

Die Aktivität von kleinen Molekülen im Rahmen der Genomik verstehen lernen**

Jun-Seok Lee und Young-Tae Chang*

Stichwörter:

Analytische Methoden · Genexpression · Genomik · Konnektivitätskarte · Molekülprofile

Techniken, um mit kleinen Molekülen biologische Reaktionswege gezielt zu verändern, haben für die Entdeckung und Entwicklung von Arzneien eine große Bedeutung. Wegen der Komplexität von biologischen Systemen können die biologische Aktivität und die Ziele von einzelnen Molekülen jedoch nur durch mühselige und zeitraubende Versuche ermittelt werden. Diese Problematik veranlasste die Forscher, Verfahren zur Hochdurchsatzanalyse der Aktivität kleiner Moleküle zu entwickeln. Nach einem Jahrzehnt Forschung können wir die Funktion von kleinen Molekülen in Bezug auf Krankheiten und Genexpression wesentlich besser verstehen. Als neues vielversprechendes Verfahren zur Erstellung von Expressionsprofilen für kleine Moleküle stellten Golub et al. kürzlich die „Konnektivitätskarte“ vor.^[1–3] Dieser Karte – oder Datenbank – wird unter den bislang verwendeten Methoden momentan das größte Potenzial zugesprochen dabei zu helfen, das Zusammenspiel von Medikament und Medikament, Medikament und Gen sowie Medikament und Krankheit besser verstehen zu lernen.

Vor dieser Konnektivitätskarte war der „US National Cancer Institute 60 Drug Screen“ (NCI60) die umfangreichste Plattform, um die Aktivität von

kleinen Molekülen systematisch zu untersuchen. Seit seinem Start in den späten 1980er Jahren wurde ein breites Spektrum an synthetischen kleinen Molekülen und Naturstoffen in zahlreichen Krebszelllinien durchmustert. Die erhaltenen Daten lieferten wertvolle Informationen über die Aktionsweise der getesteten Verbindungen.^[4,5] Beim NCI-Verfahren wird die Cytotoxizität an 60 verschiedenen Krebszelllinien gemessen. Aus diesen Daten werden dann Muster der Aktivität kleiner Moleküle erstellt, die mithilfe von Prioritätenkarten („Heat Maps“) und hierarchischer Clusterbildung ausgewertet werden.^[6] Eine Einschränkung dieser Analyse von Molekülprofilen liegt darin, dass das Muster nur auf einer einzigen finalen phänotypischen Ausprägung im Zellsystem beruht. Gesamt muster weisen höchstwahrscheinlich koordinierte Fingerabdrücke für jede Substanz auf, die aber nicht direkt mit dem biologischen Reaktionsweg verknüpft sind.

Im Jahr 2000 stellten Hughes et al. ein Kompendium-Verfahren vor, bei dem für eine Profilerstellung anstelle der Cytotoxizität die Genexpressionssignatur verwendet wird.^[7] Nach Vorstellung der Autoren lässt sich das gesamte Genexpressionsprofil als Kompendium des Zellzustands auffassen, ein Konzept, das die Autoren zumindest an Hefe verifizieren konnten. Ermutigt durch diese ersten Studien, entwickelten Golub et al. ein robusteres System (die so genannte Konnektivitätskarte), in dem die Genexpressionssignatur von kleinen Molekülen in Säugerzellen mithilfe einer neuartigen Datenverarbeitungsme thode analysiert wird.^[1]

Die technische Seite des Konnektivitätskarten-Systems ist einfach. Die

Autoren wählten 164 kleine Moleküle aus. Jedes dieser Moleküle wurde 6 h in einer bestimmten Konzentration (meist 10 µM) in Säugerzellen (hauptsächlich MCF7-Brustkrebszellen) inkubiert. Alle Genexpressionsmuster wurden verglichen, und die Aktivitäten jeder Substanz in Form einer Rangliste von 22 283 Genen dargestellt. Nach Einrichtung der Referenzdatenbank (453 Einzelbeispiele) führten die Forscher eine Datenbankabfrage durch, um den Grad der Konnektivität zu errechnen. Die Abfragesignatur und die Beziehungen der Aktivitäten der Referenzmoleküle untereinander wurden als Balkendiagramm dargestellt, das die Rangfolge der Vergleichssubstanzen nach ihrer Konnektivität angibt (Abbildung 1).

