

Dekodierung kryptischer Translationsereignisse identifiziert sORFs als neue Quelle von MHC Klasse I Antigenen

Als obligat intrazelluläre Pathogene ohne eigene Ribosomen müssen Viren die Translationsmaschinerie des Wirts für ihre Proteinbiosynthese zweckentfremden. Die Translation stellt somit eine essentielle Stellgröße der widerstrebenden Interessen von Virus und Wirt dar (1). Neue Verfahren zur Analyse von mRNAs, die zum Zeitpunkt der Messung translatiert wurden, liefern nun wichtige Erkenntnisse über die Komplexität der Translation in virusinfizierten Zellen. Da das adaptive Immunsystem virusinfizierte Zellen über MHC-präsentierte Peptide erkennen kann, haben diese Befunde auch wichtige immunologische Implikationen.

Das humane Genom beinhaltet nur etwa 20.000 protein-kodierende Gene. Welche mRNAs von diesen Genen gebildet und von Ribosomen translatiert werden, lässt sich mittels Ribosomen Profiling (Ribo-seq) bestimmen (2). Die Methode beruht auf der Aufreinigung und anschließenden Identifikation und Quantifikation von kleinen mRNA-Fragmenten (28-30 Nukleotide lang), die zum Zeitpunkt der Zelllyse aktuell translatiert und dadurch im Ribosom vor einem Nuklease-Verdau geschützt wurden. RNAs, die nicht translatiert werden, sind auf Grund des fehlenden Schutzes durch Ribosomen Nuklease-sensitiv. Die schrittweise fortschreitende Translation der Ribosomen von Codon zu Codon hinterlässt dabei in den Ribo-seq-Daten charakteristische Nukleotidtripletpuren. Dies ermöglicht es, die vom Ribosom translatierten Codons sowie die zugrunde liegenden offenen Leserrahmen („*open reading frames* [ORFs]“) zu bestimmen. Ribo-seq führte in den letzten Jahren zur Identifikation tausender neuer kleiner ORFs (sORFs) in etwa 50% unserer Gene (3). Der überwiegende Teil dieser sORFs ist allerdings auf Proteinebene z.B. mittels Massenspektrometrie nicht nachweisbar. Daher vermutet man, dass sORF-kodierte Polypeptide unmittelbar nach ihrer Translation wieder abgebaut werden und meist wohl keine intrinsische Funktionsrelevanz haben. Die Translation der kurzen vorgeschalteten sORFs reguliert jedoch die Expressionsstärke der ORFs, die auf der gleichen mRNA folgen (4). Diese sORF-bedingte Translationskontrolle stellt einen wichtigen zellulären Regulationsmechanismus insbesondere bei zellulären Stressreaktionen dar (5).

Eine Herausforderung bei der Analyse von Ribosome-Profiling-Daten stellt das Hintergrundrauschen dar. Dies beruht auf der inkorrekten Zuweisung von sequenzierten RNA-Fragmenten zu translatierten Codons. Hiervon sind insbesondere Codons von überlappenden ORFs betroffen. Zudem macht dies die zuverlässige Identifizierung von ORFs mit nicht-kanonischen (non-AUG) Startcodons schwierig. Ausgehend von der Erkenntnis, dass die Länge und Position der mRNA-Fragmente ein stochastisches und somit modellierbares Ereignis darstellt, konnten Erhard *et al.* mit einem neuen bioinformatischen Verfahren (genannt „PRICE“; <https://github.com/erhardlab/price>) die Menge der Codons bestimmen, deren Translation mit der größten Wahrscheinlichkeit zu den sequenzierten RNA-Fragmenten geführt hat (6). Sie erhielten mit diesem Ansatz ein viel schärferes Bild der zellulären Translation bei substantiell besserer Ausnutzung der Datensätze. Anhand von MHC-I-Peptidome-Daten (7) konnten sie im Folgenden zeigen, dass die Frequenz der Peptide von zellulären sORFs zwar in Gesamtproteom-Analysen unterhalb der Grenze der Fehlerrate (*false discovery rate*) liegt, in MHC-I-Peptidome-Analysen jedoch klar oberhalb davon, so dass von einer effizienten Prozessierung und Präsentation über MHC-I-Moleküle an der Zelloberfläche auszugehen ist. Diese Daten beweisen so erstmals die physische Existenz zahlreicher zellulärer sORFs. Passend zu dieser MHC-Präsentation konnten in Patienten CD4+ und CD8+ T-Zellantworten gegen solche neuen viralen ORFs nachgewiesen werden (8). Der Unterschied zwischen MHC-I-Peptidome und Vollproteomdaten wirft spannende immunologische Fragen auf, z.B. ob diese sORF-kodierten Antigene aufgrund ihrer geringen Abundanz in der Zelle für Kreuz-Präsentation zur Verfügung stehen.

In einer wegweisenden Arbeit führte Ribosome Profiling 2012 zu einer umfassenden Neu-Annotierung des Genoms des humanen Zytomegalievirus (HCMV) (9). Bis dahin ging man davon

aus, dass HCMV in dem 236 kBp großen Genom etwa 170 verschiedene virale Proteine kodiert. Ribosome Profiling bestätigte knapp 95% der annotierten viralen Genprodukte und lieferte zusätzlich Hinweise für die Existenz von mehr als 500 neuen viralen Proteinen und Polypeptiden, darunter etwa 250 sORFs. Mit ihrem optimierten Verfahren haben Erhard *et al.* nun die damals publizierten Daten erneut analysiert. Dabei konnten sie Translationsereignisse für alle 170 annotierten Proteine nachweisen. Von den ca. 500 neu vorausgesagten Proteinen wurde ca. die Hälfte von der neuen Methode untermauert, während knapp die andere Hälfte nicht unterstützt wurde. Das bedeutete jedoch nicht, dass die Translation von HCMV nun wieder etwas übersichtlicher ist. PRICE sagte 500 weitere virale ORFs und sORFs voraus, die überwiegend von AUG Startcodons initiieren. Hiervon werden etwa 200 ähnlich stark exprimiert, wie die bisher bekannten viralen ORFs. Mit diesen neuartigen Erkenntnissen zur unerwarteten Komplexität der Translation in infizierten Zellen eröffnen diese Arbeiten unzählige neue immunologische und virologische Betätigungsfelder.

Referenzen

1. D. Walsh, M. B. Mathews, I. Mohr, Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb Perspect Biol* **5**, a012351 (2013).
2. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).
3. J. I. Pueyo, E. G. Magny, J. P. Couso, New Peptides Under the s(ORF)ace of the Genome. *Trends Biochem Sci* **41**, 665-678 (2016).
4. K. Wethmar, The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **5**, 765-778 (2014).
5. S. R. Starck *et al.*, Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867 (2016).
6. F. Erhard *et al.*, Improved Ribo-seq enables identification of cryptic translation events. *Nat Methods* **15**, 363-366 (2018).
7. M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, M. Mann, Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics* **14**, 658-673 (2015).
8. N. T. Ingolia *et al.*, Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep* **8**, 1365-1379 (2014).
9. N. Stern-Ginossar *et al.*, Decoding human cytomegalovirus. *Science* **338**, 1088-1093 (2012).

Mirko Trilling
Institut für Virologie
Universität Duisburg Essen
mirko.trilling@uni-due.de