

Genomweite Annotation von Translationsinitiationsstellen
in der humanen Monozyten-Zelllinie THP-1

Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Claudia Fritsch

Kiel, im April 2013

Erste/r Gutachter/in: Prof. Dr. Karin Krupinska

Zweite/r Gutachter/in: Prof. Dr. Jochen Hampe

Tag der mündlichen Prüfung: 20.06.2013

Zum Druck genehmigt: 20.06.2013

gez. Prof. Dr. Wolfgang J. Duschl, Dekan

Für meine Familie

Diese Arbeit entstand von Februar 2010 bis April 2013 in der Arbeitsgruppe Genomische Gastroenterologie des Universitätsklinikums Schleswig-Holstein, Campus Kiel.

Zusammenfassung

In der vorliegenden Arbeit wurde die erste genomweite Annotation von Translationsinitiationsstellen (TISs) in einer humanen Zelllinie durchgeführt. Insgesamt wurden 7251 TISs in 5062 Transkripten, welche die definierten Analyse Kriterien erfüllten, detektiert.

Die Identifizierung von TISs basierte bis vor einiger Zeit hauptsächlich auf bioinformatischen Analysen, wobei alternative non-AUG Startstellen nicht erfasst wurden. Zahlreiche Studien dokumentieren, dass die Translation nicht nur an der kanonischen CDS sondern auch an alternativen AUG und non-AUG Startcodonen in der 5'-UTR der Transkripte initiiert. Diese *upstream* lokalisierten Translationsinitiationsstellen repräsentieren *Upstream-open-Reading-Frames* (uORFs) und CDS überlappende uORFs (ovuORFs), die entscheidende regulatorische Funktionen für die Genexpression haben können, sowie N-terminale Proteinextensionen, welche für die subzelluläre Proteinsortierung maßgeblich sein können. Infolge einer funktionellen Klassifizierung der 7251 vorhergesagten TISs wurden 2305 kanonische Starts, 2994 neue uORFs, 1406 neue ovuORFs und 546 neue N-terminale Proteinextensionen dokumentiert.

Als Voraussetzung für eine erfolgreiche Durchführung musste zunächst das Ribosomen-Profiling etabliert werden, das auf der Sequenzierung von durch Ribosomen geschützten ca. 30 nt langen mRNA-Fragmenten basiert. Die akkurate Identifizierung der Startstellen erfolgte anschließend durch die Behandlung der Zellen mit Puromycin und Cycloheximid, wodurch die Ribosomen an den TISs der Transkripte angereichert wurden.

Mit Hilfe eines künstlichen neuronalen Netzes, das mit ribosomalen Footprints an zuvor annotierten AUG Startstellen trainiert wurde, wurden die TISs daraufhin systematisch annotiert.

Um die Qualität der Methode und die Zuverlässigkeit der Ergebnisse dieser Arbeit zu validieren, wurden die hier bestimmten TISs mit zuvor in der Literatur beschriebenen alternativen non-AUG Startstellen und uORFs verglichen.

Letztendlich wurde die Konservierung der Translationsinitiationscodonen (TICs) analysiert. Sowohl AUG- als auch non-AUG-Startcodonen sind zwischen Primaten und zwischen Menschen und Nagetieren stark konserviert.

Zusammenfassend wurde in dieser Arbeit verdeutlicht, wie unerwartet oft die Translation an alternativen TISs initiiert und wie komplex und relevant demzufolge die translationale Regulation für die Genexpression ist.

Abstract

In the present study, the first genome-wide annotation of translation initiation sites (TISs) was performed in a human cell line. In total 7251 TISs were predicted in 5062 transcripts that met the defined analysis criteria.

Until recently the identification of TISs was mainly based on bioinformatic analyzes, and alternative non-AUG starts were not be recorded. Numerous studies document that translation not only initiates at the canonical CDS, but also from alternative AUG and non-AUG start codons in the 5' UTR of the transcripts. These upstream localized translation initiation sites represent upstream open reading frames (uORFs) and CDS overlapping uORFs (ovuORFs), that can have critical regulatory functions for gene expression, as well as N-terminal protein extensions, which may be relevant for the subcellular protein sorting. Based on a functional classification of the 7251 predicted TISs, 2305 canonical starts, 2994 new uORFs, 1406 new ovuORFs and 546 new N-terminal protein extensions were documented.

As a prerequisite for the successful performance, the ribosome profiling method, based on deep sequencing of ribosome protected mRNA fragments, had to be established.

For accurate identification of the start sites, ribosomes were enriched at TISs of the transcripts by adding puromycin and cycloheximide to the cells. TISs were subsequently annotated systematically using a neural network, which was trained on the ribosomal footprints observed at previously annotated AUG start sites.

In order to validate the quality of the method and the reliability of the results of this work, the predicted TISs were compared with alternative non-AUG start sites and uORFs, which were previously described in the literature.

Following that, the conservation of translation initiation codons (TICs) was analyzed. Both the predicted AUG and non-AUG start codons are highly conserved among primates and between humans and rodents.

In summary, it was possible to illustrate in this work, how unexpectedly often translation initiates at alternative TISs and how complex and consequently translational regulation is for gene expression.

Inhaltsverzeichnis

Zusammenfassung	I
Abstract	II
Inhaltsverzeichnis	III
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
Abkürzungs- und Symbolverzeichnis	IX
1 Einleitung	1
1.1 Regulation der Genexpression bei Eukaryoten	1
1.2 Translation	2
1.2.1 Initiation der Translation bei Eukaryoten nach dem <i>Scanning</i> -Mechanismus.....	2
1.3 Regulation der Translationsinitiation.....	5
1.3.1 Regulation der globalen Translationsinitiation durch eIFs	5
1.3.2 Regulation der Translationsinitiation durch <i>cis</i> -regulatorische Elemente	6
1.3.3 Regulation der Startcodonwahl	6
1.4 Das Kodierungspotential von eukaryotischen Transkripten	8
1.4.1 Gesteigerte Proteinviefalt durch <i>Leaky Scanning</i> und Reinitiation	8
1.4.2 uORFs.....	9
1.4.3 Alternative non-AUG Startcodonen.....	11
1.5 Die Identifizierung von Translationsinitiationsstellen.....	11
1.6 Zielsetzung	13
2. Material und Methoden	14
2.1 Allgemeine Chemikalien	14
2.2 Oligonukleotide	14
2.3 Zellkultur	14
2.3.1 Kultivierung und Passagieren von Zellen	15
2.3.2 Zellernte	15
2.3.3 Zellzahlbestimmung mittels Neubauerzählkammer	16

2.3.4 Kryokonservierung und Revitalisierung von Zellen	16
2.3.5 Mycoplasmentest.....	17
2.4 Gelelektrophorese von Nukleinsäuren.....	17
2.4.1 Agarosegelelektrophorese	17
2.4.2 Denaturierende TBE-Urea-Gele.....	18
2.5 Isolierung und Aufreinigung von Nukleinsäuren	19
2.5.1 Phenol-Chloroform-Extraktion von RNA und cDNA	19
2.5.2 Isolierung von Gesamt-RNA mittels Trizol.....	19
2.5.3 Alkoholpräzipitation von Nukleinsäuren	20
2.5.4 Qualitätskontrolle der RNA mittels Agarosegelelektrophorese.....	20
2.5.5 Konzentrationsbestimmung von Nukleinsäuren	20
2.6 Untersuchung der Translationsinitiation.....	22
2.6.1 Fixierung von Ribosomen auf der mRNA	23
2.6.2 Anreicherung von Ribosomen an TISs	23
2.6.3 Isolierung von Polysomen und Monosomen.....	23
2.6.4 Hydrolyse der Polysomen mit verschiedenen Ribonukleasen	24
2.6.5 Dichtegradientenzentrifugation und Fraktionierung von Monosomen und Polysomen	25
2.6.6 Anreicherung und Aufreinigung der RPFs.....	26
2.7 Footprint-Bibliotheken Präparation	27
2.7.1 Dephosphorylierung der Footprint-RNA	28
2.7.2 Größenselektion und Aufreinigung der dephosphorylierten Footprint-RNA	28
2.7.3 Polyadenylierung der Footprint-RNA.....	29
2.7.4 Footprint-cDNA-Synthese	29
2.7.5 Größenselektion und Aufreinigung der Footprint-cDNA.....	30
2.7.6 Zirkularisierung der Footprint-cDNA	31
2.7.7 Relinearisierung der zirkularisierten Footprint-cDNA	31
2.7.8 Amplifizierung der RPFs mittels PCR.....	32
2.8 Etablierung und Optimierung des Ribosomen-Profilings für humanes Lebergewebe	33
2.8.1 Subtraktive Hybridisierung von rRNAs.....	33
2.8.2 Abwandlung der cDNA-Synthese.....	34
2.8.3 Eluierung der Footprint-RNA und cDNA aus TBE-Urea-Gelen.....	35

2.8.4 Amplifizierung der RPFs mittels PCR unter Verwendung von Multiplex-Primern	36
2.9 Illumina <i>Next-Generation-Sequencing-System</i>	36
2.10 Analyse der sequenzierten Fragmente	36
2.10.1 Sequenzanalyse	37
2.10.2 Vorhersage von Translationsinitiationsstellen mittels eines neuronalen Netzes	37
2.10.3 Interpretation der vorhergesagten TISs	40
2.10.4 Analyse der Sequenzkonservierung der identifizierten TICs.....	41
3. Ergebnisse	43
3.1 Etablierung des Ribosomen-Profilings.....	43
3.1.1 Hydrolyse von THP-1 Polysomen mit verschiedenen Ribonukleasen.....	43
3.1.2 Etablierung der Footprint-Bibliotheken Präparation.....	44
3.1.3 Analyse der sequenzierten Footprint-Fragmente	45
3.1.4 Sequenzabdeckung der Transkripte	46
3.2 Identifizierung und Klassifizierung von TISs.....	49
3.2.1 Anreicherung von RPFs an den TISs mit Hilfe von Puromycin	49
3.2.2 Analyse der RPFs	50
3.2.3 Anreicherung von RPFs an TISs	52
3.2.4 Vorhersage von TISs mit Hilfe eines neuronalen Netzes	53
3.2.5 Funktionelle Klassifizierung der vorhergesagten TISs	56
3.2.6 Häufigkeit und Verteilung der TICs.....	59
3.2.7 Validierung der Identifizierten TISs	60
3.2.8 Ermittlung einer Konsensussequenz der vorhergesagten TISs	66
3.2.9 Konservierung der TIC-Sequenzen zwischen Primaten.....	67
3.2.10 Konservierung der TIC-Sequenzen zwischen Mensch und Maus	69
3.2.11 Annotationstrack der TISs.....	70
3.3 Etablierung des Ribosomen-Profilings für humanes Lebergewebe	71
4. Diskussion.....	73
4.1 Anreicherung von RPFs an den TISs mit Hilfe von Puromycin.....	74

4.2 Etablierung und Optimierung des Ribosomen-Profilings für humanes Lebergewebe	75
4.3 Diversität und Komplexität der neu annotierten TISs	76
4.4 Validierung der neu annotierten TISs	77
4.5 Konservierung der TIC-Sequenzen.....	78
4.6 Biologische Bedeutung von alternativen TISs.....	79
4.7 Perspektive.....	80
5. Literaturverzeichnis	82
6. Anhang.....	94
6.1 Oligonukleotide	94
6.2 Analyse der RPFs.....	95
6.3 Reproduzierbarkeit der Ribosomen-Profilings-Methode	96
6.4 Häufigkeit und Verteilung der TICs	97
6.5 Validierung der identifizierten TICs.....	98
6.6 Konservierung der TIC-Sequenzen zwischen Primaten	100
6.7 Vergleich der TIC-Verwendung zwischen Mensch und Maus.....	101
Erklärung	103
Lebenslauf	104
Publikationen	105
Danksagung.....	106

