Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

Working with sparse time series data in computational biology

Benedikt Venn

CSB
COMPUTATIONAL SYSTEMS BIOLOGY

- Biology bachelor and master studies @ RPTU Kaiserslautern
- PhD student @ Computational systems Biology (RPTU)

CSB
COMPUTATIONAL SYSTEMS BIOLOGY

BioFSharp

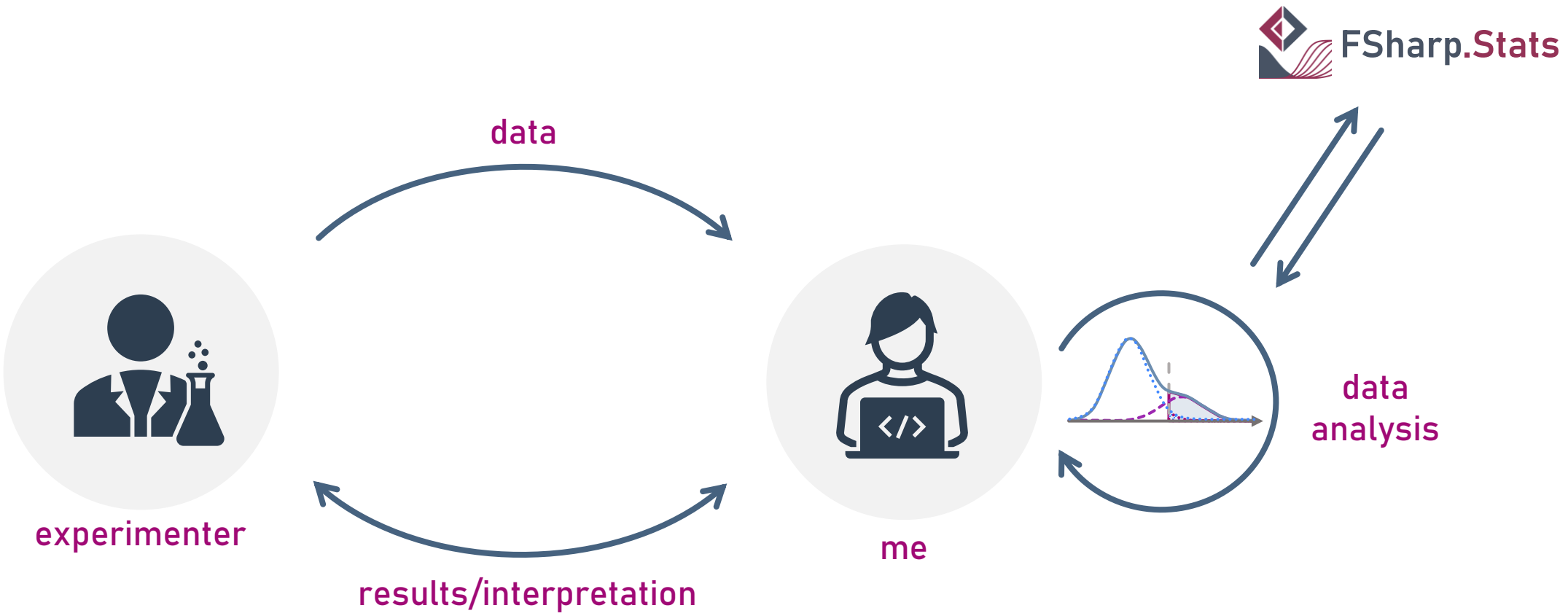
Plotly.NET

Graphoscope

FSharp.Stats

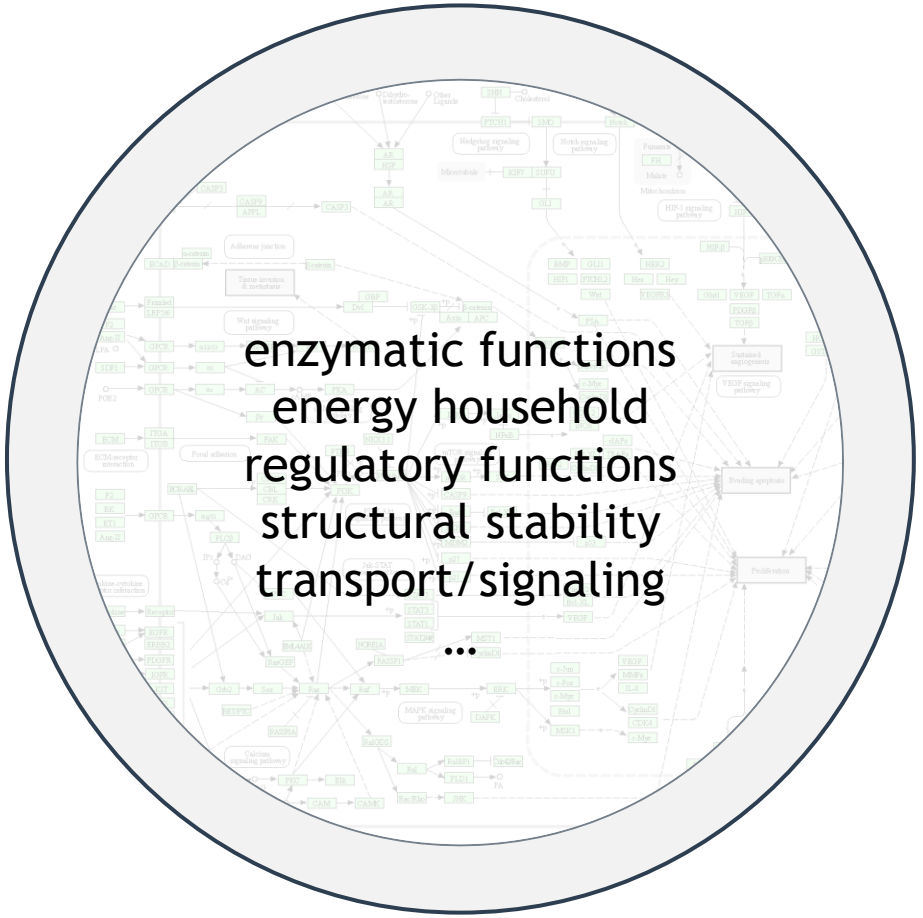
Cytoscape.NET

Daily business



Case study – plant heat acclimation

culture of algae/plants/cells



abundance of each protein

Case study – plant heat acclimation

heat treatment

0.0 h

measure protein abundance

	T0
P001	342
P002	723
P003	305
…	…

Case study – plant heat acclimation

measure protein abundance →

	T0	T1
P001	342	497
P002	723	923
P003	305	629
…	…	…

Case study – plant heat acclimation

measure protein abundance →

	T0	T1	T2
P001	342	497	85
P002	723	923	482
P003	305	629	539
…	…	…	…

Case study – plant heat acclimation

measure protein abundance →

	T0	T1	T2	T3
P001	342	497	85	477
P002	723	923	482	500
P003	305	629	539	281
…	…	…	…	…

Case study – plant heat acclimation

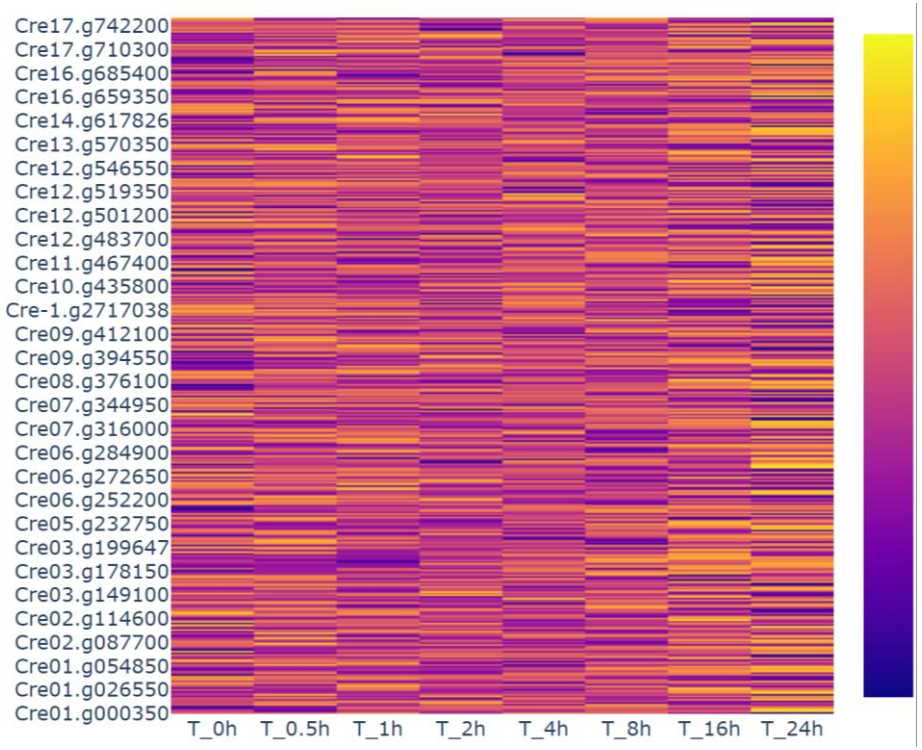
measure protein abundance →

	T0	T1	T2	T3	T4	T5	T6	T7
P001	342	497	85	477	785	821	505	427
P002	723	923	482	500	856	889	335	175
P003	305	629	539	281	429	601	42	595
…	…	…	…	…	…	…	…	…

m*n protein abundance matrix

	0.0 h	0.5 h	1.0 h	2.0 h	4.0 h	8.0 h	16.0 h	24.0 h
P0001	342	497	85	477	785	821	505	427
P0002	723	923	482	500	856	889	335	175
P0003	305	629	539	281	429	601	42	595
P0004	584	54	477	991	10	302	63	52
P0005	870	90	184	195	856	254	611	992
…	…	…	…	…	…	…	…	…
P2000	501	414	416	89	27	720	794	780

