

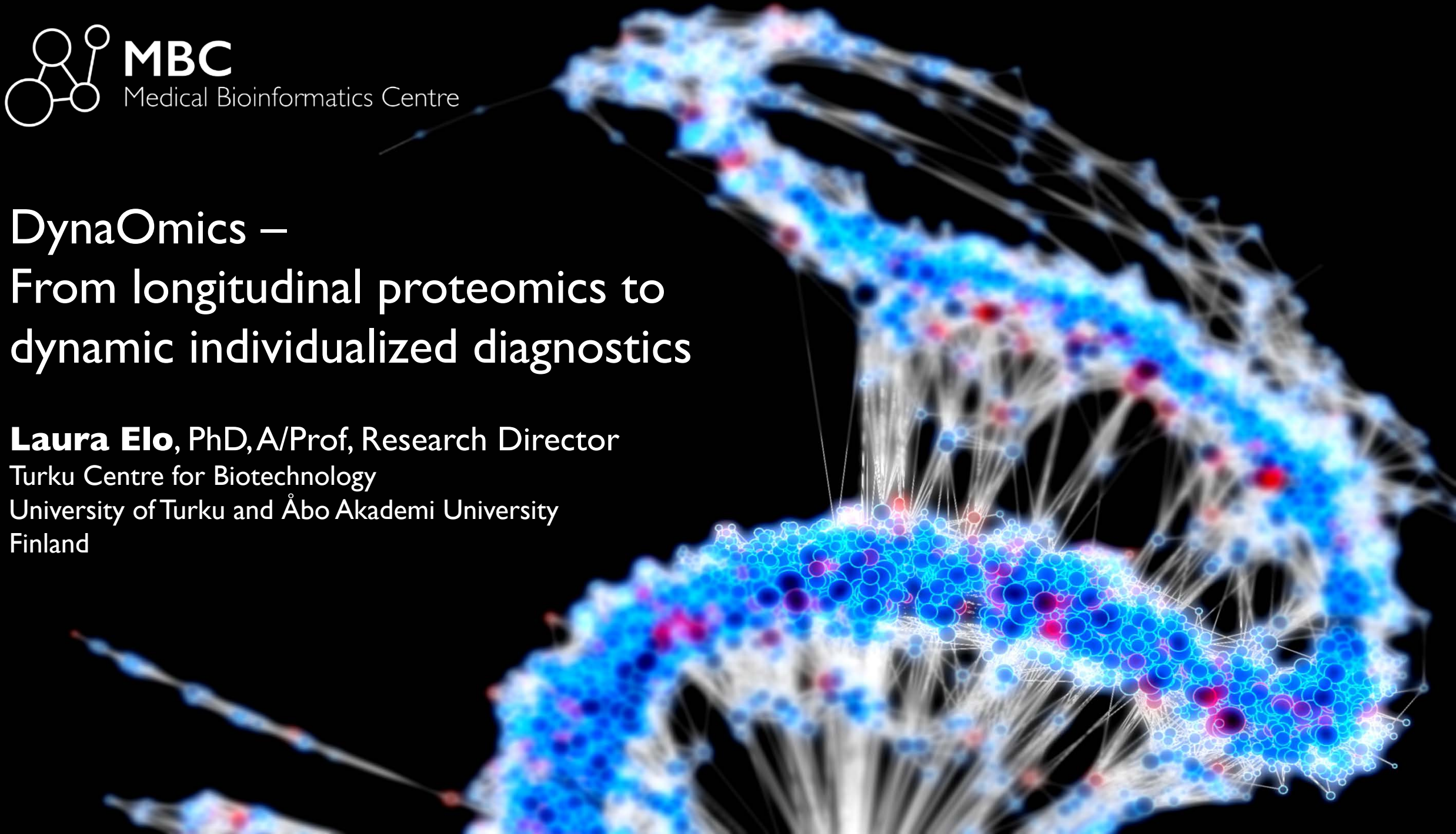


MBC

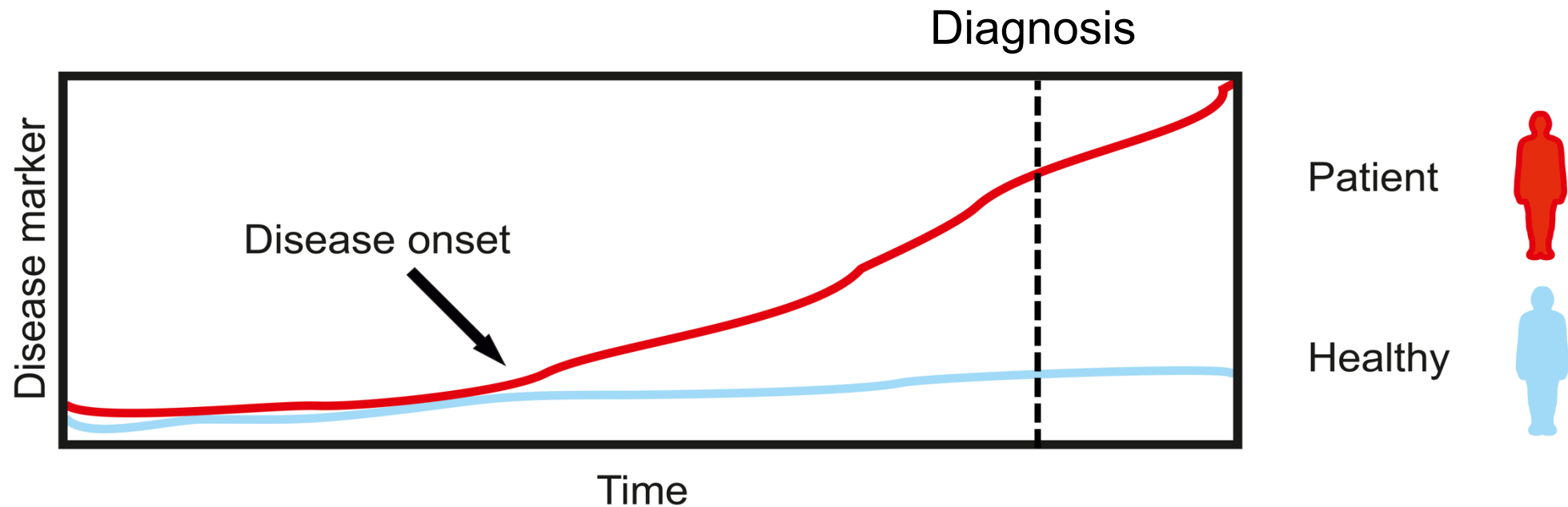
Medical Bioinformatics Centre

DynaOmics – From longitudinal proteomics to dynamic individualized diagnostics

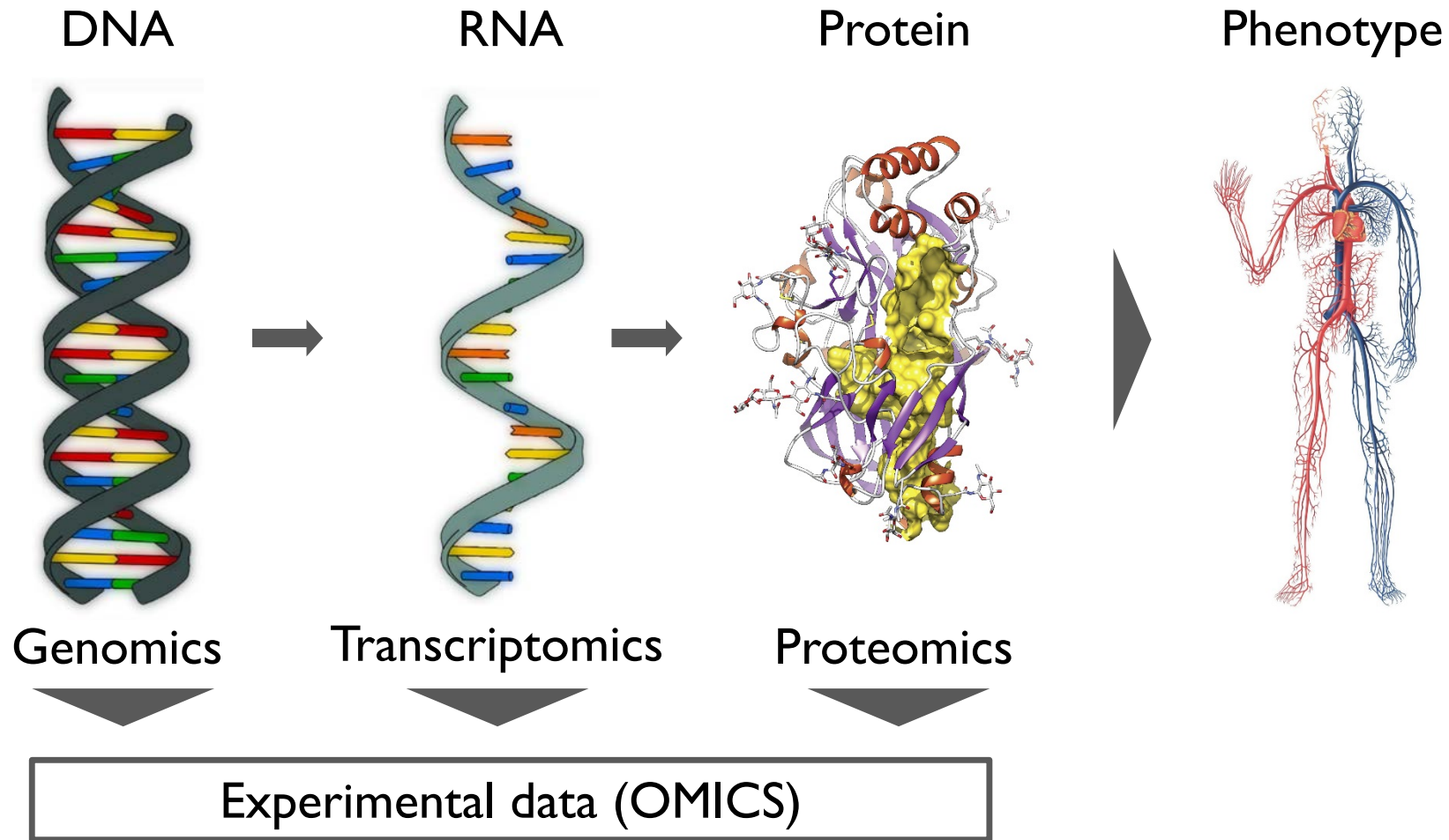
Laura Elo, PhD, A/Prof, Research Director