Abbildungsverzeichnis

Abb. 1: Regulation der Genexpression bei Eukaryoten	1
Abb. 2: Translationsinitiation nach dem <i>Scanning</i> -Mechanismus	4
Abb. 3: Schematische Abbildung von alternativen TISs.....	10
Abb. 4: Schematische Zusammenfassung der Methode	22
Abb. 5: Schematische Zusammenfassung der Präparation von Footprint-Bibliotheken....	27
Abb. 6: Möglicher Aufbau eines neuronalen Netzes.....	38
Abb. 7: Funktionelle Klassifizierung der vorhergesagten TISs.	41
Abb. 8: THP-1 Polysomenprofile von verschiedenen Hydrolyseansätzen.....	44
Abb. 9: Denaturierende TBE-Harnstoffgele bei der Footprint-Bibliotheken Präparation .	45
Abb. 10: Sequenzabdeckung von <i>ACTB</i> und <i>TUBB</i>	47
Abb. 11: Sequenzabdeckung von <i>LDHB</i>	48
Abb. 12: Schematische Darstellung zur Theorie des Versuchs.....	49
Abb. 13: Polysomenprofile von Kontroll- und mit Puromycin behandelten Proben	50
Abb. 14: Längenverteilung der sequenzierten Fragmente.....	51
Abb. 15: Reproduzierbarkeit der Verteilung der RPFs über die Transkripte	51
Abb. 16: Sequenzabdeckung von <i>TTP1</i> in Kontrollzellen und mit Puromycin behandelten Zellen.....	52
Abb. 17: Anreicherung von RPFs durch Puromycinbehandlung	53
Abb. 18: ROC-Kurven der 10 neuronalen Netze	54
Abb. 19: Verteilung der TISs pro Transkript	55
Abb. 20: Funktionelle Klassifizierung der vorhergesagten TISs	57
Abb. 21: Häufigkeit und Verteilung der TICs	59
Abb. 22: Schematische Abbildung vom <i>SP3</i> uORF.....	62
Abb. 23: Sequenzkontexte der TISs verschiedener funktioneller Kategorien.....	66
Abb. 24: Sequenzkontexte der TISs in Abhängigkeit von den verschiedenen Startcodonen	67
Abb. 25: Konservierung der Sequenzen identifizierter TICs zwischen Primaten.....	68
Abb. 26: Vergleich der Codonverteilung zwischen Mensch und Maus	70
Abb. 27: <i>Screenshot</i> eines Annotationstracks am Beispiel von <i>AMD1</i>	71
Abb. 28: Reproduzierbarkeit der Verteilung der RPFs über die Transkripte am Beispiel von <i>TTP1</i>	96

Tabellenverzeichnis

Tabelle 1: Ribonukleasen unterschiedlicher Hersteller	24
Tabelle 2: Zusammensetzung und Inkubationszeiten der Reversen Transkription.....	30
Tabelle 3: PCR-Reaktionsansatz und PCR-Programm	32
Tabelle 4: Zusammensetzung und Inkubationszeiten der Reversen Transkription.....	35
Tabelle 5: Kategorien von Translationsinitiationsstellen	40
Tabelle 6: Analyse der sequenzierten Fragmente.....	46
Tabelle 7: Zusammenfassung der vom neuronalen Netz ermittelten positiven TIS-Signale.....	55
Tabelle 8: Funktionelle Klassifizierung der identifizierten TISs	56
Tabelle 9: Verteilung der klassifizierten TISs über die Transkripte	58
Tabelle 10: Validierung der identifizierten non-AUG initiierten N-terminalen Proteinextensionen unter Verwendung von Literaturdaten	63
Tabelle 11: Validierung der identifizierten uORFs in humanen Genen unter Verwendung von Literaturdaten.....	64
Tabelle 12: Konservierung der TIC-Verwendung zwischen Mensch und Maus	69
Tabelle 13: Analyse der Leber-Footprint-Sequenzen.....	72
Tabelle 14: Verwendete Oligonukleotide.....	94
Tabelle 15: Analyse der sequenzierten Fragmente.....	95
Tabelle 16: TICs der identifizierten TISs	97
Tabelle 17: Gene mit experimentell nachgewiesenen N-terminalen Proteinextensionen und uORFs.....	98
Tabelle 18: Konservierung der identifizierten TIC-Sequenzen zwischen Primaten	100
Tabelle 19: Vergleich der TIC-Verwendung zwischen Mensch und Maus	101
Tabelle 20: Vergleich der TIC-Verwendung von uORFS zwischen Mensch und Maus ..	101
Tabelle 21: Vergleich der TIC-Verwendung von ovuORFs zwischen Mensch und Maus	102

Abkürzungs- und Symbolverzeichnis

A	Adenin
Abb.	Abbildung
ADP	Adenosindiphosphat
AS	Aminosäure
ATP	Adenosintriphosphat
bp	Basenpaar(e)
C	Cytosin
5'-Cap	5'-7-Methyl-Guanosin-Kappe
cDNA	engl. <i>Complementary DNA</i>
CDS	engl. <i>Coding Sequence</i>
Da	Dalton
DNA	engl. <i>Deoxyribonucleic Acid</i>
dNTP	Desoxynukleosidtriphosphat
eIF	eukaryotischer Initiationsfaktor
FCS	engl. <i>Fetal-Calf-Serum</i>
g	Erdgravitation, als Maßeinheit für die Kraftwirkung bei der Zentrifugation
G	Guanin
GDP	Guanosindiphosphat
GTP	Guanosintriphosphat
h	Stunde(n)
Met	Methionin
mRNA	engl. <i>Messenger-RNA</i>
MW	Molekulargewicht
N	jede Base (A, T, G, C)
NMD	engl. <i>Nonsense-mediated-mRNA-Decay</i>
nt	Nukleotid(e)
OD	optische Dichte
ORF	engl. <i>Open-Reading-Frame</i>
ovuORF	engl. <i>CDS-overlapping Upstream-open-Reading-Frame</i>
PABP	Poly(A)-Bindeprotein
PCR	engl. <i>Polymerase-Chain-Reaction</i>
PIC	prä-Initiationskomplex
PNK	Polynukleotidkinase

R	Purinbase (A/G)
RefSeq	Referenzsequenz
RNA	engl. <i>Ribonucleic Acid</i>
RNase	Ribonuklease
RPF	engl. <i>Ribosome-protected-Fragment</i>
rpm	engl. <i>Rounds per Minute</i>
rRNA	ribosomale RNA
RT	Reverse Transkriptase
S	Sedimentationskonstante
SNP	engl. <i>Single-Nucleotide-Polymorphism</i>
T	Thymin
TC	engl. <i>Ternary-Complex</i>
TIC	Translationsinitiationscodon
TIS	Translationsinitiationsstelle
tRNA	transfer-RNA
U	Units
U	Urazil
uORF	engl. <i>Upstream-open-Reading-Frame</i>
UTR	untranslatierte Region
V	A, C oder G
V	Volt
v/v	Volumen / Volumen
w/v	Gewicht / Volumen
Y	Pyrimidin (C, T)
z.B.	zum Beispiel

1 Einleitung

1.1 Regulation der Genexpression bei Eukaryoten

In Eukaryotenzellen wird die Synthese eines Proteins streng und auf verschiedenen Ebenen reguliert. Zu den Kontrollpunkten der Genexpression zählen unter anderem Chromatinstruktur (Ito 2007), Transkriptionsinitiation (Maston et al. 2006), Prozessierung und Modifizierung von mRNA-Transkripten (Black 2003, Lutz & Moreira 2010), Transport von mRNAs ins Zytoplasma (Cheng et al. 2006), Stabilität und Abbau von mRNAs (Buchan & R. Parker 2009, Balagopal et al. 2012), Initiation und Elongation der Translation (Hinnebusch & Lorsch 2012) und der intrazelluläre Transport und Abbau des exprimierten Proteins (Flick & Kaiser 2012).

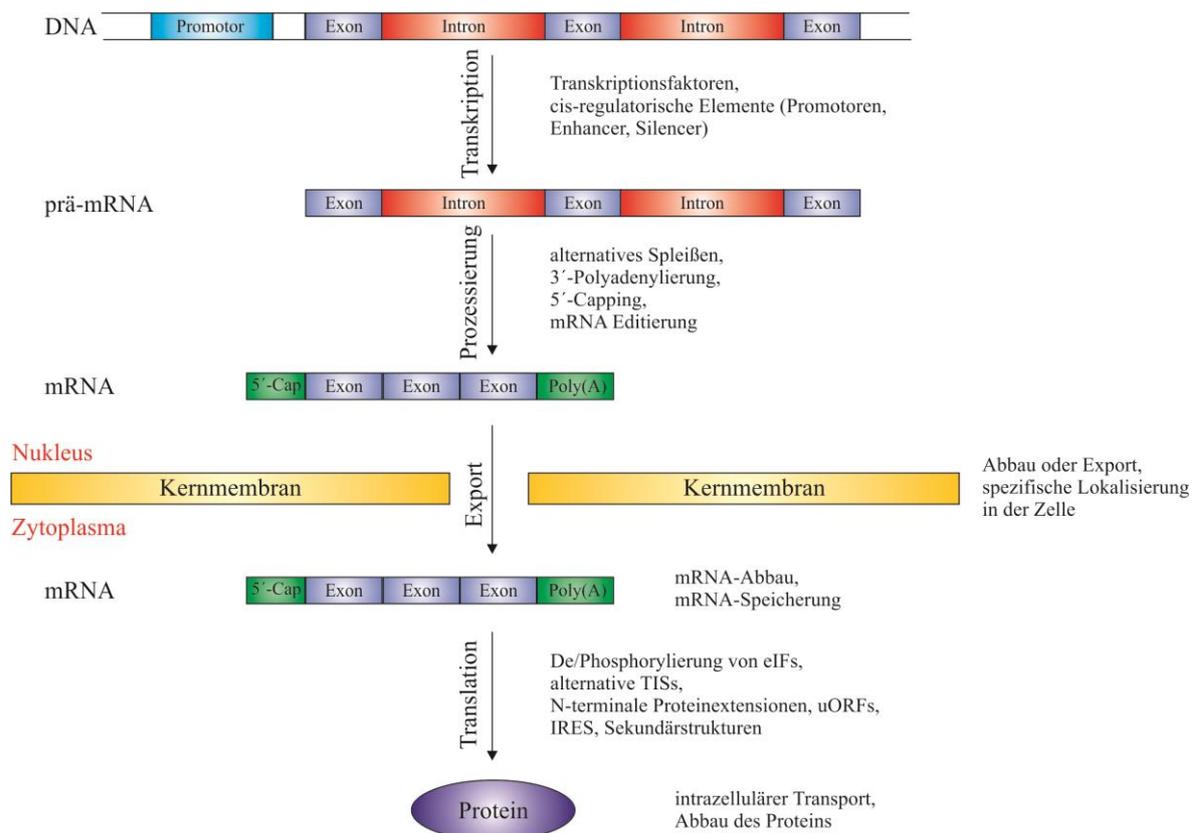


Abb. 1: Regulation der Genexpression bei Eukaryoten. Die Abbildung wurde modifiziert nach Zhang (2002).

Während der Transkription wird die Information eines Gens mit Hilfe der DNA-abhängigen RNA-Polymerase II in ein einzelsträngiges Primärtranskript (prä-mRNA) umgeschrieben. Noch während der Transkription wird an das 5'-Ende der prä-mRNA eine besondere Nukleotidsequenz angefügt, die 5'-Methyl-Guanosin-Kappe (5'-*Cap*). Des Weiteren wird die prä-mRNA gespleißt, d.h. die nicht proteinkodierenden Introns werden entfernt und die Exons wieder zusammengefügt. Am Ende der Transkription erfolgt ebenfalls im Zellkern die Anheftung eines poly(A)-Schwanzes an das freie 3'OH-Ende der mRNA. Dieser Vorgang erfordert ein Polyadenylierungssignal (AAUAAA), welches einige Nukleotide vor dem 3'-Ende der mRNA liegt (Hocine et al. 2010). Nach der Prozessierung der prä-mRNA spricht man von einer reifen mRNA. Diese wird aus dem Zellkern ins Zytoplasma transportiert. Hier wird entschieden, ob sie gespeichert, abgebaut oder an den Ribosomen als Matrize für die Translation verwendet wird (Mignone et al. 2002, Thomas et al. 2011), bei der die genetische Information in eine Aminosäuresequenz übersetzt wird. Die entstehende Aminosäurekette bildet Sekundär- und Tertiärstrukturen und faltet sich damit zum reifen Protein (Abb. 1).

1.2 Translation

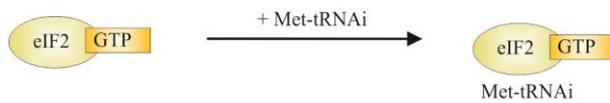
Während der Translation wird die genetische Information der mRNA von den Ribosomen abgelesen und in eine Aminosäuresequenz übersetzt. Dieser Vorgang findet im Zytoplasma der Zellen statt (Day and Tuite 1998) und wird in Initiation, Elongation und Termination unterteilt, wobei vor allem die Initiationsphase der Translation von essentieller Bedeutung für die Regulation der Genexpression bei Eukaryoten ist (Valasek 2012).

1.2.1 Initiation der Translation bei Eukaryoten nach dem *Scanning*-Mechanismus

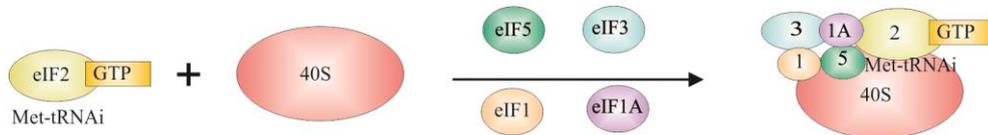
Die Initiation der Translation der meisten eukaryotischen mRNAs erfolgt nach dem *Scanning*-Mechanismus (Kozak 1980). Hierbei bindet die 40S ribosomale Untereinheit an die 5'-7-Methyl-Guanosin-Kappe der mRNA und wandert linear von 5' nach 3' auf dem Transkript entlang, bis sie auf das erste AUG-Codon in einem optimalen Konsensus trifft. Durch Basenpaarungen mit dem Anti-Codon der Methionin-tRNA (Met-tRNA_i oder Initiator-tRNA) wird das Startcodon für die Translation erkannt (Cigan et al. 1988).

