Integrated analysis of metabolite, gene expression and other profile data: Missing value estimation, dimension reduction, clustering, and classification

Joachim Selbig

University of Potsdam
and
Max Planck Institute of Molecular Plant Physiology

Leicester, August 24, 2006

Potsdam-Golm Campus

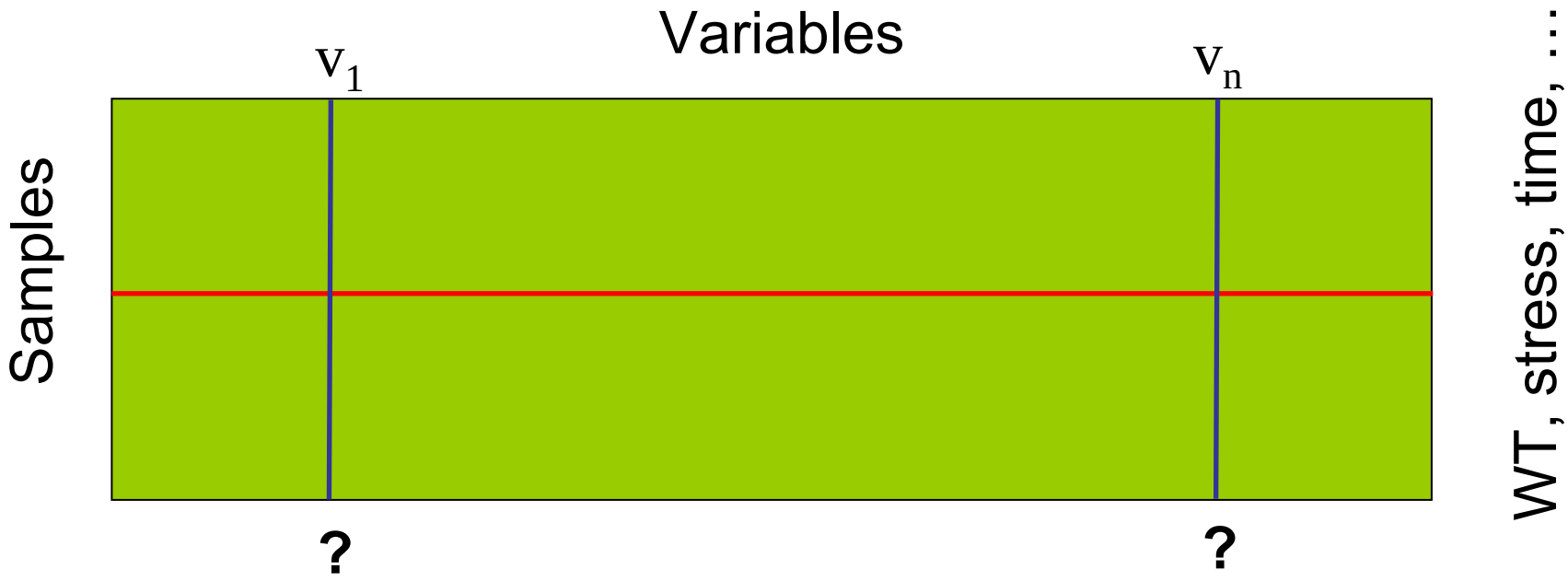
Potsdam University

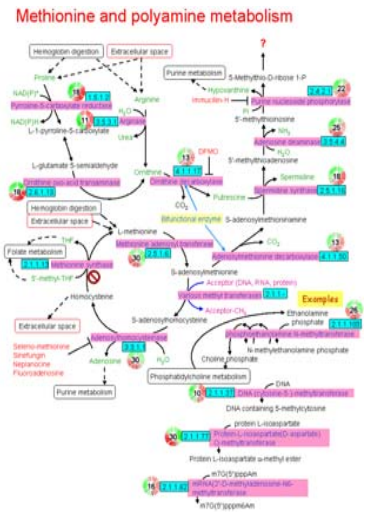
MPI Campus

Bioinformatics group

Circle a symbol representing an
open system between
magic and calculus …

Biological profile data





Genes, metabolites,
or proteins, …

Issues

- # samples << # metabolites / genes / proteins …
  ‚Curse of dimensionality‘
  Dimensionality reduction: ~ 1.000 → 2 … 10
  and visualisation
- Normalisation
- Missing value substitution
- Non-linearity
- …

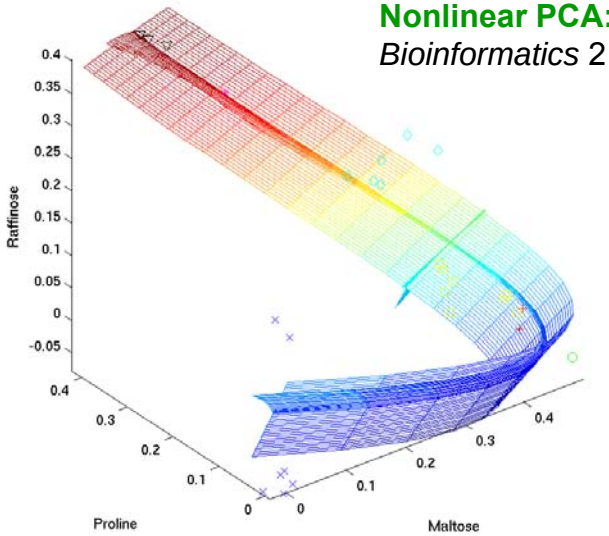
Missing value estimation

missing				
999.00	−0.05	0.08	0.74	999.00
−0.16	−0.14	0.01	0.41	−0.04
−0.05	0.01	0.12	0.79	−0.09
−0.17	−0.08	0.06	0.72	−0.18
−0.14	−0.10	0.03	0.76	−0.22
−0.13	−0.05	0.11	0.86	−0.09
−0.30	−0.39	−0.13	0.80	0.06
−0.17	−0.12	0.07	0.72	−0.14
−0.11	−0.20	−0.01	0.78	−0.19
0.10	999.00	0.19	999.00	−0.14
−0.07	999.00	0.21	999.00	999.00
−0.05	−0.03	0.22	0.60	−0.10
−0.11	−0.11	0.06	0.75	−0.16
−0.10	−0.03	0.10	999.00	−0.18
0.32	0.81	0.43	999.00	−0.14
−0.13	−0.11	0.07	0.69	−0.13
0.27	−0.42	999.00	0.17	−0.41
−0.28	−0.33	−0.10	0.78	0.05
999.00	0.62	−0.60	−0.37	1.04
999.00	−0.52	−0.21	0.34	−0.39
−0.10	−0.06	0.09	999.00	−0.16

- Deleting rows
- Row Average or filling with zeros
- Singular Value Decomposition (SVD)
- Weighted K-nearest neighbors (KNN)
- Linear regression using Bayesian gene selection
- Non-linear regression using Bayesian gene selection

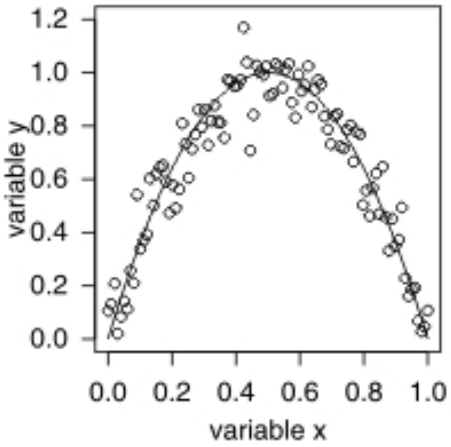
M Scholz, F Kaplan, CL Guy, J Kopka, J Selbig (2005)
Nonlinear PCA: a missing data approach.
Bioinformatics 21(3887-3895).

Non-linear Principal Component Analysis (NLPCA) provides better results than other methods

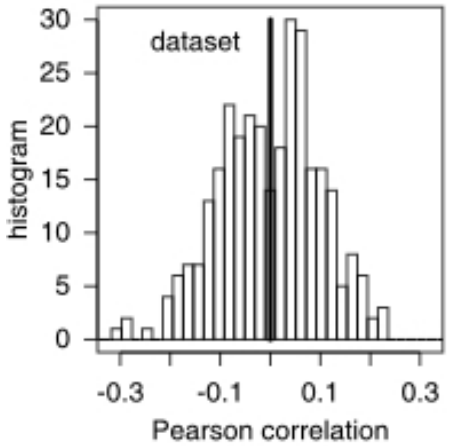


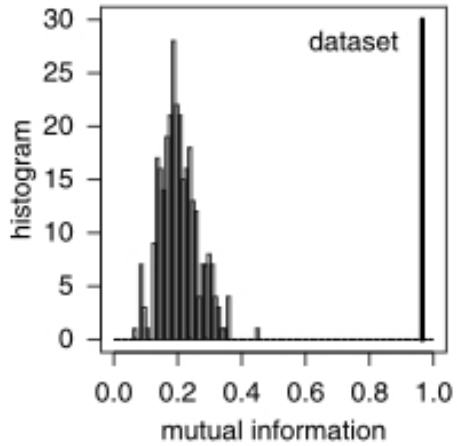
Data
A.th. cold stress data
Electrospray/QTOF mass spectra

Non-linearity / significance



$$S := \frac{MI(X,Y)^{\text{data}} - < MI(X,Y)^{\text{surr}} >}{\sigma^{\text{surr}}}$$

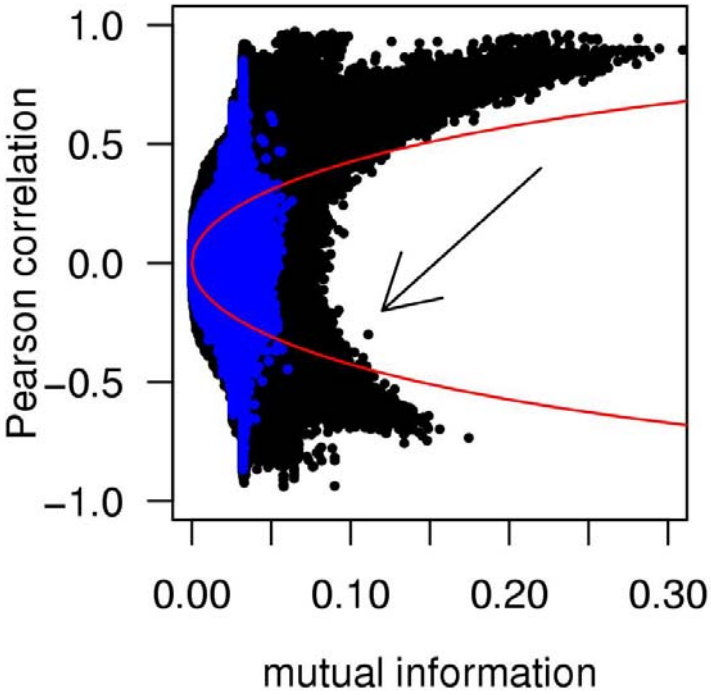


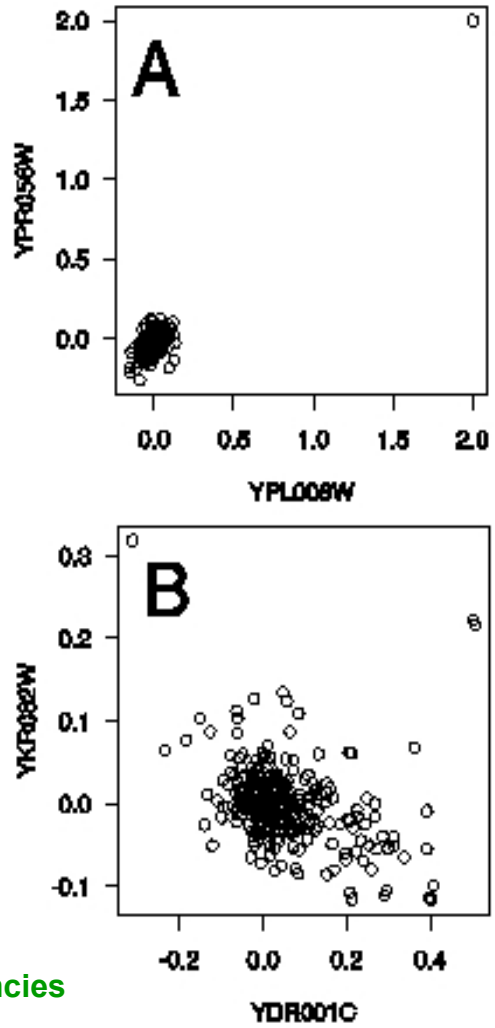


C Daub, R Steuer, J Selbig, S Kloska (2004)
Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data.
BMC Bioinformatics 5:118.

Application to Data

Yeast dataset (Hughes et al., 2000)
gene expression for ~6000 genes
under 300 experiments

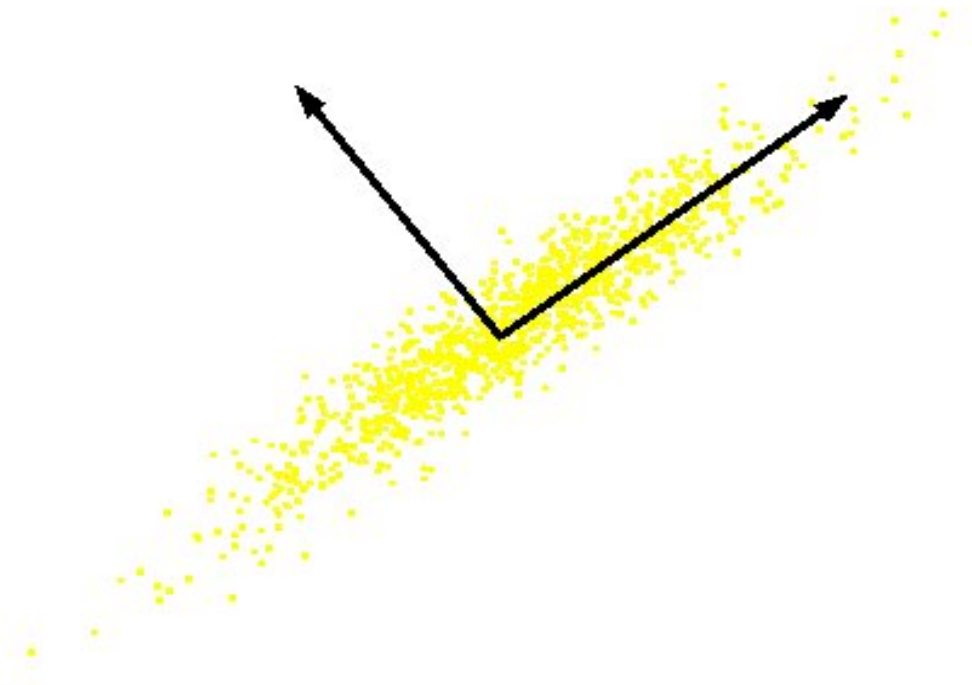




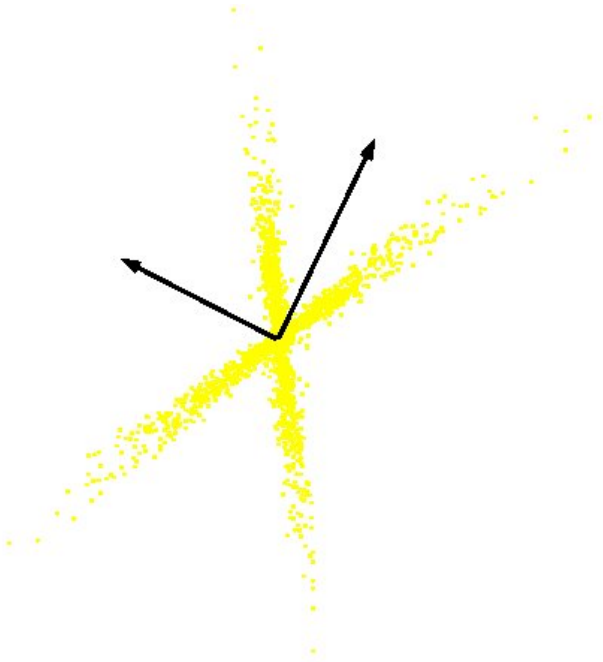
R Steuer, J Kurths, C Daub, J Weise, J Selbig (2002)
The mutual information: Detecting and evaluating dependencies between variables.
Bioinformatics 18(S231 - S240).

