

Transcriptomics

■ **Verantwortlich:** Dr. Oliver Dittrich-Breiholz

Tel.: 0511/532-5814 • E-Mail: dittrich.oliver@mh-hannover.de • www.mh-hannover.de/Transcriptomics.html

Forschungsprofil



Abb. 1: .

Die Research Core Unit Transcriptomics

Die Research Core Unit Transcriptomics (RCUT) steht allen Kliniken und Instituten der MHH zur Durchführung Microarray-basierter Transkriptomanalytik offen. Die beteiligten MitarbeiterInnen verfügen über eine breite Erfahrung rund um Planung, methodische Durchführung, Auswertung und Veröffentlichung von mRNA-Expressionsstudien. Angeboten wird eine große Bandbreite an Microarray-Formaten zur Erfassung von mRNA- oder microRNA-Expression für verschiedene Organismen unter Einsatz der Technologie-Plattformen von Agilent Technologies oder Affymetrix.

Software-Bereitstellung zur Analyse von RNA-Expressionsdaten

Im Jahr 2013 lag einer der Schwerpunkte der Arbeit von RCUT in der Bereitstellung von Software zur Analyse von RNA-Expressionsdaten. Für drei kommerzielle Programme - GeneSpring, Ingenuity Pathway Analysis und Qlucore Omics Explorer - wurden Einzellizenzen angeschafft und ein Outlook-basiertes Buchungssystem wurde etabliert. Nach initialer Registrierung und Buchung können diese Programme von MHH-MitarbeiterInnen zur Auswertung ihrer Expressionsdaten genutzt werden. Neben diesen kommerziellen Lösungen steht ab sofort das erste offizielle Release der in der Core Unit entwickelten Software „RCUTAS“ zur Verfügung.

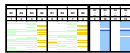



	Qualitätskontrolle	Intuitive Nutzbarkeit	Initiale Datenanalyse / Pilotexperimente	Auswahl verfügbarer Analyse-Optionen	Pathway / Netzwerk-Analysen	Visualisierungs-Optionen	Analyse großer Datensätze
RCUTAS 	+	+	+	+		+	+
GeneSpring Expression 	+		+	+	+	+	+
Qlucore Omics Explorer 	+	+	+	+		+	+
Ingenuity Pathway Analysis 		+			+	+	

Abb. 2: Durch RCUT unterstützte Analyseprogramme und deren jeweilige Stärken.

Erstes Release der Analyse-Software RCUTAS (v1.0)

Die Excel-basierte Analyse-Software RCUTAS (Research Core Unit Transcriptomics Analysis System) stellt eine wichtige Ergänzung zu kommerziell verfügbaren Auswerteprogrammen dar. RCUTAS dient der routinemäßigen Übergabe generierter Expressionsdaten durch RCUT an die Nutzer zum Zwecke weiterführender Analyse. Übergeben wird nicht die Gesamtheit angefallener Rohdaten, sondern ein sorgfältig ausgewählter Datenextrakt. Dieser Extrakt wird in einem durch RCUT-Mitarbeiter durchgeführten Importprozess in die Analyse-Software eingeladen. Auf Basis langjähriger Erfahrung wird hierfür der hinreichend informative Teil der Rohdaten ausgewählt, wobei das Augenmerk auf eine gute Balance zwischen Informationsgehalt und Übersichtlichkeit der übergebenen Daten ausgerichtet ist. Parallel werden Annotations-Informationen zu den gemessenen RNAs zusätzlich in RCUTAS integriert.

Nach Übergabe der somit bereitgestellten Daten eröffnen sich dem Nutzer eine Reihe von Möglichkeiten, den Informationsgehalt der Expressionsdaten zu erschließen. RCUTAS besitzt umfangreiche Auswahloptionen, um neben den eigentlichen Messwerten Zusatzinformationen zu visualisieren (z.B. Annotationen, Qualitätsparameter, ...). Verhältniswerte zu vergleichender Proben (Ratios) können flexibel berechnet, formatiert und angeordnet werden. Weiterhin lassen sich komplexe Filterungen anhand selbst festgelegter Schwellenwerte durchführen und die Ergebnisse als farbkodierte Heatmap exportieren.

Das Grundkonzept von RCUTAS basiert auf 4 Anforderungen, welche die Software gleichzeitig von gängigen kommerziellen - sowie vielen frei verfügbaren Analyse-Programmen abgrenzen.