Die Konnektivitätskarte hat fünf klare Pluspunkte:

- 1) Zur Profilermittlung dient die Genexpressionssignatur, was einige Vorteile gegenüber der NCI-Methode bietet. Der mRNA-Expressionsgrad steht in direkter Beziehung zu den biologischen Reaktionswegen, und das Genexpressionsprofil selbst kann zur Untersuchung von Genfunktionen dienen. Allerdings ist die Handhabung dieser Profildaten nicht einfach, da eine enorme Datensmenge bewältigt werden muss.^[8]
- 2) Die Musteranalyse nutzt einen nichtparametrischen, rangfolgebasierten Musterabgleich durch Kolmogorov-Smirnov-Statistik. Der Hauptunterschied zwischen der konventionellen hierarchischen Methode und der neuen Methode (Gen Set Enrichment Analysis, GSEA)^[9] liegt darin, dass die erste den Expressionsunterschied („Fold Change Value“) analysiert, die zweite jedoch

[*] J.-S. Lee, Prof. Dr. Y.-T. Chang
Department of Chemistry
New York University
New York, NY 10003 (USA)
Fax: (+1) 212-995-4203
E-Mail: yt.chang@nyu.edu

[**] J.-S.L. wurde durch einen Korea Research Foundation Grant finanziell unterstützt, dessen Mittel vom Korean Government MOEHRD Basic Research Promotion Fund kommen (KRF-2005-C00088).

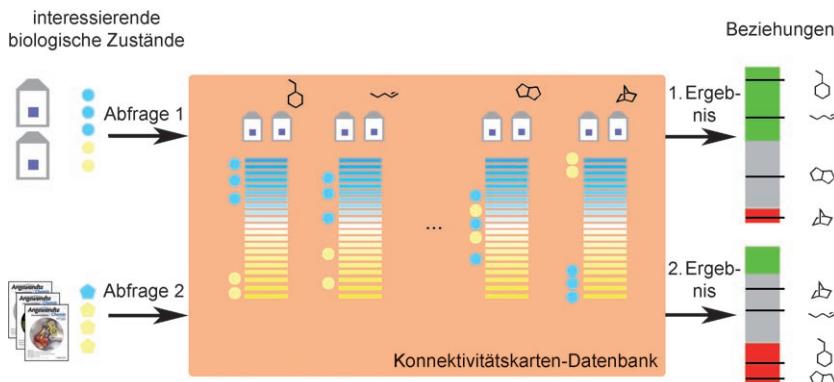


Abbildung 1. Darstellung des Konnektivitätskarten-Abfrageverfahrens (leichte Abänderung der Originalabbildung^[1]). Eine Abfragesignatur kann nicht nur aus Mikroarray-Experimenten erstellt werden (Abfrage 1), sondern auch aus veröffentlichten Informationen zu Genregulationsmustern (Abfrage 2).

die Einstufung in einer Genliste. Konventionelle Methoden zur Musteranalyse an Daten aus Mikroarray-Experimenten konzentrieren sich zumeist auf die Gene, bei denen extrem große Expressionsunterschiede auftreten. Diese Expressionsunterschiede gehen dann direkt in die Mustererkennung ein.^[10] Allerdings kann der Expressionsunterschied von Experiment zu Experiment unterschiedlich sein, sodass die Clusterbildungsergebnisse durch experimentelle Abweichungen beeinträchtigt werden. Dieses durch die experimentellen Bedingungen ausgelöste Rauschen kann drastisch reduziert werden, wenn anstelle der absoluten Expressionsunterschiede das Rangfolgesystem der Konnektivitätskarte eingesetzt wird, welches das Vergleichen von Datensätzen aus unterschiedlichen experimentellen Ansätzen möglich macht. Hierarchische Verfahren haben außerdem den Nachteil, dass es keine Standardmöglichkeit gibt, Entscheidungslinien zu bestimmen. So lässt sich bei einer hierarchischen Prioritätenkarte zur Clusterbildung kaum sagen, wo der Grenzwert für eine positive Korrelation liegt. Bei der Konnektivitätskarte wird dagegen immer ein Konnektivitätsmaß ausgegeben. Dessen Vorzeichen gibt an, ob zwei Signaturen positiv oder negativ korreliert sind (Null bedeutet keine Korrelation).

3) Die Datenbankabfrage ist bei der Konnektivitätskarte unabhängig von der experimentellen Plattform. Für

die Abfragesignatur sind nur zwei Listen von Gensätzen notwendig (Satz hochregulierter Gene und Satz herabregulierter Gene). Nicht nur die Ergebnisse von Mikroarray-Experimenten, sondern auch die von „nassen“ biologischen Experimenten können für eine Abfrage herangezogen werden.

- 4) Mit dieser Methode ist es möglich, kleine Moleküle, Gene und Krankheiten in Form von Genexpressionsprofilen umfassend zu verknüpfen. Dies bedeutet eine erhebliche konzeptionelle Ausweitung der NCI-Analyse, denn diese gestattet nur einen Vergleich zwischen Substanzen.
- 5) Schließlich sind sämtliche Vergleichsdaten und Analysentechniken über das Internet zugänglich (<http://www.broad.mit.edu/cmap/>). Dies hat allerdings weniger mit der eigentlichen Technik zu tun als vielmehr mit der Sichtweise der Autoren. Die Allgemeingültigkeit, d. h. das Tolerieren vieler unterschiedlicher experimenteller Formate, und die interaktiven, offenen Ressourcen sind große Pluspunkte der Konnektivitätskarte.