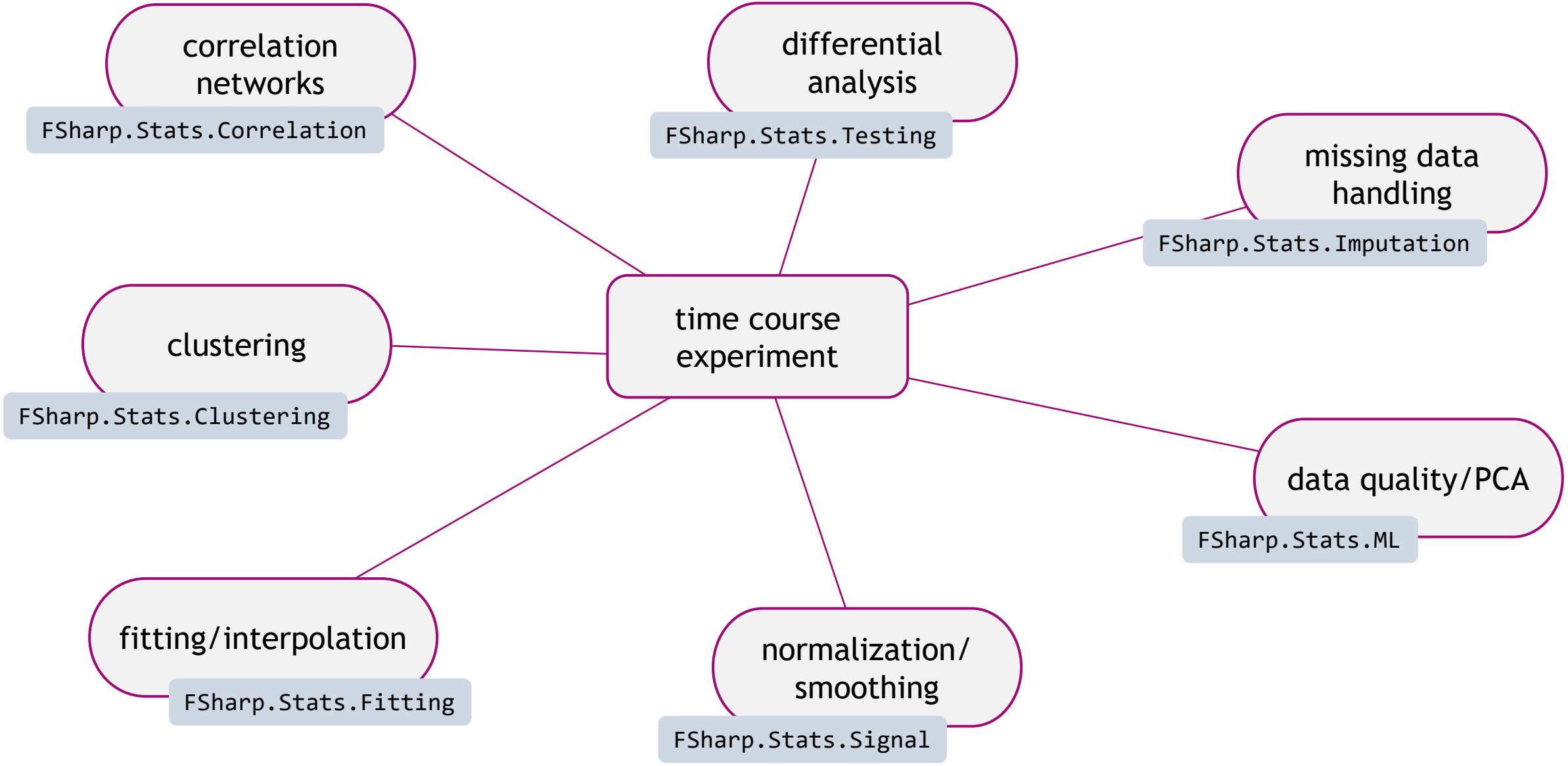
typical biological time course



Zhang et al., 2022

FSharp.Stats

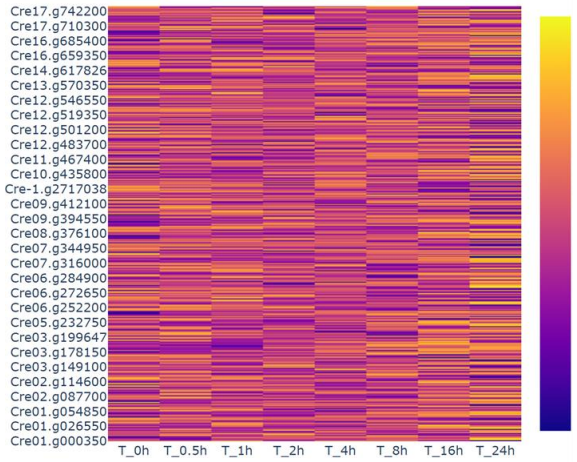
- fslab package for statistical computing
  - Hypothesis testing
  - Signal processing
  - Interpolation
  - Regression
  - Descriptive statistics
  - …

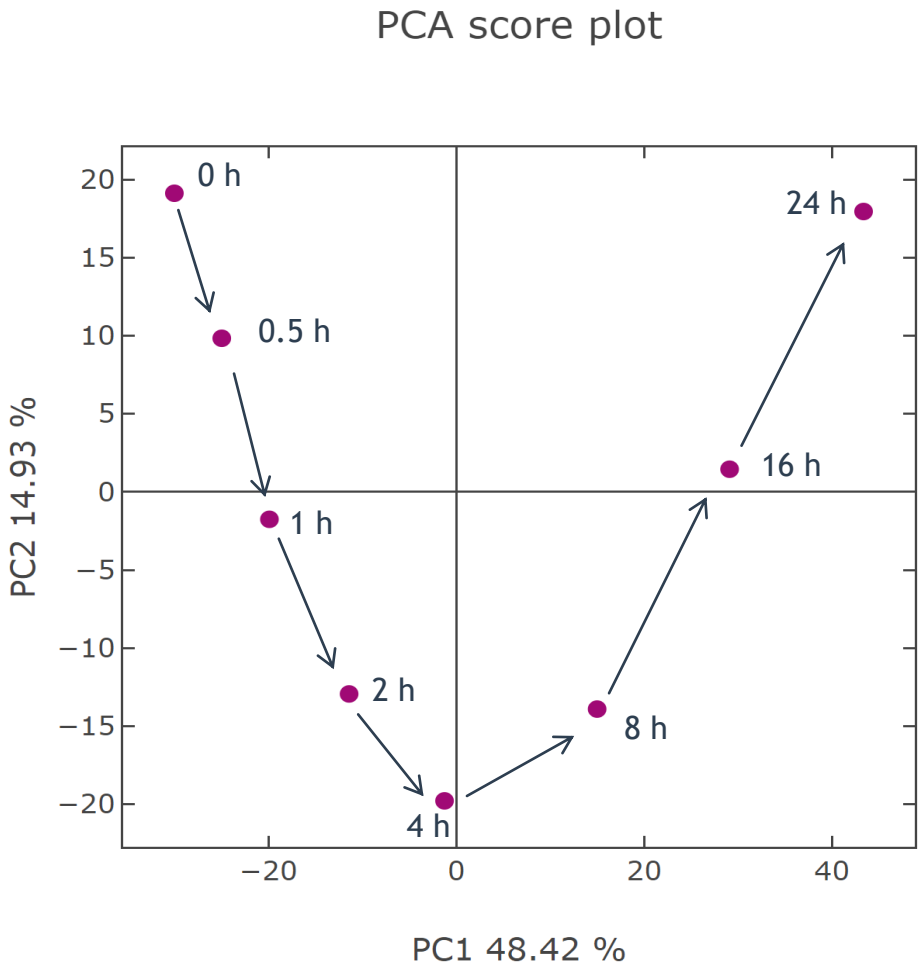

correlation networks
FSharp.Stats.Correlation
differential analysis
FSharp.Stats.Testing
missing data handling
FSharp.Stats.Imputation
clustering
FSharp.Stats.Clustering
time course experiment
data quality/PCA
FSharp.Stats.ML
fitting/interpolation
FSharp.Stats.Fitting
normalization/ smoothing
FSharp.Stats.Signal

data quality/PCA

`FSharp.Stats.ML`

- dimensionality reduction to reveal
  - global trends
  - data quality/replicate reproducibility