Der erste Schritt für die Initiation der Translation nach dem *Scanning*-Mechanismus ist die Assemblierung des *Ternary*-Komplexes, welcher sich aus der GTP-gebundenen Form des eukaryotischen Initiationsfaktors 2 (eIF2) und der Methionin-tRNA zusammensetzt (Abb. 2). Stimuliert durch die Initiationsfaktoren eIF1, eIF1A, eIF5 und eIF3 bindet dieser an die kleine ribosomale 40S Untereinheit. Der entstehende Komplex wird als 43S Prä-Initiationskomplex (43S PIC) bezeichnet und bindet dann an die 5'-7-Methyl-Guanosin-Kappe der mRNA (Hinnebusch 2011). Hierfür muss zuvor der eukaryotische Initiationsfaktorkomplex eIF4F, welcher aus dem *Cap*-Bindeprotein eIF4E, dem Gerüstprotein eIF4G und der Helikase eIF4A besteht, an die 5'-Kappe binden. Da eIF4G sowohl mit dem Poly(A)-Bindeprotein (PABP) als auch mit eIF4E, eIF4A und der mRNA Bindungen eingeht, kommt es zur Ausbildung einer stabilen *Closed-Loop*-Struktur, bei der die 5'-Kappe der mRNA mit dem Poly(A)-Schwanz der mRNA interagiert (Mangus et al. 2003, Lorsch & Dever 2010). Der an die 5'-Kappe gebundene 43S PIC scannt dann die mRNA von 5' nach 3' Richtung ab, bis er auf ein geeignetes Startcodon für die Initiation der Translation trifft (Kozak 1978). Für diese scannende Bewegung ist eine offene Konformation der ribosomalen 40S Untereinheit nötig, welche durch die Initiationsfaktoren eIF1 und eIF1A gewährleistet wird. Nach der Erkennung des Startcodons, unter anderem durch Codon-Anticodon-Wechselwirkungen, wird durch eine Änderung der Konformation die Scan-Bewegung verhindert, wodurch die Initiation begünstigt wird (Maag et al. 2005, Passmore et al. 2007). Hierbei verdrängt das GTPase aktivierende Protein eIF5 den Initiationsfaktor eIF1, welcher daraufhin dissoziiert. Des Weiteren gelangt eIF5 vermutlich in räumliche Nähe von eIF2-GTP und aktiviert dessen GTPase-Funktion, was wiederum die Hydrolyse von GTP mit sich bringt (Das et al. 2001). Nach der Dissoziation der Initiationsfaktoren eIF2-GDP, eIF1A, eIF3 und eIF5 katalysiert eIF5B-GTP die Anlagerung der großen 60S ribosomalen Untereinheit und somit die Assemblierung des 80S Initiationskomplexes. Nach der Hydrolyse vom GTP und der Dissoziation von eIF5B startet die Elongation (Lorsch & Dever 2010, Hinnebusch 2011).

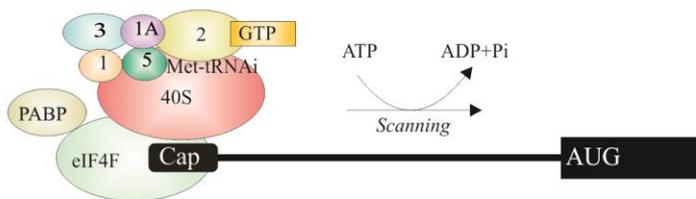
1. Assemblierung des *Ternary*-Komplexes



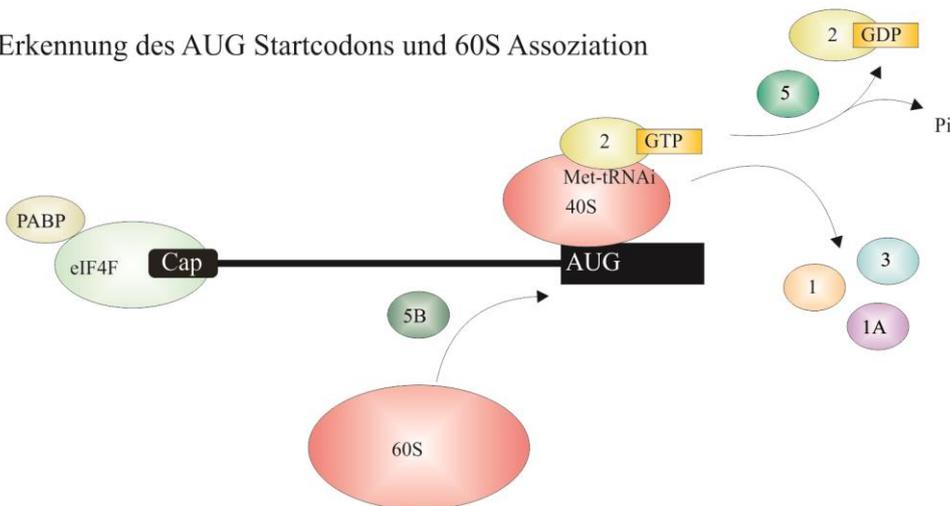
2. 43S PIC Assemblierung



3. Anlagerung vom 43S PIC an die 5'-Kappe und *Scanning*



4. Erkennung des AUG Startcodons und 60S Assoziation



5. Recycling von eIF2-GDP

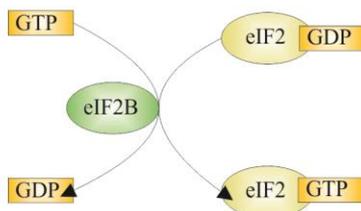


Abb. 2: Translationsinitiation nach dem *Scanning*-Mechanismus. Die großen roten Kreise symbolisieren die ribosomalen 40S und 60S Untereinheiten, wohingegen die kleineren bunten Kreise die Initiationsfaktoren der Translation repräsentieren. In schwarz wurde der 5'-Bereich einer mRNA mit 5'-Kappe, 5'-UTR und dem ORF-Beginn schematisiert. Die Abbildung wurde modifiziert nach Jackson et al. (2010) und Hinnebusch & Lorsch (2012).

1.3 Regulation der Translationsinitiation

Die Translationsinitiation wird über verschiedene Mechanismen gesteuert, die in zwei große Kategorien eingeteilt werden können. Auf der einen Seite wird die globale Proteinsynthese über die Verfügbarkeit und Aktivität von Initiationsfaktoren, die über De/Phosphorylierungen reguliert wird, kontrolliert (Jackson et al. 2010, Sonenberg & Hinnebusch 2009, Abschnitt 1.3.1). Demgegenüber wird die Translation von spezifischen Transkripten durch die Aktivität von *cis*-regulatorischen Elementen, wie z.B. *Internal-Ribosome-Entry-Sites* (IRES), uORFs, Sekundärstrukturen und RNA-Bindeproteinmotiven in den untranslatierten Regionen (UTRs) gesteuert (Keene & Tenenbaum 2002, Spriggs et al. 2010, Wethmar et al. 2010, Abschnitt 1.3.2).

1.3.1 Regulation der globalen Translationsinitiation durch eIFs

Die globale Translationsinitiation wird durch Proteinfaktoren reguliert, deren biologische Aktivität durch Phosphorylierungen und Dephosphorylierungen kontrolliert wird.

Voraussetzung für die Initiation der globalen Proteinsynthese ist die Assemblierung des *Ternary*-Komplexes, wofür die GTP gebundene Form des eukaryotischen Proteinfaktors eIF2 essentiell ist. In Säugetieren wird der von eIF2B katalysierte Austausch vom eIF2 gebundenen GDP mit GTP durch Phosphorylierungen von eIF2 α am Ser-51 verhindert (Hershey 1991). Infolgedessen kommt es unter Stressbedingungen zu einer Reduktion des eIF2-GTP Levels und somit auch zu einer verringerten Bildung des *Ternary*-Komplexes in der Zelle, wodurch die globale Translation inhibiert wird (Jiang et al. 2007, Jackson et al. 2010).

Der *Ternary*-Komplex ist wiederum Bestandteil des 43S PIC. Für eine Bindung vom 43S PIC an die 5'-Kappe ist eine vorangegangene Assoziation von eIF4F und der 5'-Kappe erforderlich, wobei die intrazelluläre eIF4F-Konzentration indirekt durch Phosphorylierungen von eIF4E-Bindeproteinen (4E-BP) reguliert wird (Ma & Blenis 2009). Hypophosphorylierte 4E-BP binden an eIF4E, wodurch die Assemblierung des eIF4F-Komplexes an der 5'-Kappe gehemmt wird. Dementsprechend kommt es zu einer Hemmung der 5'-*Cap* abhängigen Translation. Unter guten Wachstumsbedingungen werden die 4E-BP an verschiedenen Stellen durch mTOR (*Mammalian Target of Rapamycin*) phosphoryliert. Diese Phosphorylierungen führen zur Dissoziation dieser

Bindeproteine von eIF4E, wodurch der eIF4F-Komplex assoziieren und die Proteinsynthese initiieren kann (Hay & Sonenberg 2004, Ma & Blenis 2009, Pópulo et al. 2012).

1.3.2 Regulation der Translationsinitiation durch *cis*-regulatorische Elemente

Viele Transkripte, die für regulatorische Proteine wie Wachstumsfaktoren und Transkriptionsfaktoren kodieren, haben häufig lange UTRs mit zahlreichen *cis*-regulatorischen Elementen, welche die Effizienz der Translationsinitiation regulieren. Hierzu zählen unter anderem Sekundärstrukturen in den 5'-UTRs, welche den scannenden 43S PIC verlangsamen oder blockieren können, wodurch die Wahrscheinlichkeit einer Erkennung des Startcodons erhöht wird (Kozak 1990).

Neben den Sekundärstrukturen sind in den 5'-UTRs zahlreicher Transkripte uORFs lokalisiert (Abschnitt 1.4.2), welche die Translation der hauptkodierenden Sequenz regulieren und im Extremfall komplett inhibieren können (Calvo et al. 2009, Morris 2000). Bei immer mehr zellulären mRNAs wird weiterhin ein vom 5'-*Cap* unabhängiger Translationsinitiationsmechanismus beschrieben, welcher die Translationsinitiation unter Stressbedingungen moduliert, wenn die globale *Cap*-abhängige Translation inhibiert ist. Bei diesem alternativen Mechanismus werden die Ribosomen von hochstrukturierten Sequenzen, den IRES, direkt zu den TISs rekrutiert (Fitzgerald & Semler 2009).

Zusätzlich befinden sich in den UTRs zahlreicher RNAs spezifische Sequenzen, an die regulatorische Proteine binden können. Ein Beispiel hierfür sind die *Iron-Response-Elements* (IRE) in der 5'-UTR von mRNAs, deren Proteine für den Eisenmetabolismus wichtig sind und die die Translation in Abhängigkeit von der Eisenkonzentration kontrollieren (Goossen et al. 1990).

1.3.3 Regulation der Startcodonwahl

Die korrekte Erkennung eines spezifischen Codons auf der mRNA als Startstelle für die Translation ist von essentieller Bedeutung für Zellen. Fehler bei der Initiation können z.B. die Entstehung von verkürzten oder verlängerten Proteinen zur Folge haben, welche andere katalytische Eigenschaften aufweisen oder in andere Kompartimente der Zelle transportiert

werden. Des Weiteren würde eine Initiation im falschen Leserahmen den Verlust eines Proteins mit sich bringen. Das kooperative Zusammenspiel von Initiationsfaktoren, den ribosomalen Untereinheiten und *cis*-regulatorischen Elementen der mRNA gewährleistet die korrekte Identifizierung der Initiationscodonen und deren Verwendung als Startstellen für die Translation.

Ob ein AUG-Codon als Initiationsstelle verwendet wird, hängt stark von seinem Nukleotidkontext ab. Für eine effiziente Translation in Säugetieren ist CGCC(A/G)CCAUGG die optimale Konsensussequenz, welche nach ihrer Entdeckerin als Kozak-Sequenz bezeichnet wird (Kozak 1986, Kozak 1987). Hierbei sind vor allem die Nukleotide an der Position -3 und +4 (relativ zum A vom AUG) von großer Bedeutung. Mutationen in dieser Sequenz führen zu einer starken Reduktion der Initiationsfrequenz (Kozak 1986). Somit kann ein am weitesten 5' gelegenes AUG von der kleinen ribosomalen Untereinheit überlesen werden, wenn es von einer schwachen Konsensussequenz flankiert wird. Diese Ausnahme zur *First-AUG-Regel* wird als *Leaky-Scanning* (Abschnitt 1.4.1) bezeichnet (Kozak 1986).