Turku Centre for Biotechnology
University of Turku and Åbo Akademi University
Finland



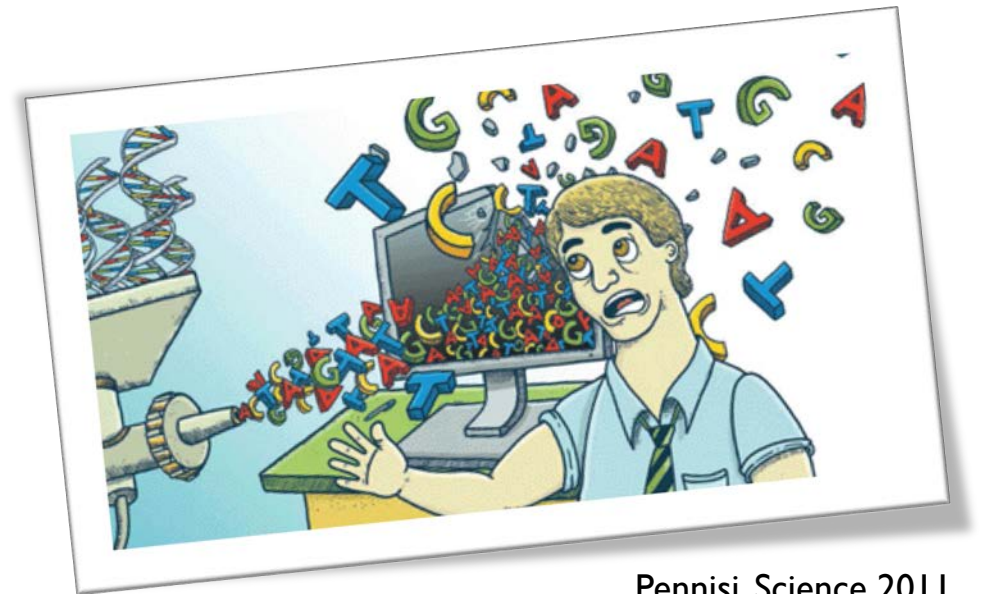
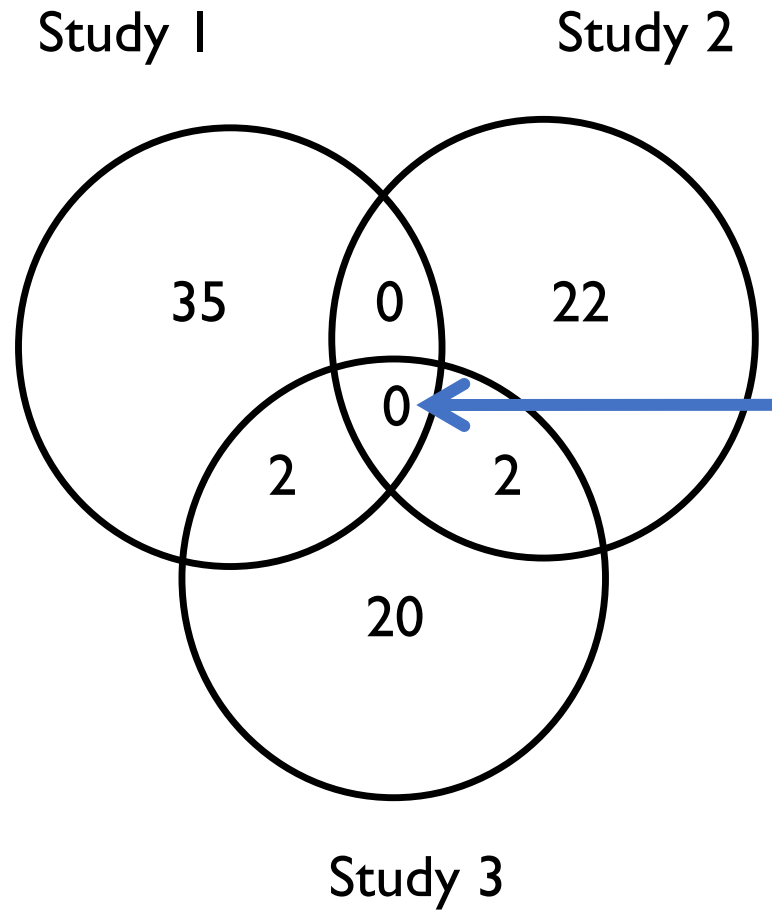
DynaOmics – From longitudinal proteomics to dynamic individualized diagnostics