1) Universelle Verfügbarkeit

RCUTAS läuft unter Microsoft Excel 2003 und 2010. Die implementierten Filter- und Visualisierungsoptionen sind mittels Visual Basic Macros realisiert. Somit kann jeder MHH-Mitarbeiter das Programm unmittelbar nutzen. Einzig die Sicherheitseinstellungen in Excel müssen dahingehend angepasst werden, dass Macros für die Analyse mittels RCUTAS aktiviert bzw. zugelassen werden müssen.

2) Sofortiger Einstieg in Datensichtung

Zum Zeitpunkt der Übergabe an den Nutzer, liegen die Daten einer Studie bereits in umfangreich aufgearbeiteter Form in RCUTAS vor. Wir sprechen hierbei von einer ersten „Konfiguration“ der Studiendaten. Diese erste Konfiguration stellt das Ergebnis einer Vielzahl von Entscheidungen zu den bereits erfolgten Rohdatenprozessierungsschritten dar. Ausgehend von langjähriger Erfahrung und unter Einbeziehung des speziellen Studienhintergrundes werden die Daten durch die RCUT-Leitung (mittels speziell entwickelter Programm-Module) vorformatiert und damit für die RCUTAS-Funktionalität nutzbar gemacht. Genauigkeit, relative Vergleichbarkeit und Datengüte sind zu diesem Zeitpunkt bereits auf einem so hohen Niveau, dass der Nutzer sich sofort (also ab Übergabe) auf inhaltliche Aspekte fokussieren kann und kaum Gefahr läuft, groben Fehlinterpretationen aufgrund inadäquater Rohdatenprozessierung aufzuliegen. Sollte der Nutzer es für erforderlich halten, so kann er beliebige weitere Konfigurationen erstellen und dabei den Fokus seiner Analyse auf andere Schwerpunkte verlagern oder gegebenenfalls andere Datentransformations-Optionen für seine Studie wählen oder ausprobieren.

3) Intuitiver Zugang

Ein wesentlicher Schwerpunkt der RCUTAS-Konzeption liegt im Aspekt einer intuitiven Zugänglichkeit. Nutzern mit fundierten wissenschaftlichen Kenntnissen soll durch RCUTAS ein intuitiver Zugang zu Stärke und Umfang transkriptomweiter Veränderungen ermöglicht werden ohne dezidierte Vorkenntnisse in der Microarray-Datenanalyse vorauszusetzen. Ein hierfür maßgeblicher Unterschied zu den meisten Analyse-Programmen stellt das Format der Messwerte in RCUTAS dar. Es handelt sich um prozessierte Fluoreszenzintensitäten im nicht-logarithmischen Format. Dieser Sachverhalt wird

hier deshalb besonders betont, weil üblicherweise logarithmierte Messwerte verwendet werden, die oftmals zusätzlich einer „Baseline Transformation“ unterzogen werden. Als Resultat derartiger Transformationen ergeben sich Zahlenformate, die relative Veränderungen von Transkripten - wenn auch teilweise mathematisch sinnvoll - intuitiv unanschaulich repräsentieren. Verloren geht der absolute Bezug zum Intensitätsniveau, auf dem sich Veränderungen manifestieren. Konkret bedeutet dies, ein zweifacher Unterschied im Expressionsniveau einer RNA, wird identisch gewichtet, egal ob er durch einen Anstieg von 10 auf 20 - oder von 10000 auf 20000 Einheiten erfolgt – und - das Intensitätsniveau kann retrograd nicht mehr rekapituliert werden. In RCUTAS bleiben die Werte nicht-logarithmiert und können mit einer als „BarGraph-Tool“ bezeichneten Funktionalität außerdem in Form eines Intensitätsprofils über die Gesamtheit aller Samples für jedes Transkript individuell visualisiert und anschließend exportiert werden.

