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

*iCrea
INSTITUCIÓ CATALANA DE RECERCA I ESTUDIS AVANÇATS
IBE CSIC UPF INSTITUT de BIOLOGIA EVOLUTIVA
INB
prbb Barcelona Biomedical Research Park
CRG Centre for Genomic Regulation
upf. Universitat Pompeu Fabra Barcelona

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

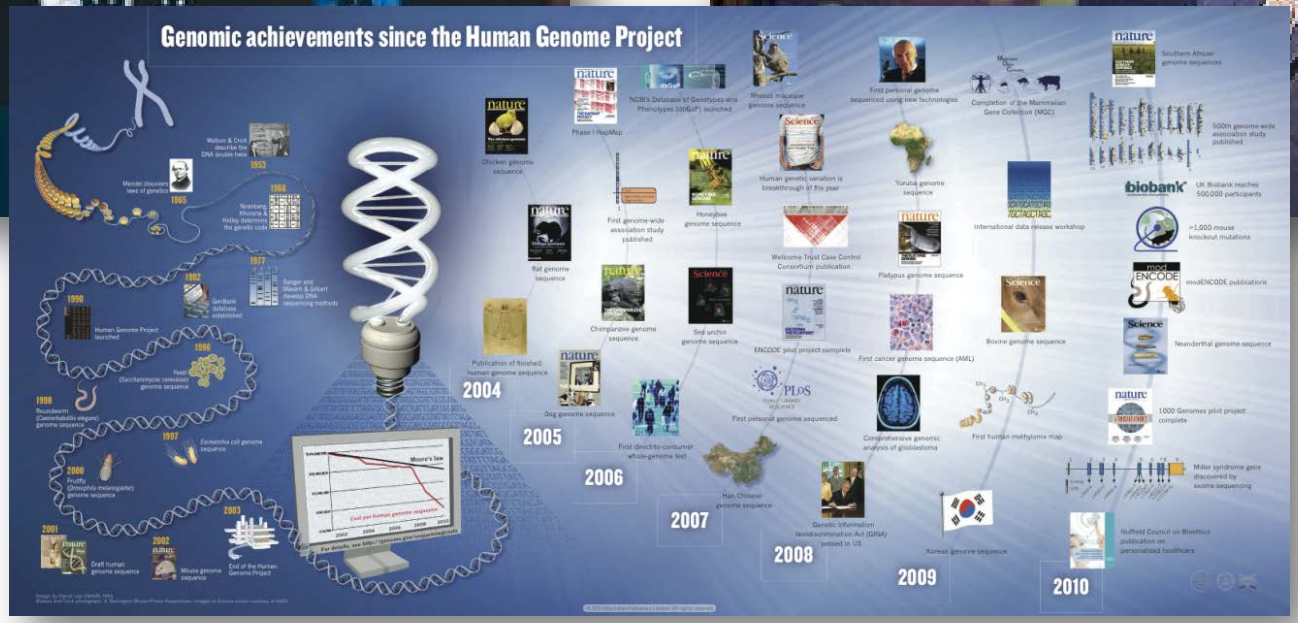
o10’ on global -Omics data

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

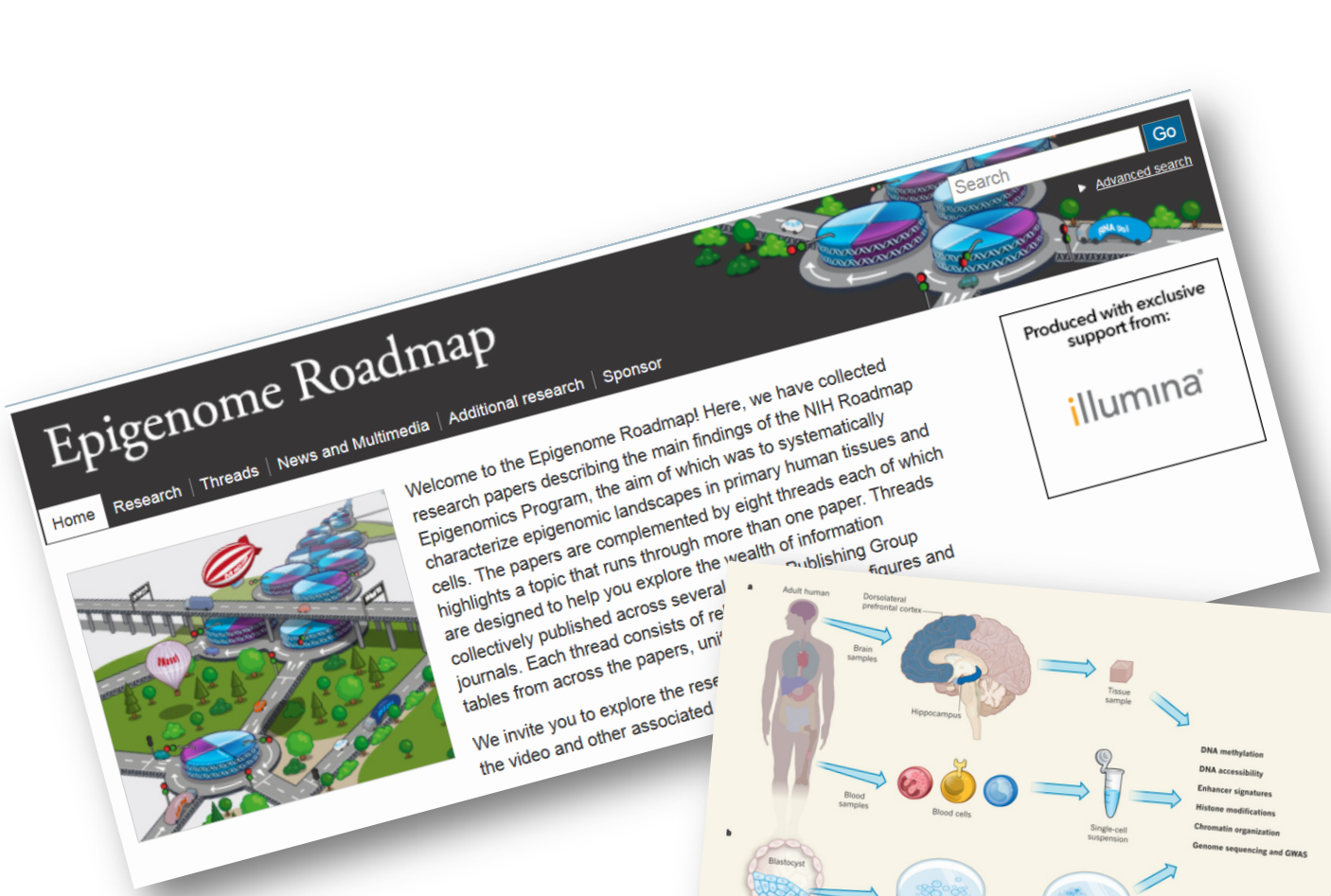
o10’ on a project done so far with very little data

- 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

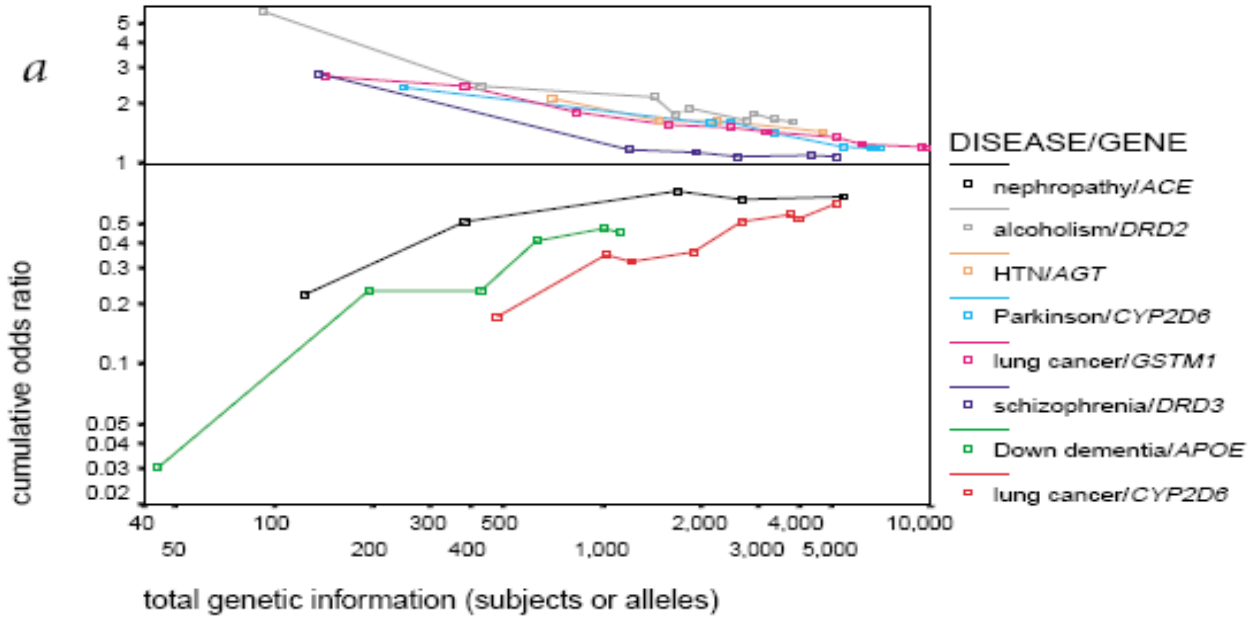


Genome Wide Association Studies (GWAS) have unveiled hundreds of SNPs associated to complex disease –*Omics* data has gone BIG