Modern high-throughput biotechnology

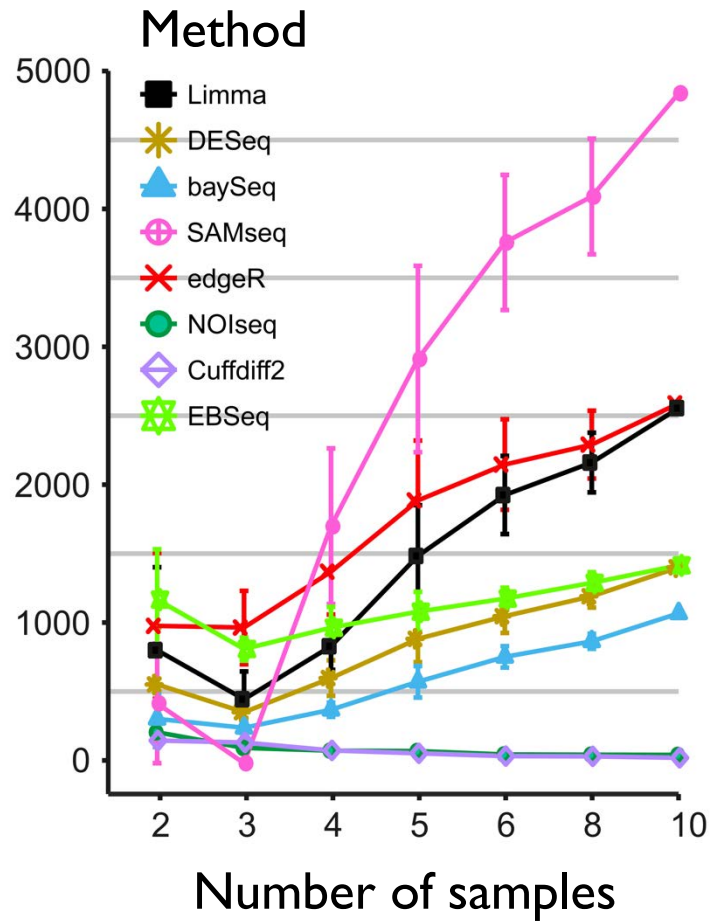


Current challenge: Failures to detect reproducible markers



Current challenge: Failures to detect reproducible markers

Number of detections



Nearly 5000 detections



No detections

More reliable markers: reproducibility optimization

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE
ROTS: An R package for reproducibility-optimized statistical testing

Tomi Suomi^{1,2*}, Fatemeh Seyednasrollah^{1,2}, Maria K. Jaakkola^{1,3}, Thomas Faux¹, Laura L. Elo^{1*}

¹ Turku Centre for Biotechnology, University of Turku and Abo Akademi University, Turku, Finland, ² Department of Future Technologies, University of Turku, Turku, Finland, ³ Department of Mathematics and Statistics, University of Turku, Turku, Finland

* tomi.suomi@utu.fi (TS); laura.elo@utu.fi (LLE)

Abstract

Differential expression analysis is one of the most common types of analyses performed on various biological data (e.g. RNA-seq or mass spectrometry proteomics). It is the process that detects features, such as genes or proteins, showing statistically significant differences between the sample groups under comparison. A major challenge in the analysis is the choice of an appropriate test statistic, as different statistics have been shown to perform well in different datasets. To this end, the reproducibility-optimized test statistic (ROTS) adjusts a modified *t*-statistic according to the inherent properties of the data and provides a ranking of the features based on their statistical evidence for differential expression between two groups. ROTS has already been successfully applied in a range of different studies from transcriptomics to proteomics, showing competitive performance against other state-of-the-art methods. To promote its widespread use, we introduce here a Bioconductor R package for performing ROTS analysis conveniently on different types of omics data. To illustrate the benefits of ROTS in various applications, we present three case studies, involving proteomics and RNA-seq data from public repositories, including both bulk and single cell data. The package is freely available from Bioconductor (<https://www.bioconductor.org/packages/ROTS>).

OPEN ACCESS

Citation: Suomi T, Seyednasrollah F, Jaakkola MK, Faux T, Elo LL (2017) ROTS: An R package for reproducibility-optimized statistical testing. *PLoS Comput Biol* 13(5): e1005562. <https://doi.org/10.1371/journal.pcbi.1005562>

Editor: Timothée Poisot, Université de Montréal, CANADA

Received: December 7, 2016
Accepted: May 10, 2017
Published: May 25, 2017

Copyright: © 2017 Suomi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The ROTS R-package is available from <https://www.bioconductor.org/packages/ROTS>, the CPTAC technology assessment data from <https://cptac-data-portal.genepattern.edu/cptac/study/001700gen-Phase-L>, the data of the SEQC project from <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47774> and the single-cell data from <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59114>.

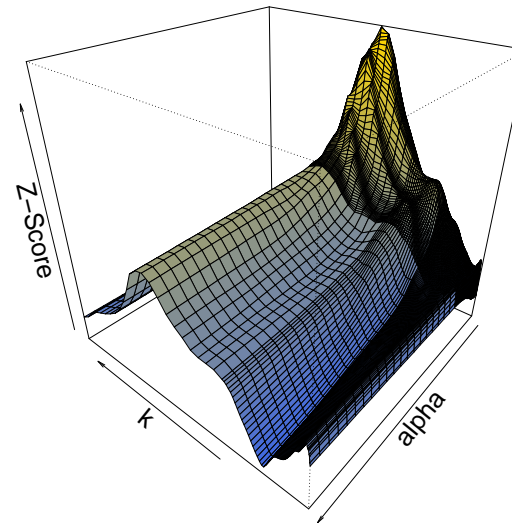
Funding: LLE reports grants from the European Research Council (ERC) (677943), European Union's Horizon 2020 research and innovation

"This is a PLOS Computational Biology Software paper."

Introduction

Differential expression analysis between two groups of samples is perhaps the most common type of analysis that is performed on various types of omics data. The aim of differential expression analysis is to detect features (e.g. genes or proteins) showing statistically significant changes between the groups. A commonly used approach has been the Student's *t*-test, which

PLOS Computational Biology | <https://doi.org/10.1371/journal.pcbi.1005562> May 25, 2017 1 / 10



$$d_{\alpha} = \frac{|\bar{x}_1 - \bar{x}_2|}{\alpha_1 + \alpha_2 s}$$

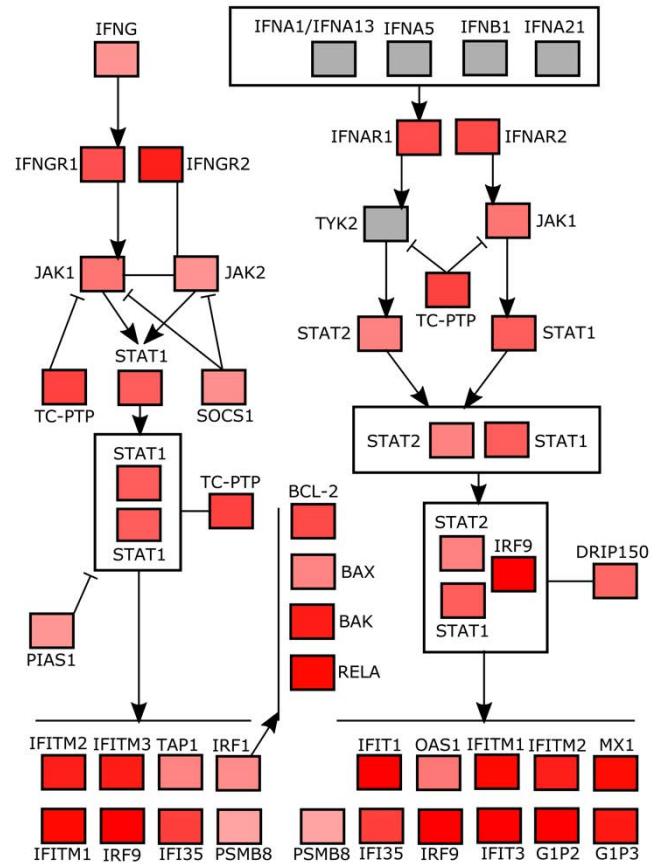
$$Z_k(d_{\alpha}) = \frac{R_k(d_{\alpha}) - R_k^0(d_{\alpha})}{s_k(d_{\alpha})}$$



"For unfiltered data sets, many methods struggled to correctly control the type I error, and the best performance was obtained by ROTS..."

