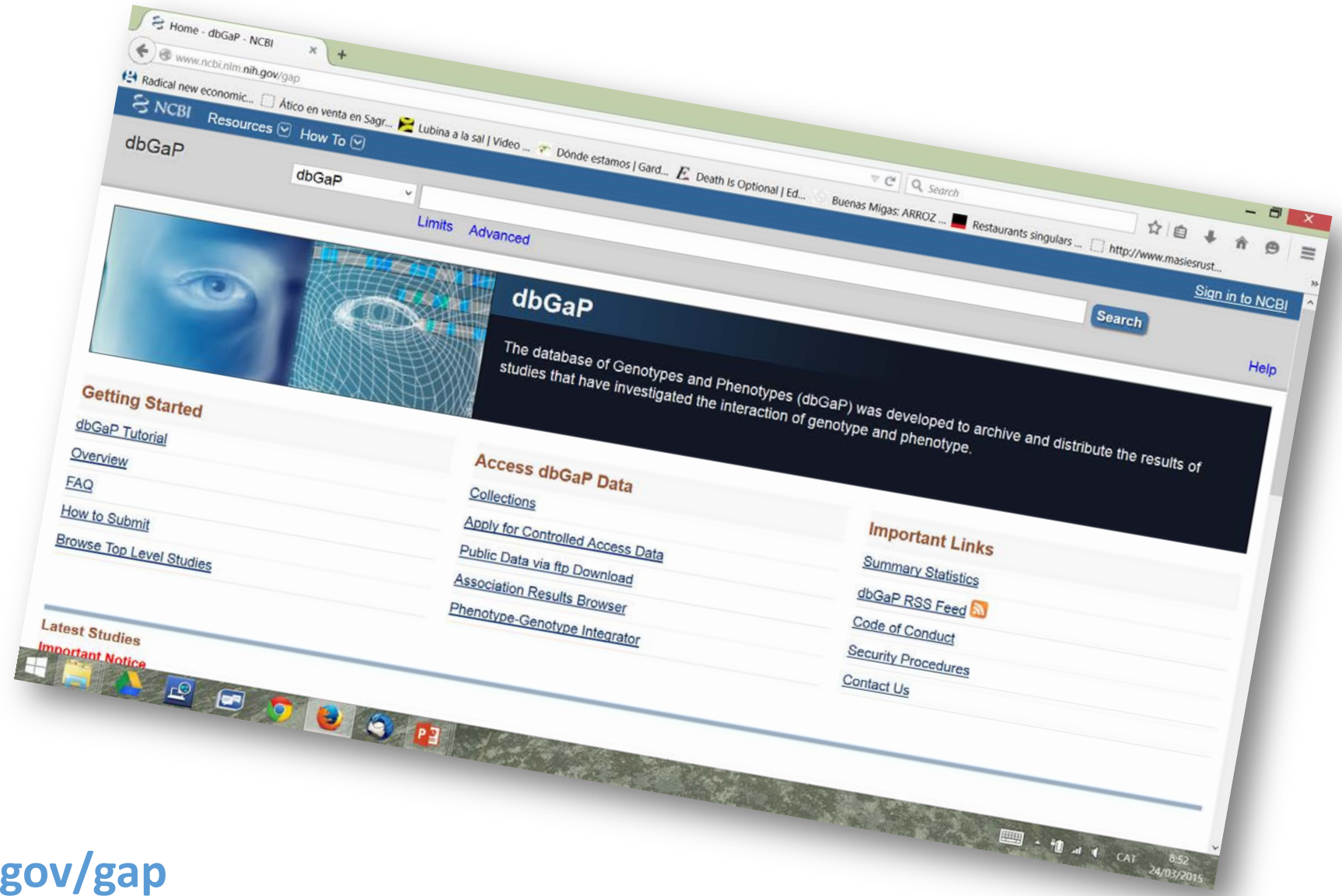


And prediction ability from GWAS results (i.e. variance explained )  
keeps increasing with time

[here prediction of the original WTCCC cases and controls with newer results]



# The USA solution: dbGaP



<http://www.ncbi.nlm.nih.gov/gap>

# The EU solution: EGA

<http://ega.crg.eu>

<http://www.ebi.ac.uk/ega>

**European Genome-phenome Archive**

EGA home | About | Studies | Datasets | Data access committees | Data providers

The European Genome-phenome Archive (EGA) allows you to explore **datasets** from genomic **studies**, provided by a range of **data providers**. Access to datasets must be approved by the specified **Data Access Committee (DAC)**.

### Studies

Studies are experimental investigations of a particular phenomenon or trait.

[Browse all studies](#)

### Learn about the EGA

- [Introduction to the EGA](#)
- [How to obtain an account with the EGA](#)
- [Using your EGA account](#)

### Datasets

The EGA archives a large number of datasets, the access to which is controlled by a Data Access Committee (DAC).

[Browse all datasets](#)

[Browse all control datasets](#)

### Data Access Committees

Providers may be involved in the creation, submission and management of Data Access Committees.

### Services

- By topic
- By name (A-Z)
- Help & Support

### Research

- Overview
- Publications
- Research groups
- Postdocs & PhDs

EMBL-EBI | News | Brochures | Contact us | Intranet

**EGA** European Genome-phenome Archive

Helpdesk | Login

HOME | BROWSE | REQUEST | SUBMISSION

### Latest studies

Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids – March 14 2014

Pulmonary neuroendocrine tumours of the lung. The molecular alterations underlying the pathogenesis of these tumours have not been systematically studied so far. Here [Read more](#) →

Study 1 / 5

[Next study](#)

Published in: [Nature Communications](#)

### What is in the EGA?

Studies in the EGA by disease

Click on a column to view category subgroups

DISEASES	TECHNOLOGY	SAMPLE TYPES
Inflammatory	267	31
Cancer	24	9
Cardiovascular	118	62
Infectious		
Neurological		
Other		

Click below to show studies: [Neurological: 21.0% of total](#)

If applicable, a study may be included in more than one category

Search...

[BROWSE](#)

- Studies
- Datasets
- Centers

[I want to access data](#)

[I have data to submit](#)

Need some [tips](#) on how to search?

**HELP**

- FTP & Aspera
- Tools
- About the EGA

Powered by: [CRG](#) [upf](#) [BSC](#)

Funded by: [Generalitat de Catalunya](#) [Obra Social Fundación "la Caixa"](#) [EXCELENCIA SEVERO OCHOA](#) [EMBL-EBI](#) [iñb](#)

© CENTRE DE REGULACIÓ GENÒMICA (CRG) · EDIF. PRBB, DR. AIGUADER, 88 · 08003 BARCELONA · SPAIN · TEL. +34 93 310 01 00 · FAX +34 93 310 00 99