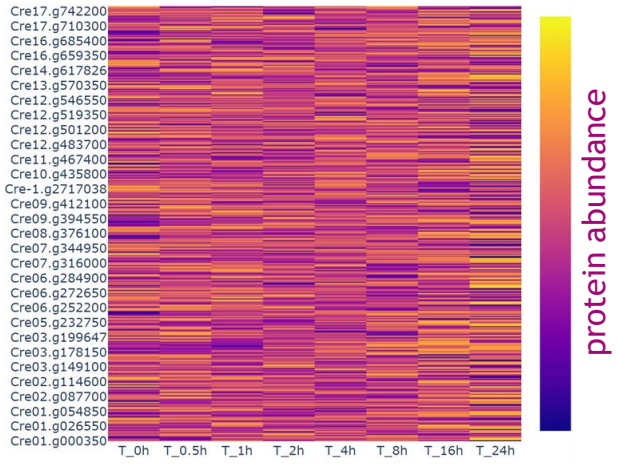


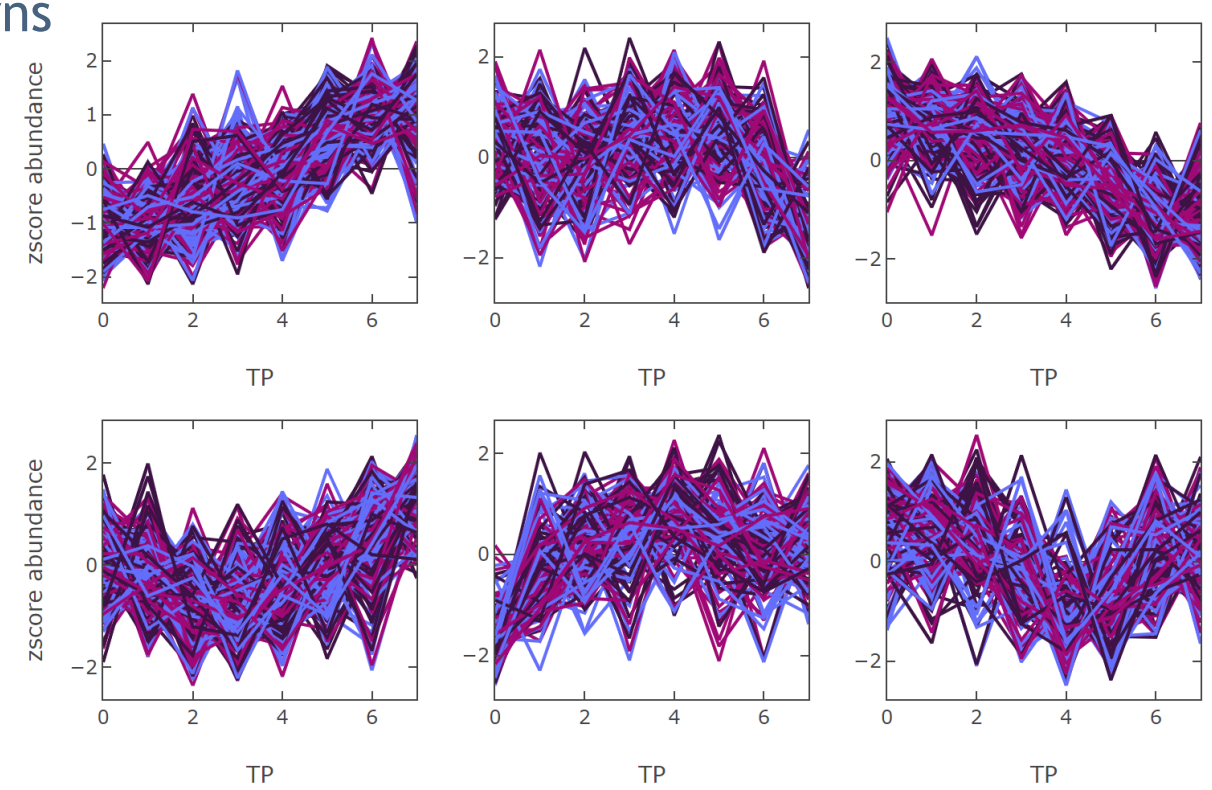


clustering

`FSharp.Stats.Clustering`

- cluster data set in groups of similar behaviour
  - k means clustering with k=6
- enrichment analyses identify common patterns





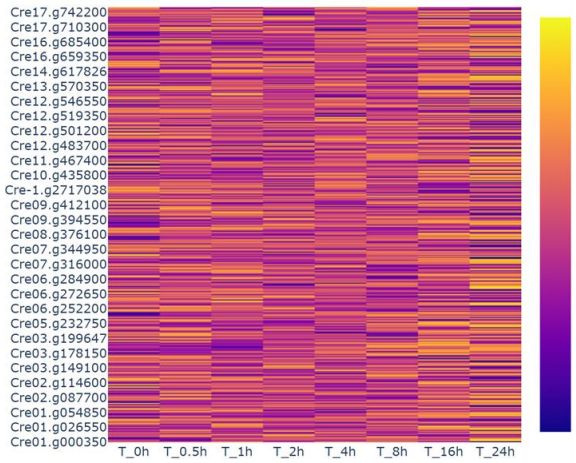
correlation networks

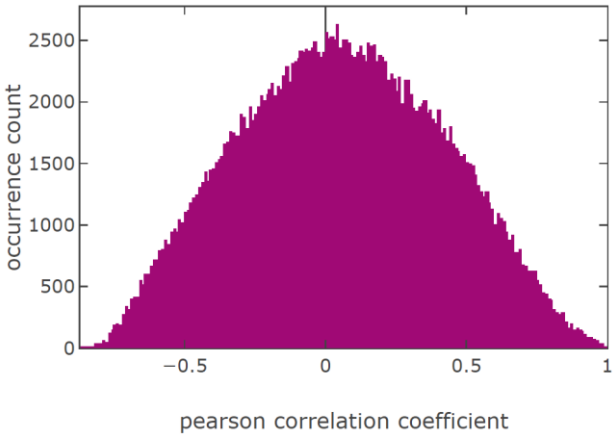
`FSharp.Stats.Correlation`

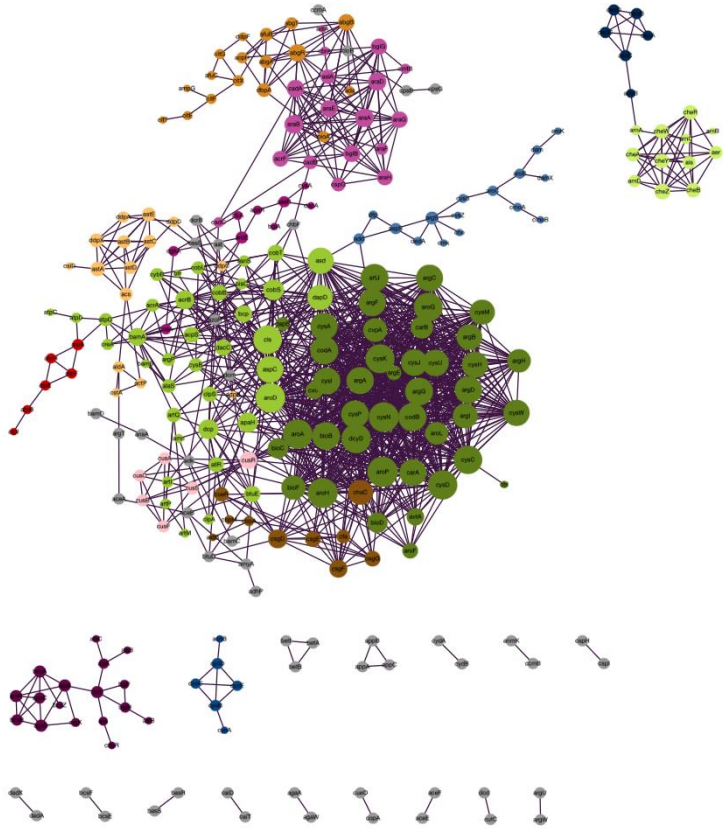
Cytoscape.NET Graphoscope

fslab - correlation-network
CSBlog - community-detection-vs-pca

- construction of co-expression network
- allows identification of protein co-regulation and regulatory clusters