Soneson & Robinson (**Nature Methods**, 2018)

More reliable markers: pathways and networks



Briefings in Bioinformatics, 17(2), 2016, 336–345

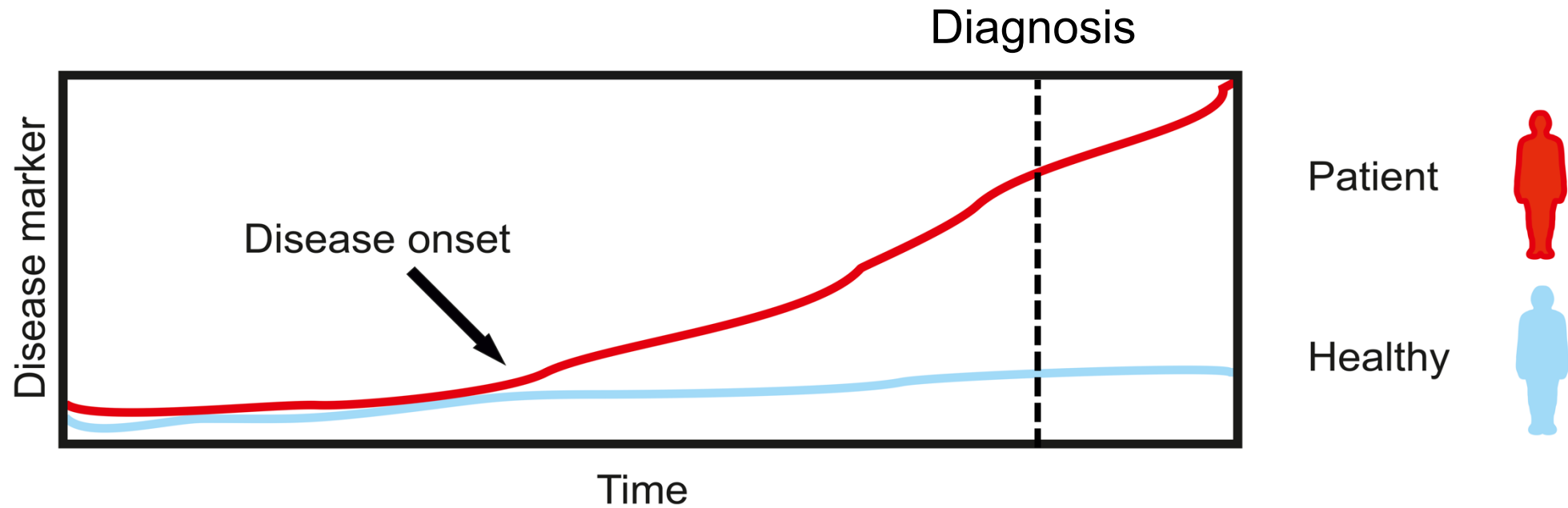
doi: 10.1093/bib/bbv049
 Advance Access Publication Date: 21 July 2015
 Paper

Empirical comparison of structure-based pathway methods

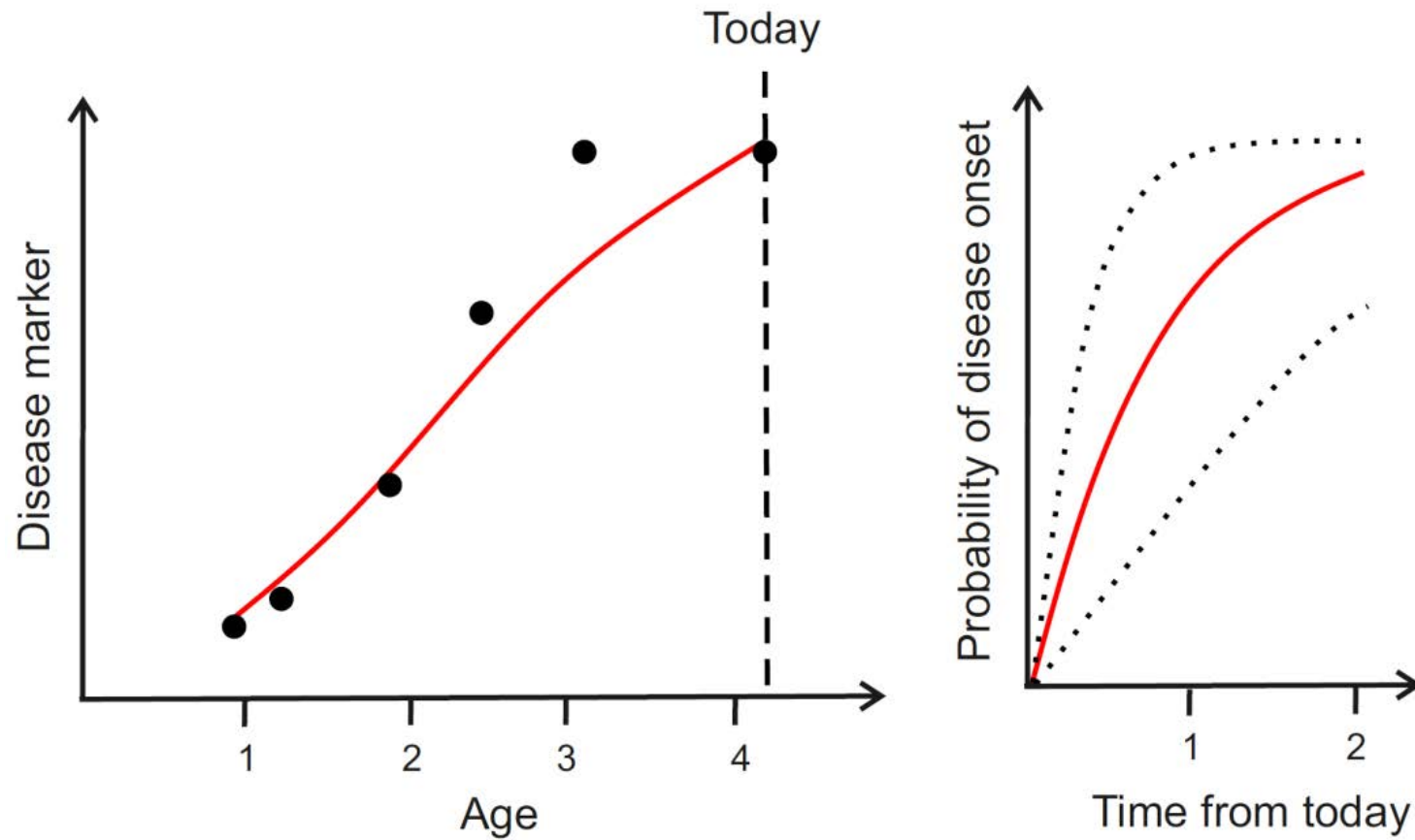
Maria K. Jaakkola and Laura L. Elo

Corresponding author: Laura L. Elo, Turku Centre for Biotechnology, and Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland. Tel.: +358-2-333-8009; Fax: +358-2-251 8808; E-mail: laliele@utu.fi

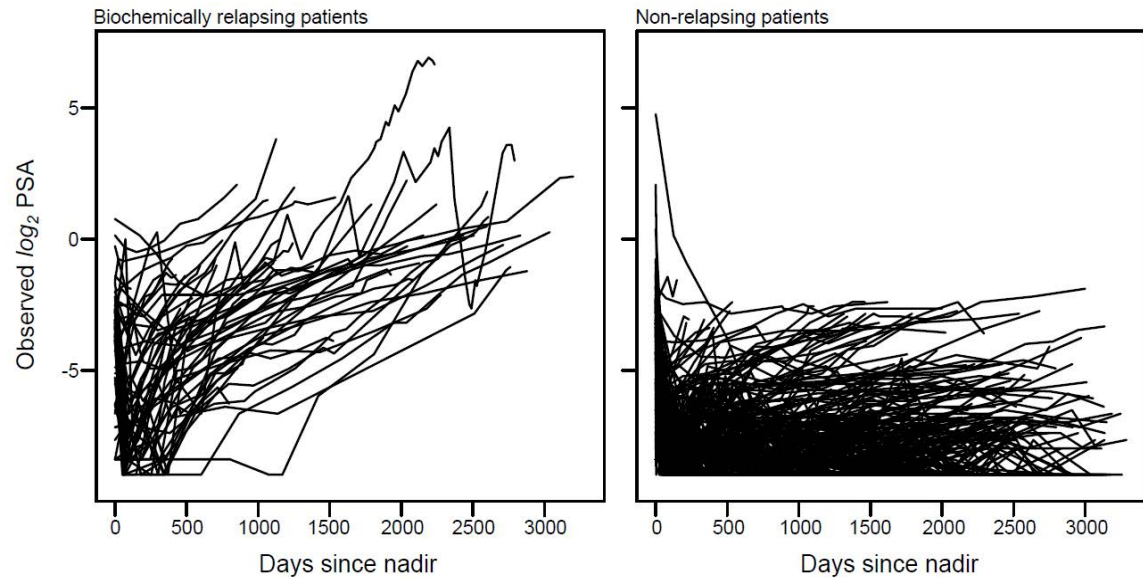
More reliable markers: longitudinal data