Mit zwei Fragen evaluieren die Forscher die Möglichkeiten der Konnektivitätskarte: 1) Kann die Konnektivitätskarte die Aktivität von kleinen Molekülen richtig vorhersagen? 2) Kann dieses Modellsystem Zusammenhänge zwischen Krankheitsbild und kleinen Molekülen herausfinden?

Die Antwort auf die erste Frage gab den Forschern schon der erste Datensatz, der eine hohe Korrelation zwischen drei bekannten Histon-Desacetylase-(HDAC)-Inhibitoren und ebenso eine Korrelation zwischen Östrogenrezeptor(ER)-Agonisten und -Antagonisten zeigte. Dies spricht dafür, dass die Genexpressionssignatur eine ganz ausgezeichnete Sonde sein kann, um ein Profil von Molekülaktivitäten zu erstellen. Für das große Spektrum an potentiellen Nutzern gibt es jedoch noch andere Faktoren zu bedenken. Die Autoren wiesen in ihrem Bericht darauf hin, dass die Inkubationszeit, der Zelltyp im Assay und auch die Medikamentendosis die Genexpressionssignatur ganz erheblich beeinflussen. Genauer gesagt unterscheiden sich einige Zelltypen in der Proteinexpression. Obwohl für die Signaturen von HDAC-Inhibitoren, die aus fünf verschiedenen Zelllinien stammen, hohe Konnektivitäten beobachtet werden, können in manchen Fällen die unterschiedlichen Bedingungen der Zelllinien das Profil der Molekülaktivität beeinflussen. Nachgewiesen wurde dies im Fall der Signatur des ER-Liganden für PC3- und HL60-Zellen, die keine ERs exprimieren.^[1] Die Autoren betonten, dass das Konnektivitätsmaß von Östradiol (einem ER-Agonisten) in ER-expressionsfreien Zelllinien die Notwendigkeit eines erweiterten Vergleichsdatensatzes belegt.

Ein ebenfalls schwieriges Thema sind die Dosierung der Substanzen und die Inkubationszeit. Bei chemisch induzierten Änderungen in den interzellulären Signalwegen handelt es sich um dynamische Phänomene. Bei Mikroarray-Studien zu Dosierungen und Inkubationszeiten wurde beobachtet, dass ein kleines Molekül je nach Inkubationszeit und Dosierung unterschiedliche Genexpressionssignaturen entwickeln kann.^[11,12] Um die Bedingungen zu optimieren, damit eine Signatur allgemeingültig werden kann, sind daher noch einige Anstrengungen nötig (Tabelle 1).

Im Hinblick auf die zweite Frage wird als besonderes Merkmal der Konnektivitätskarte das offene Abfragesystem genannt. Die Autoren versuchten, aus genetischen Studien zu bekannten Krankheiten die Daten zu sammeln, und erzeugten dann Abfragesignaturen für

Tabelle 1: Vergleich von NCI60-Methode und Konnektivitätskarte.

Merkmal	NCI60-Methode	konventionelles Genexpressionsprofil	Konnektivitätskarte
erwartetes Ergebnis	Substanzvergleich	Substanzvergleich, Identifizierung des Signalwegs	Identifizierung der Verknüpfung Medikament-Genkrankheit und des Signalwegs
Analysenparameter	Cytotoxizität	verschiedene Genaordnungen	Expressionsprofil von 22 283 Genen (Affymetrix-Array)
Abfragequelle	GI50-Daten von 60 Krebszellen	Genexpressionänderung	Paar von Genlisten (aufwärts und abwärts) aus mehreren Formaten
Analysenmethode	hierarchische Analyse	hierarchische Analyse	nichtparametrische, rangfolgebasierte statistische Analyse
Vorteil	großer Vergleichsdatensatz	informationsreich	geringes Rauschen
Nachteil	begrenzte Information	Unterschiede je nach Ansatz (großes Rauschen)	weitere Validierung erforderlich

Modellkrankheiten wie die ernährungsinduzierte Fetsucht und Alzheimer. Die Alzheimer-Abfragen aus zwei unabhängigen Quellen ergaben eine statistisch signifikante negative Konnektivität für 4,5-Dianilinophthalimid (DAPH). Einer unabhängigen Studie zufolge reduziert DAPH *in vitro* spezifisch den Gehalt an β -Faltblättern bei aggregierenden A β 1-42-Peptiden.^[13] Die Ergebnisse von Golub et al. stützen somit die Annahme, dass DAPH eine neue Leitstruktur für Wirkstoffe gegen Alzheimer sein könnte. Durch weitere praktische Anwendungen der Konnektivitätskarte konnten auch eine neuartige Klasse von Modulatoren des Reaktionswegs von Heat-Shock-Protein 90 (Gedunin und Celastrol)^[2] und die Medikamentenresistenz von Rapamycin bei akuter lymphoblastischer Leukämie identifiziert werden.^[3] Diese Beispiele zeigen, wie die Konnektivitätskarte dabei helfen kann, mithilfe von Genexpressionssignaturen Vorhersagen zu Zielen in der chemischen Genetik zu treffen.

