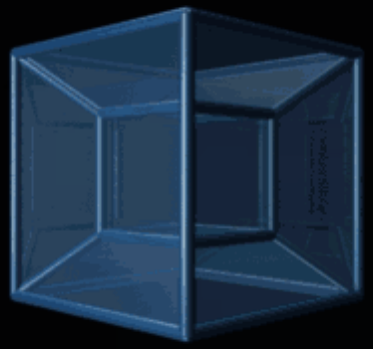
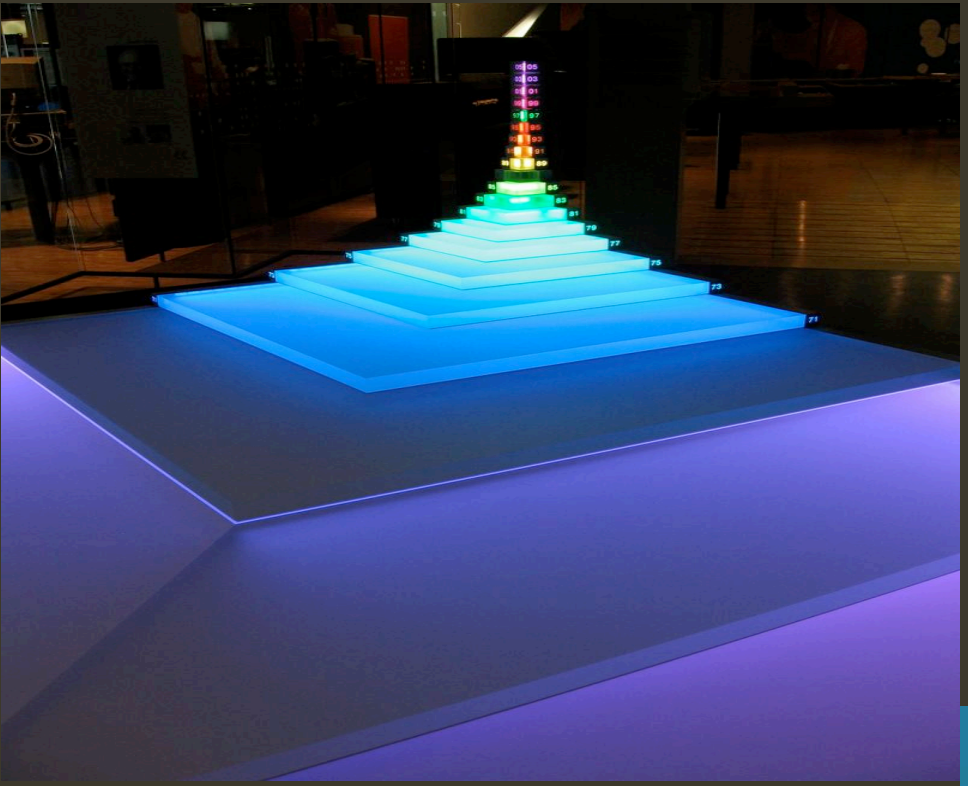
Big Data in Health

Prof. Dr. Bart De Moor

bart.demoor@kuleuven.be

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes

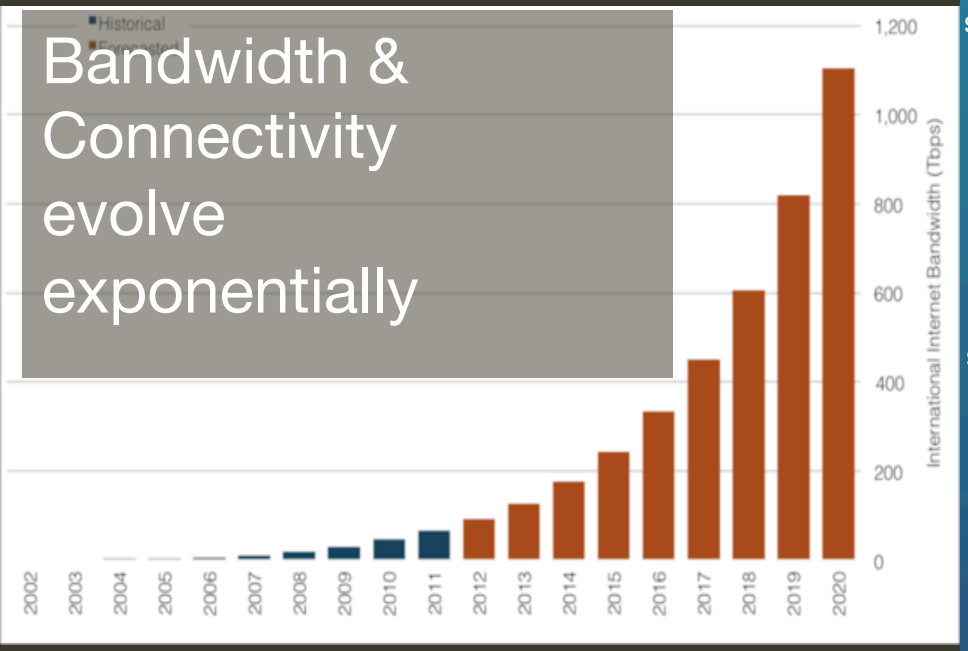


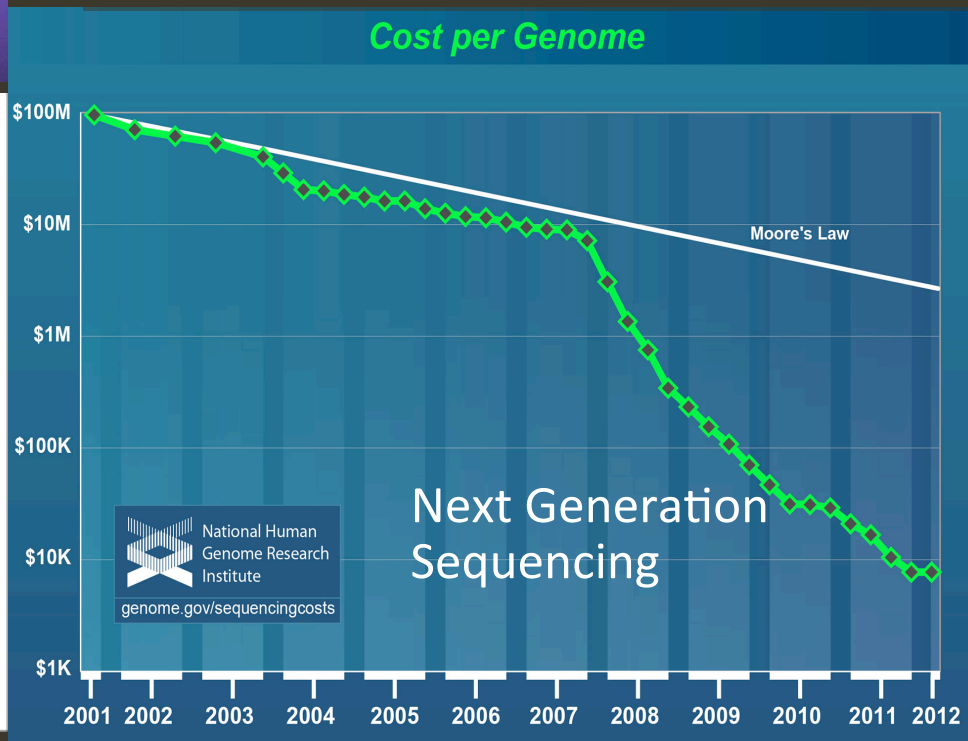


Moore's law:
computing power
doubles
every 18 months

Carlson's law:
complexity/cost
evolves
exponentially

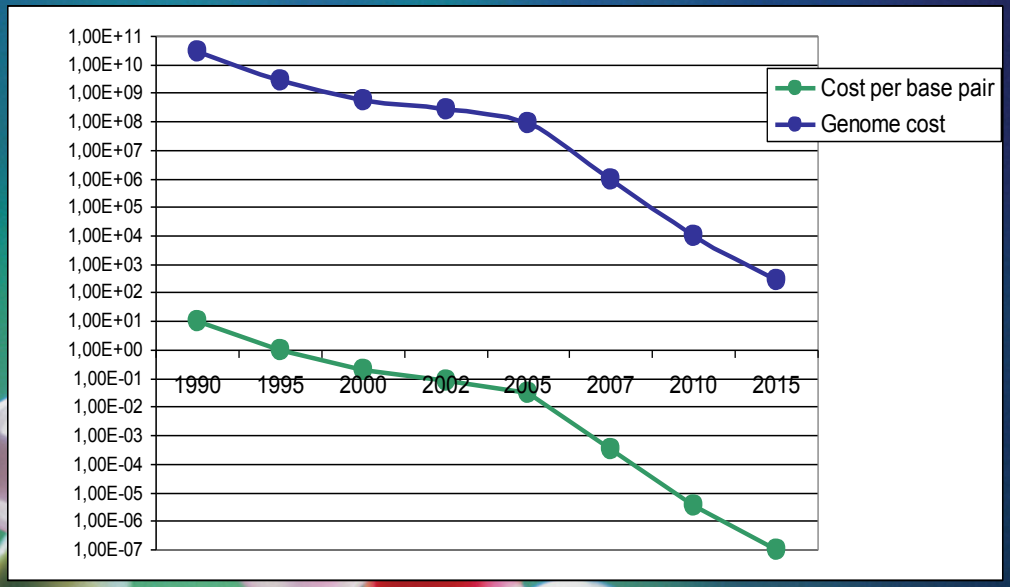
Bandwidth &
Connectivity
evolve
exponentially





Genome data

- Human genome project
  - Initial draft: June 2000
  - Final draft: April 2003
  - 13 year project
  - $300 million value with 2002 technology
- Personal genome
  - June 1, 2007
  - Genome of James Watson, co-discoverer of DNA double helix, is sequenced
    - $1.000.000
    - Two months
- €1000-genome
  - Expected 2012-2020



YearΩ	Cost per base pair	Genome cost
1990	10	3E+10
1995	1	3.000.000.000
2000	0.2	600.000.000
2002	0.09	270.000.000
2005	0.03	90.000.000
2007	0.000333333	1.000.000
2010	3.33333E-06	10000
2015	0.0000001	300

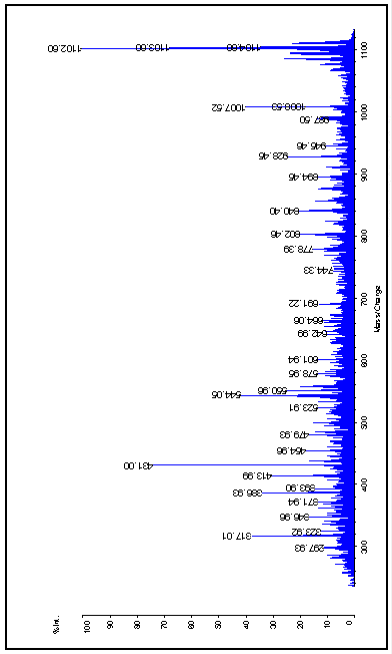
Data tsunami

Computer Tomography

GS-FLX Roche
Applied Science 454

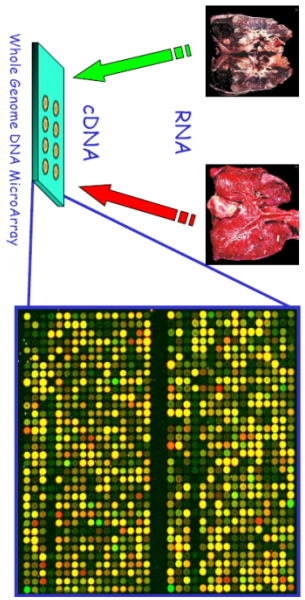
Sequencers

```
ACACATTAAATCTTATATGC
TAAAACTAGGTCTCGTTTTA
GGGATGTTTATAACCATCTT
TGAGATTATTGATGCATGGT
TATTGGTTAGAAAAAATATA
CGCTTGTTTTTCTTTCCTAG
GTTGATTGACTCATACATGT
GTTTCATTGAGGAAGGAAC
TTAACAAAACTGCACTTTTT
TCAACGTCACAGCTACTTTA
AAAGTGATCAAAGTATATCA
AGAAAGCTTAATATAAAGAC
ATTTGTTTCAAGGTTTCGTA
AGTGCACAATATCAAGAAG
ACAAAAATGACTAATTTTGT
TTTCAGGAAGCATATATATT
ACACGAACACAAATCTATTT
TTGTAATCAACACCGACCAT
GGTTCGATTACACACATTAA
ATCTTATATGCTAAAACTAG
GTCTCGTTTTAGGGATGTTT
ATAACCATCTTTGAGATTAT
TGATGCATGGTTATTGGTTA
GAAAAAATATACGCTTGTTT
TTCTTTCCTAGGTTGATTGA
```

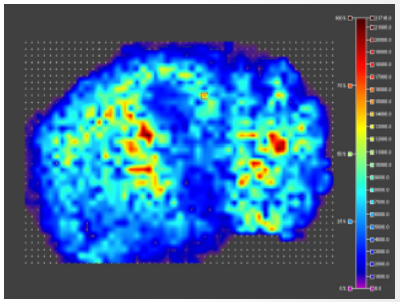


Magnetic resonance

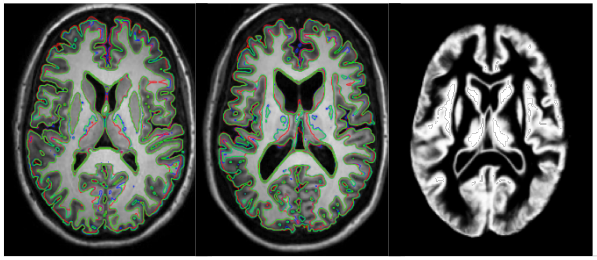
maf VIB
MicroArray Facility



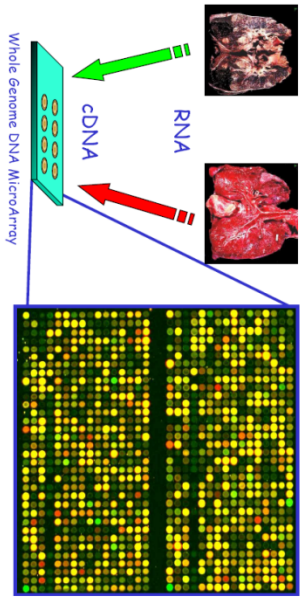
Microarrays
(DNA chips)