Dynamic disease risk prediction: new modelling strategies



Dynamic disease risk prediction



www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN Longitudinal modeling of ultrasensitive and traditional prostate-specific antigen and prediction of biochemical recurrence after radical prostatectomy

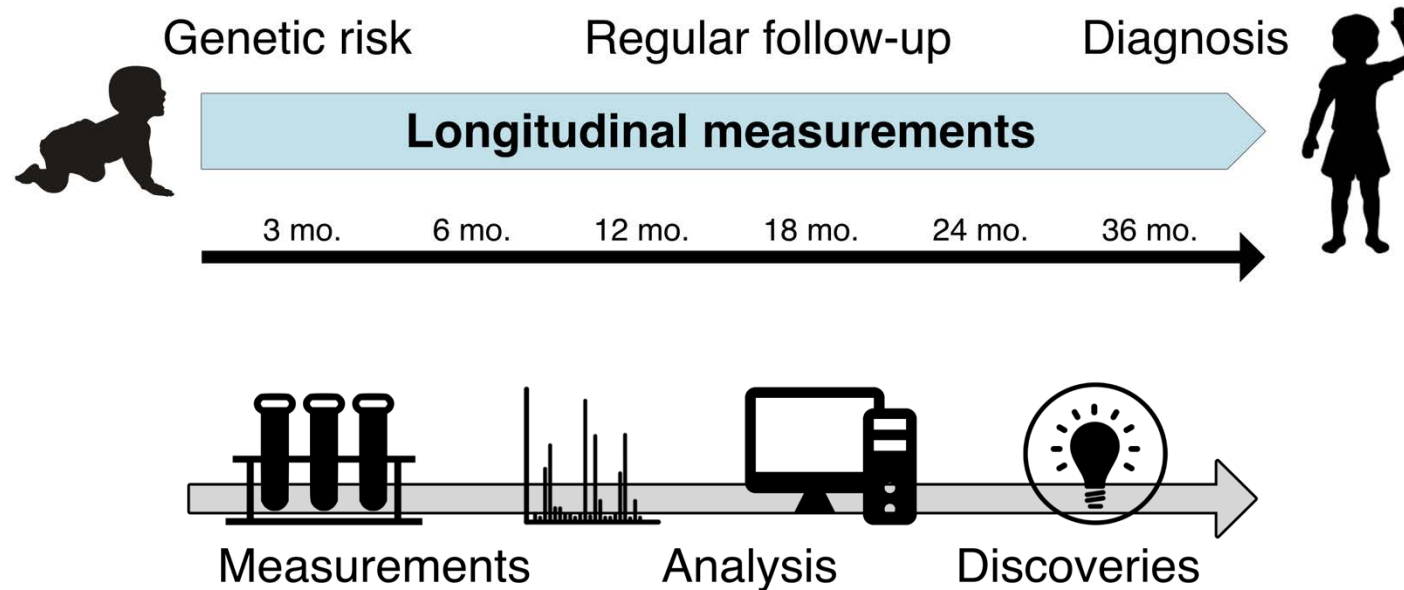
Received: 26 May 2016
Accepted: 03 October 2016
Published: 02 November 2016

Teemu D. Laajala^{1,2,*}, Heikki Seikkula^{3,4,*}, Fatemeh Seyednasrollah sadat^{1,2}, Tuomas Mirtti^{5,6}, Peter J. Boström^{*} & Laura L. Elo¹

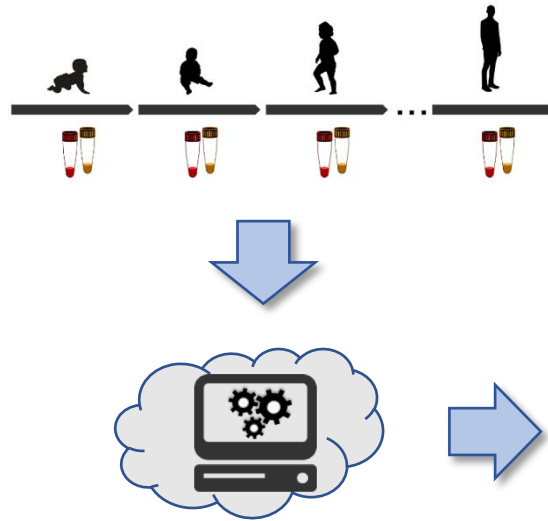
Early prediction of type I diabetes

DIPP – Type I Diabetes Prediction and Prevention project

A unique resource of prospective samples collected since 1994
>200,000 infants screened

