

Twenty years bioinformatics

Yves Moreau

KU Leuven
ESAT-STADIUS

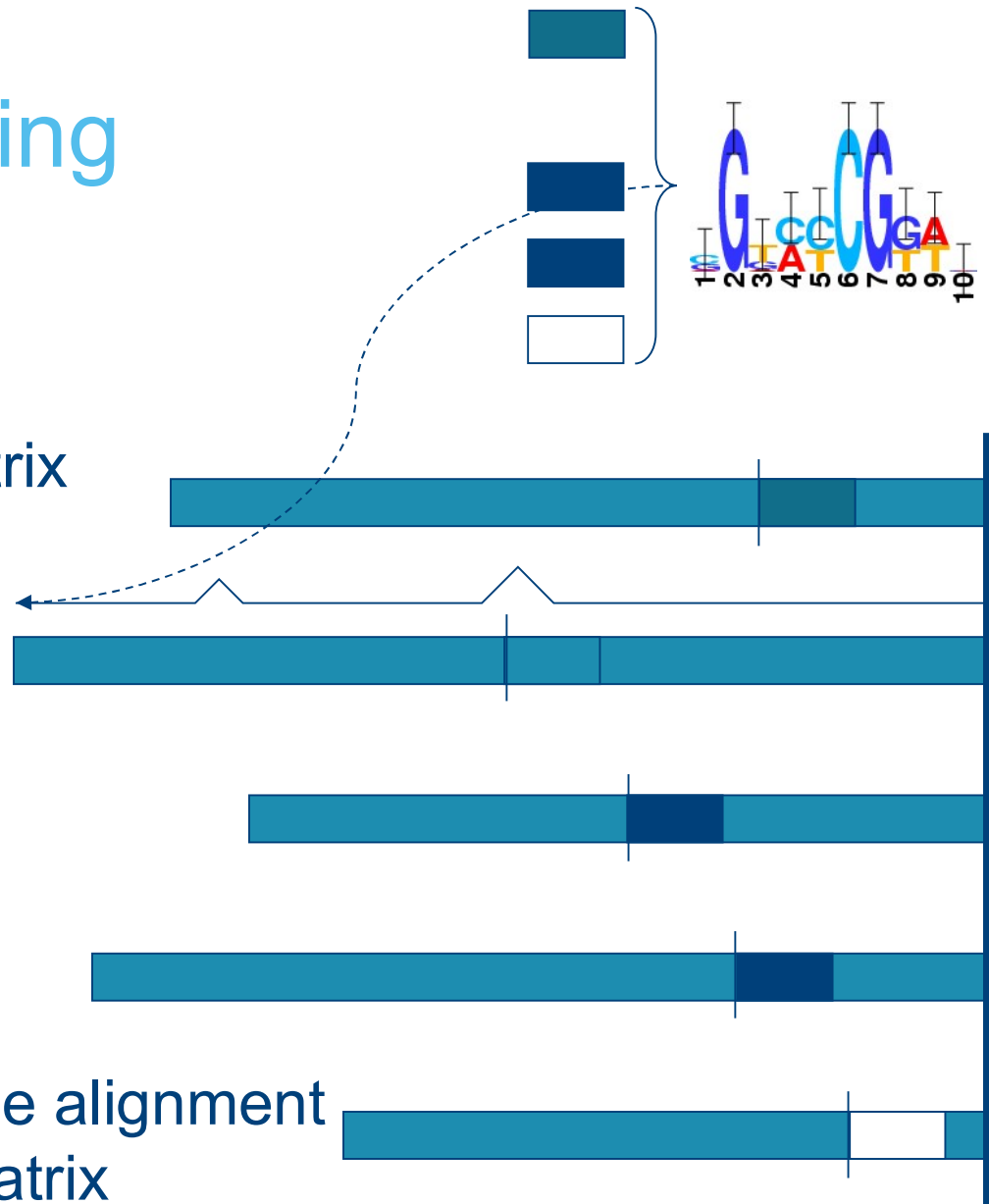
KU LEUVEN

How it all started



Gibbs motif finding

- Initialization
 - Sequences
 - Random motif matrix
- Iteration
 - Sequence scoring
 - Alignment update
 - Motif instances
 - Motif matrix
- Termination
 - Convergence of the alignment and of the motif matrix



How it all started

Use of a Multi-Layer Perceptron to Predict Malignancy in Ovarian Tumors

**Herman Verrelst,
Yves Moreau and Joos Vandewalle**
Dept. of Electrical Engineering
Katholieke Universiteit Leuven
Kard. Mercierlaan 94
B-3000 Leuven, Belgium

Dirk Timmerman
Dept. of Obst. and Gynaec.
University Hospitals Leuven
Herestraat 49
B-3000 Leuven, Belgium

Abstract

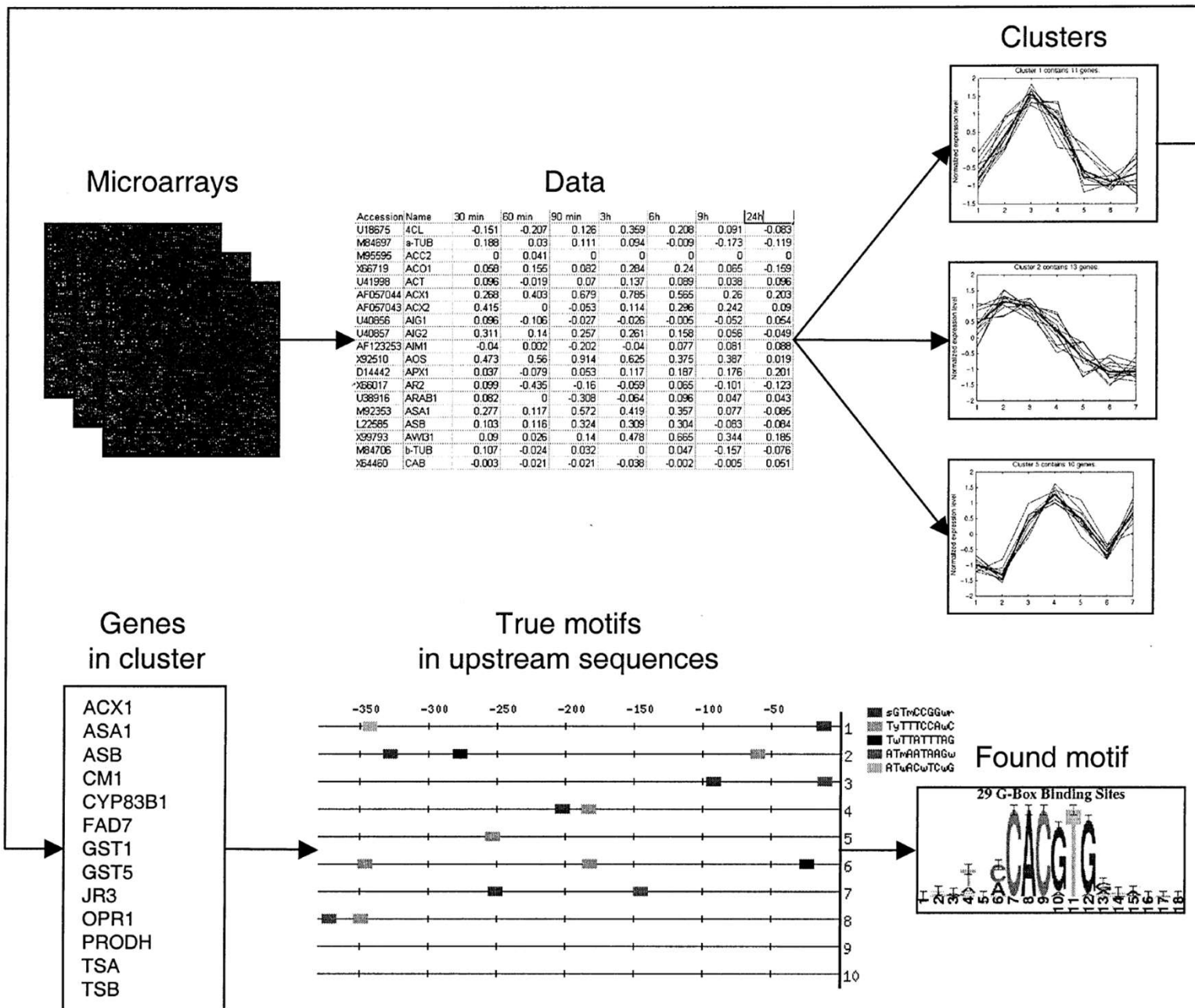
We discuss the development of a Multi-Layer Perceptron neural network classifier for use in preoperative differentiation between benign and malignant ovarian tumors. As the Mean Squared classification Error is not sufficient to make correct and objective assessments about the performance of the neural classifier, the concepts of sensitivity and specificity are introduced and combined in Receiver Operating Characteristic curves. Based on objective observations such as sonomorphologic criteria, color Doppler imaging and results from serum tumor markers, the neural network is able to make reliable predictions with a discriminating performance comparable to that of experienced gynecologists.