GENERAL INFORMATION ABOUT THE CRG · EVENTS · CONTACT US · LEGAL NOTICE

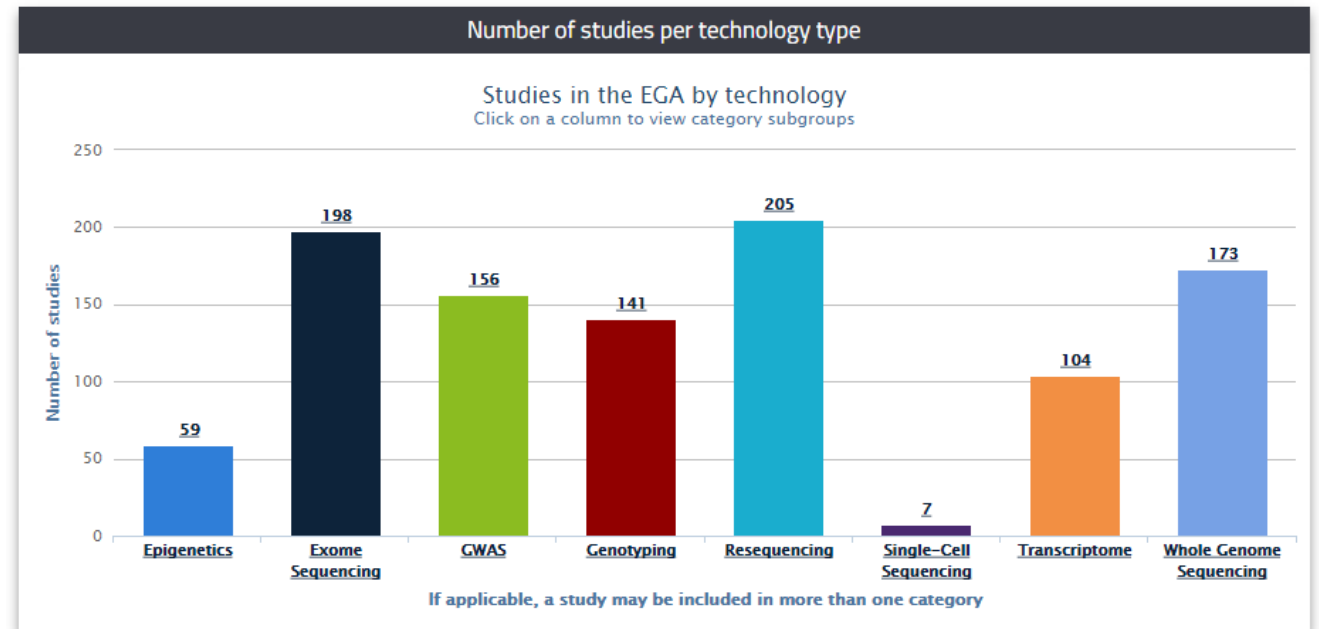
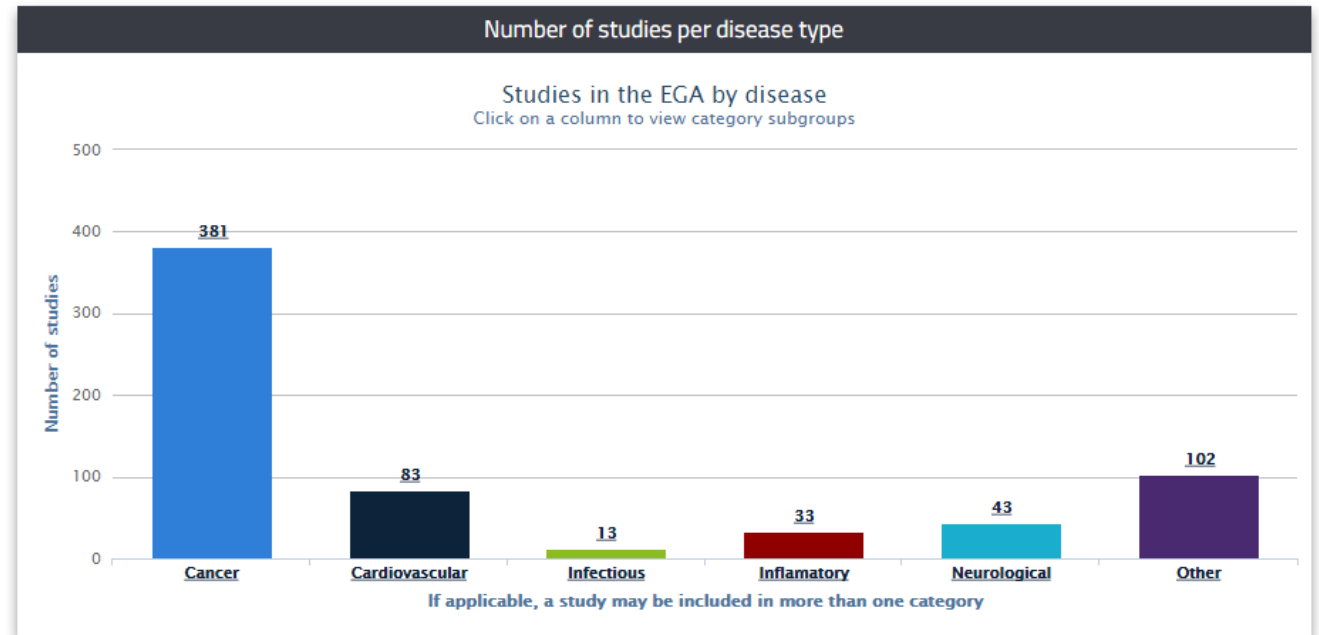


# The EGA in numbers

- > 700 Studies
- > 500,000 individuals
- > 1,200 Datasets
- > 250 Data providers
- > 6,000 Data Requesters