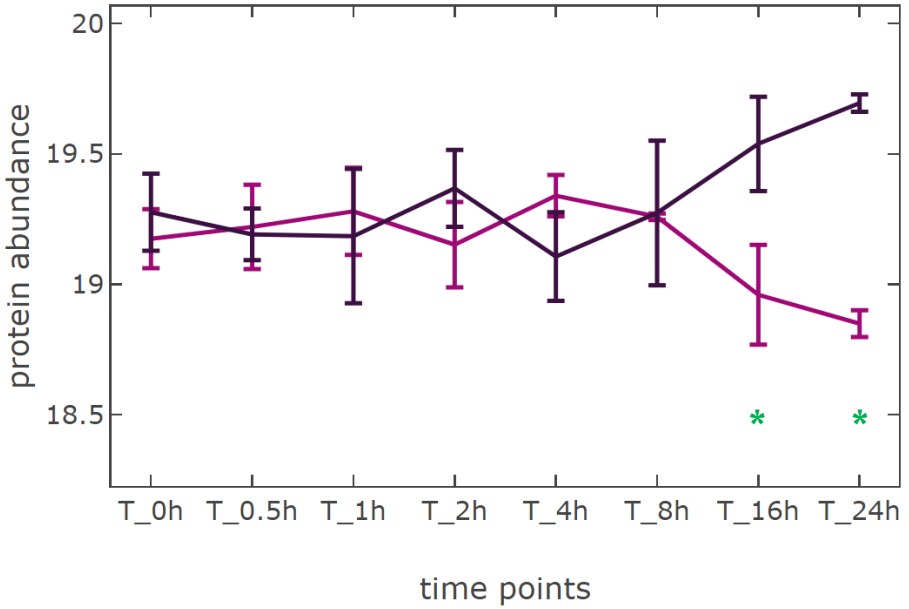


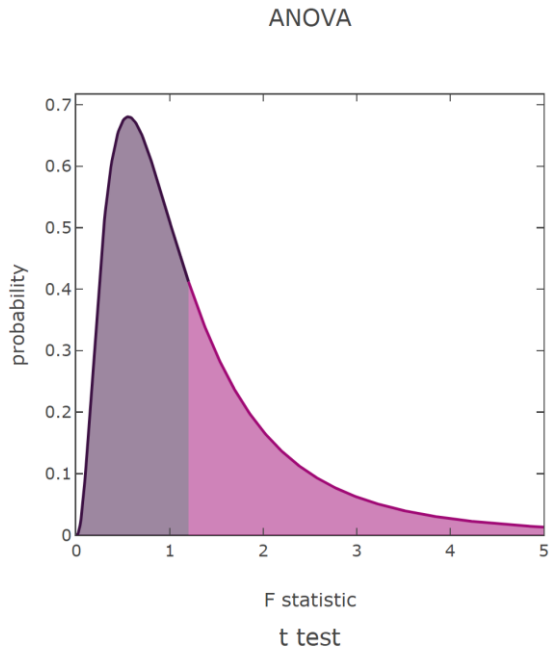


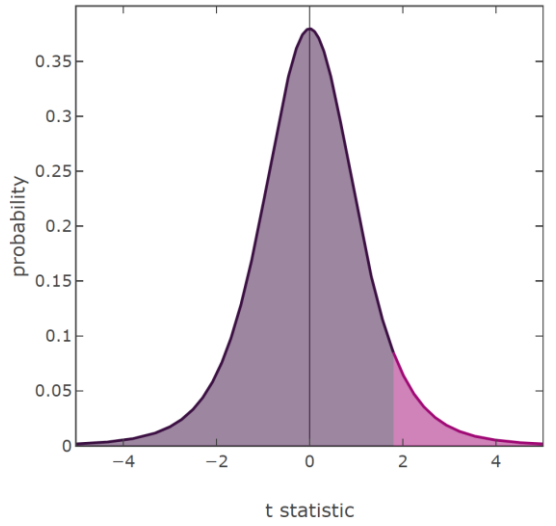
differential analysis

`FSharp.Stats.Testing`

- statistical testing reveals differential protein abundance

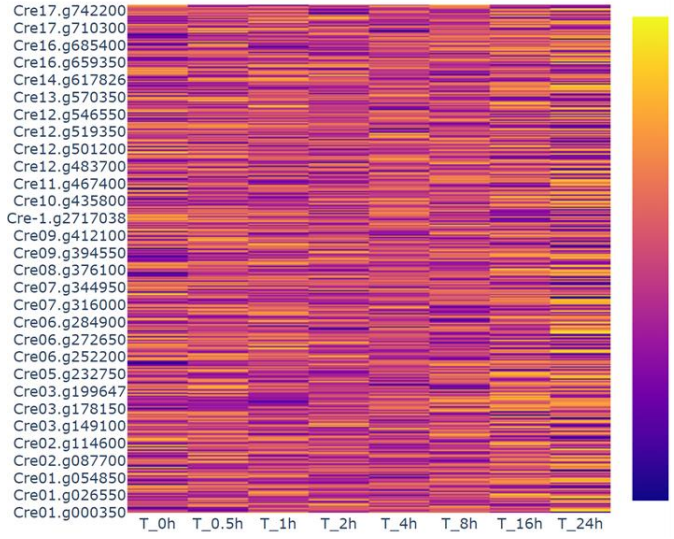


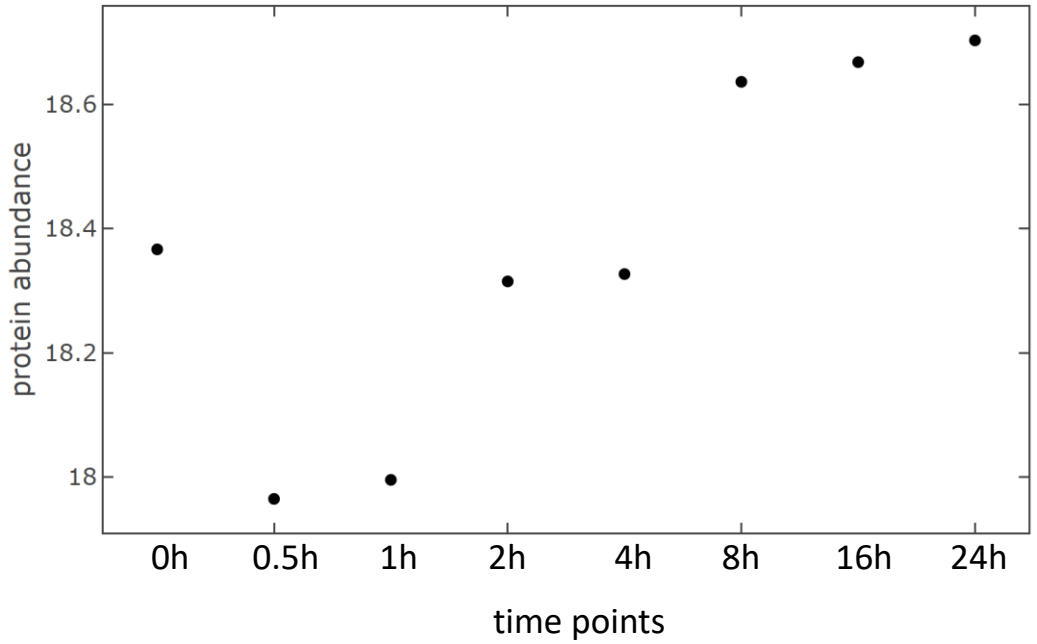




fitting/interpolation

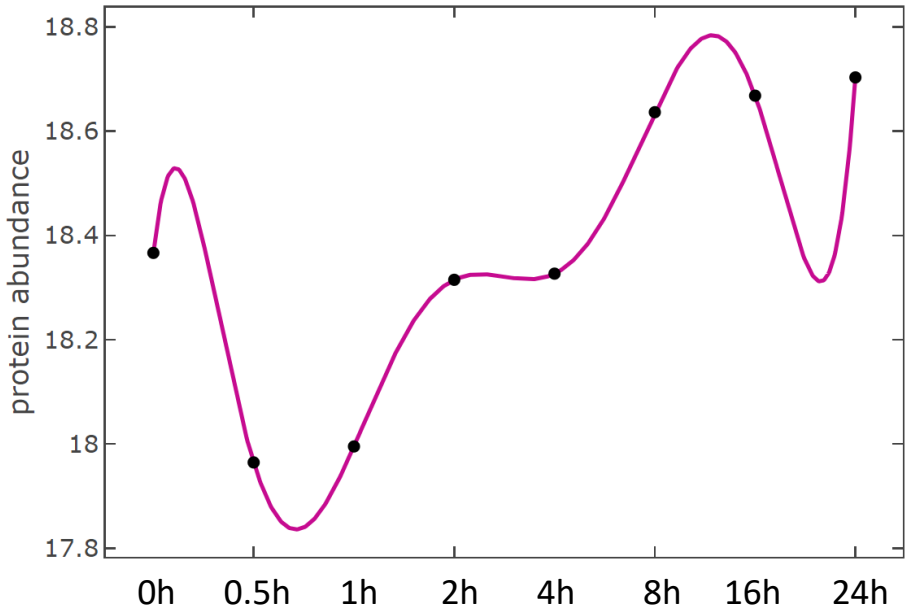
`FSharp.Stats.Interpolation`





fitting/interpolation

`FSharp.Stats.Interpolation`

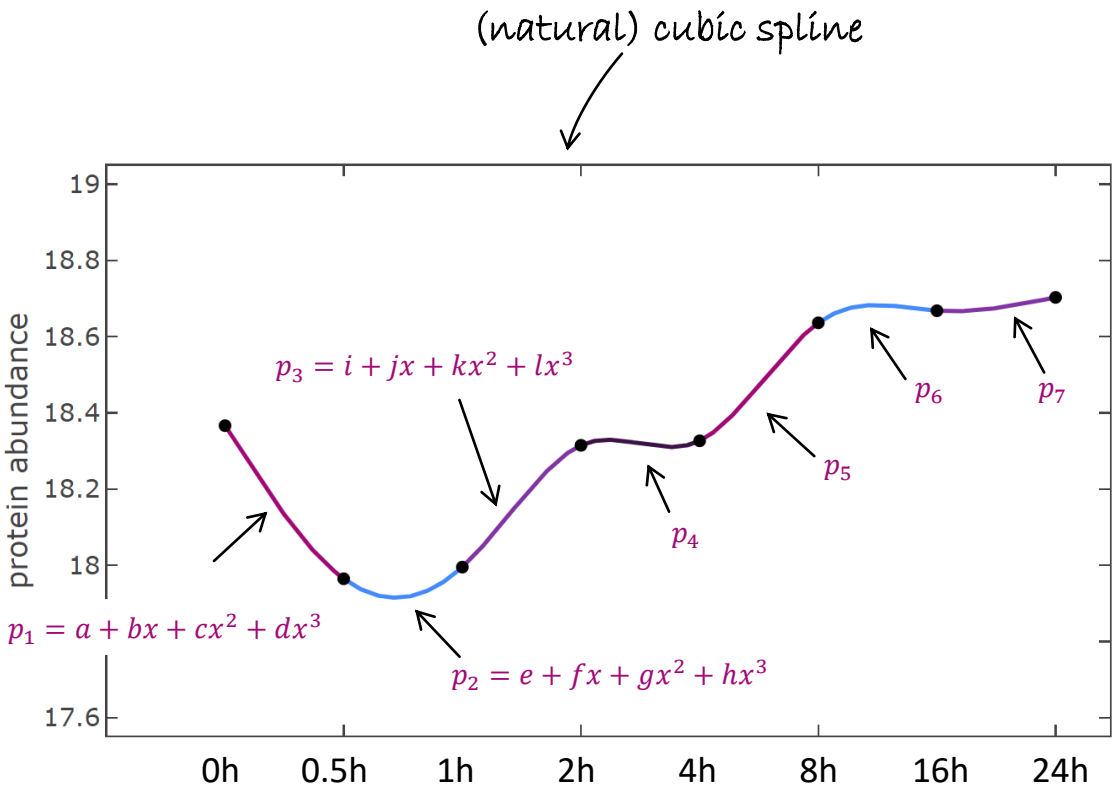


7^{th} order polynomial