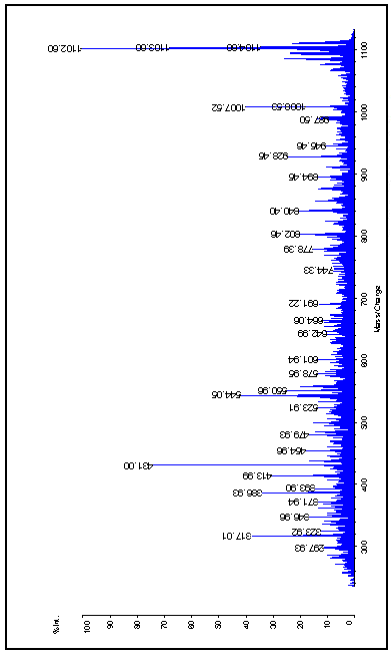


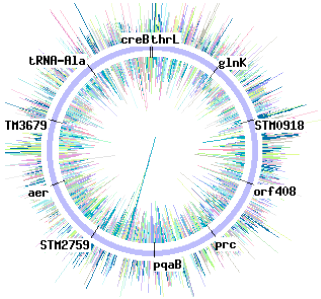
Mass spectrometry

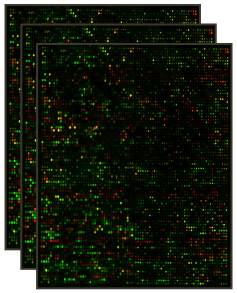


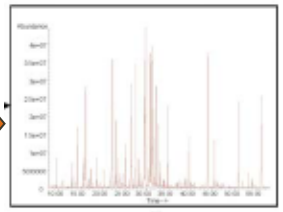
ACACATTAAATCTTATATGC
TAAAACTAGGTCTCGTTTTA
GGGATGTTTATAACCATCTT
TGAGATTATTGATGCATGGT
TATTGGTTAGAAAAAATATA
CGCTTGTTTTTCTTTCCTAG
GTTGATTGACTCATACATGT
GTTTCATTGAGGAAGGAAC
TTAACAAAACTGCACTTTTT
TCAACGTCACAGCTACTTTA
AAAGTGATCAAAGTATATCA
AGAAAGCTTAATATAAAGAC
ATTTGTTTCAAGGTTTCGTA
AGTGCACAATATCAAGAAG
ACAAAAATGACTAATTTTGT
TTTCAGGAAGCATATATATT
ACACGAACACAAATCTATTT
TTGTAATCAACACCGACCAT
GGTTCGATTACACACATTAA
ATCTTATATGCTAAAACTAG
GTCTCGTTTTAGGGATGTTT
ATAACCATCTTTGAGATTAT
TGATGCATGGTTATTGGTTA
GAAAAAATATACGCTTGTTT
TTCTTTCCTAGGTTGATTGA

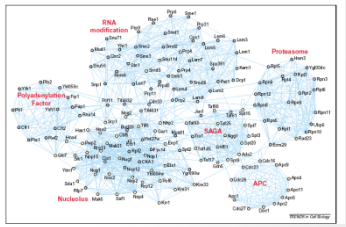










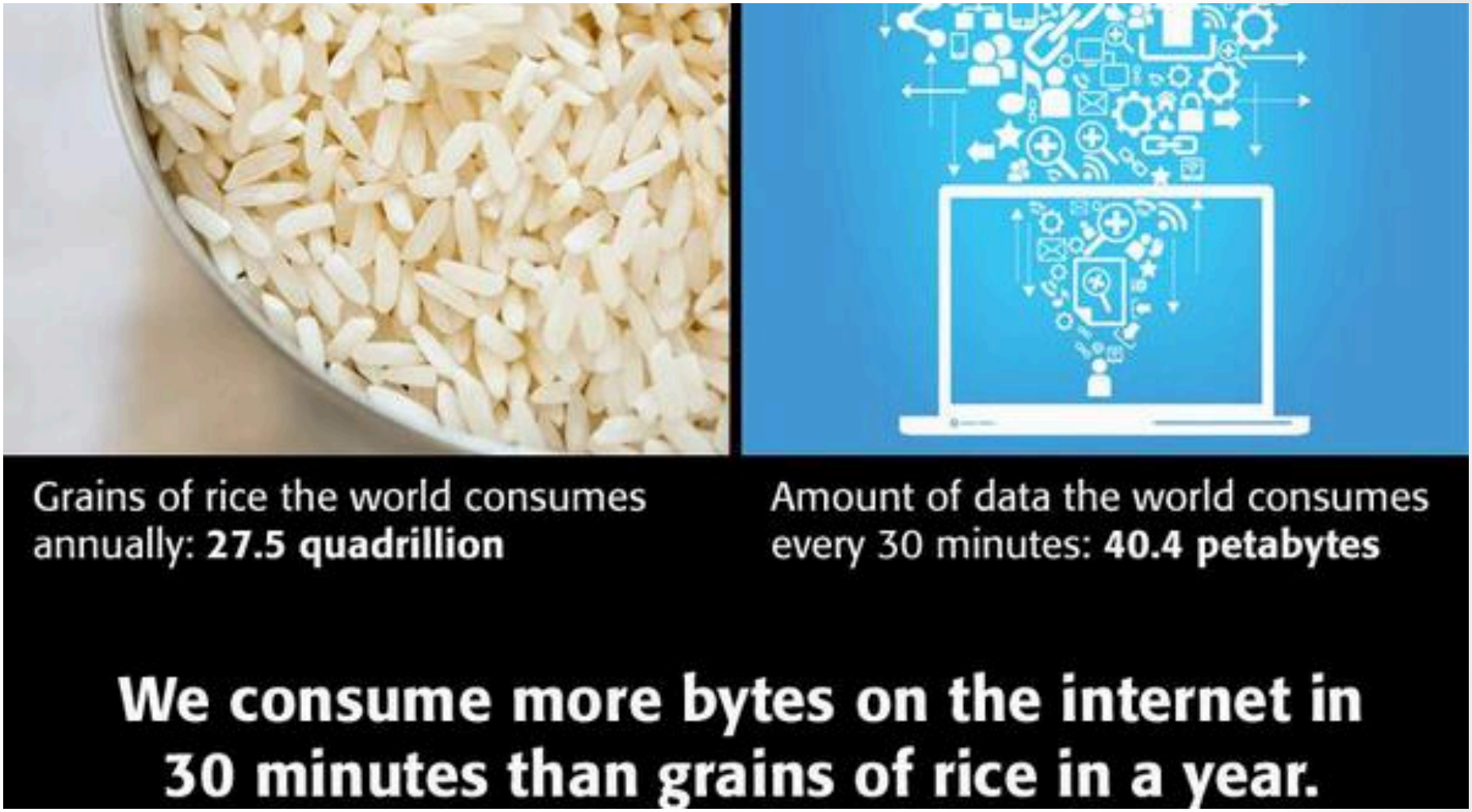


genome **transcriptome** **proteome** **metabolome** **interactome**

GS-FLX Roche
Applied Science 454

maf VIB
MicroArray Facility

Prometa



1 million = 1 000 000
1 billion = 1 000 000 000
1 trillion = 1 000 000 000 000
1 quadrillion =
1 000 000 000 000 000

1 kB = 1 000
1 MB = 1 000 000
1 GB = 1 000 000 000
1 TB = 1 000 000 000 000
1 PB = 1 000 000 000 000 000

1 TB
= large university library
= 212 DVD discs
= 1430 CDs
= 3 year music CD quality

Tsunami of medical data resulting from technological progress

1 kB = 1000
1 MB = 1 000 000
1 GB = 1 000 000 000
1 TB = 1 000 000 000 000
1 PB = 1 000 000 000 000 000

sequencing all newborns by 2020 (125k births / year)

125 PetaByte / year

index of 20 million Biomedical PubMed records

23 GigaByte

raw NGS data of 1 full genome

1 TeraByte

Genomics core HiSeq 2000 full speed exome sequencing

1 TeraByte / week

1 small animal image

1 GigaByte

1 slice mouse brain MSI at 10 µm resolution

81 GigaByte

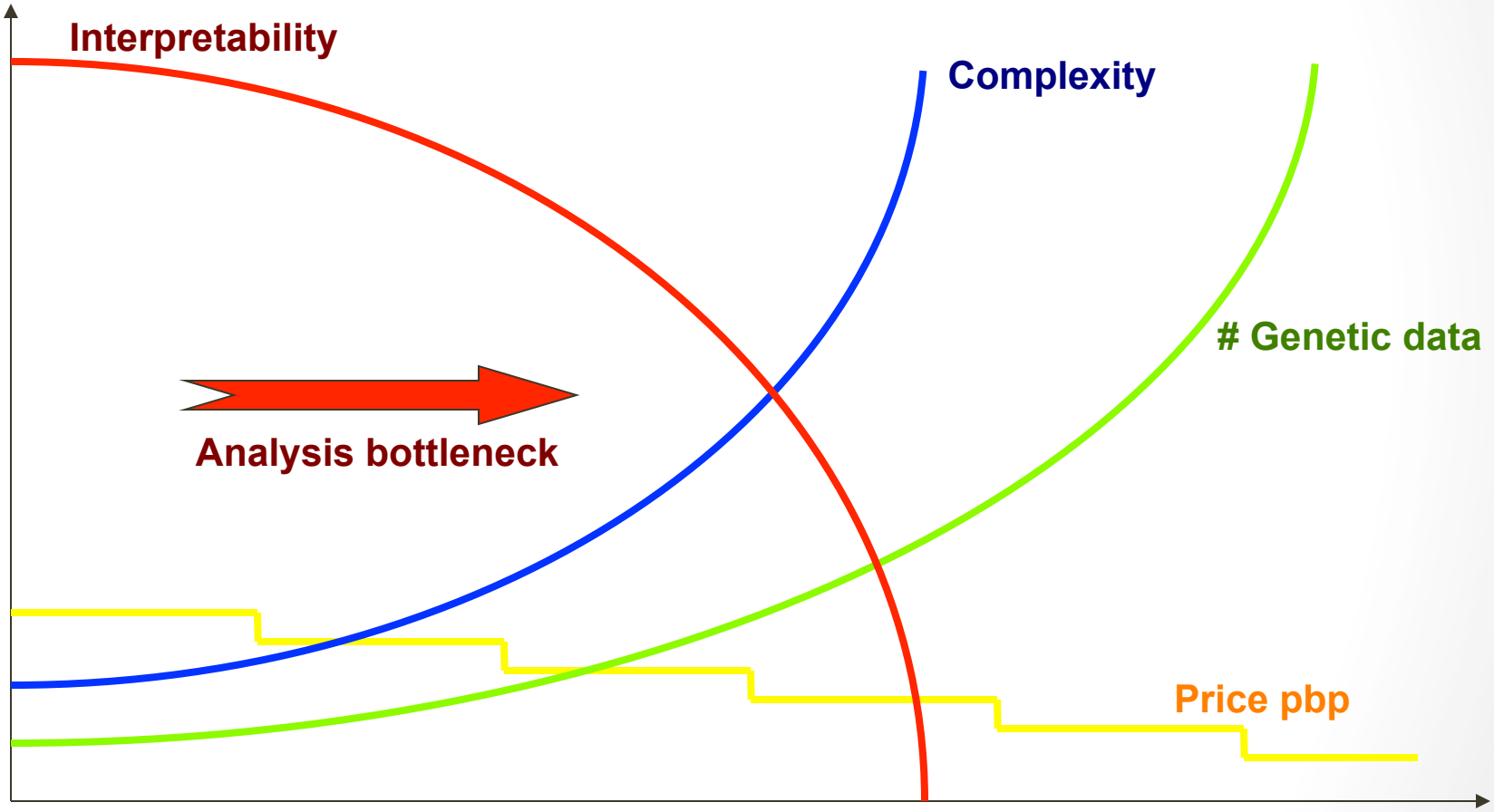
PACS UZ Leuven

1,6 PetaByte

1 CD-ROM

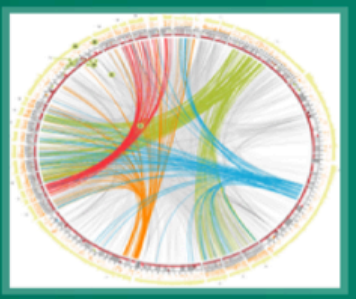
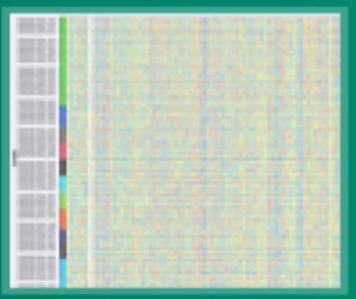
750 MegaByte

Interpretability of Genomic data



If we care about the future of care



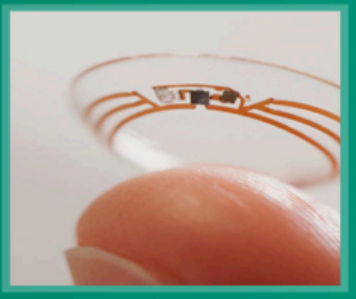







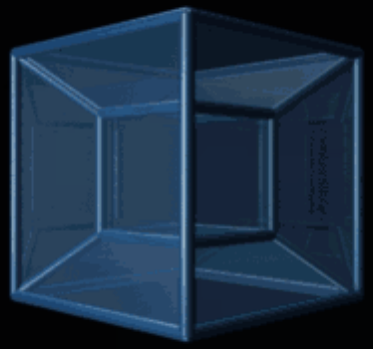






… technology will be key.

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes

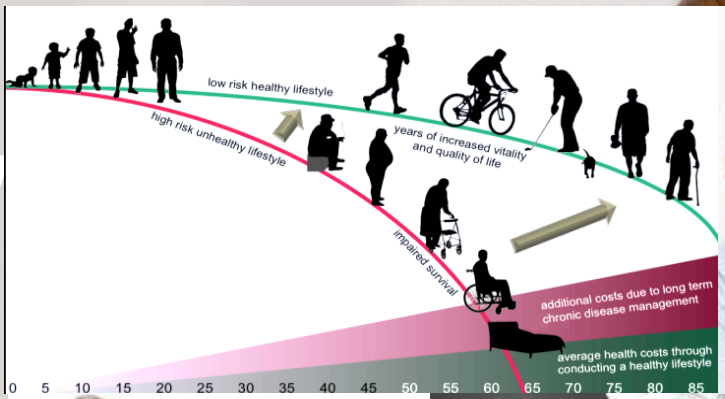


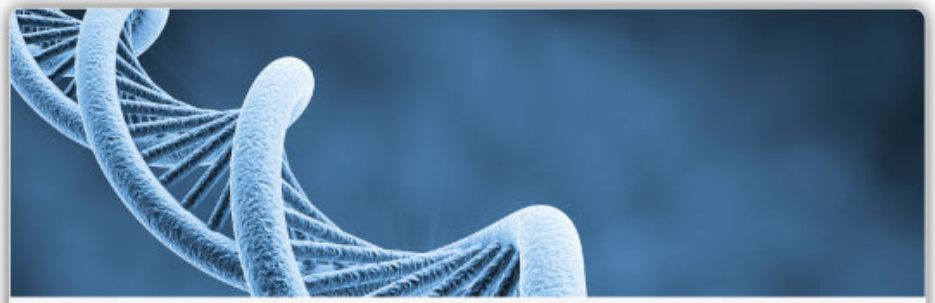
Demography

and evolving quality of life standards

generate

growing health care demands





Personalized

Customized diagnosis and treatment

Preventive

Better than curation

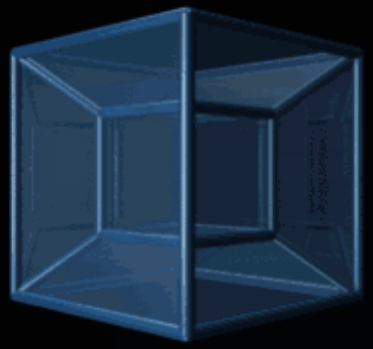
Predictive

Determine risk profiles & predict outcome

Participative

Involve the patient

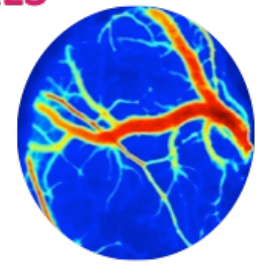
1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



Who is in demand ? The 3Ps !

