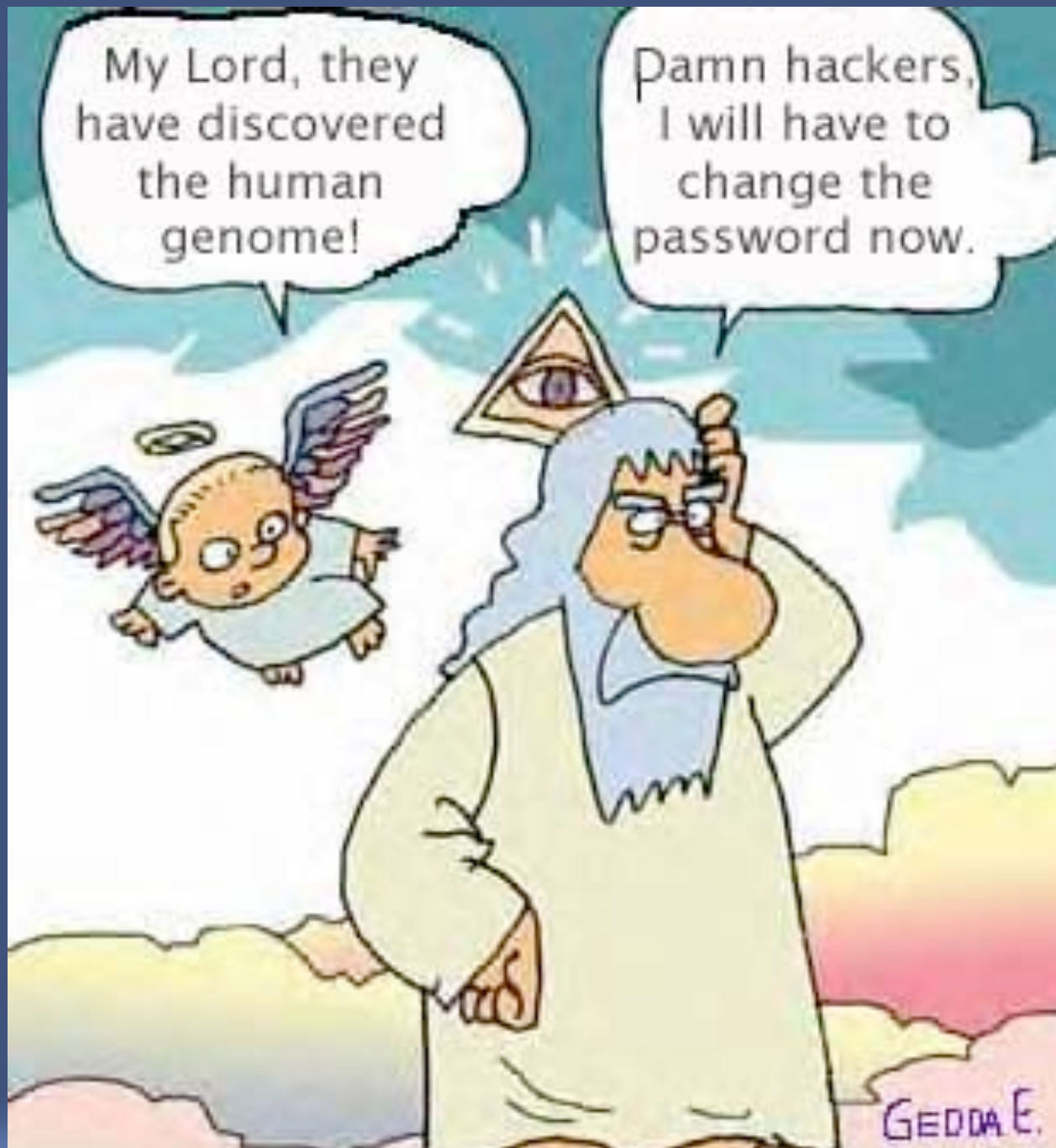


Biologisen tiedon saatavuus Postgenomiaikana

Leena Palotie

Helsingin yliopisto ja Kansanterveyslaitos
Broad Institute, MIT, Cambridge, USA





After Human Genome Project



Basic tools exist to characterize the biology behind human diseases

Completion of 468 Genome Projects

- 347 Bacterial
- 26 Archeal
- 95 Eucaryotic

Genome project(s) paved the way towards

- Immediate data release
- Open access to the data
- Data QC standards
- Rapid data integration

- Global collaborative effort

Biobanks

- Genome profiles
- Clinical data (health care information)
- Epidemiological data

- Life event data

European competitive niche in this era: societal, not technical

- Well characterized populations
- Reliable health care infrastructure
- High quality, equal education
- Top level expertise in genetics, epidemiology, clinical medicine, IT