Universität Potsdam

Example of PCA



PCA versus ICA



PCA
(orthogonal coordinate)

ICA
(non-orthogonal coordinate)

PCA

- One well-established technique for dimensionality reduction and visualization is the classical principal component analysis (PCA), where the extracted information is represented by a set of new variables, termed components or features. Diamantaras and Kung (1996) give a good overview of different PCA-algorithms.
- In the field of metabolomics, PCA became a popular tool for visualizing datasets and for extracting relevant information (Ward *et al.*, 2003; Urbanczyk-Wochniak *et al.*, 2003).
- However, **PCA is 'only' powerful if the biological question is related to the highest variance in the dataset.** If this is not the case, other techniques of statistics or related fields may be more helpful, depending on the biological question, as shown by Goodacre *et al.* (2003) and Johnson *et al.* (2003) for supervised techniques in combination with validation and pre-processing.

ICA

- In ICA, an independence condition is optimized, which often gives more meaningful components than optimization of only the variance, as is done by PCA.
- Because of this the components of ICA are termed independent components (ICs), meaning that different ICs represent different non-overlapping information.
- For applying ICA we assume that the observed data have been determined by **some unknown fundamental factors, which are independent of each other**.
- By searching for components as statistically independent as possible these required factors can be detected. These fundamental factors are often termed sources and the application field is called blind source separation, BSS.

Dimensionality reduction: Data

Arabidopsis thaliana

Parents	Co10	C24
Co10	Co10 x Co10	Co10 x C24
C24	Co10 x C24	C24 x C24

Total samples: 96

↓ **MS**

Electrospray/QTOF mass spectra

↓ **Weighted density function**

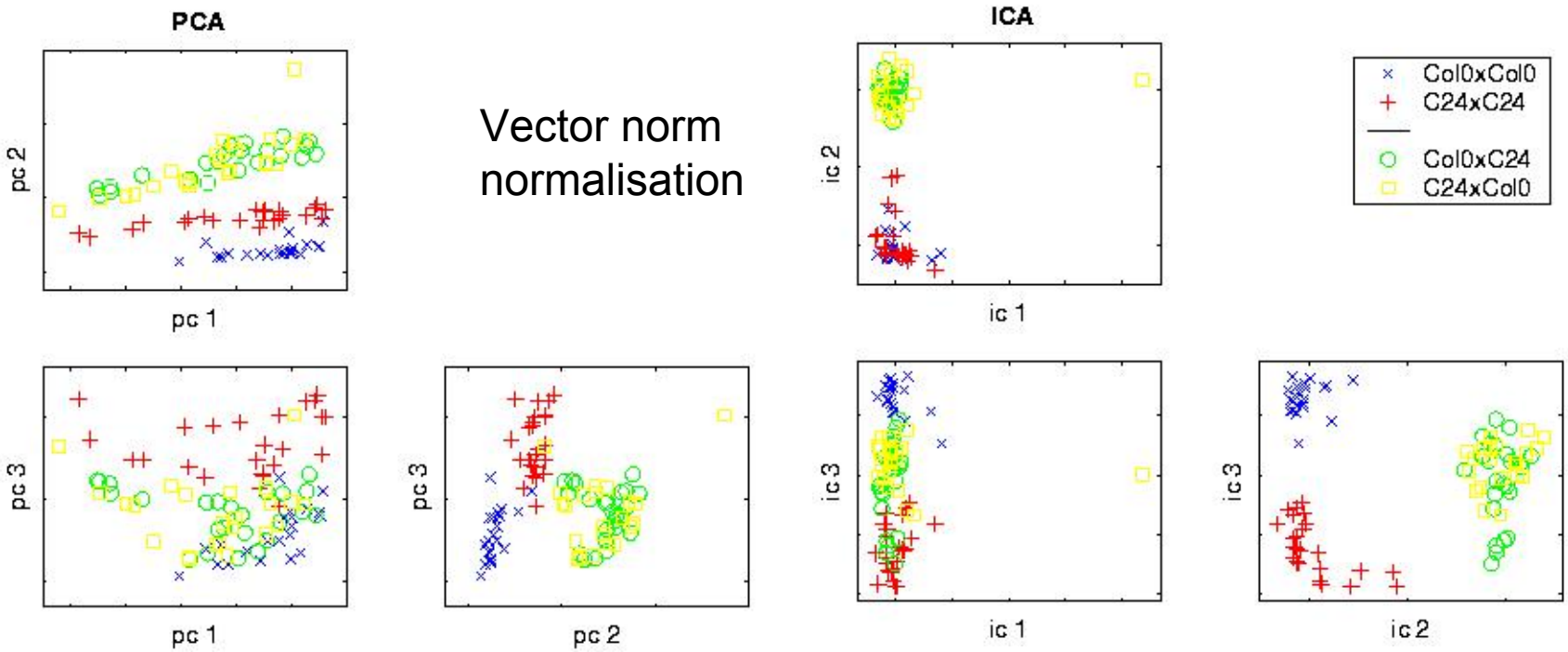
763 variables, 96 samples

Hybrid vigour or Heterosis:
display interesting features such as higher growth, better fitness and improved resistance against biotic and abiotic stress factors.

Therefore, we expected to find the largest distance between the F1 groups and the parents, the second largest difference between the two parents and just a small difference or none at all between the two F1 genotypes.

J Taylor, RD King, T Altmann, and O Fiehn (2002)
Application of metabolomics to plant genotype discrimination using statistics and machine learning.
Bioinformatics 18(S241-248).

PCA versus ICA (II)



M Scholz, S Gatzek, A Sterling, O Fiehn, J Selbig (2004)
Metabolite fingerprinting: detecting biological features by Idependent Component Analysis.
Bioinformatics 20(2447-2454).

ICA details

Model: $x = As$

mixing matrix

$s = Wx$

separating matrix

Universität Potsdam

…continued

Driving idea for finding sources: $\mathbf{s_1}$, $\mathbf{s_2}$ are *statistically independent* ==
information about one gives no knowledge over the other.

Not just *uncorrelated*: covariance = 0

$$\begin{aligned} cov(s_1, s_2) &= E\{(s_1 - \bar{s}_1) \cdot (s_2 - \bar{s}_2)\} \\ &= E\{s_1 \cdot s_2\} E\{s_1\} E\{s_2\} \end{aligned}$$

== PCA

…continued

If ***independent*** as well, the pdf is separable:

$$p(s_1, s_2) = p_1(s_1)\, p_2(s_2)$$

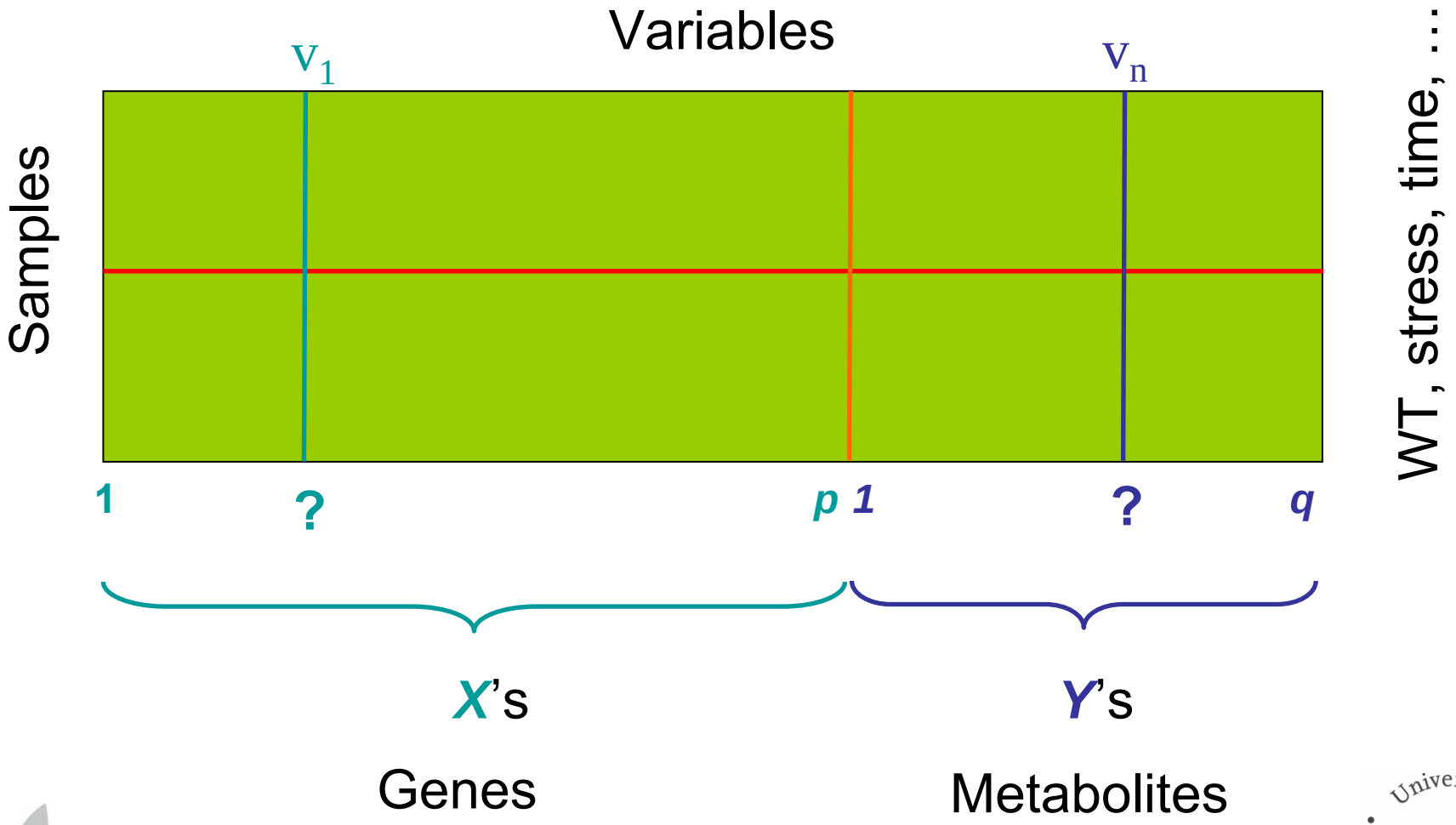
joint pdf — marginal pdf's

which implies

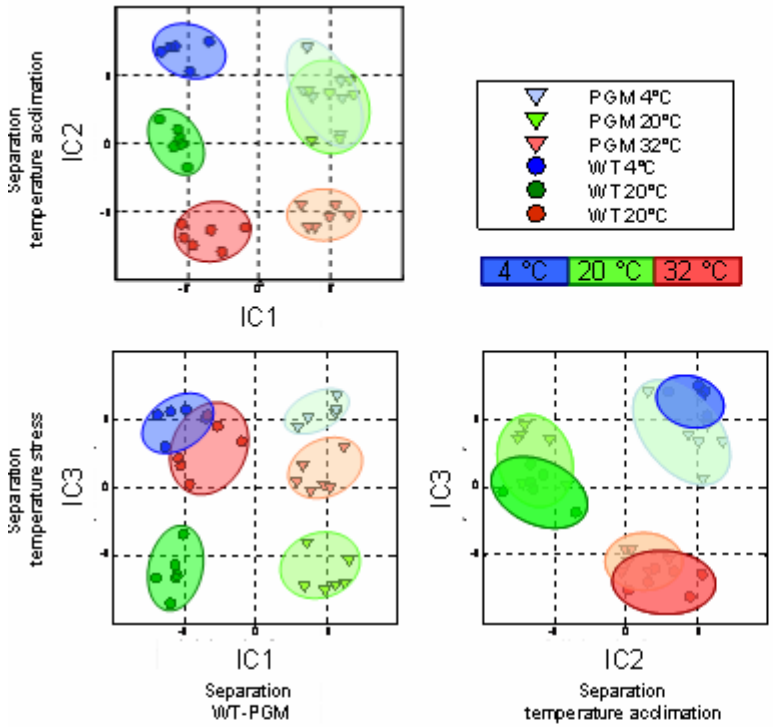
$$E\{f_1(s_1)\, f_2(s_2)\} - E\{f_1(s_1)\}E\{f_2(s_2)\} = 0$$

for ***any*** functions f_1 f_2 $\Rightarrow$ useful for solving.

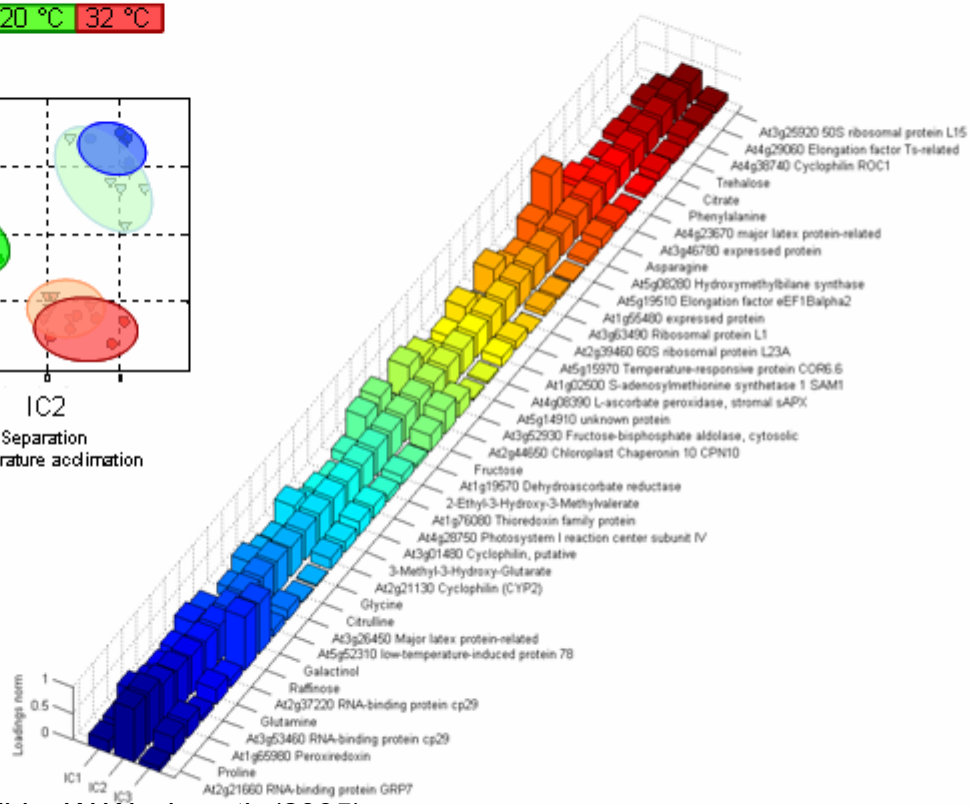
Biological profile data (II)



1st approach: Merging protein and metabolite profiles



A. th. temperature acclimation experiment PGM / WT



6 x 6 replicates
280 metabolites
200 proteins

K Morgenthal, S Wienkoop, M Scholz, J Selbig, W Weckwerth (2005)
Correlative GC/TOF/MS based metabolite profiling and LC/MS based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection. *Metabolomics* 1(109-121).