PATIENTS

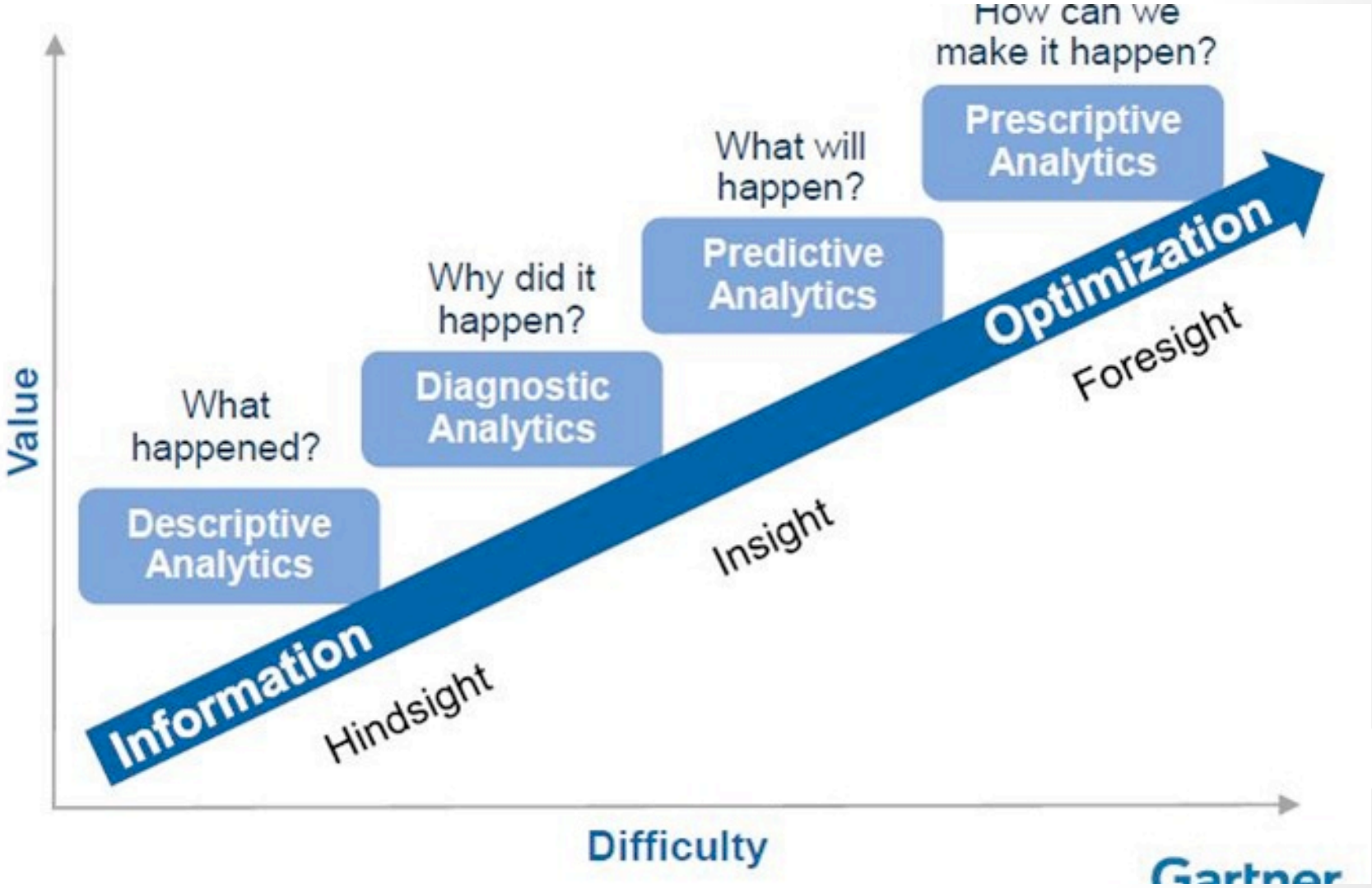
PROFESSIONALS



POLICY MAKERS

RIZIV

Analytics are mandatory



What is a CDSS?

- based on evidence-based medicine (EBM) to relay the best up to date information available to the physician or doctor;
- Result of top-notch medical research, biomedical know-how welded/ blended into information technology and computer science;
- Inference based machine learning algorithms
- In (increasing number of) environments where accurate interpretation of (lots of) data is mandatory.

Dr. Algorithm is coming

"In the next 10 years, data science and software will do more for medicine than all the biological sciences together."

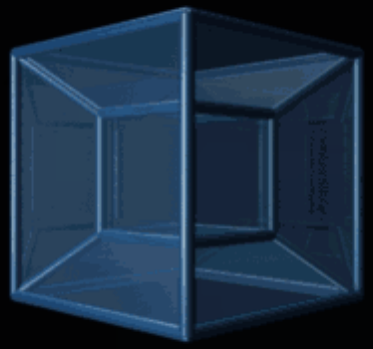
– Vinod Khosla, Khosla Ventures

http://techcrunch.com/2013/09/11/vinod-khosla-in-the-next-10-years-data-science-will-do-more-for-medicine-than-all-biological-sciences-combined/

Generic data processing tasks

- Data preprocessing, denoising, normalization
- Clustering and classification; feature detection; profiling;
- Relevance detection, ranking
- Dynamic modelling, time series, longitudinal modelling
- Decorrelation, modelling, (Kalman) filtering
- Predictive analytics
- Vizualisation
- Heterogeneous data fusion
- Prediction, processing and monitoring

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



Demand driven projects: Solid tradition of working with medical doctors

Intensive Care Unit

Radiology

Urology

Cardiology

Pneumology

Orthopaedics

Neonatology

Dentistry

Forensics

Rehabilitation

Gynaecology

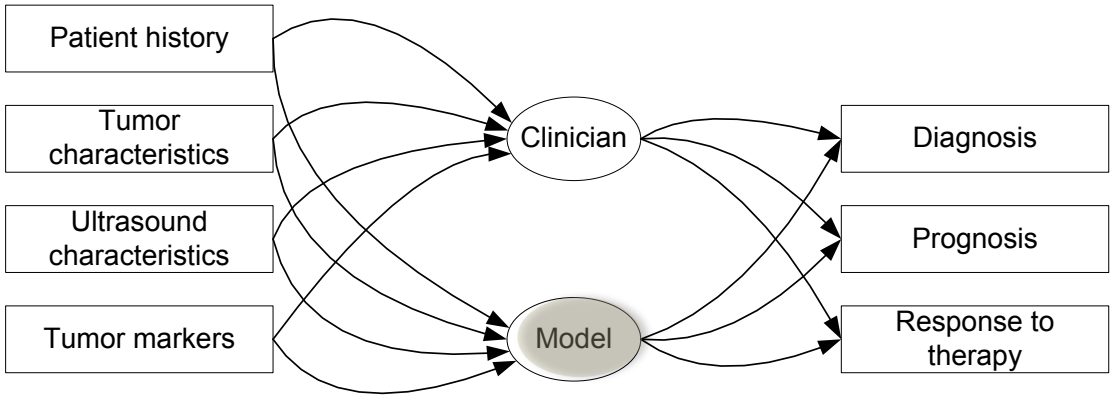
Radiotherapy

Neurology

Human Genetics

Oncology

CDSS: an example



standardize ultrasonographic ovarian tumor analysis → models giving an indication of the probability of malignancy of an ovarian tumour based on 6 to 12 observed parameters

IOTA app to assess ovarian tumour malignancy: population based / standardized

IOTA 2013 International Ovarian Tumour Analysis

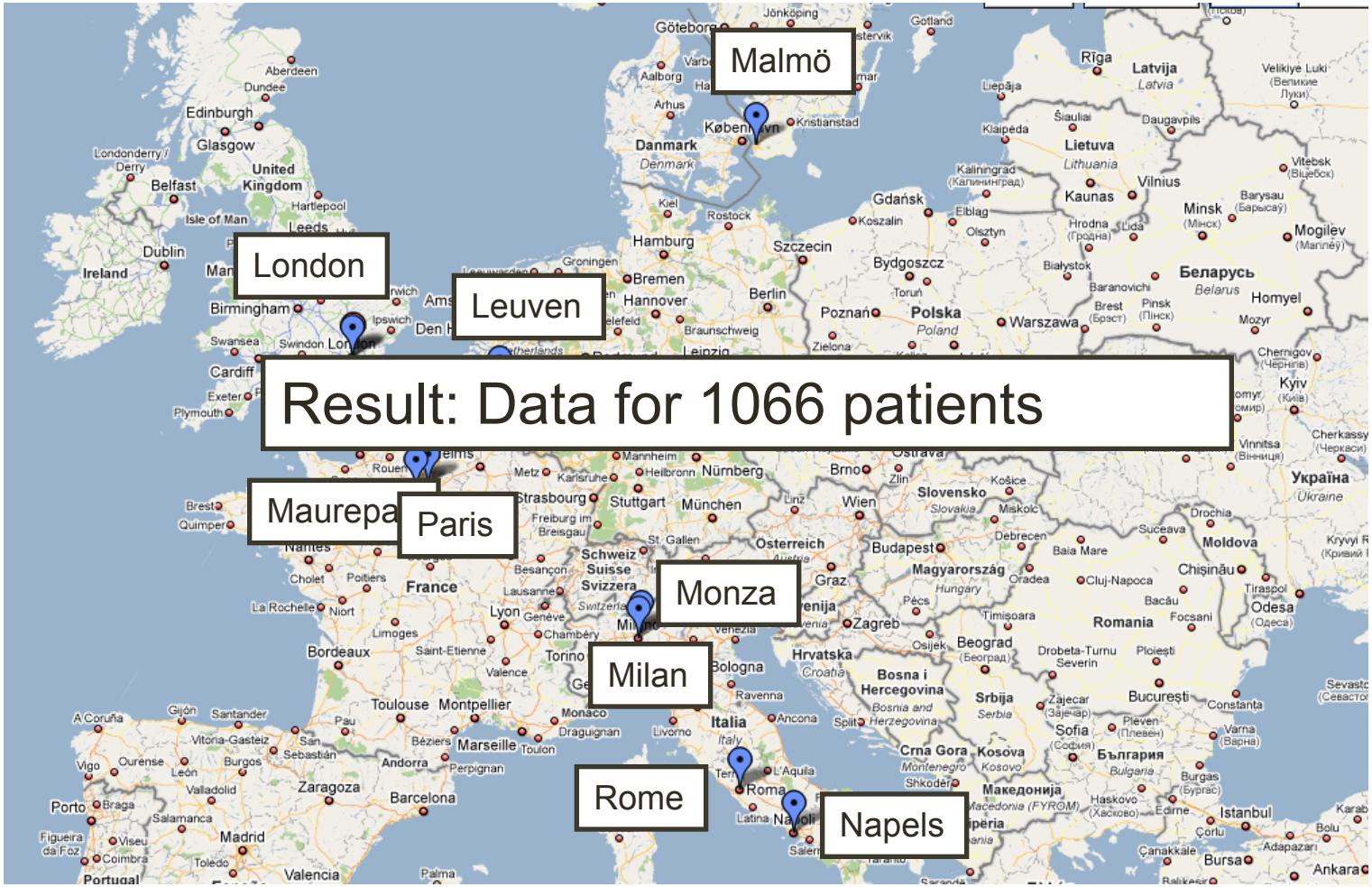
General challenges & opportunities:

- Integration of various heterogeneous data sources
- Connect with Electronic Medical Records
- Need for population data

IOTA app available in iTunes app store and on http://homes.esat.kuleuven.be/~sistawww/biomed/iota/