Die Entscheidung für eine Initiation an einem spezifischen Codon wird zusätzlich von mRNA-Sekundärstrukturen beeinflusst. Stabile Sekundärstrukturen, die sich 13-17 nt *downstream* vom Startcodon befinden, verzögern z.B. die Scan-Bewegung des Ribosoms und erhöhen somit die Wahrscheinlichkeit einer Initiation an einem AUG in einem suboptimalen Kontext (Kozak 1990, Kochetov et al. 2007). Überdies verschafft das verlangsamte *Scanning* einem non-AUG-Codon mehr Zeit, mit einer Met-tRNA_i zu hybridisieren, was zu einer Initiation der Translation an einen non-AUG-Codon führen kann (Kozak 2002).

Neben den *cis*-regulatorischen Elementen spielen auch translationale Initiationsfaktoren eine bedeutende Rolle bei der Wahl des Startcodons. *In vitro* und *in vivo* Experimente konnten zeigen, dass hierfür vor allem der Initiationsfaktor eIF1 essentiell ist (Pestova & Kolupaeva 2002, Ivanov et al. 2010). Dieser Faktor bindet in der Nähe der P-Stelle der kleinen ribosomalen Untereinheit (Lomakin et al. 2003) und begünstigt eine offene, scannende und die Initiation inhibierende Konformation des PIC (Passmore et al. 2007). Die Dissoziation von eIF1 bewirkt eine Konformationsänderung zu einem geschlossenen PIC. Dieser kann nicht scannen, was die Wahrscheinlichkeit für eine Erkennung des Startcodons erhöht (Nanda et al. 2009).

In weiteren Studien konnte der eukaryotische Initiationsfaktor eIF5 als Gegenspieler von eIF1 bei der Wahl des Startcodons identifiziert werden, dessen Überexpression in humanen

Zellen eine verstärkte Initiation an Startcodonen in einem nicht optimalen Konsensus und an non-AUG Startcodonen begünstigt (Loughran et al. 2011).

Der Initiationsfaktor eIF1A, dessen N- und C-terminale Enden in die P-Stelle der kleinen ribosomalen Untereinheit reichen, hat entgegengesetzte Effekte auf die Wahl des Startcodons. In der C-terminalen Domäne dieses Proteins befinden sich *Scanning-Enhancer*-Elemente, die mit eIF1 kooperieren um eine offene Konformation des PIC zu stabilisieren. Demgegenüber fördert das N-terminale Ende eine geschlossene Konformation und die Dissoziation von eIF1, was wiederum den Start der Translation begünstigt (Fekete et al. 2007, Yu et al. 2009).

1.4 Das Kodierungspotential von eukaryotischen Transkripten

1.4.1 Gesteigerte Proteinvielfalt durch *Leaky Scanning* und Reinitiation

Die vor einigen Jahren postulierte *First-AUG*-Regel vermittelte, dass das scannende Ribosom generell die Translation an dem 5'-proximalsten AUG-Codon initiiert (Kozak 1978), aber mittlerweile werden immer mehr Gene identifiziert, deren Translation von dieser Regel abweicht. Reinitiation und *Leaky Scanning* sind zwei Mechanismen, durch die das Umgehen der *First-AUG*-Regel bei der Translationsinitiation gewährleistet wird. So können aufgrund dieser Mechanismen neben den am weitesten 5'-gelegenen Startcodon auf einer mRNA noch zusätzliche *downstream* gelegene Codonen als Translationsinitiationsstellen verwendet werden (Kozak 2002). Reinitiation und kontextabhängiges *Leaky Scanning* tragen somit zur Erhöhung der Proteindiversität bei.

Beim *Leaky Scanning* können Startcodonen von den scannenden PICs überlesen werden. Das Überlesen ermöglicht die Produktion von zwei verschiedenen Proteinen von nur einer mRNA (Kozak 2002). Haben alle Startcodonen einer mRNA den gleichen Leserahmen, entstehen, wenn kein Stoppcodon in diesem Leserahmen in der 5'-UTR lokalisiert ist, N-terminal verkürzte oder verlängerte Isoformen eines Proteins, die z.B. in verschiedene subzelluläre Kompartimente sortiert werden (Leissring et al. 2004, Souciet et al. 1999) oder antagonistische Funktionen haben (Descombes & Schibler 1991, Kozak 2002). Das Ausmaß des *Leaky Scanning* ist stark vom Sequenzkontext abhängig. Wenn das am weitesten 5'-gelegene AUG-Startcodon von einer starken Kontextsequenz flankiert ist, stoppen alle scannenden Ribosomen an dieser Stelle und initiieren die Translation. Ist das

erste AUG-Codon jedoch von einer schwachen Kontextsequenz umgeben, führt dies zu einer Initiation von nur wenigen Ribosomen. In diesem Fall überlesen die meisten PICs diese Startstelle und stoppen an einem weiter *downstream* gelegenen Startcodon. Ein gut beschriebenes Beispiel für diesen Mechanismus ist die *c-myc* mRNA, von der drei verschiedene Proteinisoformen generiert werden. Aufgrund einer Initiation an zwei verschiedenen AUG-Codonen entstehen lange (65 kDa) und N-terminal verkürzte (50 kDa) *c-myc* Proteinisoformen (Spotts et al. 1997). Zusätzlich initiieren einige Ribosomen an einem *upstream* gelegenen non-AUG (CTG), was zur Bildung der dritten Proteinvariante (68 kDa) führt (Hann et al. 1988). Das Verhältnis der Proteinvarianten ist von großer Bedeutung für die normale *c-myc* Funktion in der Zelle. So konnte eine veränderte relative Expression der verschiedenen *c-myc* Proteinvarianten in einigen Tumorzelllinien nachgewiesen werden (Spotts et al. 1997).

1.4.2 uORFs

Mit Hilfe bioinformatischer Analysen wurden in 35-50% der Transkripte von Nagetieren und Mensch neben dem hauptkodierenden offenen Leserahmen ein oder mehrere kurze uORFs identifiziert (Iacono et al. 2005, Matsui et al. 2007, Abb. 3). Sie sind vor allem gehäuft in Genen zu finden, deren Produkte wichtige regulatorische Funktionen für Wachstum und Differenzierung haben (Kozak 1987). Diese *cis*-regulatorischen Elemente befinden sich in der 5'-untranslatierten Region von mRNAs und zeigen in ihrem Aufbau und ihrer Funktionsweise eine hohe Diversität. Sie variieren in ihrer Länge, Anzahl pro Transkript (0-13), ihrer Lage relativ zur 5'-Kappe und zum Startcodon, Sequenz und Sekundärstruktur. Des Weiteren unterscheidet man zwischen uORFs, welche *upstream* vom hauptkodierenden ORF enden oder mit diesem überlappen (Wethmar, Smink, and Leutz 2010). Eine wichtige Funktion dieser kleinen Elemente ist die Regulation der Translation an der hauptkodierenden Sequenz. Abhängig von den uORF-Eigenschaften kann diese um 30 bis 80% reduziert werden (Calvo, Pagliarini, and Mootha 2009). In einigen Fällen kann die Aktivierung eines uORFs zu einer kompletten Inhibition der Translation am *downstream* gelegenen Start führen (Morris & Geballe 2000).

Ein gut beschriebenes Beispiel, wie uORFs die Translation eines Proteins regulieren, ist der Transkriptionsfaktor ATF4, welcher unter Stressbedingungen in der Zelle translatiert wird (Harding et al. 2000, Harding et al. 2003). Die globale Proteinsynthese wird unter

Stressbedingungen, bedingt durch die Phosphorylierung der α -Untereinheit von eIF2, in Zellen inhibiert (Hershey 1991). Diese Phosphorylierung verhindert den Austausch des an eIF2 gebundenen GDP durch GTP. Infolge der Reduktion des eIF2-GTP-Levels kommt es zu einer Verringerung der TC-Bildung. Dennoch werden einige mRNAs als Antwort auf Stress verstärkt translatiert (Hinnebusch 1997, Vattem & Wek 2004). Der Transkriptionsfaktor ATF4 enthält zwei uORFs in seiner 5'-UTR. Unter normalen Bedingungen translatiert das Ribosom den ersten uORF. Da unter diesen Gegebenheiten der TC schnell wieder rekrutiert werden kann, reinitiiert die Ribosomen am zweiten inhibitorischen uORF und dissoziieren nach dessen Translation. Auch unter Stressbedingungen wird der erste uORF von ATF4 translatiert. Aufgrund der reduzierten TC-Menge benötigt das Ribosom aber mehr Zeit, bevor es durch die Bindung mit dem TC seine Initiationskompetenz wieder erhält. Infolgedessen scannt die ribosomale 40S Untereinheit am zweiten uORF vorbei und initiiert erst am ATF4-Startcodon. Das Resultat ist eine erhöhte ATF4-Proteinsynthese und eine ATF4-vermittelte Stressantwort der Zelle (Vattem & Wek 2004).

uORFs regulieren die Proteinexpression von tausenden Genen und Variationen in diesen Elementen sind mit phänotypischen Veränderungen und der Entstehung von Krebs (Liu et al. 1999) und anderen Krankheiten wie Thrombozythämie (Wiestner et al. 1998), Alzheimer (Zhou & Song 2006) und dem Unna-Syndrom (Wen et al. 2009) assoziiert. Da diese Elemente in vielen Transkripten identifiziert wurden, die wichtige regulatorische Proteine kodieren, kann man eine große physiologische Bedeutung annehmen (Kozak 1991, Calkhoven et al. 2002).

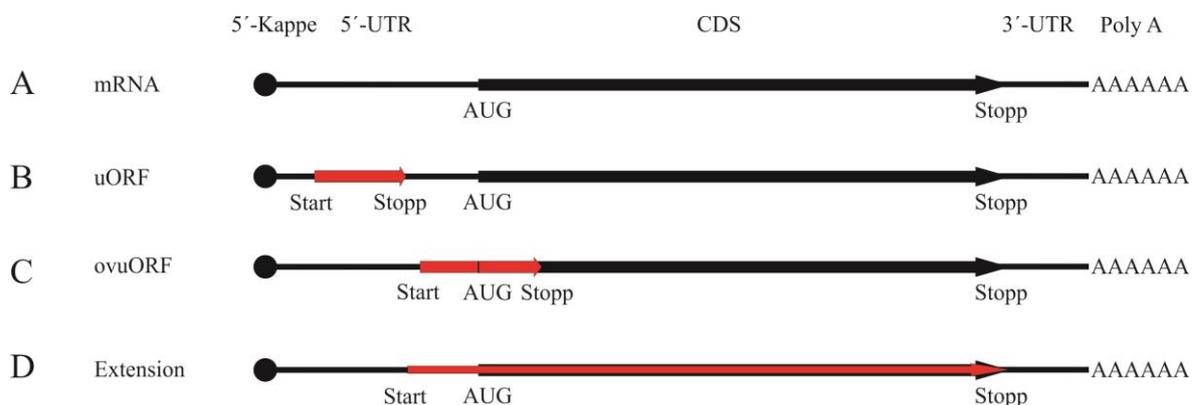


Abb. 3: Schematische Abbildung von alternativen TISs. A: In schwarz ist ein Transkript mit 5'-Kappe, 5'-UTR, CDS, 3'-UTR und Poly(A)-Ende dargestellt. Die proteinkodierende Sequenz ist als dicke Linie gezeichnet. Der rote Pfeil symbolisiert die Lage des zusätzlichen Leserahmens B: eines uORFs, C: eines ovuORFs, D: einer N-terminalen Proteinextension. Die Abbildung wurde modifiziert nach Calvo et al. (2009).

1.4.3 Alternative non-AUG Startcodonen

Immer mehr Studien belegen, dass die Translation nicht nur an AUGs initiiert sondern auch an *near-cognate* Codonen (= GUG, CUG, UUG, AAG, AGG, AUA, AUC, AUU und ACG), welche durch ein Einzelnukleotid vom AUG abweichen. In Säugetieren gibt es 23 experimentell nachgewiesene und beschriebene Gene, deren Transkripte von einem non-AUG translatiert werden (Tikole & Sankararamakrishnan 2006). In den meisten Fällen liegen diese non-AUG-Startstellen *upstream* vom ersten AUG-Codon, so dass durch deren Verwendung N-terminal verlängerte Proteinvarianten entstehen.

Experimente an embryonalen Mausstammzellen zeigten, dass vor allem CUG- und GUG-Codonen häufig als Startstellen verwendet werden, was auf eine wichtige biologische Relevanz dieser Codonen schließen lässt (Ingolia et al. 2011). Aber auch UUG, AAG, AGG, AUA, AUC, AUU und ACG fungieren als TISs (Ivanov et al. 2011). Die Effizienz mit der diese alternativen Startstellen genutzt werden, scheint taxonspezifisch zu sein (Peabody 1989) und variiert unter verschiedenen physiologischen Bedingungen. Dies unterstützt wiederum die Hypothese, dass die Initiation an non-AUG-Codonen eine wichtige regulatorische Rolle in eukaryotischen Zellen spielt (Ivanov et al. 2011).