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



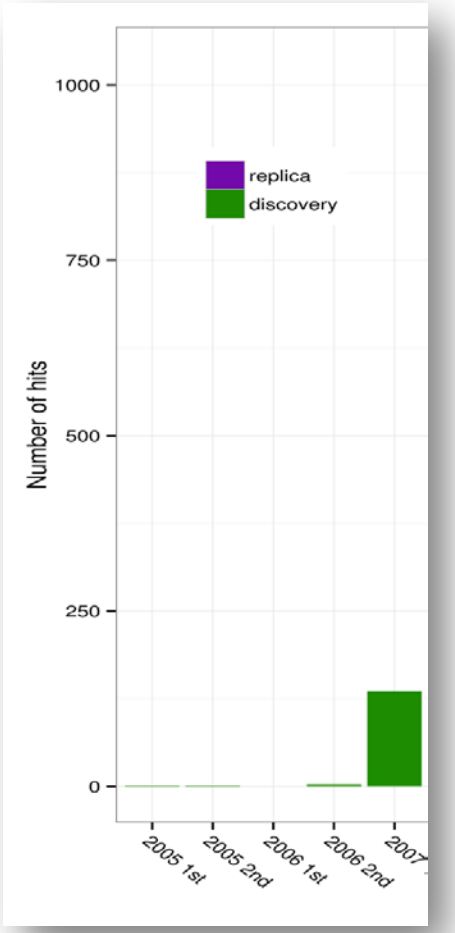
Replication validity of genetic association studies

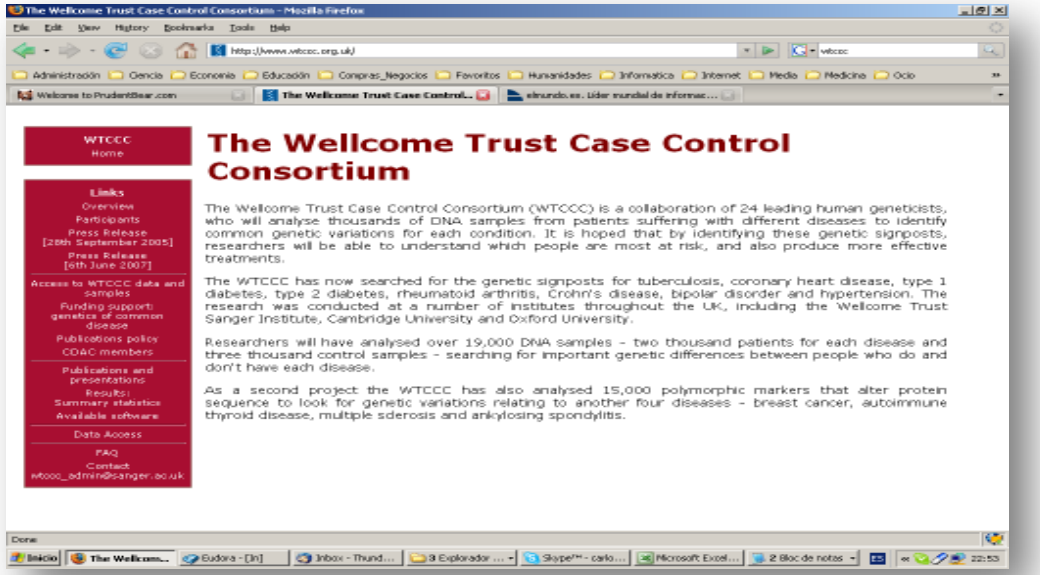
John P.A. Ioannidis[1–3], Evangelia E. Ntzani[1], Thomas A. Trikalinos[1] & Despina G. Contopoulos-Ioannidis[1,4]

nature genetics • volume 29 • november 2001

Have GWAS been replicated?