$$p(t) = 4.7 + 35.3t - 34.6t^2 + 16.8t^3 - 4.5t^4 + 0.7t^5 - 0.05t^6 + 0.002t^7$$

fitting/interpolation

`FSharp.Stats.Interpolation`



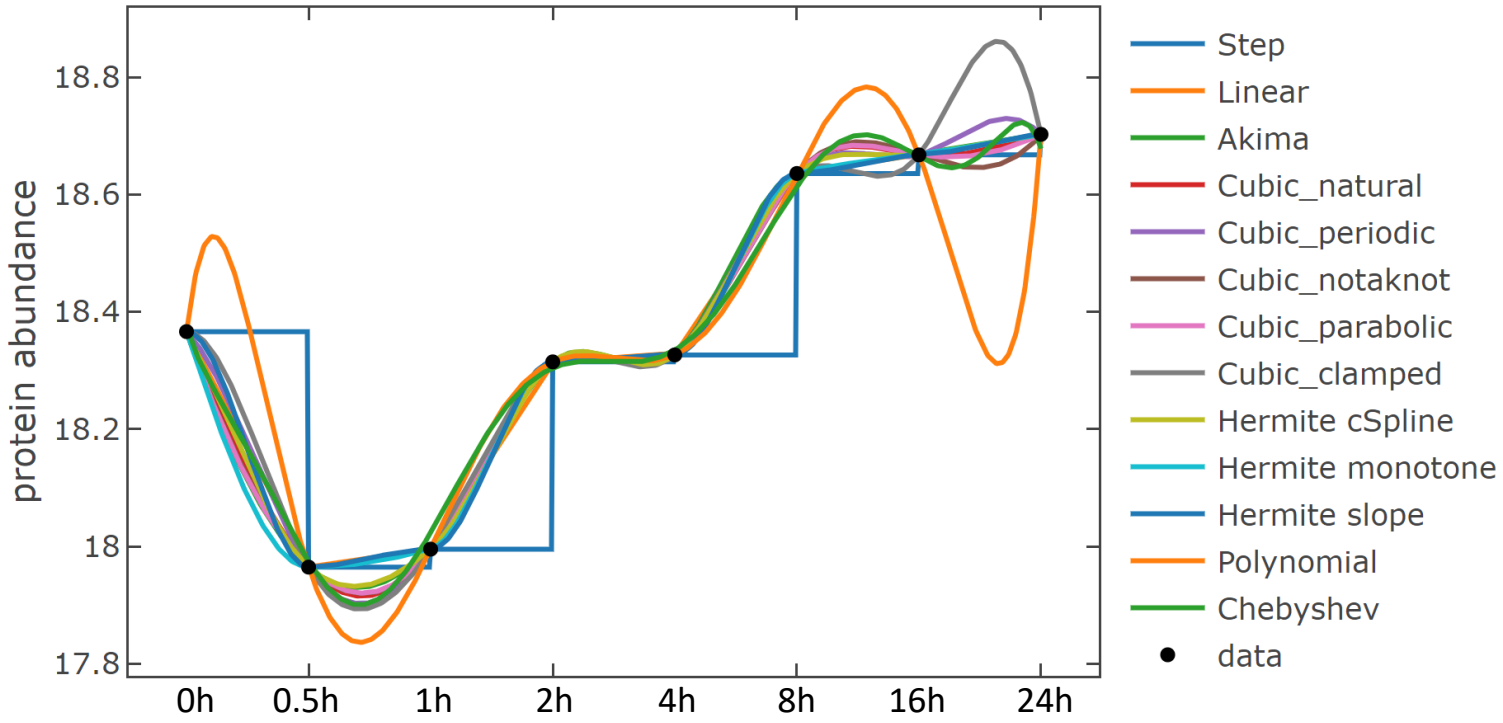
- instead of 8 coefficients (polynomial) 4x7=28 coefficients are needed
  - 14 interpolation constraints
  - 6 slope continuity
  - 6 curvature continuity
  - 2 boundary conditions

$$p_1''(1) = 0$$
$$p_7''(8) = 0 \quad \leftarrow \text{natural cubic spline}$$

$$p_1'(1) = p_7'(8)$$
$$p_1''(1) = p_7''(8) \quad \leftarrow \text{periodic cubic spline}$$