The Rise and Fall

The Nasdaq Composite Index, daily close





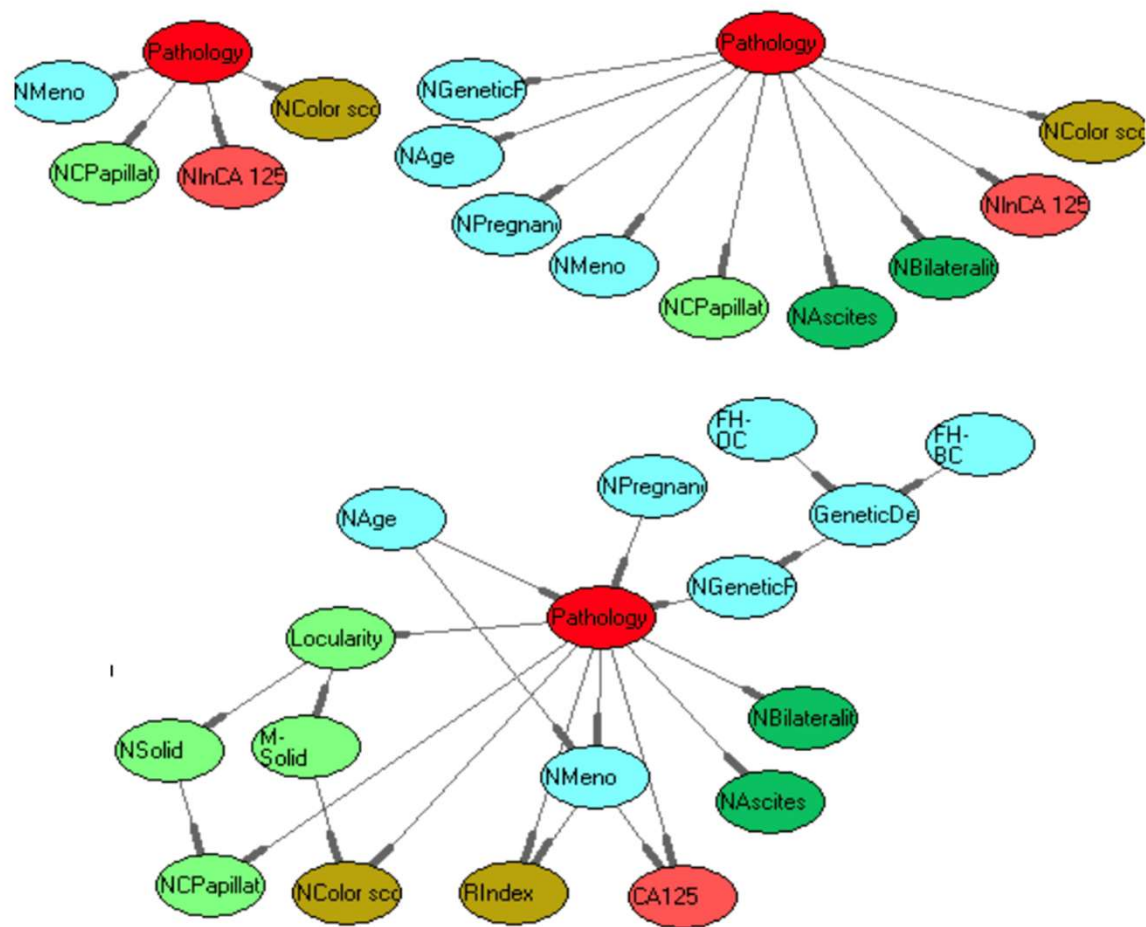
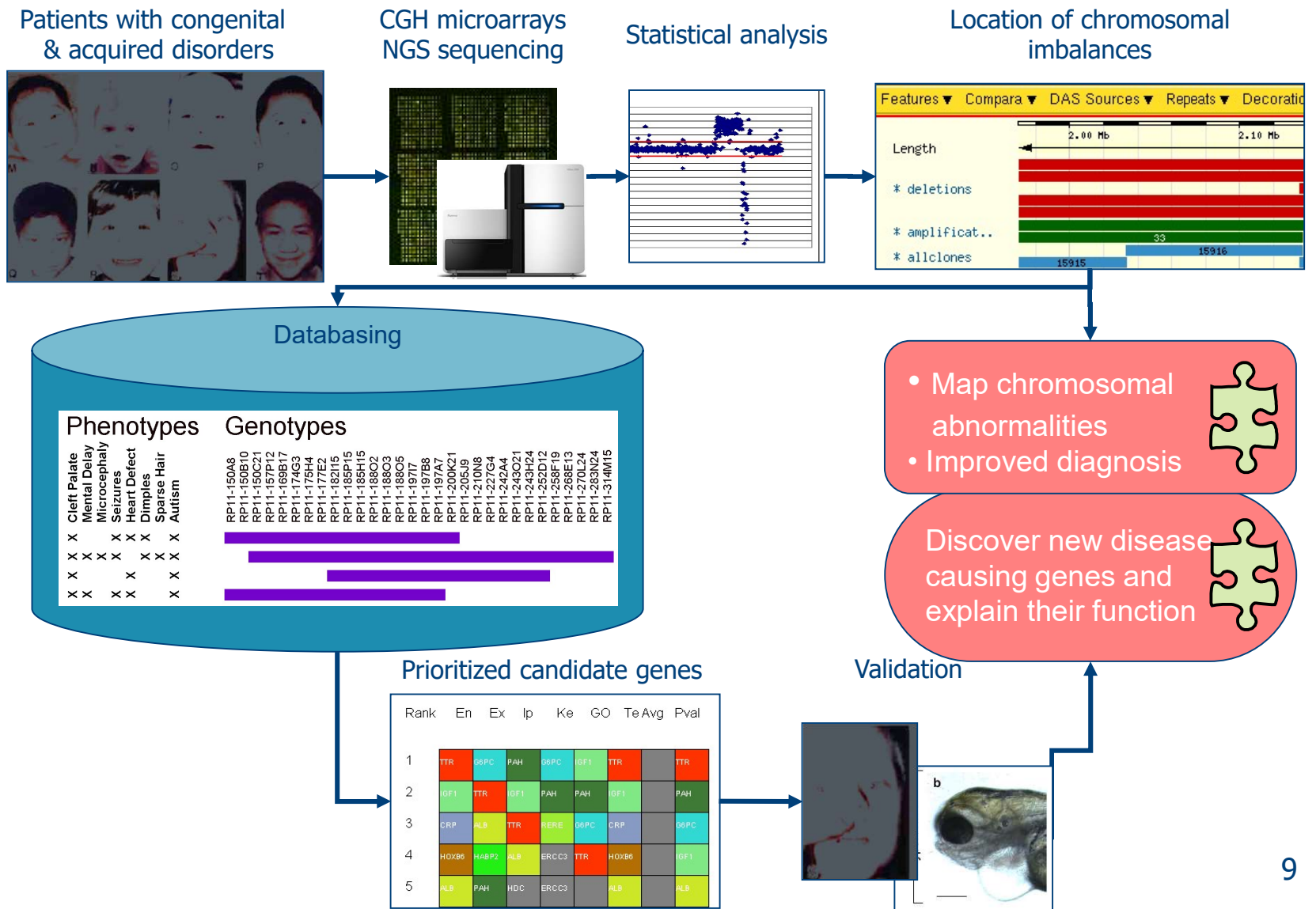