Die Voraussetzung für eine Initiation an non-AUG-Codonen sind neben einem starken Sequenzkontext (Kozak 1989) Sekundärstrukturen, 15 nt *downstream* von der TIS, welche die Effizienz der Initiation an non-AUGs signifikant steigern (Kozak 1990). Mit Hilfe von *in vitro* und *in vivo* Experimenten konnte darüber hinaus belegt werden, dass eIF1 essentiell für die Unterscheidung für die Initiation an AUG und non-AUG-Codonen ist (Pestova & Kolupaeva 2002, Cheung et al. 2007). Eine Überexpression dieses Initiationsfaktors in Säugetierzellen führte zu einer reduzierten Initiation an non-AUG-Codonen (Ivanov et al. 2010), wohingegen eine Überexpression vom Gegenspieler eIF5 eine gesteigerte Translationsinitiation an non-AUGs mit sich bringt (Loughran et al. 2011).

1.5 Die Identifizierung von Translationsinitiationsstellen

Die systematische Identifizierung von TISs und uORFs basierte noch vor einiger Zeit ausschließlich auf bioinformatischen Untersuchungen von Transkriptsequenzen (Iacono et al. 2005, Volkova & Kochetov 2010, Bazykin & Kochetov 2011, Ivanov et al. 2011). Ein

wesentlicher Nachteil ist hierbei, dass alternative non-AUG-Startcodonen mit diesen Methoden nur schwer ermittelt werden können.

In einer Studie von 2009 wurde das Ribosomen-Profiling für *Saccharomyces cerevisiae* etabliert. Diese Methode basiert auf der Sequenzierung von durch Ribosomen geschützten mRNA-Fragmenten (Ingolia et al. 2009). Die vor einer Hydrolyse geschützten ca. 30 nt langen RNA-Bereiche werden als Ribosomen-Footprints oder *Ribosome-protected-Fragments* (RPFs) bezeichnet. Aufgrund der Möglichkeit, anhand der Footprint-Sequenzen die präzise Lage der Ribosomen auf einem mRNA-Molekül zu kartieren, kann man mit dieser Methode die proteinkodierenden Regionen von Transkripten definieren.

In einer Studie von 2011 wurde das Ribosomen-Profiling modifiziert und verwendet, um neue Initiationsstellen der Translation in embryonalen Mausstammzellen zu identifizieren (Ingolia 2011). Hierfür wurden die Ribosomen an den TISs angereichert, indem die Zellen mit dem Elongationsinhibitor Harringtonin behandelt wurden. Dieser Inhibitor bewirkt die Dissoziation von Ribosomen nach der ersten Elongationsrunde, wohingegen neu initiiierende Ribosomen nicht beeinflusst werden. Indem die Footprint-Sequenzen auf das Mausgenom kartiert wurden, konnten die TISs genomweit identifiziert werden. Ein entscheidender Vorteil dieser Methode ist die Möglichkeit auch alternative Startstellen akkurat zu detektieren, was zuvor allein mit Hilfe bioinformatischer Analysen nicht realisierbar war.

1.6 Zielsetzung

Zahlreiche Studien dokumentieren, dass das Kodierungspotential von Transkripten komplexer ist, als zuvor angenommen wurde, denn immer häufiger wird beschrieben, dass die Translation nicht nur an den annotierten AUG-Startcodonen der hauptkodierenden Sequenz, sondern auch an alternativen AUG- und non-AUG-Codonen initiiert. Diese alternativen Translationsereignisse führen zu uORFs, CDS überlappende uORFs und N-terminale Proteinextensionen, die von großer regulatorischer Relevanz sein können.

Bis vor einiger Zeit basierte die Annotation von proteinkodierenden Transkriptbereichen hauptsächlich auf bioinformatischen Analysen, wobei die relevanten alternativen TISs nicht berücksichtigt wurden. Aus diesem Grund soll im Rahmen der vorliegenden Arbeit eine transkriptomweite Übersicht von Translationsinitiationsstellen in einer humanen Monozyten-Zelllinie erstellt werden.

Hierfür soll als erstes das 2009 in Hefe beschriebene Ribosomen-Profiling für eine humane Zelllinie etabliert werden. Diese Methode basiert auf der Sequenzierung von durch Ribosomen geschützten mRNA-Fragmenten, nach einer Hydrolyse von nicht geschützten mRNA-Bereichen. Indem die sequenzierten Fragmente auf eine humane Referenzsequenz kartiert werden, sollen die tatsächlich translatierten Transkriptbereiche genau definiert und kategorisiert werden.

Um eine akkurate Identifizierung der TISs zu gewährleisten, sollen die Ribosomen-Footprints an den TISs angereichert werden, indem die Zellen mit dem Elongationsinhibitor Puromycin behandelt werden, welcher die Dissoziation elongierender Ribosomen bewirkt, wohingegen neu initiiierende Ribosomen nicht beeinflusst werden. Durch die anschließende Behandlung mit Cycloheximid werden neu initiiierende Ribosomen an den TISs arretiert. Mit Hilfe eines künstlichen neuronalen Netzes sollen diese TISs systematisch vorhergesagt und kategorisiert werden.

Zur Validierung der Ergebnisse sollen beschriebene und experimentell bestätigte alternative Translationsstarts, welche zu N-terminalen Proteinextensionen führen oder uORFs repräsentieren, mit den hier identifizierten TISs verglichen werden. Um die Richtigkeit der in dieser Arbeit vorhergesagten TISs zusätzlich zu fundieren, soll letztendlich der Konservierungsgrad der identifizierten Startcodonen zwischen nicht humanen Primaten und zwischen Mensch und Maus analysiert werden.

Letztendlich soll das Ribosomen-Profiling für ein humanes Gewebe optimiert und etabliert werden.

2. Material und Methoden

2.1 Allgemeine Chemikalien

Alle verwendeten Chemikalien wurden, wenn nicht anders angegeben, bei der Firma Carl Roth GmbH & Co. KG, Karlsruhe bestellt. Bidestilliertes Wasser und selbst hergestellte Lösungen wie Puffer und Medien wurden durch Autoklavieren oder Filtration durch einen 0,2 µm Cellulose Acetat Filter (VWR, Darmstadt, Deutschland) sterilisiert. Für Arbeiten mit RNA wurde RNase freies Wasser verwendet. Hierfür wurde bidestilliertes Wasser mit 0,1% (v/v) Diethylpyrocarbonat (DEPC) versetzt und über Nacht gerührt. Anschließend wurde DEPC durch Autoklavieren inaktiviert.

2.2 Oligonukleotide

Alle verwendeten Oligonukleotide wurden von der Firma Microsynth AG (Balgach, Schweiz) bezogen.

Für diese Arbeit kamen die in der Studie von Ingolia et al. 2009 beschriebenen Oligonukleotide zum Einsatz. Die Sequenzen, die Verwendung und spezifische Modifikationen sind im Anhang 6.1 aufgelistet.

2.3 Zellkultur

Alle Arbeitsschritte mit Zellen wurden unter sterilen Bedingungen unter einer Sicherheitswerkbank (Safe 2020, Thermo Scientific, Langenselbold, Deutschland) durchgeführt. Die verwendeten, in Suspension wachsenden, THP-1 Zellen wurden von der Deutschen Sammlung für Mikroorganismen und Zellkulturen GmbH (DSMZ, Braunschweig, Deutschland) bezogen. Diese humane Zelllinie wurde aus dem Blut eines, an akuter monozytärer Leukämie erkrankten, einjährigen, Jungen isoliert.

2.3.1 Kultivierung und Passagieren von Zellen

4×10^6 THP-1 Zellen wurden in 20 ml RPMI1640 Medium (PAA Laboratories GmbH, Cölbe, Deutschland) + 10% (v/v) fetalem Kälberserum (FCS, Biochrom AG, Berlin, Deutschland) ausgesät. Die Zellen wurden in 15 cm großen, mit Kollagen beschichteten Kulturschalen (BD Falcon™, Becton Dickinson GmbH, Heidelberg, Deutschland) in einem Brutschrank (Heraeus® BBD 6220, Thermo Scientific, Langenselbold, Deutschland) unter konstanten Bedingungen bei 37°C, einem CO₂-Gehalt von 5% und gesättigter Luftfeuchte für 2 Tage kultiviert. Anschließend erfolgte die Zellernte bzw. die Stimulation der Zellen.

Um das Wachstum der Zellen zu gewährleisten, waren ein regelmäßiger Mediumwechsel und das Passagieren der Zellen von der Stammhaltung erforderlich. Die Zellen wurden in ein 50 ml Zentrifugenröhrchen (Sarstedt AG & Co., Nümbrecht, Deutschland) überführt und 5 min bei 500 x g sedimentiert (Zentrifuge 5804 R, Eppendorf AG, Köln, Deutschland). Das Zellpellet wurde anschließend mit 37°C warmen PBS (PAA Laboratories GmbH, Cölbe, Deutschland) gewaschen und erneut für 5 min und 500 x g zentrifugiert. Anschließend wurde PBS entfernt und das Zellpellet in neuem vorgewärmten Medium + 10% (v/v) FCS resuspendiert. Abhängig von der gewünschten Zellzahl, der Wachstumsgeschwindigkeit und der Zelldichte wurde die Zellsuspension auf neue Kulturschalen aufgeteilt und mit der entsprechenden Menge Medium aufgefüllt.

2.3.2 Zellernte

Für die Ernte wurden die Zellen in 50 ml Reaktionsgefäße (Sarstedt AG & Co., Nümbrecht, Deutschland) überführt und für 5 min bei 4°C und 500 x g sedimentiert (Zentrifuge 5804 R, Eppendorf AG, Köln, Deutschland). Der Überstand wurde vorsichtig abgenommen und das Pellet mit 20 ml eiskalten PBS + 100 µg/ml (w/v) Cycloheximid gewaschen. Zellen aus vier 15 cm Kulturschalen wurden zu einem Ansatz vereinigt. Nach erneuter Zentrifugation wurde das Zellpellet in flüssigem Stickstoff eingefroren und bei -80°C bis zur weiteren Verwendung aufbewahrt oder direkt in 1 ml Polysomenpuffer für anschließende Arbeiten resuspendiert.

2.3.3 Zellzahlbestimmung mittels Neubauerzählkammer

Von der Zellsuspension wurden 10 µl für die Zellzahlbestimmung in ein neues Reaktionsgefäß überführt und mit 10 µl Trypanblaulösung (Sigma-Aldrich, Steinheim, Deutschland) versetzt. Mittels Trypanblau ist eine Unterscheidung von toten und lebenden Zellen möglich, da es nur die Zellmembranen von toten Zellen passieren kann und das Zytosol tiefblau anfärbt, was unter dem Inversmikroskop (Olympus CK2, Olympus GmbH, Hamburg, Deutschland) beobachtet werden kann. 10 µl dieser Lösung wurden in eine Neubauer Zählkammer (VWR International, Darmstadt, Deutschland) eingebracht und unter dem Mikroskop wurden vier Großquadrate ausgezählt. Anschließend konnte die Zellzahl für je 1 ml Suspension nach folgender Formel berechnet werden:

$$\text{Zellzahl [ml}^{-1}\text{]} = Z \times V \times 10^4$$

Z = durchschnittliche Zellzahl der Großquadrate
V = Verdünnungsfaktor = 2

2.3.4 Kryokonservierung und Revitalisierung von Zellen

Für die Konservierung von Suspensionszellen wurden diese in ein Zentrifugenröhrchen überführt und für 5 min bei 4°C und 500 x g pelletiert. Während der Zentrifugation wurden die Kryoröhrchen (CryoPure, Sarstedt, Nümbrecht, Deutschland) mit Einfriermedium (FCS + 10% (v/v) DMSO) befüllt. Nach vollendeter Zellsedimentation wurde der Überstand abgesaugt und das Zellpellet in das Einfriermedium überführt. Um die zelltoxische Wirkung von DMSO einzuschränken wurde auf Eis gearbeitet. Anschließend wurden die Kryoröhrchen in einen Kunststoffbehälter, der mit Isopropanol (Sigma-Aldrich, Steinheim, Deutschland) gefüllt war, gestellt und so bei -80°C eingefroren. Nach 24 h wurden die Röhrchen aus dem Isopropanol gefüllten Behälter entnommen und bei -80°C weggefroren. Für die Revitalisierung von Zellen wurde Medium + 20% (v/v) FCS auf 37°C erwärmt. Daraufhin mussten die Zellen in einem 37°C warmen Wasserbad (JB Aqua Plus, Grant Instruments, Shepreth, England) schnell aufgetaut und sofort in das vorgewärmte Medium überführt werden. Die Suspension wurde in Zellkulturschalen überführt und im Kulturschrank unter Standardbedingungen inkubiert.

2.3.5 Mycoplasmentest

Um Kontaminationen der Zellen mit Mycoplasmen zu überprüfen, kam das PCR Mycoplasma Test Kit I/C (Promokine, Heidelberg, Deutschland) zur Anwendung, nach dessen Standardprotokoll gearbeitet wurde. Diese Methode basiert auf der Amplifizierung von mycoplasmaspezifischer DNA. Hierfür wurde von der Zellkultur etwas Überstand abgenommen und als Template für eine PCR eingesetzt. Zusätzlich wurde eine Positiv- und Negativkontrolle durchgeführt, so dass die amplifizierten PCR-Produkte aller drei Proben auf einem 1%igen Agarosegel verglichen und Kontaminationen bestätigt oder ausgeschlossen werden konnten. Mit Mycoplasmen kontaminierte Kulturen wurden sofort entsorgt.