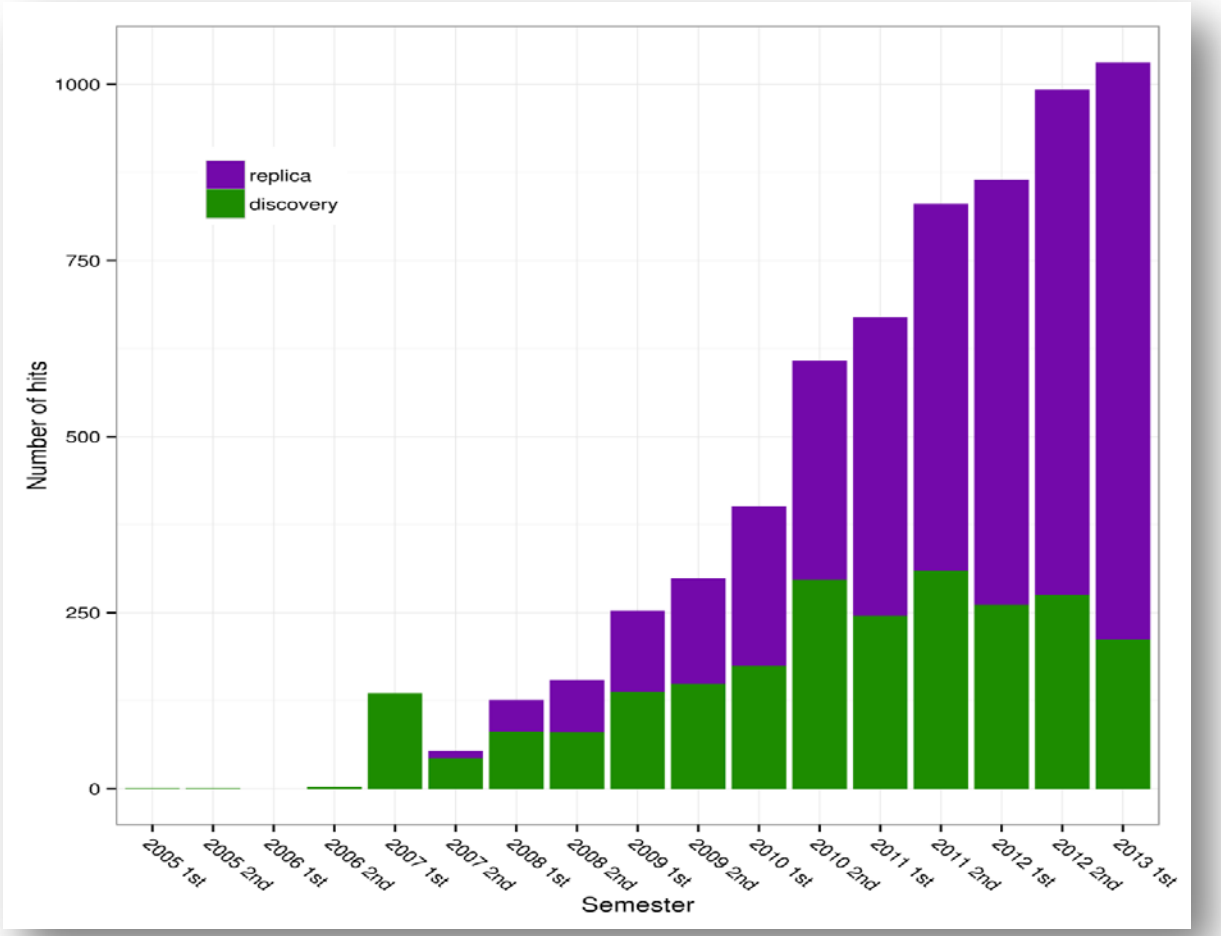
June 2007, publication of the WTCCC





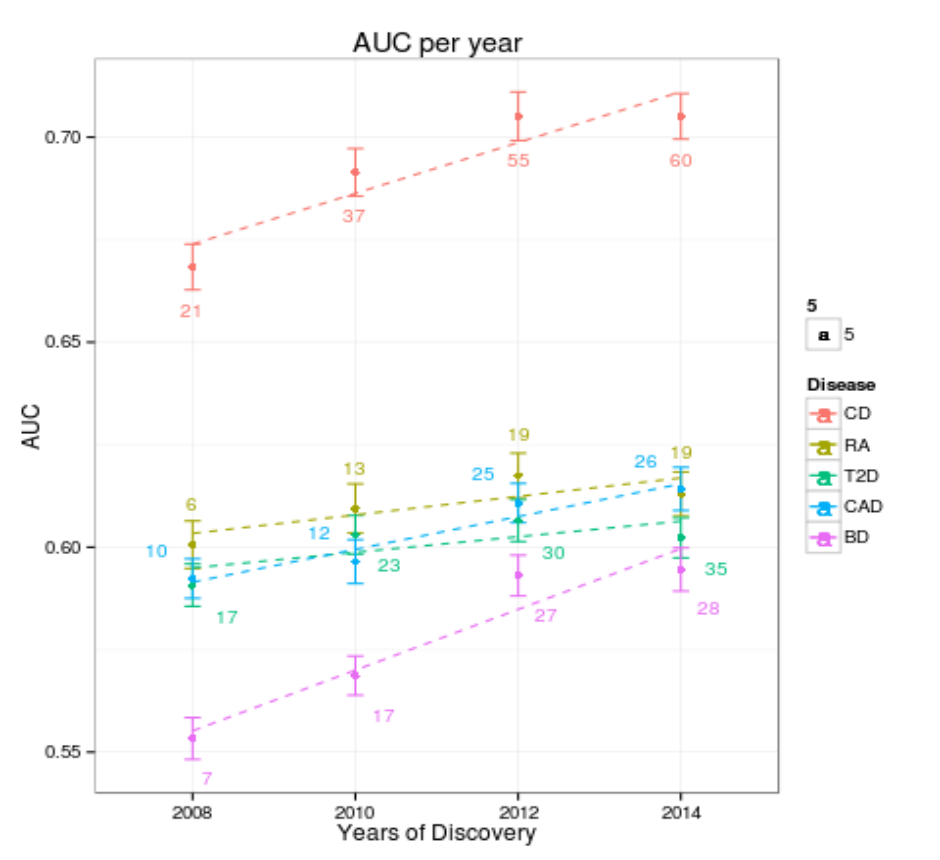
Have GWAS been replicated?

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

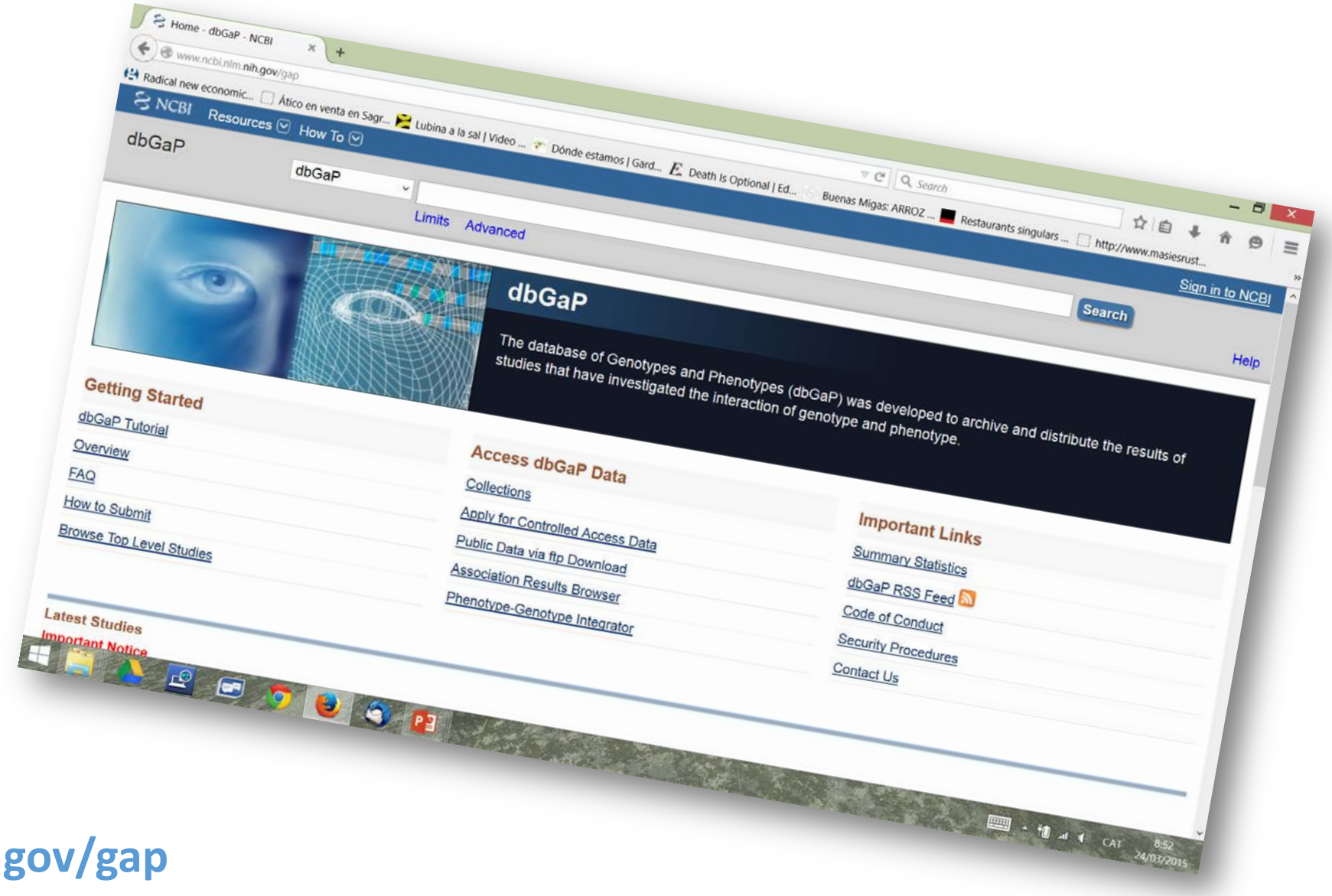


And prediction ability from GWAS results (i.e. variance explained) keeps increasing with time

[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

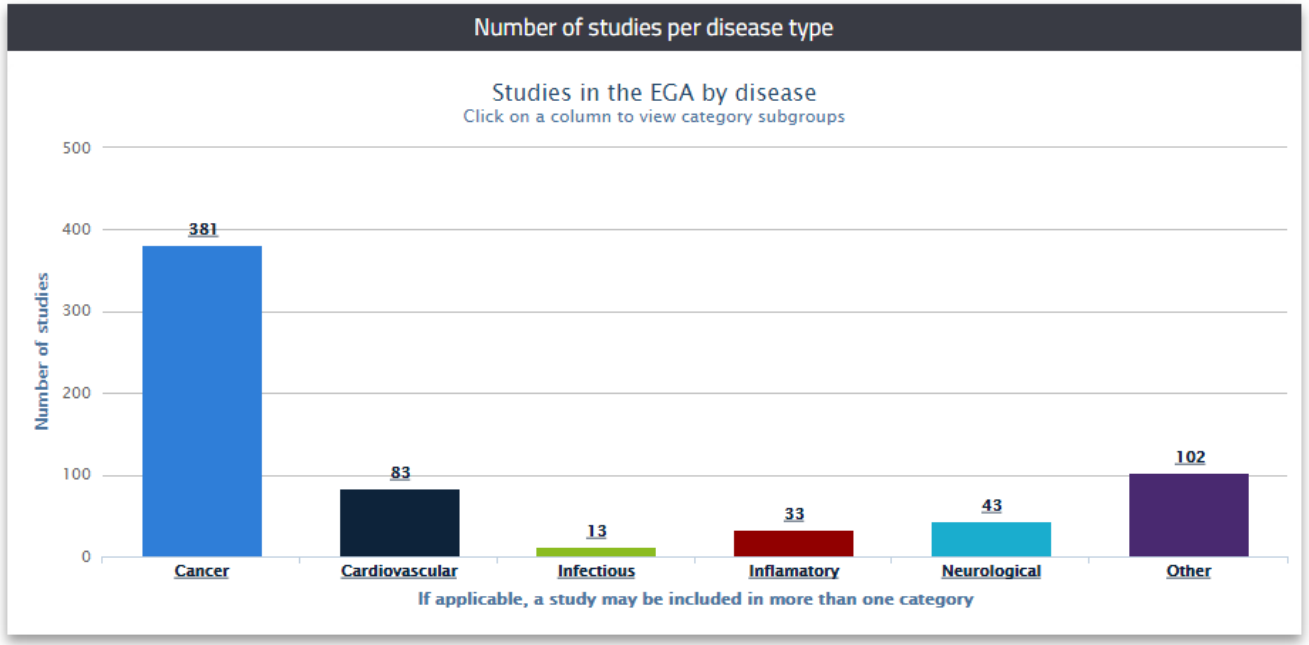
http://ega.crg.eu
http://www.ebi.ac.uk/ega

