

26. HÜLSENBERGER GESPRÄCHE

Konzepte der genomweiten molekularen Analyse und Datenintegration in der Biologie

Prof. Dr. Wolfram Weckwerth, Department für Ökogenomik und Systembiologie, Universität Wien, Österreich

Systembiologie ist eine moderne Entwicklung in der Biologie, die genomweite molekulare Analysen, z.B. Metabolomics, Proteomics und/oder Transkriptomics, mit Computer-gestützten mathematischen und statistischen Modellen verknüpft, um einerseits kausale Mechanismen vom Molekül zum Organismus abzuleiten und andererseits Vorhersagemodelle für die Merkmalsausprägung bzw. die Genotyp-Phenotyp-Beziehung zu erhalten. Diesen Ansatz hat bereits Ludwig von Bertalanffy 1944 in seinem Buch „Vom Molekül zur Organismenwelt“ angedeutet [1]. In einer seiner wichtigsten Publikationen „Der Organismus als physikalisches System betrachtet“ beschreibt Bertalanffy bereits 1940 die mathematische Modellierung eines sich selbst regulierenden Systems von biochemischen Pathways eines „offenen“ Organismus [2]. In den folgenden Jahren hat Bertalanffy seine Theorie verallgemeinert und als auf alle komplexen nicht-linearen Systeme, z.B. der Biologie, Ökologie oder auch Ökonomie anwendbare „Allgemeine Systemtheorie“, definiert [3].

Die technischen Beschränkungen Bertalanffys waren zu seiner Zeit im Hinblick auf unsere Möglichkeiten heutzutage schier unüberwindbar: Das System, mit dem er sich 1940 auseinandersetzte, bestand aus vier (fünf) Komponenten mit vier Differentialgleichungen. Dieses System lässt sich analytisch beschreiben und man kann daraus grundlegende Prinzipien der Selbstregulation oder Selbstorganisation ableiten [2].

Wo stehen wir heutzutage mit einem „typischen“ Organismus, Pflanze oder Tier (Mensch)? Nach der Erkenntnis der molekularen Struktur und Prinzipien der Informationsspeicherung der DNA 1953 [4], dem zentralen Dogma der Molekularbiologie von Crick 1973 [5] und der rapiden Entwicklung von „next generation sequencing (NGS)“ [6] seit der Publikation des ersten Humangenoms und des ersten Pflanzengenoms 2000 sind zur Zeit ca. 80.000 Genomprojekte und deren Datenbanken verfügbar. Eine genomische Rekonstruktion eines typischen tierischen oder pflanzlichen Stoffwechsels umfasst ca. 2.500 Reaktionen und wesentlich mehr kaum abzuschätzende metabolische Komponenten [7, 8]. In anderen Worten, wir müssen mindestens 2.500 vernetzte Reaktionen als Differentialgleichungen darstellen und modellieren, um eine kausale Verknüpfung des Systems zu beschreiben, bzw. Vorhersagen des dynamischen molekularen Phenotyps aus der Genomsequenz abzuleiten. Diese Art von mathematischer Beschreibung eines komplexen nicht-linearen Systems ist erst mit Hilfe der hochmodernen Computertechnologie, lange nach Bertalanffy, möglich geworden und man ist heutzutage in der Lage, solche Systeme numerisch zu lösen bzw. zu approximieren [7]. Desweiteren haben sich Technologien für die genomweite molekulare Analyse entwickelt, von denen Bertalanffy keine Vorstellungen haben konnte: RNAseq, Proteomics und Metabolomics [9]. Diese bioanalytischen Verfahren orientieren sich an dem molekularen Dogma von Crick und sind in der Lage, hochkomplexe Gemische aus Transkripten (RNAseq), Proteinen (Proteomics) und Metaboliten (Metabolomics) aufzulösen, viele Komponenten zu identifizieren und zu quantifizieren. Schließlich werden in diesen Daten Interaktionen der molekularen Komponenten gesucht, die molekulare oder auch andere phenotypische Merkmale erklären können. Hierzu werden hochkomplexe, multivariate statistische Verfahren eingesetzt, die letztendlich eine Datenintegration und -interpretation ermöglichen [10]. In einem letzten Schritt müssen diese statistischen Modelle mit den mathematischen Modellen verknüpft werden, um aussagekräftige Genotyp-Phenotyp-Modelle zu generieren [9, 11].

In meinem Vortrag werde ich Methoden der molekularen Hochdurchsatzanalyse vorstellen sowie einige Aspekte der mathematischen und statistischen Modellierung von molekularen Hochdurchsatzdaten und deren Verknüpfung mit genomweiten biochemischen Netzwerken erläutern.

26. HÜLSENBERGER GESPRÄCHE

References

- [1] Bertalanffy, L. v., Vom Molekül zur Organismenwelt. *Akademische Verlagsgesellschaft Athenaion, Potsdam* 1944.
- [2] Bertalanffy, L. v., Der Organismus als physikalisches System betrachtet. *Naturwissenschaften* 1940, 33, 522-531.
- [3] Bertalanffy, L. v., General System Theory. *George Braziller, New York* 1969.
- [4] Watson, J. D., Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953, 171, 737-738.
- [5] Crick, F., Central dogma of molecular biology. *Nature* 1970, 227, 561-563.
- [6] Metzker, M. L., Sequencing technologies - the next generation. *Nat Rev Genet* 2010, 11, 31-46.
- [7] Nagele, T., Mair, A., Sun, X., Fragner, L., Teige, M., Weckwerth, W., Solving the differential biochemical Jacobian from metabolomics covariance data. *Plos One* 2014, 9, e92299.
- [8] Nagele, T., Weckwerth, W., Mathematical modeling of plant metabolism-from reconstruction to prediction. *Metabolites* 2012, 2, 553-566.
- [9] Weckwerth, W., Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J Proteomics* 2011, 75, 284-305.
- [10] Weckwerth, W., Morgenthal, K., Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 2005, 10, 1551-1558.
- [11] Weckwerth, W., Unpredictability of metabolism--the key role of metabolomics science in combination with next-generation genome sequencing. *Anal Bioanal Chem* 2011, 400, 1967-1978.