

Eine genomumspannende Korrelation von mRNA und Proteinen einer Einzelzelle**

Edward S. Yeung*

Einzelmoleküluntersuchungen · Einzelzellen ·
Genexpression · mRNA · Proteine

Seit der Vollendung des Humangenomprojekts vor rund zehn Jahren gab es große Fortschritte hinsichtlich der Analyse und des Verständnisses von Transkriptom und Proteom. Letztlich ist die DNA ja nicht mehr als eine Matrize, die systematisch in mRNA umgewandelt wird, die wiederum die Proteinsynthese und damit das Leben selbst steuert. Eine Hauptschwierigkeit bei Arbeiten zu Transkriptom und Proteom besteht darin, dass die Genexpression in jeder Zelle einzigartig ist. Selbst innerhalb eines Gewebes tragen das Alter der Zelle, der Zellzyklus, die Umgebung, die Signaltransduktion und andere stochastische Ereignisse zu dieser Vielgestaltigkeit bei. Da die Nachweisgrenzen der gewöhnlichen Analysemethoden für eine umfassende Quantifizierung der mRNAs und Proteine im Regelfall eine große Zahl an Zellen erfordern und es so etwas wie die „Durchschnittszelle“ nicht gibt, ist unser Verständnis des biologischen Systems unvollständig. Fortschritte im Bereich der Transkriptomik und Proteomik auf der Ebene der Einzelzelle sind daher dringend erforderlich.

Die Erfassung von Nucleinsäuren und Proteinen in Einzelzellen reicht ein halbes Jahrhundert zurück. Der Gesamtnucleinsäuregehalt wurde durch Absorptionsphotometrie ermittelt,^[1] und mehrere Hämoglobine – die Hauptproteine roter Blutkörperchen – wurden auf Einzelzellniveau mithilfe der Faserelektrophorese^[2] und später durch Kapillarelektrophorese^[3] getrennt und identifiziert. Lediglich die Hauptbestandteile jedes Typs zu erfassen ist jedoch weit von der umfassenden „Omik“, die eine Zelle definiert, entfernt. Enzyme sind hochspezifische und effiziente biochemische Katalysatoren; deshalb kann ein Protein, auch wenn es nur in wenigen Exemplaren in einer Zelle vorhanden ist, eine entscheidende Rolle für die Lebensfunktionen haben. Bei einem solchen Protein ist logischerweise dann auch nur eine geringe Menge mRNA für die Translation notwendig. Die Herausforderung, die sich hieraus ergibt, besteht in der Identifizierung und

Quantifizierung aller mRNA-Sorten und aller Proteine in einer Zelle, unabhängig davon, ob es sich um Haupt- oder Spurenbestandteile handelt.

Eine Zwischenlösung auf dem Weg zu einer umfassenden Proteomik von Einzelzellen ist das Erfassen von Proteinprofilen (protein profiling).^[4] Der Informationsgehalt kann dabei extrem hoch sein, wie Studien mit der konventionellen zweidimensionalen Gelelektrophorese (2D-PAGE) belegen. Um die kleinen in Einzelzellen vorhandenen Mengen handhaben zu können, wird die Proteintrennung kapillarelektrophoretisch durchgeführt; die laserinduzierte Fluoreszenz erlaubt dann einen empfindlichen Nachweis. Die Implementierung der beiden Trenntechniken (isoelektrische Fokussierung und SDS-Polyacrylamid-Größenbestimmung) im Kapillarformat ist nach wie vor schwierig, insbesondere da die Proteine zur Detektion fluoreszenzmarkiert sein müssen; damit bleibt die Identifizierung und Korrelierung mit traditionellen 2D-Gel-Daten ein Fernziel. Die Trennleistung liegt ebenfalls unter der der Platten-Gelelektrophorese; als Folge davon leidet der Informationsgehalt.

