

Netherlands eScience Center

ICT Synergy Hub, Amsterdam

Research & Innovation in the Big Data Era

CWI in Bedrijf

Centrum Wiskunde & Informatica

Op 5 oktober 2012

Prof. dr. Jacob de Vlieg ^{1 2}

1. CEO & Scientific Director of Netherlands eScience Center, NWO-SURF

2. Head Computational Design & Discovery, Group, CMBI, Radboud University, Medical Center, Nijmegen, Netherlands

netherlands

eScience center

by SURF & NWO

Science itself is changing ...We need to change with it...

Agenda presentation

- Netherlands eScience Center (NLeSC)
 - Bridging Research and ICT
- Public-private *basic* research projects in the life sciences
- Outsourcing basic discovery (pharma, biotech, other)



The Global Pharmaceutical Dilemma



R&D Productivity in Big Pharma is falling

- Increasing cost of drug development (~\$1.7 billion/per approved drug)
- Patient safety and effectiveness remains a major concern

Reasons for lack of output and increased R&D cost

- Increased regulatory requirements
- Complexity of human biology not recognized
- Model systems in preclinical insufficiently predictive for efficacy and safety in man

The Global Pharmaceutical Dilemma



R&D Productivity in Big Pharma is falling

- Increasing cost of drug development (~\$1.7 billion/per approved drug)
- Patient safety and effectiveness remains a major concern

Reasons for lack of output and increased R&D cost

- Increased regulatory requirements
- Complexity of human biology not recognized
- Model systems in preclinical insufficiently predictive for efficacy and safety in man

New opportunities & challenges

- Biobanking: availability of patient samples
- New technologies: omics, imaging, simulations, etc
- Difficulties to realize full potential of new technologies due to data problem ; Data-Data-Data

Not the generation but the management of data has become the central challenge

The recent World Economic Forum has deemed *'data to be a new form of currency'*

The
Economist

Business

Welcome to the yotta world

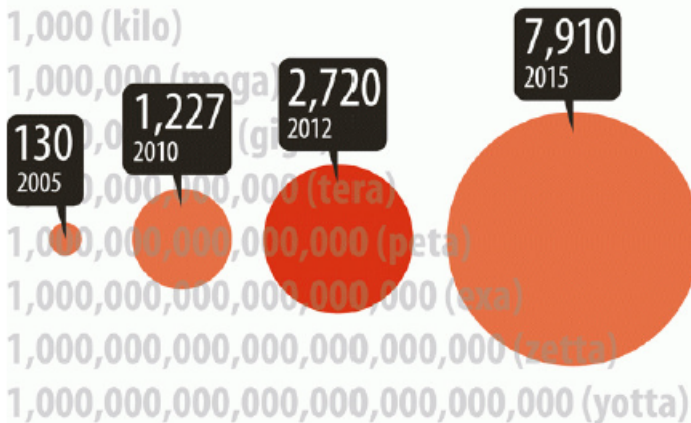
Big Data will flood the planet

Nov 17th 2011 | from The World In 2012 print edition

Even if you still have to think twice about the meaning of "giga" and "tera" in

Exponential

Quantity of global digital data, exabytes



Source: EMC/IDC Digital Universe Study, 2011



Digital Information Created, Captured, Replicated Worldwide



Source: IDC, 2006

eScience: *enhanced* Science

1. Huge amounts of data produced in all scientific disciplines
 2. Cross-seeding of technologies inspired by new collaborations
- New techniques needed to explore & connect massive datasets
 - Cross-type data integration
 - Data-driven & multi-models simulations
 - Visualization & analytics
 - Extreme computing: *connected computers & fast networks.*
 - *Any combinations thereof*
 - Reinventing science: new ways to do science not possible without computing



Netherlands eScience Center



Netherlands organization
for scientific research:



Principal Dutch body for
ICT innovation for research &
business processes



**Synergy ICT hub for
research; SARA, EGI**

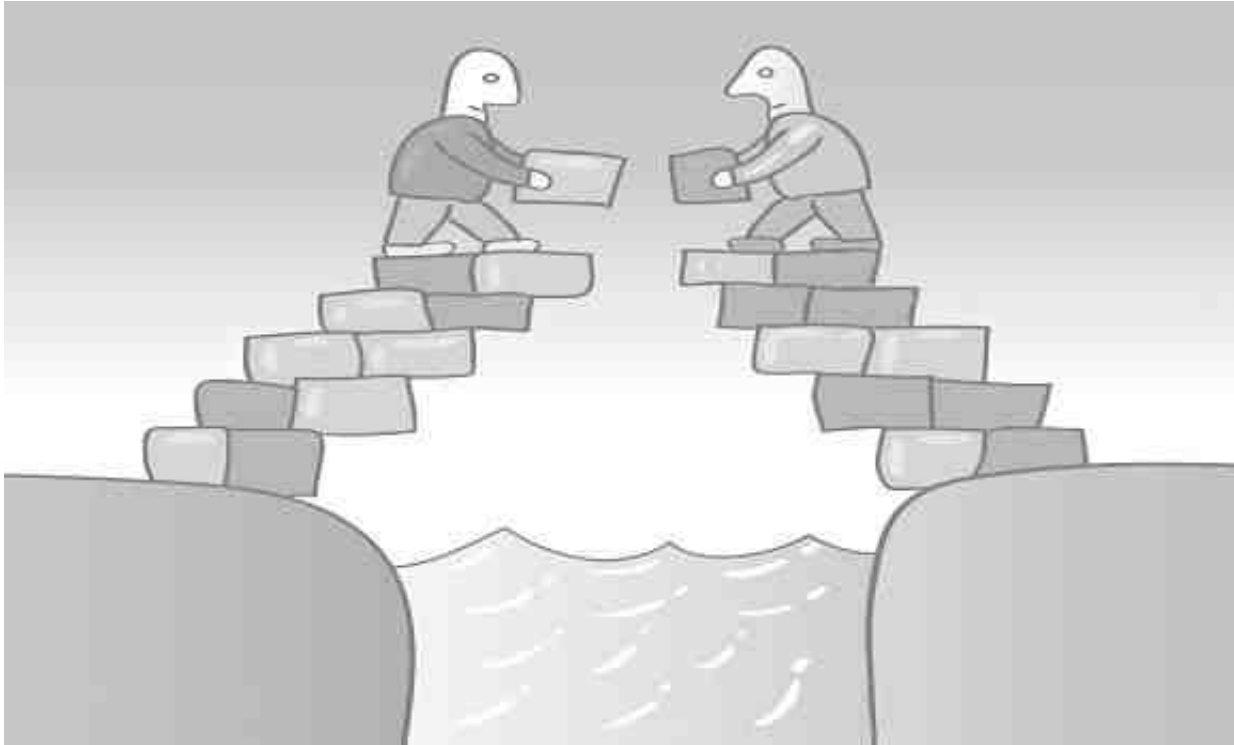
Network organisation:

**CWI, UvA, VU, Radboud,
,KNAW, CTMM, NBIC,
companies, etc`, etc**

Expert Centre for Big Data Analysis



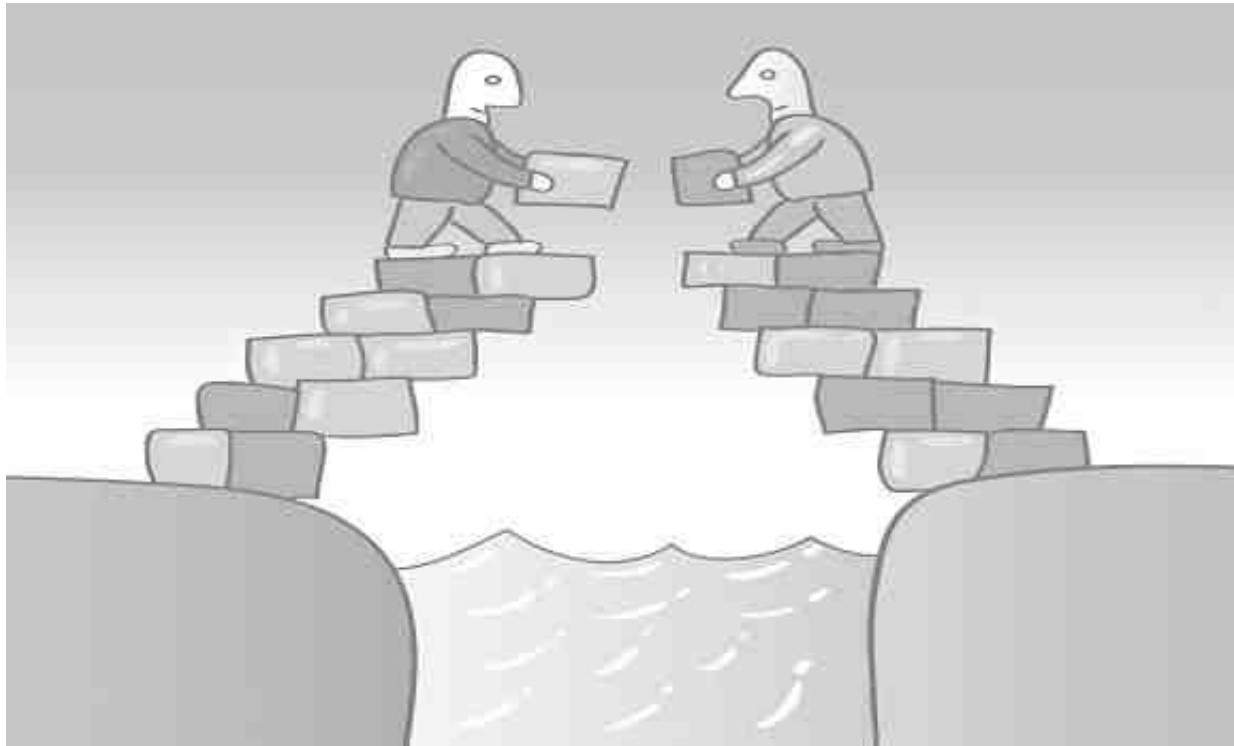
NLeSC: Innovation with ICT



Deliver output in terms of science & business value



NLeSC: Innovation with ICT



Deliver output in terms of science & business value

Valorization is a fundamental component of the vision of the NLeSC

e.g. sustainable solutions, hotel functions; eScience engineers, collaboratorium, etc.

Cyber-common: a facility for 21st century data-driven research and multidisciplinary team work

To link minds and information



Cyber-common: a facility for 21st century data-driven research and multidisciplinary team work

Connect Demand and Supply

To link minds and information



NLeSC themes: *research projects inspired by new collaborations*

- **Sustainability & Environment**

- Climate
- Water management
- Energy
- Ecology

- **Chemistry & Materials**

- Chemistry

- **Humanities & Social Sciences**

- Humanities
- Social Sciences

- **Life Sciences**

- Green Genetics
- Translational Research IT
- Foods
- Cognition/Neuroscience

- **eScience Methodology & 'Big Data'**

- eScience Methodology
- Astronomy



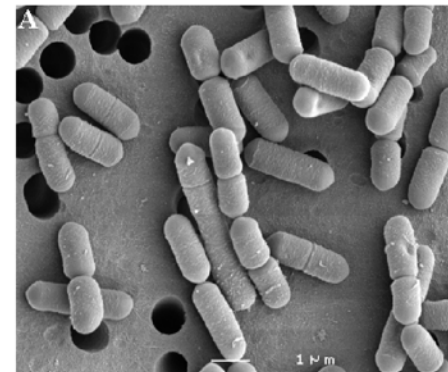
NLeSC Project Portfolio: *Can scientists from digital humanities help food researchers?*

Food Research: Food Specific Ontologies for Food Focused Text Mining

Project Leader: Wynand Alkema



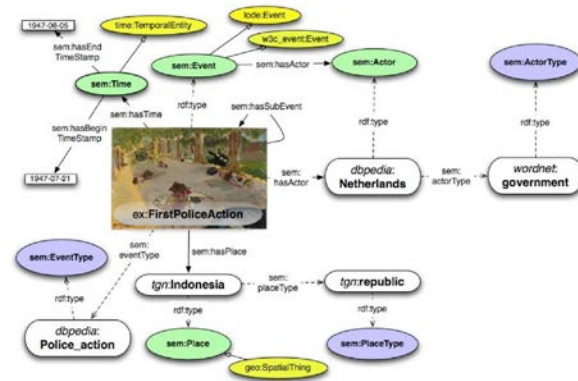
Addressing absence of domain specific structured vocabularies which limits the use of data mining & knowledge management methods in food research.



eHumanities: BiographyNED

Project Leader: Guus Schreiber

Will improve current version of the Biography Portal by incorporating analytical tools to show interconnections, trends, geographical maps and time lines.



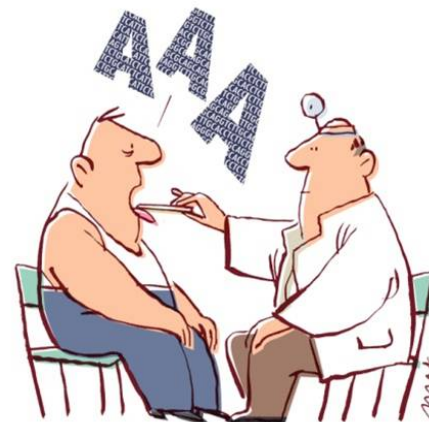
NLeSC Project Portfolio

Life Sciences: TraIT (Translational Research IT)

Project Leader: Jan Willem Boiten



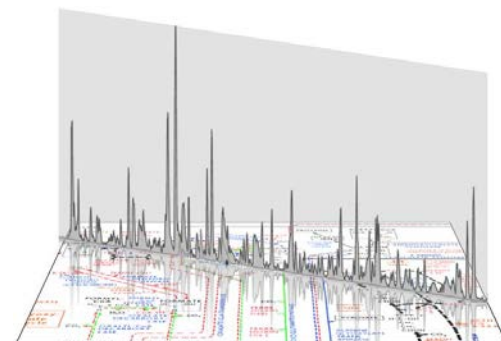
Adapting existing software solutions to provide the Netherlands with an IT infrastructure to facilitate translational research.



eChemistry: Integrative Chemical Metabolomics Data Analysis

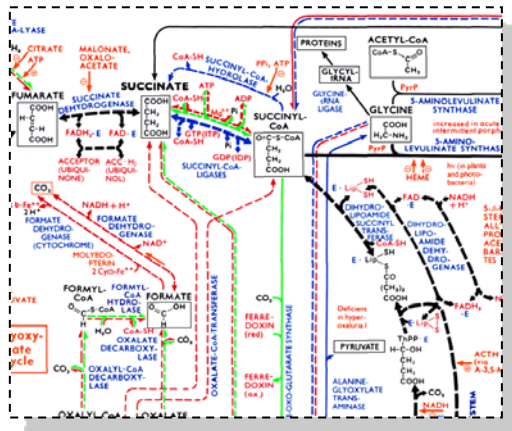
Project Leader: Lars Ridder

Developing a computational workflow to improve and accelerate metabolite identification and biochemical pathway reconstruction for metabolomics.