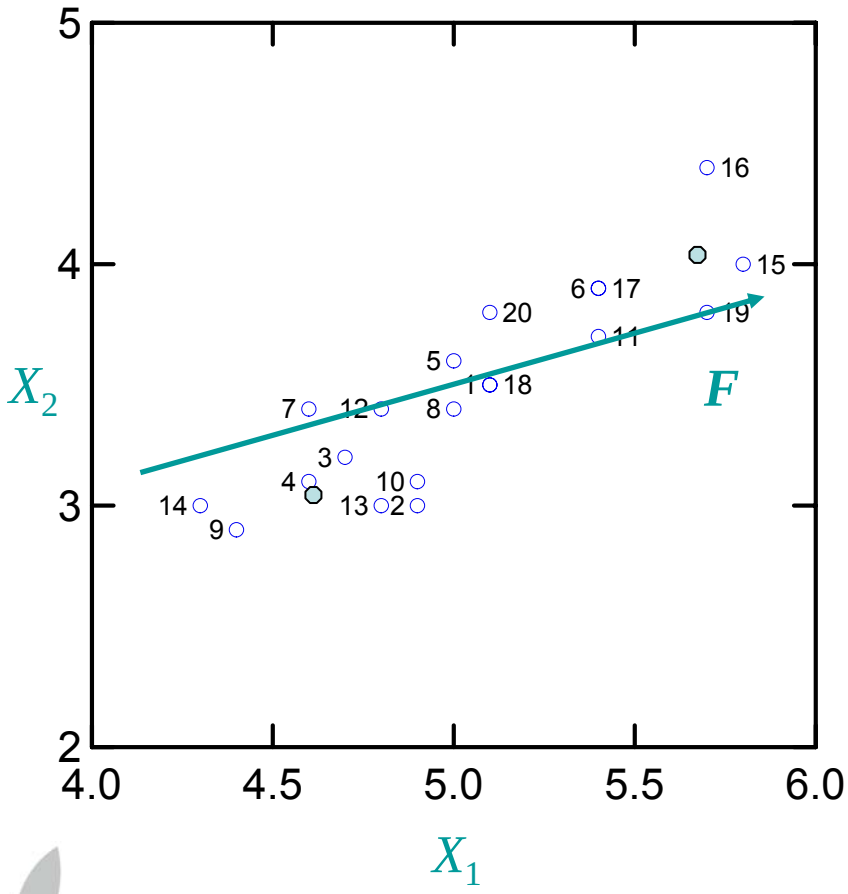
Universität Potsdam

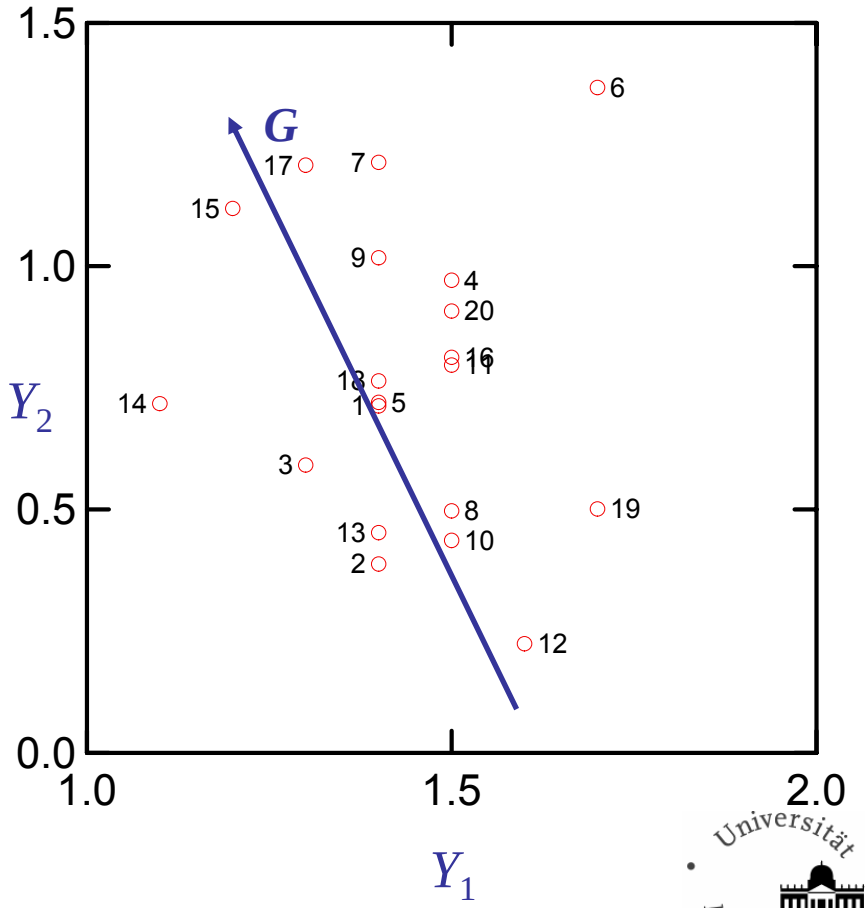
2nd appoach: Canonical Correlation Analysis

- Multivariate extension of correlation analysis
- Looks at relationship between ***two*** sets of variables

CCA

Maximize $r(F,G)$





CCA (2)

Given a linear combination of X variables:

$$F = f_1X_1 + f_2X_2 + ... + f_pX_p$$

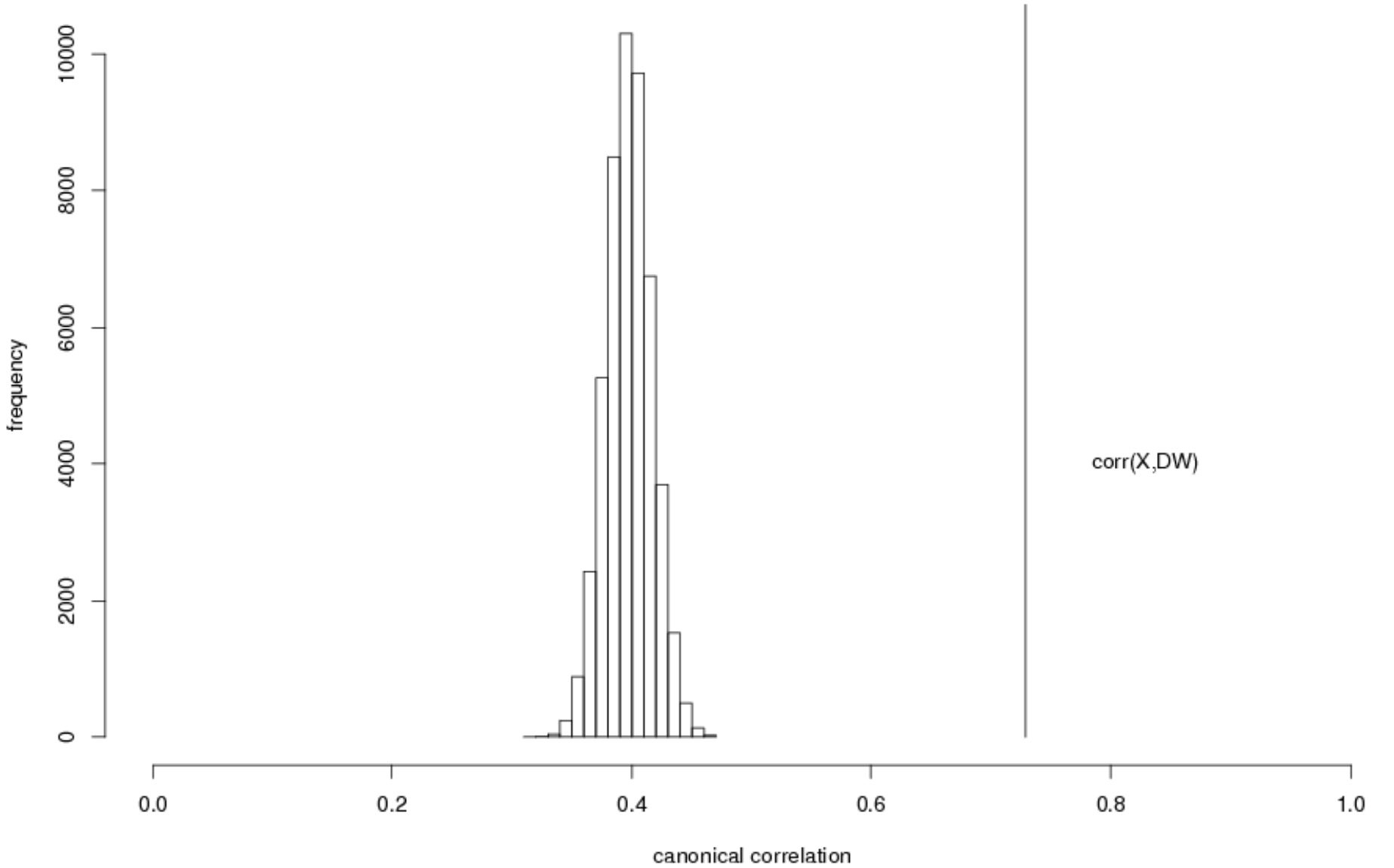
and a linear combination of Y variables:

$$G = g_1Y_1 + g_2Y_2 + ... + g_qY_q$$

The **first canonical correlation** is:
Maximum correlation coefficient between F and G,
for all F and G

$F_1=\{f_1,f_2,...,f_p\}$ and $G_1=\{g_1,g_2,...,g_q\}$
are corresponding **canonical variates**

A. th. Heterosis (Steinfath et al., in prep.)



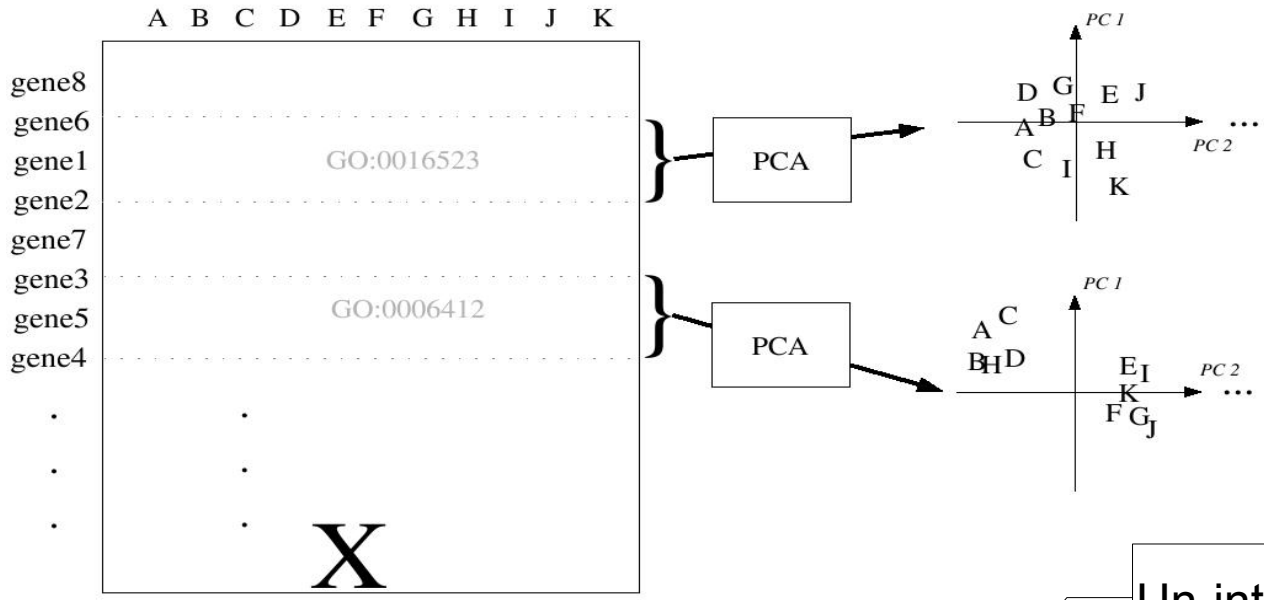
CAPIU

Clustering using **A**-**P**riori **I**nformation via **U**nsupervised decision trees