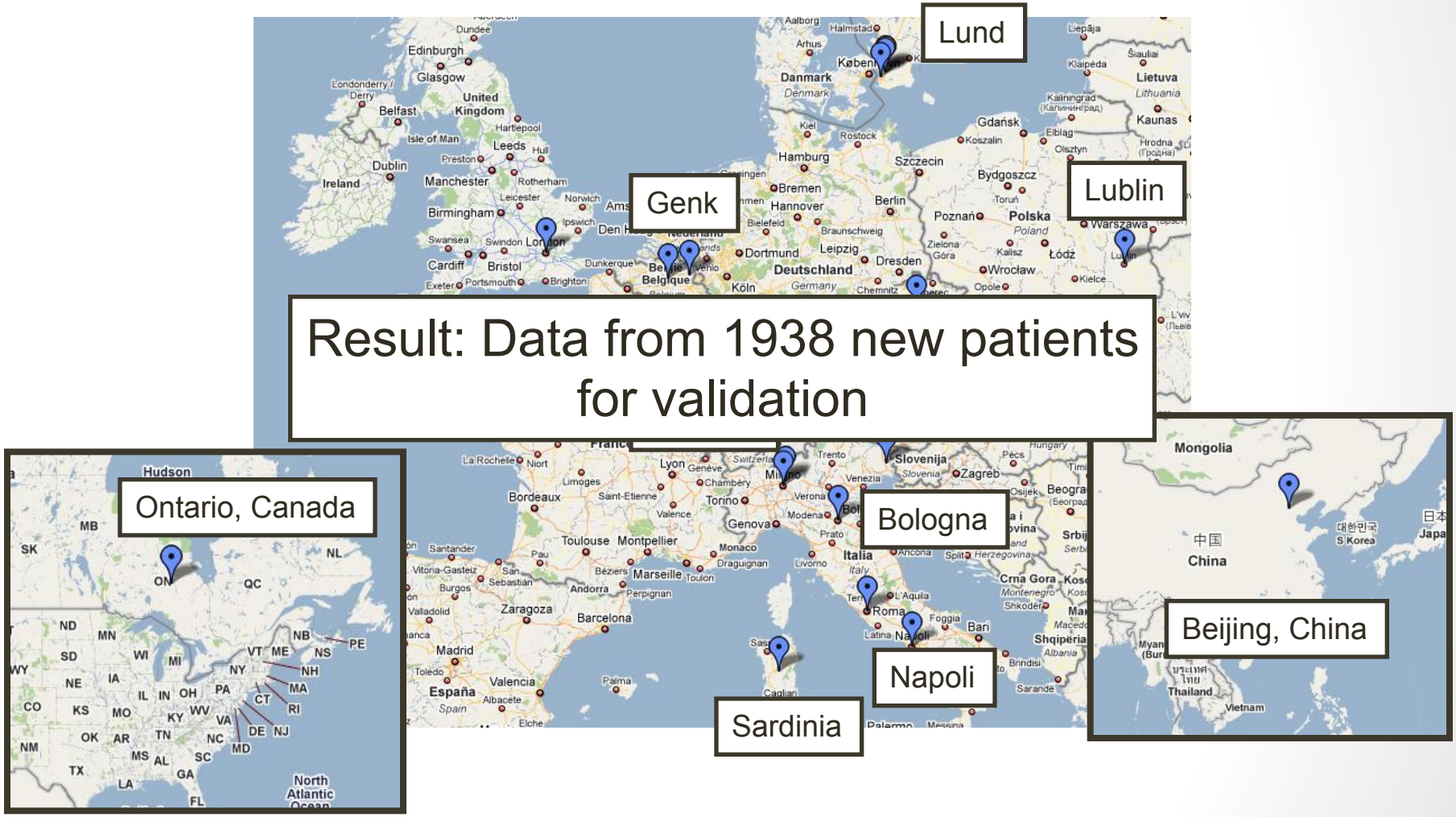
IOTA phase 1 centers

9 centers, 60 variabelen/patient

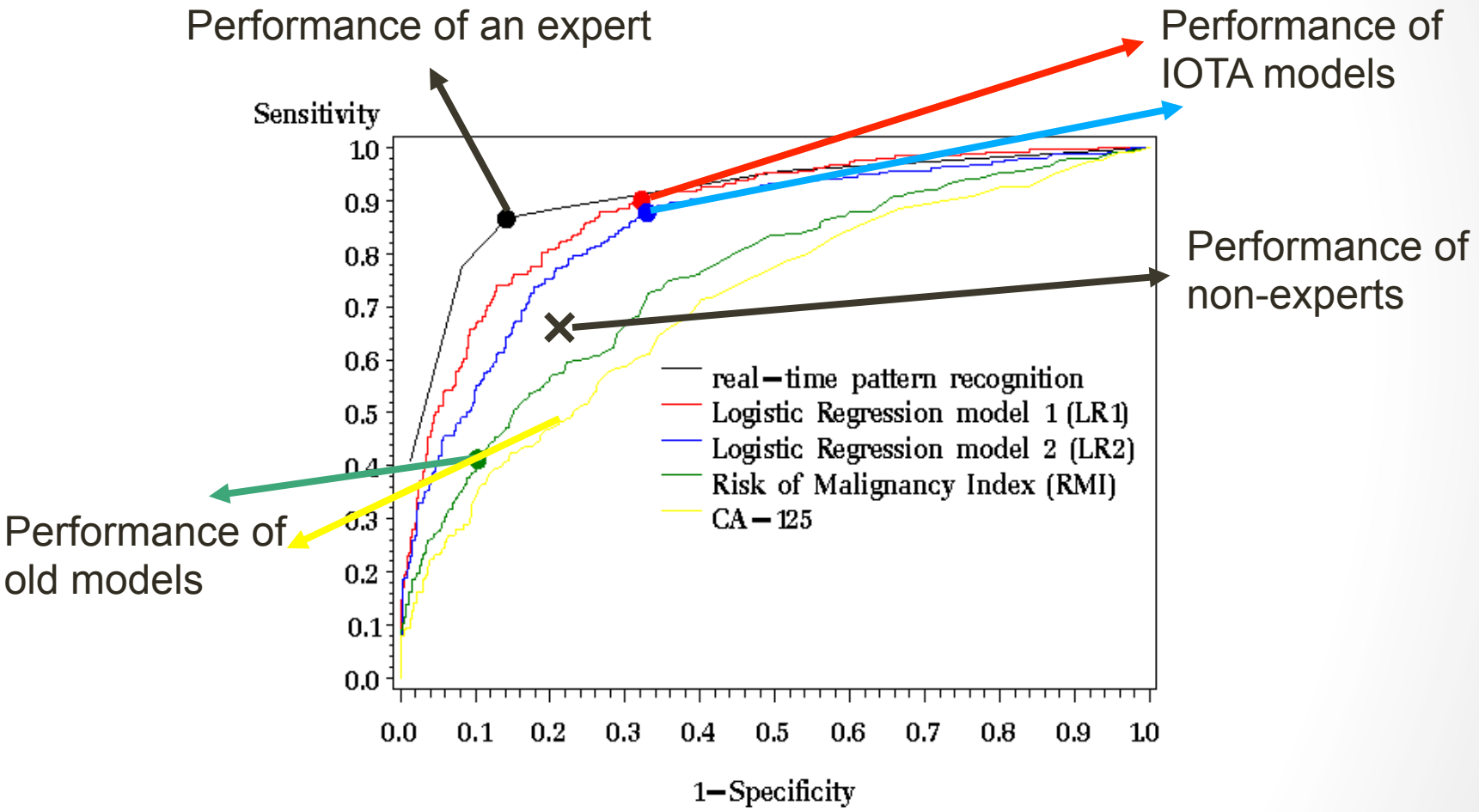


IOTA phase 2 centers

12 new centers

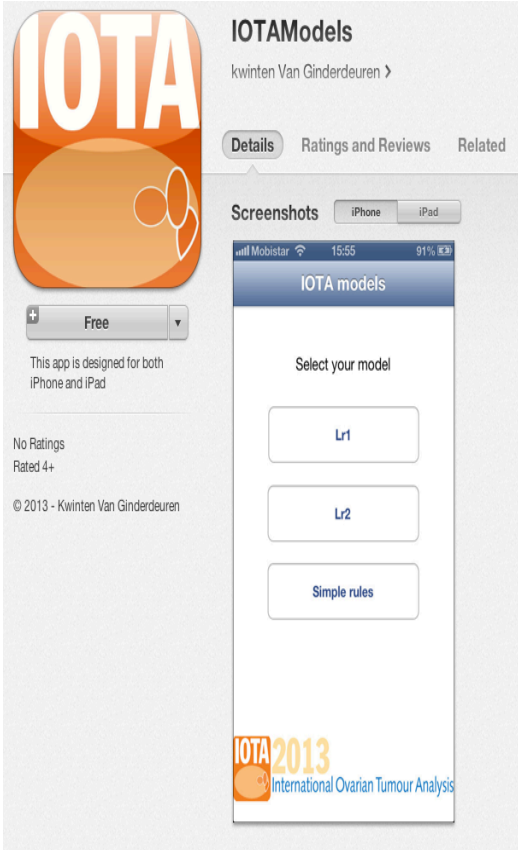


Performance comparison

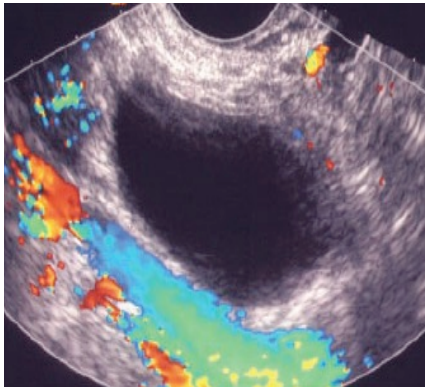


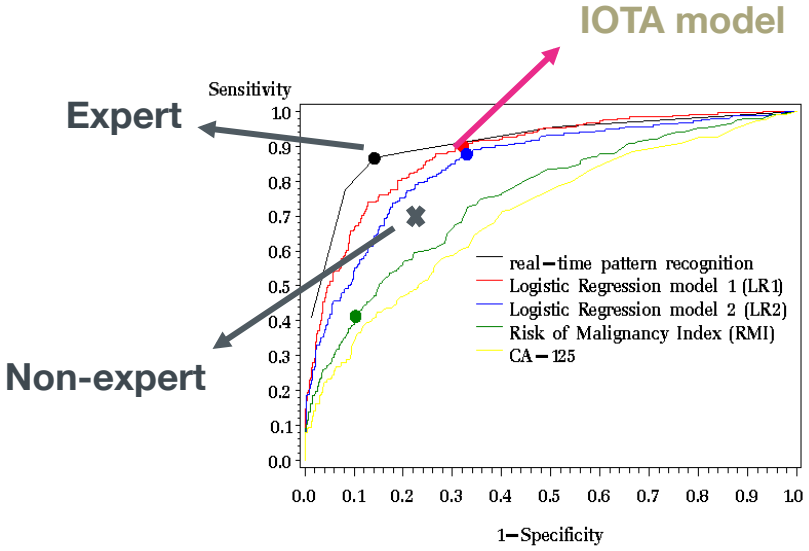
You share, we care !

PROFESSIONAL for clinicians



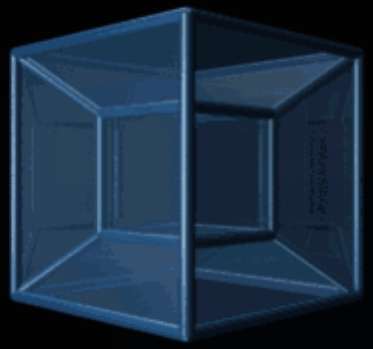
IOTA app to **assess ovarian tumour malignancy**: population based & standardized





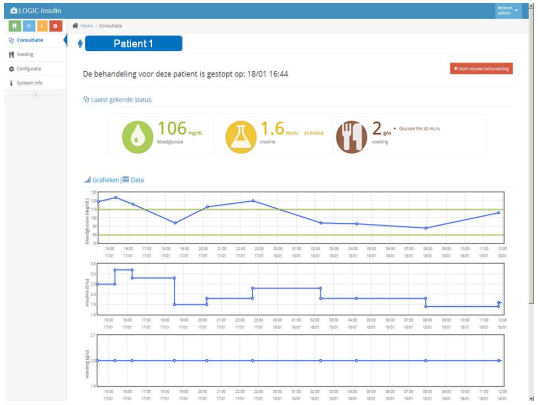
IOTA app available in iTunes app store and on http://homes.esat.kuleuven.be/~sistawww/biomed/iota/

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



Automation Algorithms

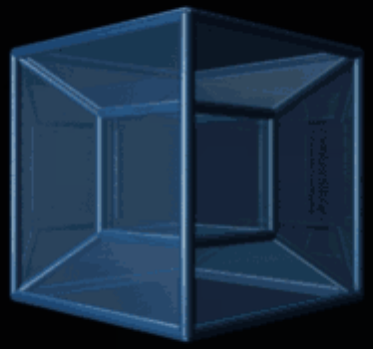
- 10 mio adult ICU patients / year (EU + US) (1-2 b$ market)
- 'Tight Glycemic Control (TGC) in intensive care unit lowers mortality'
  - implement through LOGIC-Insulin: semi-automatic control system that advises nurse on insulin dosage and blood sampling interval aiming at TGC and avoiding hypoglycemia
- LOGIC-I randomized clinical trial (single-centre): compared with expert nurses, LOGIC-Insulin showed improved efficacy of TGC without increasing rate of hypoglycemia
- LOGIC-II randomized clinical trial (multi-centre): Start February 2014

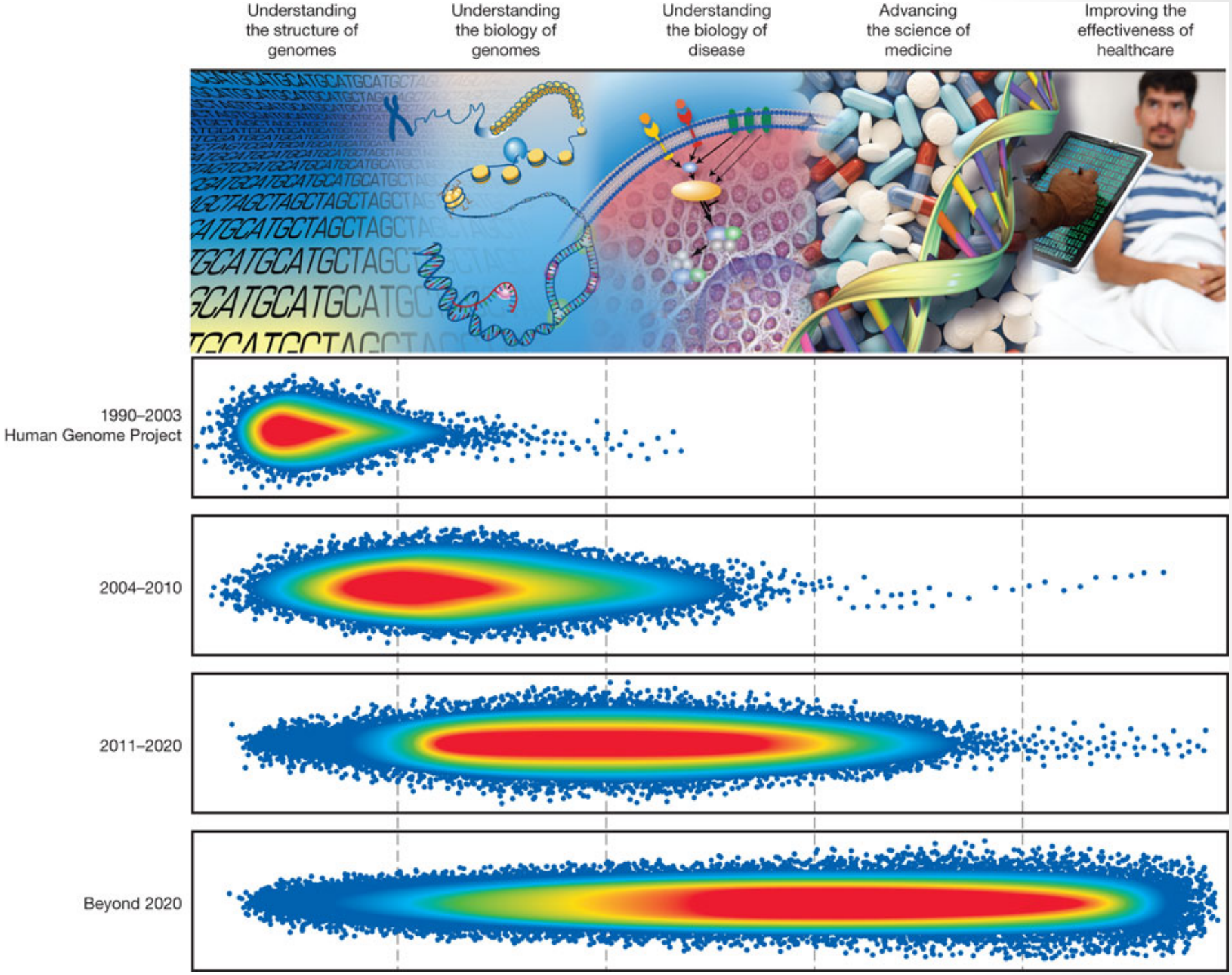


in collaboration with ICU UZ Leuven

UZ LEUVEN

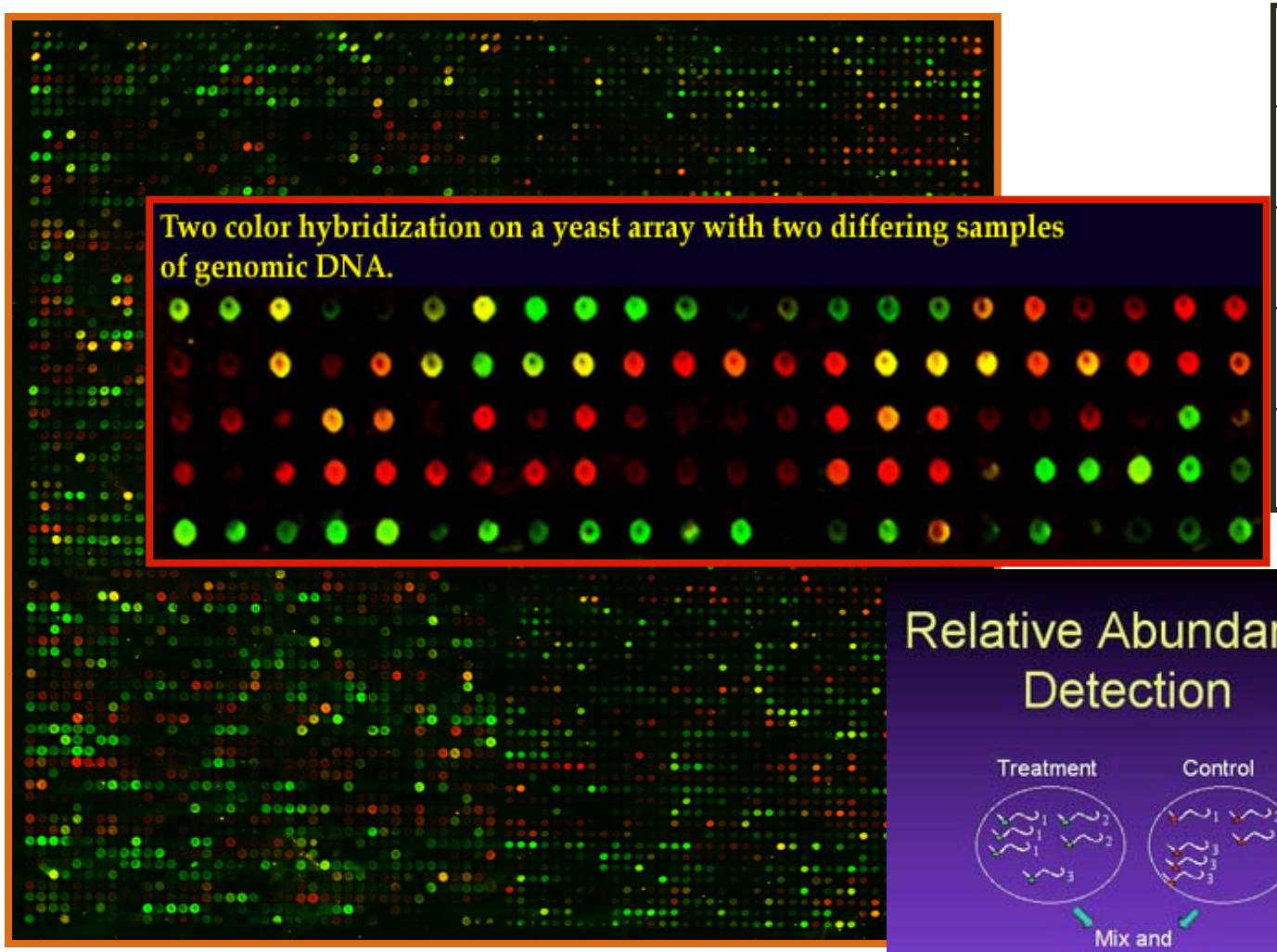
1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