2.4 Gelelektrophorese von Nukleinsäuren

2.4.1 Agarosegelelektrophorese

10 x Ladepuffer

45% (v/v) Glycerin

0,02% (w/v) Bromphenolblau (Sigma-Aldrich, Steinheim, Deutschland)

0,02% (w/v) Xylencyanol (Sigma-Aldrich, Steinheim, Deutschland)

10 x TBE

1 M Tris

0,9 M Borsäure

0,01 M EDTA

Für die Auftrennung von PCR-Produkten kamen 1%ige, für die Auftrennung von Gesamt-RNA 1,2%ige Agarosegele zur Anwendung. Hierfür wurde die entsprechende Menge Agarose (Biozym Scientific GmbH, Oldendorf, Deutschland) in 0,5 x TBE-Puffer gelöst. Weiterhin wurden 0,00005% (w/v) Ethidiumbromid dazugegeben. Das erstarrte Gel wurde dann in eine mit 0,5 x TBE gefüllte Elektrophoresekammer (EasyPhor Medi, Biozym Scientific GmbH, Oldendorf, Deutschland) überführt, wo es mit den entsprechenden Proben beladen wurde. Die PCR-Produkte oder Gesamt-RNA-Proben mussten zusätzlich

mit einem 10-fach Ladepuffer versetzt werden. Als Größenmarker diente der SmartLadder (Eurogentec, Köln, Deutschland). Nach der Beladung des Gels wurde für 1 h ein elektrisches Feld bei 110 V angelegt (Elektrophoresespannungsgerät Consort EV243, Biozym Scientific GmbH, Oldendorf, Deutschland). Die Proben konnten danach mit Hilfe einer UV-Geldokumentationsanlage (AlphaImager[®] EP, Biozym Scientific GmbH, Oldendorf, Deutschland) optisch ausgewertet und fotografiert werden.

2.4.2 Denaturierende TBE-Urea-Gele

Für die Größenauftrennung von kleinen Footprint-RNAs, Footprint-cDNAs und Footprint-PCR-Reaktionen kamen 10%ige und 15%ige denaturierende TBE-Urea-Gele zum Einsatz. Die Gele wurden entweder selbst hergestellt oder es kamen Novex[®] TBE-Urea-Fertiggele (Life Technologies, Darmstadt, Deutschland) zum Einsatz.

<u>selbst hergestellte TBE-Urea-Gele</u>		<u>denaturierender 2 x Ladepuffer</u>	
8 M	Harnstoff	8 M	Harnstoff
1 x	10 x TBE	20 mM	EDTA
25% oder 37,5% (v/v)	Rotiphorese [®] Gel 40 (19:1)	5 mM	Tris pH 7,5
0,05% (v/v)	TEMED	0,02% (w/v)	Bromphenolblau
0,6% (v/v)	10% APS	0,02% (w/v)	Xylencyanol

Die Fertiggele wurden in eine mit 1 x TBE Laufpuffer gefüllte XCell SureLock[™] Mini-Cell (Life Technologies, Darmstadt, Deutschland), die selbst hergestellten Gele in eine vertikale Polyacrylamid-Gelelektrophorese-Apparatur (Biometra GmbH, Göttingen, Deutschland) überführt. Anschließend wurde für 30 min ein elektrisches Feld bei 170 Volt angelegt. Während dieser Zeit wurden die Proben vor dem Beladen mit einem denaturierenden 2 x Ladepuffer versetzt. Bromphenolblau und Xylencyanol diente dabei zur Visualisierung der Lauffront der Nukleinsäuren. Nach der Auftrennung der Nukleinsäuren bei 170 V wurde das Gel in 100 ml 1 x TBE + 0,01% (v/v) SYBR[®] Gold (Life Technologies, Darmstadt, Deutschland) für 10 min gefärbt und anschließend mit Hilfe einer UV-Geldokumentationsanlage optisch ausgewertet und fotografiert.

2.5 Isolierung und Aufreinigung von Nukleinsäuren

2.5.1 Phenol-Chloroform-Extraktion von RNA und cDNA

Um RNA und DNA zu isolieren, kam die Phenol-Chloroform-Extraktion zum Einsatz, wobei für RNA-Extraktionen Roti[®]-Aqua-Phenol und für DNA-Extraktionen Roti[®]-Phenol verwendet wurde. Bei dieser Methode werden Nukleinsäurelösungen nacheinander mit einem gleichen Volumen Phenol (oder Aquaphenol), Phenol-Chloroform-Isoamylalkohol (25:24:1, v/v) und Chloroform-Isoamylalkohol (24:1, v/v) versetzt. Zwischen den einzelnen Mischschritten wurden die Ansätze für 10 min bei 16.000 x g und 4°C zentrifugiert. Die wässrige Phase, welche die Nukleinsäuren enthält, wurde vorsichtig in ein neues Reaktionsgefäß überführt. Die Interphase, in der sich Proteine sammeln sowie die untere organische Phase wurden verworfen.

Im Anschluss wurde eine Alkoholfällung durchgeführt (Abschnitt 2.5.3). Die gefällte DNA oder RNA wurde darauf folgend mit 1 ml 70% (v/v) Ethanol gewaschen, getrocknet und in einem gewünschten Volumen DEPC-Wasser resuspendiert.

2.5.2 Isolierung von Gesamt-RNA mittels Trizol

Für die Isolierung von Gesamt-RNA aus Zellen oder Geweben kam das Reagenz TRIzol[®] (Life Technologies, Darmstadt, Deutschland) zur Anwendung. Hierbei handelt es sich um eine Lösung aus Phenol und Guanidin-Isothiozyanat, so dass das Prinzip der RNA-Isolation mittels Phenol-Chloroform entspricht. Die geernteten Zellen, bzw. das zerkleinerte Gewebe wurden in 1 ml Trizol resuspendiert und für 10 min bei Raumtemperatur inkubiert. Es folgte die Zugabe von 200 µl Chloroform und ein erneuter Inkubationsschritt für 5 min bei Raumtemperatur. Nach der anschließenden Zentrifugation für 10 min bei 4°C und 12.000 x g wurde die obere wässrige Phase, welche die RNA enthielt, in ein neues Reaktionsgefäß überführt. Durch Zugabe von der gleichen Volumeneinheit Isopropanol fällte diese bei guter Durchmischung der Lösung innerhalb von 10 min bei Raumtemperatur spontan aus. Durch Zentrifugation für 5 min bei 4°C und 7500 x g wurde die RNA pelletiert. Um Salze und Isopropanolreste zu entfernen wurde das RNA-Pellet mit 1 ml 75% (v/v) Ethanol gewaschen, anschließend getrocknet und in 30 µl DEPC-Wasser gelöst. Um die Qualität der RNA zu überprüfen, erfolgte eine

Konzentrationsbestimmung (Abschnitt 2.5.5) und eine Gelauswertung der RNA (Abschnitt 2.4.1).

2.5.3 Alkoholpräzipitation von Nukleinsäuren

Eine häufige Methode der Molekularbiologie zur Entsalzung und Konzentrierung von Nukleinsäuren ist die Alkoholfällung. Hierfür wurden die Nukleinsäurelösungen mit einem gleichen Volumen Isopropanol und 1/10 Volumen 3 M NaAc, pH 4,5 (für RNA) oder 5 M NaCl (für DNA) versetzt und bei -20°C über Nacht oder bei -80°C für 30 min inkubiert. Für die Fällung von den niedrig konzentrierten Footprint-Fragmenten wurden zusätzlich 50 µg Glykogen (Life Technologies, Darmstadt, Deutschland) zu der Lösung gegeben. Glykogen hat ein ähnliches Präzipitationsverhalten wie die Nukleinsäuren, so dass diese mit einer erhöhten Effizienz co-präzipitiert wurden. Im Anschluss erfolgte ein Zentrifugationsschritt für 10 min bei 4°C und 16.000 x g. Das sedimentierte DNA- oder RNA-Pellet wurde anschließend mit 1 ml 70% (v/v) Ethanol gewaschen, erneut für 10 min bei 4°C und 16.000 x g zentrifugiert, luftgetrocknet und in einem gewünschten Volumen DEPC-Wasser resuspendiert.

2.5.4 Qualitätskontrolle der RNA mittels Agarosegelelektrophorese

Zur Überprüfung der Gesamt-RNA-Qualität wurde 1 µg RNA auf einem 1,2%igen Agarosegel bei 110 V für 45 Minuten aufgetrennt (Abschnitt 2.5.1). Ein Indiz für intakte RNA sind zwei Banden, die bei 1,9 und 4,7 kb deutlich auf dem Gel erkennbar sein müssen. Diese Banden repräsentieren die 18S und 28S ribosomale RNA. Im Gegensatz dazu sind unsaubere und unklare Banden ein Hinweis für degradierte oder kontaminierte RNA-Proben.

2.5.5 Konzentrationsbestimmung von Nukleinsäuren

Die Konzentrationsbestimmung der Nukleinsäureproben erfolgte mittels spektrophotometrischer Messung (NanoPhotometer[®], Implem GmbH, München,

Deutschland), bei der die optische Dichte der Probe bei 260 nm gemessen wurde. Ein OD-Wert von 1 bei einer Wellenlänge von 260 nm entspricht einer Konzentration von 50 µg/ml DNA oder 40 µg/ml RNA. Somit lässt sich die Nukleinsäurekonzentration nach folgender Formel berechnen:

- DNA-Konzentration (µg/ml) = $OD_{260} \times 50 \text{ µg/ml} \times \text{Verdünnungsfaktor}$
- RNA-Konzentration (µg/ml) = $OD_{260} \times 40 \text{ µg/ml} \times \text{Verdünnungsfaktor}$

Für die Referenzmessung wurde das Lösungsmittel der Probe verwendet.

Des Weiteren war die Bestimmung der ungefähren Reinheit von den Proben durch die Berechnung von OD-Verhältnissen möglich. Hierfür wurde zusätzlich die optische Dichte der Probe bei 230 nm (Absorptionsmaximum von Polysacchariden), und 280 nm (Absorptionsmaximum von Proteinen) gemessen. Das ideale Verhältnis aus OD_{260}/OD_{280} liegt für RNA zwischen 1,8 und 2,0 und für DNA zwischen 1,8 und 1,95. Kleinere Werte deuten auf mögliche Verunreinigungen durch Proteine oder Phenol hin. Eine hohe Absorption bei 230 nm ist ein Indiz für Verunreinigungen von Kohlenhydraten. Eine Probe ist frei von Polysacchariden und gilt als rein, wenn der Quotient der OD-Werte bei 260 nm und 230 nm $\geq 2,0$ ist.

2.6 Untersuchung der Translationsinitiation

Für die vorliegende Arbeit wurde die humane Monozyten-Zelllinie THP-1 verwendet und mit Cycloheximid (Abschnitt 2.6.1) oder mit Puromycin und Cycloheximid (Abschnitt 2.6.2) versetzt. Cycloheximid bewirkt eine Fixierung von elongierenden Ribosomen auf der mRNA, wohingegen die kombinierte Behandlung der Zellen mit Puromycin und Cycloheximid zu einer Dissoziation von elongierenden Ribosomen von den Transkripten und zu einer Fixierung und Anreicherung von neu initiiertender Ribosomen an den TISs führt. Darauf folgend wurde die mRNA hydrolysiert (Abschnitt 2.6.4), so dass nur die von Ribosomen geschützten, ca. 30 nt langen Footprint-Fragmente erhalten blieben (Abb. 4). Diese wurden aufgereinigt (Abschnitt 2.6.6), enzymatisch prozessiert, amplifiziert (Abschnitt 2.7) und sequenziert (Abschnitt 2.9).