# The EGA in volume

- > 250,000 files
- > 1,5 Petabytes

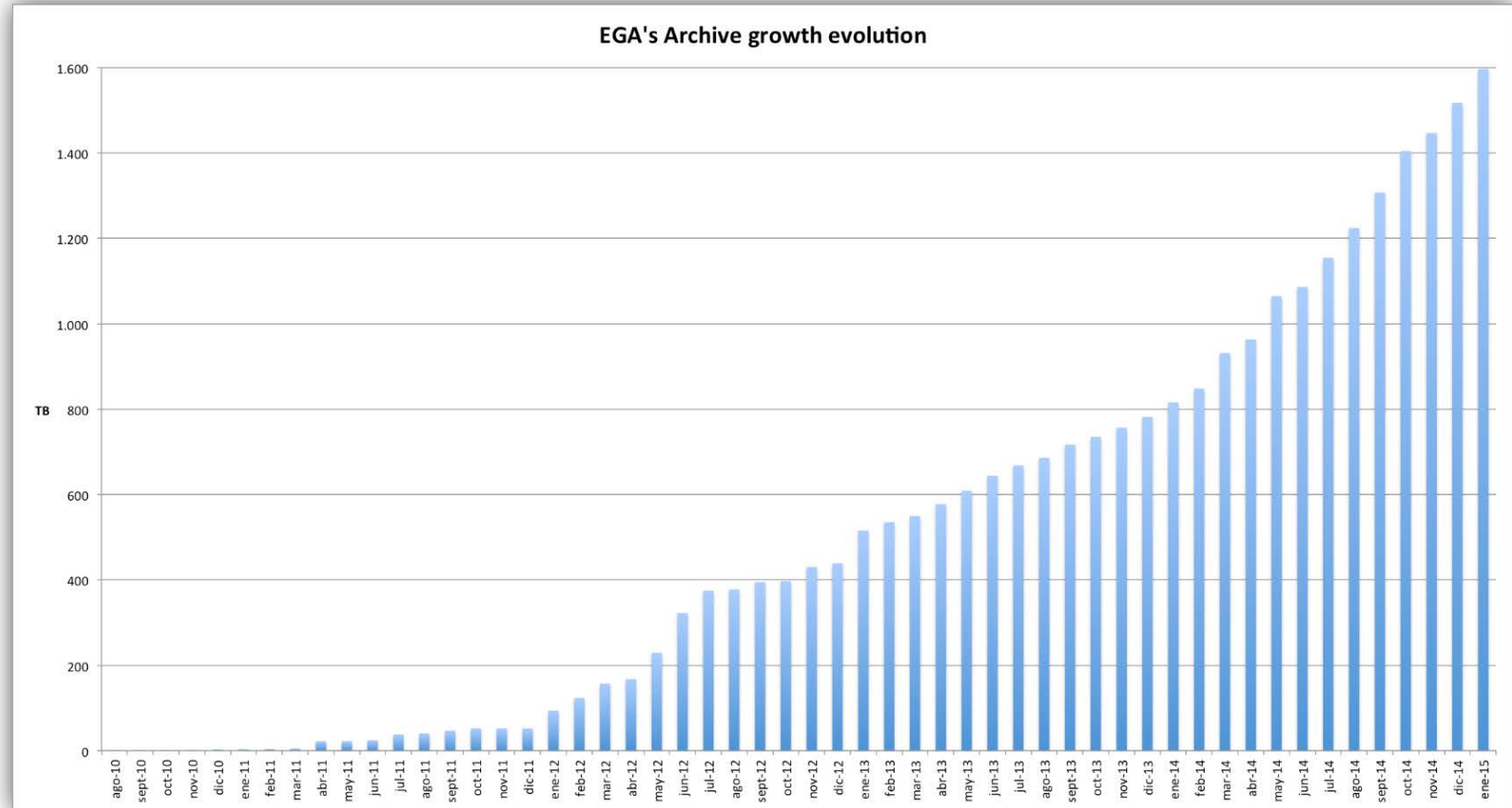


# The EGA in numbers

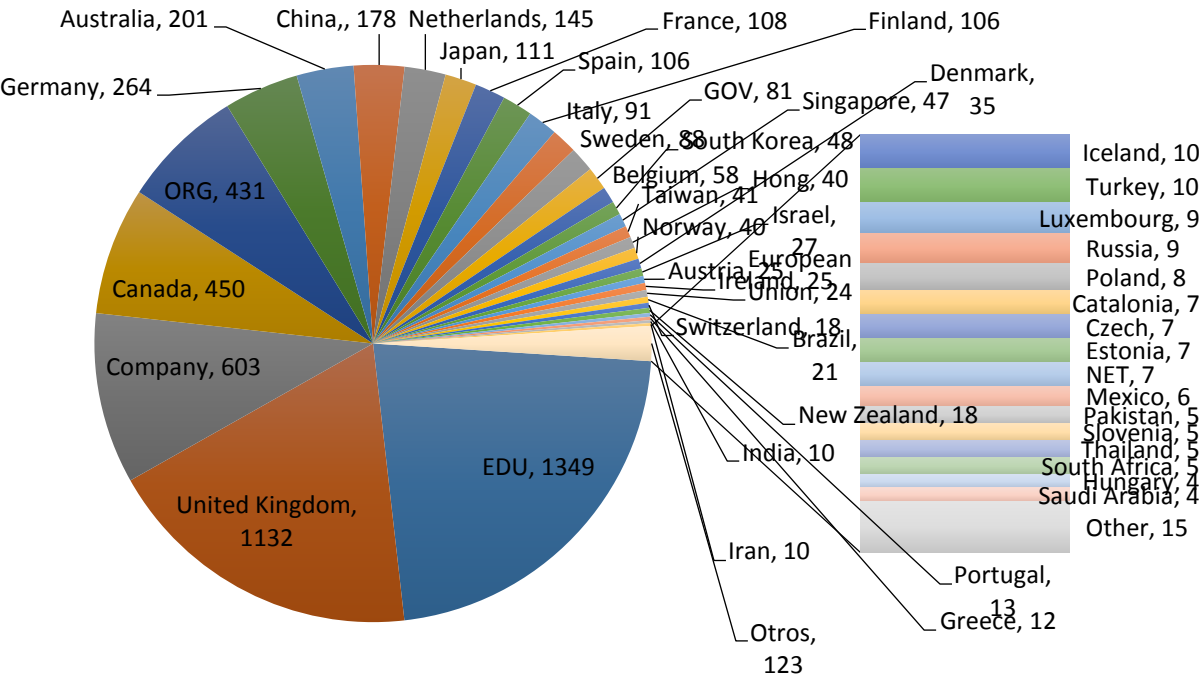
- > 700 Studies
- > 500,000 individuals
- > 1,200 Datasets
- > 250 Data providers
- > 6,000 Data Requesters

# The EGA in volume

- > 250,000 files
- > 1,5 Petabytes



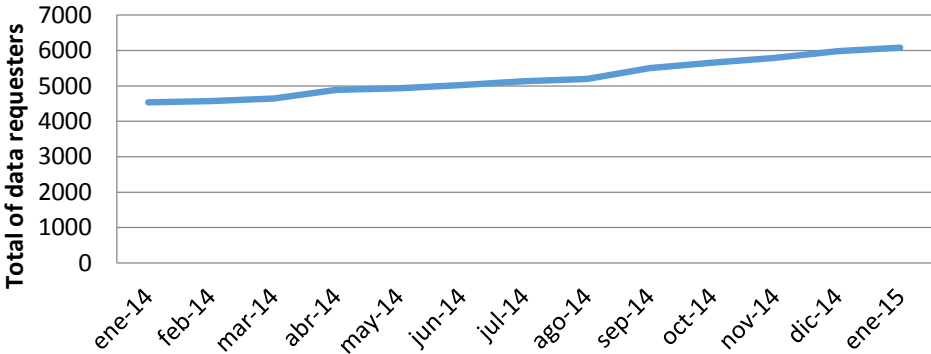
# Data Requesters



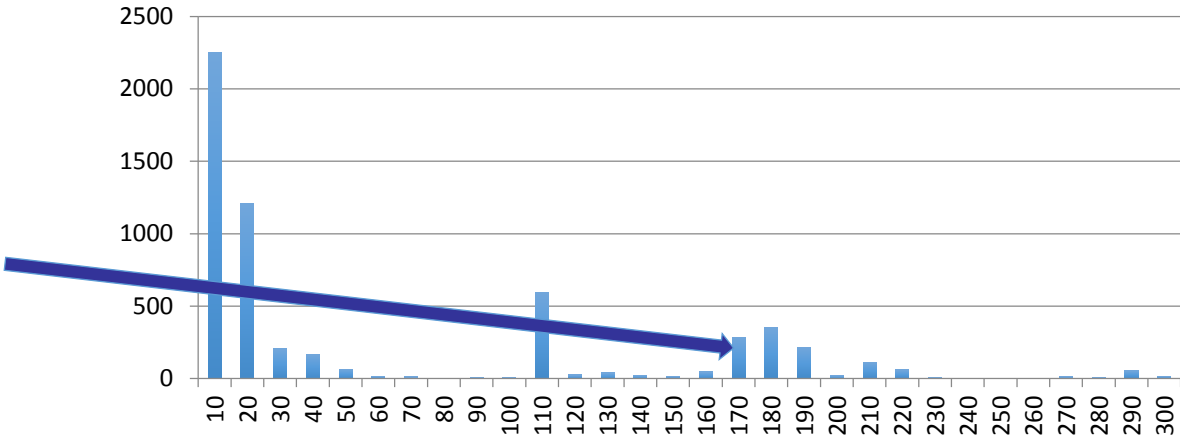
5877 total requesters  
986 ICGC users (16%) access to 185\* datasets  
Most of the users have access to 10 or less datasets