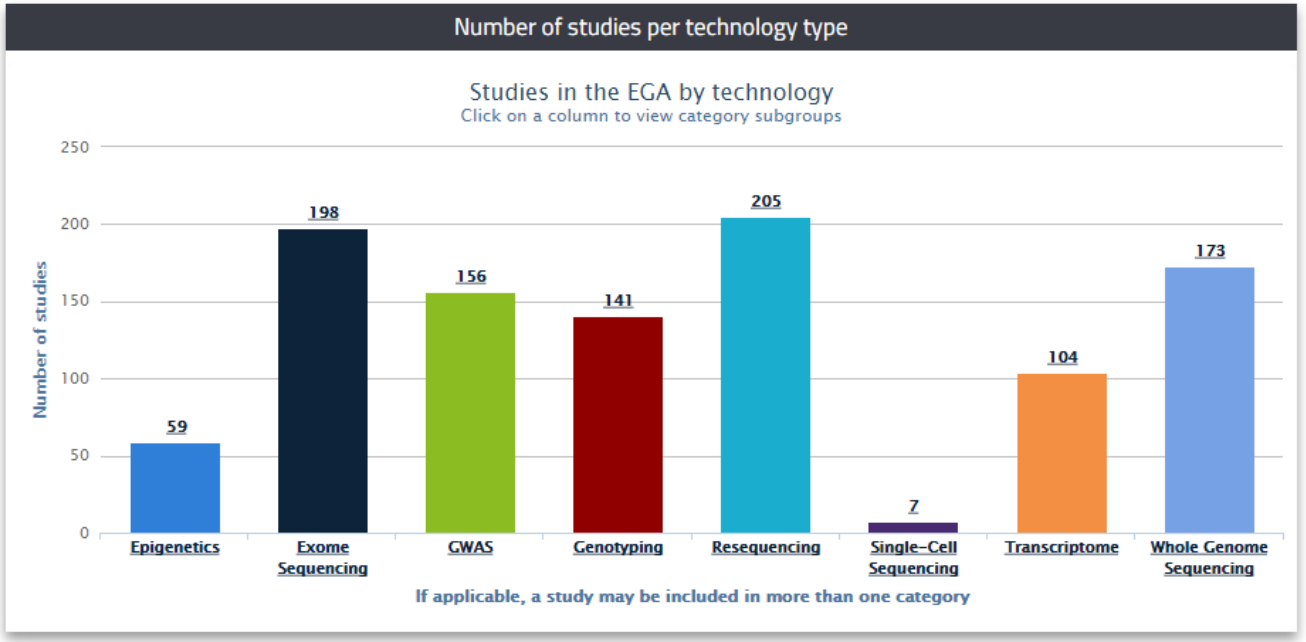
The EGA in numbers

- > 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



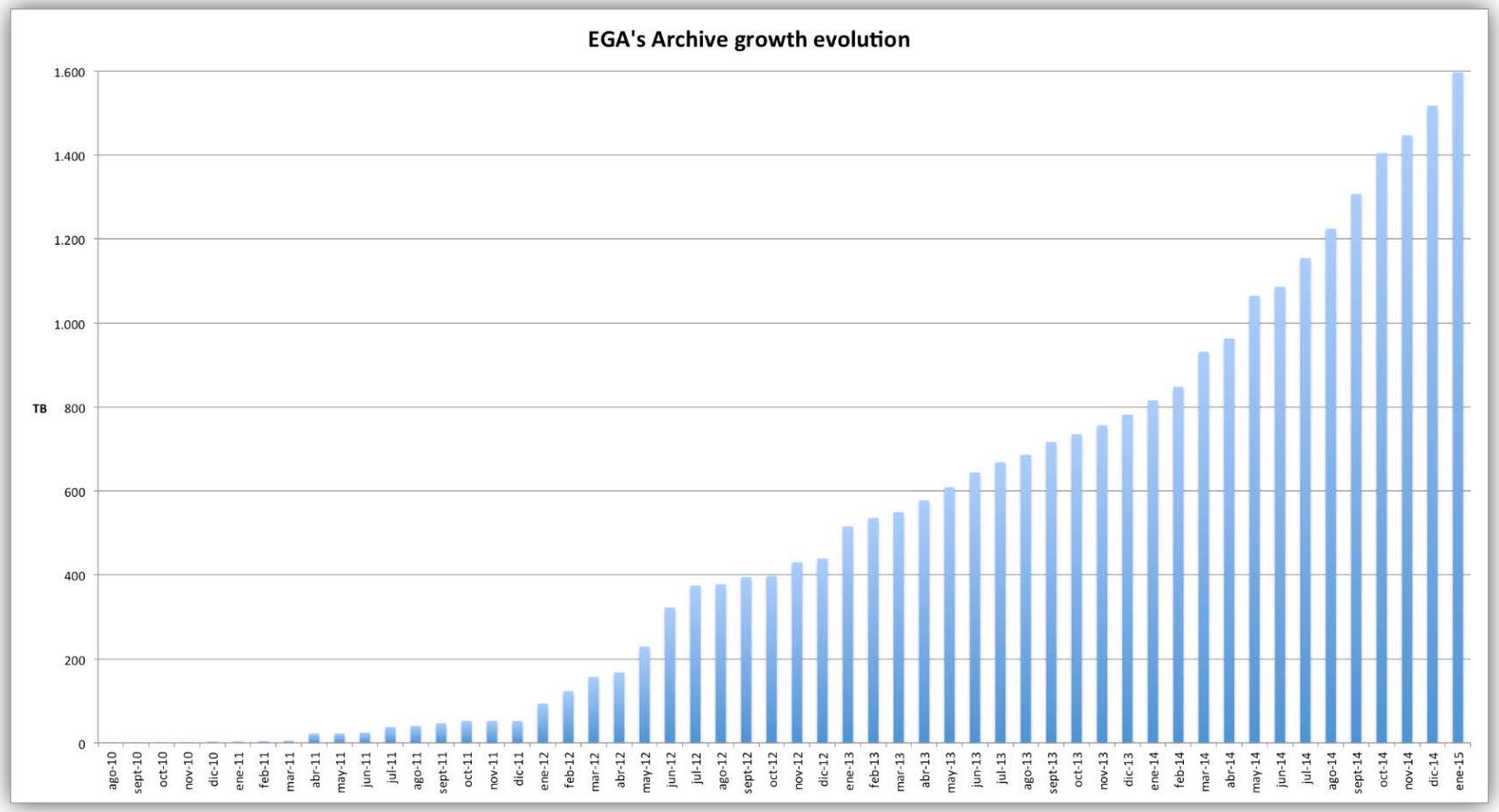


The EGA in numbers

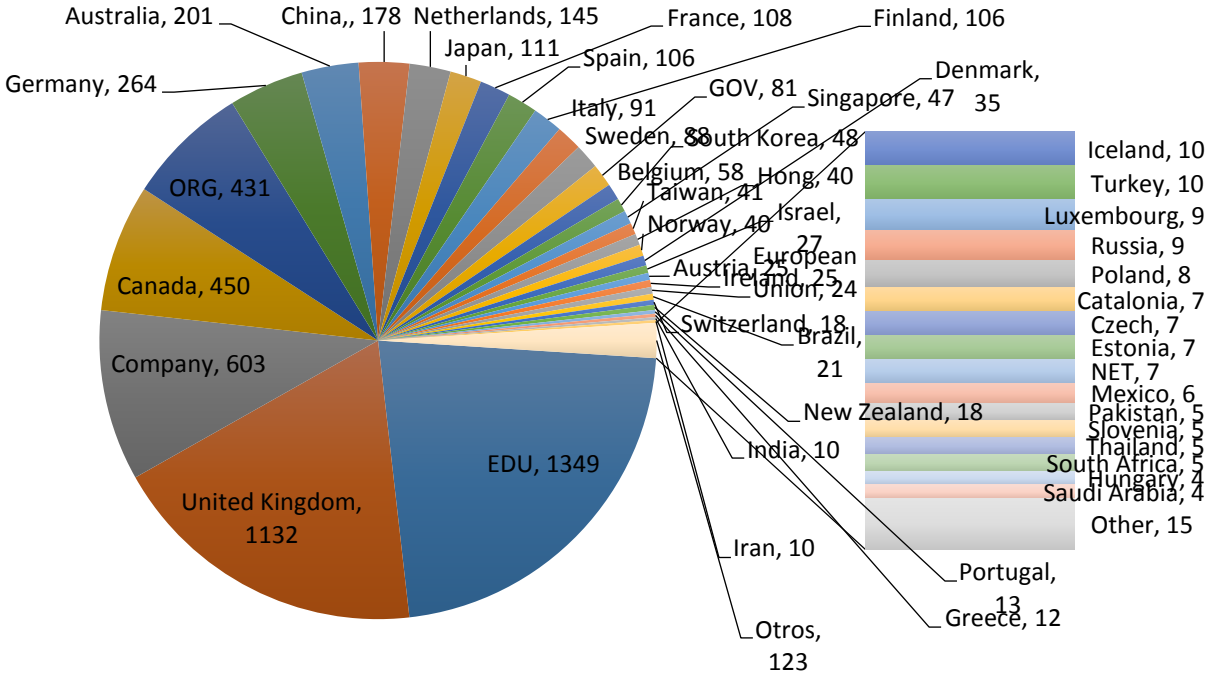
- > 700 Studies
- > 500,000 individuals
- > 1,200 Datasets
- > 250 Data providers
- > 6,000 Data Requesters