Figure 2.10. The *small naïve* (top left), *large naïve* (top right), and *standard* (bottom) Bayesian networks.

Disease gene discovery in rare congenital disorders






cartagenia

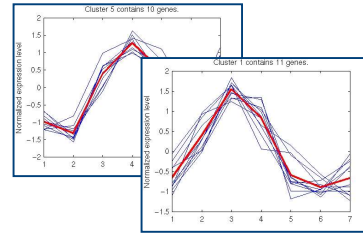
Candidate gene prioritization

High-throughput genomics



Array CGH – CNVs
Exome seq.
GWAS – SNPs
Expression

Data analysis



Candidate genes

Name	Ensembl
TTR	ENSG00000118271
PAH	ENSG00000171759
G6PC	ENSG00000131482
IGF1	ENSG0000017427
ALB	ENSG00000163631
CRP	ENSG00000132693
HABP2	ENSG00000148702
IF	ENSG00000138799
FST	ENSG00000134363
ARAF1	ENSG00000078061
HMG2	ENSG00000149948
C9	ENSG00000113600
PCBP2	ENSG00000111406
HOXB6	ENSG00000108511
RERE	ENSG00000142599
HOXA11	ENSG00000005073
CLIC1	ENSG00000096238
ERCC3	ENSG00000163161
ERCC3	ENSG00000163161
TLL2	ENSG00000095587
BYT4	ENSG00000132872
BYT4	ENSG00000132872
PIK4CB	ENSG00000143393
PKD2	ENSG00000118762
ANKRD3	ENSG00000081026
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	ENSG00000100000
C10orf101	ENSG00000100000
	ENSG00000100000
PDCD1	ENSG00000100000
PDGFRA	ENSG00000100000
NR2F1	ENSG00000100000
NFYA	ENSG00000100000
	ENSG00000100000
	ENSG00000100000
	ENSG00000100000
MMP3 MTM1	ENSG00000149968
	ENSG00000100000

Information sources

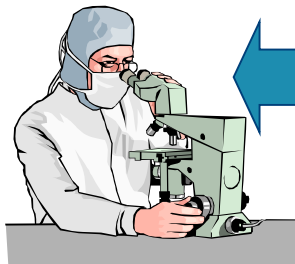
Information sources include:

- DNA methylation flowchart
- Microarray image
- Text snippet: "In their response, at 11, the why Michael Jackson performing an induction and told Angelina that she wanted to be their dog. This, said Angelina, who appeared here to make 18 minutes old, during which she underwent a shocking makeover that included platinum blonde, shorter hair and cape for the Swan Queen that had persuaded a "Kanye West" interview to talk about "The Love Train"."
- Diagram of p53 and Irf1 sites
- Protein structure
- Phylogenetic tree

