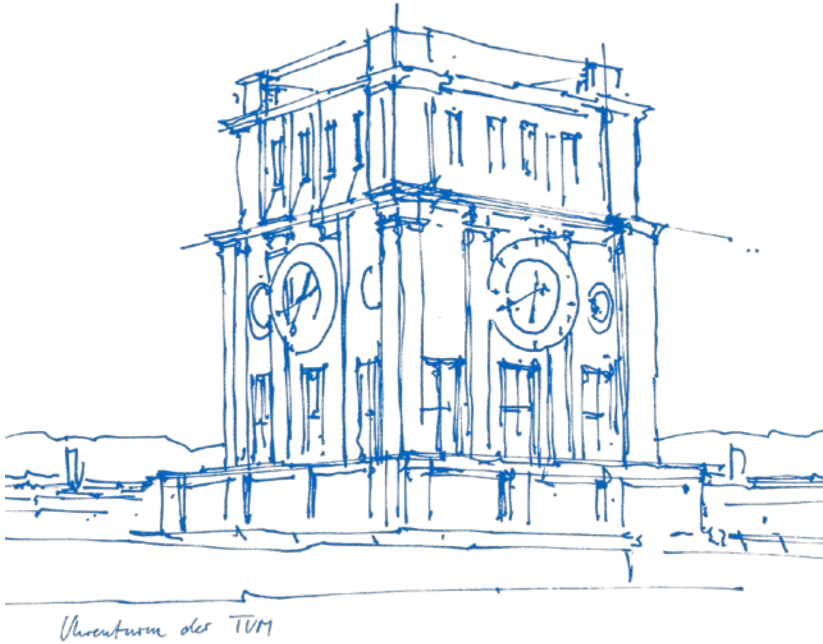
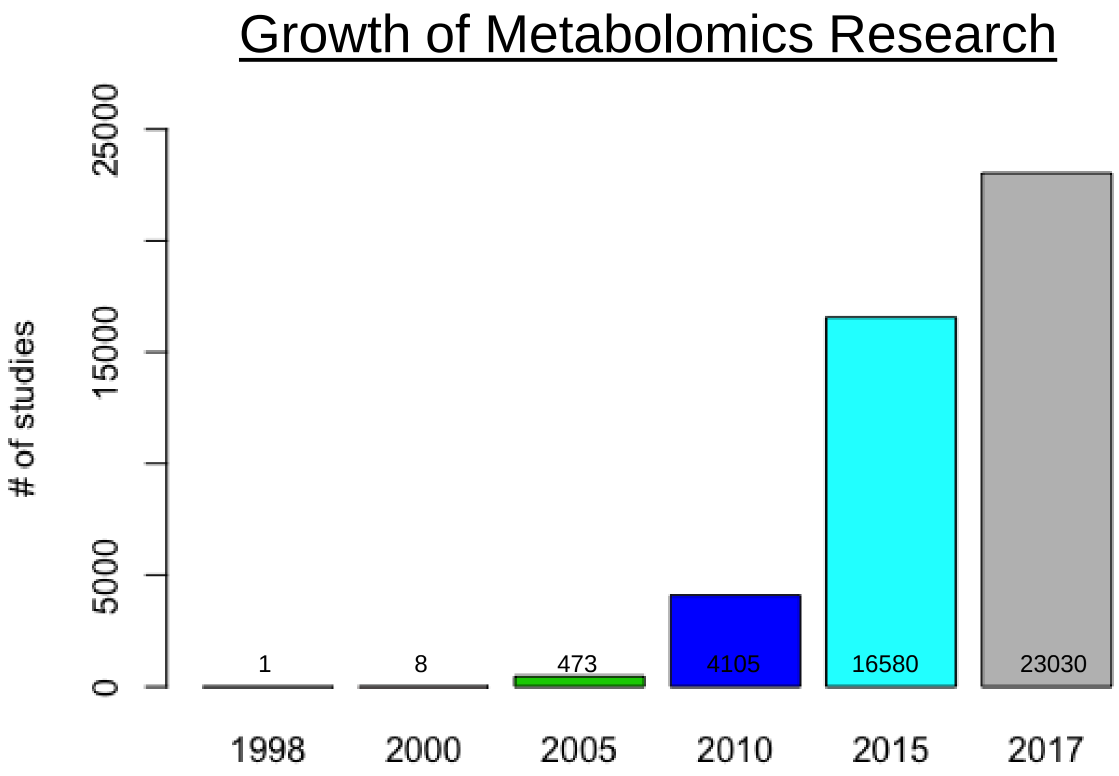
Embracing the Uncorrelated:

Dynamic Time Warping for the Exploration of High Dimensional Time-Series Metabolomics Data from the HuMeT Study