4) Eignung für Pilotexperimente

Niemand wir die Tatsache bestreiten, dass die Verlässlichkeit wissenschaftlicher Daten mit zunehmender Anzahl von Replikaten pro Versuchsbedingung steigt. Dies verhält sich bei Microarray-Studien nicht anders. Eine Herausforderung stellen hier allerdings die vergleichsweise hohen Kosten pro Sample dar. In vielen Fällen ist es sinnvoll, vor der Konzeption einer umfangreichen finalen Studie ein Pilotexperiment durchzuführen. Hierbei gibt es Fälle, bei denen eine (vorläufige) Umsetzung mit $n=1$ Sinn machen kann. Gerade solche Pilot-Experimente lassen sich mit vielen Analyse-Programmen und Standard-Applikationen (z.B. Clustering, Principal Component Analysis, Signifikanztests) kaum sinnvoll auswerten. An dieser Stelle kommen insbesondere die Stärken von RCUTAS zum Tragen. Das RCUTAS-Kernmodul zur Auswertung basiert auf Filteralgorithmen, welche, bei adäquat eingestellten Thresholds, auch bei Experimenten mit $n=1$ hinreichend verlässliche Ergebnisse liefern. Durch die Möglichkeit, sich informative Qualitätsparameter neben den Daten anzeigen zu lassen und/oder diese als Kriterien bei durchgeführten Filterprozessen mit heran zu ziehen, manifestiert sich eine weitere Stärke von RCUTAS. Die heutige Generation von Microarrays weist aufgrund generell hoher Performance einen relativ kleinen Anteil technisch beeinträchtigter Messwerte auf. Diese Beeinträchtigungen sind in der Regel sporadischer- und nicht systematischer Natur. Aus diesem Grund stellen diese bei Studien mit $n=3$ oder höher keine nennenswerten Probleme dar. Gibt es pro Bedingung nur einen einzigen Messwert (wie bei Pilot-Studien) muss diesen technischen Ausreißern bei der Datenanalyse eine höhere Beachtung geschenkt werden – genau dies kann RCUTAS in diesen Fällen gewährleisten.

Publikationen unter Beteiligung der Zentralen Forschungseinrichtung Transcriptomics

Bertram S, Dijkman R, Habjan M, Heurich A, Gierer S, Glowacka I, Welsch K, Winkler M, Schneider H, Hofmann-Winkler H, Thiel V, Pöhlmann S. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J Virol* 2013;87(11):6150-6160

Dhamija S, Winzen R, Doerrie A, Behrens G, Kuehne N, Schauerte C, Neumann E, Dittrich-Breiholz O, Kracht M, Holtmann H. Interleukin-17 (IL-17) and IL-1 activate translation of overlapping sets of mRNAs, including that of the negative regulator of inflammation, MCP1. *J Biol Chem* 2013;288(26):19250-19259

Haas DA, Bala K, Büsche G, Weidner-Glunde M, Santag S, Kati S, Gramolelli S, Damas M, Dittrich-Breiholz O, Kracht M, Rückert J, Varga Z, Keri G, Schulz TF. The Inflammatory Kinase MAP4K4 Promotes Reactivation of Kaposi's Sarcoma Herpesvirus and Enhances the Invasiveness of Infected Endothelial Cells. *PLoS Pathog* 2013;9(11):e1003737

Handschiek K, Beuerlein K, Jurida L, Bartkuhn M, Müller H, Soelch J, Weber A, Dittrich-Breiholz O, Schneider H, Scharfe M, Jarek M, Stellzig J, Schmitz ML, Kracht M. Cyclin-Dependent Kinase 6 Is a Chromatin-Bound Cofactor for NF-kappaB-Dependent Gene Expression. *Mol Cell* 2014;53(2):193-208

Scharf M, Neef S, Freund R, Geers-Knörr C, Franz-Wachtel M, Brandis A, Krone D, Schneider H, Groos S, Menon MB, Chang KC, Kraft T, Meissner JD, Boheler KR, Maier LS, Gaestel M, Scheibe RJ. Mitogen-Activated Protein Kinase-Activated Protein Kinases 2 and 3 Regulate SERCA2a Expression and Fiber Type Composition To Modulate Skeletal Muscle and Cardiomyocyte Function. *Mol Cell Biol* 2013;33(13):2586-2602

Tran DD, Saran S, Dittrich-Breiholz O, Williamson AJ, Klebba-Färber S, Koch A, Kracht M, Whetton AD, Tamura T. Transcriptional regulation of immediate-early gene response by THOC5, a member of mRNA export complex, contributes to the M-CSF-induced macrophage differentiation. *Cell Death Dis* 2013;4:e879