Die aktuellen Versionen der Konnektivitätskarte sehen keine besondere „Form“ wie eine minimale Zahl an Tags für die Abfragesignatur vor, was volle Flexibilität in den unterschiedlichen Anwendungsbereichen garantiert. Jede Forschungsgruppe kann ausgehend von der gegebenen Datenbank ihre eigene Anwendungsmethode entwickeln. Die Konnektivitätskarte ist zudem nicht auf eine bestimmte experimentelle Plattform beschränkt, und sämtliche Datenbanken sind frei zugänglich. Um die Si-

gnatur- und Profildatenbank von kleinen Molekülen weiter auszubauen, sind allerdings gemeinsame Anstrengungen und eine gute Zusammenarbeit der Forscher unabdingbar. Es verbleiben unter anderem die Aufgaben, eine umfassende Datensammlung zu Medikamenten-Screenings aufzubauen sowie die Dosis-Zeit-Abhängigkeit und die Zelltypenspezifität zu bestimmen. Sobald die Konnektivitätskarte aber etabliert und populär geworden ist, haben wir ein Werkzeug in der Hand, dessen universelle Sprache hilft, die Aktivität von kleinen Molekülen in Rahmen der Genomik besser zu verstehen.

Online veröffentlicht am 3. April 2007

- [1] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, *Science* **2006**, *313*, 1929–1935.
- [2] H. Hieronymus, J. Lamb, K. N. Ross, X. P. Peng, C. Clement, A. Rodina, M. Nieto, J. Du, K. Stegmaier, S. M. Raj, K. N. Maloney, J. Clardy, W. C. Hahn, G. Chiosis, T. R. Golub, *Cancer Cell* **2006**, *10*, 321–330.
- [3] G. Wei, D. Twomey, J. Lamb, K. Schlis, J. Agarwal, R. W. Stam, J. T. Opferman, S. E. Sallan, M. L. den Boer, R. Pieters, T. R. Golub, S. A. Armstrong, *Cancer Cell* **2006**, *10*, 331–342.
- [4] K. D. Paull, R. H. Shoemaker, L. Hodes, A. Monks, D. A. Scudiero, L. Rubinstein, J. Plowman, M. R. Boyd, *J. Natl. Cancer Inst.* **1989**, *81*, 1088–1092.
- [5] J. N. Weinstein, K. W. Kohn, M. R. Grever, V. N. Viswanadhan, L. V. Rubinstein, A. P. Monks, D. A. Scudiero, L. Welch, A. D. Koutsoukos, A. J. Chiausa, K. D. Paull, *Science* **1992**, *258*, 447–451.
- [6] J. N. Weinstein, T. G. Myers, P. M. O'Connor, S. H. Friend, A. J. Forname, Jr., K. W. Kohn, T. Fojo, S. E. Bates, L. V. Rubinstein, N. L. Anderson, J. K. Buolamwini, W. W. van Osdol, A. P. Monks, D. A. Scudiero, E. A. Sausville, D. W. Zaharevitz, B. Bunow, V. N. Viswanadhan, G. S. Johnson, R. E. Wittes, K. D. Paull, *Science* **1997**, *275*, 343–349.
- [7] T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stephanants, D. D. Shoemaker, D. Gachotte, K. Chakraburty, J. Simon, M. Bard, S. H. Friend, *Cell* **2000**, *102*, 109–126.
- [8] D. E. Bassett, Jr., M. B. Eisen, M. S. Boguski, *Nat. Genet.* **1999**, *21*, 51–55.
- [9] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomerooy, T. R. Golub, E. S. Lander, J. P. Mesirov, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550.
- [10] D. K. Slonim, *Nat. Genet.* **2002**, *32 Suppl.*, 502–508.
- [11] Y. Zhou, F. G. Gwadry, W. C. Reinhold, L. D. Miller, L. H. Smith, U. Scherf, E. T. Liu, K. W. Kohn, Y. Pommier, J. N. Weinstein, *Cancer Res.* **2002**, *62*, 1688–1695.
- [12] J. N. Weinstein, Y. Pommier, *Nat. Biotechnol.* **2006**, *24*, 1365–1366.
- [13] B. J. Blanchard, A. Chen, L. M. Rozenboom, K. A. Stafford, P. Weigle, V. M. Ingram, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14326–14332.