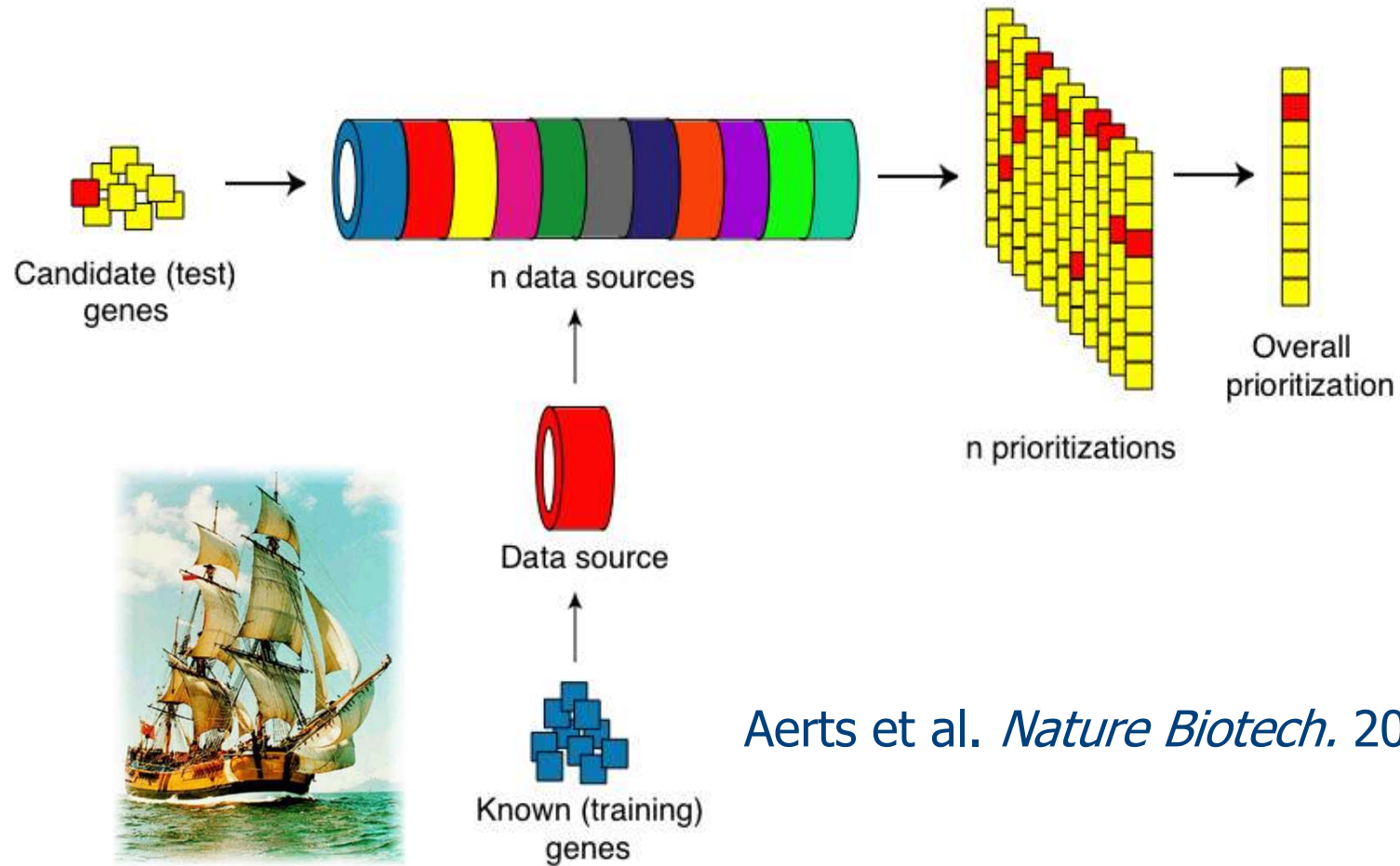
Candidate prioritization

Rank	En	Ex	Ip	Ke	GO	Te	Avg	Pval
1	TTR	G6PC	PAH	G6PC	IGF1	TTR		
2	IGF1	TTR	IGF1	PAH	PAH	IGF1		
3	CRP	ALB	TTR	RERE	G6PC	CRP		
4	HOXB6	HABP2	ALB	ERCC3	TTR	HOXB6		
5	ALB	PAH	HDC	ERCC3		ALB		
6	NR4A2	IF	TLL2	ANKRD3		HMG2		
7	PAH		C10R1	ARAF1	HDC	NR4A2		
8	HOXA11	IGF1	G6PC	PKD2	F13A1	PAH		
9	NFYA	CRP	HABP2	MTMR1	KCNN3	HOXA11	C13orf7	FST
10	C9	ARAF1	IF	HDC	CLIC1	NFYA	TTR	ARAF1