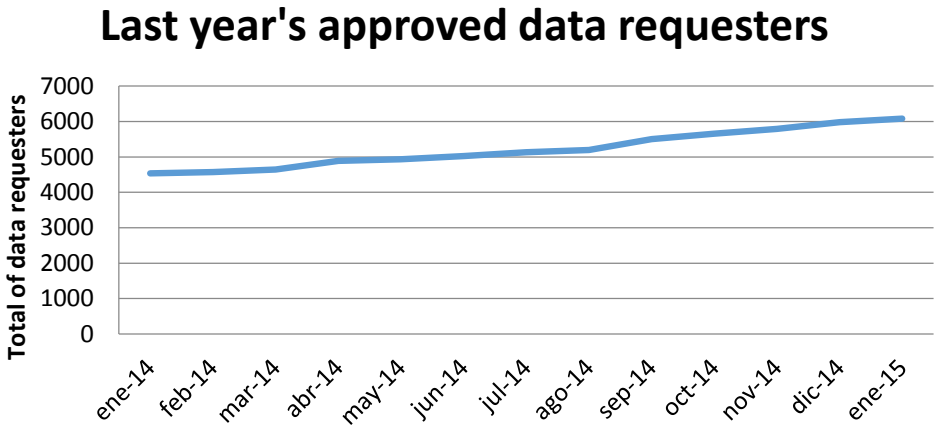
The EGA in volume

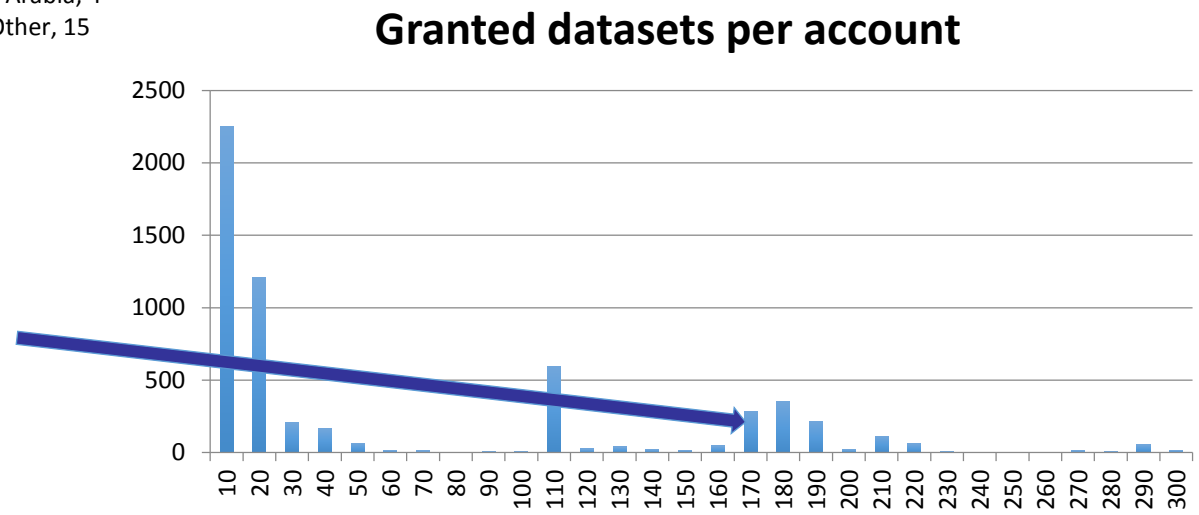
- > 250,000 files
- > 1,5 Petabytes



Data Requesters



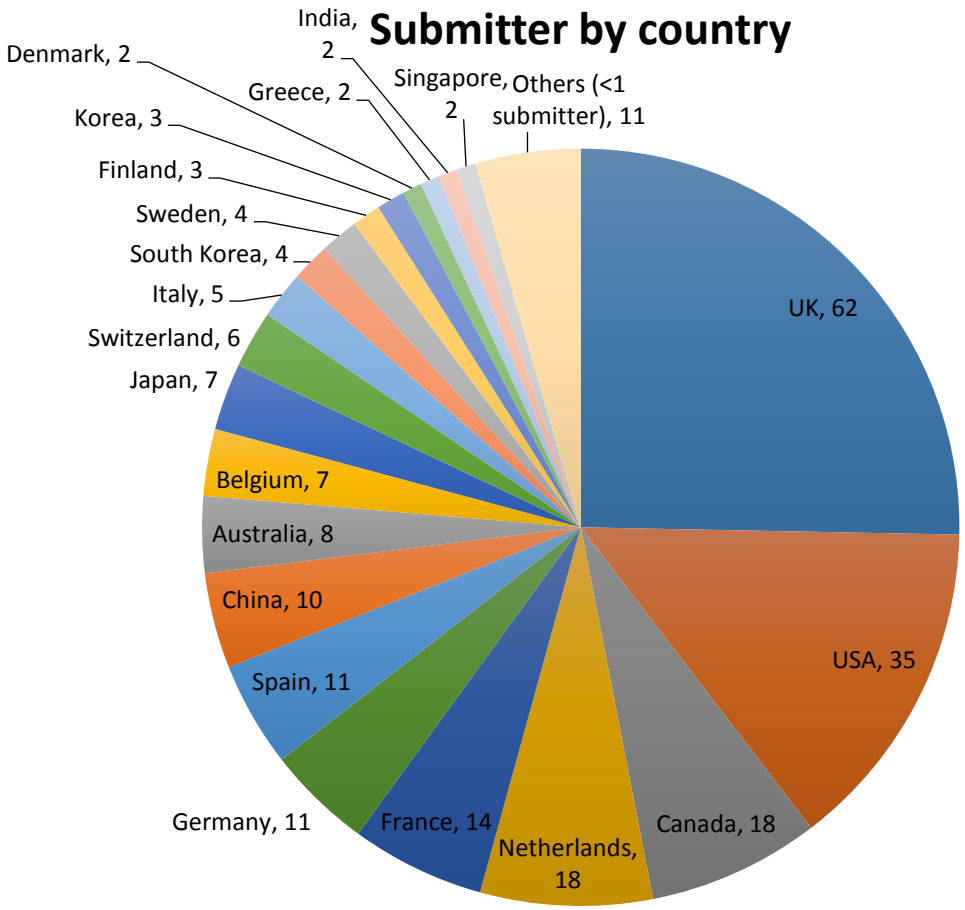


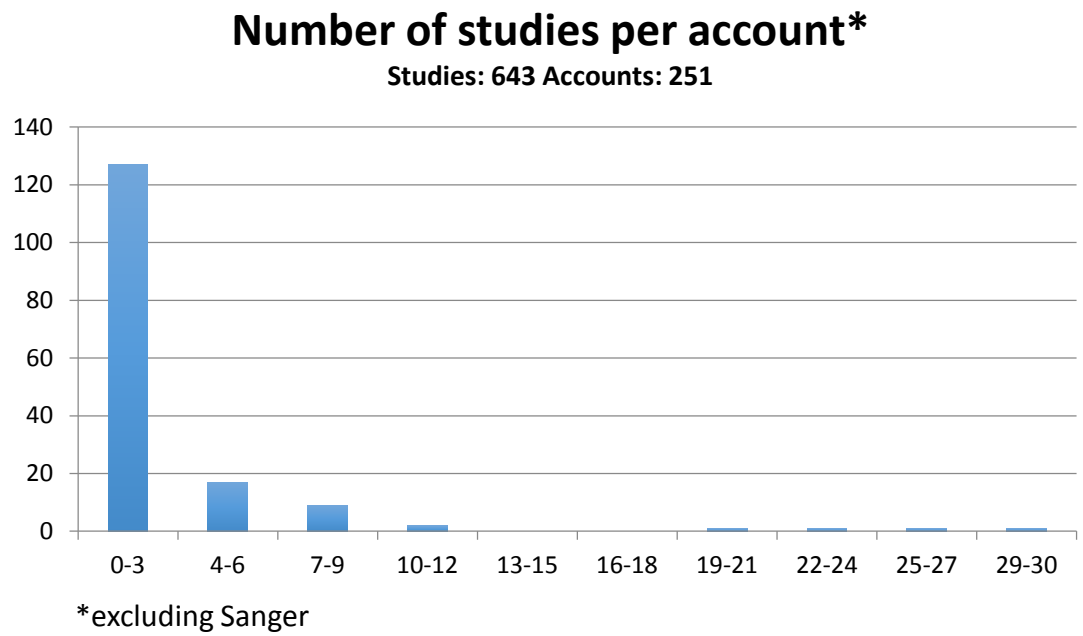


5877 total requesters
986 ICGC users (16%) access to 185* datasets
Most of the users have access to 10 or less datasets