Unique possibilities in health care-related
genome research

BIOBANKS

Genomic, transcriptomic, proteomic
metabolomic analyses

New
element

Epidemiological information

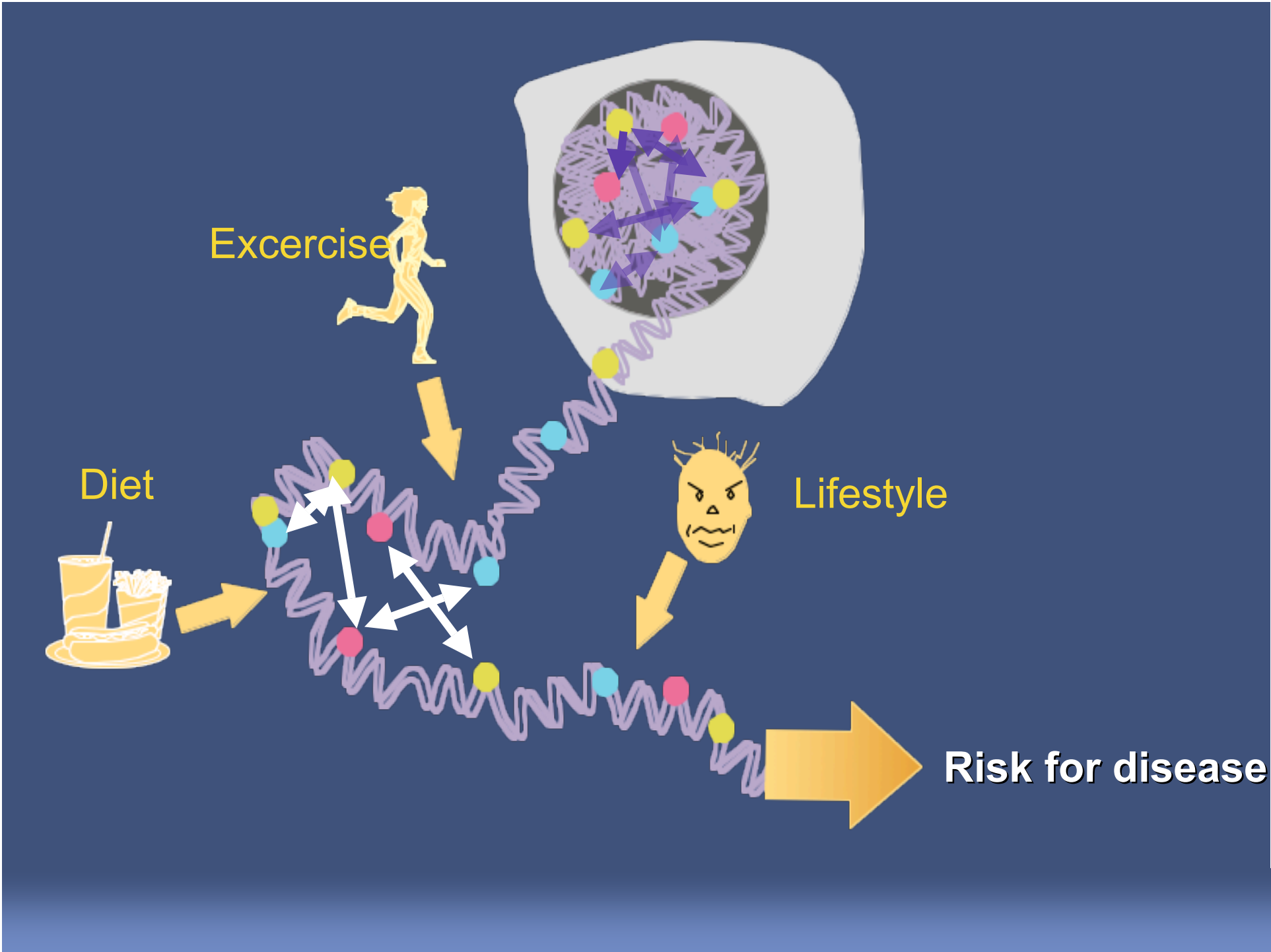
Free medication
registers

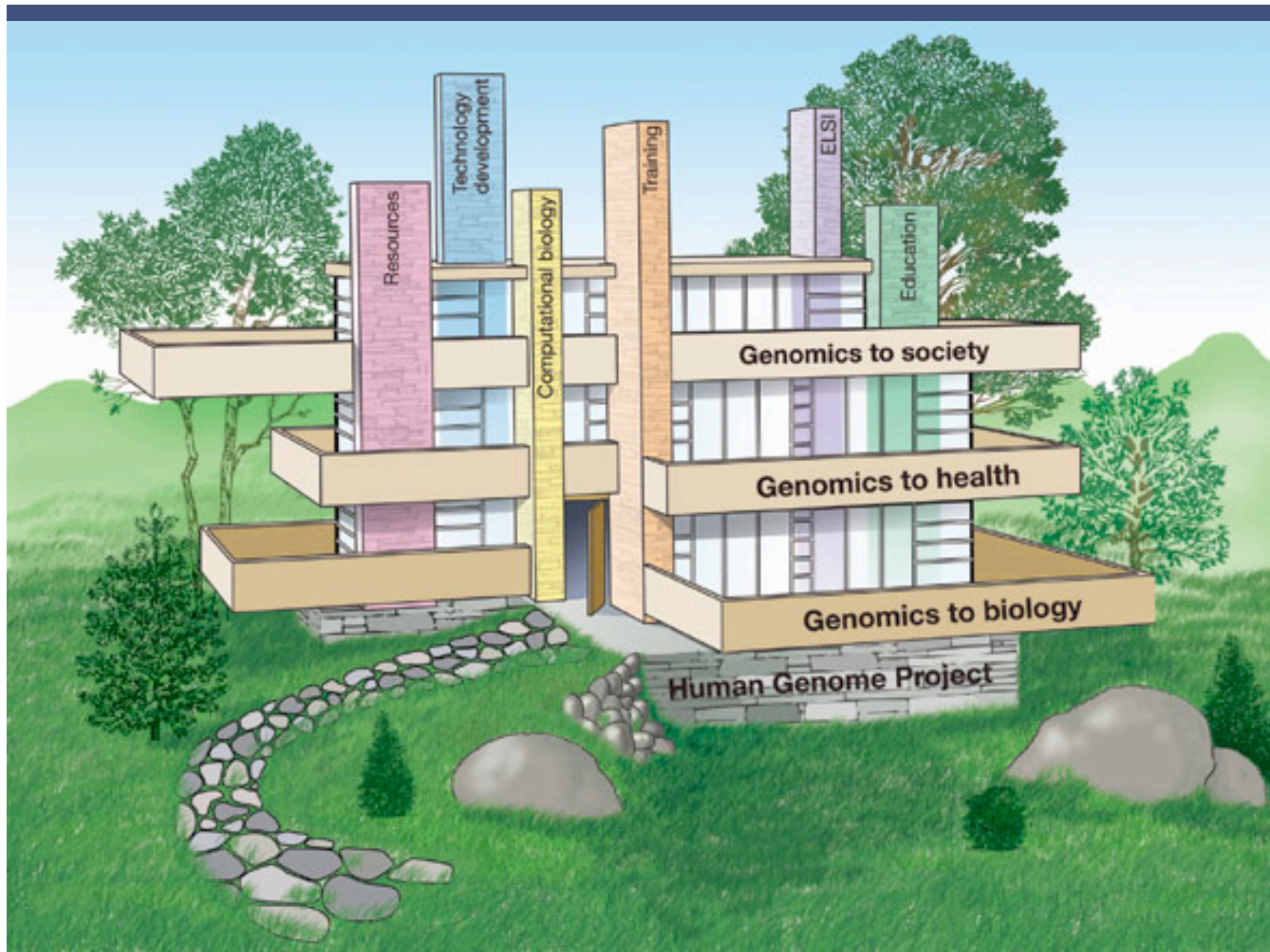
Health Care Records

Hospital discharge
registers

Old and new biobanks with varying study designs and IT infrastructures

- UK Biobank
- Estonian Genome project
- DeCode (Iceland)
- Swedish National Biobank Program (Lifegene)
- National birth cohorts (Norway, Finland..)
- "Trait focused" banks: EPIC, Carthagen, MORGAM, Kora, Finrisk...
- Genomeutwin cohorts
- NIH Prospective cohort based population study





Why massive data analyses are needed

- Effect of individual gene variant is small
- Multiple biological and life style variants contribute to the phenotype



Necessity of large cohorts

Genome era

- New type of research -> new challenges
- Use of public databases is critical
- Who is the inventor?
- Who owns the data?
- What data provides commercial value?
- Time to discovery?
- Data sharing across the centers

Critical bottlenecks:

- Relative lack of experts: We must build attractive centers to be able to recruit globally
- Lack of data harmonization: Demand for shared data and standards and facilitating databases
- Expensive technology: Platforms with high running costs and constant updating of large databases require intensified collaboration between academic institutes (and industry)
- The most expensive part, huge amount of clinical and life style data has been collected by the investments of our societies. It should be used!