Es ist ganz klar, dass der Einzelmolekülnachweis und die Identifizierung von mRNAs und Proteinen notwendig sind, um Transkriptom und Proteom einer Einzelzelle zu erfassen und zu charakterisieren. Fortschritte in der Einzelmolekülspektroskopie in den beiden zurückliegenden Jahrzehnten haben zur Entwicklung einfacher und robuster fluoreszenzspektroskopischer Analysengeräte wie dem Durchflusszytometer oder dem Fluoreszenzmikroskop geführt. Wenn man Stabilitätsaspekte außer Acht lässt, kann mRNA durch geeignete (komplementäre) Hybridisierungs sonden mit hoher Spezifität nachgewiesen werden, und dies sogar ohne unterstützende Amplifikation durch die Polymerasekettenreaktion (PCR). Ein derartiger Ansatz ist die In-situ-Hybridisierung mit einem fluoreszenzmarkierten Marker (FISH).^[5] Die Fluoreszenzantwort ist über einen großen Dynamikbereich linear, solange Photobleichung vernachlässigbar ist.

Ein analoges Vorgehen zum Nachweis einzelner Proteine in Einzelzellen funktioniert nicht sehr gut, denn fluoreszierende Antikörper sind zwar wirkungsvolle Sonden für Proteine, doch ihre Bindungsaffinitäten reichen vielfach für einen Einzelmolekülnachweis nicht aus. Darüber hinaus stehen umfassende Antikörperbibliotheken nicht zur Verfügung. Glücklicherweise wird durch das Genspleißen die Coexpression des Wunschproteins mit einem signalgebenden Protein, das auf der Einzelmolekülebene detektierbar ist, wie β -Ga-

[*] Prof. E. S. Yeung
Ames Laboratory-USDOE, Iowa State University
118 TASF, Ames, IA 50011 (USA)
Fax: (+1) 515-294-0105
E-Mail: yeung@ameslab.gov

[**] Das Ames Laboratory wird auf der Grundlage des Vertrages DE-AC02-07CH11358 von der Iowa State University für das US-Energieministerium betrieben. Die vorliegende Arbeit wurde vom Director of Science, Office of Basic Energy Sciences, Division of Chemical Sciences gefördert.

lactosidase,^[6] Luciferase^[7] und dem grün fluoreszierenden Protein (GFP),^[8] erleichtert. Von diesen sind die ersten beiden hochempfindlich, da sich mit der Zeit eine große Menge der signalgebenden Produktmoleküle anreichern kann; die Tatsache, dass eine weitere Reaktion erforderlich ist, bedeutet aber, dass die Quantifizierung eine indirekte ist, von den Milieubedingungen abhängt und den Zusatz von Fremds substraten erfordert. Beim ursprünglichen GFP-System dauert es circa zwei Stunden, bis sich das durch Translation gebildete Protein in die fluoreszierende Konformation gefaltet hat; zudem bleicht dieses System leicht photochemisch aus. Mit der Herstellung einer vollständigen, chromosomal affinitäts-markierten Bibliothek von *Escherichia coli*-Zellen^[9] wurde die Einzelproteomik zur Realität, da YFP-Fusionsproteine (YFP: gelb fluoreszierendes Protein) eingesetzt werden können.^[10]

Mit diesen technischen Innovationen ausgestattet, haben Taniguchi, Li et al. eine elegante und umfassende analytische Bestimmung des Proteoms von Einzelzellen vorgenommen. Dabei kam eine YFP-Fusionsbibliothek des als Modellorganismus dienenden Bakteriums *E. coli* zum Einsatz,^[11] und es wurden intakte Zellen fluoreszenzmikroskopisch ausgewertet (Abbildung 1). Natürlich muss bei dieser Form der umfas-

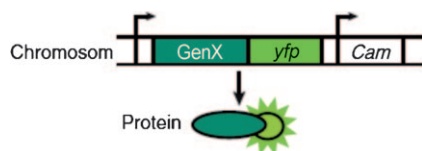


Abbildung 1. Quantitative Untersuchung einer YFP-Fusionsbibliothek.^[11] Jeder Bibliotheksbakterienstamm bildet ein YFP-Fusionsprotein, dessen YFP-Teil mit dem C-Terminus eines bakterieneigenen Proteins fusioniert ist. Das zugrundeliegende Fusionsgen liegt am nativen chromosomal Ort des Bakteriengens. *Cam*: Chloramphenicol-Acetyltransferase-Gen.