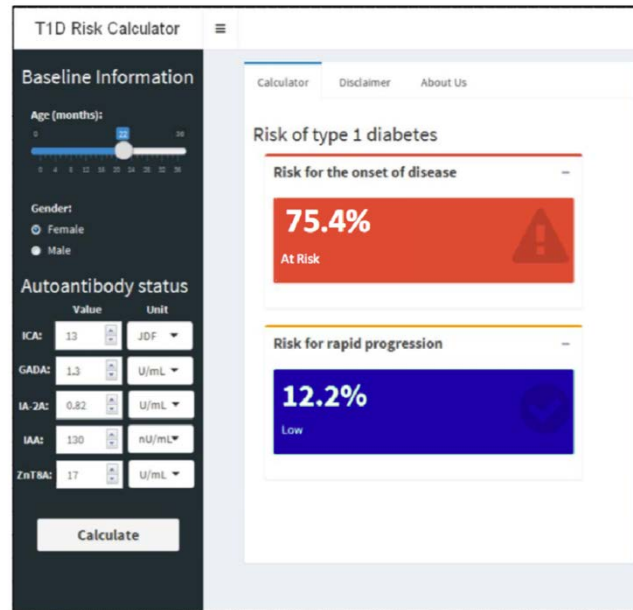


Practical utilization



Data mining and machine learning

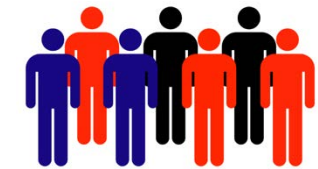
Risk calculator



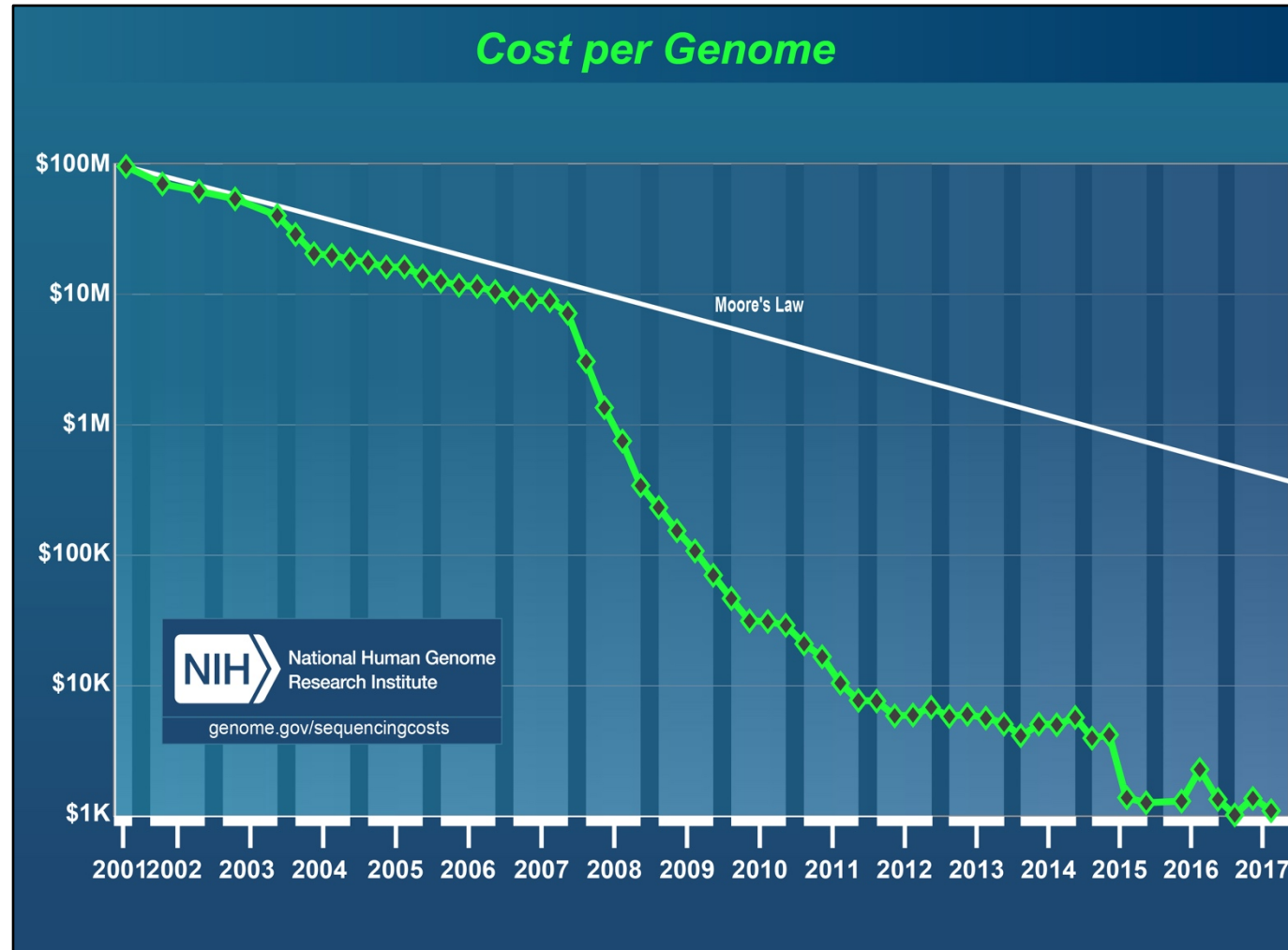
Personalized treatment



Improved clinical trials

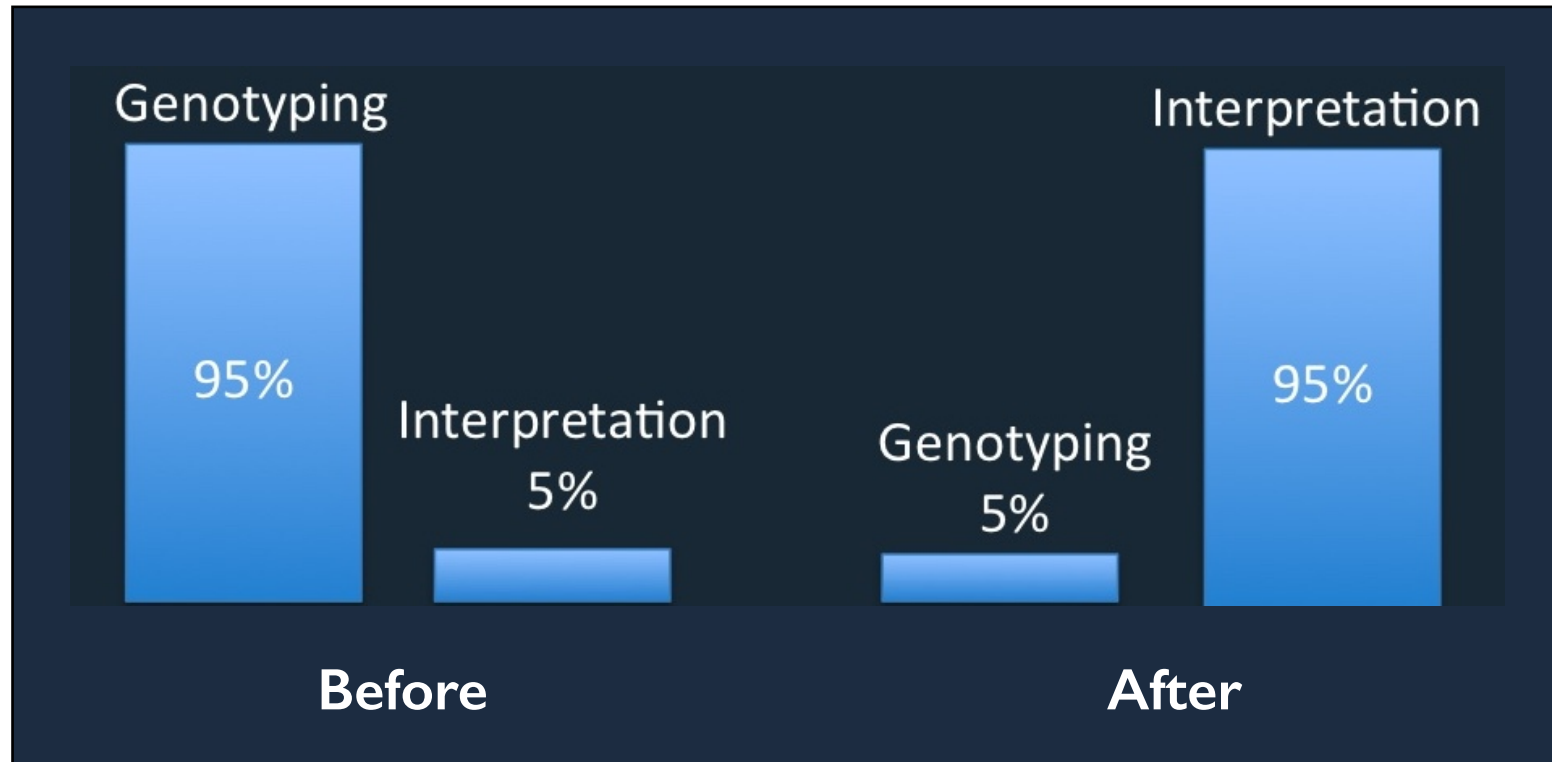


Data explosion



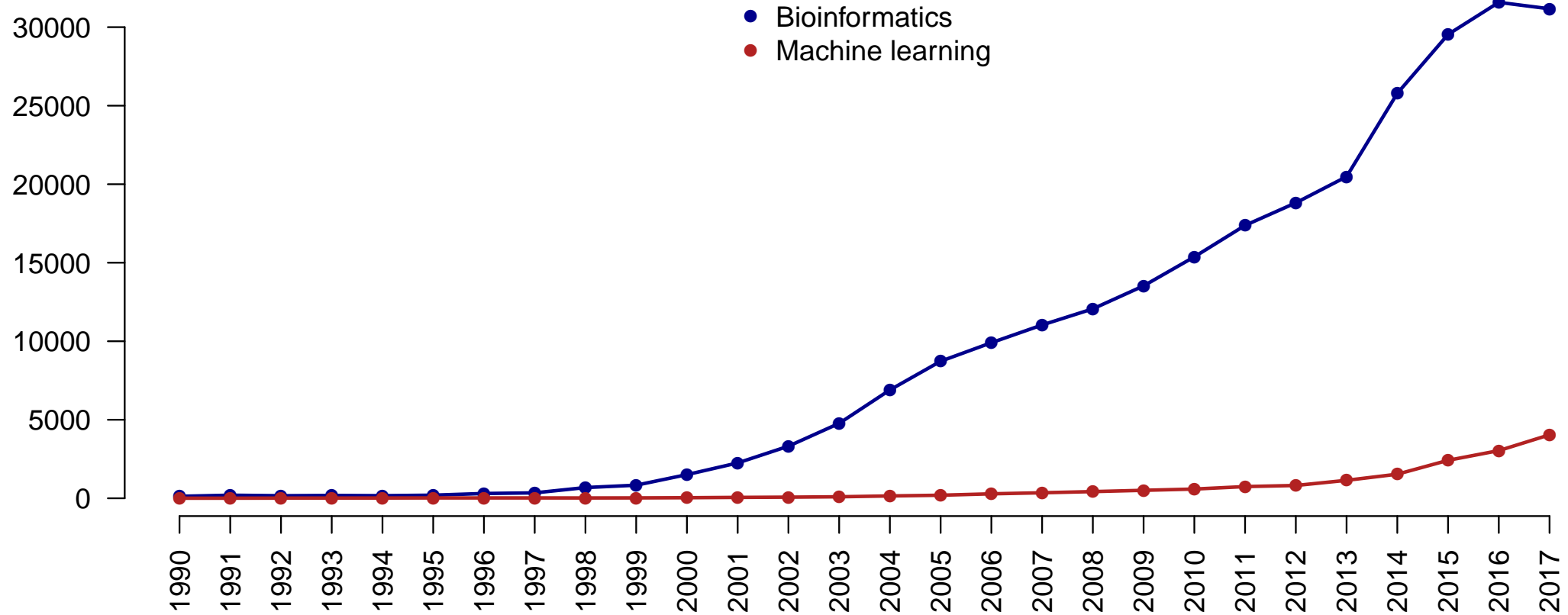
National Human Genome Research Institute, <https://www.genome.gov/>