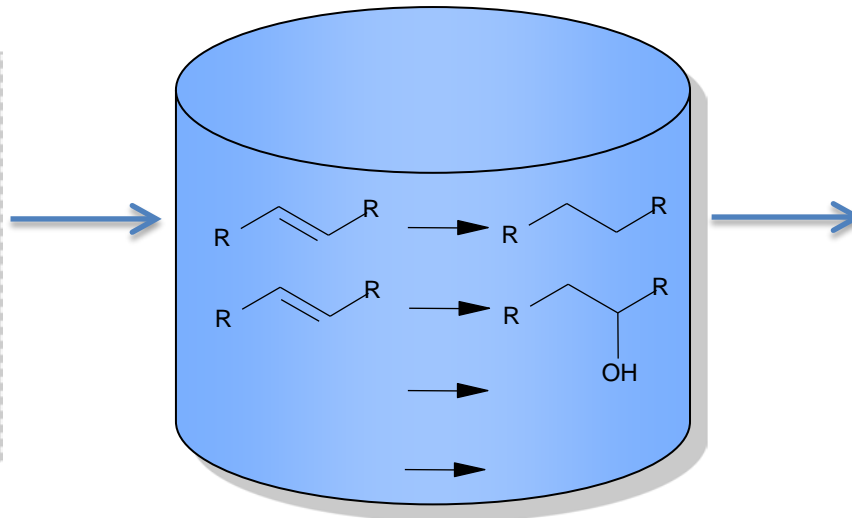


Bridging chemistry & biology in the computer

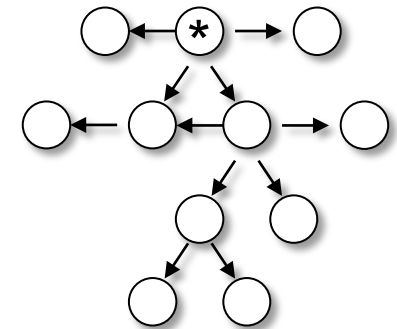
Biochemical knowledge



Derive generic biochemical rules

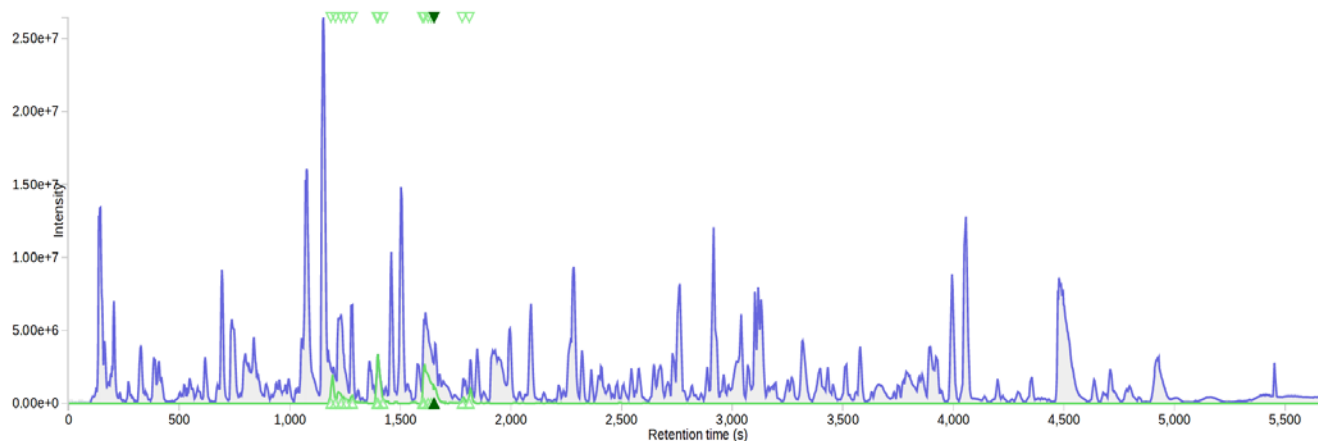
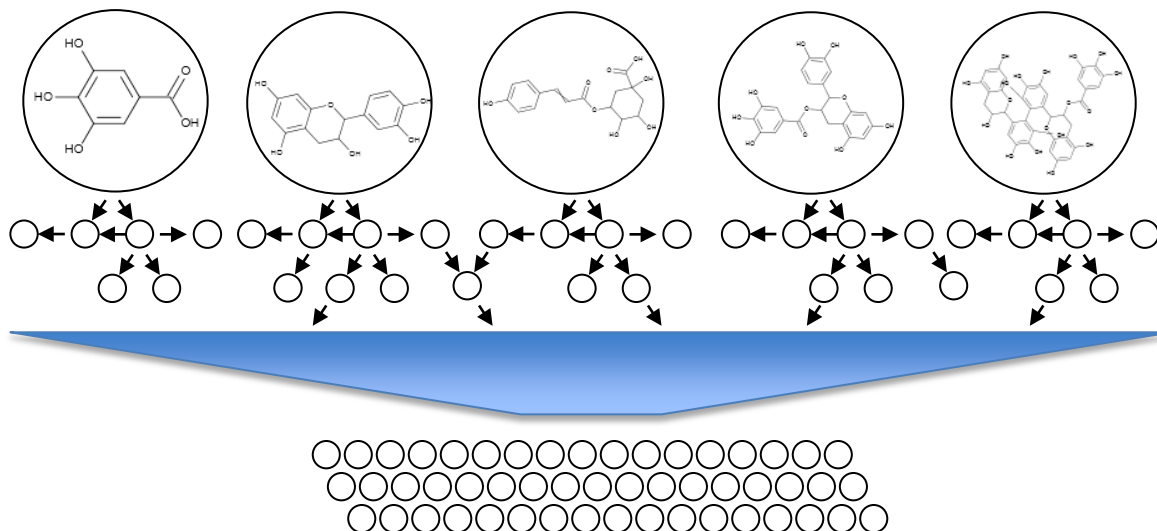