Validation



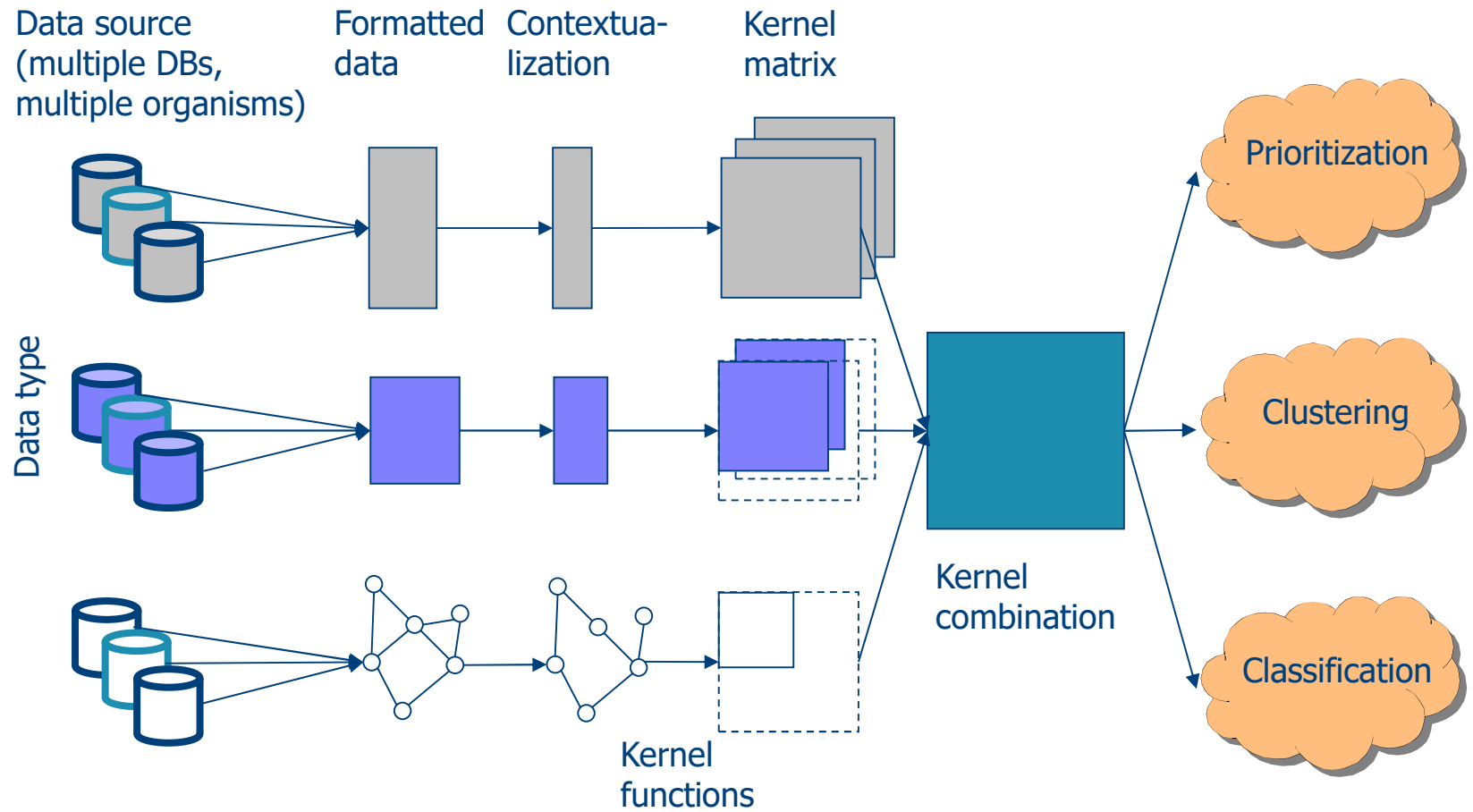
Data fusion with order statistics



Endeavour

Aerts et al. *Nature Biotech.* 2006

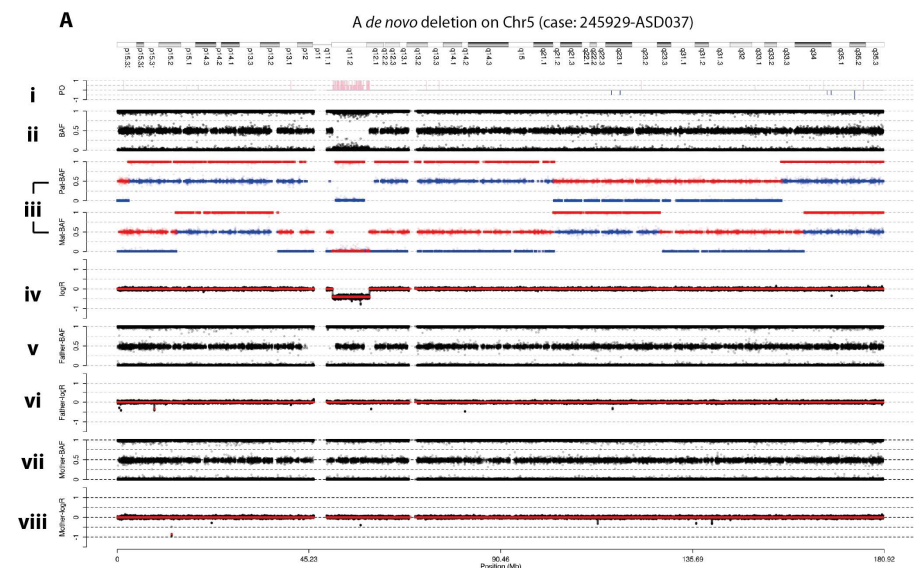
Kernel data fusion



Chromosome instability is common in human cleavage-stage embryos

Evelyne Vanneste^{1,2,9}, Thierry Voet^{1,9}, Cédric Le Caignec^{1,3,4}, Michèle Ampe⁵, Peter Konings⁶, Cindy Melotte¹, Sophie Debrock², Mustapha Amyere⁷, Miikka Vikkula⁷, Frans Schuit⁸, Jean-Pierre Fryns¹, Geert Verbeke⁵, Thomas D’Hooghe², Yves Moreau⁶ & Joris R Vermeesch¹

Chromosome instability is a hallmark of tumorigenesis. This study establishes that chromosome instability is also common during early human embryogenesis. A new array-based method allowed screening of genome-wide copy number and loss of heterozygosity in single cells. This revealed not only mosaicism for whole-chromosome aneuploidies and uniparental disomies in most cleavage-stage embryos but also frequent segmental deletions, duplications and amplifications that were reciprocal in sister blastomeres, implying the occurrence of breakage-fusion-bridge cycles. This explains the low human fecundity and identifies post-zygotic chromosome instability as a leading cause of constitutional chromosomal disorders.



Patient #1, Leuven, BE
6 *de novo* mutations
Chr 11, ~66Mb: PACS1?

T

TAC AAG AAT CGG ACC ATC
Tyr.Lys.Asn.*Trp*.Thr.Ile

Arg

Federated analytics
1,000,000 genomes



Patient #2, Nijmegen, NL
7 *de novo* mutations
Chr 11, ~66Mb: **PACS1!**

T

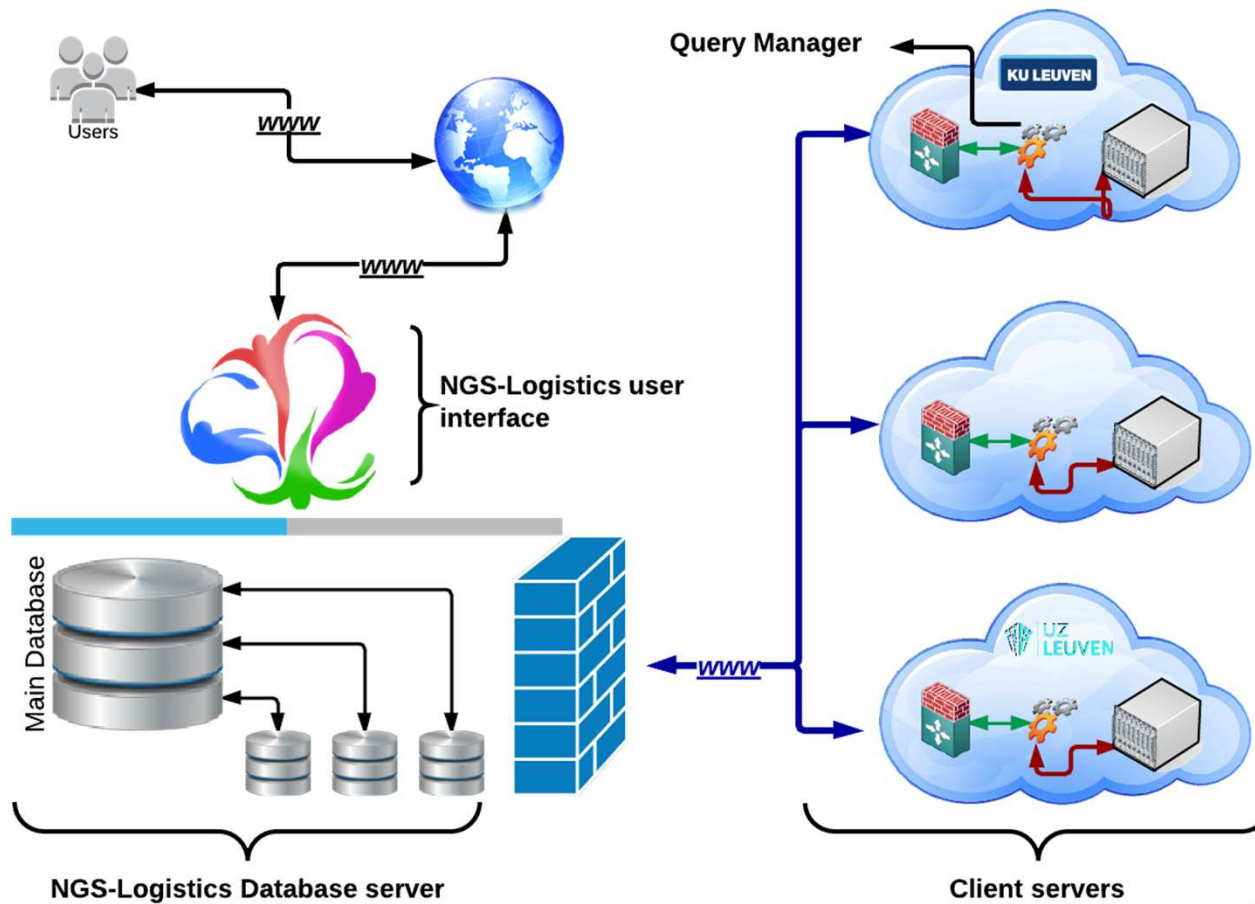
TAC AAG AAT CGG ACC ATC
Tyr.Lys.Asn.*Trp*.Thr.Ile