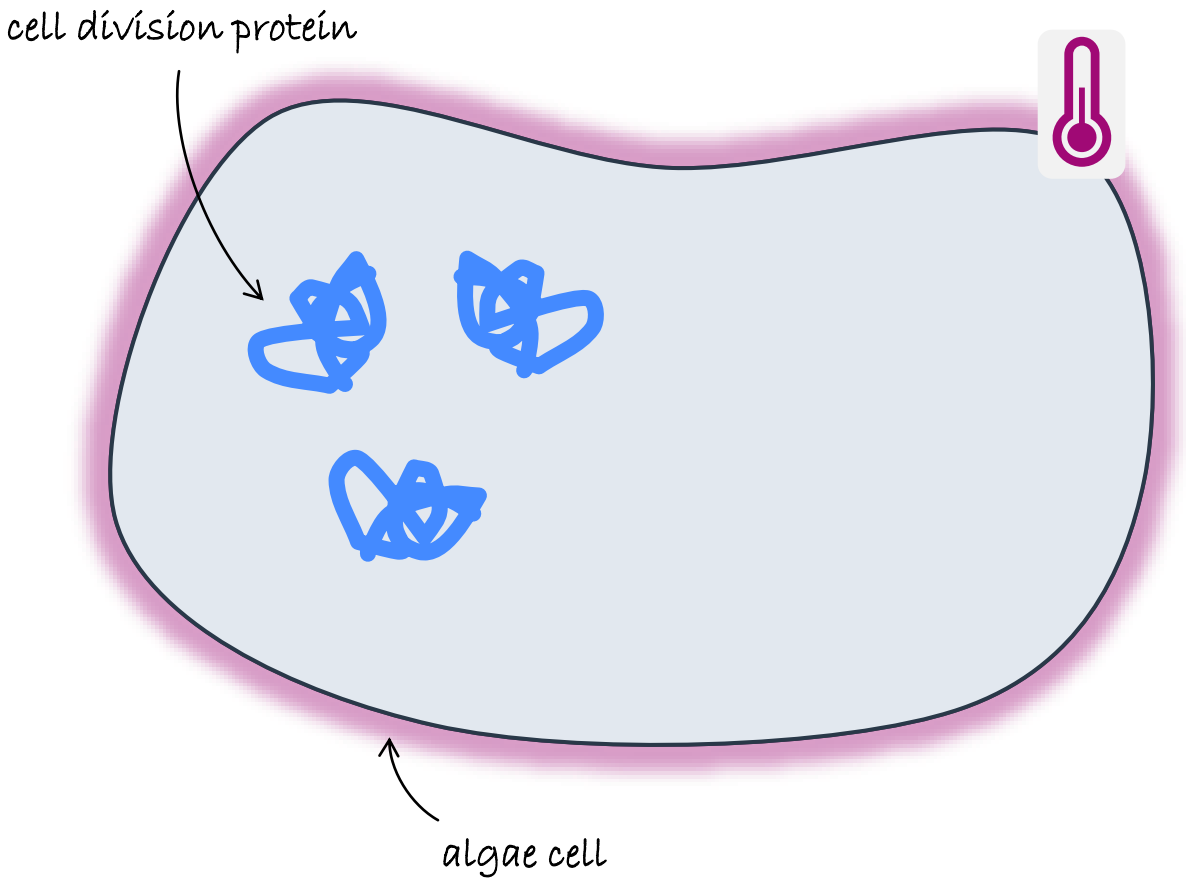
fitting/interpolation

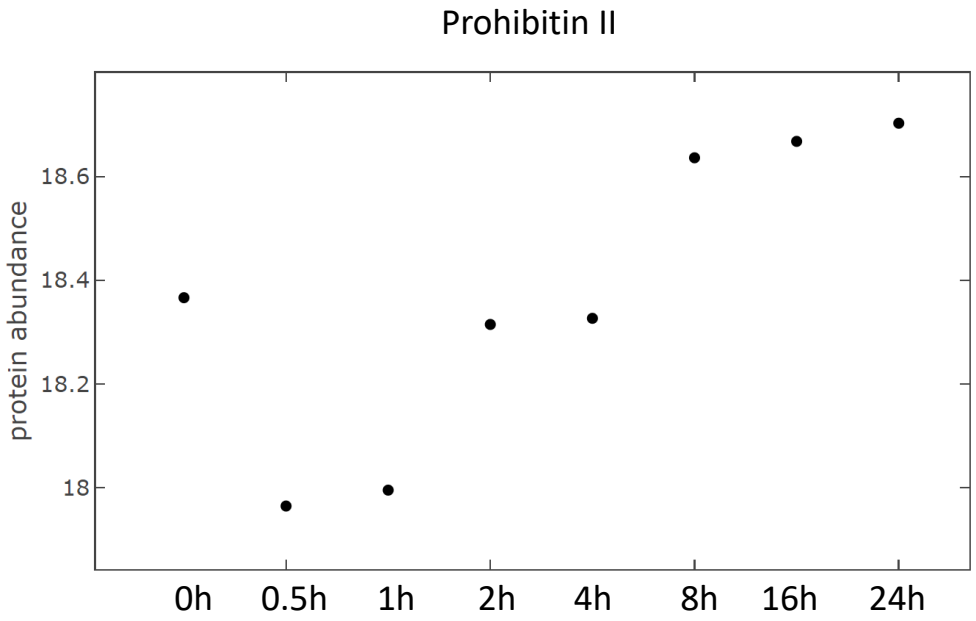
`FSharp.Stats.Interpolation`



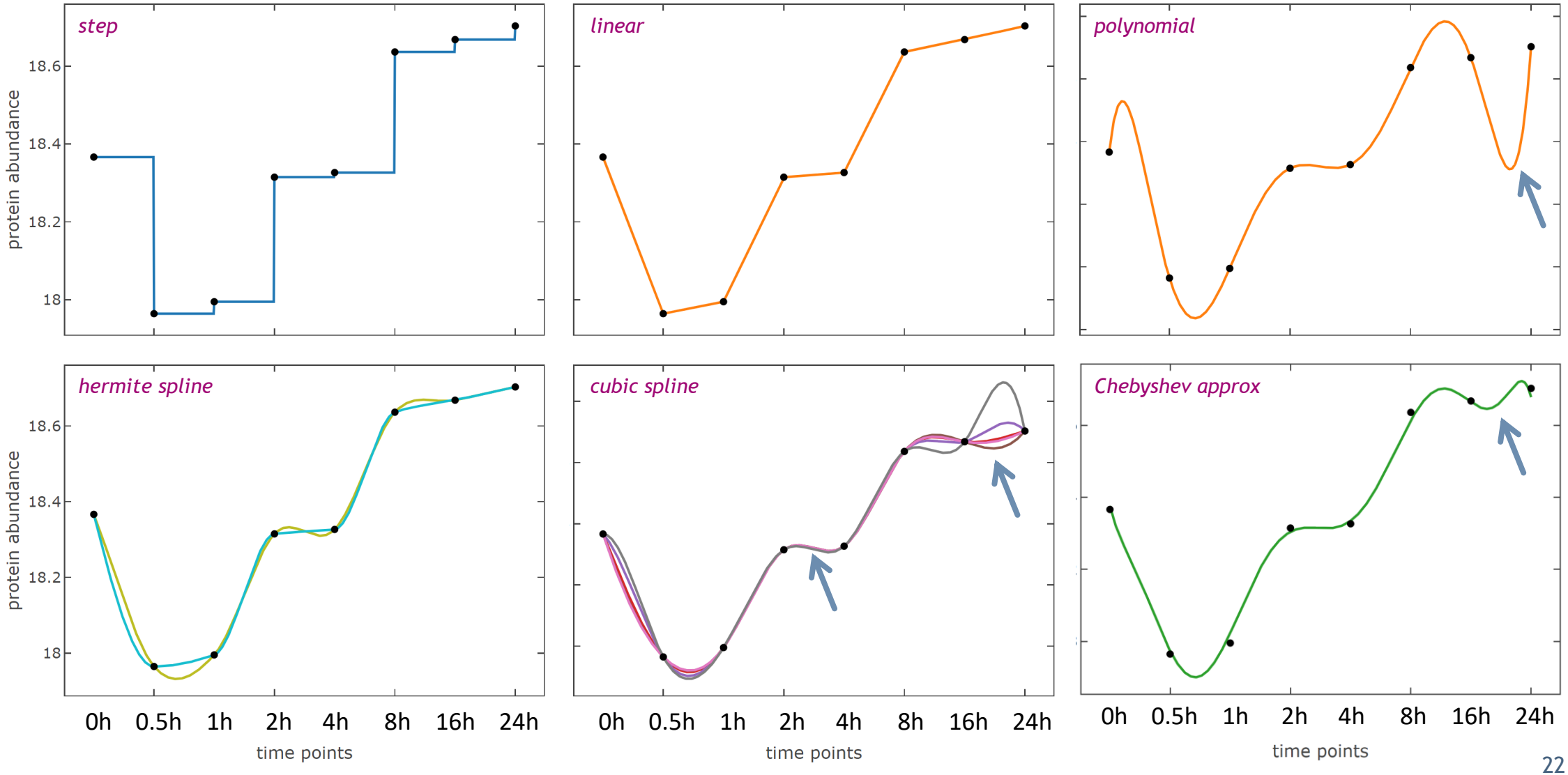

```
open FSharp.Stats
let x,y,slopes = ...
Interpolation.interpolate(x,y,InterpolationMethod.Step)
Interpolation.interpolate(x,y,InterpolationMethod.LinearSpline)
Interpolation.interpolate(x,y,InterpolationMethod.AkimaSubSpline)
Interpolation.interpolate(x,y,InterpolationMethod.CubicSpline CubicSpline.BoundaryCondition.Natural)
Interpolation.interpolate(x,y,InterpolationMethod.HermiteSpline (HermiteMethod.WithSlopes slopes))
Interpolation.interpolate(x,y,InterpolationMethod.Polynomial)
...
```