Aaron Novikoff
Master student - Technische Universität München



Metabolomics

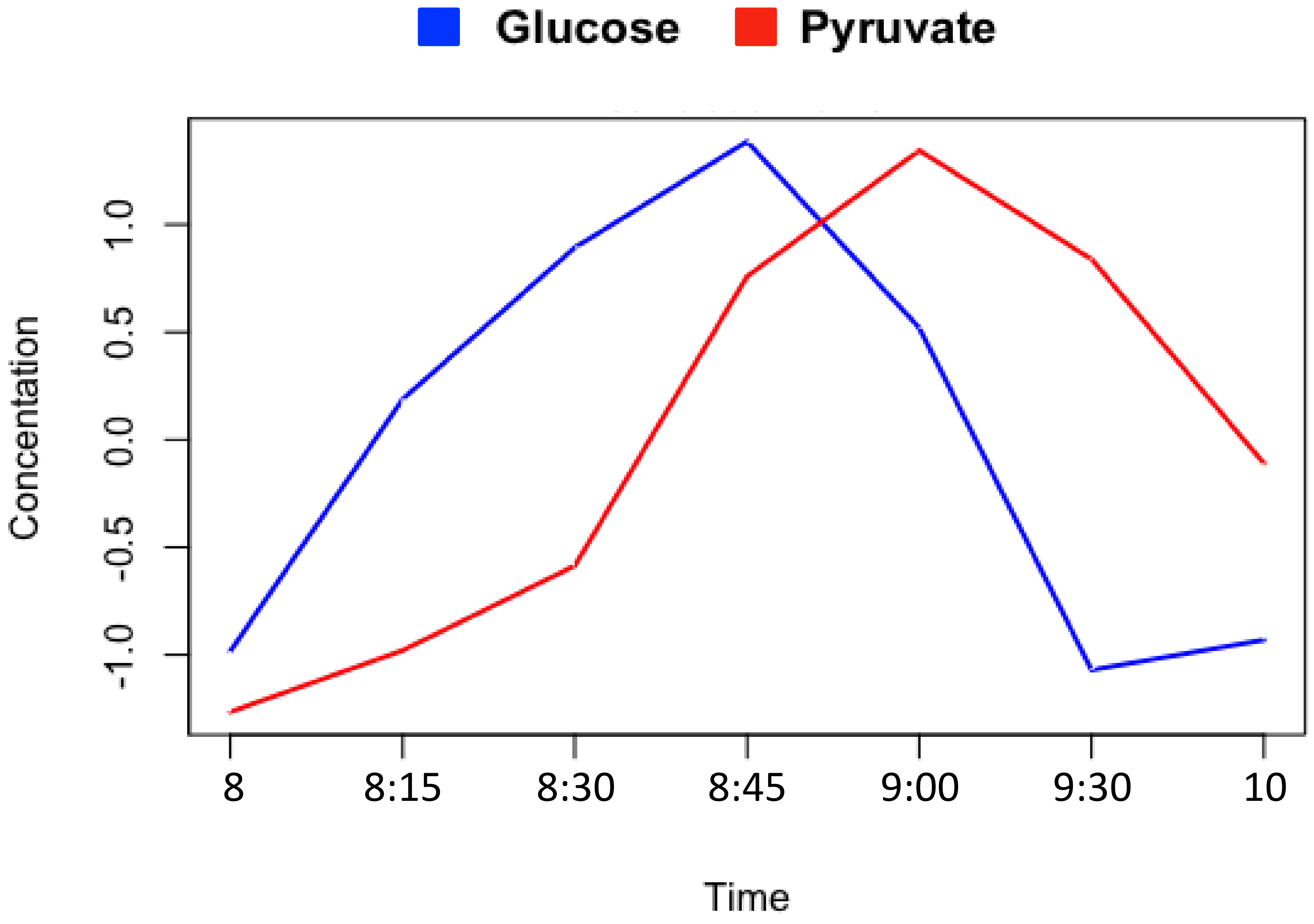


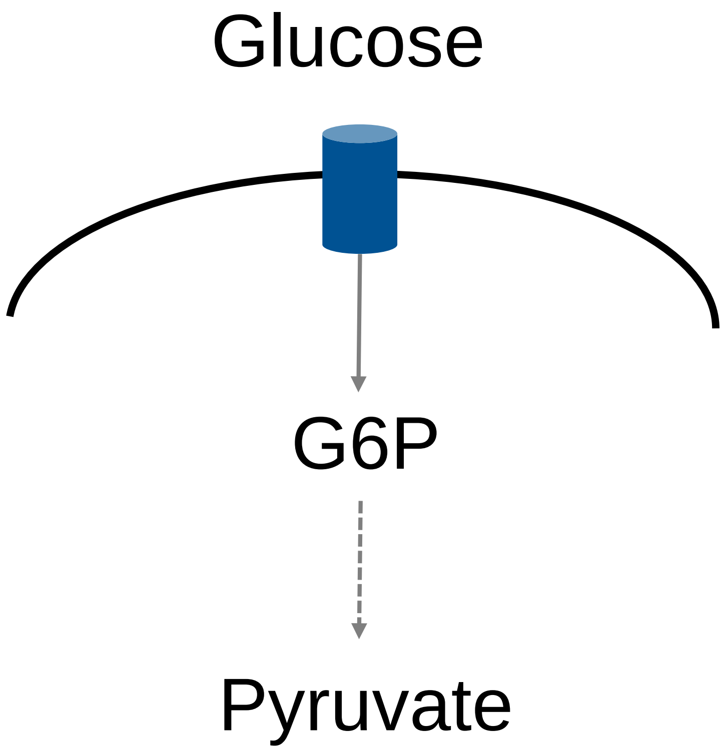
Major Contributors:

- Epidemiology
- Pharmacometabolomics
- Oncology (cancer)
- Lung and Cardiovascular

Time-Series metabolomics

The Fundamental Issue





Correlation: 0.26

Primary Question

How do we identify latent trend line shape similarities?

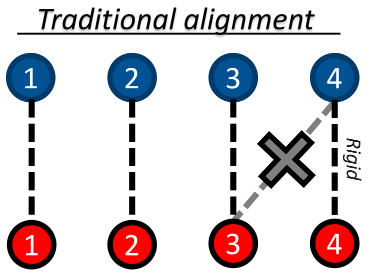
By first understanding the fundamentals

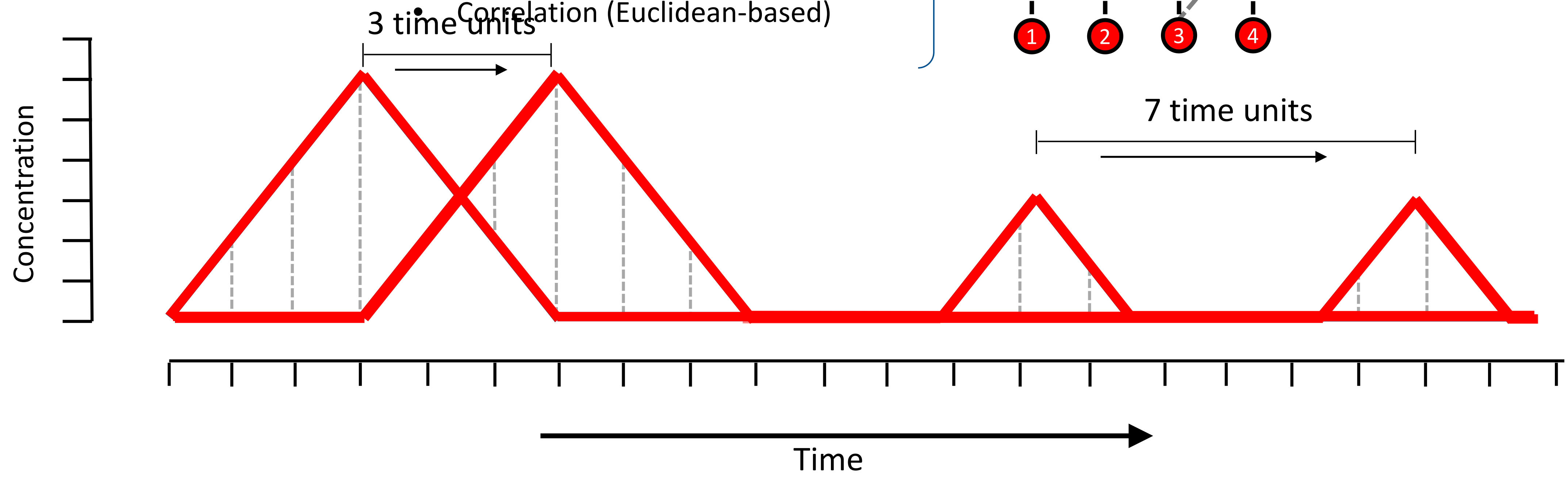
Key Points:

- Non-uniform lag
- Traditional distance measures rely on rigid pairwise alignment

Methods using rigid alignment

- Distance (Euclidean, Manhattan)
- Correlation (Euclidean-based)





Metabolite #1 Metabolite #2

Key Points:

- Non-uniform lag
- Traditional distance measures rely on rigid pairwise alignment

What is Dynamic Time Warping?

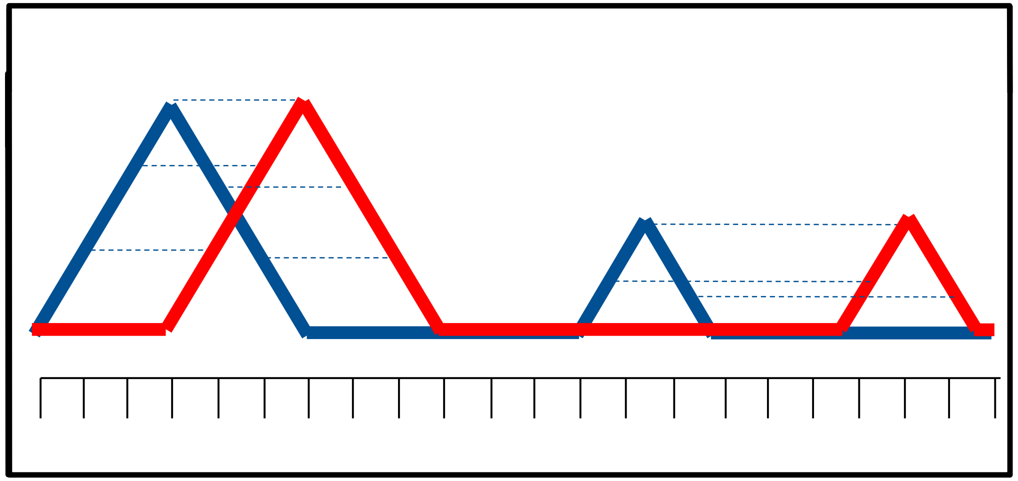
- DTW identifies the alignment producing the lowest distance value
- Global and local alignment restrictions are possible