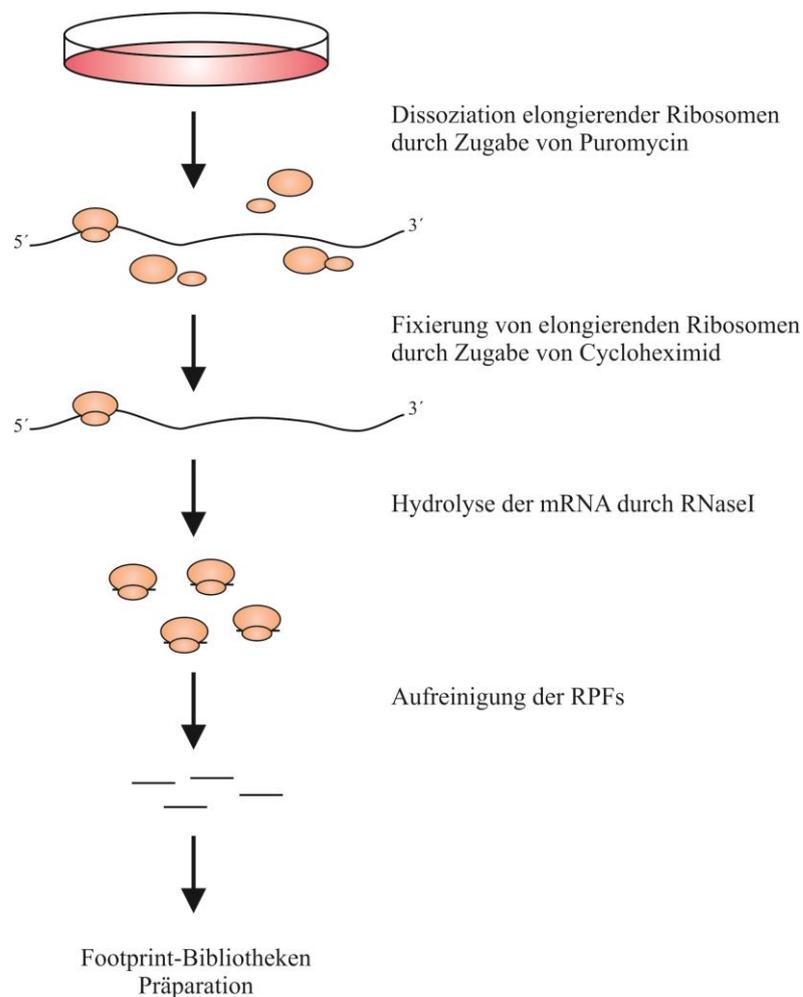


Abb. 4: Schematische Zusammenfassung der Methode

2.6.1 Fixierung von Ribosomen auf der mRNA

Um elongierende Ribosomen auf der mRNA zu fixieren, wurden die Zellen 48 h nach dem aussäen mit 100 µg/ml (w/v) Cycloheximid behandelt und für 2 min inkubiert. Cycloheximid bindet an die E-Stelle der großen ribosomalen Untereinheit, wodurch die Ablösung der deacetylierten tRNA verhindert wird. Infolgedessen kommt es zu einer Inhibition der Ribosomentranslokation und somit zu einer Hemmung der Elongation (Schneider-Poetsch et al. 2010). Im Anschluss erfolgte die Zellernte.

2.6.2 Anreicherung von Ribosomen an TISs

Für die Anreicherung von Ribosomen an den TISs wurden die Zellen 48 h nach dem aussäen mit 100 µg/ml Puromycin (InvivoGen, San Diego, USA) behandelt und für 15 min bei 37°C inkubiert. Puromycin ist ein Elongationsinhibitor, welcher als Aminoacyl-tRNA-Analoga fungiert und an Stelle einer Aminosäure in die wachsende Peptidkette eingebaut wird. Dieser Einbau führt zum Abbruch der Elongation und letztendlich zur Dissoziation von elongierenden Ribosomen. Die Initiation neuer Ribosomen wird hierbei nicht beeinflusst. Durch die anschließende Zugabe von 100 µg/ml Cycloheximid für 2 min wurden die neu initiiierenden Ribosomen an den TISs fixiert. Infolgedessen kam es zu einer Anreicherung von Ribosomen an den Translationsstartstellen. Danach erfolgte die Zellernte.

2.6.3 Isolierung von Polysomen und Monosomen

Polysomenpuffer

20 mM	Tris / HCl, pH 8
35 mM	MgCl ₂
140 mM	KCl
1% (v/v)	Triton X-100

Für die Isolierung von Polysomen und Monosomen aus THP-1 Zellen wurde das Zellpellet aus ursprünglich vier 15 ml Kulturschalen in 1 ml Polysomenpuffer resuspendiert.

Unmittelbar vor der Verwendung wurde eine Tablette Complete ULTRA (Roche, Ponsberg, Deutschland) pro 10 ml und 100 µg/ml (w/v) Cycloheximid zum Polysomenpuffer hinzugefügt. Das Zellysate wurde zehnmal durch eine 19 Gauge Nadel (B. Braun Melsungen AG, Melsungen, Deutschland) gezogen und anschließend für 10 min auf Eis inkubiert. Nach einer Zentrifugation für 8 min bei 4°C und 1300 x g wurde der Überstand, welcher die Polysomen bzw. Monosomen enthielt, in ein neues Reaktionsgefäß überführt.

2.6.4 Hydrolyse der Polysomen mit verschiedenen Ribonukleasen

Um die RNA-Bereiche, welche nicht durch gebundene Ribosomen geschützt sind, optimal zu hydrolysieren, wurden 1 ml Polysomensuspensionen mit verschiedenen Konzentrationen von RNasen unterschiedlicher Hersteller versetzt (Tab. 1). Anschließend wurde der Reaktionsansatz für 60 min unter leichtem Schütteln bei 30°C oder 37°C inkubiert. Die verwendete Endonuklease RNaseA hydrolysiert einzelsträngige RNA-Moleküle nur 3' von Pyrimidinbasen, wohingegen die Endonuklease RNaseI die Spaltung jeder RNA-Dinukleotidbindung katalysiert. Dementsprechend war zu erwarten, dass die gesamte RNA durch RNaseI hydrolysiert wird, ausgenommen der 30 nt langen RPFs, welche durch die gebundenen Ribosomen bedeckt werden. Eine Kontrolle wurde mit 4 µl des Ribonukleaseinhibitors SUPERaseIn, (Life Technologies, Darmstadt, Deutschland) versetzt. Der Erfolg der Polysomenhydrolyse wurde im Anschluss mit Hilfe von Polysomen-Profilen beurteilt.

Tabelle 1: Ribonukleasen unterschiedlicher Hersteller. Die Unitangaben wurden von den Herstellern individuell definiert und sind somit nicht direkt miteinander vergleichbar.

Enzym	Hersteller	Units / µl	hydrolysiert
RNaseI	Life Technologies, Darmstadt, Deutschland	100	jedes RNA-Dinukleotid
RNaseI	Thermo Fisher Scientific, Langenselbold, Deutschland	10	jedes RNA-Dinukleotid
RNaseI	Epicentre, Wisconsin, USA	10	jedes RNA-Dinukleotid
RNaseA	Thermo Fisher Scientific, Langenselbold, Deutschland	50	3' von Pyrimidinen

2.6.5 Dichtegradientenzentrifugation und Fraktionierung von Monosomen und Polysomen

10 x Gradientenpuffer

200 mM	Tris / HCl, pH 8
1,4 M	KCl
100 mM	MgCl ₂

10%ige Saccharoselösung

0,3 M
1 x
0,5 mM
100 µg/ml
20 U/ml

50%ige Saccharoselösung

1,5 M	Saccharose
1 x	10 x Gradientenpuffer
0,5 mM	DTT
100 µg/ml	Cycloheximid
20 U/ml	SUPERaseIn

Jeweils 1 ml der verdauten Monosomenextrakte bzw. der unverdauten Kontrollpolysomenextrakte wurden vorsichtig auf einen linearen 10-50% Saccharosedichtegradienten geschichtet und für 3 h bei 30.000 rpm und 4°C in einer Ultrazentrifuge (Optima™ L-90K Ultracentrifuge, Beckman Coulter GmbH, Krefeld, Deutschland) unter Verwendung des Rotors SW 32.1 Ti (Beckman Coulter GmbH, Krefeld, Deutschland) zentrifugiert. Dabei wandert die hoch translatierte „schwere“ Polysomen-RNA weiter durch den Gradienten als die Monosomen-RNA.

Für die Herstellung von Saccharosedichtegradienten wurde zunächst eine 10%ige (w/v) und eine 50%ige (w/v) Saccharoselösung in Gradientenpuffer hergestellt. 8 ml der 10%igen Saccharoselösung wurden in ein 17 ml Ultrazentrifugenröhrchen (Ultra Clear, Beckman Coulter GmbH, Krefeld, Deutschland) überführt und mit 8 ml der 50%igen Lösung vorsichtig unterschichtet, so dass zwei deutlich voneinander getrennte Schichten entstanden. Mit Hilfe eines Gradientenmischers (Gradient Master™, BioComp Fredericton, Kanada) wurden die Schichten so durchmischt, dass ein linearer, 10-50%iger Saccharosedichtegradient entstand, welcher vor der Verwendung über Nacht bei 4°C gelagert wurde. Im Anschluss an die Zentrifugation wurden die Gradienten mit Hilfe eines Gradientenfraktionierungssystems (Teledyne Isco, Lincoln, NE, USA) in ca. 500 µl Fraktionen getrennt. Diese Fraktionen wurden alle 30 s in einem neuen Reaktionsgefäß

aufgefangen. Parallel wurde kontinuierlich die Absorption der RNA bei 254 nm gemessen und aufgezeichnet. Fraktionen, welche die Monosomen enthielten, wurden vereinigt.

2.6.6 Anreicherung und Aufreinigung der RPFs

Ablösungspuffer

20 mM	HEPES-KOH
100 mM	KCl
10 mM	EDTA
2 mM	DTT
20 U/ml	SUPERaseIn

Für die Konzentrierung der RPFs wurden die Monosomenfraktionen auf Zentrifugalfilter (Amicon Ultra-4 Centrifugal Filters 100K, Millipore, Billerica, MA, USA) überführt und für 30 min bei 1.900 x g und 4°C zentrifugiert (Centrifuge 5804R, Eppendorf AG, Köln, Deutschland). Anschließend wurde 500 µl eiskalter Ablösungspuffer auf den Filter gegeben und der Ansatz wurde für 10 min auf Eis inkubiert. Darauf folgte ein erneuter Zentrifugationsschritt für 15 min bei 1.900 x g und 4°C, wobei das Filtrat in einem neuen Reaktionsgefäß aufgefangen wurde. Das Filtrat wurde dann mit 25 µl 20% SDS (Serva, Heidelberg, Deutschland) und 3 U Proteinase K (NEB, Frankfurt a.M., Deutschland) versetzt und für 30 min bei 42°C in einem Heizblock (Thermomixer comfort, Eppendorf AG, Köln, Deutschland) inkubiert. Mittels Phenol-Chloroform-Extraktion (Abschnitt 2.5.1) wurden die RPFs isoliert, anschließend gefällt (Abschnitt 2.5.3) und letztendlich in 18 µl DEPC-Wasser resuspendiert.

2.7 Footprint-Bibliotheken Präparation

Die vor der Hydrolyse geschützten und aufgereinigten RNA-Fragmente (Abschnitt 2.6) wurden nach einer von Ingolia et al. 2009 und Guo et al. 2010 beschriebenen aber etwas modifizierten Methode zu sequenzierbaren cDNA-Bibliotheken aufgearbeitet (Abb. 5). Hierfür ist eine Dephosphorylierung der 3'-Enden der Fragmente (Abschnitt 2.7.1) als Basis für eine anschließende Polyadenylierung (Abschnitt 2.7.3) erforderlich. Das Poly(A)-Ende diente darauf folgend als Primerbindestelle für die cDNA-Synthese (Abschnitt 2.7.4). Nach dieser Erststrangsynthese trägt die cDNA am 3'-Ende einen Adapter. Für eine Amplifizierung der Footprint-Fragmente müssen diese aber 3' und 5' von Primerbindestellen flankiert sein. Aus diesem Grund wurde die cDNA mittels einer Ligation zirkularisiert (Abschnitt 2.7.6) und anschließend an dem Spacer Furan geschnitten und infolgedessen wieder relinearisiert (Abschnitt 2.7.7). Die Footprint-Fragmente wurden dadurch von Adaptersequenzen flankiert, die als Primerbindestellen für eine anschließende Amplifizierung (Abschnitt 2.7.8) dienen. Letztendlich wurden die Footprint-Bibliotheken sequenziert (Abschnitt 2.9).