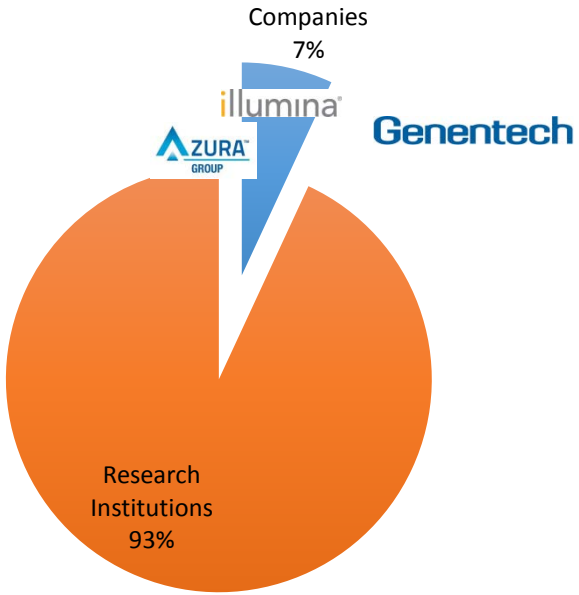
*old users do not get access to new ICGC releases automatically

Data Submitters

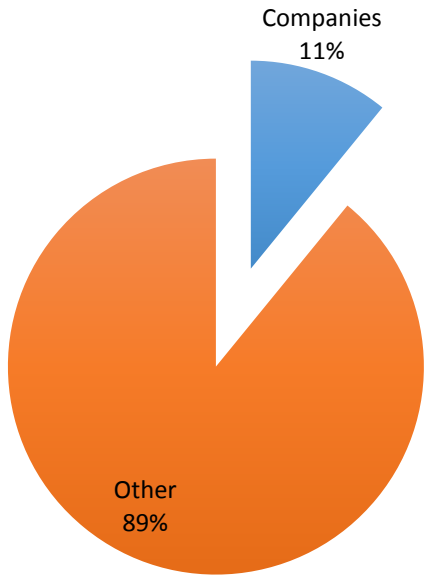




Mostly non-profit projects, but clear transference too



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)

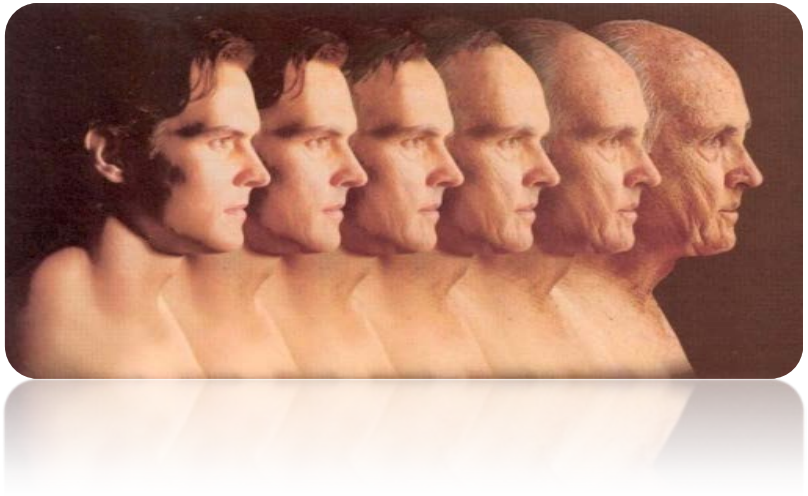
Outline

- 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 ⟷Senescence

Senescence is General… or is it?

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

Jordi Hurtado 1980s

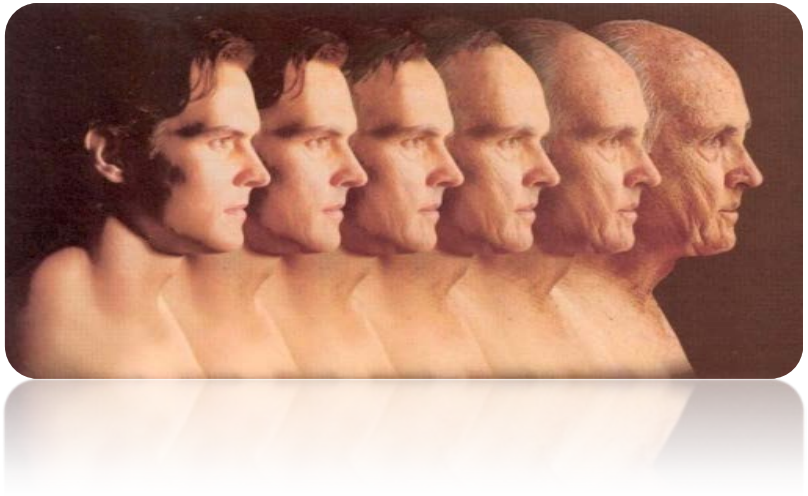
Jordi Hurtado 1990s

Jordi Hurtado 2000s

Jordi Hurtado 2014

Organismic **senescence** is decay

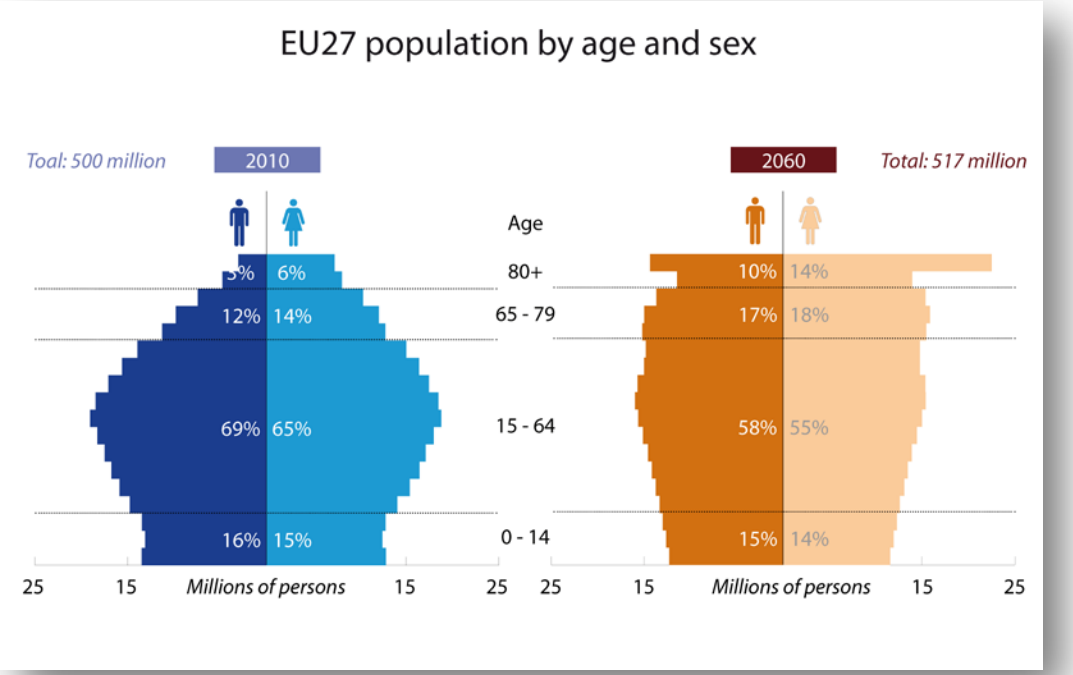
Deterioration of function, increase in mortality… It is **biological ageing** and also, among other effects, ***increased disease risk with ageing.***