\*old users do not get access to new ICGC releases automatically

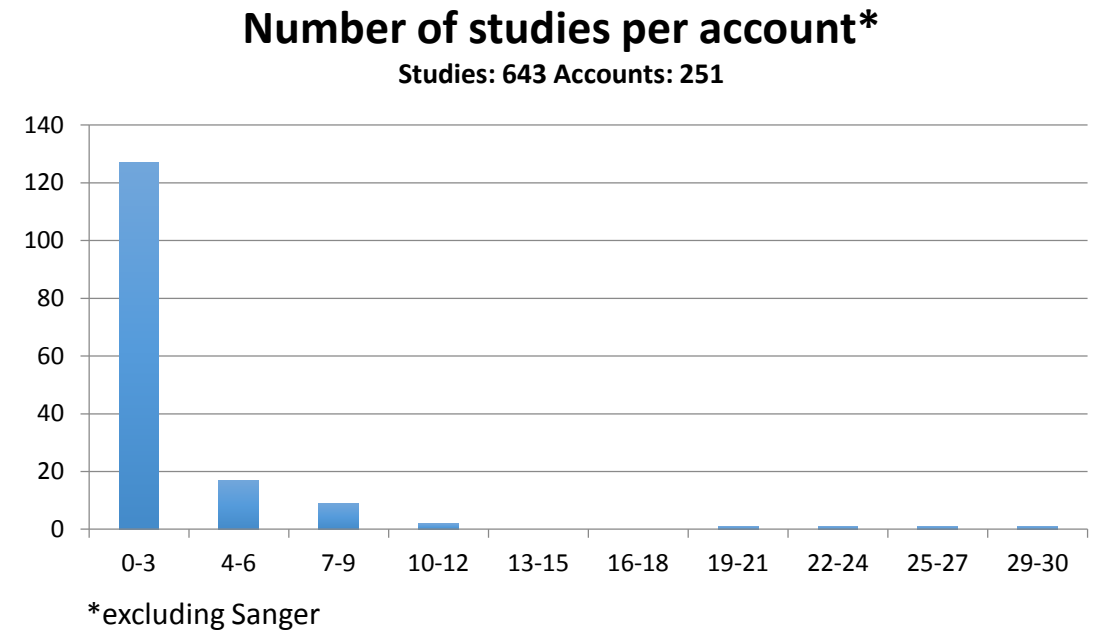
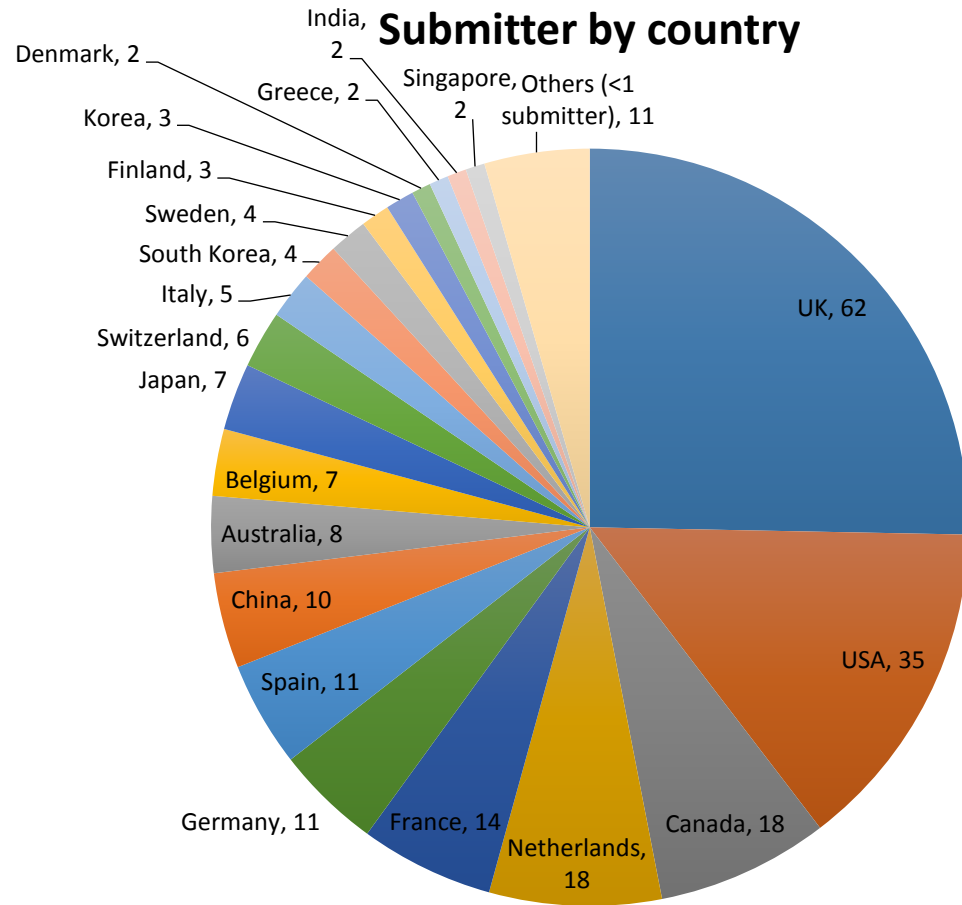
Last year's approved data requesters



Granted datasets per account

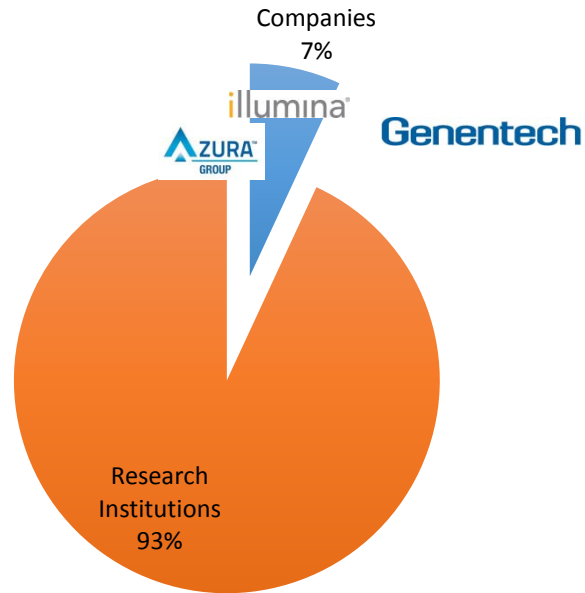


# Data Submitters

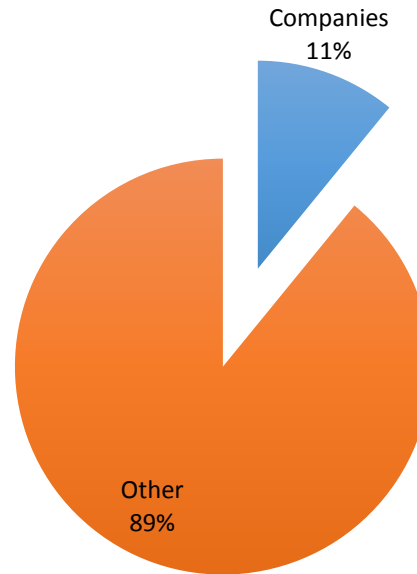




# Mostly non-profit projects, but clear transference too



Submitters  
188\*



Requesters  
1312\*

\*Figures are based on unique organizations, both public and private.  
Accounts belonging to the same organization are counted only once.