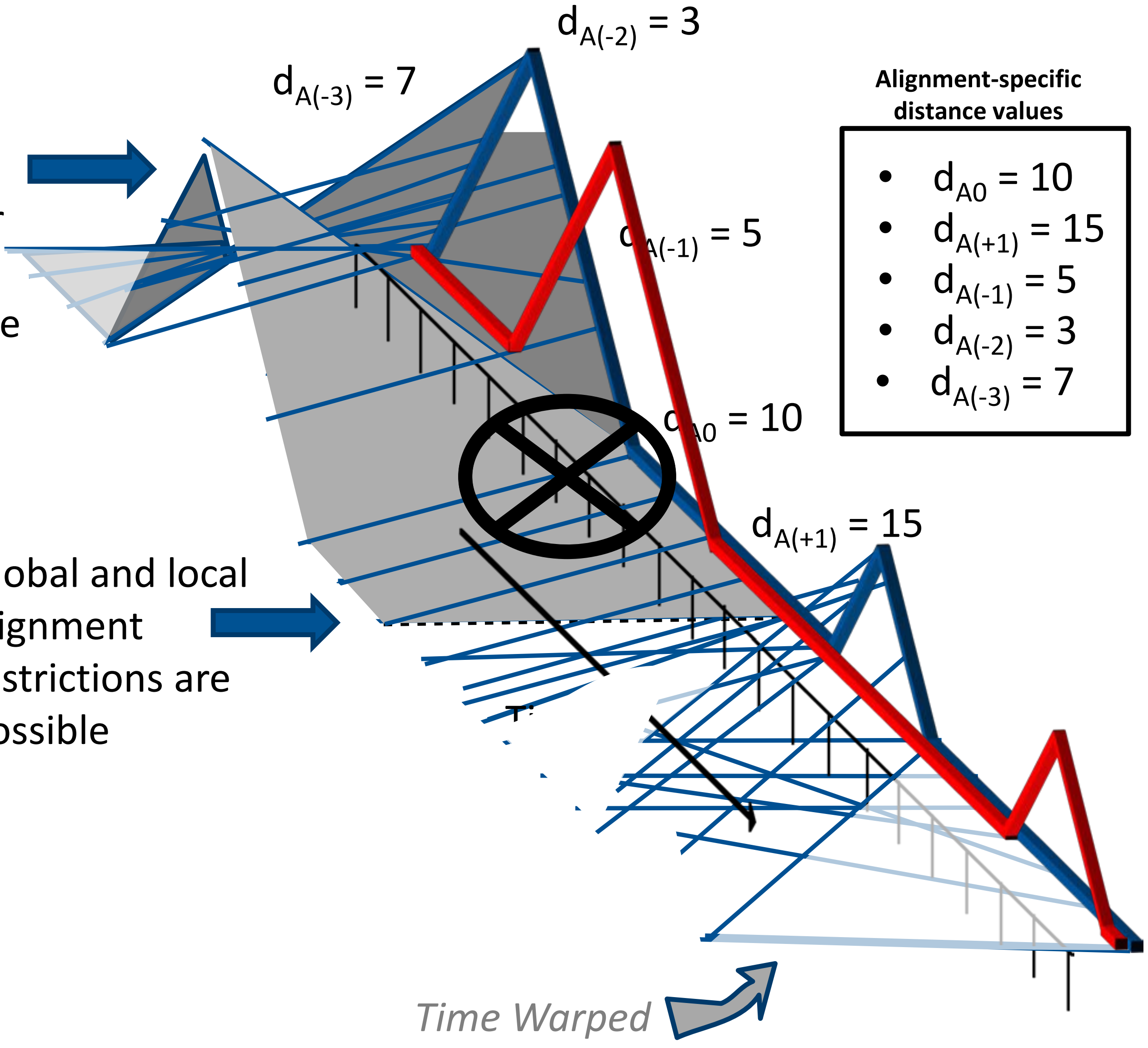
What is Dynamic Time Warping (DTW)?

- Time warp: repeating specific observations on metabolite X until metabolite Y is fitted.

- Identifies the alignment pair producing lowest distance value

- Global and local alignment restrictions are possible

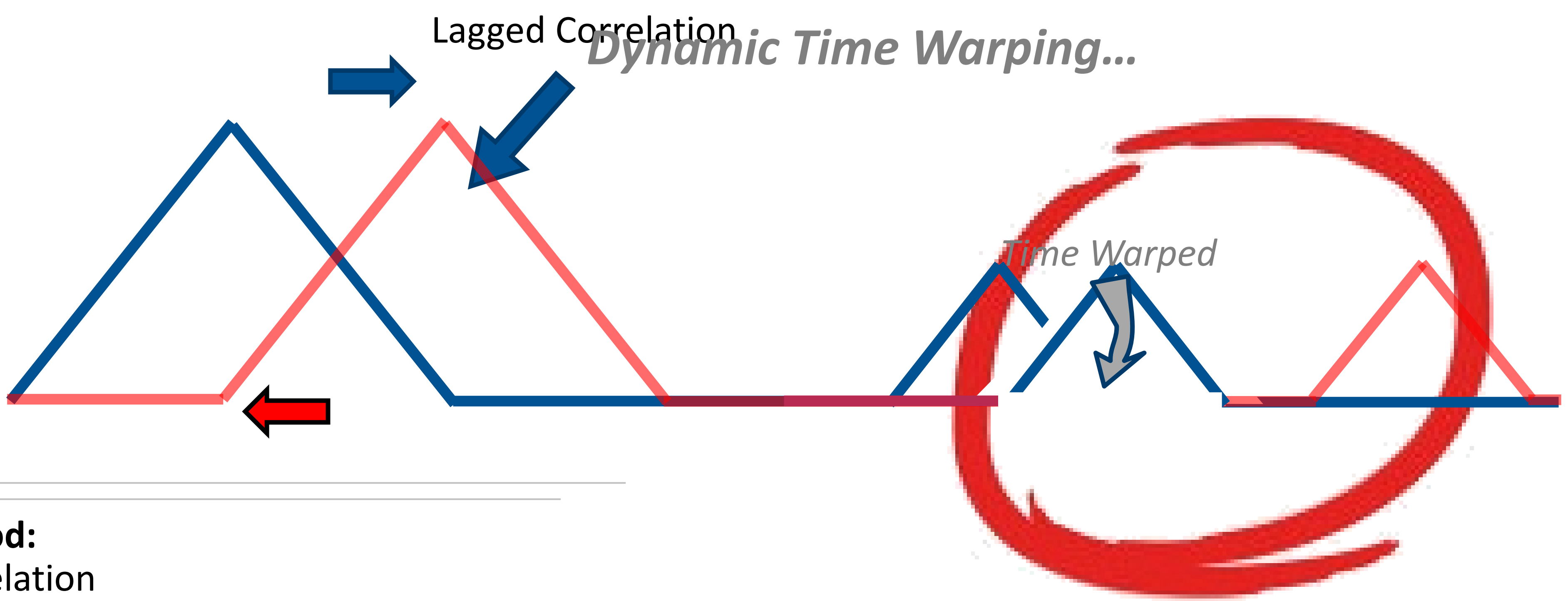




Time-Series Metabolomics Data

Metabolite #1

Metabolite #2



Similar Method:

- Cross-correlation
  - Uniform shift of variable
  - Non-equidistant alignment similarities are missed

Is capable of overcoming uniformity by warping time in a non-linear fashion to align with the nearest pairwise point

are often missed

Quick Summary

What is Dynamic Time Warping?

- DTW identifies the alignment producing the lowest distance value

Don’t similar methods exist?

- Cross-correlation
  - Uniform shift of alignment

Why would we need Dynamic Time Warping ?

- Non-uniform shifts in alignment allows us to identify correlations not seen in traditional statistics.

So, does it work?