senden Untersuchung eine große Zahl an Zellen angeschaut werden. Die Lösung besteht darin, sich das kleine Volumen und die hohe Multiplizität, die die Mikrofluidik kennzeichnen, zunutze zu machen. Gruppen von jeweils 96 Zellen wurden gleichzeitig einer automatisierten Bildauswertung unterworfen. Dieses Verfahren erzeugt eine große Datenmenge. Innerhalb von 25 s lassen sich 4000 Zellen eines Bakterienstamms untersuchen. Bei ausgewählten cytoplasmatischen, membranständigen oder DNA-bindenden Proteinen war die Lokalisierung der YFP-Fluoreszenz in den erwarteten Bereichen evident. Zunächst wurde mit einem nichtfluoreszierenden Bakterienstamm ein Referenzwert für die Autofluoreszenz ermittelt, der als „Nulllinie“ für die Bestimmung der mittleren Signalintensität pro isoliertem Membranprotein auf ähnlichen Zellbildern diente. Parallel wurden die Volumina der Zellen gemessen und aufgezeichnet, um die berechneten Proteinzahlen in intrazelluläre Konzentrationen umrechnen zu können. Dieser Ansatz wurde durch Massenspektrometrie und Immun-Blot (Western-Blot) validiert.

Die Proteine von 1009 der 1018 untersuchten Bakterienstämme zeigten eine Gamma-Verteilung, wie aus einer vor-

angegangenen Untersuchung mit einer kleineren Stichprobe vorausgesagt worden war.^[6] Die gemessenen Proteinmengen lassen sich heranziehen, um das Rauschen und die Dynamik der Genexpression zu verstehen.^[12] Die Analyse für ausgewählte Genpaare ergab, dass bei geringer Individuenzahl des betreffenden Proteins das intrinsische Rauschen (das einer Zufallsstatistik inhärent ist) dominiert, während extrinsisches Rauschen (Variabilität von Zelle zu Zelle) ein gemeinsames Merkmal aller stark exprimierten Gene ist.^[11] Ein weiterer interessanter Befund war, dass das *E. coli*-System ein stärkeres extrinsisches Rauschen zeigt als eine Hefeart,^[12] dies gilt sogar für etwa gleich häufige Proteine. Der Grund für dieses Phänomen ist bislang unklar.

Die Zahl der mRNA-Moleküle in einer jeden Zelle wurde bis hinunter zur Einzelmolekülebene analog ermittelt (Abbildung 2). Um mRNA und zugehöriges Protein (Transla-

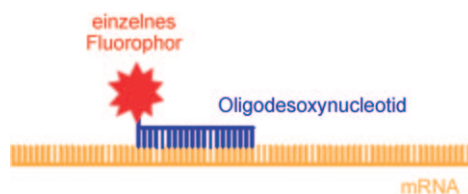


Abbildung 2. Die mRNA eines markierten Gens kann durch FISH gegen die *yfp*-mRNA-Sequenz mithilfe eines DNA-Oligomers als Sonde nachgewiesen werden, das mit einem einzelnen Atto594-Fluorophormolekül markiert ist.^[11]

tionsprodukt) parallel analysieren zu können, wurde eine zweite Farbmarkierung für eine weitere FIS-Hybridisierung mit demselben *yfp*-Gen verwendet. Der experimentelle Ansatz wurde in Kontrollexperimenten durch Einsatz von Echtzeit-PCR zur Quantifizierung der mRNA validiert. Im Unterschied zum Translationsprodukt Protein, das in Echtzeit dargestellt werden kann, muss die Zelle für eine FISH fixiert werden. Glücklicherweise ergab sich für die YFP-Signale der Zellen vor und nach der Fixierung eine gute Korrelation.