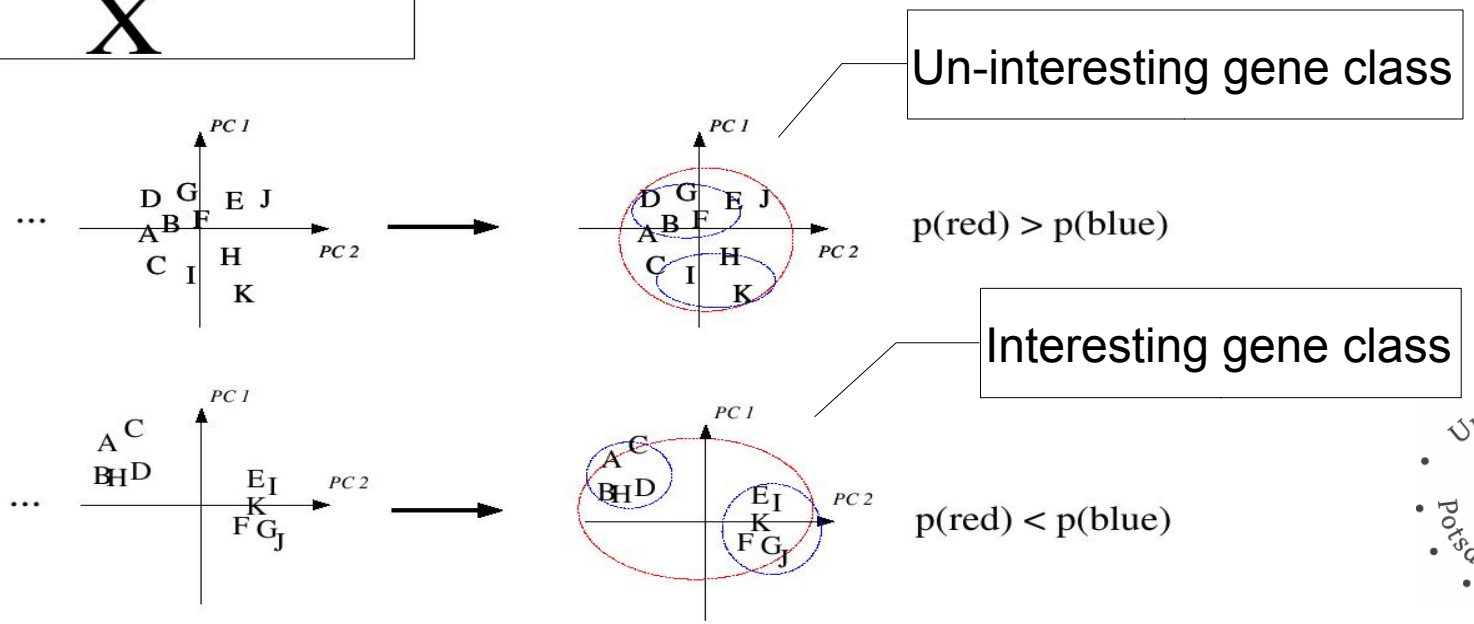
- **Clustering of *genes*** can reveal functionally related regulons, **clustering of *samples*** can reveal related treatments, knockouts, etc.
- Classical methods rely on variance filtering so any relatively low variant but informative genes are lost
  - Condense data to gene classes by **integrating functional annotations**
- Classical methods for clustering samples are hard to interprete: **What are the biological reasons for observed clusterings?**
  - Use same philosophy as for Decision Trees: one-feature-at-a-time

Select the most interesting gene classes and split the data using information from only these

The approach

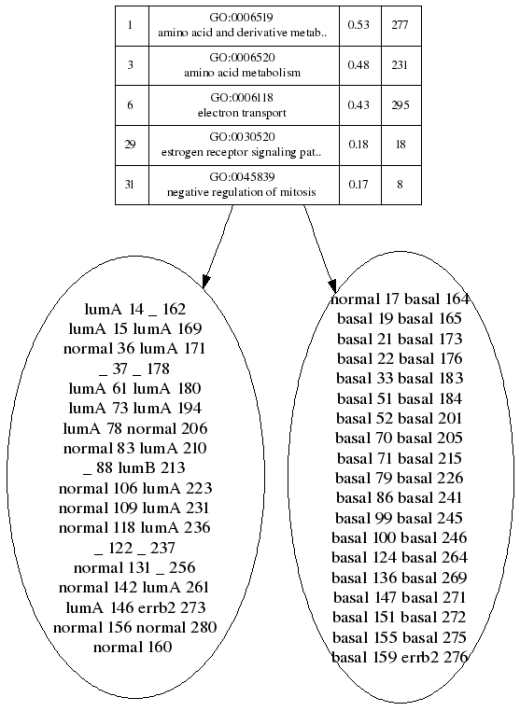


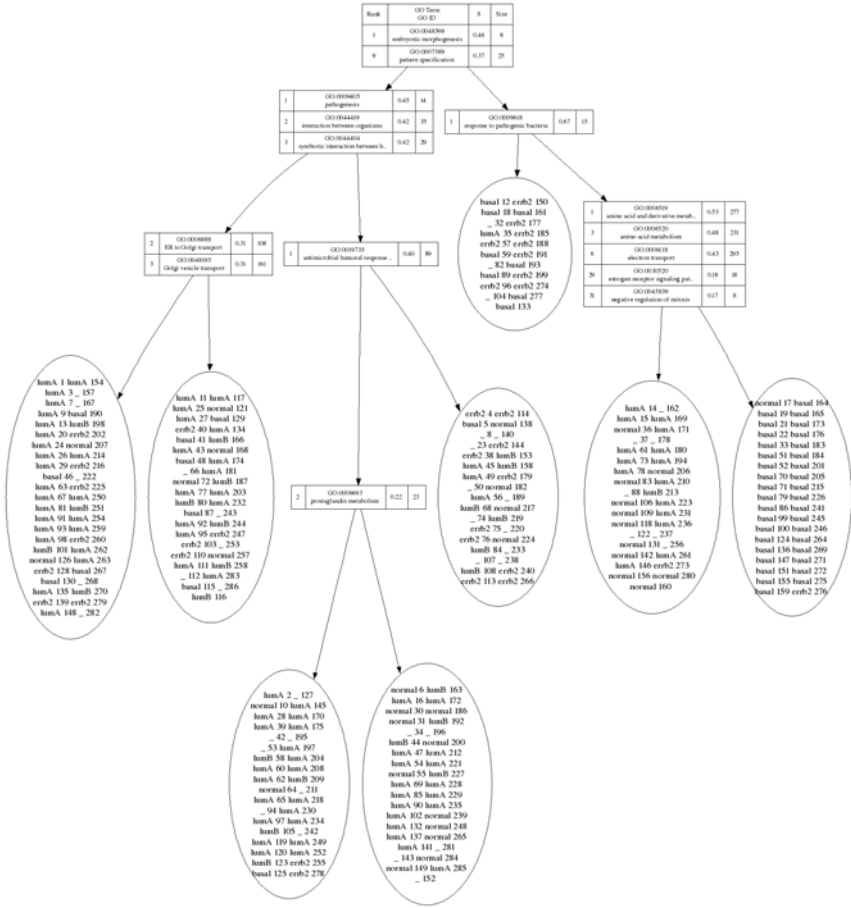
H Redestig, D Repsilber, F Sohler, J Selbig (2006)
Integrating functional knowledge during sample clustering for microarray data using unsupervised decision trees.
Accepted by *Biometrical Journal*.



Dataset I

Five types of breast cancer
Number of samples: 286
Number of genes: 17816
TYPE: five types of breast cancer (lumA, lumB, normal, errb2, basal and unclassified _)





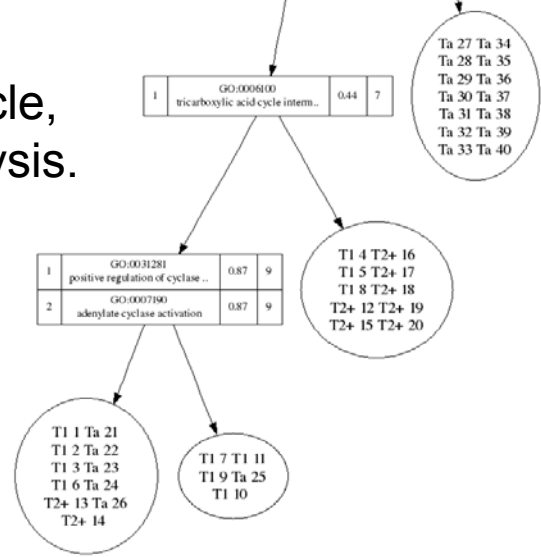
Dataset II

Three types of bladder cancer
Number of samples: 40
Number of genes: 3036
TYPE: clinical classification of tumors (T1, T2+, Ta)

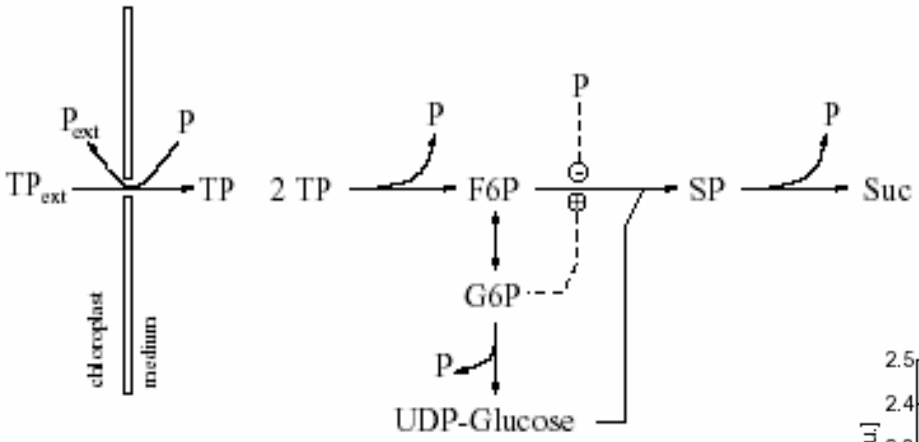
Dyrskjot et al (2003)
Identifying distinct classes of bladder carcinoma using microarrays.
Nat Genetics 33(1):90-6.

… However, because cancer cells have a defective Krebs cycle, they must derive almost all of their energy needs from Glycolysis.

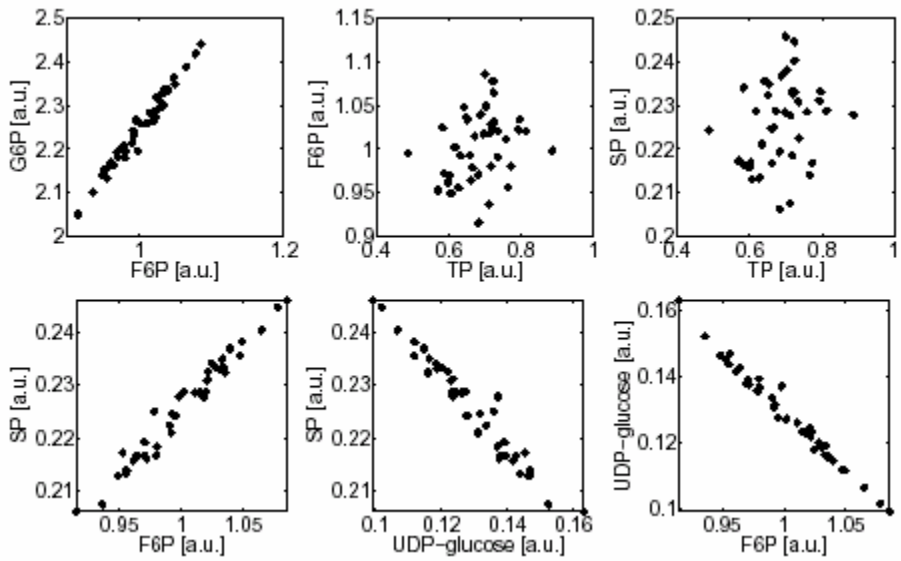
Rank	GO Term GO ID	S	Size
1	GO:0044275 cellular carbohydrate catabolism	0.38	40
2	GO:0046164 alcohol catabolism	0.33	33
3	GO:0019320 hexose catabolism	0.33	33
4	GO:0006096 glycolysis	0.33	26
5	GO:0006445 regulation of translation	0.33	33
6	GO:0008584 male gonad development	0.33	5
7	GO:0007028 cytoplasm organization and biog..	0.33	20
8	GO:0009889 regulation of biosynthesis	0.33	38
9	GO:0006417 regulation of protein biosynthe..	0.32	37
10	GO:0006007 glucose catabolism	0.32	29
11	GO:0006413 translational initiation	0.32	20
12	GO:0042254 ribosome biogenesis and assembly	0.31	14
13	GO:0044419 interaction between organisms	0.30	7
14	GO:0046661 male sex differentiation	0.30	7
16	GO:0009059 macromolecule biosynthesis	0.29	209
17	GO:0019318 hexose metabolism	0.29	50
18	GO:0006412 protein biosynthesis	0.28	194
20	GO:0007548 sex differentiation	0.27	11
21	GO:0006006 glucose metabolism	0.27	37
22	GO:0006826 iron ion transport	0.27	7
25	GO:0043037 translation	0.26	70
26	GO:0000041 transition metal ion transport	0.25	11
28	GO:0051246 regulation of protein metabolism	0.25	57



Correlations: A simple example - Glycolysis



Fluctuations in certain metabolic compounds ....



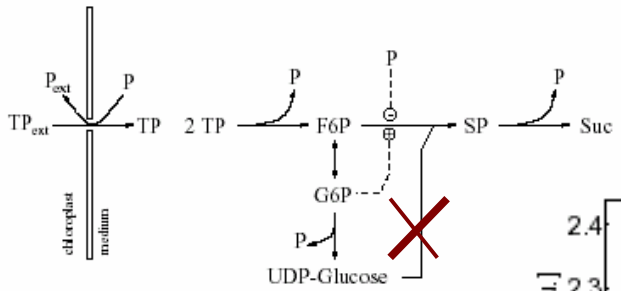
... propagate through the system and induce a specific pattern of correlations.

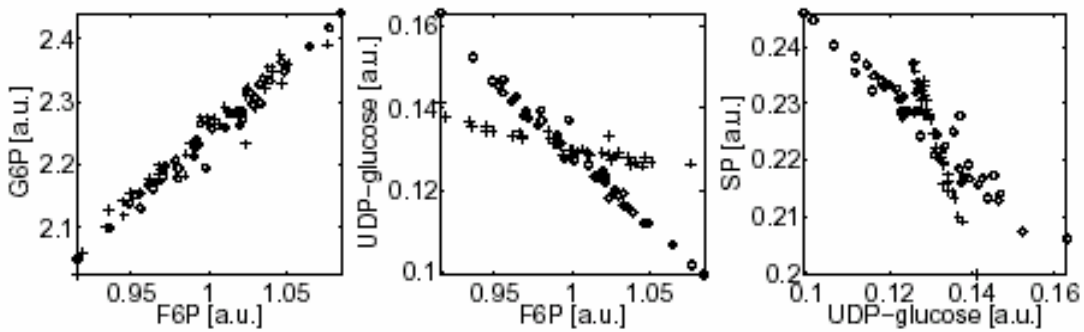
Data: Potato tubers metabolite profiles

Differential Correlations

Predictions:
Changes in 'state' will result in altered correlation patterns.
For instance, source versus sink tissue.
Some correlations will change, others not.
Some correlations are 'hard-wired' into the system.

An Example: Change regulation in the previous model.

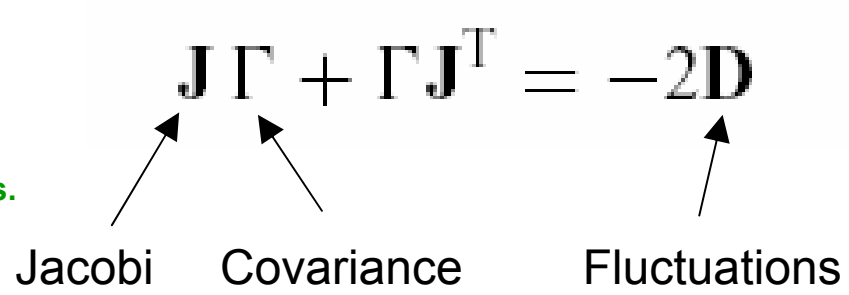




Theoretical Framework

van Kampen, 1992

R Steuer, J Kurths, O Fiehn W Weckwerth (2003)
Interpreting correlations in metabolomic networks.
Biochemical Society Transactions 31(1476-1478).