Data explosion

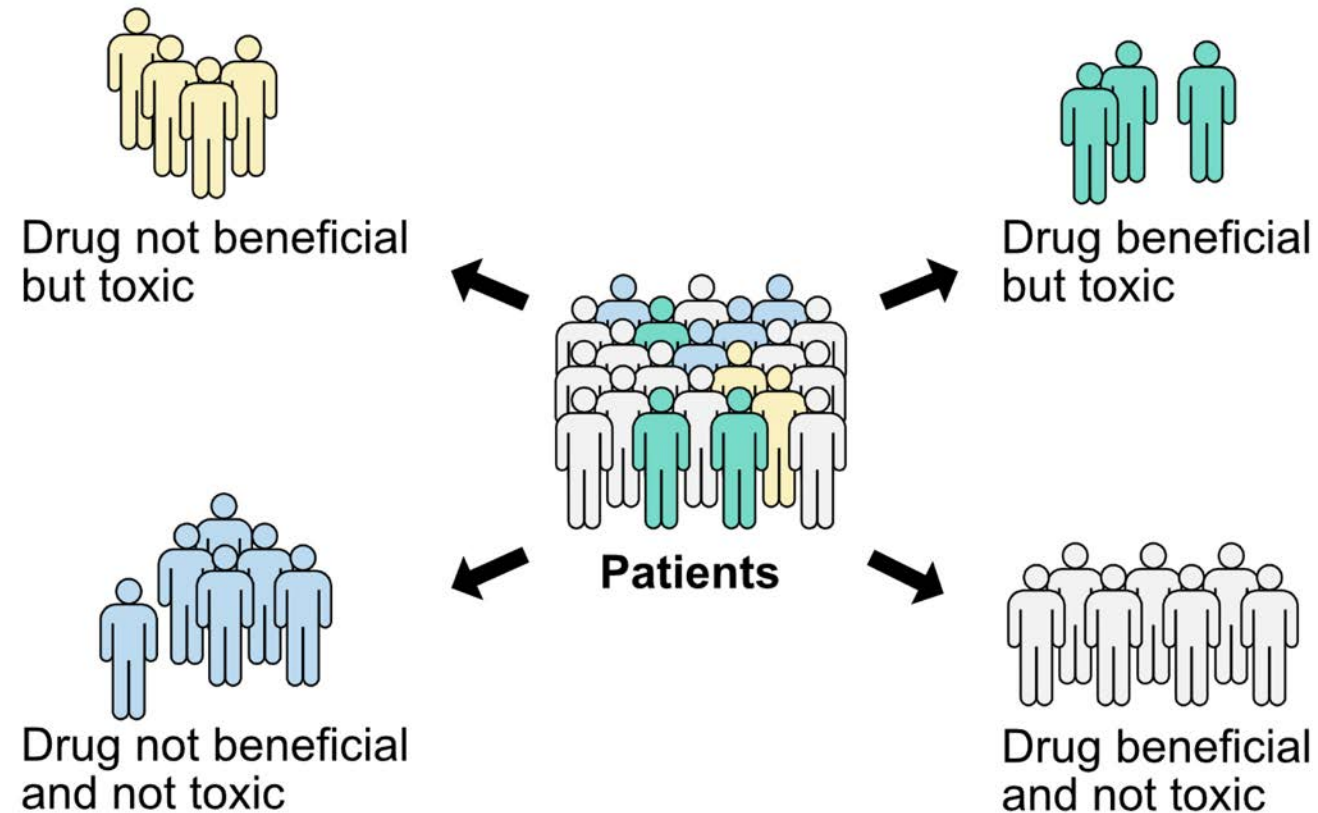


Bioinformatics and machine learning

Number of publications in Pubmed

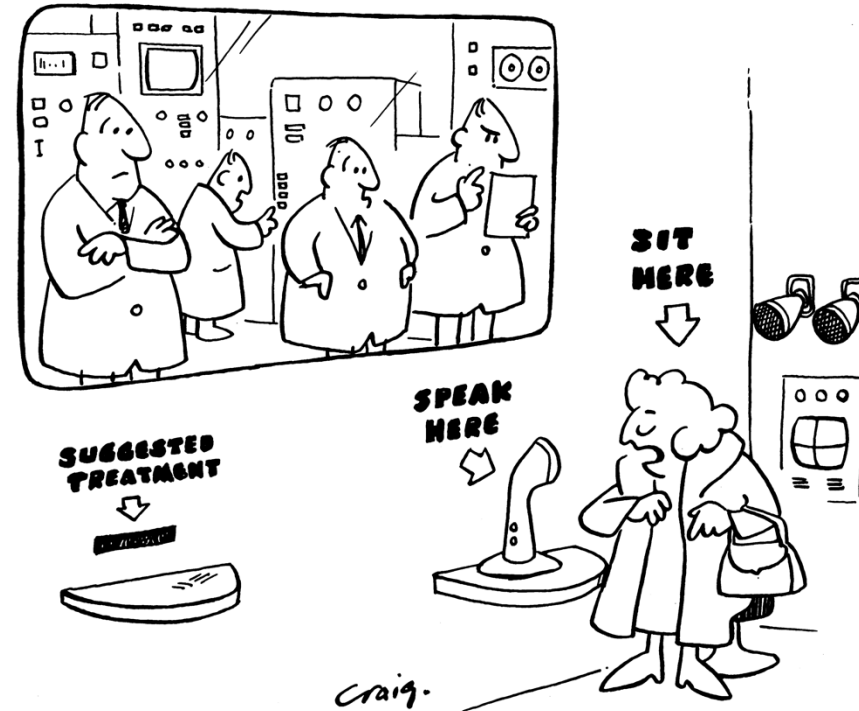


Transforming data to knowledge: precision medicine



Current challenges in translation to clinical practice

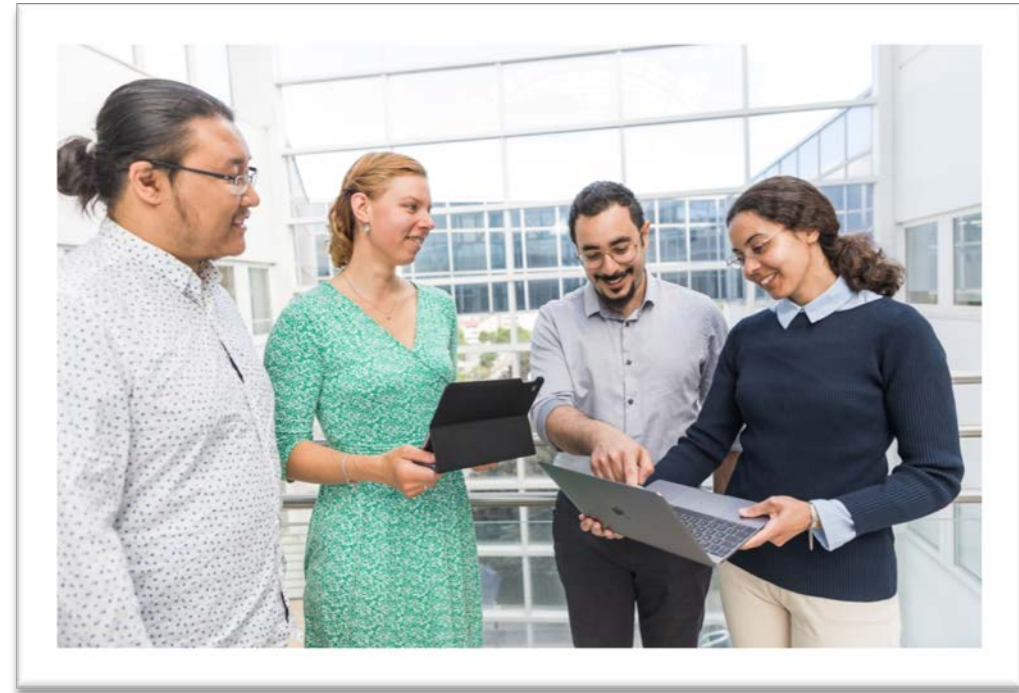
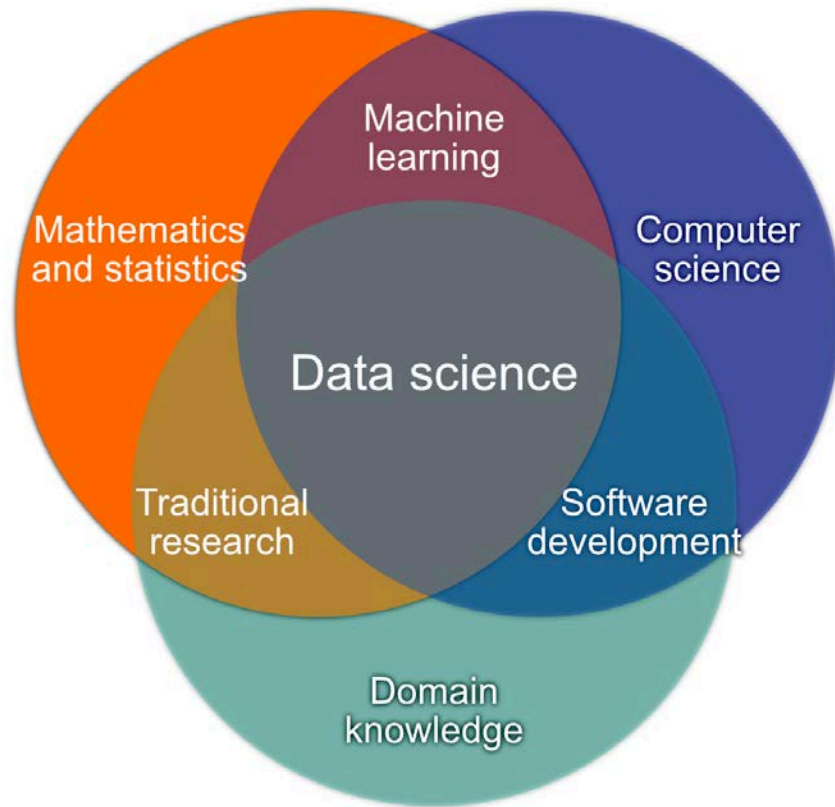
- Prediction performance
- Interpretation
- Rigorous validation
- Ethical and legal issues



”Within a few years the average doctor will likely have available to him a computer programmed for medicine for providing him with a great store of knowledge...”

Gibson (Canadian Family Physician, 1971)

Multidisciplinary research



Multidisciplinary research: mathematician in medicine

PhD

Applied Mathematics

$$P(T_i > t | \mathcal{M}_i(t), w_i) = e^{-\int_0^t h_0(s) e^{y^T w_i + a m_i(s)} ds}$$

$$h_0(t) = \sum_{q=1}^Q \xi_q I(v_{q-1} < t \leq v_q)$$

$$m_i(t) = x_i^T(t) \beta + z_i^T(t) b_i$$

Postdoc

Molecular Immunology,
Systems Biology

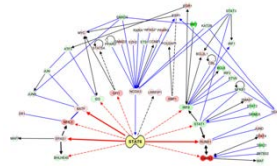


Immunity
Resource

Genome-wide Profiling of Interleukin-4 and STAT6 Transcription Factor Regulation of Human Th2 Cell Programming

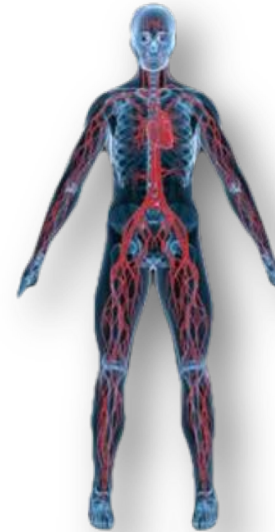
Laura L. Ebo,^{1,2,12} Hanna Järvenpää,^{1,2,12} Salla Tuomola,^{2,3,12} Sunil Raghav,^{2,12} Helena Anttonen,^{2,4} Kirsi Laurila,⁵ Bhavna Gupta,⁷ Rikka J. Lams,^{2,4} Johanna Tahvanainen,^{2,7} R. David Hawkins,² Matej Oresic,² Harri Lähdesmäki,^{1,12} Omid Rasool,⁷ Kanury V. Rao,^{11,14} Tero Aittokallio,^{2,12} and Riitta Lahesmaa^{2,12}