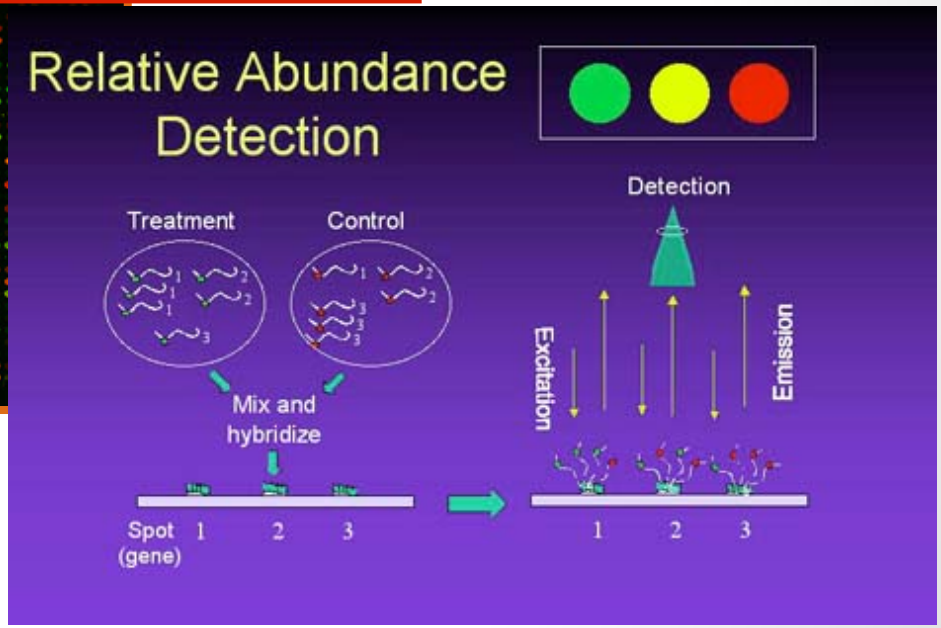


E D. Green *et al. Nature* **470**, 204-213 (2011) doi:10.1038/nature09764

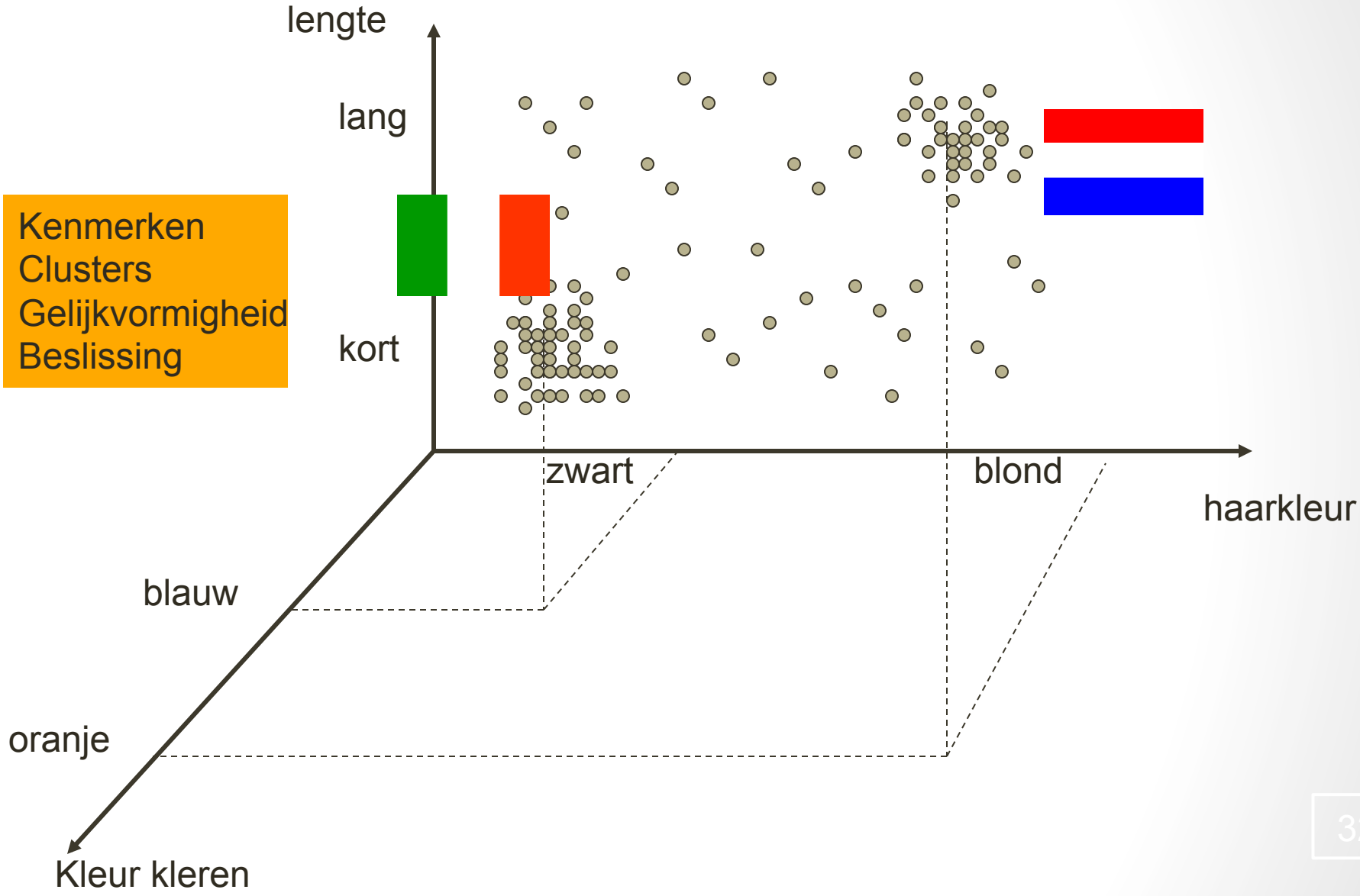
DNA-chips



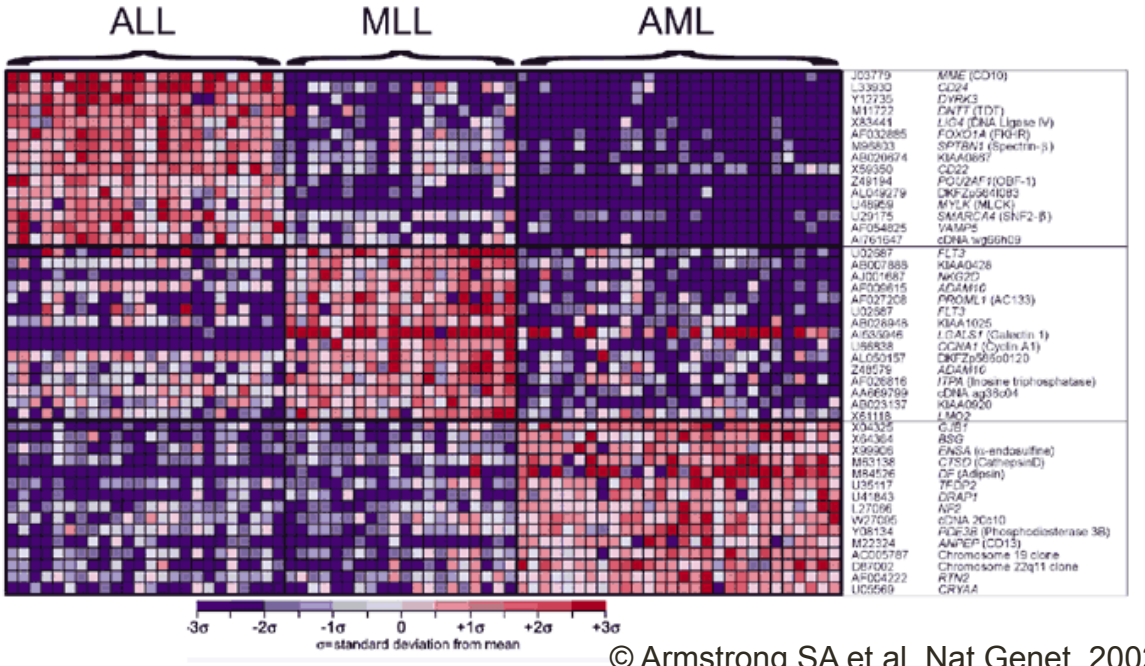
	Test	Ref.
	High	Low
	Low	High
	High	High
	Low	Low

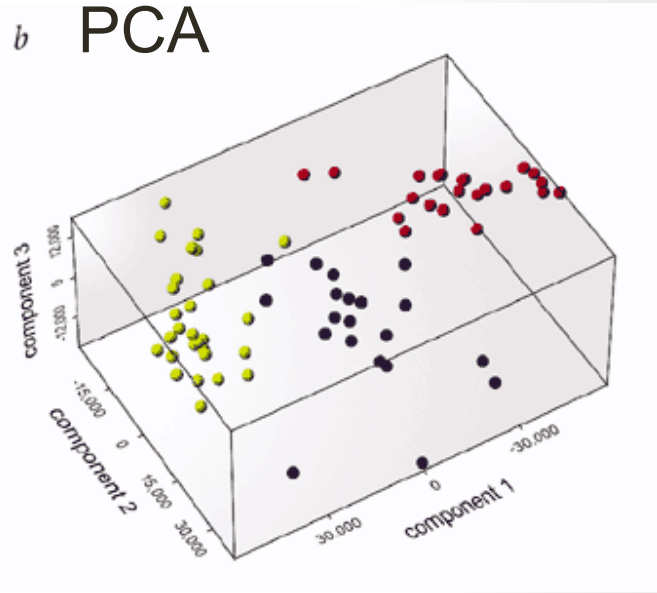


Methodes om te clusteren



ALL/AML/MLL biomarkers





© Armstrong SA et al. Nat Genet. 2002 Jan;30(1):41-7.

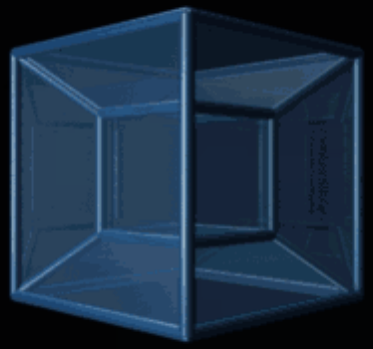
12 600 genes

72 patients:

- 28 Acute Lymphoblastic Leukemia (ALL)
- 24 Acute Myeloid Leukemia (AML)
- 20 Mixed Linkage Leukemia (MLL)

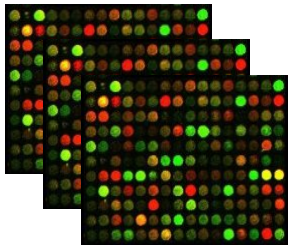
3 patients for each class used as test set

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes

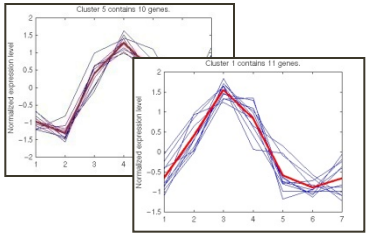


Verschillende databronnen combineren

High-throughput genomics



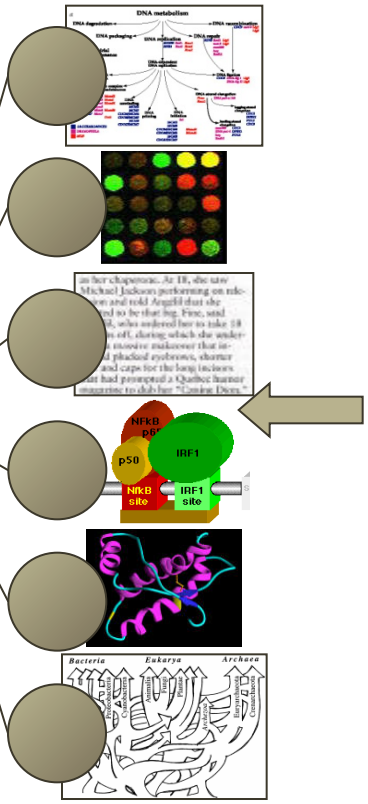
Data analysis



Candidate genes

Name	Ensembl
TTR	ENSG00000118271
PAH	ENSG00000171759
G6PC	ENSG00000131482
IGF1	ENSG00000017427
ALB	ENSG00000163631
CRP	ENSG00000132693
HABP2	ENSG00000148702
IF	ENSG00000138799
FST	ENSG00000134363
ARAF1	ENSG00000078061
HMGA2	ENSG00000149948
C9	ENSG00000113600
PCBP2	ENSG00000111406
HOXB6	ENSG00000108511
RERE	ENSG00000142599
HOXA11	ENSG00000005073
CLIC1	ENSG00000096238
ERCC3	ENSG00000163161
ERCC3	ENSG00000163161
TLL2	ENSG00000095587
SYT4	ENSG00000132872
SYT4	ENSG00000132872
PIK4CB	ENSG00000143393
PKD2	ENSG00000118762
	ENSG00000081026
ANKRD3	ENSG00000183421
F13A1	ENSG00000124491
BPAG1	ENSG00000151914
KCNN3	ENSG00000143603
GRIN2A|GRIN2B	ENSG00000150086
SIM1	ENSG00000112246
	ENSG00000174891
	ENSG00000089195
C14orf10	ENSG00000092020
STX8	ENSG00000170310
	ENSG00000107671
MSH5	ENSG00000096474
CRH	ENSG00000147571
MID1	ENSG00000101871
	ENSG00000184508
	ENSG00000113460
TGFB3	ENSG00000
C1QR1	ENS
N	ENS
PDC	ENS
PE	E
NF	E
NFYA	E
	E
	E
	ENSG
MMP3|MMP1	ENSG00000149968