Die parallele Erfassung der mRNA und des korrespondierenden Proteins für jedes Gen einer Zelle (Abbildung 3) erhellt auf bislang unerreichte Weise den Mechanismus der Expression. Vorausgegangene Studien hatten gezeigt, dass Proteine in Form von „Syntheseausbrüchen“ gebildet werden, deren Ablauf und deren Produktmenge als Funktion der

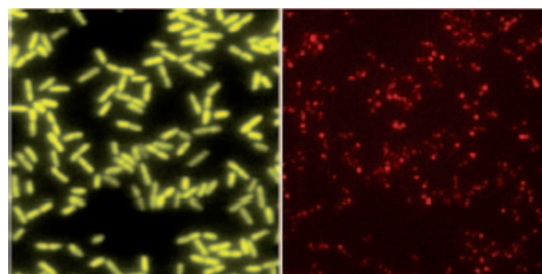


Abbildung 3. Proteine eines YFP-Fusionsstamms (gelb, links) und die korrespondierende Atto-594-markierte mRNA (rot, rechts) in Einzelzellen.^[11]

Zeit stochastisch sind. Dies ist in Einklang mit einer geringen Zahl an mRNA-Kopien.^[10] Die Frage bleibt, ob die Zahl der mRNA-Kopien und die der durch die Translation gebildeten Proteinmoleküle korreliert sind. Überraschenderweise wurde bei den 129 untersuchten Bakterienstämmen mit stark exprimierten Genen zu einem definierten Zeitpunkt, dem der Fixierung der Zellen, keine Korrelation zwischen der Zahl an mRNAs und der an Proteinmolekülen festgestellt. Dies lässt sich mit den unterschiedlichen Halbwertszeiten der beiden Biomoleküle in einer Zelle erklären. Proteine sind sich akkumulierende Produkte, mRNAs dagegen transiente Botenmoleküle, die innerhalb von Minuten abgebaut werden. Selbstverständlich würde man dennoch erwarten, dass der über die Zeit integrierte Gehalt an mRNA mit dem Proteingehalt korreliert.

Obschon diese Studie einen großen Schritt hin zur umfassenden Einzelzelltranskriptomik und -proteomik darstellt, sind weitere Entwicklungen wünschenswert. Um die Korrelation der mRNA- und der Proteinmolekülzahl zu bestimmen, würde man diese Werte gern durch einen ganzen Zellzyklus hindurch verfolgen. Während sich Proteine in Echtzeit mit bildgebenden Verfahren erfassen lassen, ist mit FISH gegenwärtig nur eine Einzelpunktmessung möglich. Weiterhin müsste man, um tatsächlich etwas über die Funktion zu erfahren, die enzymatischen Reaktionen selbst verfolgen,

nicht bloß die Proteinmengen; man müsste sich mithin um die Metaboliten kümmern. Dies stellt aber eine noch größere Herausforderung dar, weil keine passenden Markierungen zur Verfügung stehen und diese Stoffe eine enorme chemische Vielfalt aufweisen.

Eingegangen am 23. September 2010

-
- [1] J. E. Edström, *Nature* **1953**, 172, 809.
 - [2] G. T. Mantioli, H. B. Niewisch, *Science* **1965**, 150, 1824.
 - [3] S. J. Lillard, E. S. Yeung, R. M. A. Lautamo, D. T. Mao, *J. Chromatogr. A* **1995**, 718, 397.
 - [4] D. Cohen, J. A. Dickerson, C. D. Whitmore, E. H. Turner, M. M. Palcie, O. Hindsgaul, N. J. Dovichi, *Annu. Rev. Anal. Chem.* **2008**, 1, 165.
 - [5] J. M. Levisky, S. M. Shenoy, R. C. Pezo, R. H. Single, *Science* **2002**, 297, 836.
 - [6] L. Cai, N. Friedman, X. S. Xie, *Nature* **2006**, 440, 358.
 - [7] S. H. Kang, S. Lee, E. S. Yeung, *Angew. Chem.* **2010**, 122, 2657; *Angew. Chem. Int. Ed.* **2010**, 49, 2603.
 - [8] R. Y. Tsien, *Annu. Rev. Biomed.* **1998**, 67, 509.
 - [9] G. Butland, et al., *Nature* **2005**, 433, 531.
 - [10] J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, *Science* **2006**, 311, 1600.
 - [11] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, *Science* **2010**, 329, 533.
 - [12] J. R. S. Newman, et al., *Nature* **2005**, 441, 840.
-