HuMet Study

Oral Glucose Tolerance Test

Data origin: HuMet Study
Method: GC/MS; LC/MS; NMR

# of subjects:
- 15 young men selected to be as homogenous as possible

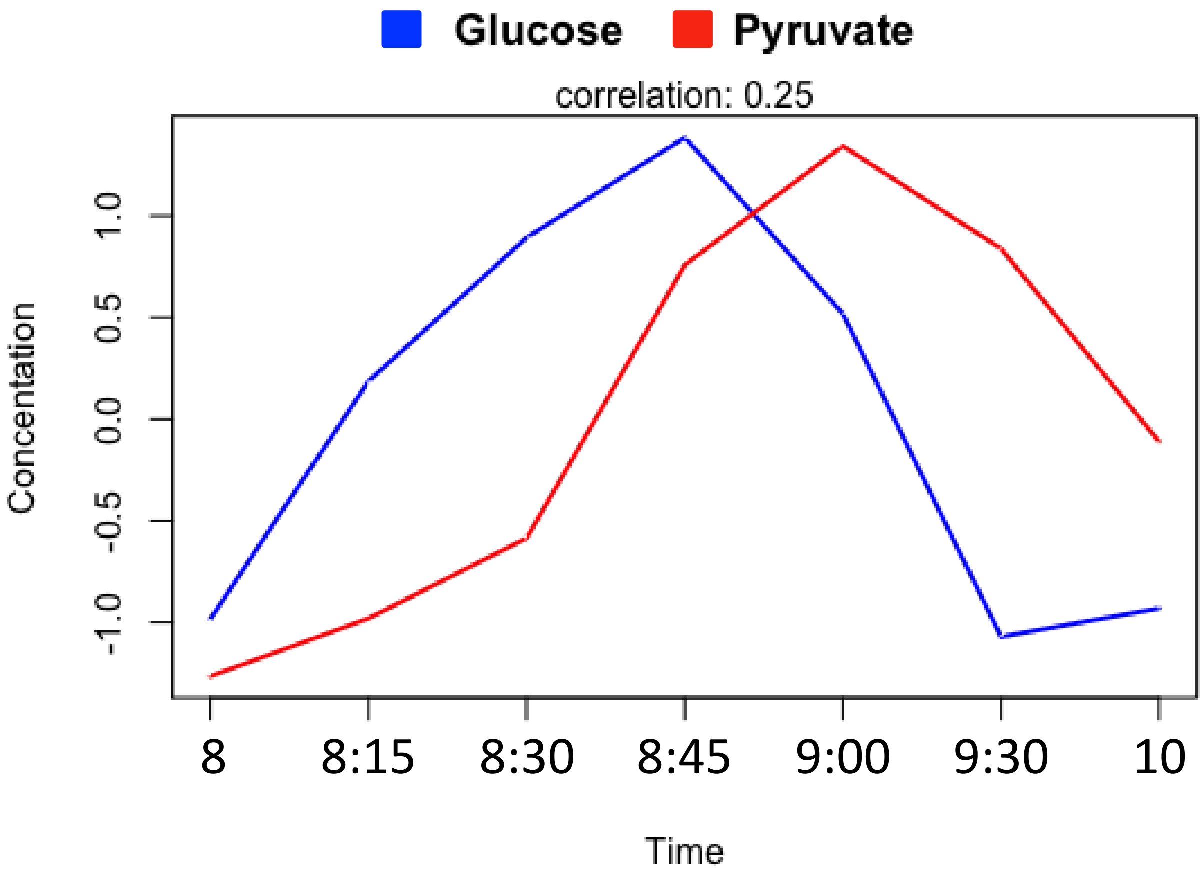
Type of samples: Plasma
Measurements by: Metabolon Inc.

# of variables: 1000+
- Amino Acids
- Carbohydrates
- Energy

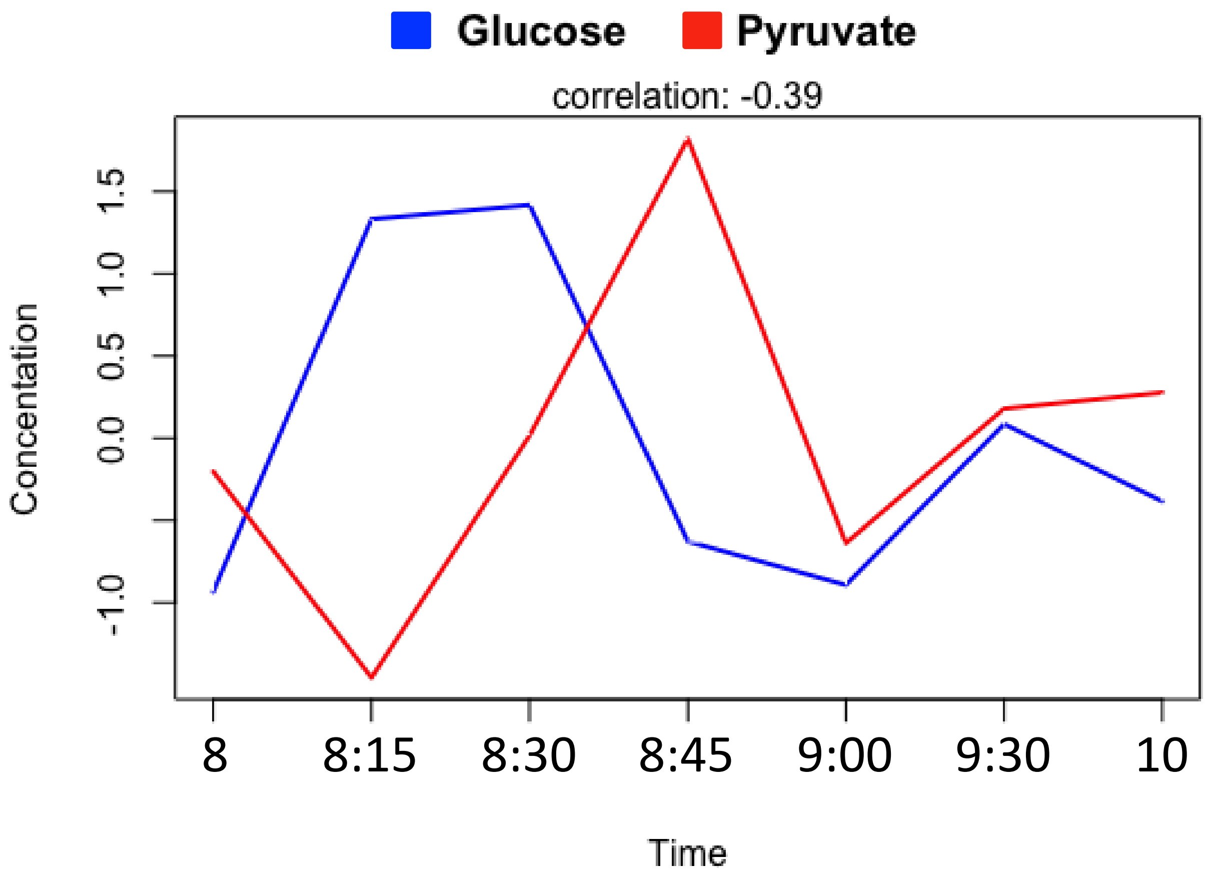
Time:

Special thanks to Prof. Daniel for the opportunity!!