Company	Number of accounts
Eli Lilly	54
Astellas Pharma	25
Personalis	20
Pfizer	20
Merck-Serono	19
Genentech	18
Bayer	13
Astra-Zeneca	12
Takeda	11
Bina Techonologies	8

## Top 10 requesting companies.

Based on number of accounts

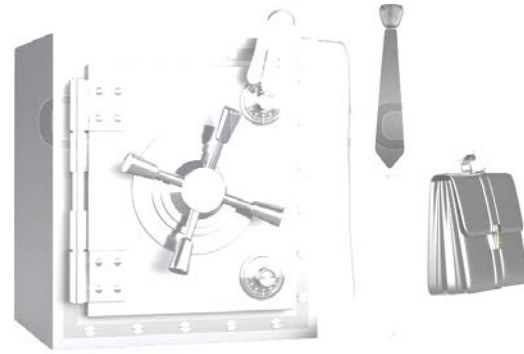
# Summary:

- Plenty of good data to mine (even if still difficult)
- Plenty of opportunities for accelerating research (even if just starting)

# Outline

## ○ 10' on global -Omics data

- Current status
- Recent and incoming novelties
  - Unified brand
  - EGA 2.0



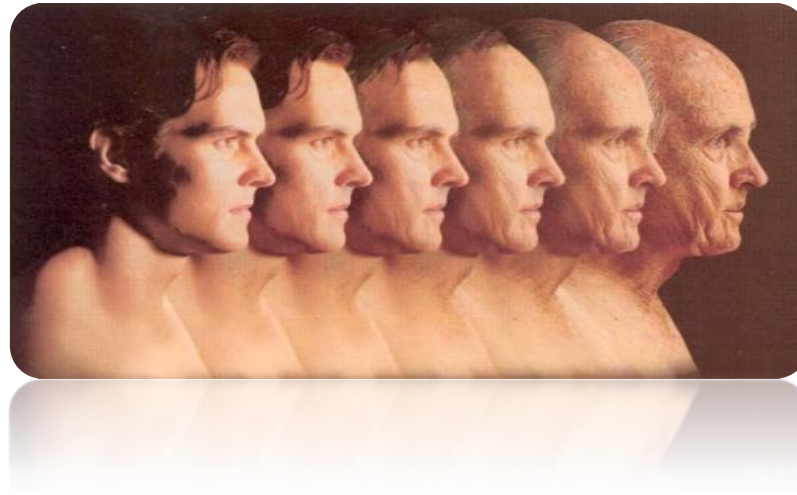
## ▪ 10' on a project done so far with very little data

- The pleiotropic theory of Senescence?
- Genome-Phenome data
- Genome-wide evidence for a theory of senescence



# Organismic **senescence** is decay

Deterioration of function, increase in mortality... It is **biological ageing** and also, among other effects, *increased disease risk with ageing*.



Late-onset complex diseases  $\leftrightarrow$  Senescence



# Senescence is General... or is it?

Most organisms deteriorate with age, but there are some clear exceptions



*Jordi Hurtado 1980s*



*Jordi Hurtado 1990s*



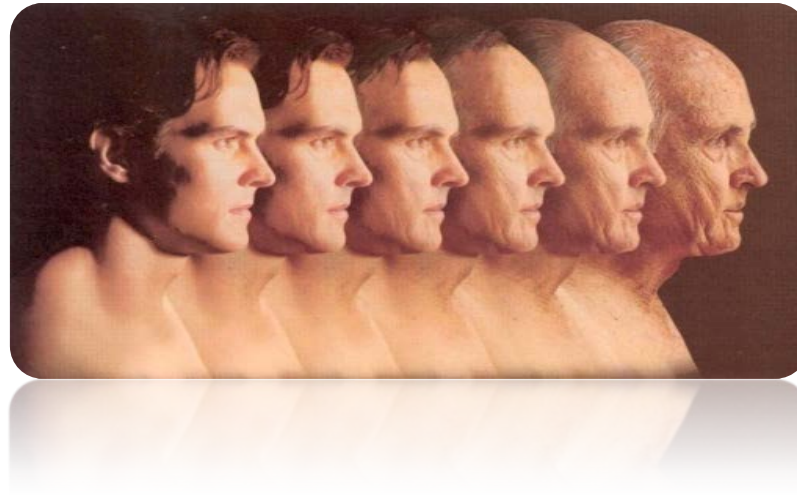
*Jordi Hurtado 2000s*



*Jordi Hurtado 2014*

# Organismic **senescence** is decay

Deterioration of function, increase in mortality... It is **biological ageing** and also, among other effects, *increased disease risk with ageing*.