Arg

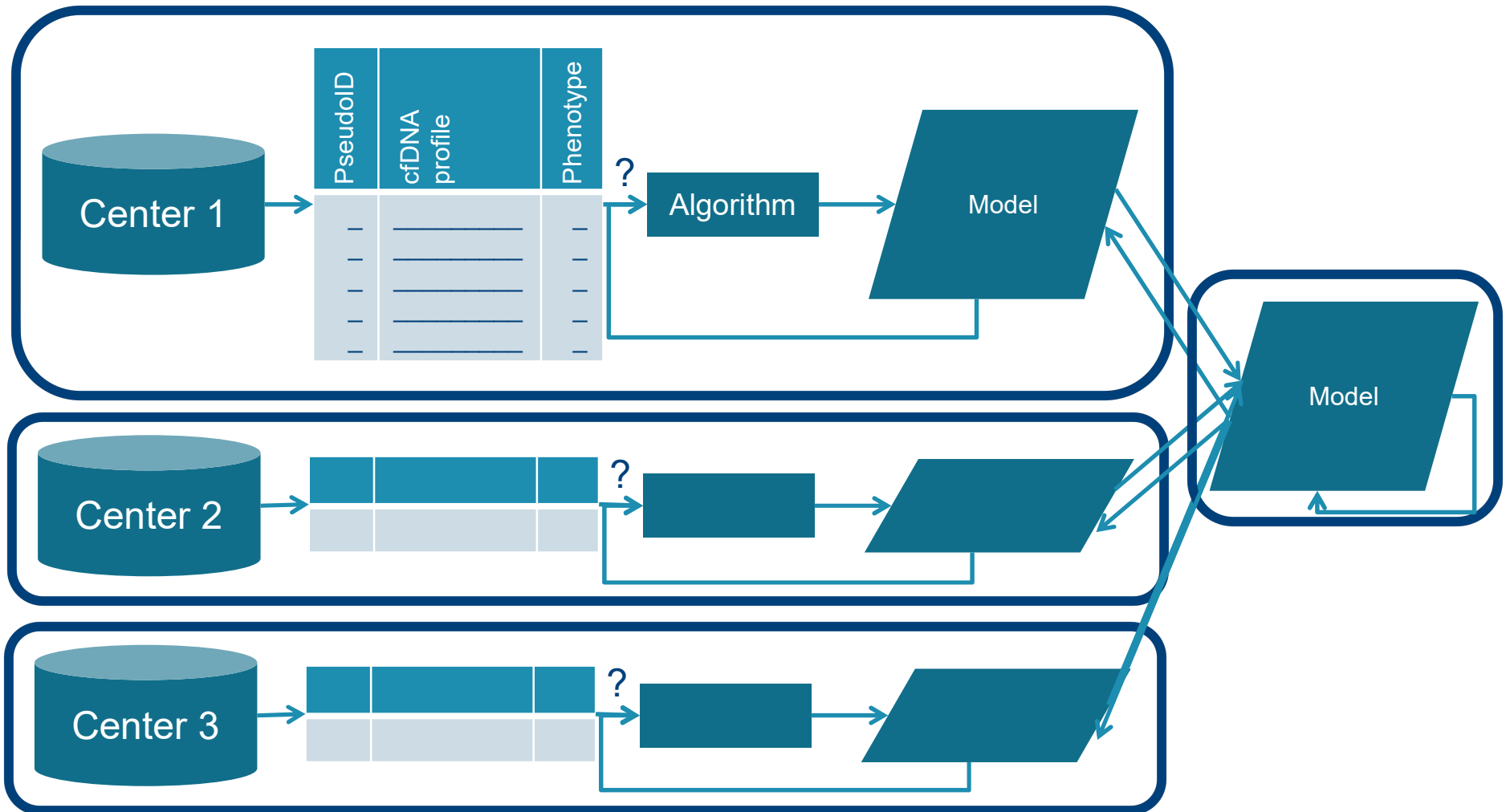


Southern
Ocean

Federated analytics

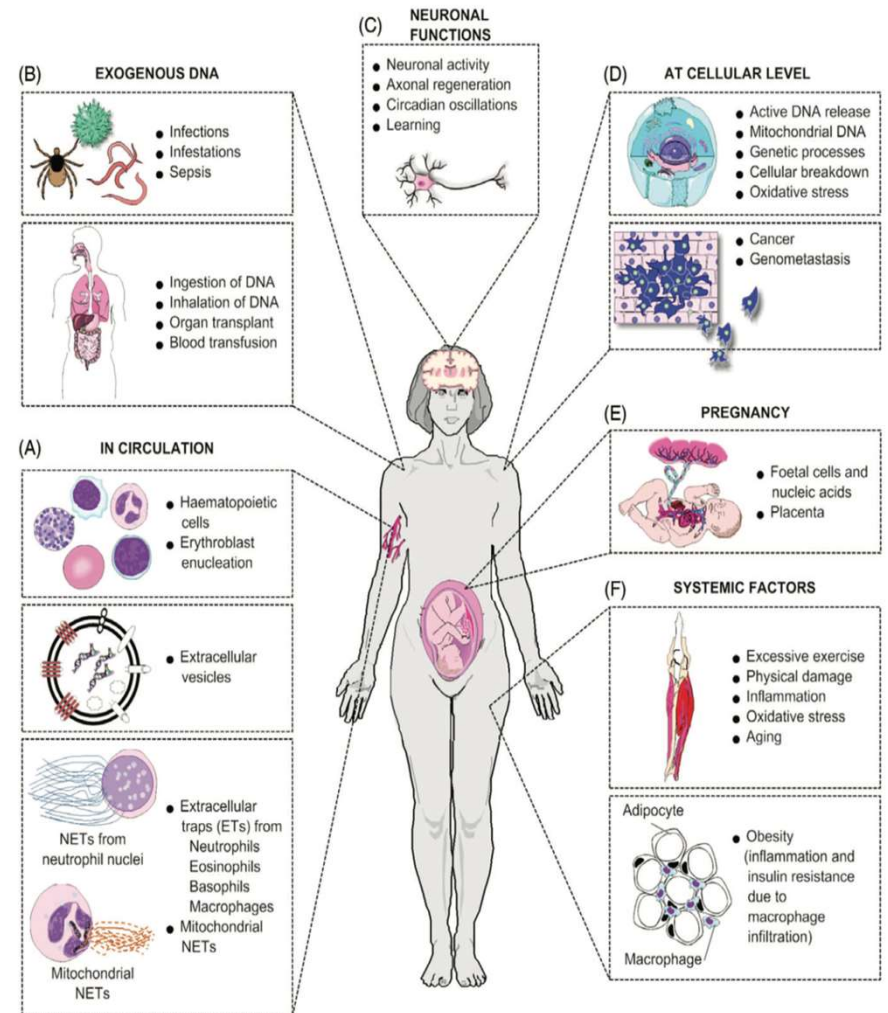


Federated machine learning



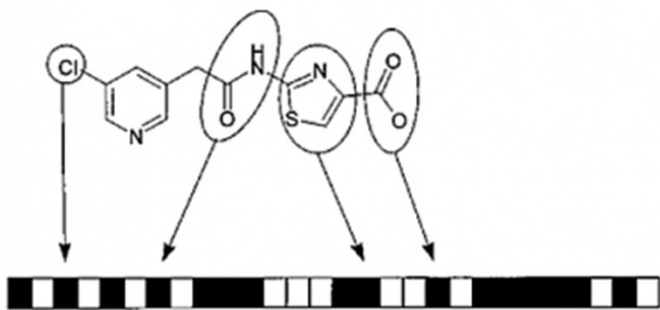
Liquid biopsies

- Cell-free genomic DNA can be shed into the blood stream or other bodily fluids for different reasons and via different pathways.
- Cell-free genomic DNA can provide a non-invasive way of detecting genomic imbalances within the body.