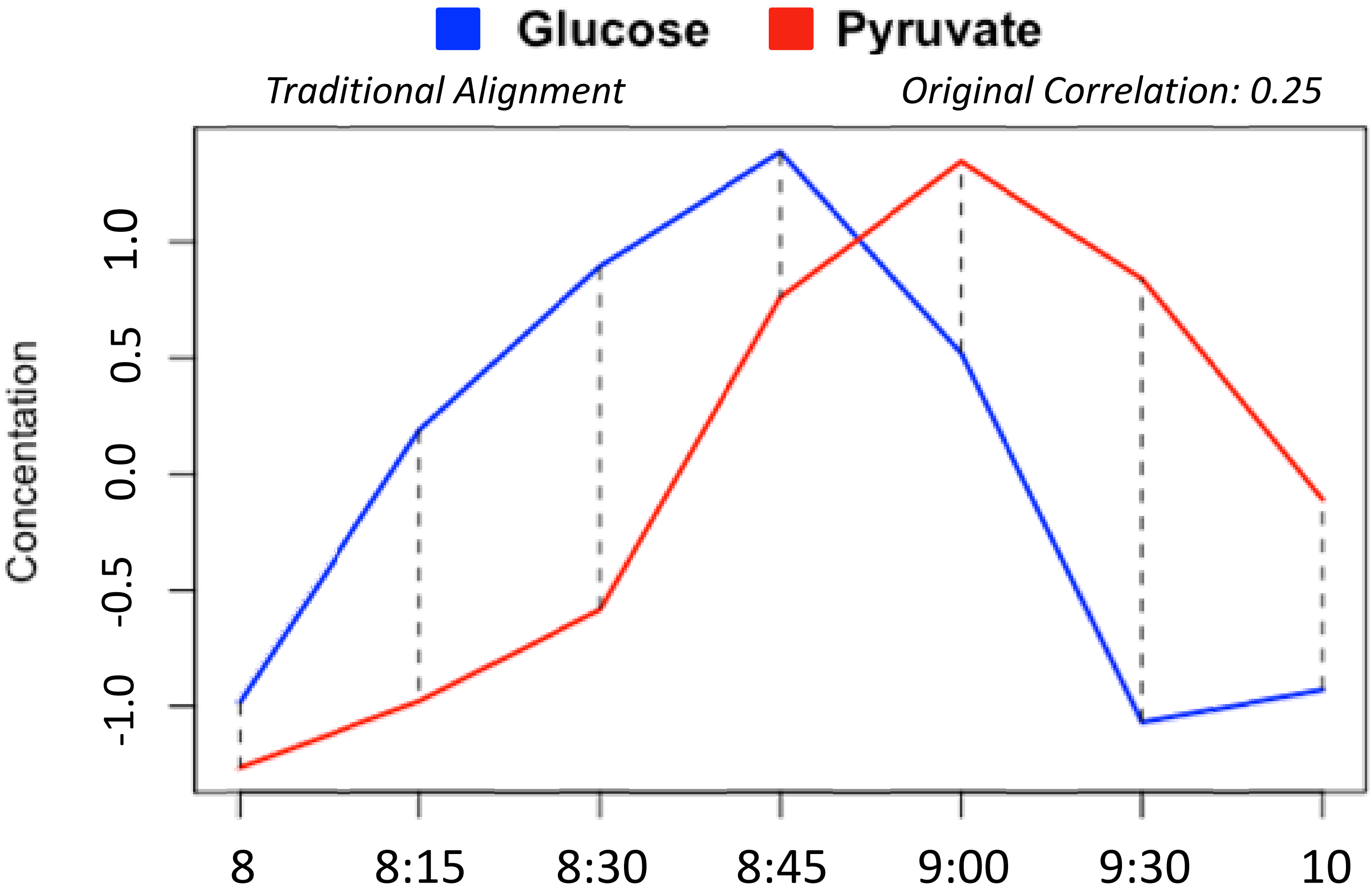
Website: www.Humet-TUM.de



Subject 1



Subject 2

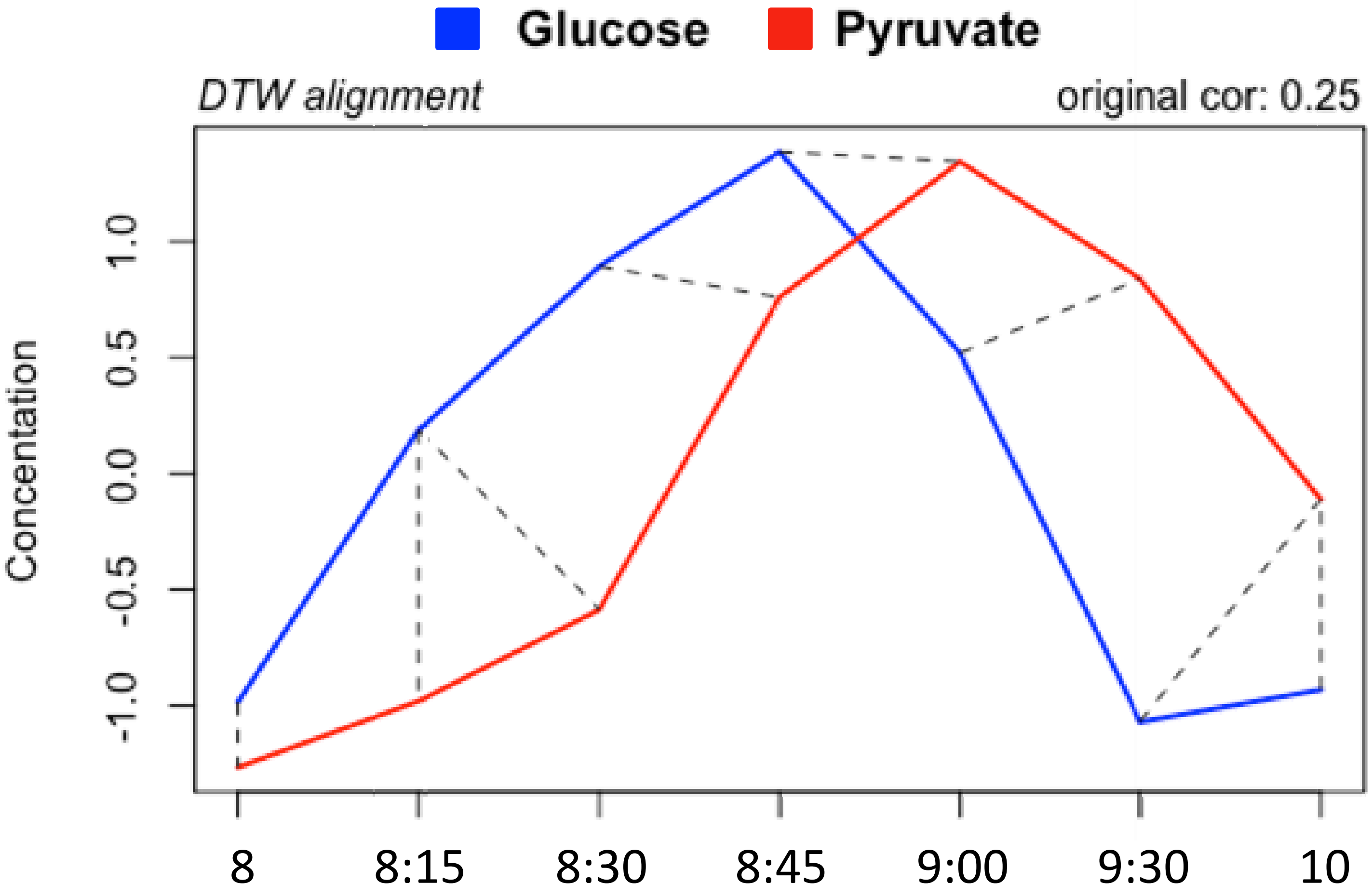


Metabolite 1

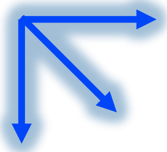
Metabolite 2	1	2	3	4	5	6	7
1	+1	+3	+5	+6	+7	+8	+9
2	+3	+2	+2	+4	+5	+6	+8
3	+5	+1	+4	+2	+5	+7	+6
4	+6	+6	+1	+2	+5	+6	+9
5	+7	+7	+6	+0	+2	+1	+4
6	+9	+8	+3	+2	+1	+5	+3
7	+1	+4	+3	+1	+3	+1	+0