Generate “biochemically feasible” molecules

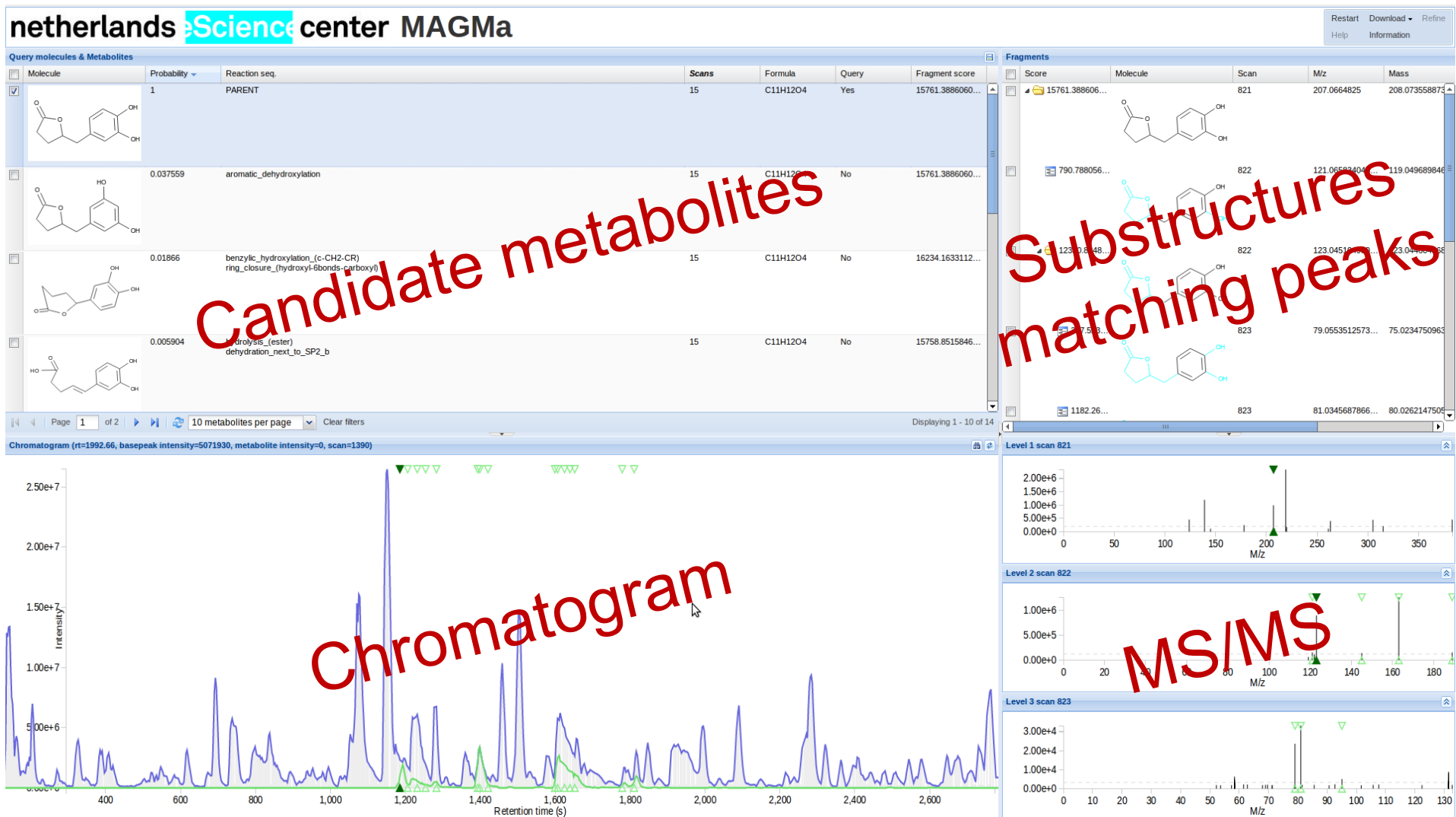


Example: urine of a tea-drinker

- Start with >50 known components of tea
- Apply human biotransformation rules
- > 19000 candidate metabolites
- Match metabolites with LC-MS peaks from urine



Easy to use web interface

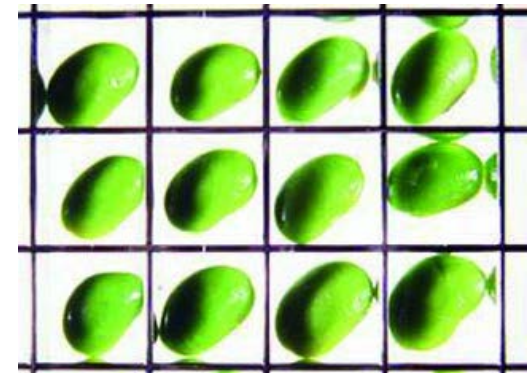


NLeSC Project Portfolio; Life Sciences

Green genetics: Virtual Lab for Plant Breeding

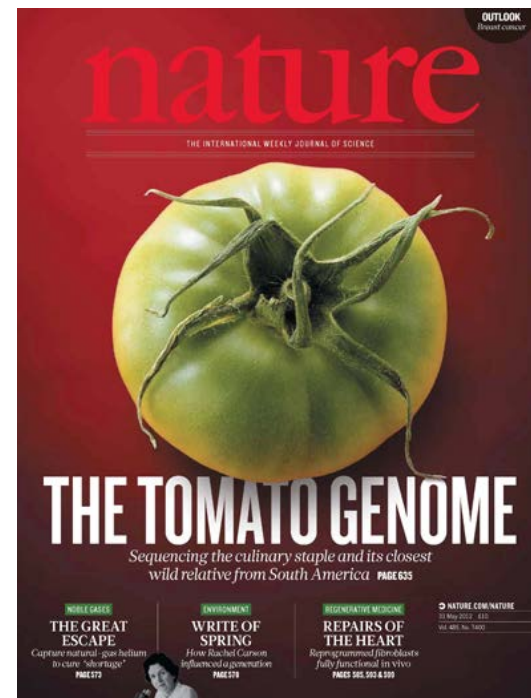
Project Leader: Bernard de Geus

Developing a virtual lab for plant breeding based on next-generation sequencing technology to support storage, integration and exploration of plant-genome data.



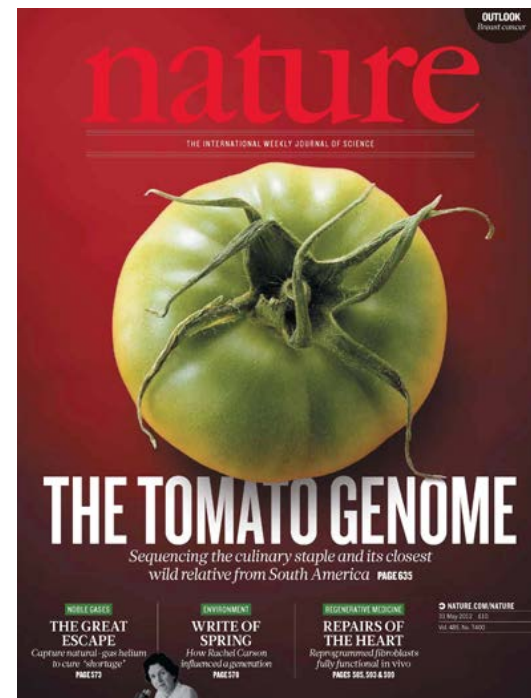
eScience & NGS are disruptive technologies for plant breeding sector

- Plant Breeding in the Netherlands is a healthy and innovative sector
- 4 plant breeding companies in top 25 of R&D investments
- Until 2000: plant breeding “trial & error”
- Insight into core genomes (tomato, rice, etc) may reverse the traditional breeding workflows
- First *in silico* mining for relevant genes -> data-driven crossing
- Opportunity to develop commercially interesting varieties faster