Linear Approximation:

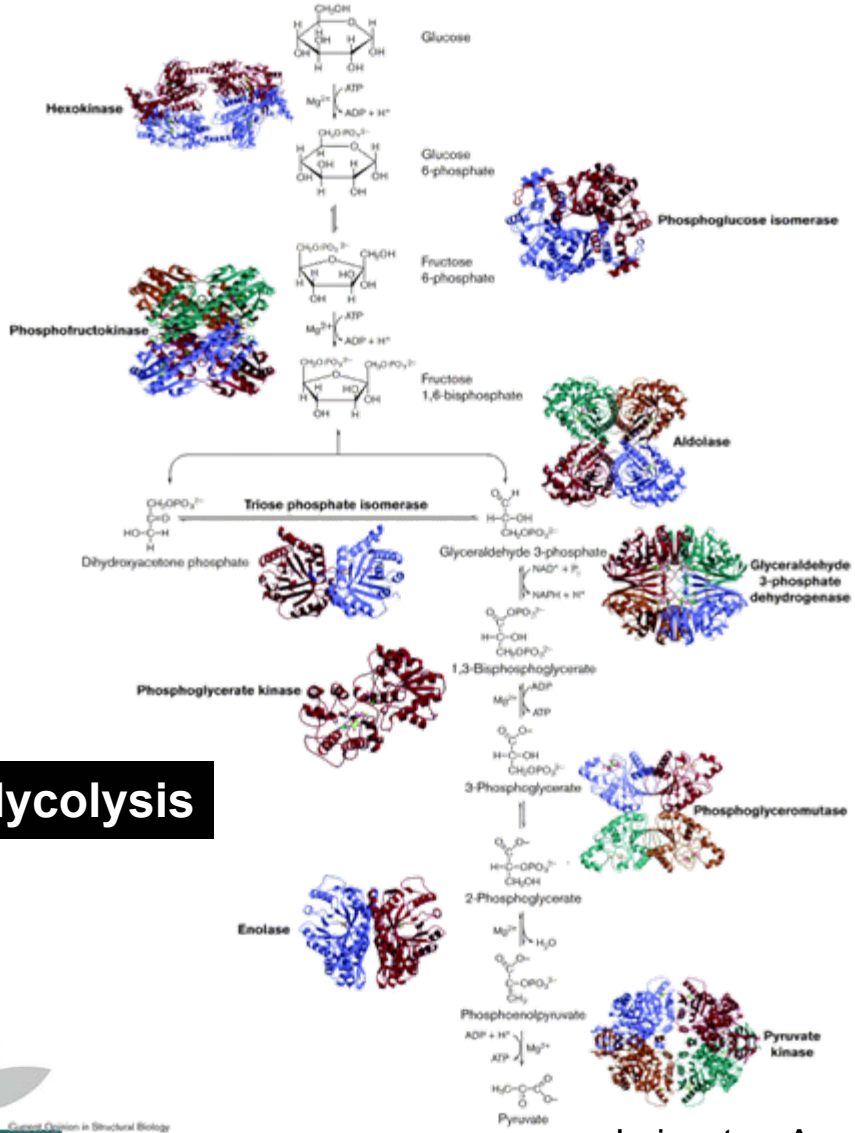
Systematic relationship between the observed correlations and the corresponding dynamical system

Correlations are global properties of the system

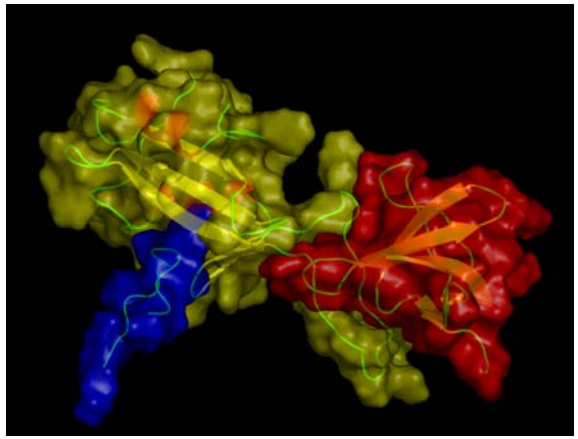
Correlations represent a global fingerprint of the system at a given point in time or in a specific state

R Steuer, T Gross, J Selbig, B Blasius (2006)
Structural Kinetic Modeling of Metabolic Networks.
PNAS 103(11868-11873).

3D structure-based analysis of networks



Glycolysis



Current Opinion in Structural Biology

Universität Potsdam

PROMI

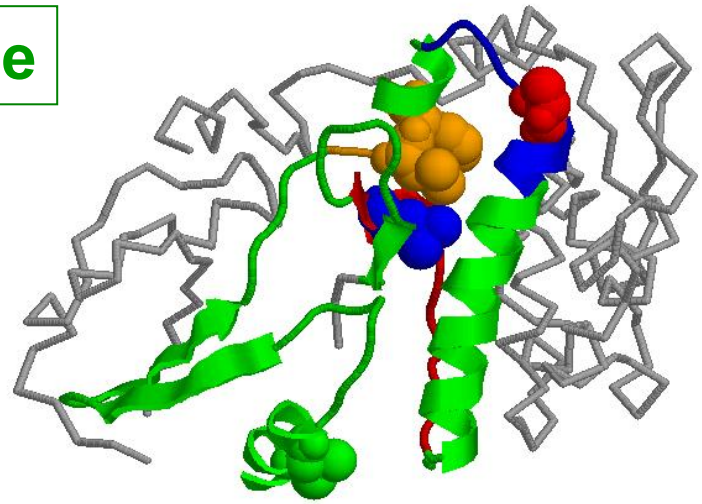
```
1                                                                50
AVDIEDVKREVAIMKHLPKSSSIVTLKEACEDDNAVHLVMELCEGGELFD
LEGKEGSMENEIAVLHKIKHPNIVALDDIYESGGHLYLIMQLVSGGELFD
  | E| | |   | H  K ||IV|L|| |E||| | L|M|L  GGELFD

RIVARGHYTERAAAGVTKTIVEVVQLCHKHGVIHRDLKPENFLFANKKEN
RIVEKGFYTERDASRLIFQVLDAVKYLHDLGIVHRDLKPENLLYYSLDED
RIV||G YTER|A| ||  ||||V|  H| G||HRDLKPEN|L| | |E|

SPLKA
SKIMI
S ||
```

Alignment of the *A. th.* calcium-dependent protein kinase fragment with the homologous fragment of the calmodulin dependent kinase from rat.

http://promi.mpimp-golm.mpg.de



Calmodulin-dependent kinase from rat (PDB enty 1A06). The fragment corresponding to the alignment is coloured green.

	F	**I**	**L**	R	**V**	X
Bacteria	1	**52**	8	1	**35**	1
Mammalian		**18**	**70**			9
Plants	**98**					1

J Hummel, N Keshvari, W Weckwerth, J Selbig (2005)
Species-specific analysis of protein sequence motifs using mutual information.
BMC Bioinformatics 6:164.

Thanks to the group and partners!

Group members
(Past* and current)

*Sven Borngräber
Roman Brunnemann
*Carsten Daub
Pavel Durek
*Toni Goßmann
*Susanne Grell
Sergio Grimbs
*Stefanie Hartmann
*Peter Humburg
Jan Hummel
Peter Krüger
Henning Redestig
*Matthias Scholz
*Sandro Schugk
*Natascha Shevchenko
*Danny Tomuschat
Dirk Walther
Daniel Weicht

Cooperations with AGs

Thomas Altmann
Ally Fernie
*Oliver Fiehn (*Martin Scholz)
Dirk Hincha
Joachim Kopka
Ute Krämer
Victoria Nikiforova (*Petra Birth)
Wolf-Rüdiger Scheible
Mark Stitt (Jan Hannemann)
Michael Udvardi
Wolfram Weckwerth
Lothar Willmitzer (Jan Lisec)
…

Potsdam University

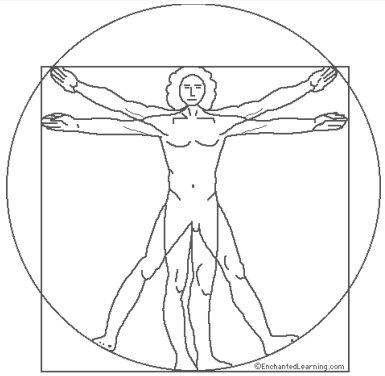
Marco Ende
*André Flöter
Andy Fohrmann
Manuela Hische
Sabine Kern
Dirk Repsilber
Wolfram Stacklies
Matthias Steinfath
Ralf Steuer

Computational Tools for the Optimization of Antiretroviral Drug Therapies

Joachim Selbig

University of Potsdam
and
Max Planck Institute of Molecular Plant Physiology

HIV infection: the players



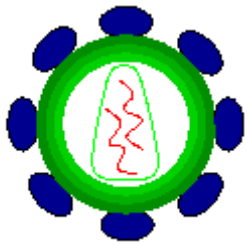
Host: human

infection

immune response

toxicity

degradation, elimination



drug resistance

inhibition

Infectious agent: HIV

Therapy: antiretrovirals

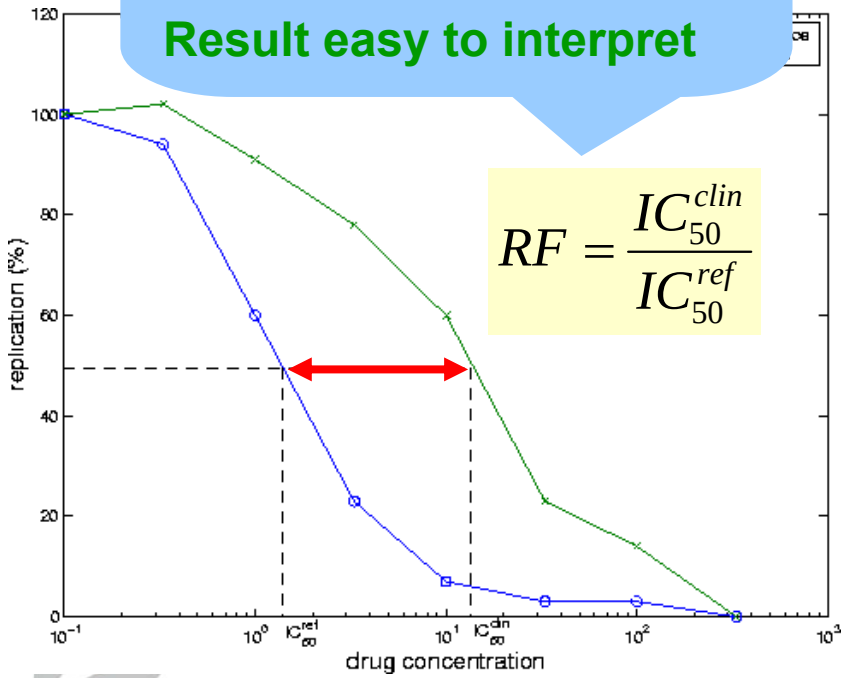
Resistance Testing

- Phenotypic Resistance Testing
  - How well does the virus replicate *in vitro* in the

Labour intensive

Takes 4-8 weeks, costs ~1500 US$

Result easy to interpret



$$RF = \frac{IC_{50}^{clin}}{IC_{50}^{ref}}$$

- Genotypic Resistance Testing
  - Which (resistance-associated) mutations do the viral drug targets harbor?
  - Cycle-sequencing assay

```
PR: PQITLWQRPLVTVKIGGQL…
RT: PISPIKTVPVRLKPGMDGP…
…   …
```