Example: Euclidean Distance

Traditional distance = 16



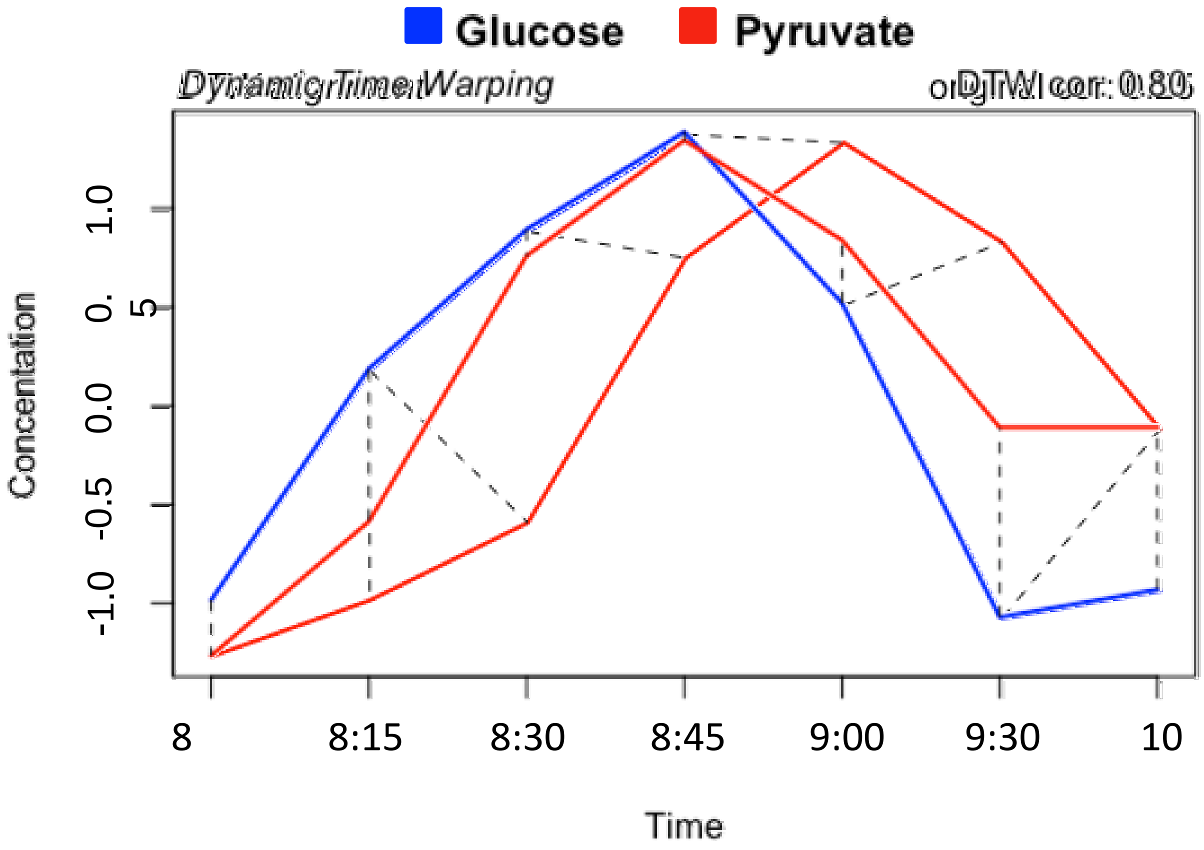
Example: Dynamic Time Warping



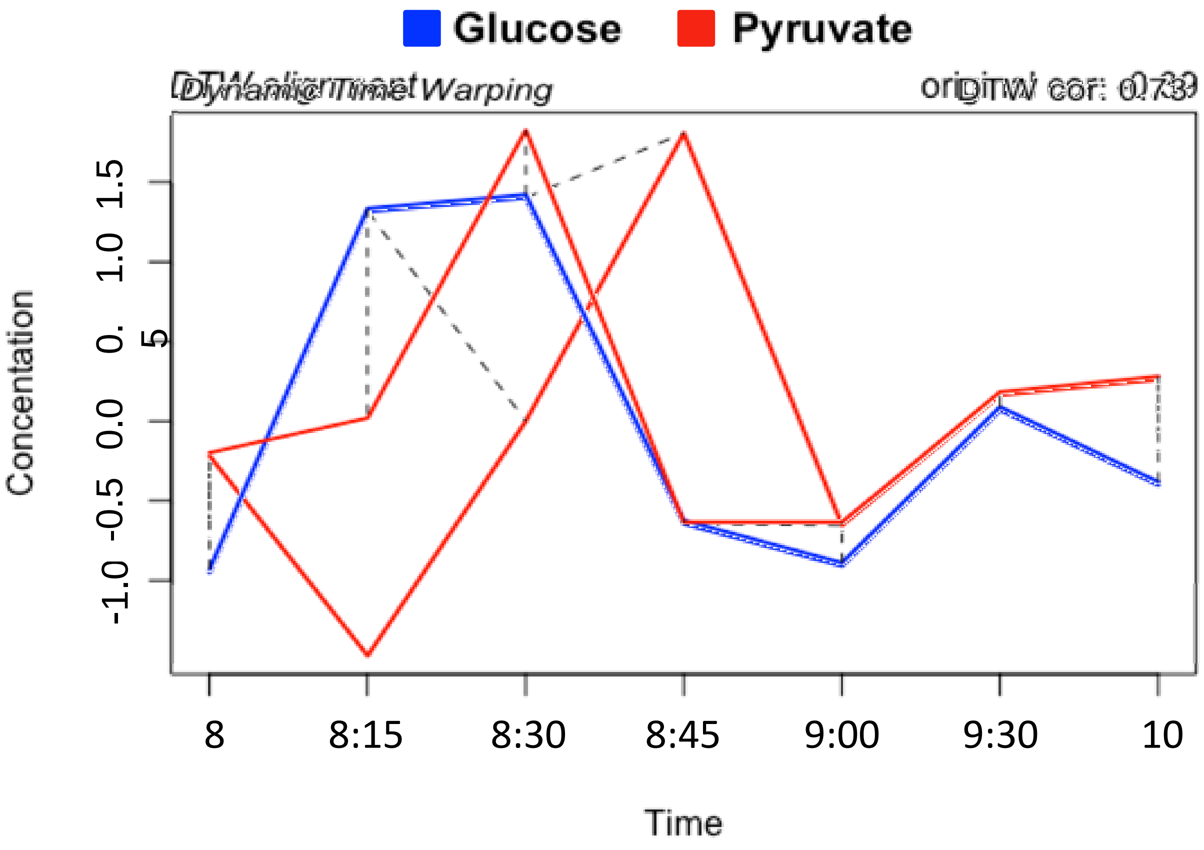
Metabolite 1

Metabolite 2	1	2	3	4	5	6	7
1	+1	+3	+5	+6	+7	+8	+9
2	+3	+2	+2	+4	+5	+6	+8
3	+5	+1	+4	+2	+5	+7	+6
4	+6	+6	+1	+2	+5	+6	+9
5	+7	+7	+6	+0	+2	+1	+4
6	+9	+8	+3	+2	+1	+5	+3
7	+1	+4	+3	+1	+3	+1	+0