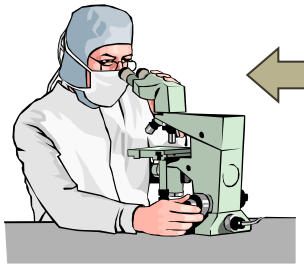
Information sources



Candidate prioritization

Rank	En	Ex	Ip	Ke	GO	Te Avg	Pval
1	TTR	G6PC	PAH	G6PC	IGF1	TTR	TTR
2	IGF1	TTR	IGF1	PAH	PAH	IGF1	PAH
3	CRP	ALB	TTR	RERE	G6PC	CRP	G6PC
4	HOXB6	HABP2	ALB	ERCC3	TTR	HOXB6	IGF1
5	ALB	PAH	HDC	ERCC3		ALB	ALB
6	NR4A2	IF	TLL2	ANKRD3		HMGA2	CRP
7	PAH		C1QR1	ARAF1	HDC	NR4A2	
8	HOXA11	IGF1	G6PC	PKD2	F13A1	PAH	IF
9	NFYA	CRP	HABP2	MTMR1	KCNN3	HOXA11	C13orf7 FST
10	C9	ARAF1	IF	HDC	CLIC1	NFYA	TTR ARAF1

Validation



Aerts et al, Nature Biotechnology, 2006

genomic data fusion:
trace disease-causing variants

20x

more accurate

Sifrim, Popovic et al,
Nature Methods, 2013

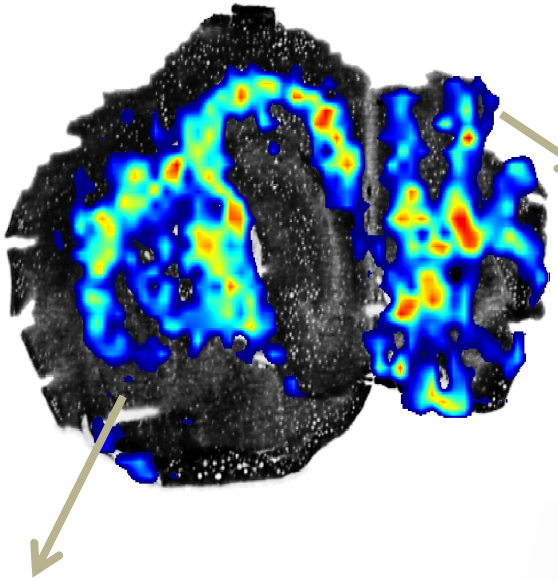
ONDERZOEK

Vlamingen sporen genetische ziektes accurater op

Onderzoekers van iMinds (het vroegere IBBT) en de KU Leuven hebben software ontwikkeld die enorme hoeveelheden genetische data kan doorzoeken en die aanbevelingen doet over de meest waarschijnlijke oorzaak van een erfelijke ziekte. Het 'eXtasy' maakt gebruik van artificiële intelligentie en werkt tot twintig keer beter dan andere software. Volgens professor Yves Moreau zit de software nog in de onderzoeksfase en is nog één tot twee jaar nodig voor de technologie commercieel beschikbaar is.

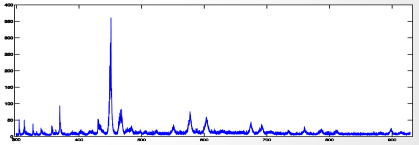
Bron:
De Tijd, woensdag 23 oktober 2013

mass spectrometry imaging:
true molecular imaging for disease characterization

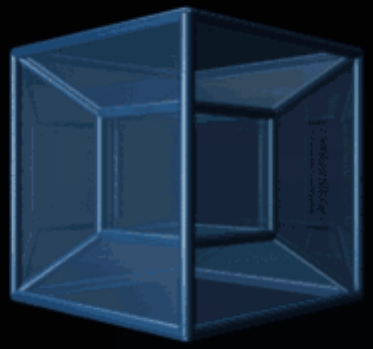


use IT to extract whatever information you need

full molecular profile for every pixel



1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



CARTAGENIA

- BENCH = web-based software and database platform for interpretation of genomic variation in routine diagnostics
- array-CGH + Next Gen Seq: first diagnostics grade solution for NGS data based diagnostics in the world
- SaaS go-to-market model
- Leuven + US office
- large customer base of diagnostic labs, private labs, academic institutes, and consortia in Europe, Northern America and Australia
- rare genetic disorders, extension towards cancer and prenatal

building on expertise in big data & machine learning

CARTAGENIA

BENCH is used in ca. 50 accredited genetic labs worldwide

Support | Careers | News | Contact

Heverlee + US office

Home | Software | Knowledge | Enterprise | Company

Knowledge, software and services for efficient patient genetics.

Cartagenia BENCH is a web-based software and database platform rich in features and knowledge sources geared at interpretation of genomic variation in routine diagnostics.

Discover Cartagenia Bench

routine diagnostics

For genetics labs

- BENCH lab NGS
- BENCH lab CNV
- Knowledge libraries

For clinicians

- BENCH clinic
- BENCH consortium
- Knowledge bases

For hospitals

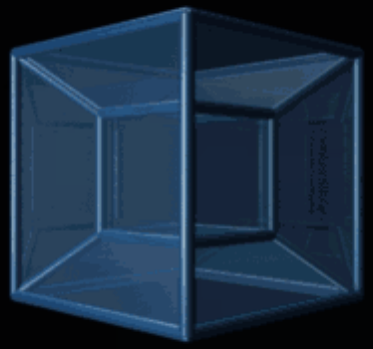
- Customized solutions
- Enterprise integration
- Consulting

tools for non-IT users

first diagnostics grade solution for Next Generation Sequencing data based diagnostics in the world

originating from research in rare genetic disorders in close collaboration with Centre Human Genetics UZ Leuven

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



health house

EXPERIENCE THE FUTURE

A first in kind exhibition platform on the future of health and care

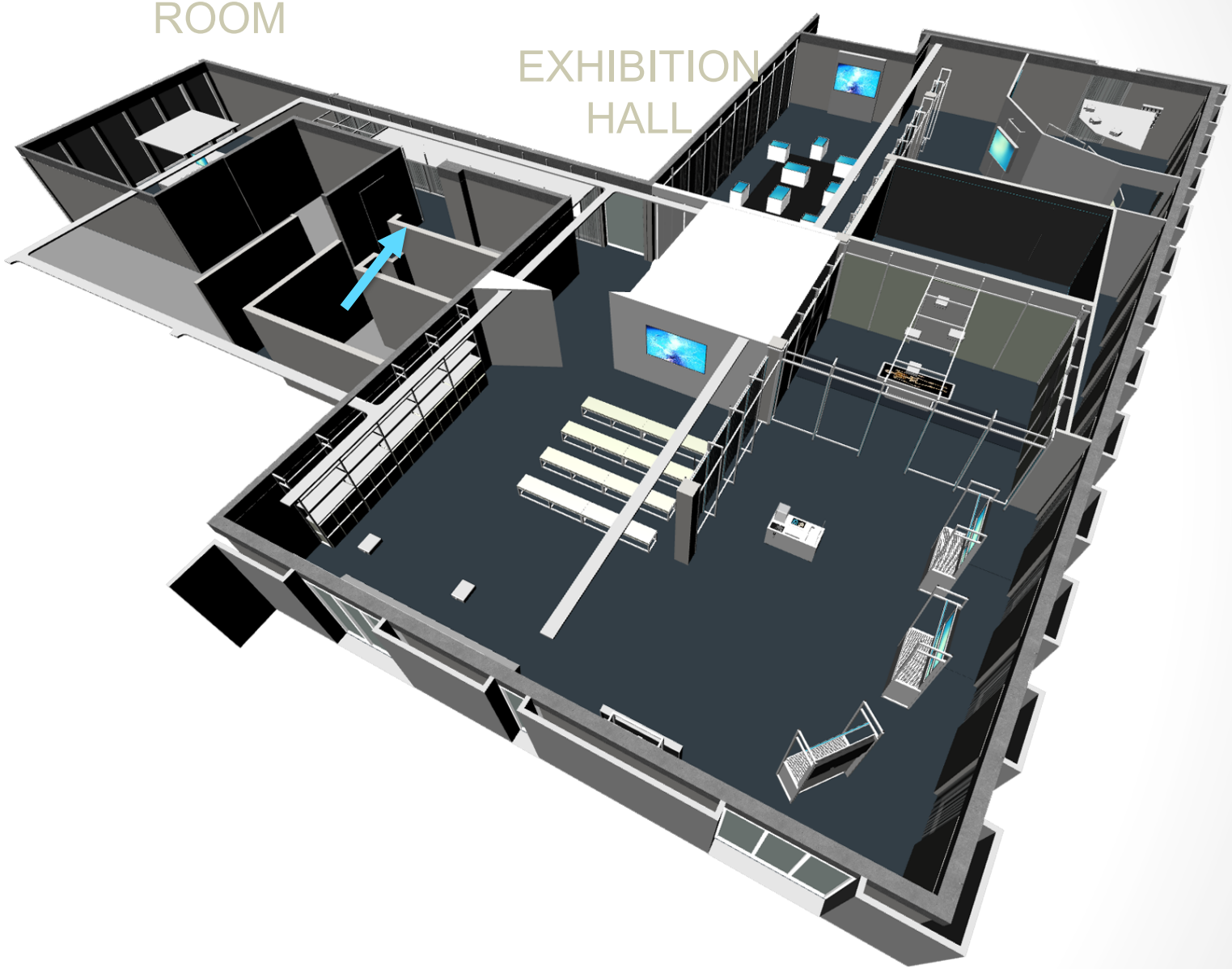
FOUNDING PARTNERS

A first-in-kind exhibition platform on the future of health & care

- Allowing flexible programming of tailored exhibitions
- Featuring class-leading content & astonishing visualizations
- Enabling organizational and thematic storytelling > dissemination > interaction
- Creating cross-domain synergies between health, technology & creativity