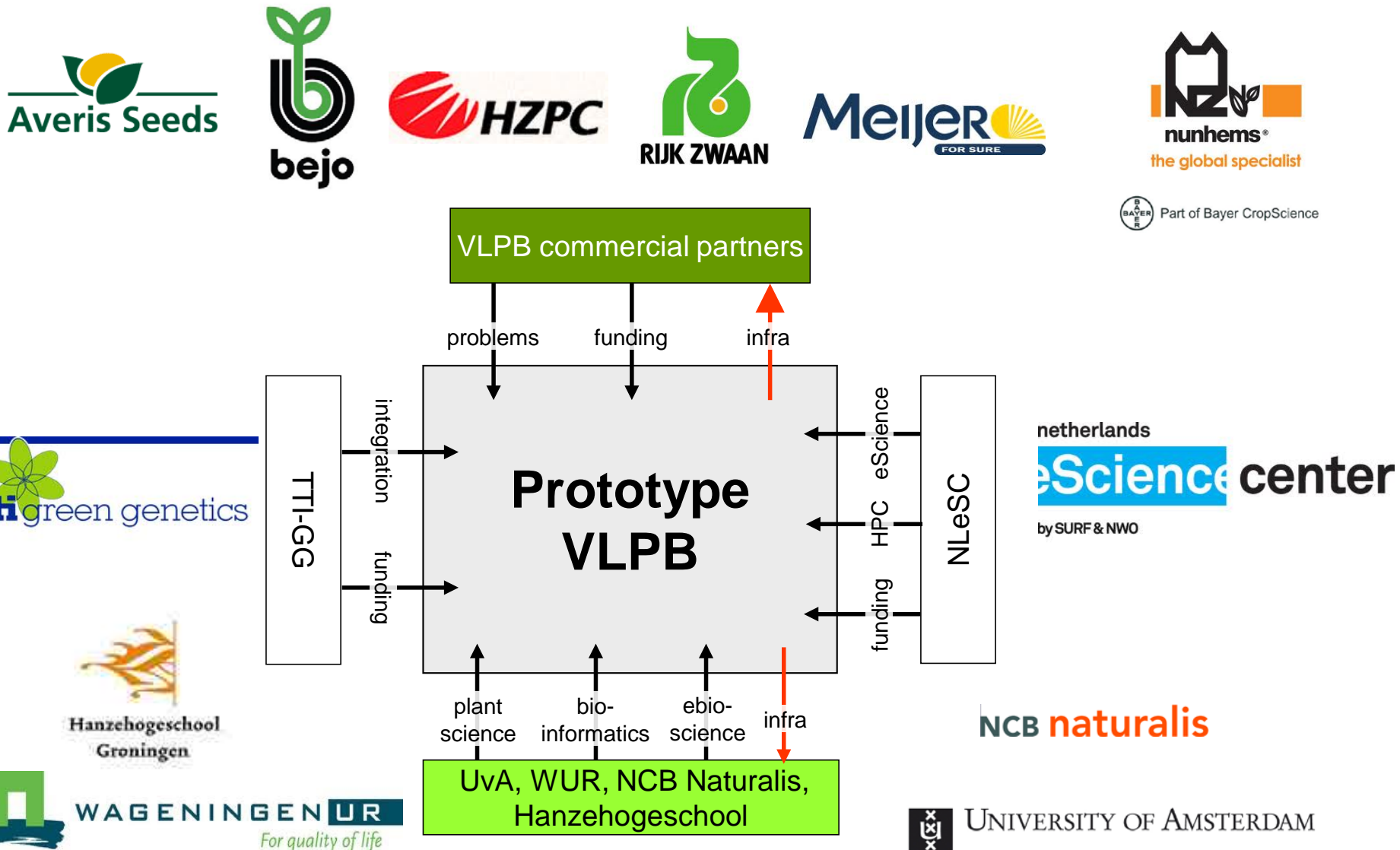
eScience & NGS are disruptive technologies for plant breeding sector

- Plant Breeding in the Netherlands is a healthy and innovative sector
- 4 plant breeding companies in top 25 of R&D investments
- Until 2000: plant breeding “trial & error”
- Insight into core genomes (tomato, rice, etc) may reverse the traditional breeding workflows
- First *in silico* mining for relevant genes -> data-driven crossing
- Opportunity to develop commercially interesting varieties faster

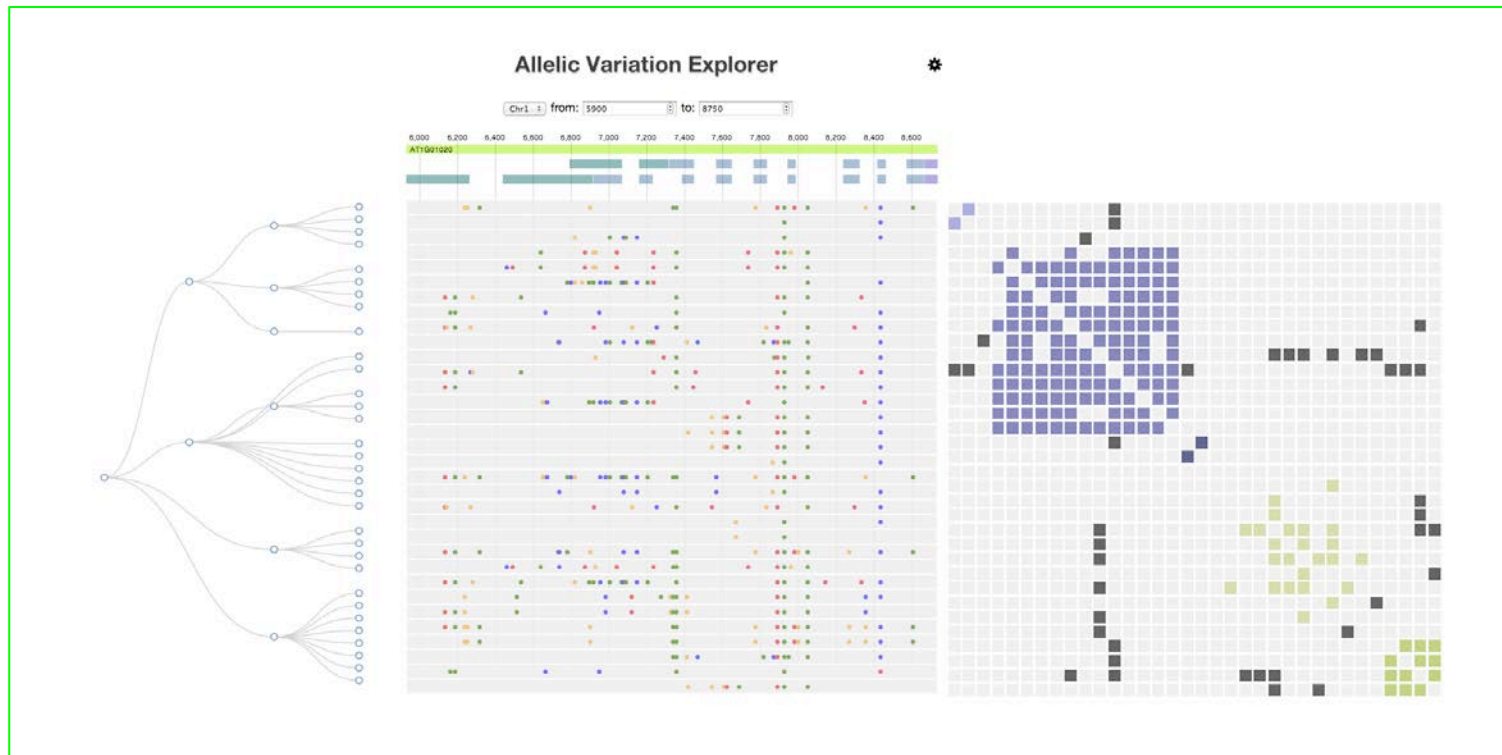


- **Big Data Challenge: Amount of sequence data coming towards the sector far too much for individual companies to cope with**
- **Acute need for effective eScience platform securing innovation power**

Unique collaboration between 6 breeding companies, 3 academic institutes, 1 HBO, TTI-GG and NLeSC to develop a *pre-competitive* eScience platform