DTW distance = 7



Subject 1



Subject 2

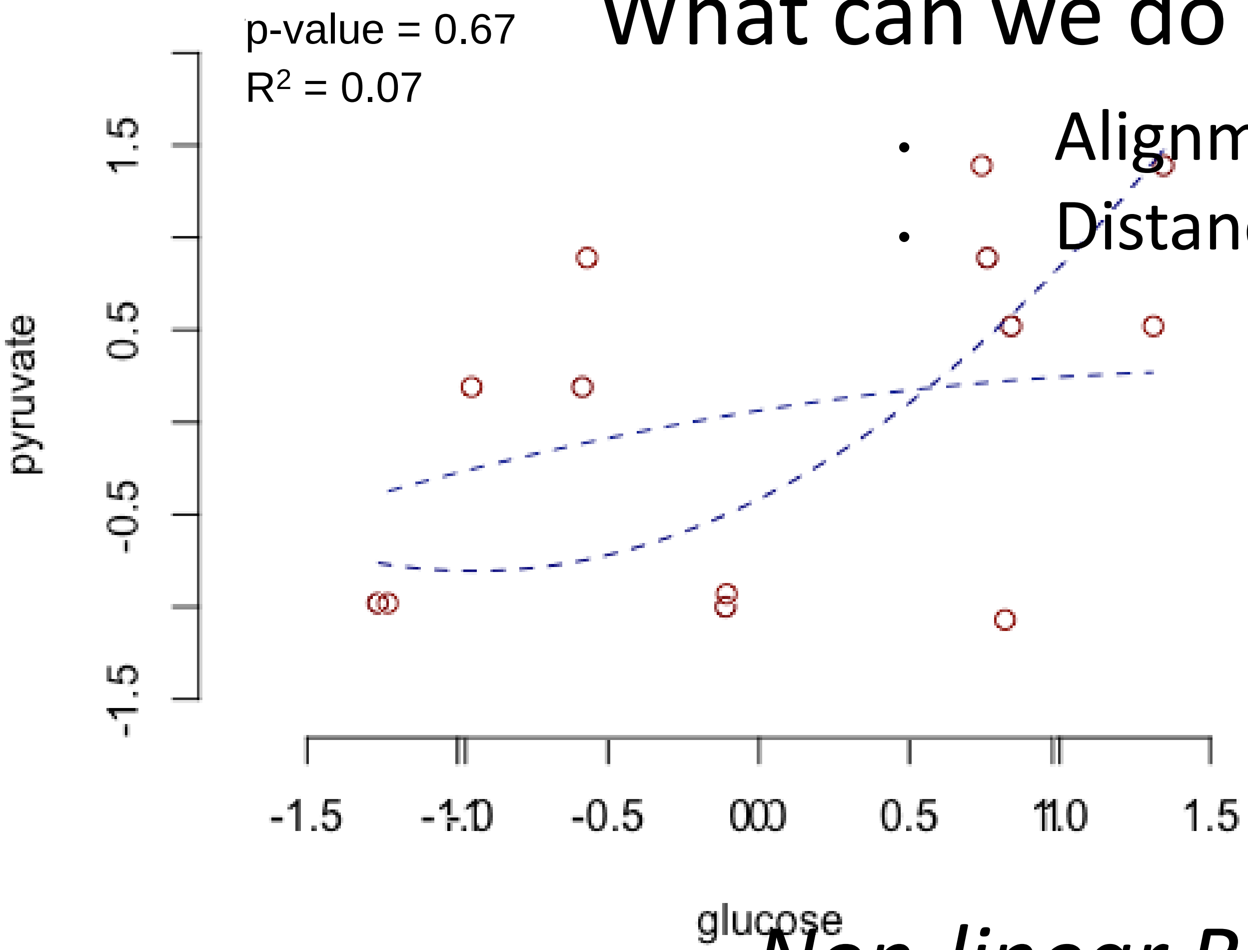
Non-linear Regression

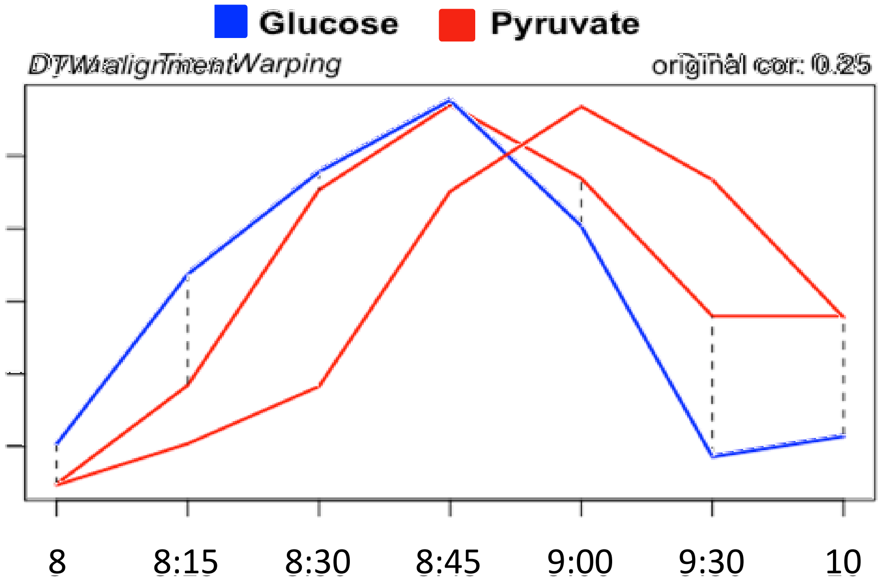
Traditional Alignment

Warped Alignment

What can we do with our new

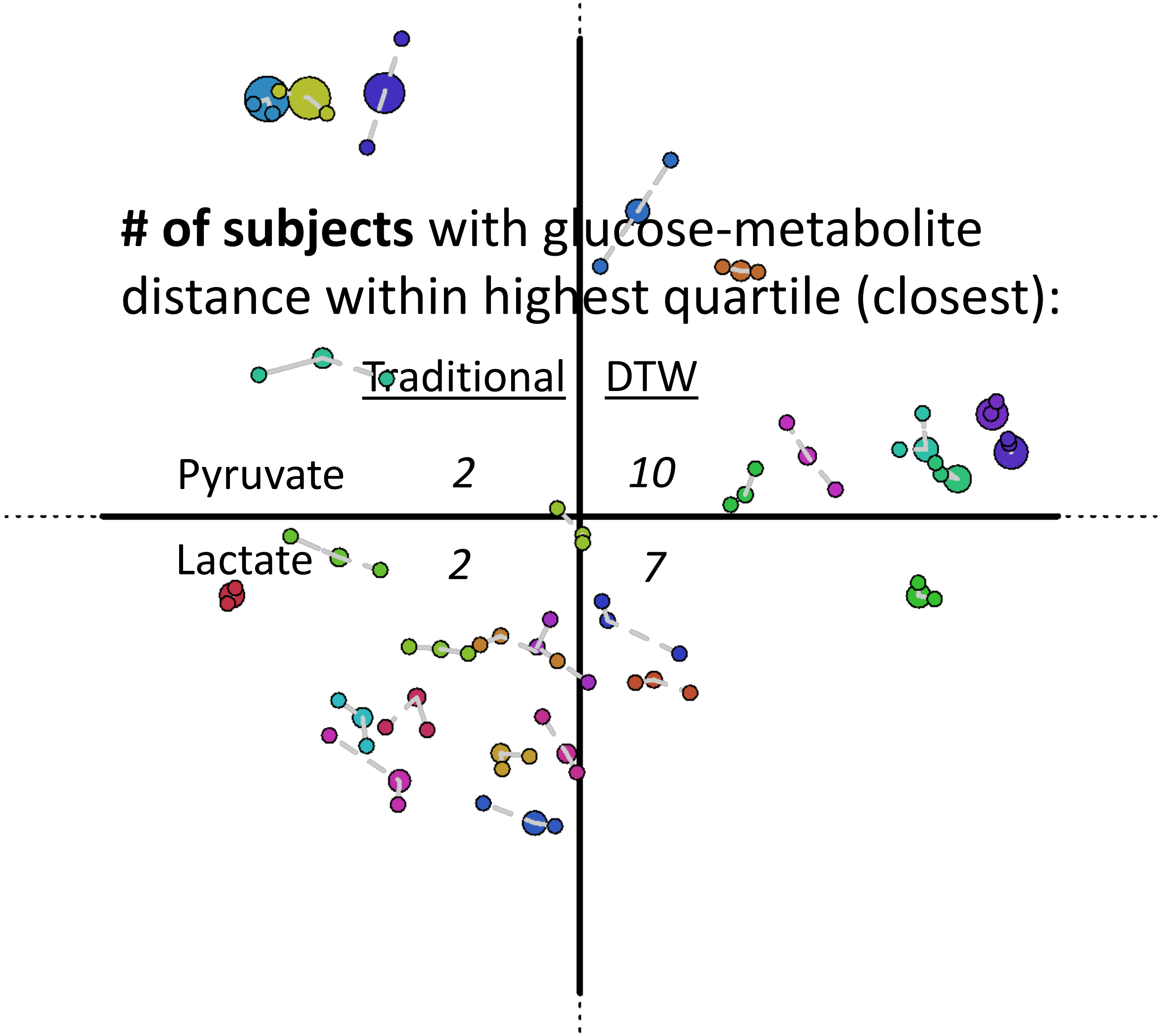
- Alignments
- Distance





Non-linear Regression

Feature Selection



# of subjects with glucose-metabolite distance within highest quartile (closest):

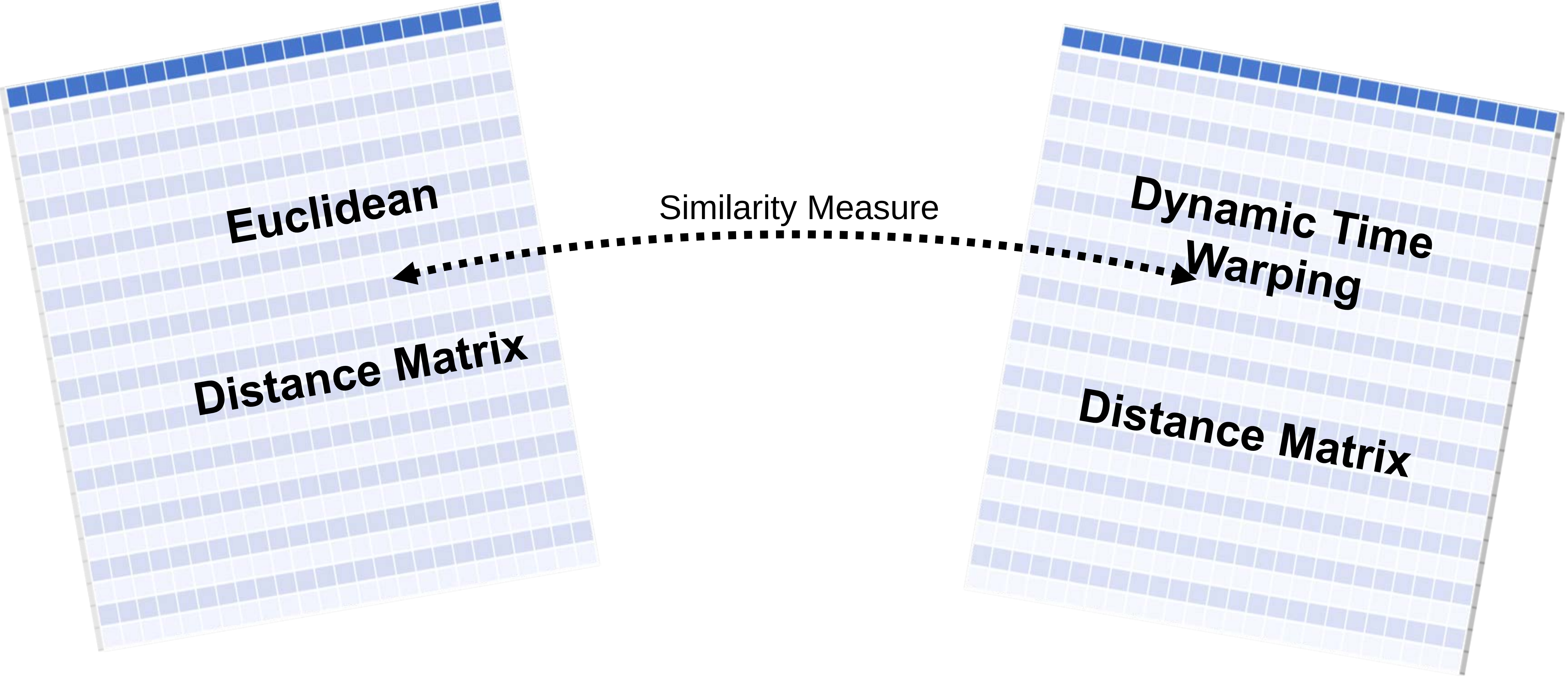
	Traditional	DTW
Pyruvate	*2*	*10*
Lactate	*2*	*7*

Ranking of Average Subject Glucose-Metabolite Distance

Metabolite	Metabolite-Glucose Rank (DTW)	Metabolite-Glucose Rank (EUC)
N-acetyltryptophan	4	28
phenyllactate (PLA)	10	84
oxalate (ethanedioate)	17	58
indoleacetate	19	62