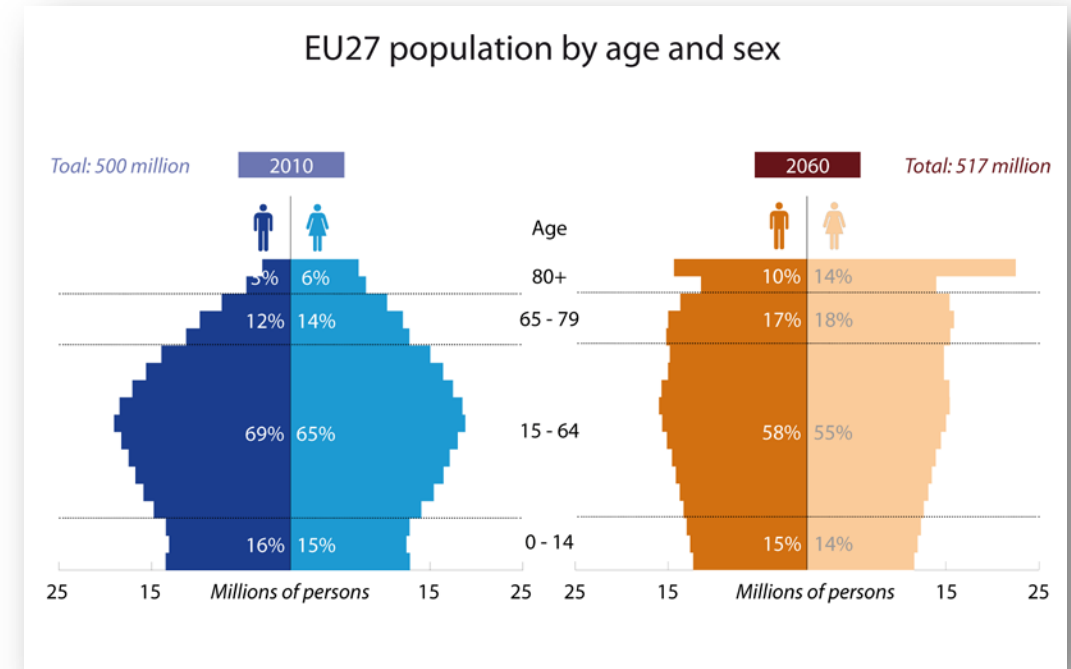


Late-onset complex diseases  $\leftrightarrow$  Senescence

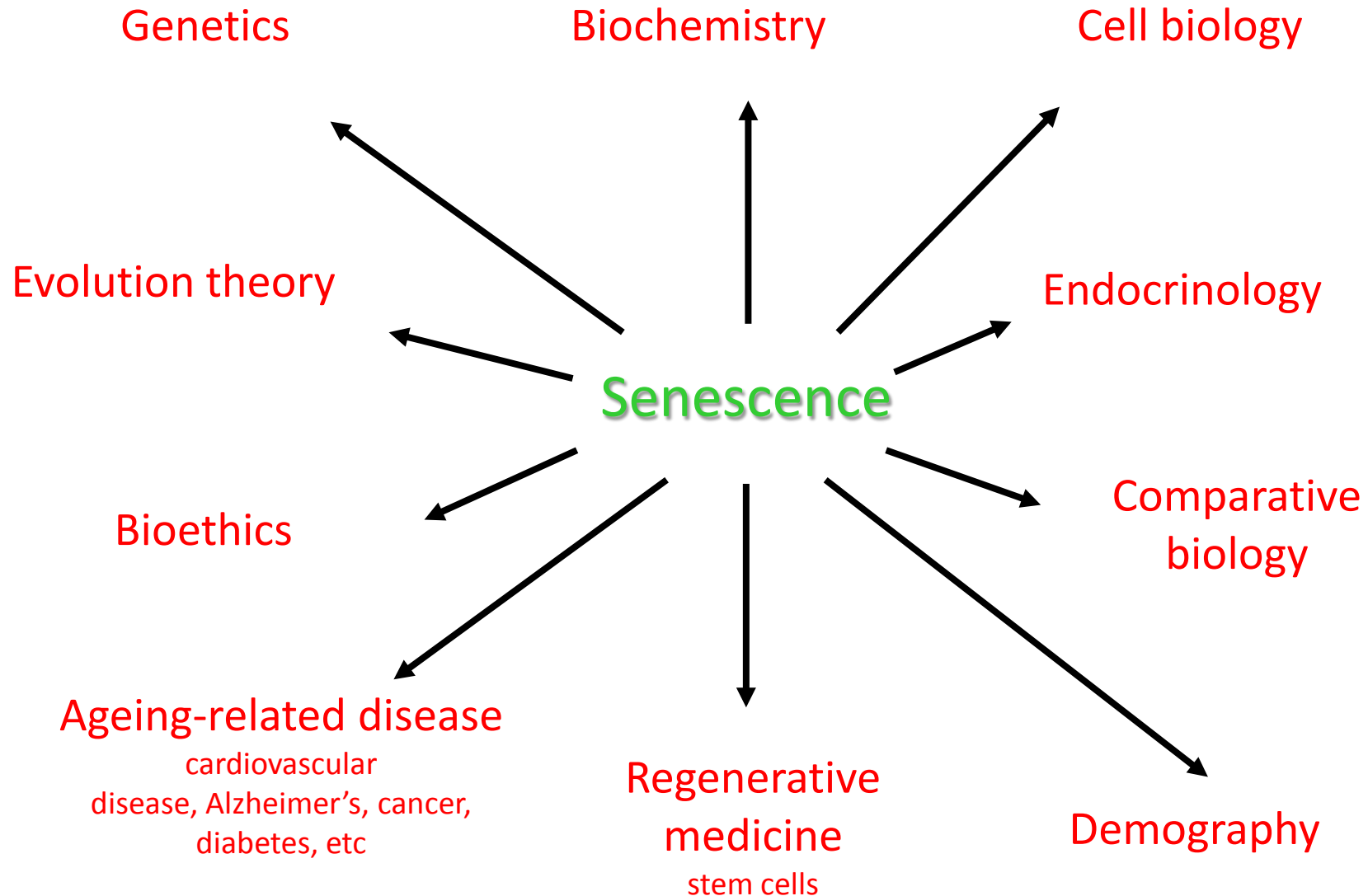
# A growing public health issue:



Mme Jeanne Calment, died 1998, aged 122



# With plenty of scientific implications





But in short, senescence is the leading  
cause of death and thus it is...

a public health issue:

- What can we expect in the near future?
- How to pay for it?

an evolutionary question:

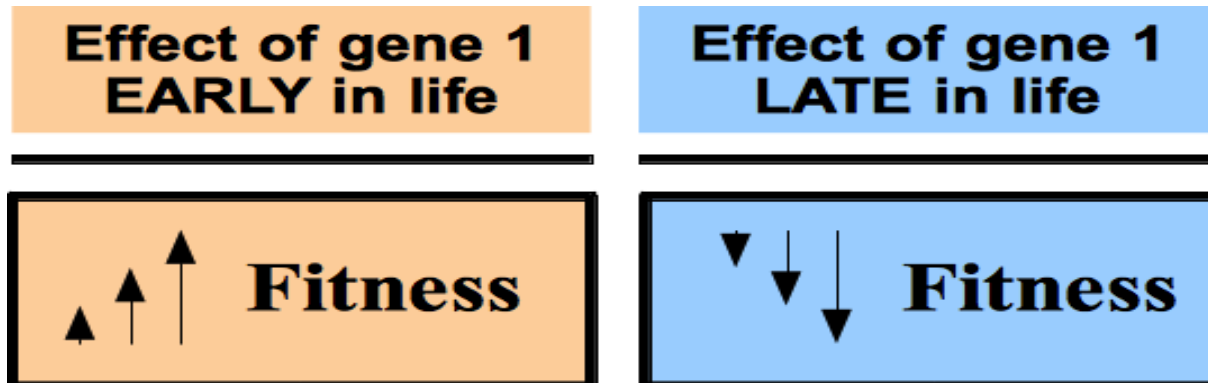
- How do we decay with age?
- *Why* are we senescent?

# The pleiotropic theory of senescence

An evolutionary theory of senescence (not the only one)

## The antagonistic pleiotropy theory

- **Pleiotropic** → mutations having effects on more than one trait
- **Antagonistic** → mutations having both, beneficial and harmful effects on the organism.