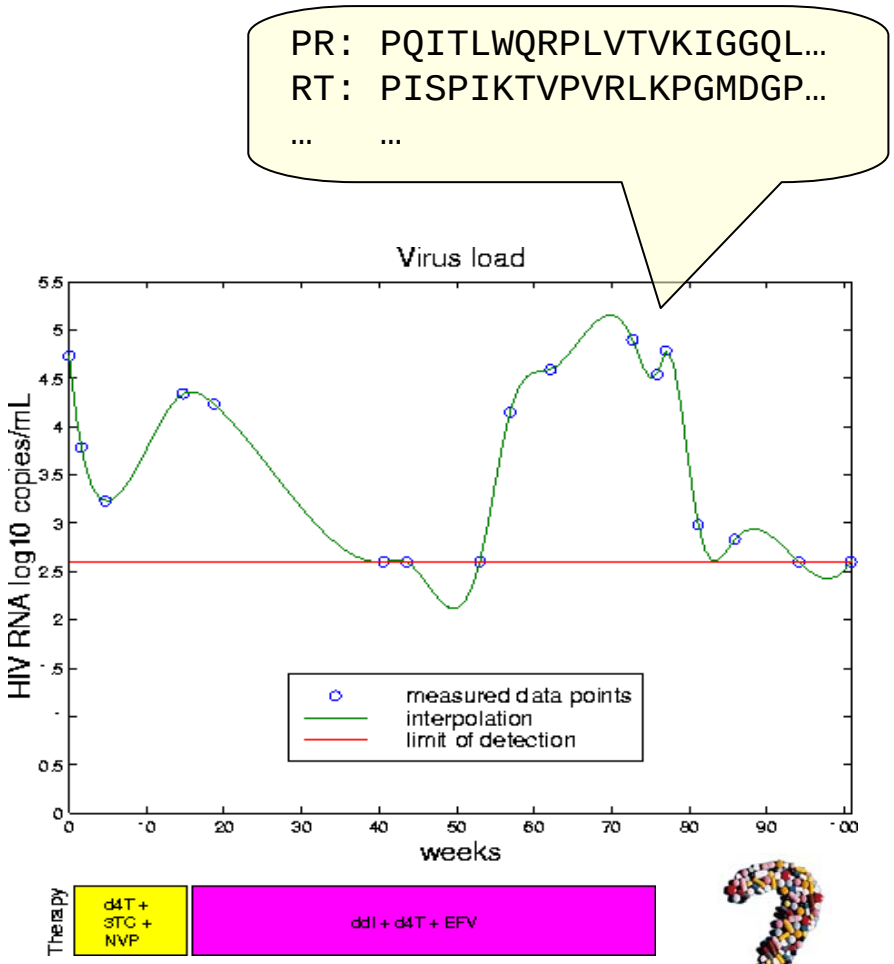
Standardized kits

Takes a few days, cheaper: ~300 US$

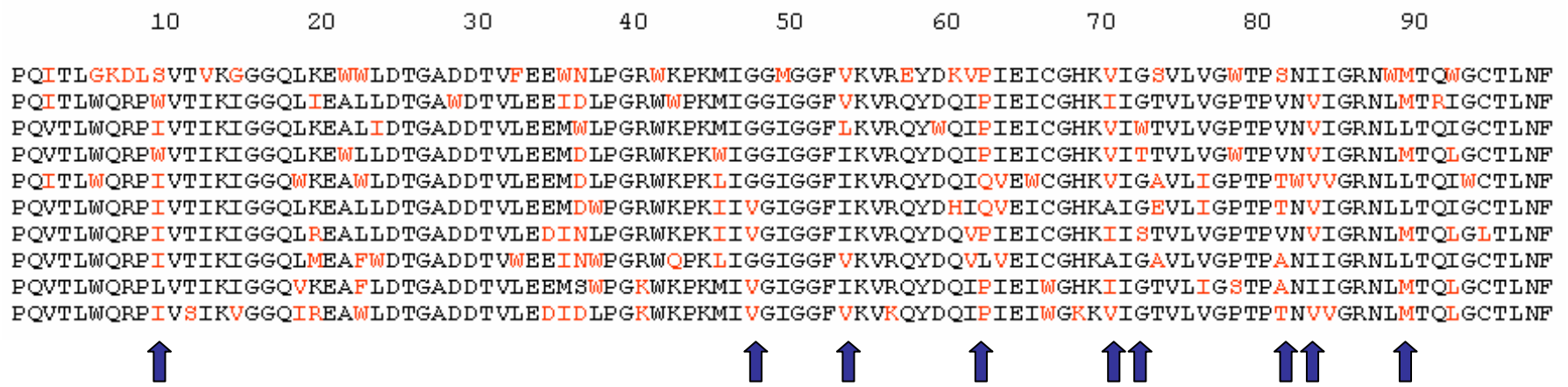
Interpretation challenging

Challenges

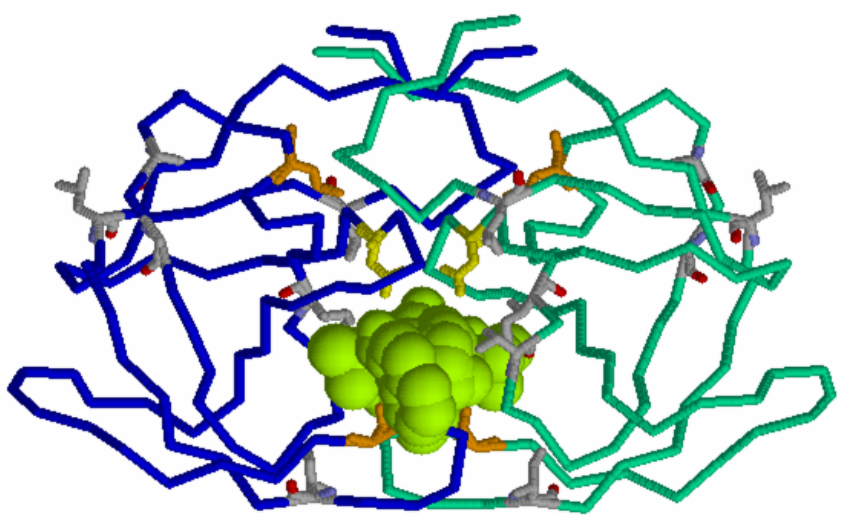
1. Data integration for patient monitoring and collaborative research.
2. What is the relationship between genetic changes in the drug targets and phenotypic drug resistance?
3. Given the viral genotype, what is the **optimal** choice of a **drug combination** (after therapy failure)?
4. How will the virus population react to the selective pressure of a certain drug (combination)?



Resistance-associated **mutation patterns**

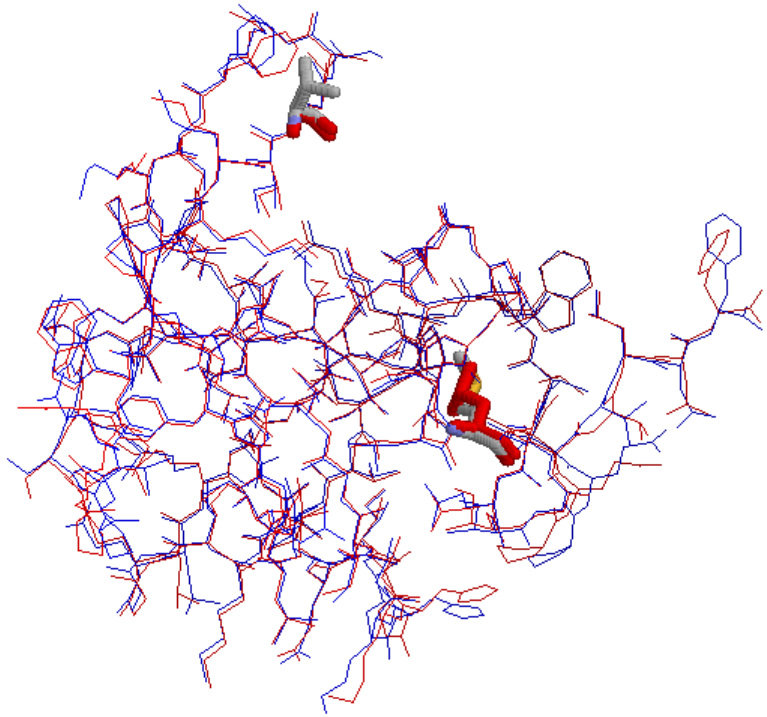


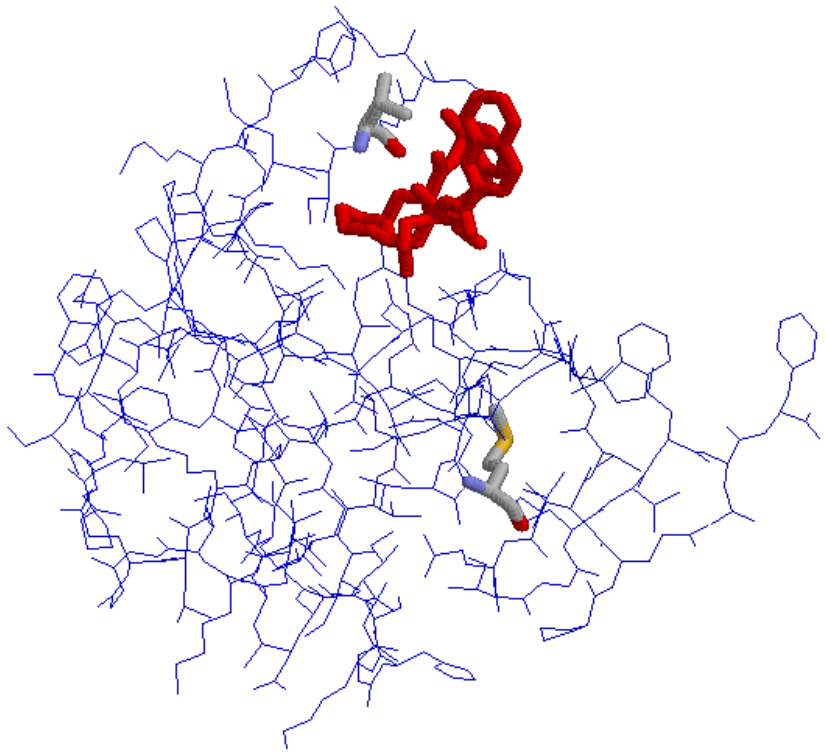
HIV-1 PR – SQV complex



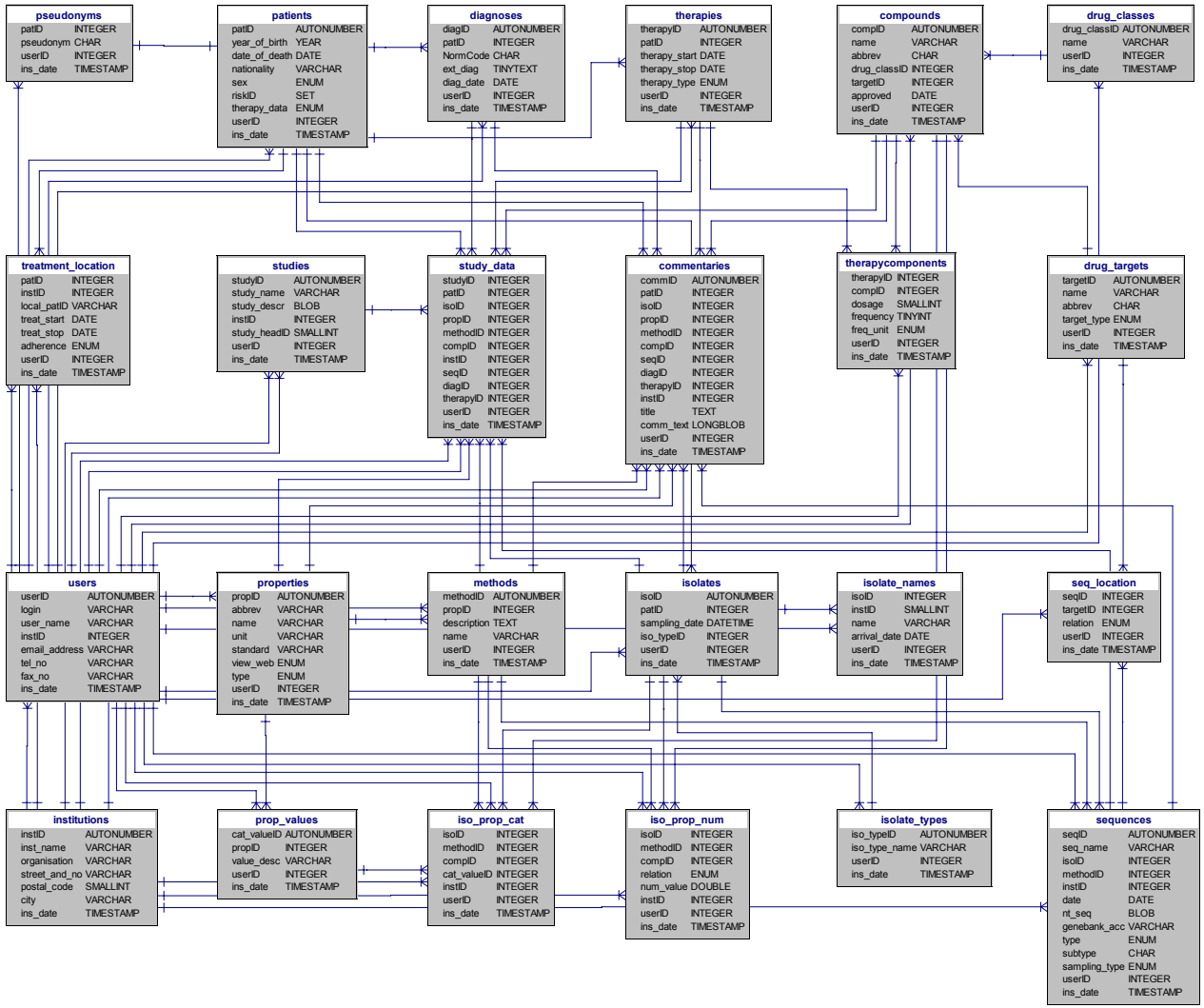
Double mutant PR – SQV complex

- G48V + L90M: ~ 400-fold increase in K_i value



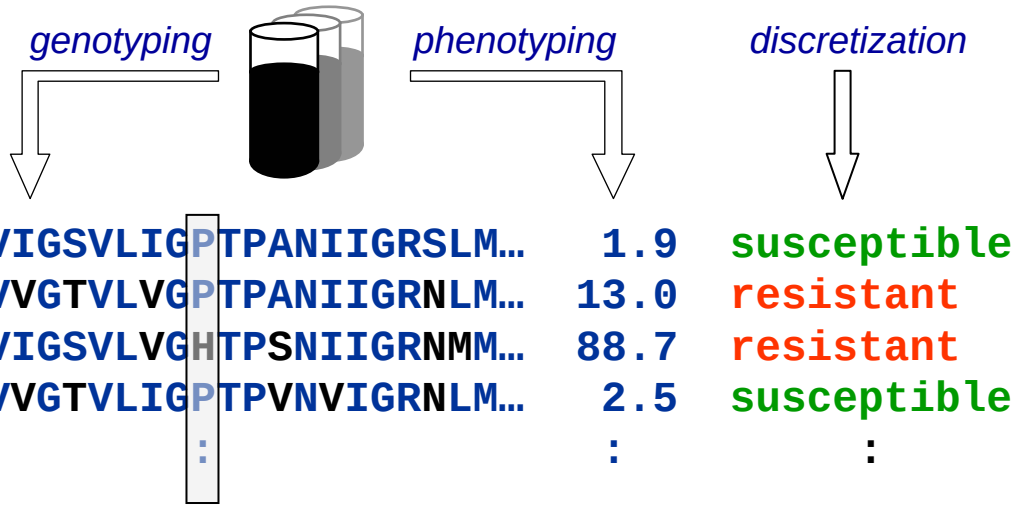


Arevir database



Genotype – phenotype

- 700 matched genotype-phenotype pairs:

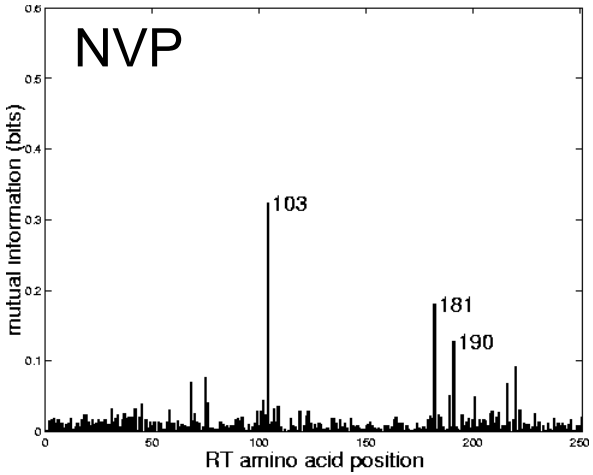


$X=(X_1, \ldots, X_i, \ldots, X_n)$ $\quad Y$

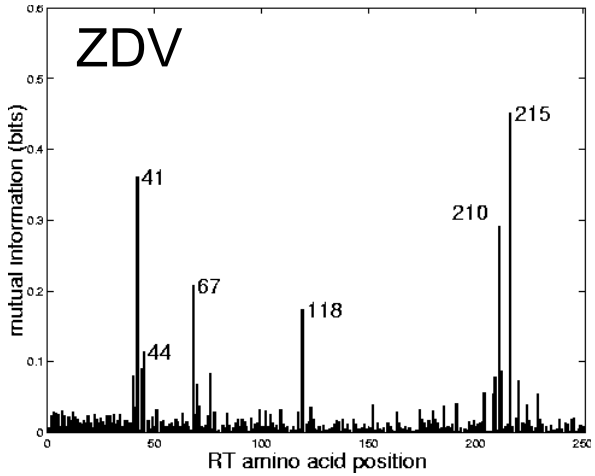
- Density estimation: $\mathrm{P}(X) = ?$
- Feature Selection: How relevant is X_i for Y ?
- Prediction: $\mathrm{P}(Y|X) = ?$
  - Regression $Y \in \mathbf{R}$
  - Classification $Y \in \{\mathbf{sus}, \mathbf{res}\}$

