

Abstract

German

Um die komplexe Vernetzung zellulärer Systeme besser zu verstehen, sind in the letzten Jahren diverse bioinformatische Methoden entwickelt worden, welche Transkriptom- und Metabolomdaten sowohl separat als auch parallel analysieren. In dieser Diplomarbeit werden Korrelationsmatrizen zur Analyse und Integration von Zeitserien-Messungen von Transkriptom- und Metabolomdaten von *Arabidopsis thaliana* verwendet. Die Datensätze umfassen Kälteakklimatisierungs-, CO_2 -Erhöhungs- und Sulfur-Defizienz-Experimente. Das Ziel dieser Diplomarbeit ist es, (1) globale Zusammenhänge zwischen Korrelationen und Molekülfunction zu analysieren und (2) regulatorische Zusammenhänge zwischen und innerhalb dieser zellulären Levels zu entdecken. Um generelle Zusammenhänge zwischen Korrelationswerten und funktioneller Verwandtschaft von Genen/Metaboliten zu bestimmen, wurden globale Korrelationsverteilungen untersucht. Allerdings haben diese Resultate nur eine geringe Aussagekraft zur Bestimmung von spezifischen biologischen Prozessen, welche unter den experimentellen Gegebenheiten dereguliert sind. Daher wurde eine neue Methode entwickelt, welche die Überrepräsentation von hohen Korrelationswerten innerhalb bzw. zwischen Gruppen von Genen/Metaboliten berechnet. Die Gruppierung von Genen und Metaboliten erfolgt durch die Einbindung von Annotationsbibliotheken. Dadurch werden funktionell verwandte Moleküle, die potenziell co- oder entgegengesetzt reguliert werden, identifiziert. Die Resultate der paarweisen Überrepräsentationsanalyse wurden in Form von Netzwerken dargestellt, in denen Knoten mit Annotationsbezeichnungen korrespondieren und Kanten die signifikante Anreicherung von hohen Korrelationswerten indizieren. Die Methode wurde auf Robustheit gegenüber Parametervariationen und Plausibilität der generierten Resultate durch Vergleich mit vorangegangenen Studien überprüft. Im Einklang mit vorhergehenden Studien konnten in allen Datensätzen Änderungen der Aminosäuren-Konzentrationen über die Zeit nachgewiesen werden. Für alle Datensätze wurden weiters Unterschiede in der Genexpression von Photosynthese, im primären Stoffwechsel sowie in der globalen Proteinzusammensetzung (z. B. durch induzierte Wachstumsprozesse) berichtet. Weiters wurden für die CO_2 - und Sulfur-Defizienz-Datensätze Deregulationen, die im Zusammenhang mit pflanzlichen Abwehrprozessen stehen, entdeckt. Insbesondere führte Sulfur-Defizienz zu hohen Korrelationen zwischen Transkripten und Metaboliten in der Biosynthese für Glucosinolat. Die Methode konnte viele dieser zellulären Adaptationsprozesse identifizieren, wodurch die prinzipielle Anwendbarkeit und Nutzbarkeit gezeigt werden konnte.

Stichwörter: Korrelation, Transkriptom, Metabolom, Überrepräsentationsanalyse;

English

To understand more thoroughly the cellular intertwinedness, several computational approaches have recently been proposed in order to analyse and integrate transcriptome and metabolome data sets separately as well as in parallel. In this thesis, correlation matrices were utilized for the analysis and integration of transcriptome and metabolome time-series data sets of *Arabidopsis thaliana*. The data sets comprise of cold acclimation, CO_2 elevation and sulphur starvation experiments. The goals of the thesis are (1) to analyse whether correlation values are in general connected to the underlying function of genes/metabolites and (2) to uncover putative regulatory dependencies among and within the cellular levels. Global correlation distributions were explored to identify a general relationship between correlation value observations and functional dependencies. However, these distributions are inappropriate to reveal particular, deregulated, biological processes under certain experimental settings. In the thesis, a novel method, which determines within or among annotation label enrichment of high correlation values, is proposed. Thereby, functionally related groups of gene/metabolites that are potentially co- or counter-regulated are identified. The results of the pairwise annotation label enrichment analysis are visualized in terms of networks, with nodes corresponding to annotation labels and edges corresponding to significantly enriched of high correlation values. The method was examined for robustness to parameter variations and plausibility of the generated results by drawing comparisons with previous studies. In concordance with previous studies, for all data sets tight correlations among amino acids were detected. Further, for all data sets, transcriptional deregulations of photosynthesis, primary metabolism as well as shifts in the global protein content were in agreement with previous reports. For CO_2 elevation and sulphur deficiency, annotation labels associated with plant defence processes were revealed in line with previous findings. In particular, for sulphur deficiency, tight correlations between metabolite and transcript levels of glucosinolate biosynthesis were revealed. These findings underline the general applicability and usefulness of the method.

Keywords: Correlation, Transcriptome, Metabolome, Enrichment Analysis;