Zeitreihenanalyse der Genexpression bei *Escherichia coli* zur Validierung von RNA-Seq Daten

Kooperation zwischen

- Prof. Dr. rer. nat. habil. T. Villmann, Hochschule Mittweida, *Fachgruppe Mathematik*
- Prof. Dr. rer. nat. habil. R. Wünschiers, Hochschule Mittweida, *Fachgruppe Biotechnologie*
- Prof. Dr. rer. nat. D. Labudde, Hochschule Mittweida, *Fachgruppe Informatik*

HOCHSCHULE MITTWEIDA UNIVERSITY OF APPLIED SCIENCES

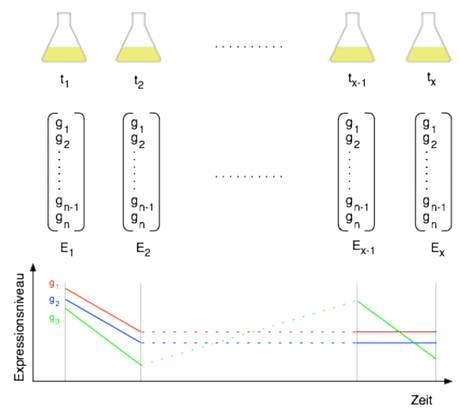
Forschungsgruppe Computational Intelligence

bigM BioInformatics Group Mittweida

Mathematik / Naturwissenschaften / Informatik

Ziel des Projektes:

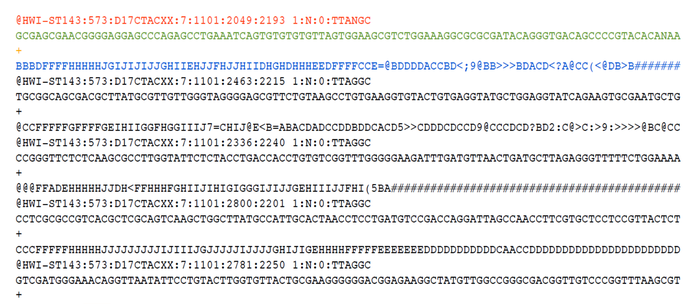
- *Extraktion und Sequenzierung hochwertiger, angereicherter Transkripte aus einer inhomogenen Biogasfermentermatrix)*
- *Identifizierung von Positionen und Promotoren im Chromosom von* E. coli *als Integrationsorte für genetische Konstrukte*
- *Untersuchung des zeitlichen Verlaufes der Genexpression*
- *Cluster- und Klassifizierungsverfahren für Genexpressionsdaten*



Realisierung

Experiment

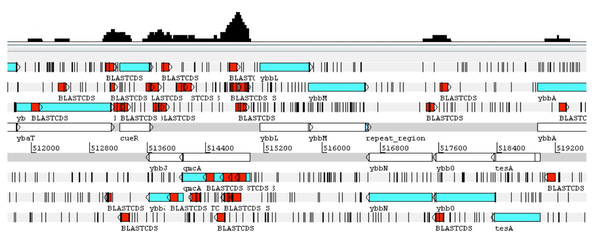
Die Rohdaten: identifier + sequence + line break + sequence quality



Molekularbiologie

Extraktion & Sequenzierung

Information

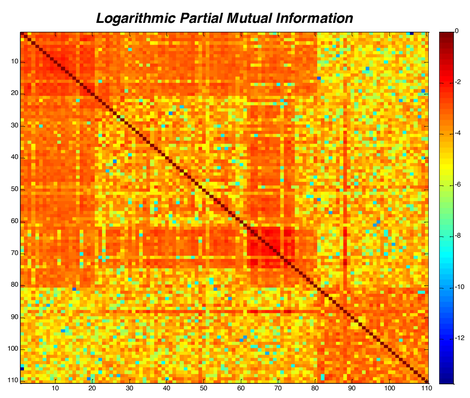


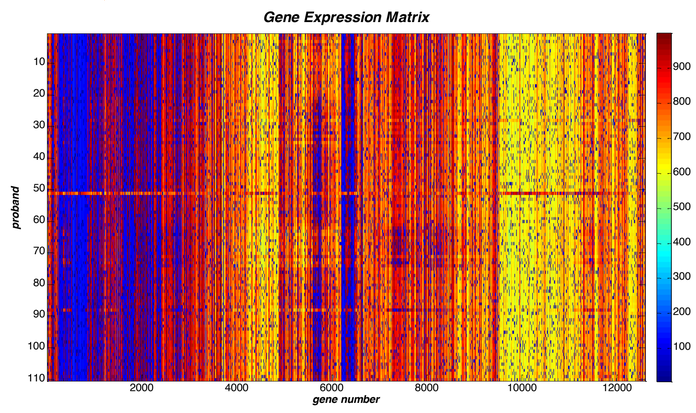
Mapping von Sequenzfragmenten aus der RNA-Sequenzierung (rote Balken, schwarzes Histogram) auf das E. coli Genom. Über die Daten im Histogram kann das Expressionsniveau ermittelt werden

Daten-Analyse

Daten-Aggregierung

Bioinformatik





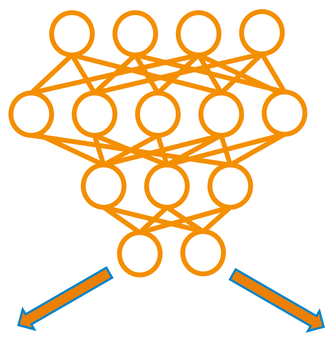
Information

Biomathematik

Ähnlichkeitsbewertung mittels Partial Mutual Information

$$I_R^{\alpha}(X,Y|Z_1\dots Z_k) = \frac{1}{(1-alpha}\int \log\left(\frac{p(x,y,\mathbf{z})^{\alpha}p_{X,Y}(\mathbf{z})^{-(\alpha+1)}}{(p_Y(x,\mathbf{z})p_X(y,\mathbf{z}))^{1-\alpha}}\right)dxdyd\mathbf{z}$$

Selbstlernender Klassifikationsalgorithmus



Prototypbasiertes neuronales Netz (LVQ)

Resultat

Datenklassifikation und Klasseninformation

Kontakte:

Molekularbiologie *wuenschi@hs-mittweida.de*

Bioinformatik *labudde@hs-mittweida.de*

Biomathematik *villmann@hs-mittweida.de*