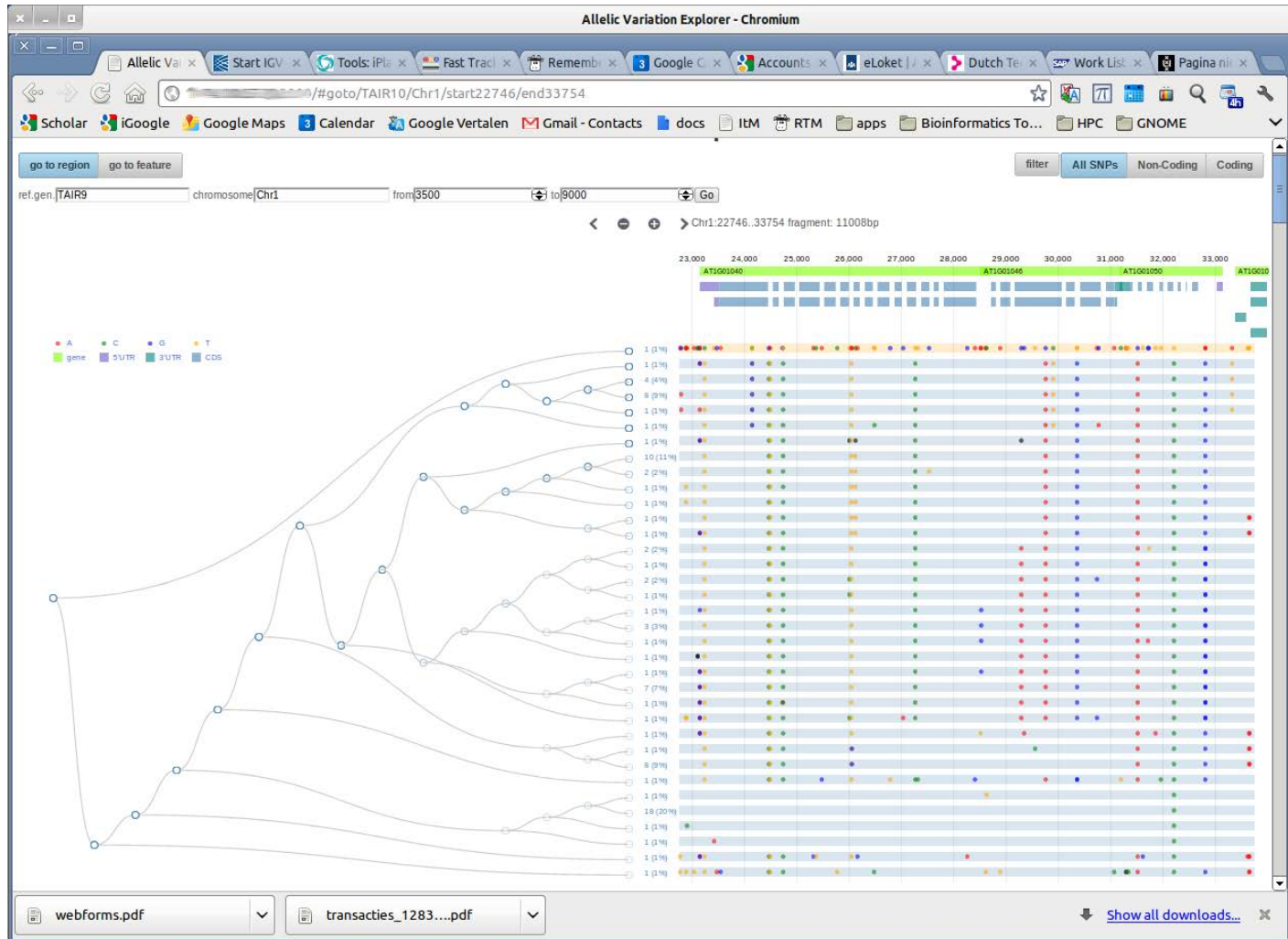


Easy to use interface: *visualization of SNPs as haplotype blocks*



optimal dialogue between scientists from public and private sectors needed to ensure eScience is applied in business process

Application in the cloud



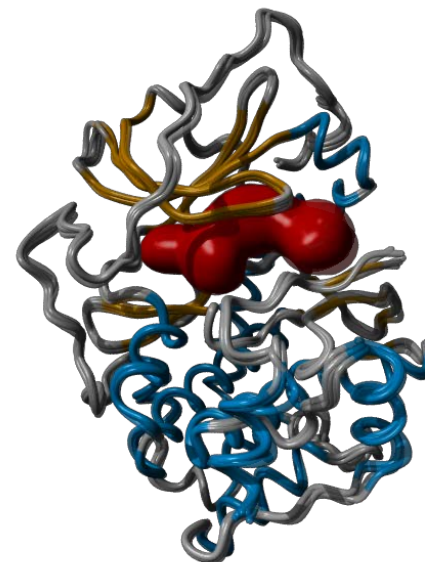
eScience technologies for multidisciplinary and remote collaboration

Unique pre-competitive public-private partnership

- Results available to all commercial and academic partners on a 'freedom to use' policy
- Only precompetitive or public Data, Information & Knowledge will be introduced in the public VLPB eScience platform

The Protein Flexibility Challenge

- Drug targets are flexible biomolecules
- Insight in protein receptor flexibility valuable for drug design & development