Global effort is needed

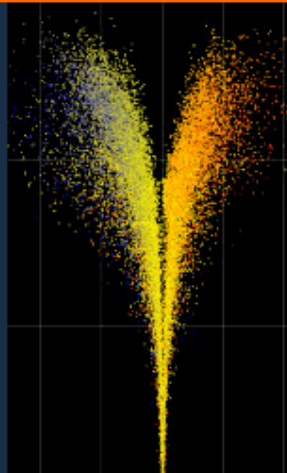
- In data harmonization: genetic profiles, phenotypes, epidemiological data
- In pooling resources and expertise in genotyping metabolomics/ proteomics and statistical analyses
- In data delivery to scientific community and public



- Public Population Projects in Genomics (www.p3g.org)
- ESFRI efforts (www.biobanks.eu)



THE FACILITY



Website launched >

Biobanks website has been opened 01.10.2007



PARTNERS

Partners supportive of a European Biobank Research Infrastructure:

- GenomEUtwin (www.genomeutwin.org)
- Finnish biobank (www.nationalbiobanks.fi)
- Swedish biobank (<http://www.biobanks.se/>)
- German biobank, KORA (http://epi.gsf.de/kora-gen/index_e.html)
- British biobank (<http://www.alspac.bris.ac.uk/welcome/index.shtml>)
- Estonian biobank: (www.geenivaramu.ee)
- UK biobank (www.ukbiobank.ac.uk)
- A Nordic Center of Excellence linking family-based collections from Nordic countries (www.ncoedg.org)
- Generation Scotland (www.generationscotland.org)
- HUNT, a Norwegian biobank targeted to cardiovascular health (www.hunt.ntnu.no)
- EPIC, (<http://www.iarc.fr/epic/>) European prospective study of

Genomics & Genetic Epidemiology

Research in
small groups

Biobanks



Research in large consortia
with significant infrastructure

Genomics

"Small research"



"Big research"

Genomics

"Small research"



"Big research"

RO1 type funding

Large investments
compare: physics

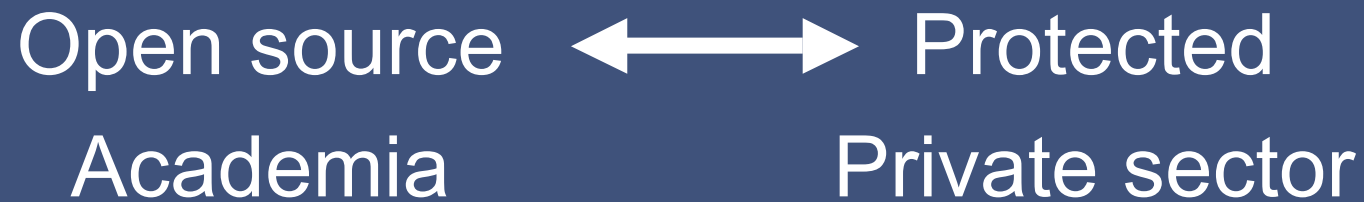
“New ideas need new structures”

Lee Roy Hood

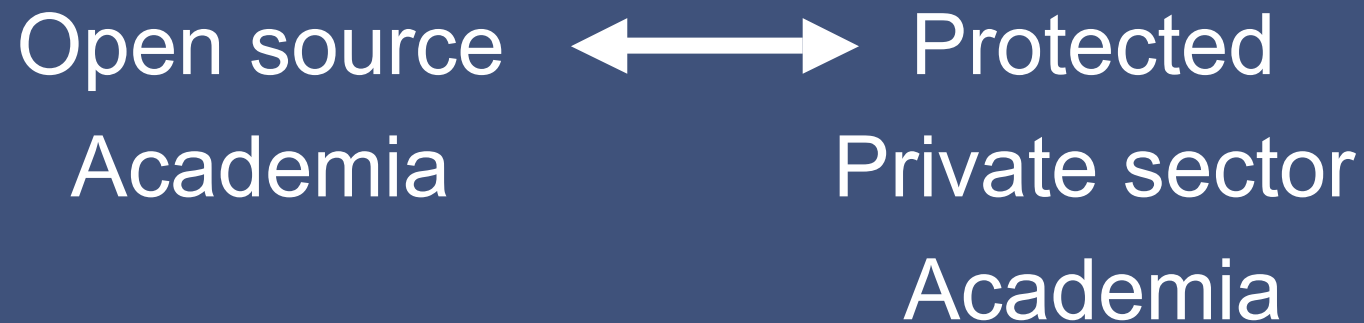
Data sharing in Biosciences

Open source ↔ Protected

Data sharing in Biosciences



Data sharing in Biosciences



Data sharing in Genomics

Open source



Protected

- SNP consortium
- Genome project
 - Hapmap data
- Whole genome association studies

- Private genome project
 - Whole genome association studies produced in companies
- Epidemiological data
 - Clinical data

“If you like the field to move forward,
place the data on the web”

Eric Lander

Data sharing requirements

- Major funding agencies are producing very strict requirements for data release
 - NIH
 - Wellcome Trust

Data integration of large population genetic projects