Back to biology and protein kinetics



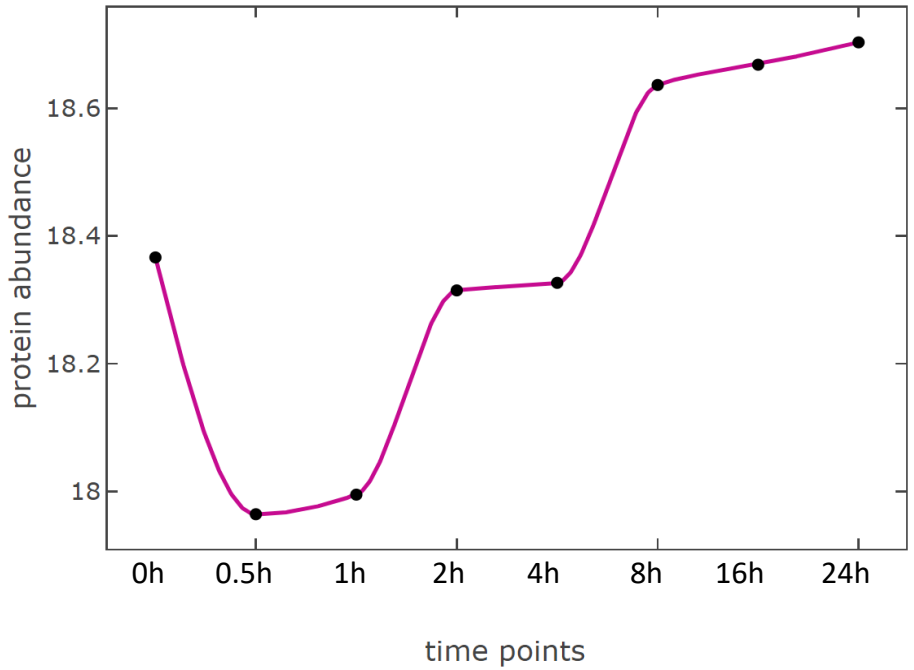


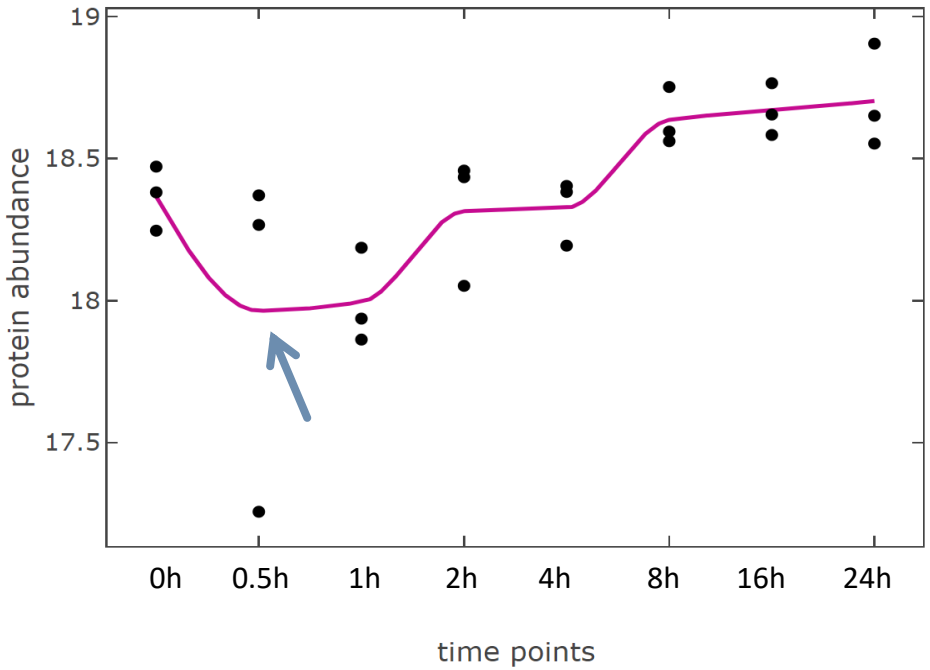
Back to biology and protein kinetics



Back to biology and protein kinetics

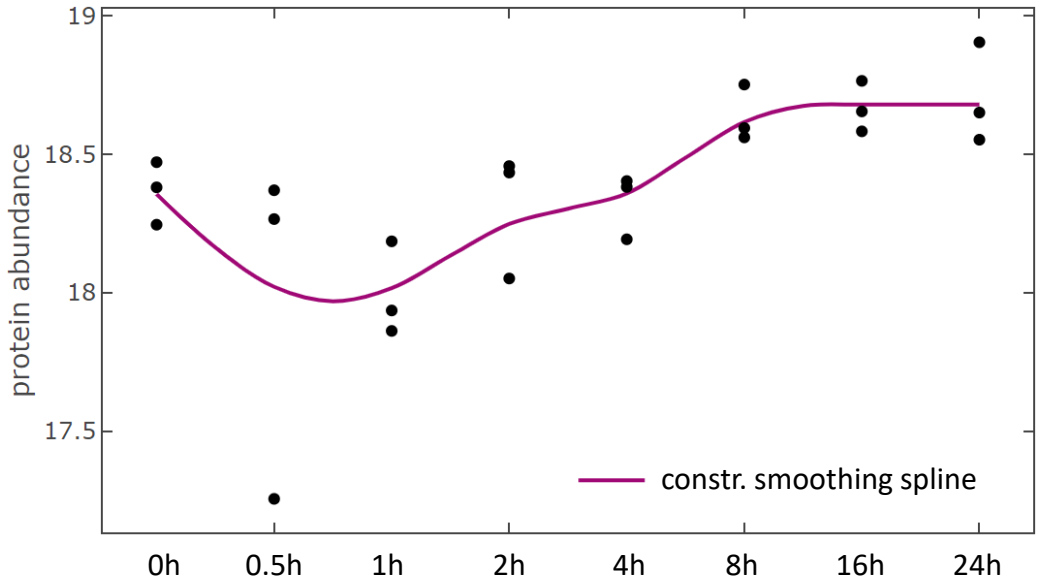
- Necessity to incorporate **smoothness** assumption
- Measurement **variance** serves as important information and should be considered
  - interpolation → regression





Back to biology and protein kinetics

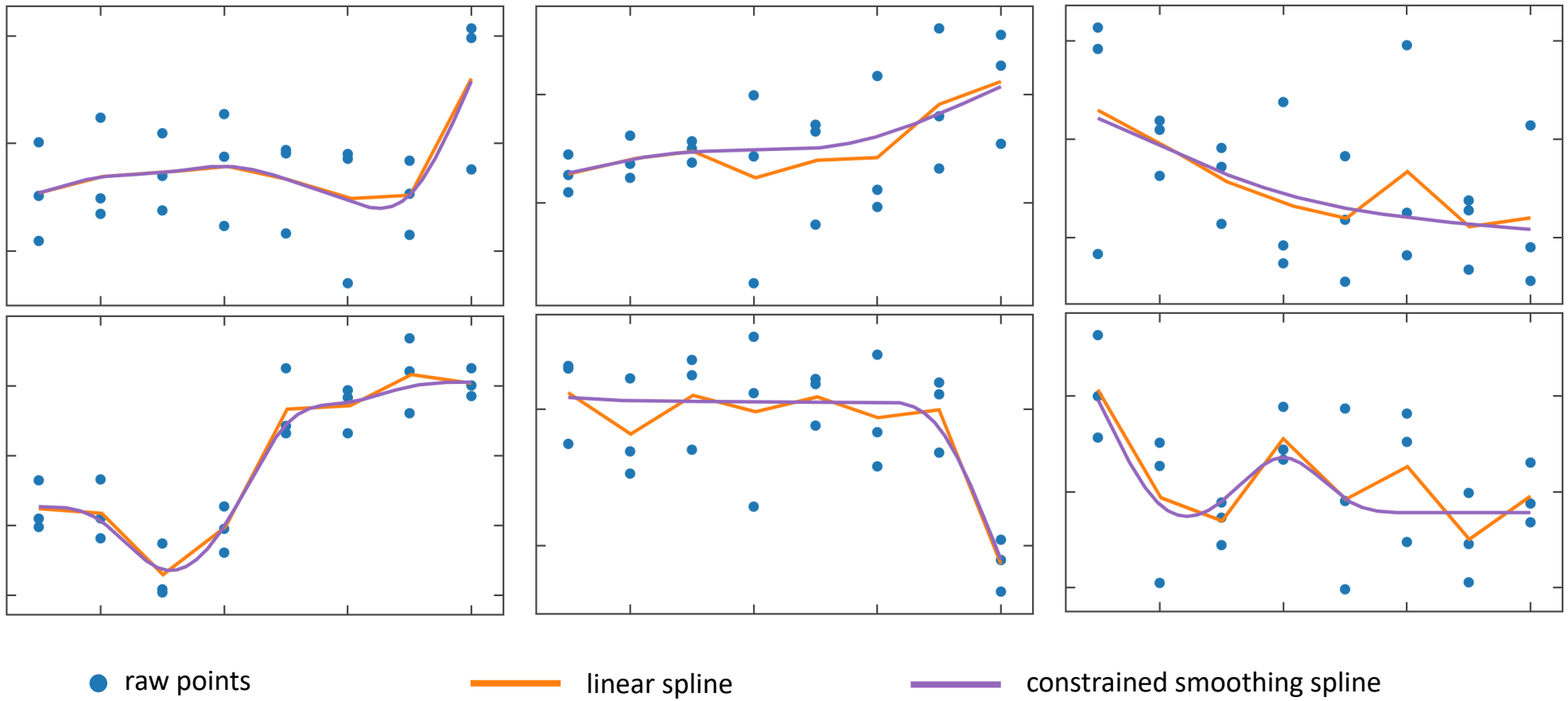
- Necessity to incorporate **smoothness** assumption
- Measurement **variance** serves as important information and should be considered
  - interpolation → regression
- Usage of constrained smoothing splines allows for accurate shape classification