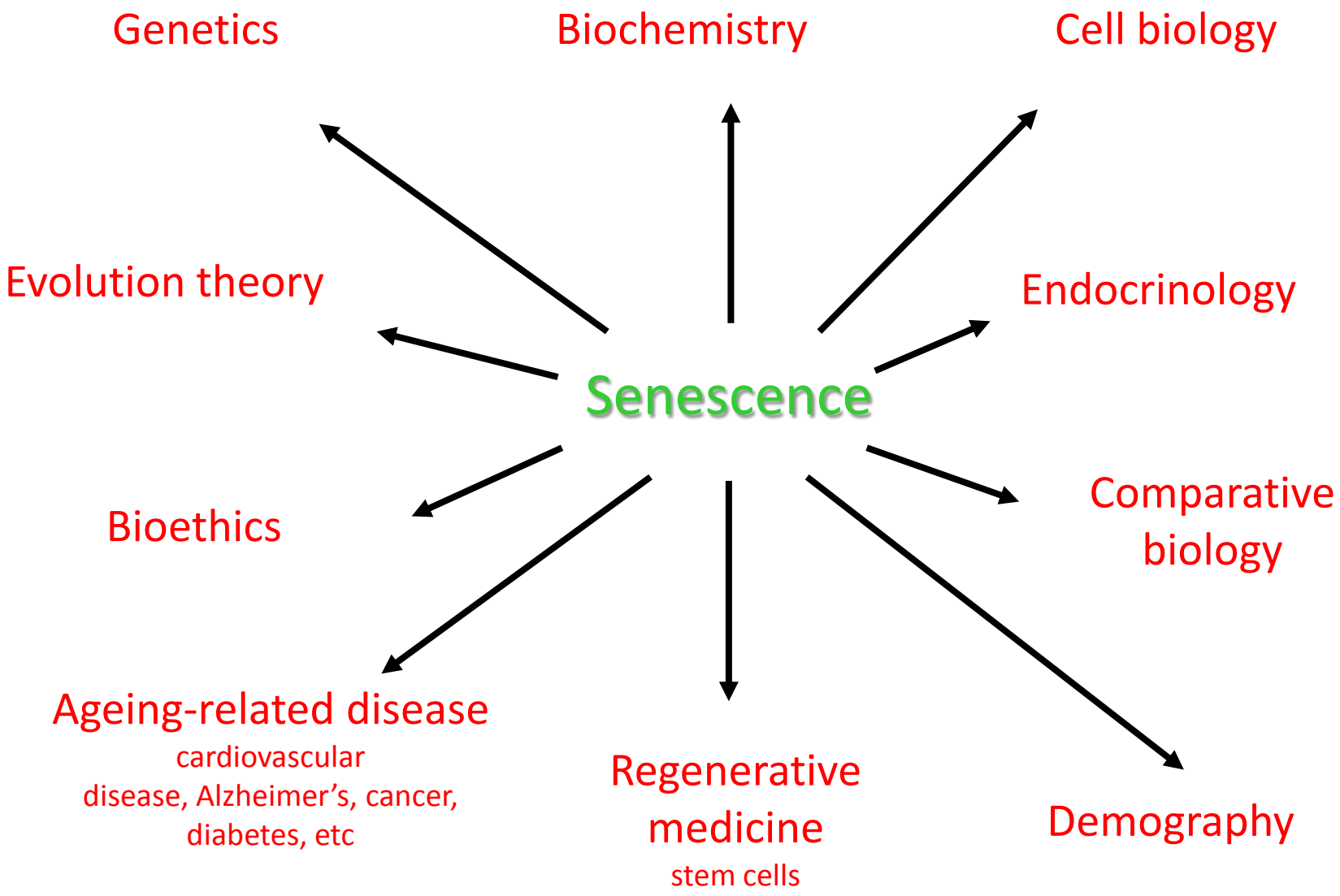
Late-onset complex diseases ⟷Senescence

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:

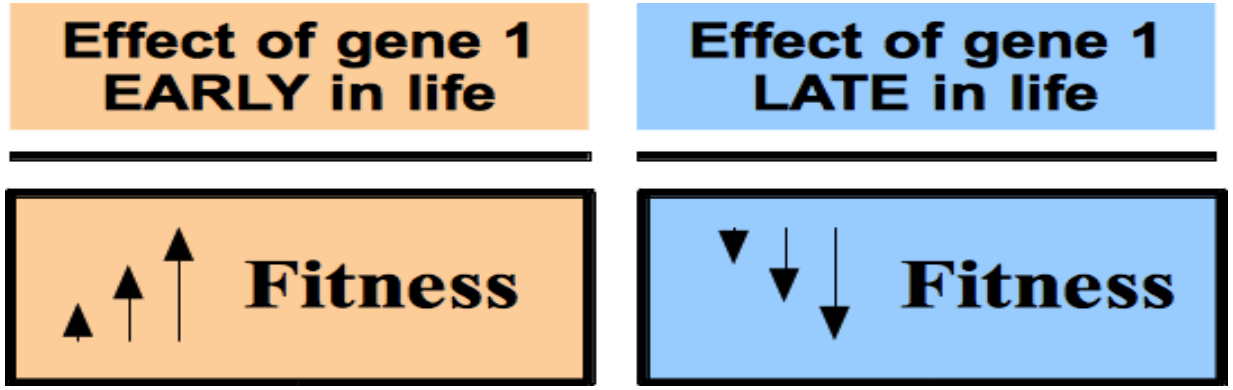
- How do we decay with age?
- *Why* are we senescent?

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…

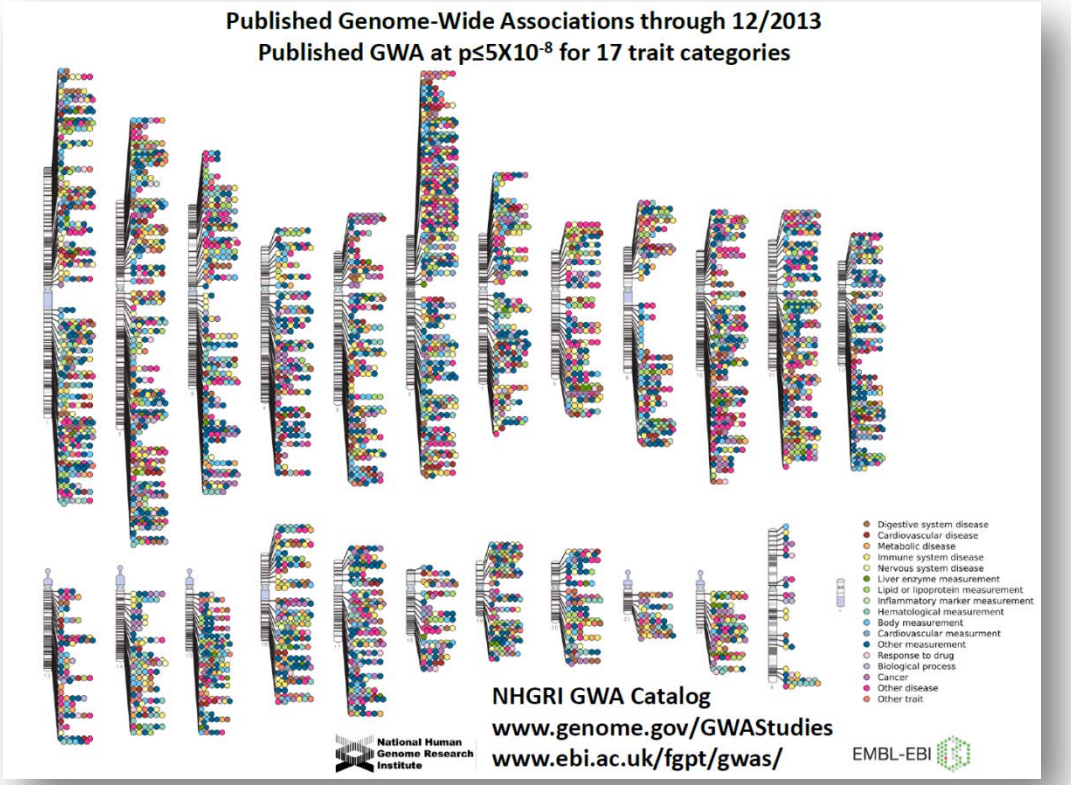
- Do we get sick when old because natural selection favoured health when young?