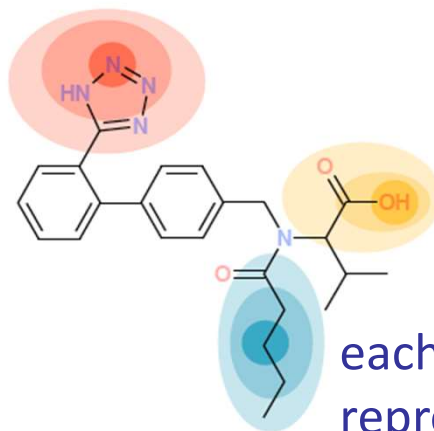
Machine learning for drug discovery

Path/key-based fingerprints
FP2 & MACCS



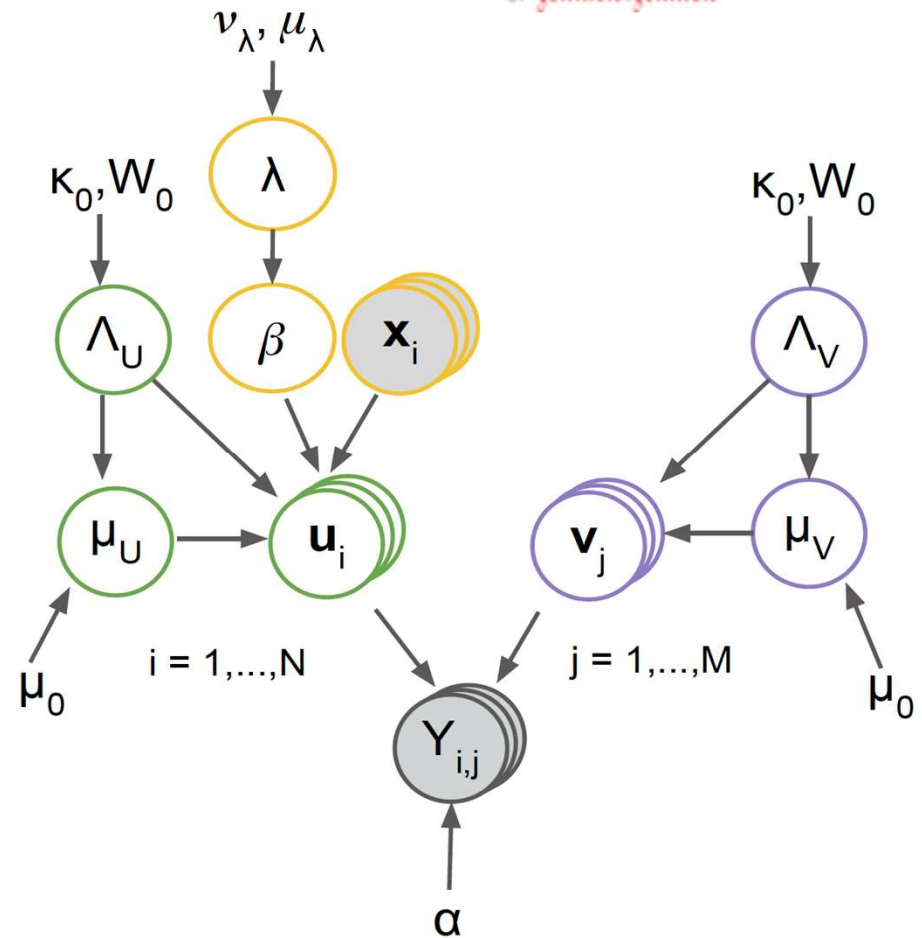
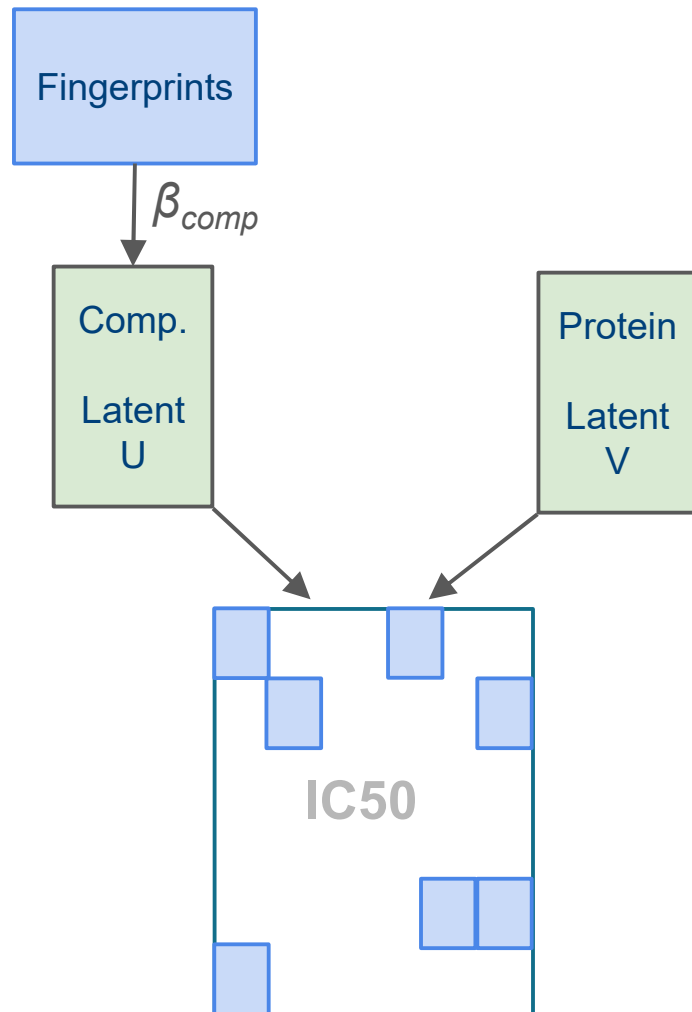
bitstrings representing presence or absence of particular substructures for each molecule, there is only one fingerprint