Information profiles

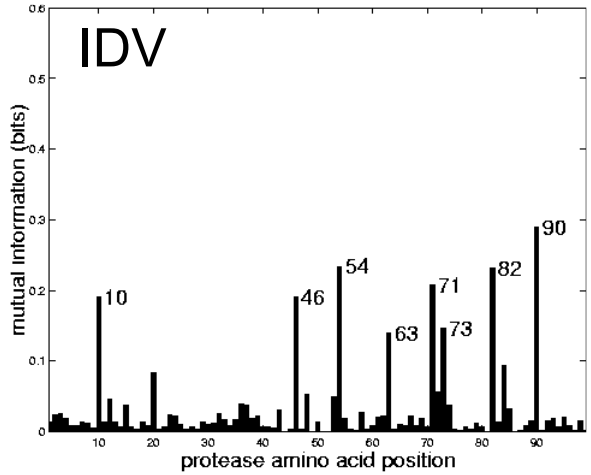
Non-nucleoside RTI



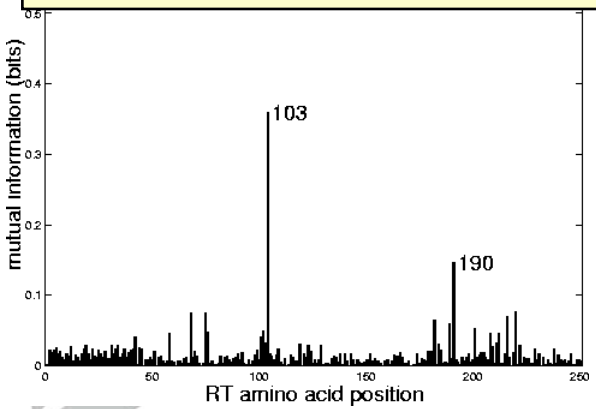
Nucleoside RTI

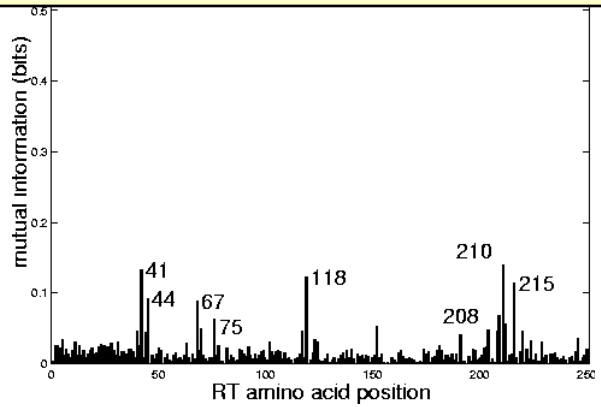


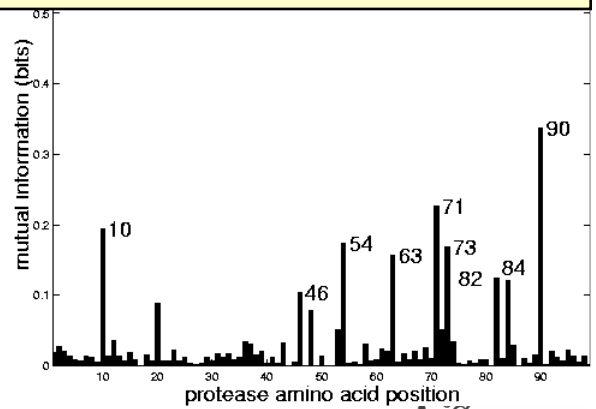
Protease inhibitors



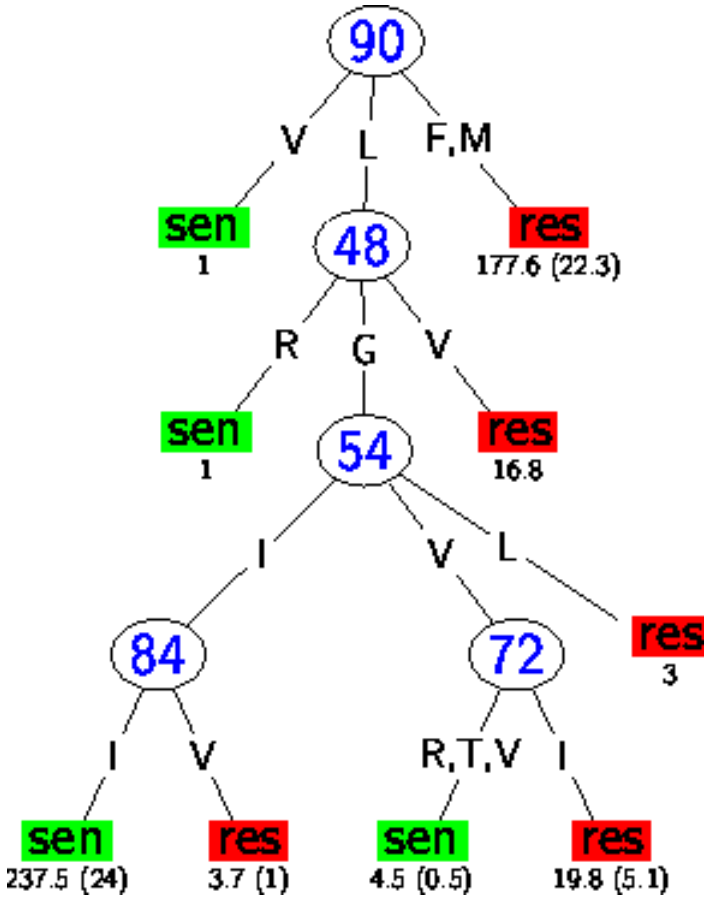
High degree of cross-resistance within drug classes!



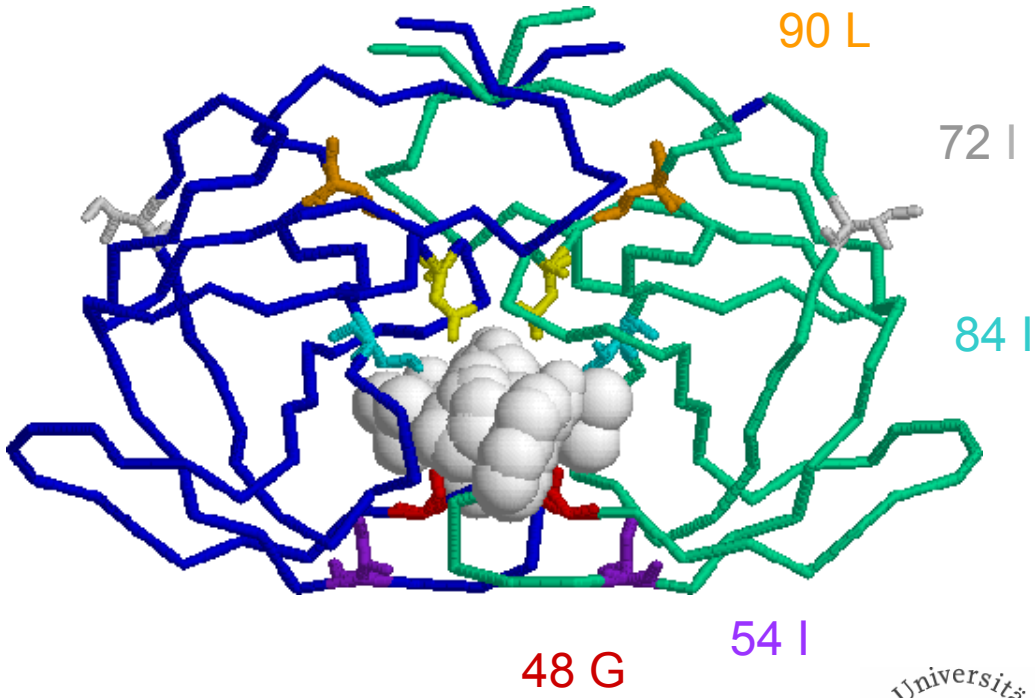




SQV decision tree

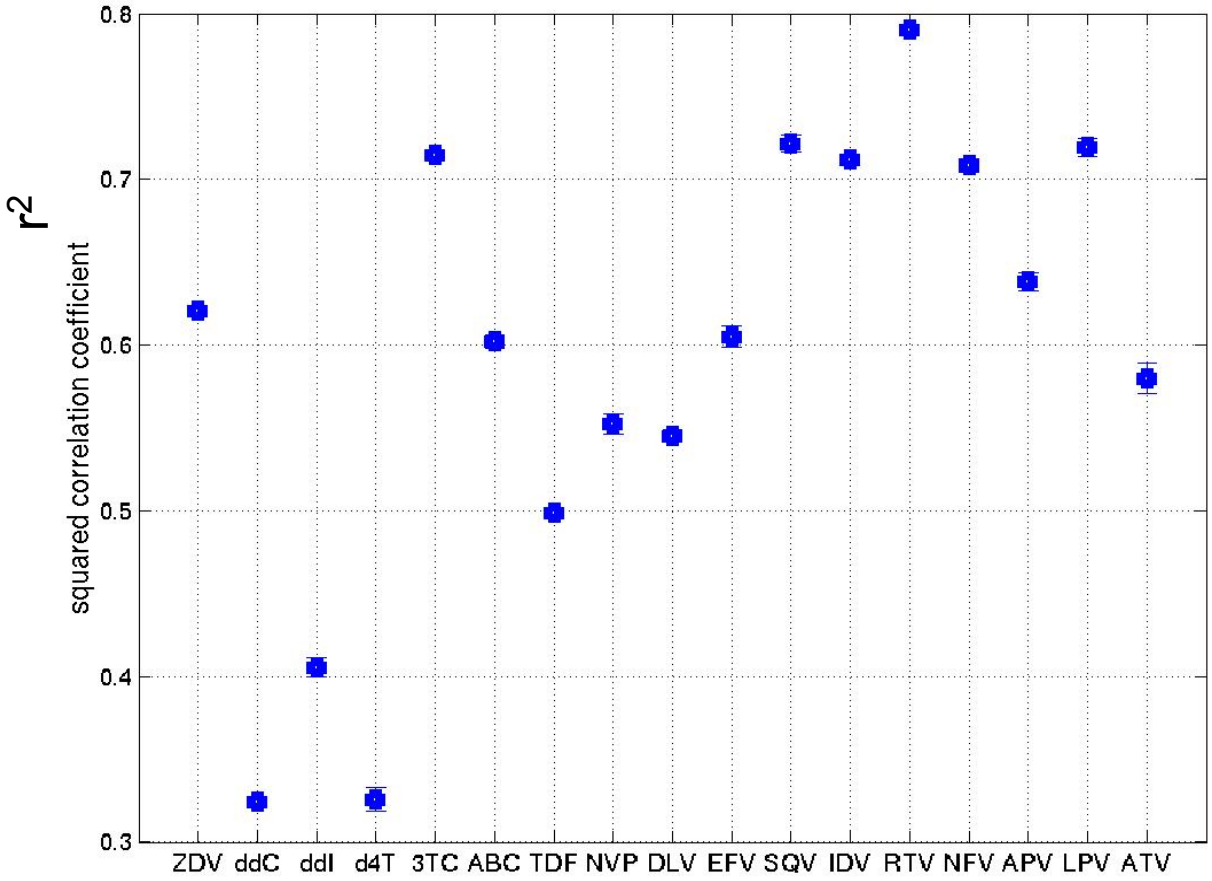


9 known resistance mutations:
10, **48**, **54**, 63, 71, 73, 82, **84**, **90**



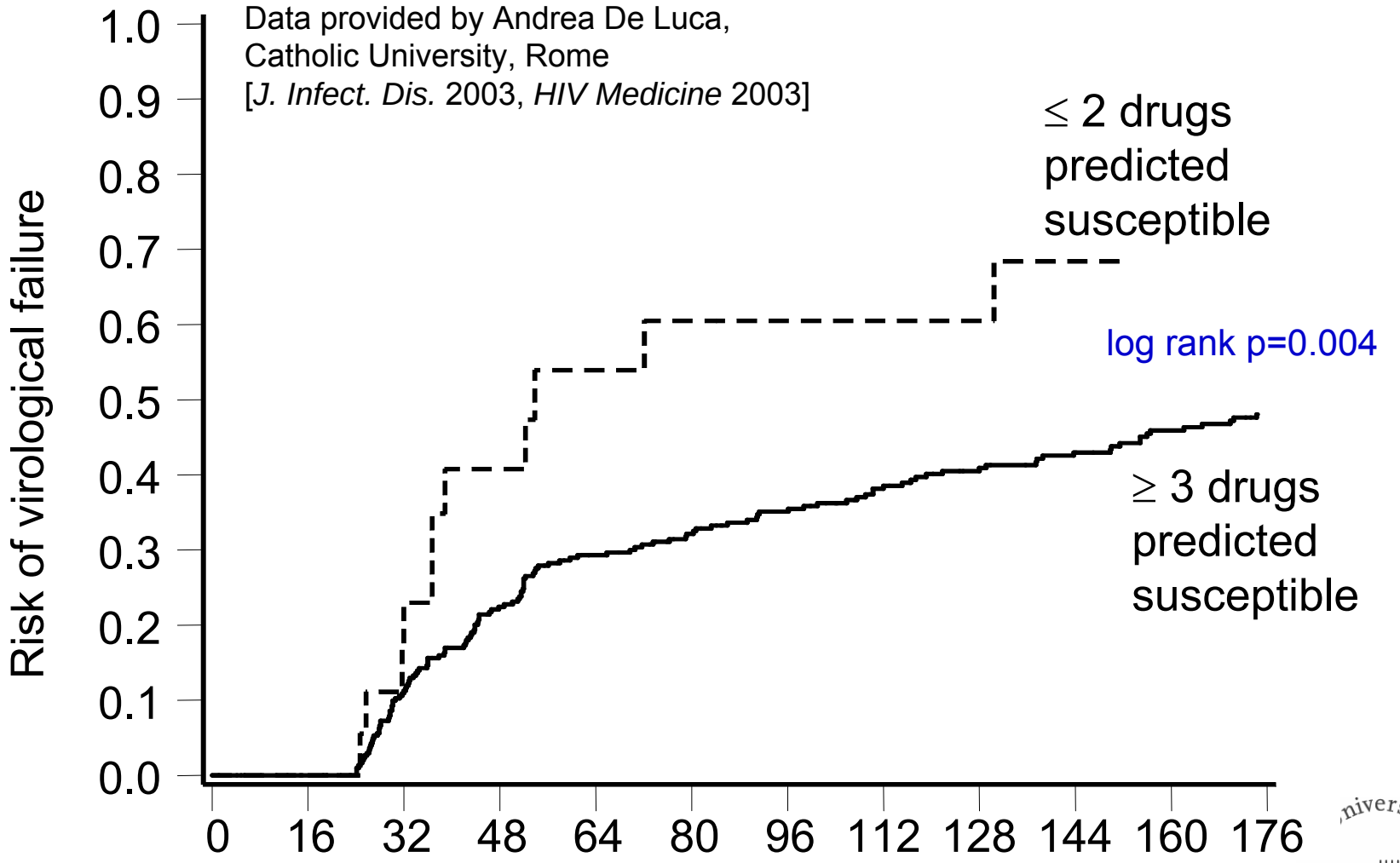
72R/T/V may reverse 54V mediated SQV resistance

Generalization error



10-fold cross-validation estimates

Clinical relevance



N = 332 untreated patients

Leicester, August 24, 2006

Netscape: geno2pheno – version 2.1

File Edit View Go Communicator Help

last update: February 2003

www.genafor.org

Description of output

Statistics

References

Contact

Phenotype Prediction

At ambiguous sequence positions (due to a mixed virus population) the resistance associated mutation has been assumed if present.