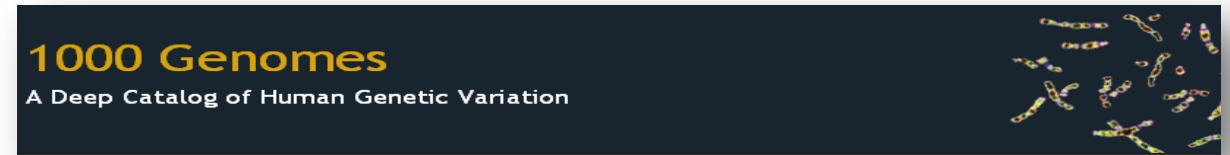
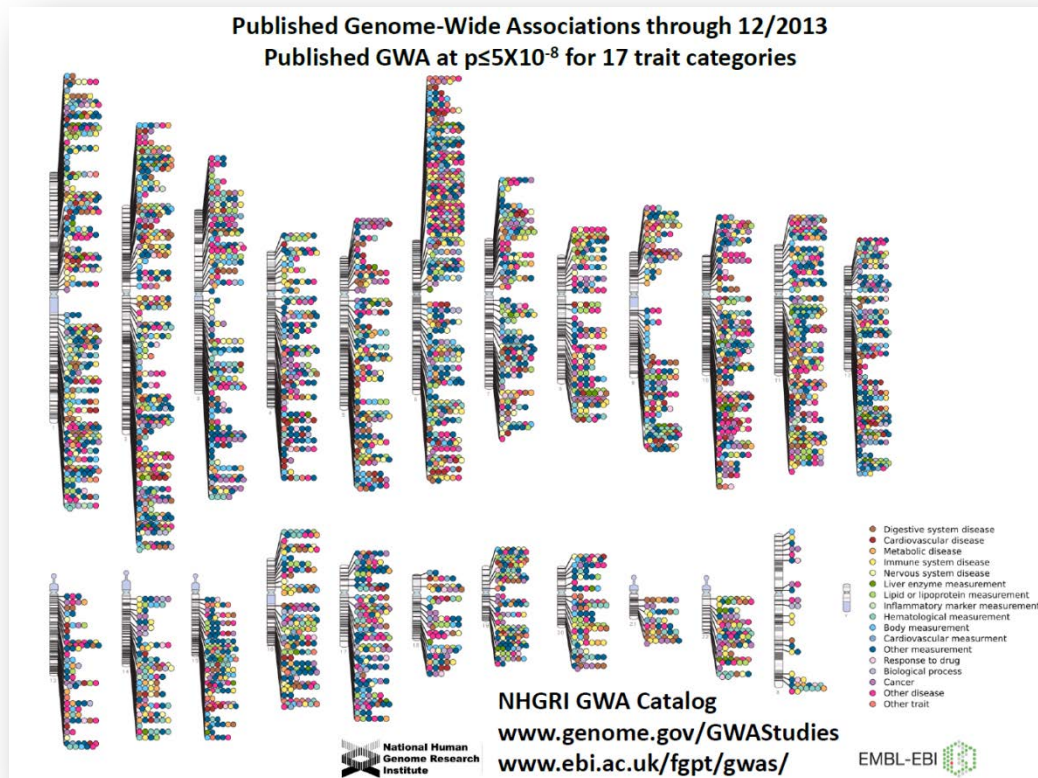


## Putting this theory in terms of health...

- Do we get sick when old because natural selection favoured health when young?
- That is, are there age-related antagonistic pleiotropies that link late onset disease and natural selection? And, if so:
  - *Up to which point?*
  - *Which diseases and genes are involved?*

Some freely accessible (i.e. non EGA) data:  
**tapping only about 1/100,000th of the SNPs available in the EGA**

The **GWAS catalogue** informs about the effects on complex disease of SNPs whose p-value is  $<10^{-5}$



The **HapMap** and the **1000 Genomes Project** inform us about population-specific allele frequencies and Linkage Disequilibrium patterns

We first classify each disease  
according to its **age of onset**

According to the info on the GWAS paper or Medscape or the WHO. For instance:

<i>Early Onset</i> ( $\leq 25$ years)	n = 43 diseases
<i>Late Onset</i> ( $> 25$ years)	n = 83 diseases

Of course, the definitions of what are evolutionary relevant early and late onsets diseases is uncertain, so we try several thresholds

Then we classify variants unveiled by GWAS according to **their effect** in **pairs of diseases**

If they increase risk of two diseases: **Agonist**

If increasing the risk of a disease while protecting from another: **Antagonist**



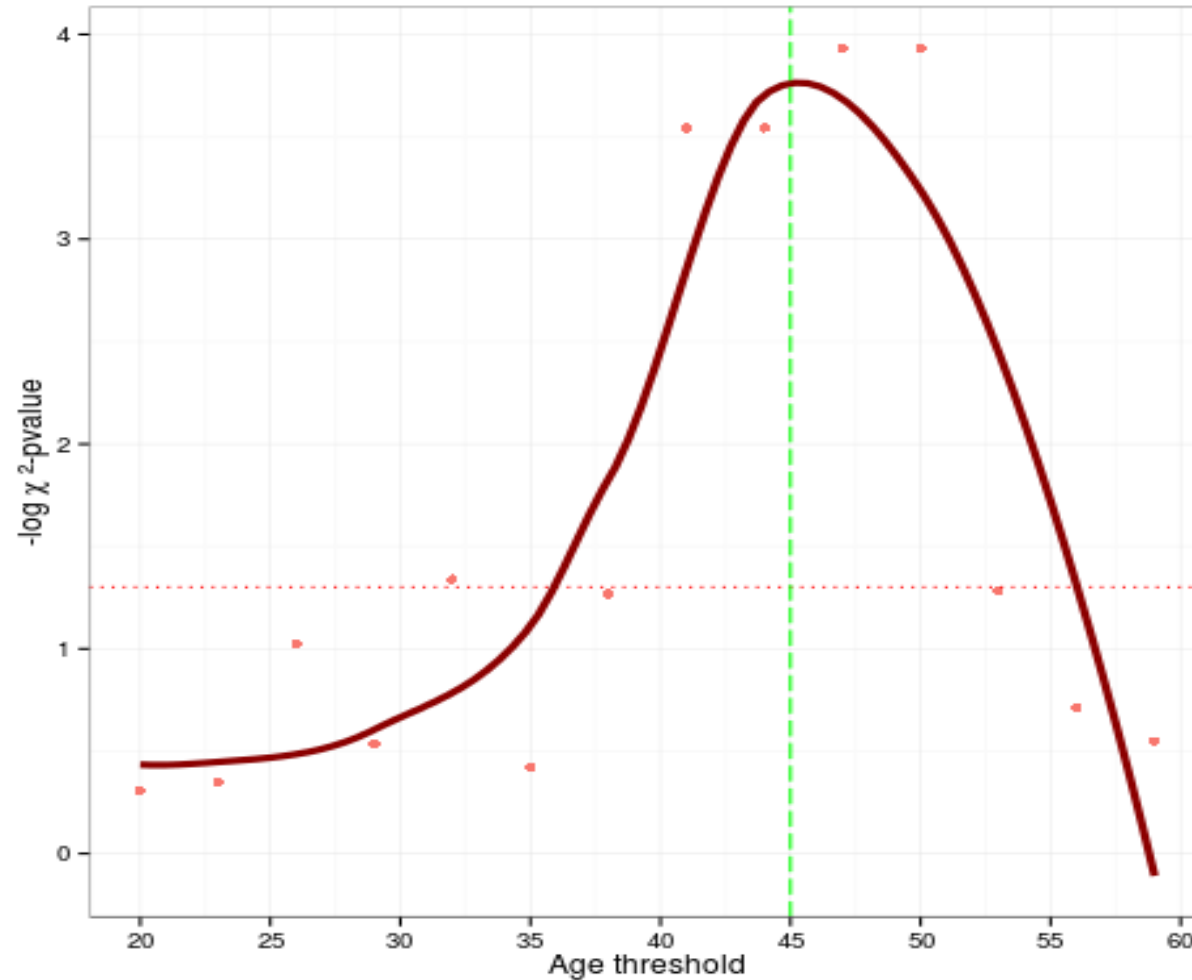
## A clear excess of Early-Late Antagonistic pleiotropies

$R^2 \geq 0.8$ 46 years thr. $X^2$ pv: 0.0001	ANTAGONIST	AGONIST
EARLY-EARLY / LATE-LATE	46	167
EARLY-LATE	26	27
Total	72	194

Comparing classes of pleiotropies: same period vs. different periods

A clearly relevant threshold **around age ~45**

0.05 significance



Distribution of X<sup>2</sup> pvalues depending on the age threshold

But... is this solid?