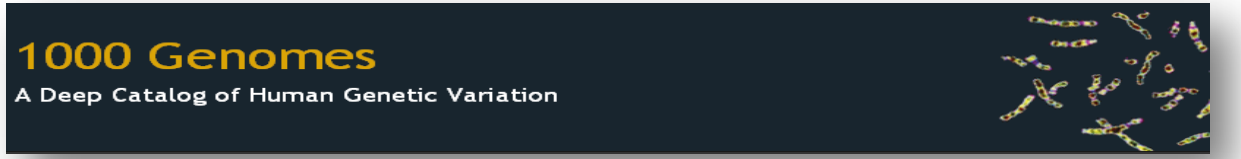
- That is, are there age-related antagonistic pleiotropies that link late onset disease and natural selection? And, if so:
    - *Up to which point?*
    - *Which diseases and genes are involved?*

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:

Early Onset (=< 25 years)	n = 43 diseases
Late Onset (> 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**

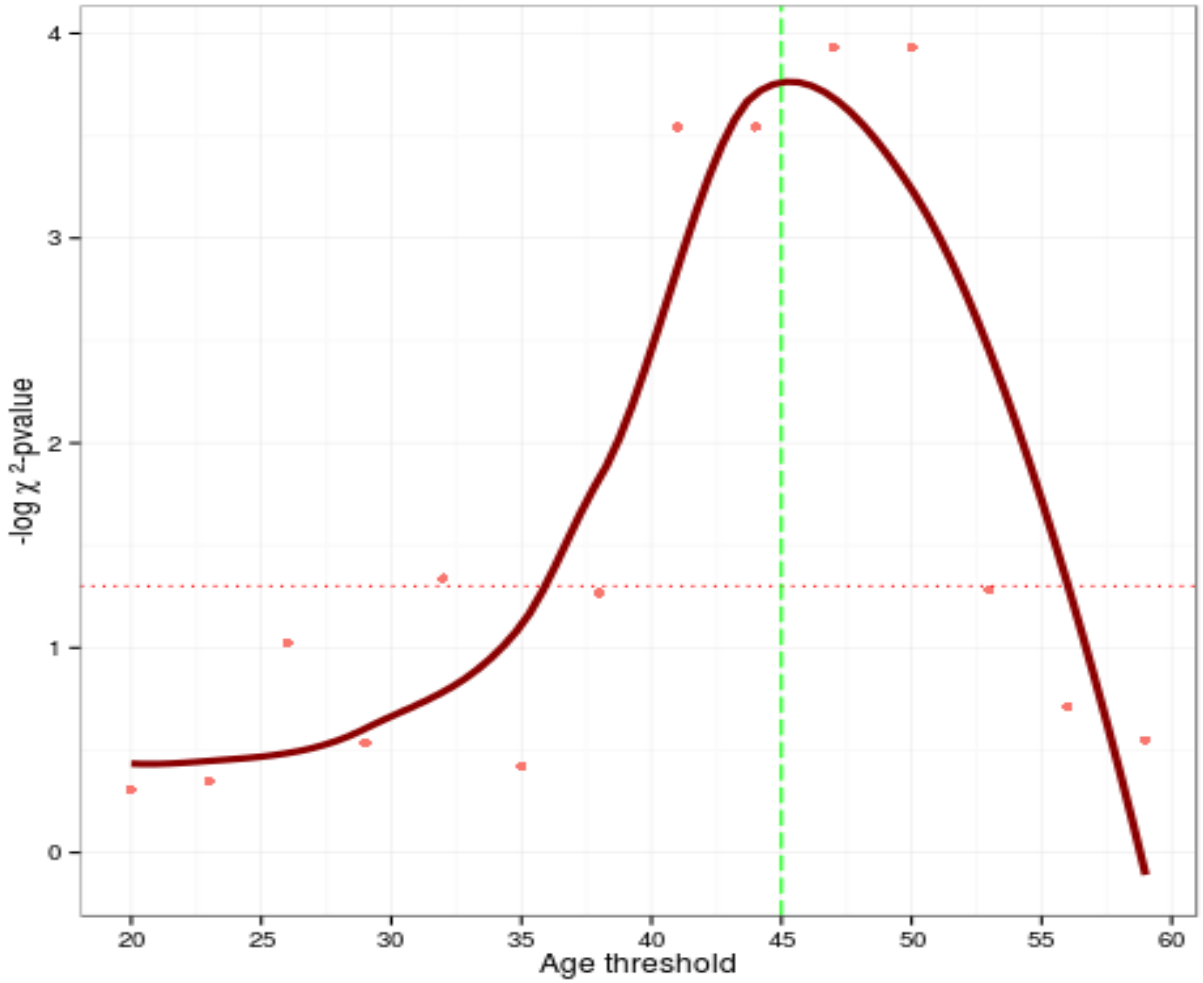
A clear excess of Early-Late Antagonistic pleiotropies

$R^2 \geq 0.8$ 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



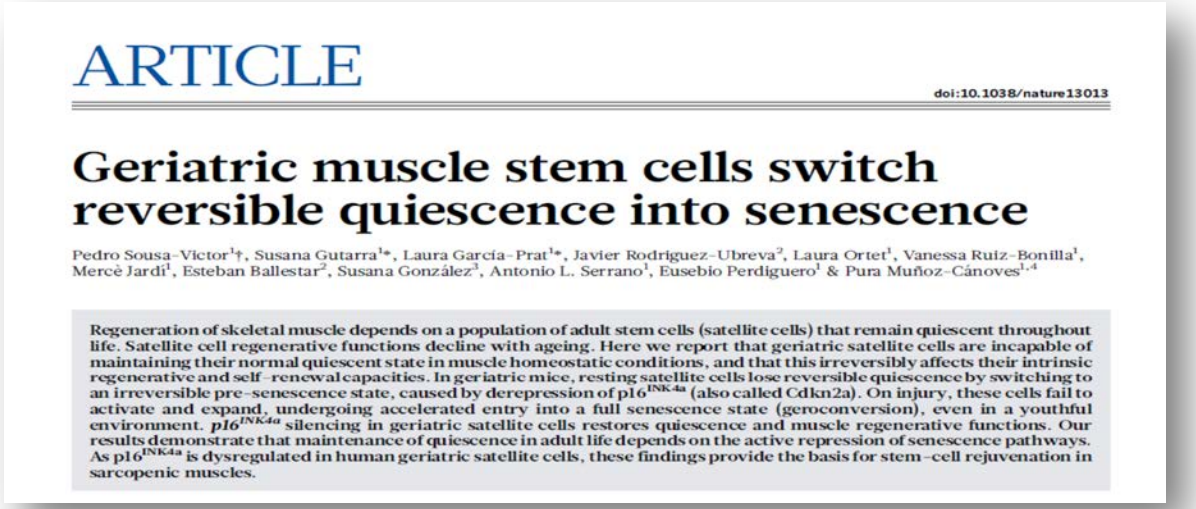
Distribution of X^2 pvalues depending on the age threshold

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

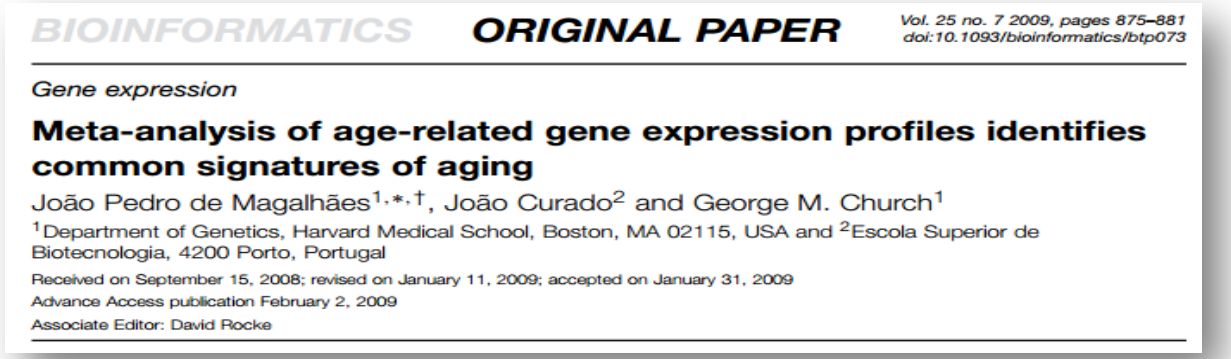
Geriatric muscle stem cells switch reversible quiescence into senescence

Pedro Sousa-Victor[1]†, Susana Gutarra[1]*, Laura García-Prat[1]*, Javier Rodriguez-Ubreva[2], Laura Ortet[1], Vanessa Ruiz-Bonilla[1], Mercè Jardí[1], Esteban Ballestar[2], Susana González[3], Antonio L. Serrano[1], Eusebio Perdiguero[1] & Pura Muñoz-Cánoves[1,4]

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 $p16^{INK4a}$ (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. *$p16^{INK4a}$* 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 $p16^{INK4a}$ 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 = $1.69 \cdot 10^{-14}$

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[1,*,†], João Curado[2] and George M. Church[1]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA and [2]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 = $1.9 \cdot 10^{-15}$

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. 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

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. 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

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!!!