alpha-hydroxyisocaproate	23	91
pyruvate	24	74
indolepropionate	26	60
beta-hydroxyisovalerate	28	44
hypotaurine	31	77
3-hydroxyisobutyrate	33	71

BCAA intermediate

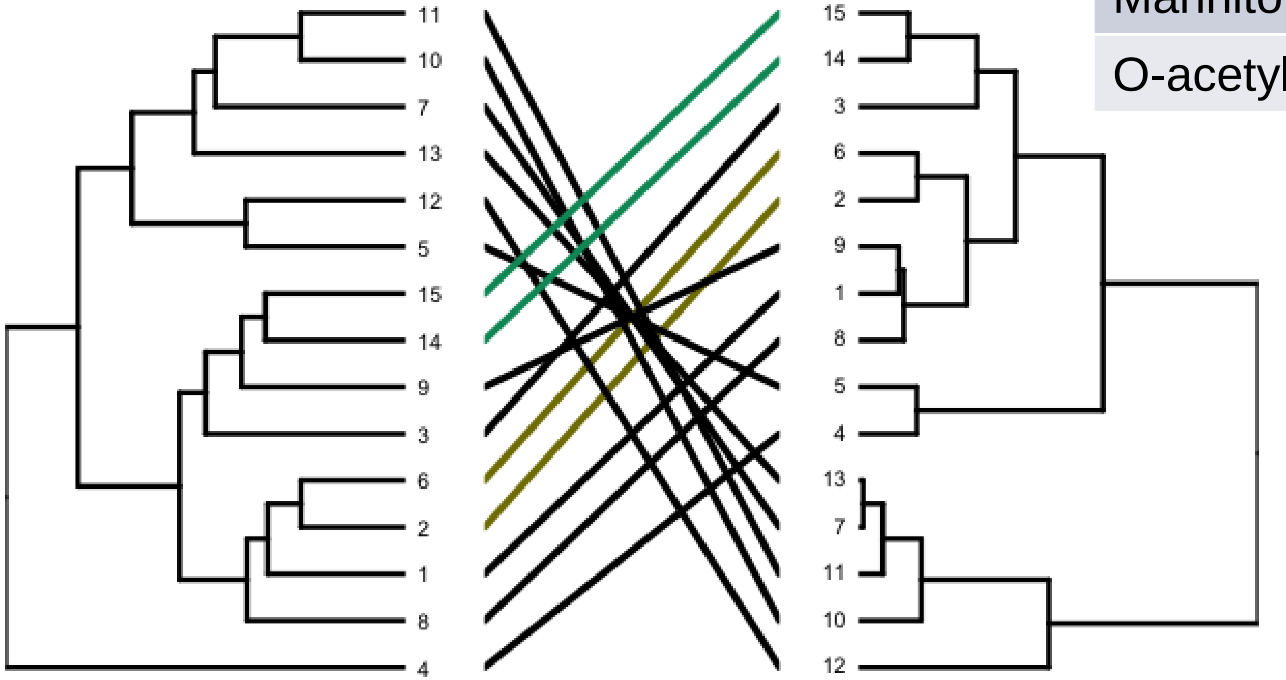
Feature Selection



Euclidean-DTW Correlation: 0.85

Subject Cluster compare

Glucose-Only Subject Clustering



Multivariate DTW vs. Glucose-only Subject Clustering Comparison

Metabolite1	Metabolite 2	Cophenetic Correlation
Mannitol	Kynurenate	0.69
O-acetylhomoserine	Kynurenate	0.62

Multivariate - Glucose DTW Clustering

A special thank you to:

- *Prof. Hannelore Daniel*
- *Dr. Jan Krumsiek*
- *Systems Medicine of Diabetes group (ICB-Helmholtz)*
- *Dr. Kurt Gedrich*
- *Dr. Pieter Giesbertz*
- *Alessio Ciurli*