Circular fingerprints
MNA & MPD

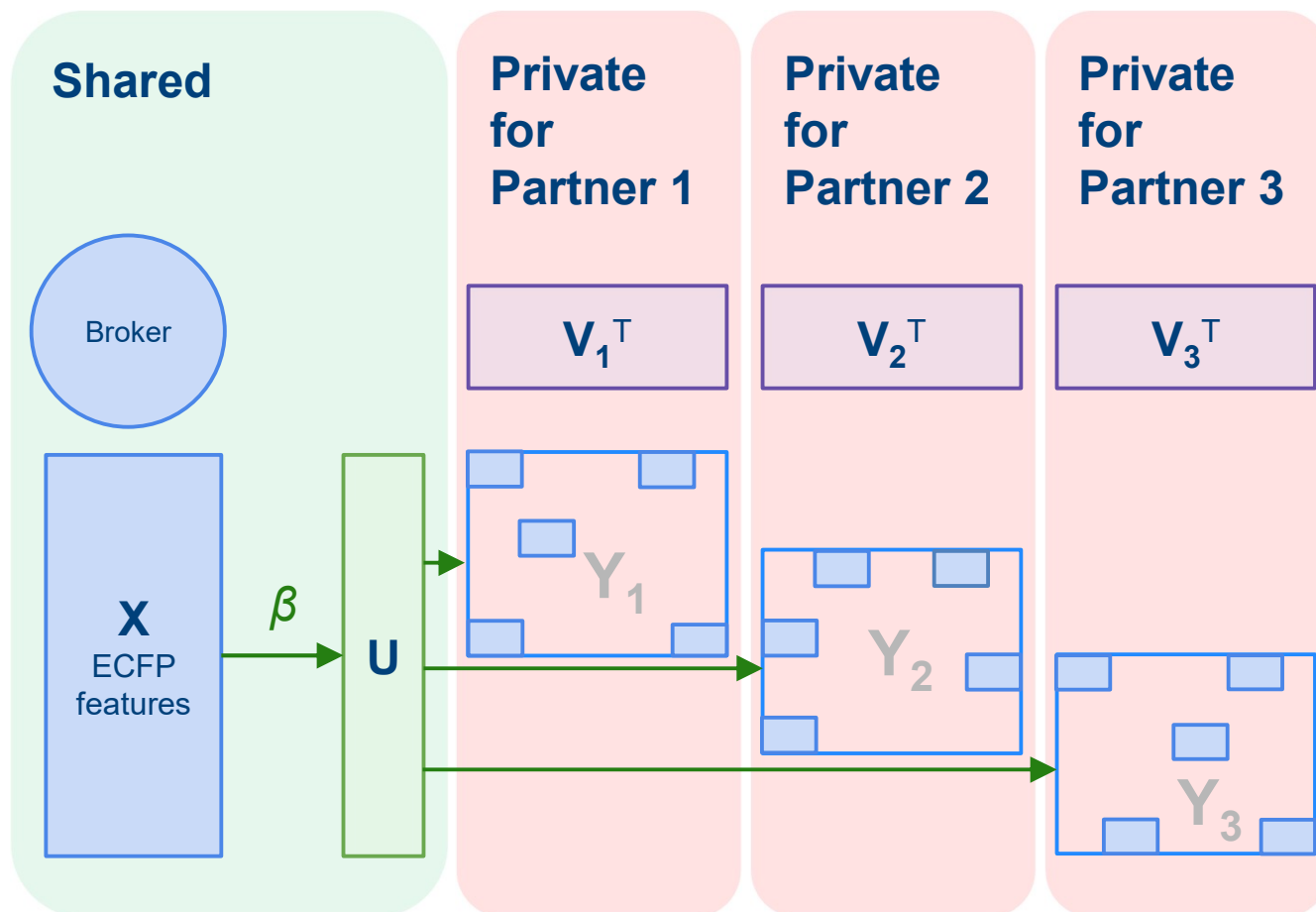


each fingerprint represents a central atom and its neighbors for each molecule, there are as many fingerprints as (heavy) atoms in the molecule

Bayesian matrix factorization



Privacy-preserving machine learning



Initialization

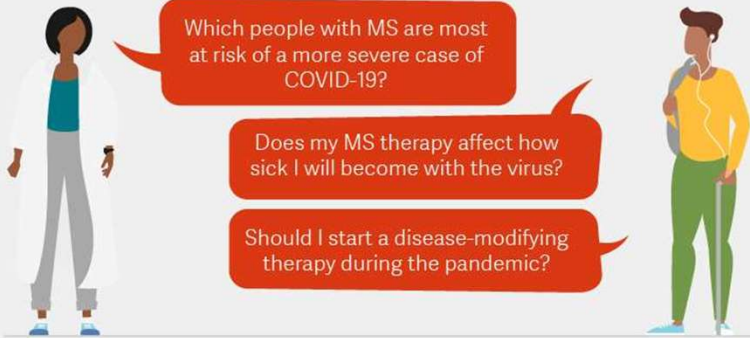
Broker receives X from each partner and aligns them

Iteration

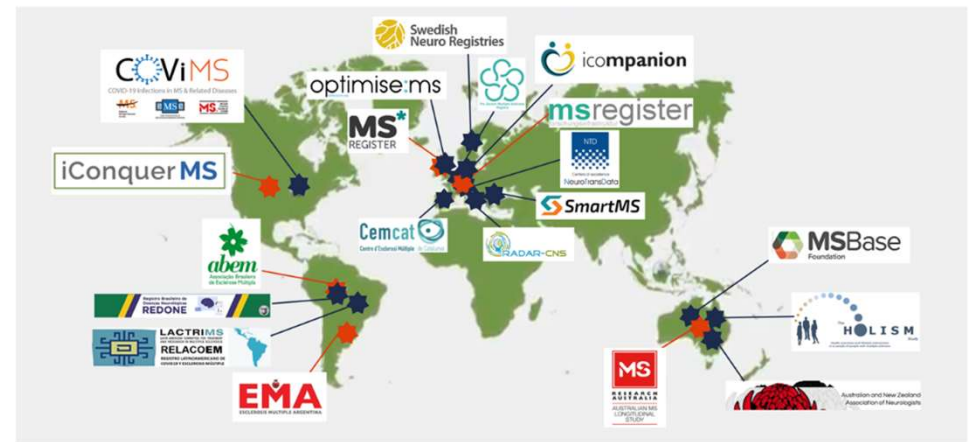
1. Partners privately update V
2. Partners send contributions for U to broker
3. Broker computes and shares U
4. Broker updates β

Federated analytics of real-world clinical data

Thanks to everyone who has taken part, this study has enabled us to answer important questions about COVID-19 in people with MS.



Over 80 countries are participating in this initiative



A large number of partners were involved



acting under the umbrella of the European Charcot Foundation



Karolinska Institutet



*Put up in a place
where it is easy to see
the cryptic admonishment
T.T.T.*

*When you feel how depressingly
slowly you climb
it's well to remember that
Things Take Time.*

— Piet Hein

Thank you to so many people!