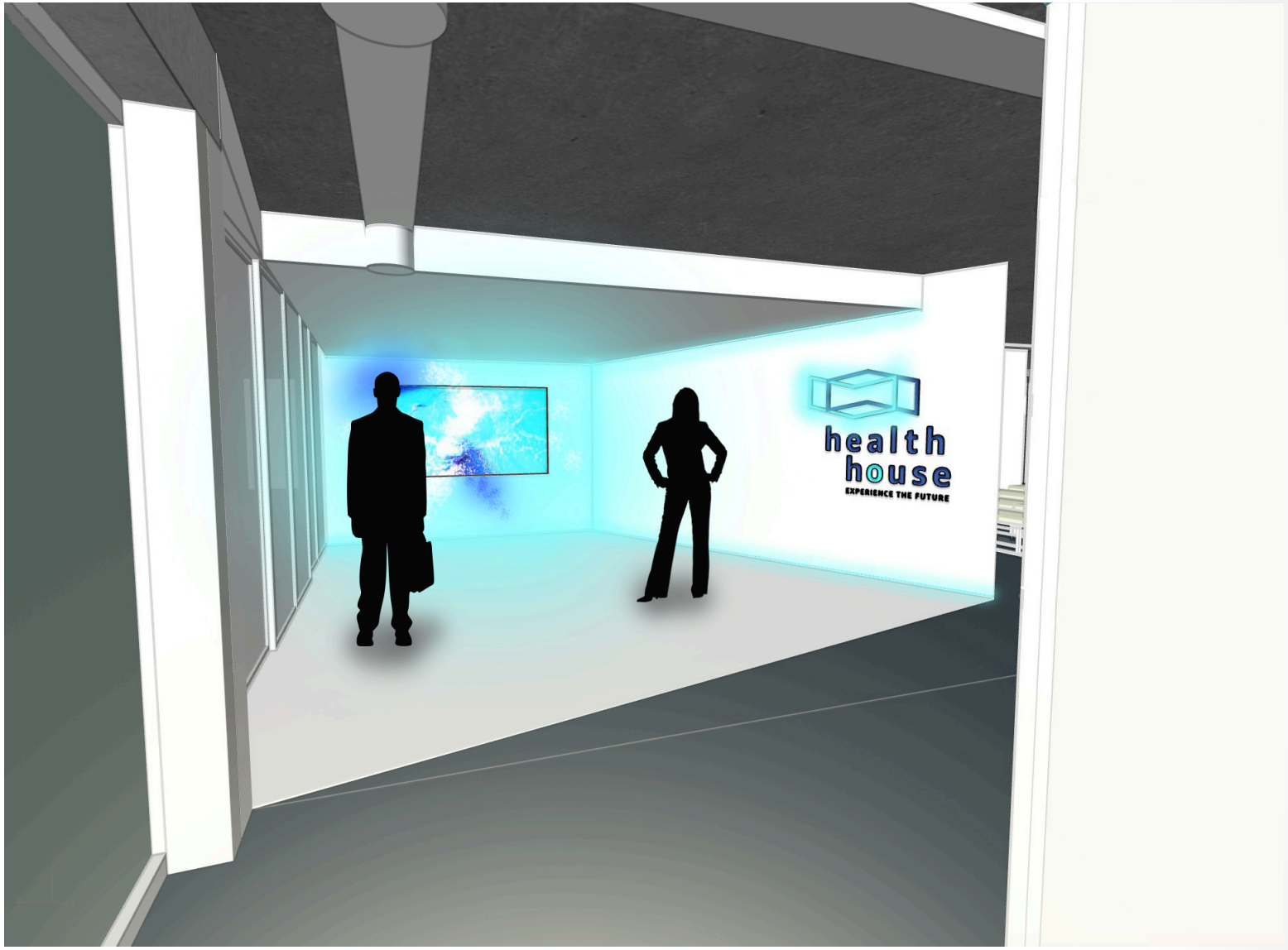
DEMO
ROOM

EXHIBITION
HALL

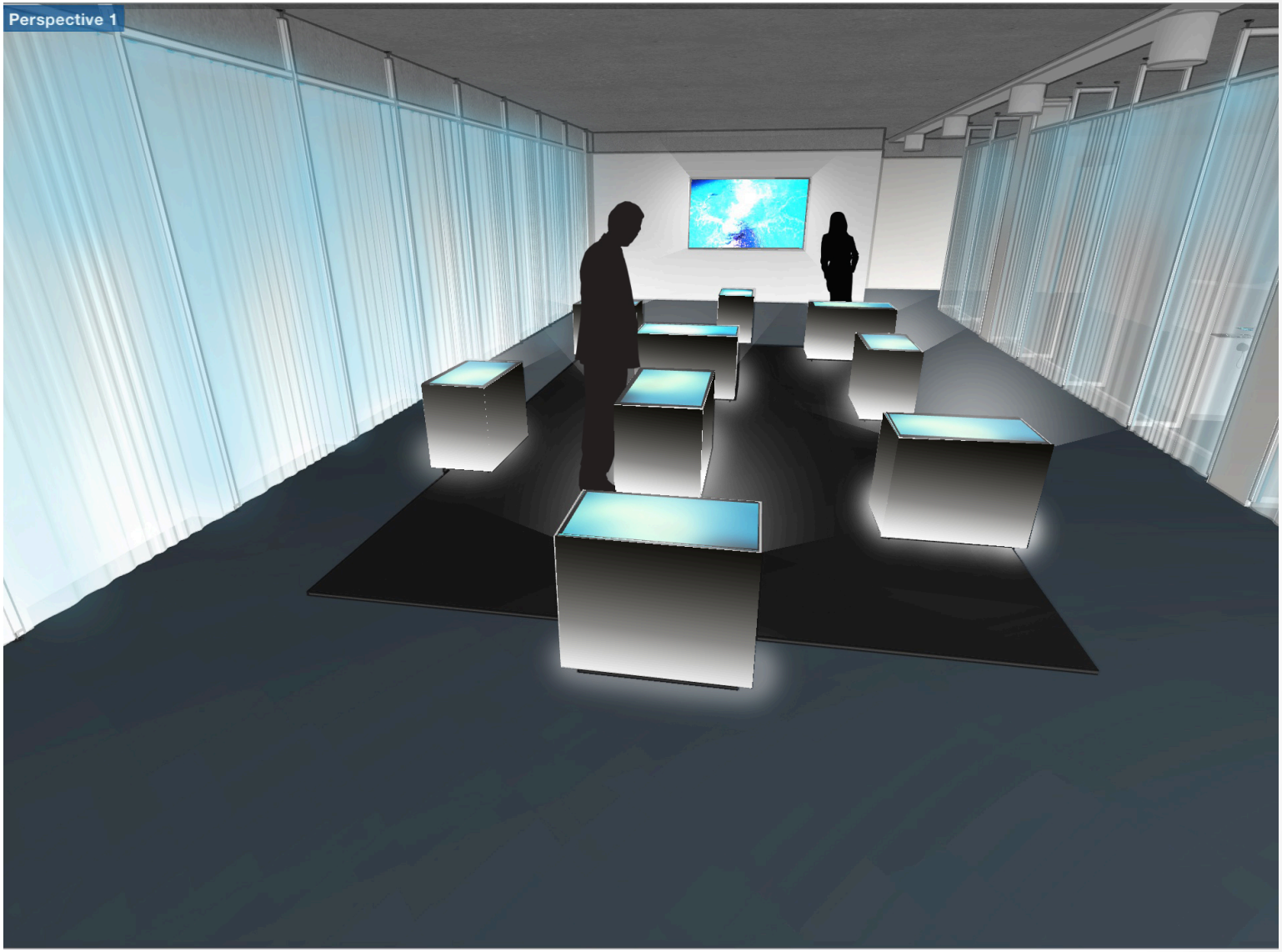


WALK THROUGH

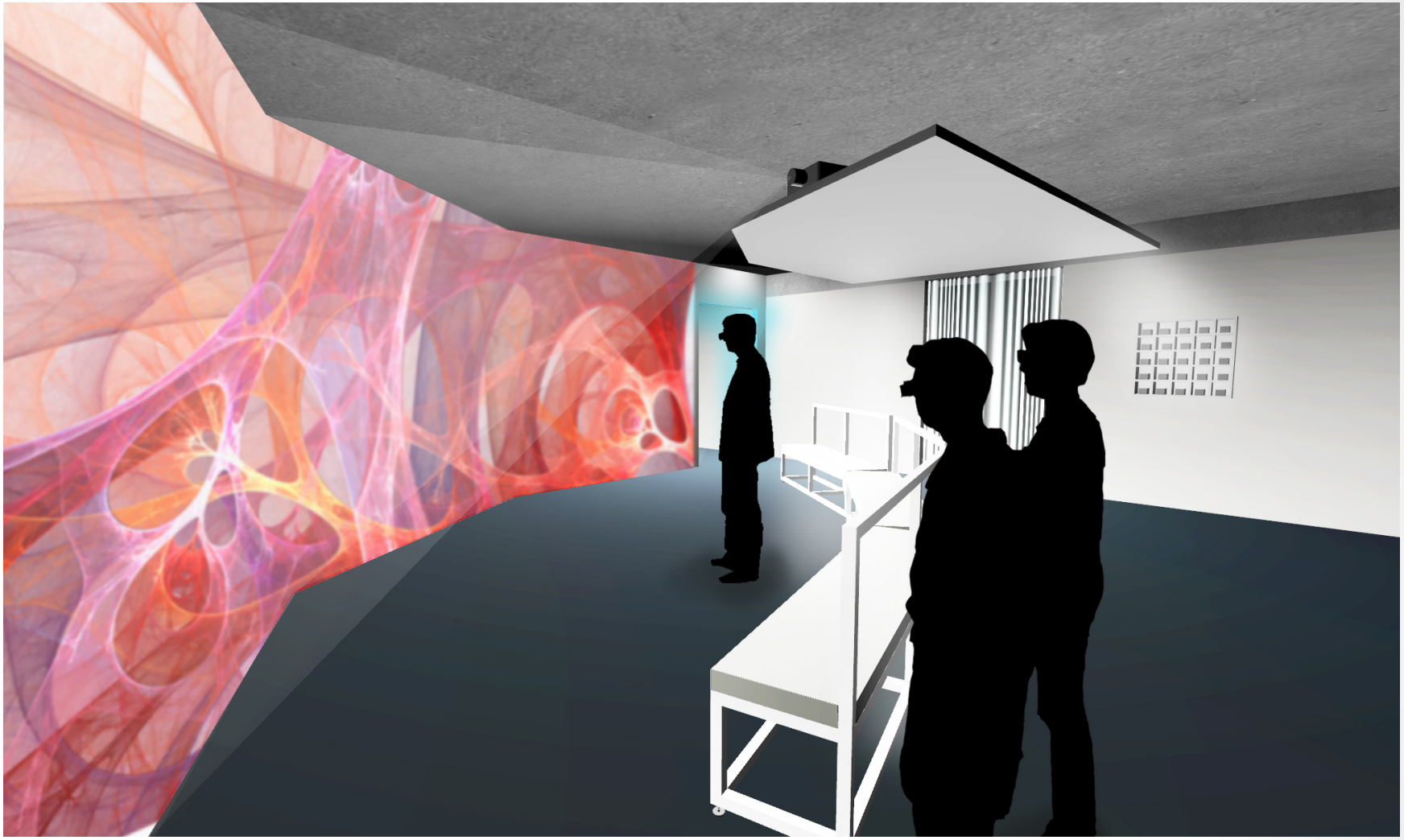
The Tunnel



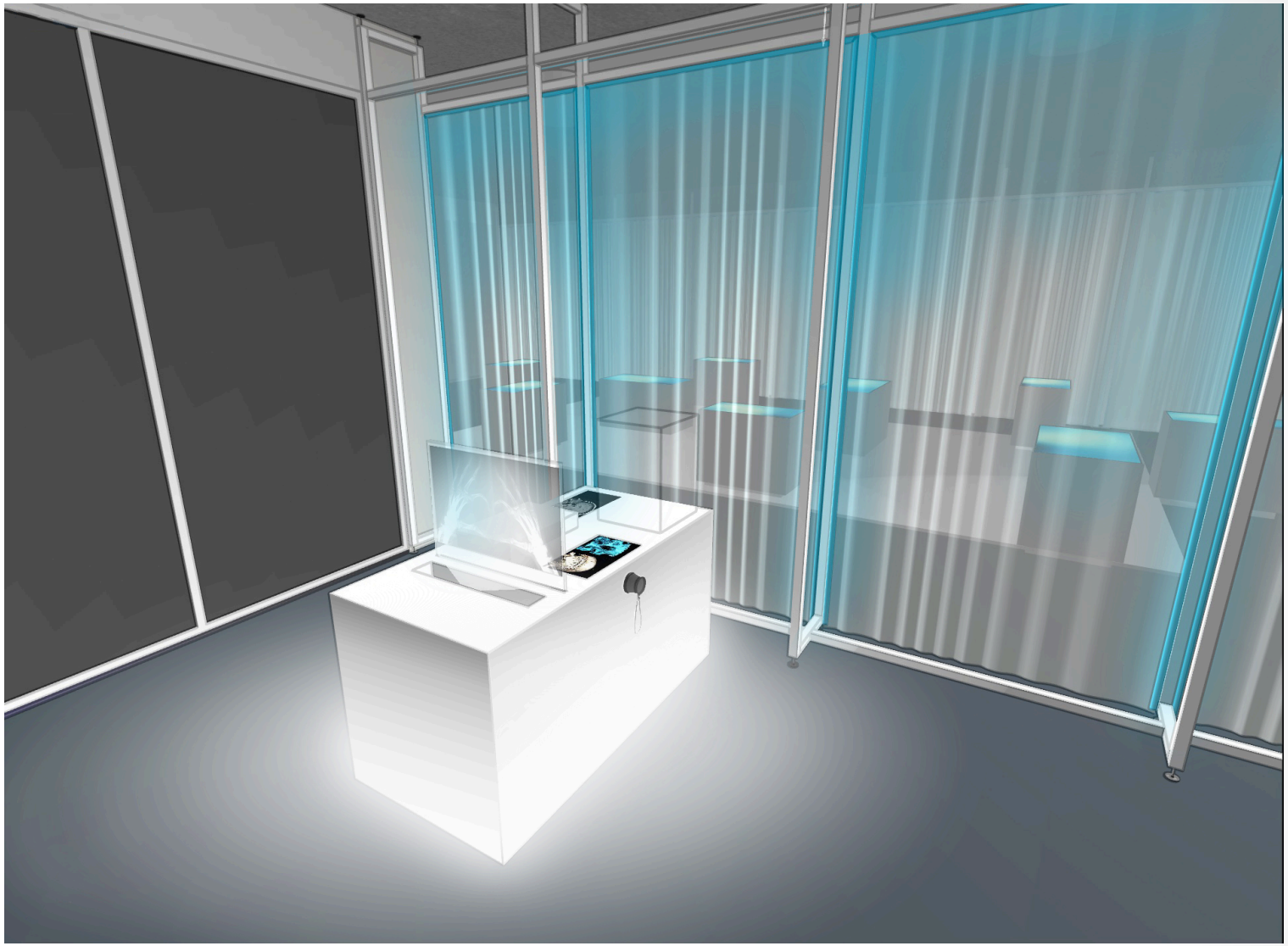
The Gateway



The Labyrinth



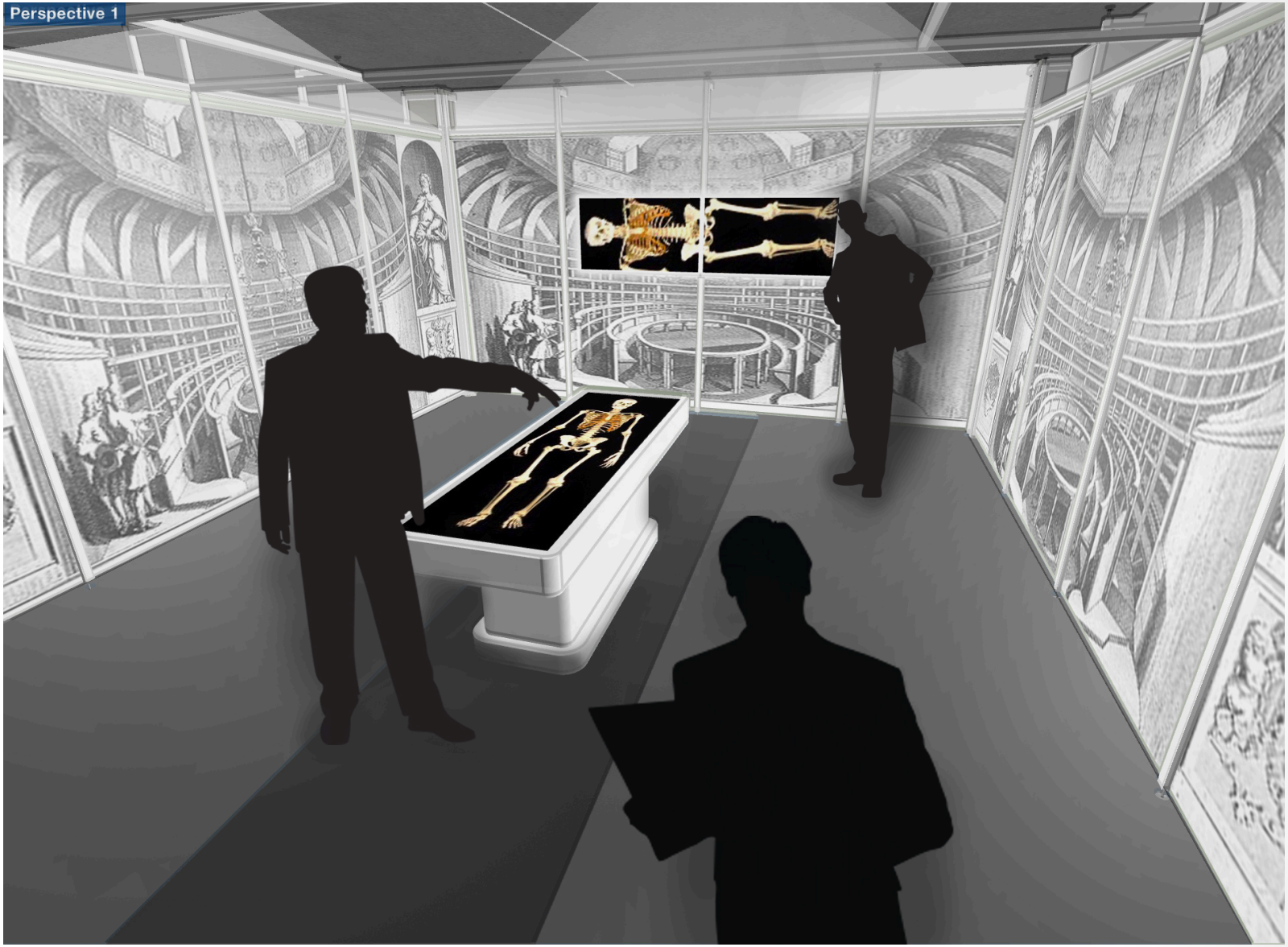
The Cave



The Artefactum I & II

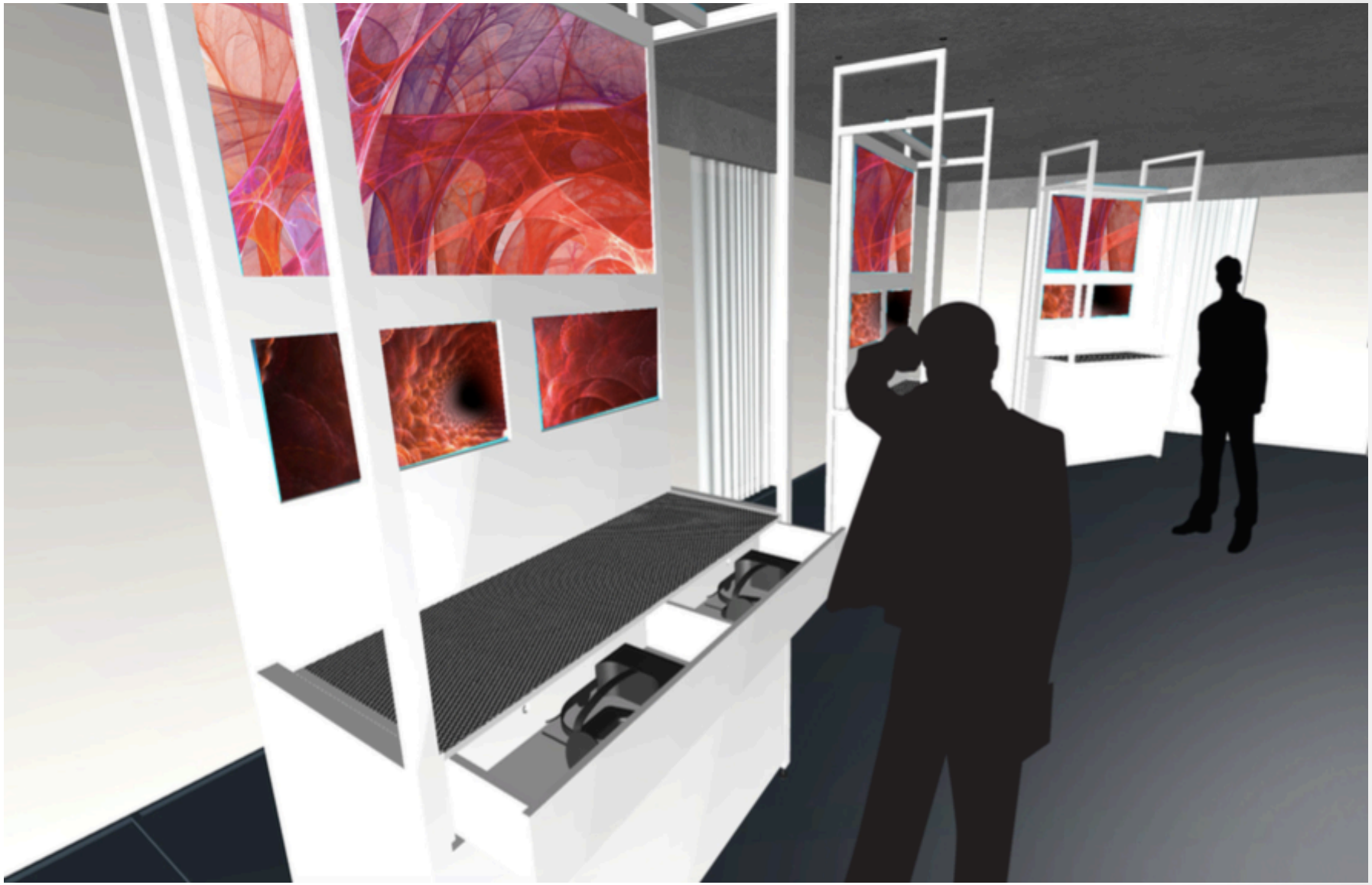
The Interactum

Kinect in medical applications



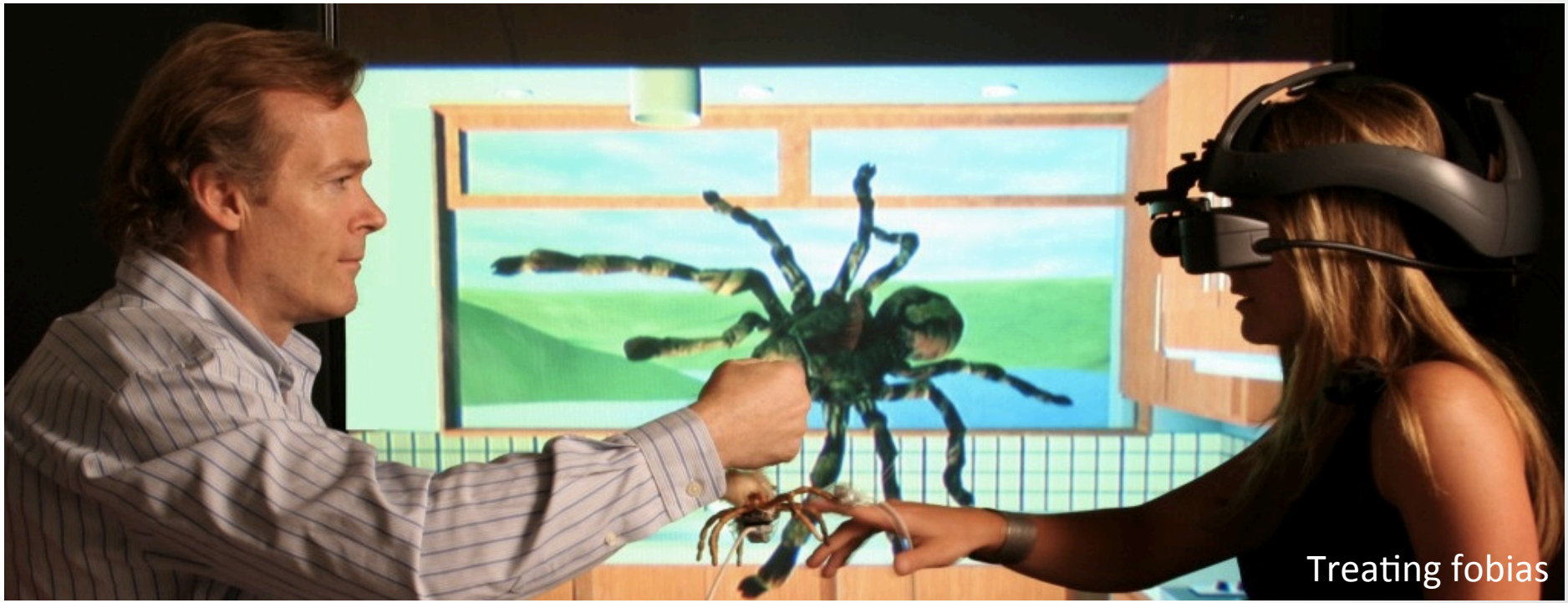
The Corpus

Anatomage Table in medical applications

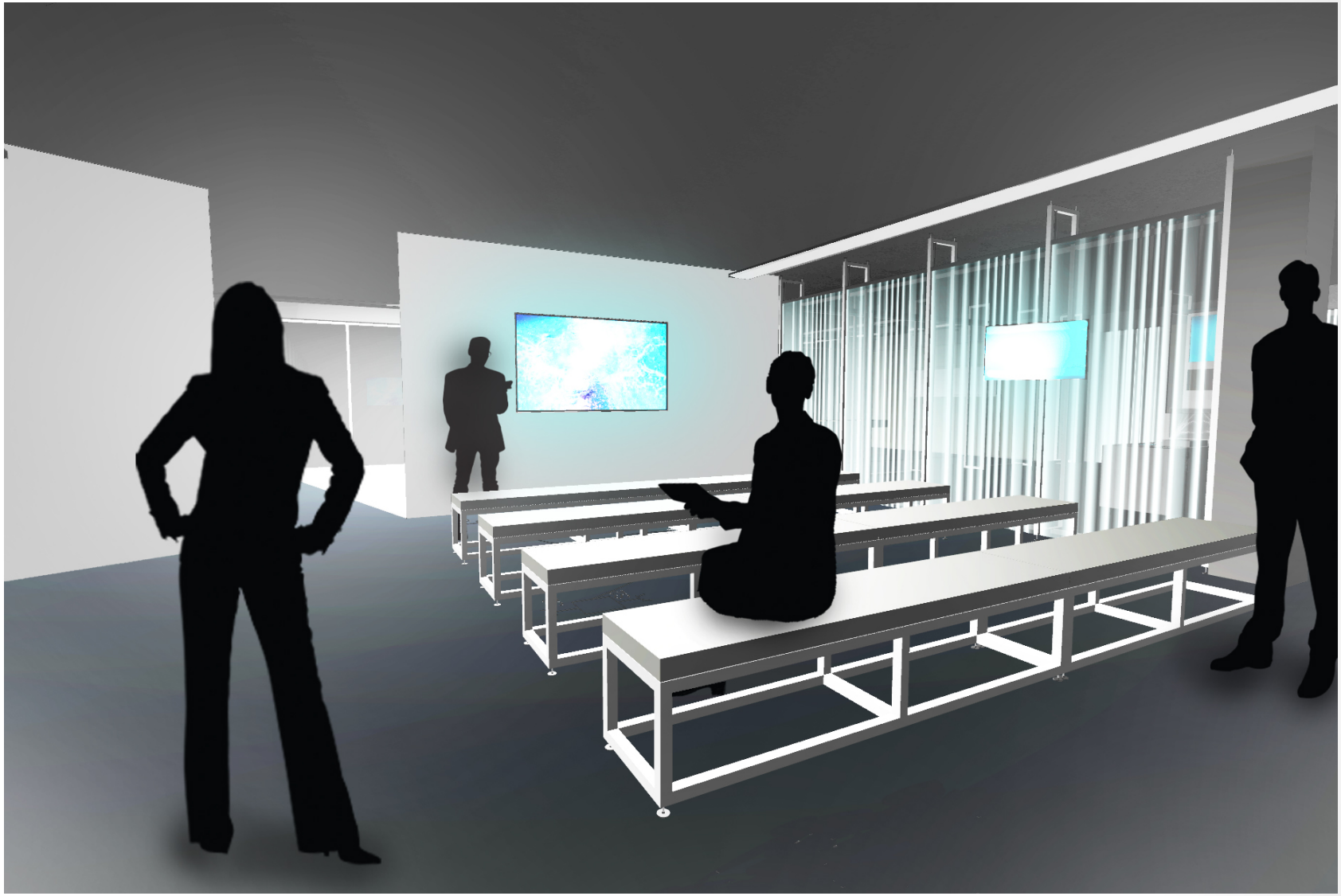


The Visionarium

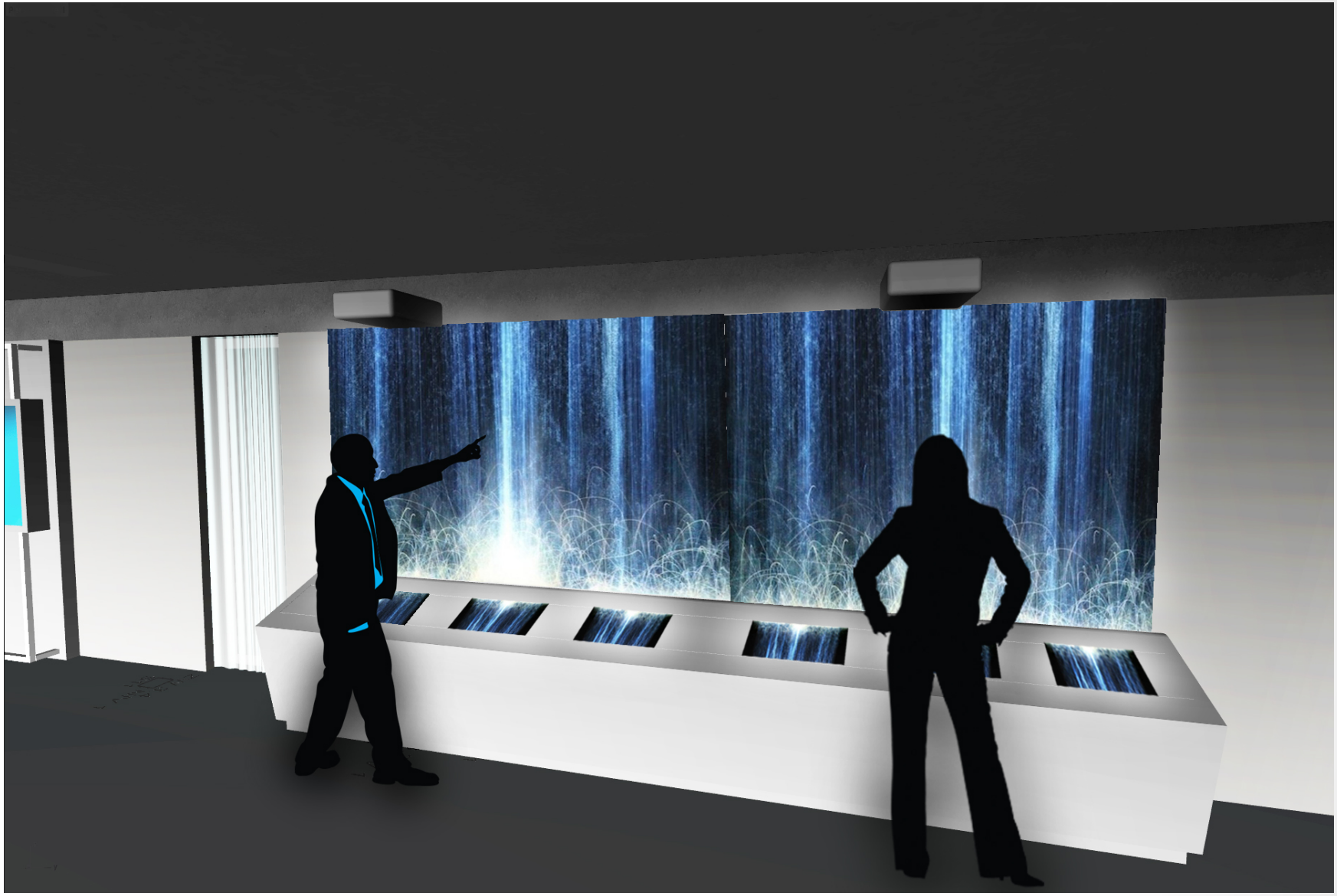
VR/AR in medical applications



The Mosaic I & II



The perspectum



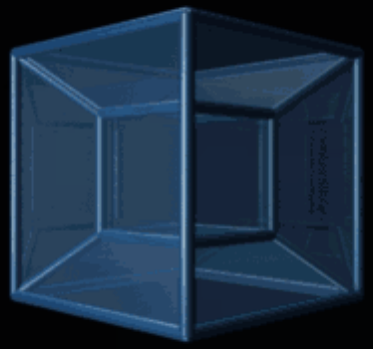
The source

The demo room

PUBLIC SECTOR	PRIVATE SECTOR	ASSOCIATIONS
• SOC's • Hospital(network)s • Universities, UC's, management schools • Foreign top tech TTO's • Flemish, Federal, EU government agencies • FIT • …	• Pharma companies • Technology companies • Medical event agencies • Banking & insurance • Trend & mktng agencies • Portfolio companies • Consultancy firms • Every thinkable vertical • …	• LOK's • Patient organizations • Startup/KMO support initiatives • Industry associations • General business networks • Health & Tech networks • …

www.health-house.be

1. Technology will be key
2. Societal trends
3. DSS for 3P
4. Examples
    1. Ovarian Cancer
    2. Glycemia in ICU
    3. Leukemia
    4. Gene prioritization
    5. Cartagenia
5. Health House
6. Obama concludes



Obama

But in order to lead in the global economy and to ensure that our businesses can grow and innovate, and our families can thrive, we're also going to have to address the shortcomings of our health care system.

The Recovery Act will support the long overdue step of ***computerizing America's medical records***, to reduce the duplication, waste and errors that cost billions of dollars and thousands of lives. But it's important to note, ***these records also hold the potential of offering patients the chance to be more active participants in the prevention and treatment of their diseases***. We must maintain patient control over these records and respect their privacy. At the same time, we have the opportunity to offer billions and ***billions of anonymous data points to medical researchers who may find in this information evidence that can help us better understand disease***.

History also teaches us the greatest advances in medicine have come from scientific breakthroughs, whether the discovery of antibiotics, or improved public health practices, vaccines for smallpox and polio and many other infectious diseases, antiretroviral drugs that can return AIDS patients to productive lives,
pills that can control certain types of blood cancers, so many others.

Because of recent progress –- ***not just in biology, genetics and medicine, but also in physics, chemistry, computer science, and engineering –- we have the potential to make enormous progress against diseases in the coming decades***. And that's why my administration is committed to increasing funding for the National Institutes of Health, including $6 billion to support cancer research -- part of a sustained, multi-year plan to double cancer research in our country. (Applause.)

http://www.whitehouse.gov/blog/09/04/27/The-Necessity-of-Science/