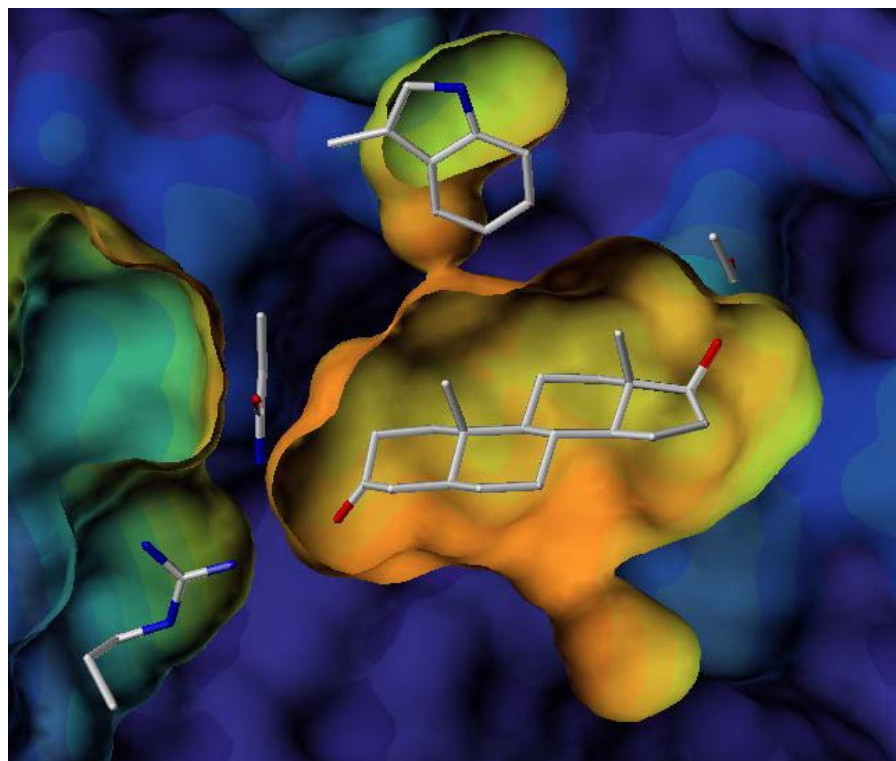


Research Collaboration Organon NV and IBM Research Zurich

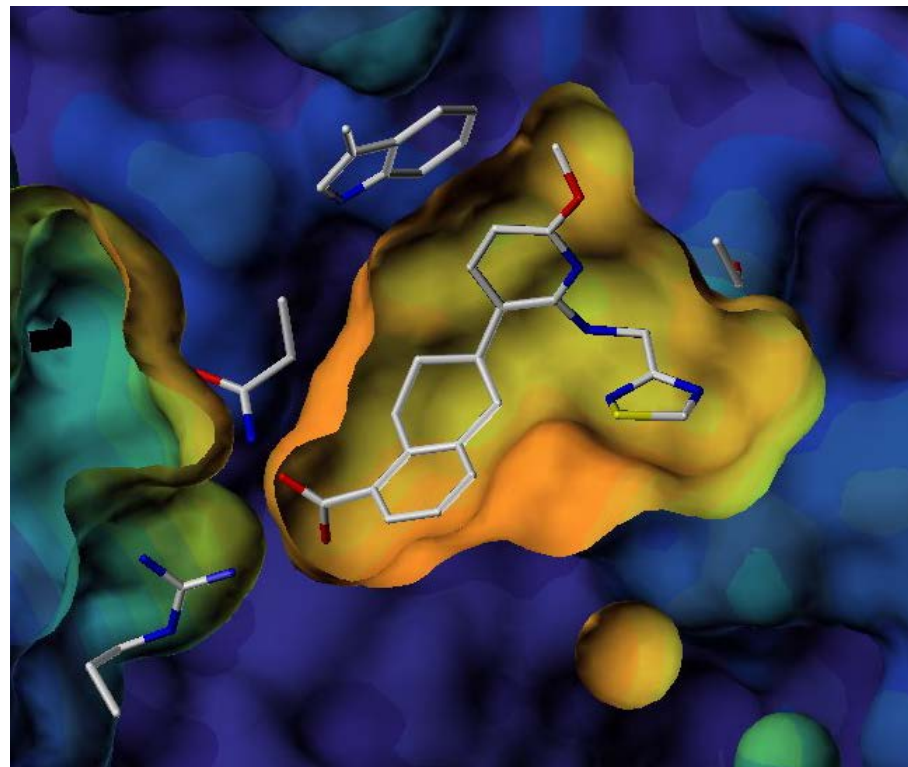
Collaboration Schering/MSD and Radboud University



Induced conformational change in binding pocket of Target Receptor



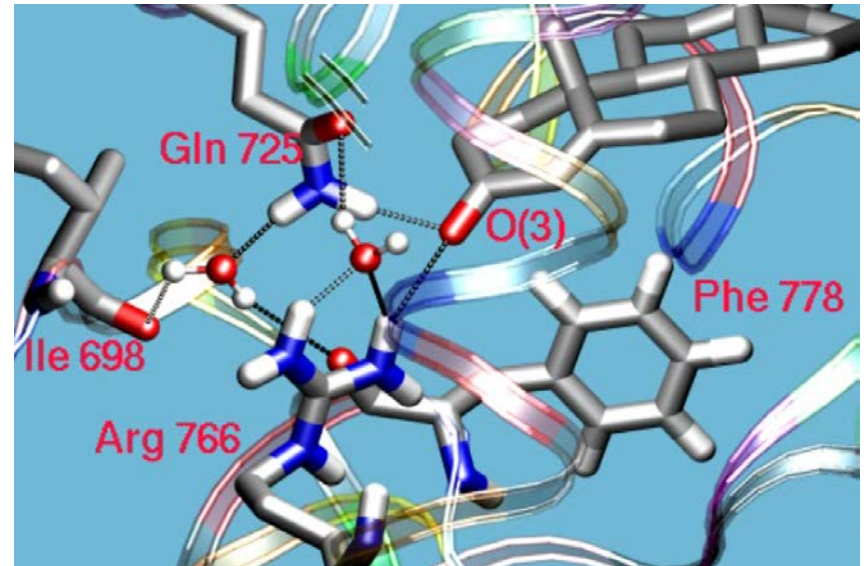
Dihydrotestosterone (DHT)



Compound X

Case study: IBM-Organon Molecular Dynamics computer simulations

- To study the highly specific progesterone-receptor interaction
- In collaboration with Prof Andreoni, IBM Research Zurich, Switzerland