Drug	Cutoff	Decision tree classification[1] [confidence factor]	SVM classification[2]	Predicted fold-resistance (res	z-score (number of standard deviations	Probability score (likelihood of belonging
ZDV	8.5	resistant [0.90]	resistant			
ddC	2.5	susceptible [0.81]	susceptible			
ddI	2.5	resistant [0.52]	resistant			
d4T	2.5	resistant [0.74]	susceptible			
3TC	8.5	susceptible [0.80]	resistant			
ABC	2.5	resistant [0.89]	resistant			
TDF	2.5	resistant [0.76]	resistant			
NVP	8.5	resistant [0.74]	resistant			
DLV	8.5	resistant [0.74]	resistant			
EFV	8.5	susceptible [0.84]	susceptible			
SQV	3.5	resistant [0.88]	resistant			
IDV	3.5	resistant [0.87]	resistant			
RTV	3.5	resistant [0.89]	resistant			
NFV	3.5	resistant [0.93]	resistant			
APV	3.5	susceptible [0.92]	susceptible			
LPV	3.5	susceptible [0.86]	resistant			
ATV	3.5	resistant [0.83]	resistant			

1: based on C5.0, Ru
2: based on LIBSVM, Copyright (c) 2

[Protease | Reverse transcriptase

backgrounds)

6. Action:

align & predict (takes a few seconds) Reset Load a

You will make prediction no. 12330.Service started December 1

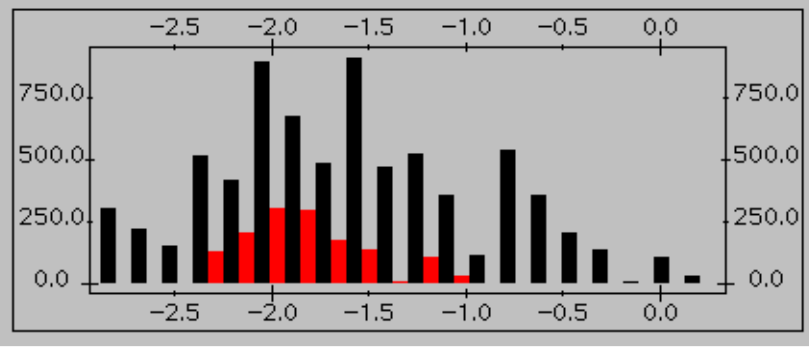
100%

Length <= 6

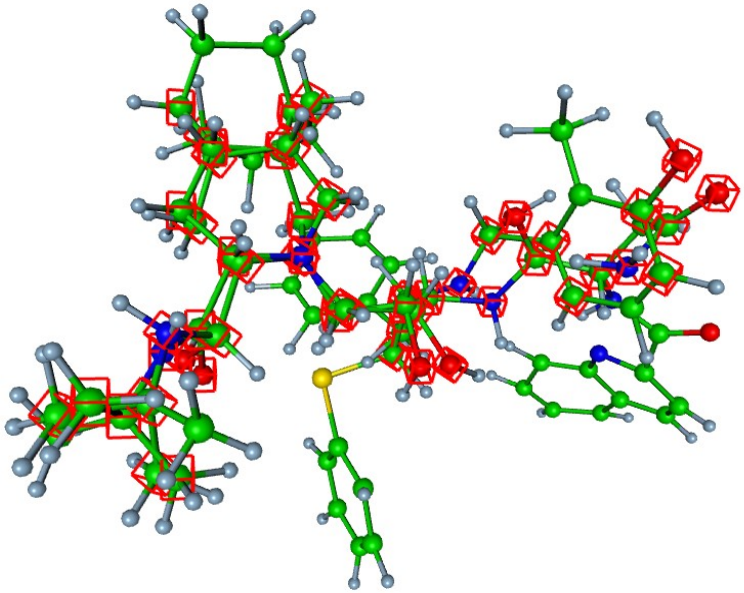
#NRTI:		#NNRTI:		#PI:	
>= -	<= -	>= -	<= -	>= -	<= -
#ZDV= -		#NVP= -		#IDV= -	
#ddC= -		#DLV= -		#RTV= -	
#ddI= -		#EFV= 1		#SQV= -	
#d4T= -				#NFV= -	
#3TC= -		Reset		#APV= -	
#ABC= -				#LPV= 0	
#TDF= -		Compute		#ATV= -	

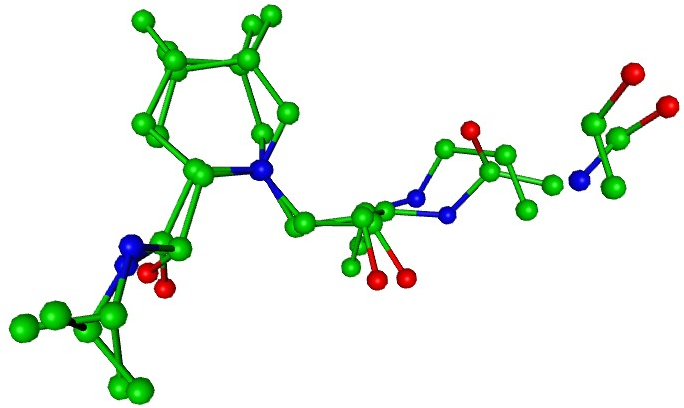
Best 20 therapies:

3TC EFV APV −2.3189607
3TC ABC EFV APV −2.3189607
d4T 3TC EFV APV −2.3189607
3TC EFV IDV APV −2.3189607



Molecular similarity of SQV and NFV

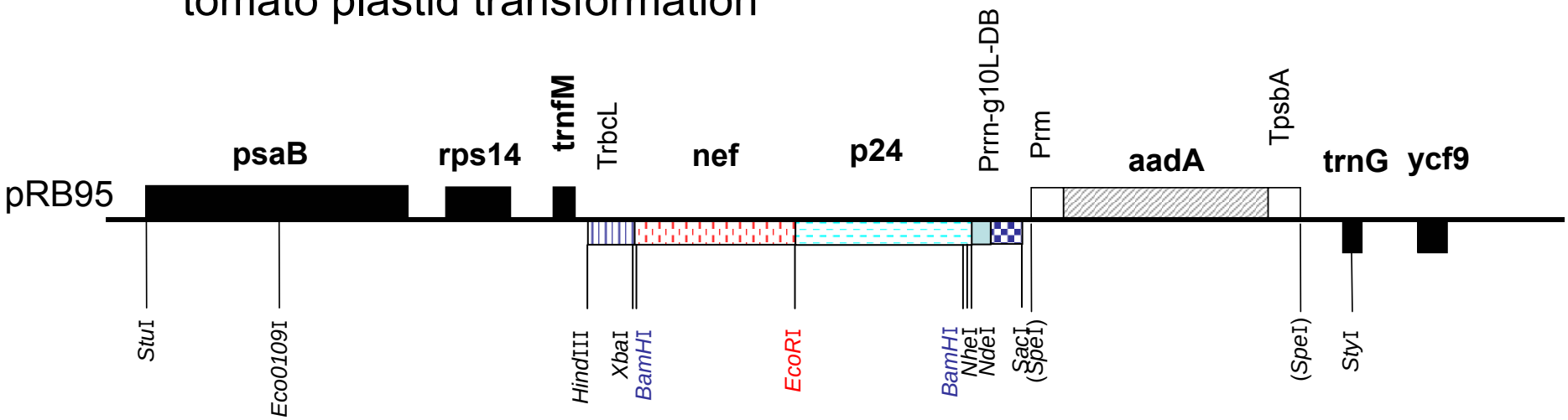




	IDV	RTV	SQV	NFV	APV	LPV	ATV
IDV		0,9204	0,7053	0,8742	0,6296	0,8483	0,9191
RTV			0,7242	0,8049	0,6971	0,8757	0,8976
SQV				0,7641	0,5273	0,6199	0,7575
NFV					0,6093	0,7308	0,848
APV						0,7052	0,7193
LPV							0,9115

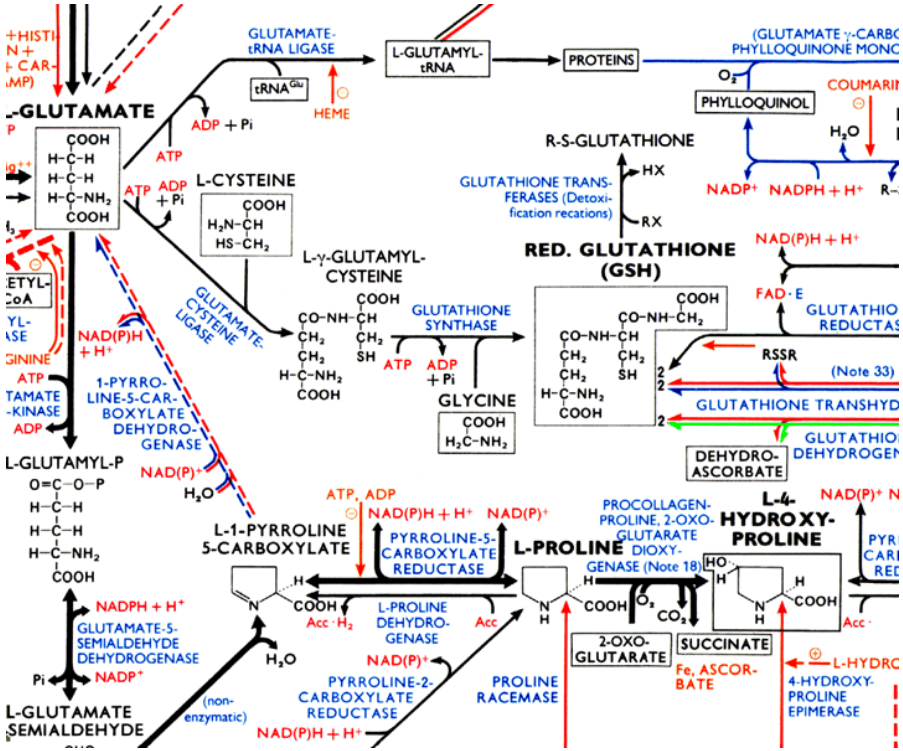
Co-expression of HIV nef and p24

Towards an oral HIV vaccine: Expression of nef and p24 by tomato plastid transformation



p24-nef fusion protein: N-terminal p24, C-terminal nef

Nutritional issues related to HIV



Glutathione deficiency contributes to oxidative stress, **which plays a key role in** aging and the pathogenesis of many diseases such as **HIV infection**.

Dietary plant polyphenols (**flavonoids**) **modulate expression of** an important enzyme in both cellular antioxidant defenses and detoxification of xenobiotics, i.e, **γ-glutamylcysteine synthetase**. **This enzyme** is rate **limiting in the synthesis of** the most important endogenous antioxidant in cells, **glutathione**.

References

N Beerenwinkel, M Däumer, T Sing, J Rahnenführer, T Lengauer, J Selbig, D Hoffmann, R Kaiser (2005)
Estimating HIV evolutionary pathways and the genetic barrier to drug resistance.
Journal of Infectious Diseases, in press, 2005.

N Beerenwinkel, J Rahnenführer, R Kaiser, D Hoffmann, J Selbig, T Lengauer (2005)
Mtreemix: a software package for learning and using mixture models of mutagenetic trees.
Bioinformatics Advance Access first published online on January 18, 2005

N Beerenwinkel, J Rahnenführer, M Däumer, D Hoffmann, R Kaiser, J Selbig, T Lengauer (2004)
Learning Multiple Evolutionary Pathways from Cross-sectional Data.
RECOMB, San Diego, March 27-31 (36-44).

K Wolf, H Walter, N Beerenwinkel, W Keulen, R Kaiser, D Hoffmann, T Lengauer, J Selbig, A-M Vandamme, K Korn, and B Schmidt (2003)
The drug resistance profile of Tenofovir: Resistance and resensitization.
Antimicrobial Agents and Chemotherapy 47(3478-3484).

N Beerenwinkel, M Däumer, M Oette, K Korn, D Hoffmann, R Kaiser, T Lengauer, J Selbig, and H Walter (2003)
Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes.
Nucleic Acids Research 31(3850-3855).

N Beerenwinkel, T Lengauer, M Däumer, R Kaiser, H Walter, K Korn, D Hoffmann, and J Selbig (2003)
Methods for optimizing antiviral combination therapies.
Bioinformatics 19(i18-i25).

N Beerenwinkel, B Schmidt, H Walter, R Kaiser, T Lengauer, D Hoffmann, K Korn, and J Selbig (2002)
Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype.
PNAS 99(8271-8276).

geno2pheno

http://genafor.org

Acknowledgements

Niko Beerenwinkel	University of California
Daniel Hoffmann	CAESAR, Bonn
Rolf Kaiser	Institute of Virology, University of Cologne
Martin Däumer	Institute of Virology,
Saleta Sierra-Aragon	University of Cologne
Tobias Nolden	
Barbara Schmidt	Institute of Clinical and Molecular Virology,
Hauke Walter	German National Reference Center for Retroviruses,
Klaus Korn	University of Erlangen-Nürnberg
Jürgen Klein	Fraunhofer Institute for Algorithms and Scientific Computing,
Eberhard Schrüfer	Sankt Augustin
Mark Oette	Dept. of Gastroenterology, University of Düsseldorf
Gerd Fätkenheuer	Dept. of Internal Medicine I, University of Cologne
Jürgen Rockstroh	Dept. of Internal Medicine I, University of Bonn
Thomas Berg	Medical Laboratory, Berlin
Patrick Braun	PZB, Aachen
Thomas Lengauer	MPI für Informatik, Saarbrücken
Jörg Rahnenführer	
Jochen Maydt	
Tobias Sing	
Frank Cordes	ZIB, Berlin
Marcus Weber	
Daniel Baum	

MPI-MP
Mark Stitt
Lothar Willmitzer
Ralf Bock
Oliver Fiehn (Martin Scholz)
Rainer Höfgen (Petra Birth)
Joachim Kopka
Ute Krämer
Mark Stitt (Henning Rdeistig)
Michael Udvardi
Wolfram Weckwerth
Sven Borngräber
*Carsten Daub
Jan Hannemann
*Stefanie Hartmann
Peter Humburg
Jan Hummel
Peter Krüger
Matthias Scholz
*Natascha Shevchenko
Danny Tomuschat
Daniel Weicht

UP
Martin Steup
Thomas Altmann
Bernd Müller-Röber
Torsten Schaub
Jürgen Kurths
Matthias Steinfath
André Flöter
Ralf Steuer

DIfE
Andreas Pfeiffer
Matthias Möhlig

DFG (Priority program *Informatics Methods for the Analysis and Interpretation of Large Genomic Data Sets*)

Leicester, August 24, 2006