¹Biomediatrics Research Group, Department of Mathematics, University of Turku, FI-20014 Turku, Finland
²Turku Centre for Biotechnology, University of Turku and Åbo Akademi, P.O. Box 123, FI-20521 Turku, Finland
³Turku Graduate School of Biomedical Sciences, Kivimyllykatu 13, FI-20020 Turku, Finland
⁴The National Graduate School in Informational and Structural Biology, Åbo Akademi University, FI-20520 Turku, Finland
⁵Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
⁶Department of Biological Science, University of Sheffield, S10 2TN, Sheffield, UK
⁷Drug Discovery Graduate School, University of Turku, FI-20014 Turku, Finland
⁸Ludwig Institute for Cancer Research, University of California, San Diego, CA 92037, USA
⁹VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 Espoo, Finland
¹⁰Department of Information and Computer Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Finland
¹¹International Centre for Genetic Engineering and Biotechnology, P.O. Box 10504, 110067 New Delhi, India
¹²Immune Disease Institute, Harvard Medical School, Boston, MA 02115, USA
¹³These authors contributed equally to this work
¹⁴Correspondence: riitta.lahesmaa@utu.fi
 DOI 10.1016/j.immuni.2010.06.011



Group Leader

Computational
Biomedicine



Research Director

Bioinformatics

Vice Director

Turku Centre for Biotechnology



European Research Council
Established by the European Commission



Thank you



Web: <http://elolab.utu.fi>

Twitter: https://twitter.com/comp_biomed/

Facebook: <https://www.facebook.com/CompBiomedI/>



European
Research
Council



Horizon 2020
European Union Funding
for Research & Innovation