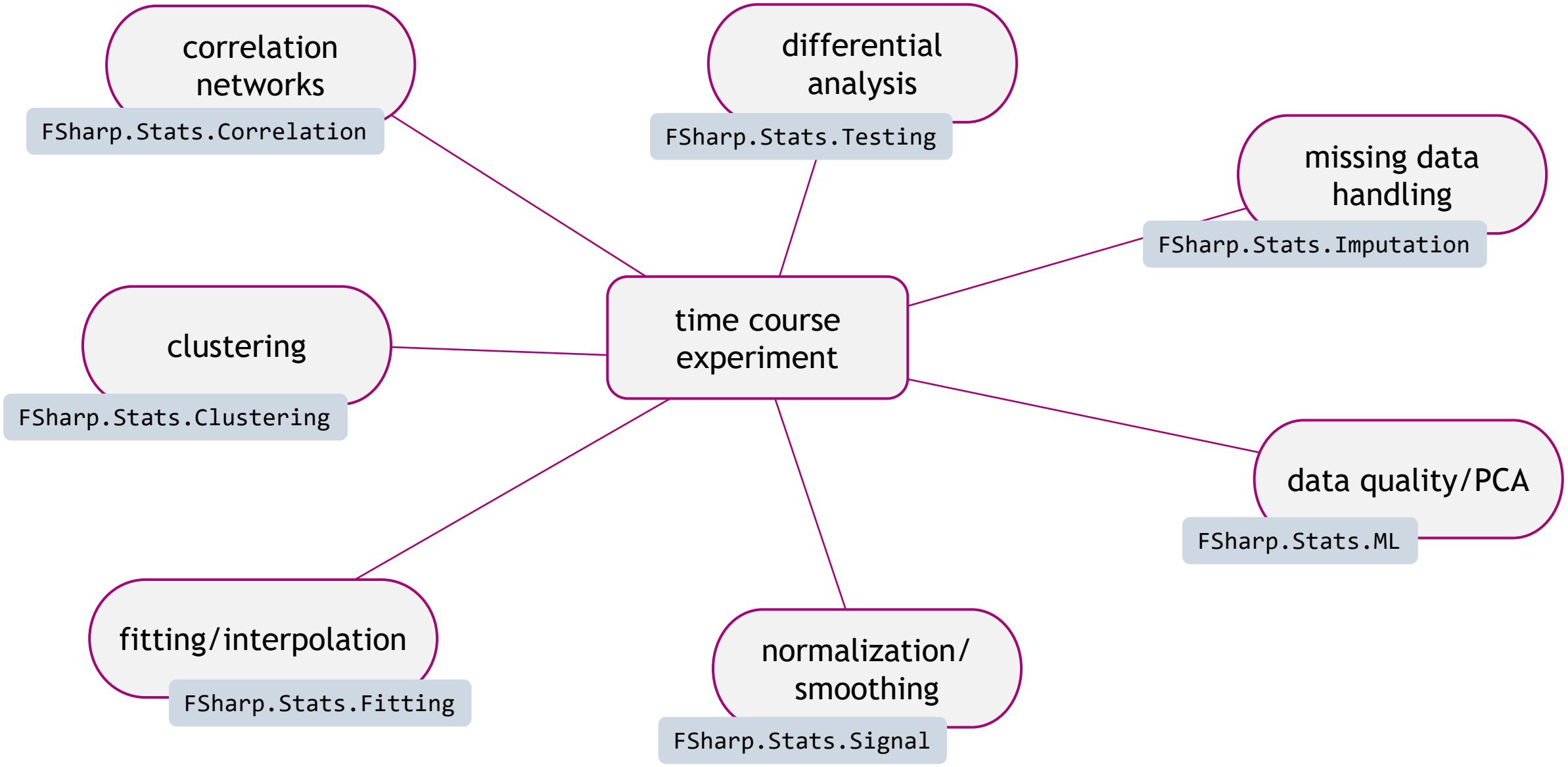


→ class: minimum@0.75 h

Back to biology and protein kinetics

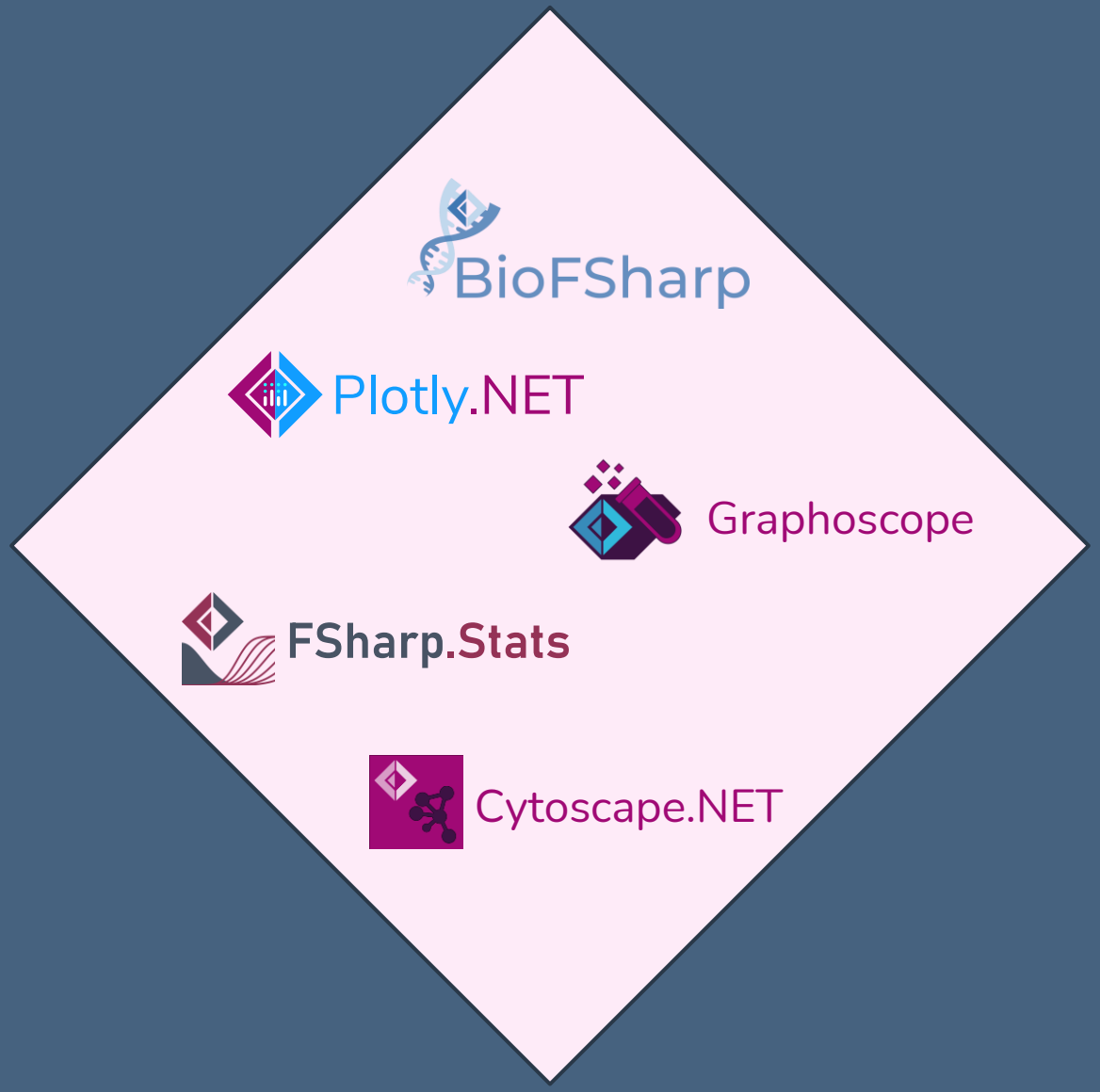
- Usage of constrained smoothing splines allows for accurate shape classification




correlation networks
FSharp.Stats.Correlation
differential analysis
FSharp.Stats.Testing
missing data handling
FSharp.Stats.Imputation
clustering
FSharp.Stats.Clustering
time course experiment
data quality/PCA
FSharp.Stats.ML
fitting/interpolation
FSharp.Stats.Fitting
normalization/ smoothing
FSharp.Stats.Signal

TU RP Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

https://fslab.org/blog/
https://csbiology.github.io/CSBlog/



CSB
COMPUTATIONAL SYSTEMS BIOLOGY