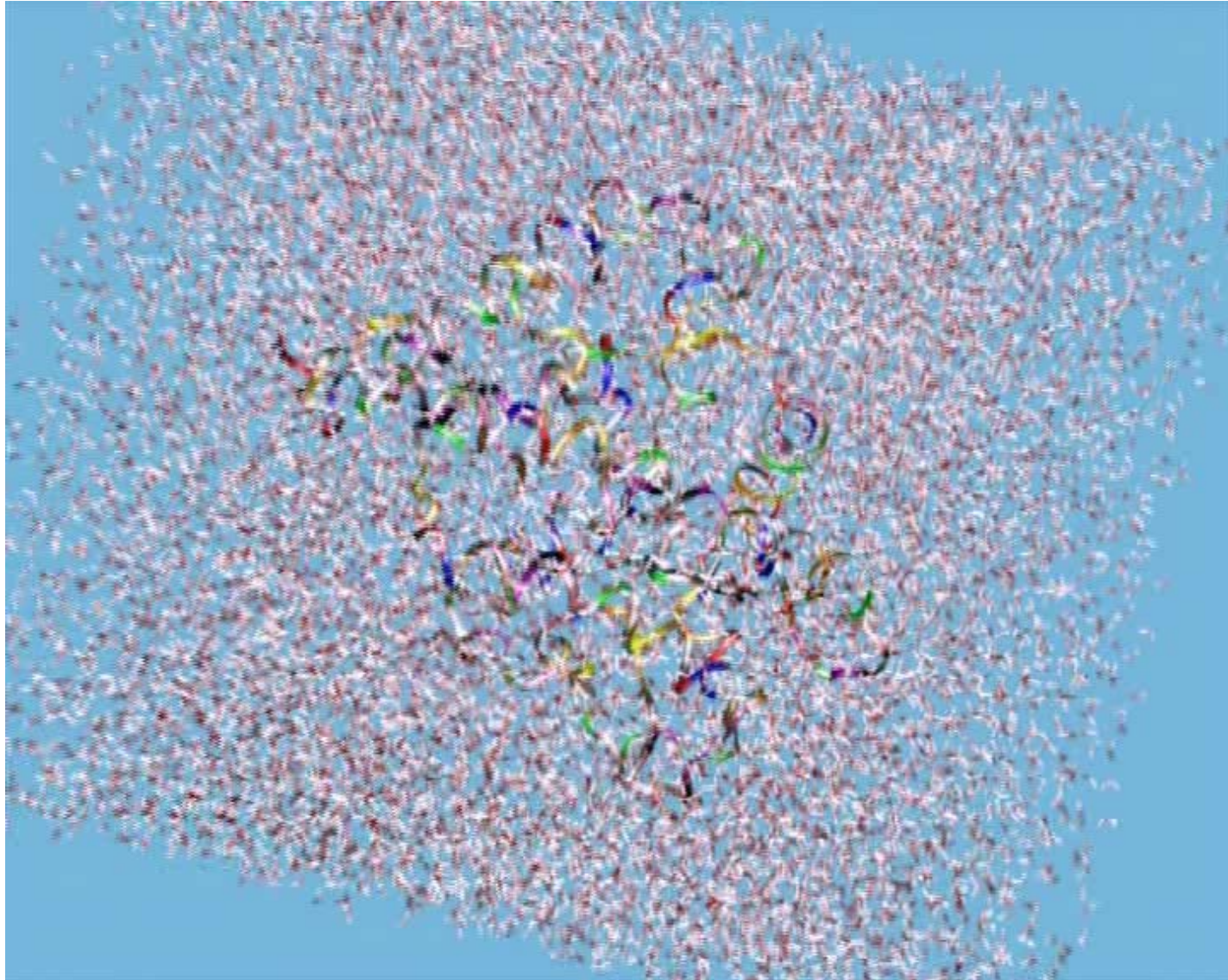


netherlands

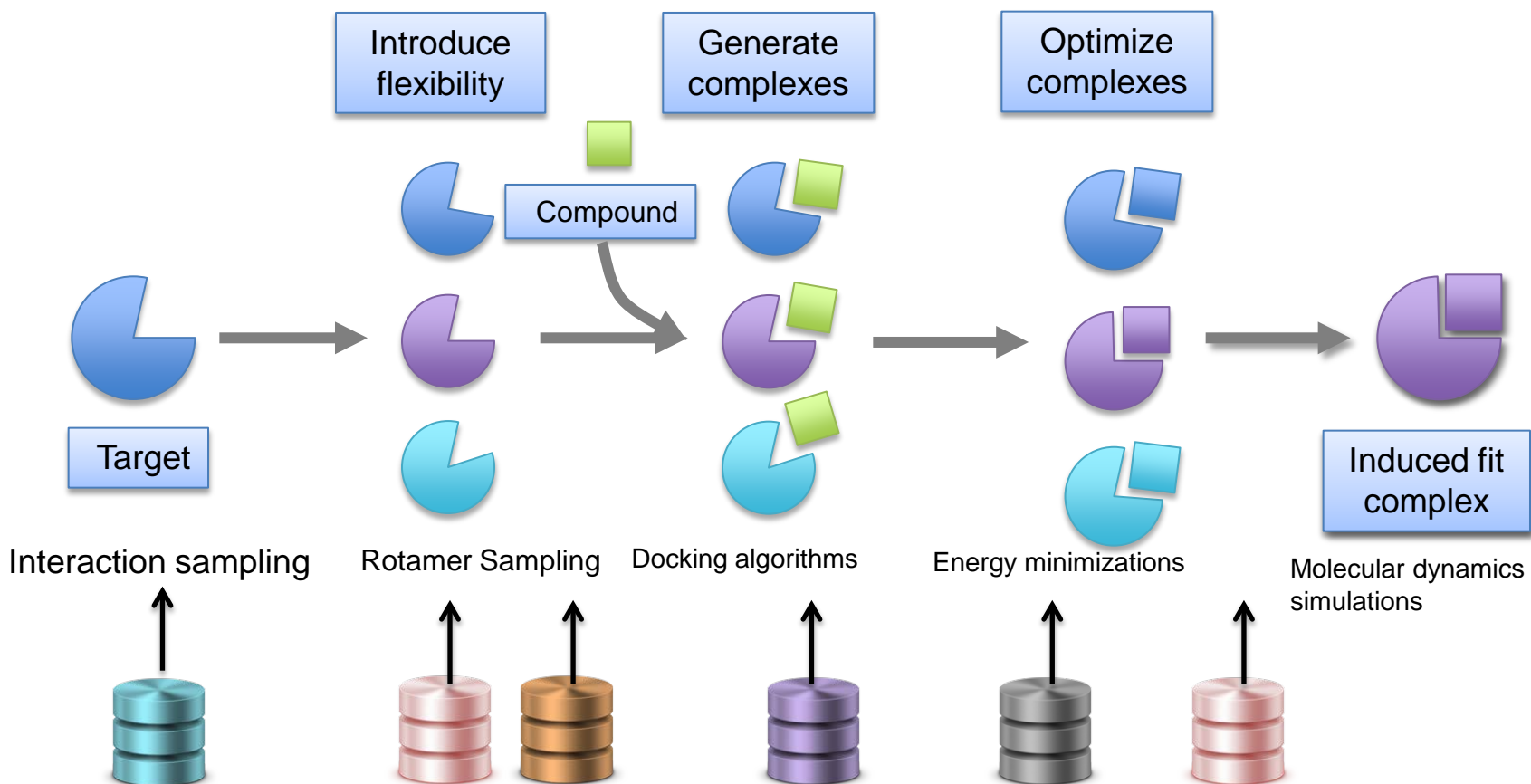
eScience center

by SURF & NWO

Water bridging interaction between receptor and ligand



Combining the best: *data-driven simulations and integrated workflow solutions*



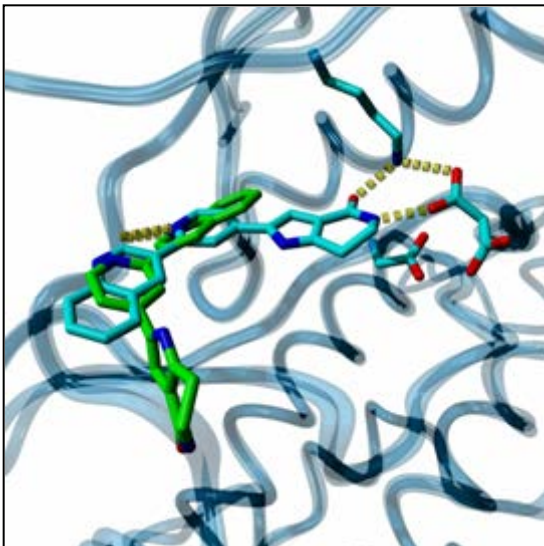
Radboud University Nijmegen / Medical Centre and Schering/MSD collaboration

Nabuurs SB, Wagener M, de Vlieg J, J Med Chem 2007 , 50:6507-6518.

NWO-Veni project

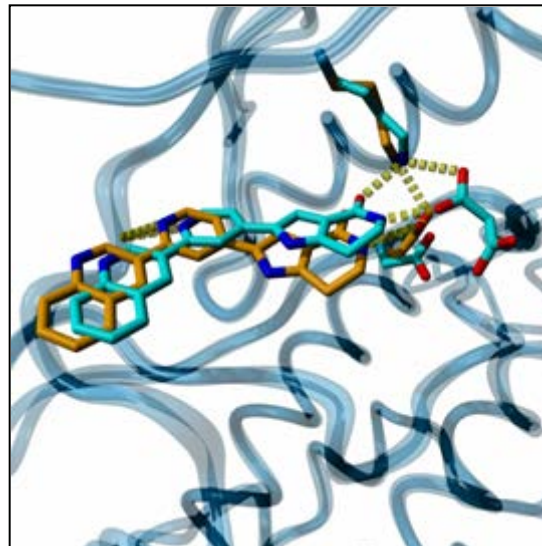
Docking compound x into flexible kinase protein target

standard docking



$\text{RMSD}_{\text{xray}} = 8.0 \text{ \AA}$

eScience protocol: combining the best



$\text{RMSD}_{\text{xray}} = 1.2 \text{ \AA}$

Binding mode confirmed by in-house protein Xray

eScience Hero

Fights for medical innovation

- **Pattern recognition**
- **Machine learning**
- **Big Data**
- **Social Media**



Andy Grove (ex-CEO Intel)

netherlands

eScience center

by SURF & NWO

Voice algorithms spot Parkinson's disease

- Machine learning algorithms that analyse voice recordings to detect Parkinson's symptoms early on ([Little et al. @ Media Lab, MIT](#))

Voice tremors



The technology works partly by tracking the motion of vocal cords

- Social Media:**



Looking for volunteers to contribute to the database to improve pattern recognition

Data-driven parkinson diagnostics project sponsored by Andy Grove

Success factors for precompetitive public-private partnerships

- Long-term & challenging scientific problem
- Validation of new scientific approach by industry partner
- Intensive dialogue between academic and industry scientists:
 - To create shared responsibility
 - to ask the correct scientific questions
- Sustainable solutions & community building

It is all about people: trust & respect

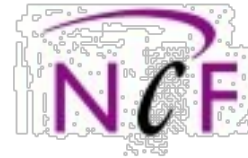
Thank you & Acknowledgements



Radboud Universiteit Nijmegen



TI PHARMA



Netherlands Genomics Initiative