**After all, it is just data mining and hand waving!!**

# Independently ascertained “Ageing” genes harbor an excess of pleiotropies

135 genes from Sousa et al. 2014 Nature 506: 316–321

ARTICLE

doi:10.1038/nature13013

Geriatric muscle stem cells switch reversible quiescence into senescence

Pedro Sousa-Victor<sup>1†</sup>, Susana Gutarra<sup>1\*</sup>, Laura García-Prat<sup>1\*</sup>, Javier Rodríguez-Ubreva<sup>2</sup>, Laura Ortet<sup>1</sup>, Vanessa Ruiz-Bonilla<sup>1</sup>, Mercè Jardi<sup>1</sup>, Esteban Ballestar<sup>2</sup>, Susana González<sup>1</sup>, Antonio L. Serrano<sup>1</sup>, Eusebio Perdiguero<sup>1</sup> & Pura Muñoz-Cánoves<sup>1,4</sup>

Regeneration of skeletal muscle depends on a population of adult stem cells (satellite cells) that remain quiescent throughout life. Satellite cell regenerative functions decline with ageing. Here we report that geriatric satellite cells are incapable of maintaining their normal quiescent state in muscle homeostatic conditions, and that this irreversibly affects their intrinsic regenerative and self-renewal capacities. In geriatric mice, resting satellite cells lose reversible quiescence by switching to an irreversible pre-senescence state, caused by derepression of p16<sup>INK4a</sup> (also called Cdkn2a). On injury, these cells fail to activate and expand, undergoing accelerated entry into a full senescence state (geroconversion), even in a youthful environment. p16<sup>INK4a</sup> silencing in geriatric satellite cells restores quiescence and muscle regenerative functions. Our results demonstrate that maintenance of quiescence in adult life depends on the active repression of senescence pathways. As p16<sup>INK4a</sup> is dysregulated in human geriatric satellite cells, these findings provide the basis for stem-cell rejuvenation in sarcopenic muscles.

	Full Genome	135 genes from Sousa et al (2014)
Disease associated SNPs	2,666	78
Pleiotropies	266	34
p-val = 1.69·10 <sup>-14</sup>		

298 genes from Magalhaes et al., 2009 Bioinformatics 25: 875-881

BIOINFORMATICS ORIGINAL PAPER

Vol. 25 no. 7 2009, pages 875–881  
doi:10.1093/bioinformatics/btp073

Gene expression

Meta-analysis of age-related gene expression profiles identifies common signatures of aging

João Pedro de Magalhães<sup>1,\*†</sup>, João Curado<sup>2</sup> and George M. Church<sup>1</sup>

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA and <sup>2</sup>Escola Superior de Biotecnologia, 4200 Porto, Portugal

Received on September 15, 2008; revised on January 11, 2009; accepted on January 31, 2009

Advance Access publication February 2, 2009

Associate Editor: David Rocke

	Full Genome	298 genes from Magalhaes et al (2009)
Disease associated SNPs	2,666	156
Pleiotropies	266	55
p-val = 1.9·10 <sup>-15</sup>		

# SNPs and diseases involved in the 26 Antagonistic Pleiotropies

A clear excess of Early-Late Antagonistic pleiotropies

$R^2 \geq 0.8$ 46 years thr. $\chi^2$ pv: 0.0001	ANTAGONIST	AGONIST
EARLY-EARLY / LATE-LATE	46	167
EARLY-LATE	26	27
Total	72	194

Comparing classes of pleiotropies: same period vs. different periods

EARLY DISEASE	LATE DISEASE
Multiple sclerosis	Age-related mac. degeneration
Duodenal ulcer	Bladder cancer
Bipolar dis. and schizophrenia	Breast cancer
Restless legs syndrome	Breast cancer
Type 2 diabetes	Breast cancer
Celiac disease	Colorectal cancer
Duodenal ulcer	Coronary heart disease
Glioma	Coronary heart disease
Graves' disease	Coronary heart disease
Migraine	Coronary heart disease
Ulcerative colitis	Dupuytren's disease
Type 2 diabetes	Endometrial cancer
Multiple sclerosis	Hepatocellular carcinoma
Crohn's disease	Hypothyroidism
Thyroid cancer	Hypothyroidism
Glioma	Idiopathic pulmonary fibrosis
Lung cancer	Idiopathic pulmonary fibrosis
Lung cancer	Melanoma
Vitiligo	Melanoma
Bipolar disorder	Osteoarthritis
Meningioma	Ovarian cancer
Duodenal ulcer	Pancreatic cancer
Graves' disease	Pancreatic cancer
Lung cancer	Pancreatic cancer
Psoriasis	Parkinson's disease
Type 2 diabetes	Prostate cancer

# SNPs and diseases involved in the 26 Antagonistic Pleiotropies

A clear excess of Early-Late Antagonistic pleiotropies

$R^2 \geq 0.8$ 46 years thr. $\chi^2$ p< 0.0001	ANTAGONIST	AGONIST
EARLY-EARLY / LATE-LATE	46	167
EARLY-LATE	26	27
Total	72	194

Comparing classes of pleiotropies: same period vs. different periods

Consistent with comorbidity analysis, but some of these relationships **never revealed** by it

EARLY DISEASE	LATE DISEASE
Multiple sclerosis	Age-related mac. degeneration
Duodenal ulcer	Bladder cancer
Bipolar dis. and schizophrenia	Breast cancer
Restless legs syndrome	Breast cancer
Type 2 diabetes	Breast cancer
Celiac disease	Colorectal cancer
Duodenal ulcer	Coronary heart disease
Glioma	Coronary heart disease
Graves' disease	Coronary heart disease
Migraine	Coronary heart disease
Ulcerative colitis	Dupuytren's disease
Type 2 diabetes	Endometrial cancer
Multiple sclerosis	Hepatocellular carcinoma
Crohn's disease	Hypothyroidism
Thyroid cancer	Hypothyroidism
Glioma	Idiopathic pulmonary fibrosis
Lung cancer	Idiopathic pulmonary fibrosis
Lung cancer	Melanoma
Vitiligo	Melanoma
Bipolar disorder	Osteoarthritis
Meningioma	Ovarian cancer
Duodenal ulcer	Pancreatic cancer
Graves' disease	Pancreatic cancer
Lung cancer	Pancreatic cancer
Psoriasis	Parkinson's disease
Type 2 diabetes	Prostate cancer

## In conclusion...

- George G. Williams was right (as usual): there is an excess of putative antagonistic pleiotropies that fits the pleiotropic theory of ageing
- Age ~45 seems has been relevant in the evolution of our species
- Plenty of unobvious relationships between diseases
- We are just beginning...

... imagine if we had used full datasets!!!