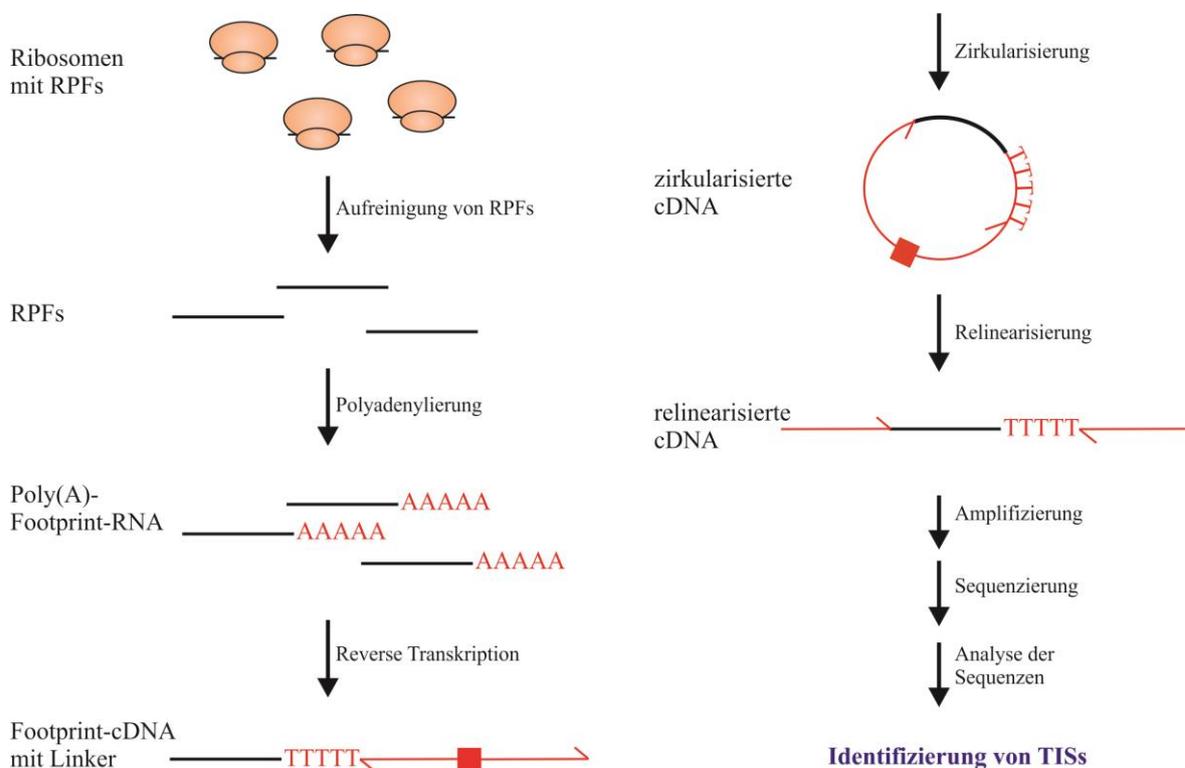


Abb. 5: Schematische Zusammenfassung der Präparation von Footprint-Bibliotheken

2.7.1 Dephosphorylierung der Footprint-RNA

Da die RPFs aufgrund der Hydrolyse durch RNaseI (Abschnitt 2.6.4) einen 3'-Phosphatrest aufweisen, wurden diese mit der T4-Polynukleotidkinase (NEB, Frankfurt a.M., Deutschland) dephosphoryliert. Neben der Kinaseaktivität hat dieses Enzym in Puffern ohne ATP eine 3'-Phosphataseaktivität, die man sich hier zunutze machte. Nach der Denaturierung der RNA für 3 min bei 75°C wurde der Reaktionsansatz auf Eis zusammengefügt und für 1 h bei 37°C inkubiert. Im Anschluss erfolgte die Inaktivierung des Enzyms bei 75°C für 5 min.

- Eine 20 µl Dephosphorylierungsreaktion enthielt die folgenden Substanzen:

16,5 µl	Footprint-RNA (Abschnitt 2.6.6)
1 x	PNK-Puffer
10 U	SUPERaseIn
10 U	T4-PNK (NEB, Frankfurt a.M., Deutschland)

2.7.2 Größenselektion und Aufreinigung der dephosphorylierten Footprint-RNA

RNA-Geleluierungspuffer

300 mM	NaAc
1 mM	EDTA
0,1 U/µl	SUPERaseIn

Frühere Studien zeigten, dass Ribosomen ca. 28 bis 30 nt lange RNA-Bereiche bedecken (Ingolia et al. 2009, Guo et al. 2010). Aus diesem Grund wurde die dephosphorylierte Footprint-RNA auf einem denaturierendem, 15%igen TBE-Urea-Gel aufgetrennt (Abschnitt 2.4.2). Der Gelbereich mit RNA-Banden zwischen 27 bis 34 nt wurde ausgeschnitten. Zur Größenkontrolle kamen die RNA-Oligonukleotide RNA_27nt und RNA_33nt zum Einsatz. Die ausgeschnittenen Gelstücke wurden in flüssigem Stickstoff gefroren und mittels einer Schwingmühle (Retsch MM301, Retsch GmbH, Haan, Deutschland) und unter Verwendung einer Stahlkugel (Retsch GmbH, Haan, Deutschland) physikalisch zerkleinert (30 s, 30 Schwingungen / s). Zur Rückgewinnung der RNA wurde

das zerkleinerte Gel über Nacht bei Raumtemperatur in 500 µl RNA-Geleluierungspuffer leicht geschüttelt. Im Anschluss erfolgte eine Phenol-Chloroform-Extraktion (Abschnitt 2.5.1) und Alkoholfällung der RNA (Abschnitt 2.5.3). Zum Resuspendieren der gefällten RNA kamen 5 µl DEPC-Wasser zum Einsatz.

2.7.3 Polyadenylierung der Footprint-RNA

Voraussetzung für eine cDNA-Synthese mit Oligo-dT-Primern ist ein Poly(A)-Schwanz am 3'-Ende der RNA, welcher als Primerbindestelle dient. Die Synthese dieses Poly(A)-Endes wurde durch die RNA abhängige *E.coli* Poly(A)-Polymerase (NEB; Frankfurt a.M., Deutschland) katalysiert. Es wurden 4,09 µl von der gelaufgereinigten Footprint-RNA (Abschnitt 2.7.2) in einer 6,6 µl Reaktion polyadenyliert. Dieser Reaktionsansatz wurde für 1 h bei 37°C inkubiert und enthielt die folgenden Komponenten:

4,09 µl	dephosphorylierte Footprint-RNA (Abschnitt 2.7.1)
1 x	Poly(A)-Puffer
0,15 mM	ATP
0,75 U/µl	SUPERaseIn
3 U	<i>E. coli</i> Poly(A)-Polymerase (NEB, Frankfurt a.M., Deutschland)

2.7.4 Footprint-cDNA-Synthese

Mittels Reverser Transkription wird die polyadenylierte RNA in eine einzelsträngige cDNA (*Complementary DNA*) umgeschrieben. Für diese Reaktion ist ein zur RNA komplementärer Primer nötig, der als Startpunkt für die cDNA-Synthese dient. Hier kam der Oligo-dT-Primer oNTI223 (5'-pGAT CGT CGG ACT GTA GAA CTC TØC AAG CAG AAG ACG GCA TAC GAT TTT TTT TTT TTT TTT TTT TVN-3') zum Einsatz, wobei p eine 5'-Phosphorylierung, Ø den dSpacer Furan, V ein A-, C- oder G-Nukleotid und N ein zufälliges Nukleotid symbolisiert (Ingolia et al. 2009).

Für die Umschreibung der Footprint-RNA in cDNA wurden 5 µl der Polyadenylierungsreaktion (Abschnitt 2.7.3) direkt als Template für eine Reverse Transkription in einem 14,25 µl Denaturierungsansatz aufgenommen (Tab. 2), welcher für

3 min bei 75°C inkubiert wurde. Im Anschluss wurde auf Eis SUPERaseIn, DTT, und Superscript III Reverse Transkriptase (Life Technologies, Darmstadt, Deutschland) zugegeben. Nach einer Inkubation von 30 min bei 48°C erfolgte die Eliminierung der Template-RNA durch Zugabe von 1 M NaOH und einer Inkubation von 15 min bei 98°C. Durch das Hinzufügen von 1 M HCl wurde der Reaktionsansatz wieder neutralisiert.

Tabelle 2: Zusammensetzung und Inkubationszeiten der Reversen Transkription. Die Konzentrationsangaben beziehen sich auf das Gesamtvolumen von 16,5 µl.

Komponente	Konzentration / Volumen	Inkubation
Poly(A)-RNA	5 µl	3 min bei 75°C Denaturierung
Tris, pH 8	37 mM	
dNTPs	0,5 mM	
Primer oNTI223	0,3 µM	
SUPERaseIn	10 U	30 min bei 48°C Reverse Transkription
DTT	5 mM	
SuperScript III RT	164 U	
Gesamtvolumen	16,5 µl	
NaOH	90 mM	15 min bei 98°C RNA-Eliminierung
HCl	90 mM	Neutralisation

2.7.5 Größenselektion und Aufreinigung der Footprint-cDNA

Um die Footprint-cDNA aufzureinigen, wurde diese auf einem 10%igen denaturierendem TBE-Urea-Gel aufgetrennt (Abschnitt 2.4.2). Der Gelbereich mit der 94 nt langen cDNA wurde ausgeschnitten, in flüssigem Stickstoff gefroren und mittels einer Schwingmühle (Retsch MM301, Retsch GmbH, Haan, Deutschland) und unter Verwendung einer Stahlkugel (Retsch GmbH, Haan, Deutschland) physikalisch zerkleinert (30 s, 30 Schwingungen / s). Zur Rückgewinnung der cDNA wurde das zerkleinerte Gel über Nacht bei Raumtemperatur in 500 µl cDNA-Geleluierungspuffer (300 mM NaCl, 1 mM EDTA) leicht geschüttelt. Der gesamte Ansatz wurde daraufhin auf ein Illustra MicroSpin G-50 Säulchen (GE Healthcare, München, Deutschland) aufgetragen. Durch Zentrifugation bei 500 x g für 2 min wurden Gelreste entfernt. Anschließend wurde die cDNA im Filtrat gefällt (Abschnitt 2.5.3). Die mit 70%igen (v/v) Ethanol gewaschene und dann

luftgetrocknete cDNA wurde direkt in Zirkularisierungspuffer resuspendiert (Abschnitt 2.7.6).

2.7.6 Zirkularisierung der Footprint-cDNA

Im Anschluss an die Gelaufreinigung (Abschnitt 2.7.5) und Fällung der Footprint-cDNA wurde diese zirkularisiert. Hierfür kam die ssDNA CircLigase (Epicentre, Madison, Wisconsin, USA) zum Einsatz. Dieses Enzym ligiert die 5'-Phosphat- und 3'-Hydroxylenden von einzelsträngigen DNA-Molekülen. Für diese Arbeit wurde das gefällte cDNA-Pellet direkt in 4,5 µl Zirkularisierungs-Mix resuspendiert. Zusätzlich wurden 50 U Ligase zur Reaktion gegeben, welche 1 h bei 60°C inkubiert wurde. Im Anschluss erfolgte die Hitzeinaktivierung des Enzyms bei 80°C für 10 min.

- Der gesamte 5 µl Reaktionsansatz enthielt die folgenden Substanzen:

	Footprint-cDNA
1 x	Zirkularisierungspuffer
50 µM	ATP
2,5 mM	MgCl ₂
50 U	ssDNA CircLigase (Epicentre, Madison, Wisconsin, USA)

2.7.7 Relinearisierung der zirkularisierten Footprint-cDNA

Die zirkulare, einzelsträngige cDNA (Abschnitt 2.7.6) wurde durch die Zugabe von 6,25 µl Relinearisierungs-Mix und 12 U der Endonuklease APE1 (NEB, Frankfurt a.M., Deutschland) während einer 1 stündigen Inkubation bei 37°C wieder relinearisiert. Da APE1 an dem modifizierten dSpacer Furan schneidet, wird die zirkulare cDNA an einer anderen Stelle geschnitten, als sie zuvor ligiert wurde. Aus diesem Grund können neue Adaptoren geschaffen werden, die das Footprint-Fragment 5' und 3' flankieren (Abb. 5).

- Ein 12,45 µl Relinearisierungsansatz enthielt die folgenden Substanzen:

5 µl	Zirkularisierungsreaktion (Abschnitt 2.7.6)
50 mM	KCl
1 mM	DTT
12 U	APE1 (NEB, Frankfurt a.M., Deutschland)

2.7.8 Amplifizierung der RPFs mittels PCR

Die Polymerase-Kettenreaktion (engl.: *Polymerase-Chain-Reaction*, PCR) ist eine einfache Methode, mit der spezifische DNA-Abschnitte stark vermehrt werden können. Am Anfang jeder PCR-Reaktion wurde die cDNA bei 94 °C denaturiert. Während der folgenden *Annealing*-Phase richtete sich die Temperatur nach der Schmelztemperatur (T_M) der verwendeten Oligonukleotide. Die Synthese des komplementären Stranges wurde durch die Platinum[®] *Taq* DNA Polymerase High Fidelity (Life Technologies, Darmstadt, Deutschland) bei 68°C katalysiert. Alle Reaktionen wurden zyklisch mehrmals wiederholt, so dass die amplifizierte DNA-Menge mit jeder Wiederholung exponentiell zunahm. Die PCR-Reaktion wurde in einem Thermocycler T3000 (Biometra, Göttingen, Deutschland) durchgeführt. Die verwendeten Komponenten des Reaktionsansatzes sowie das PCR-Programm sind der Tabelle 3 zu entnehmen.

Tabelle 3: PCR-Reaktionsansatz und PCR-Programm: Für die PCR kam die Platinum[®] *Taq* DNA Polymerase High Fidelity und die dazugehörigen Puffer (Life Technologies, Darmstadt, Deutschland) zum Einsatz. Informationen zu den verwendeten Primern finden sich im Anhang 6.1.

Komponente	Konzentration / Volumen	Temperatur [°C]	Zeit	Zyklusanzahl
High Fidelity PCR Puffer	1 x	94	30 s	1
dNTPs	0,2 mM	94	30 s	30
MgSO ₄	2 mM	63	10 s	
Primer oNTI200	0,2 µM	68	10 s	
Primer oNTI201	0,2 µM	68	2 min	1
Polymerase	1 U	4	Pause	1
Template cDNA (Abschnitt 2.7.7)	10 µl			
Gesamtvolumen	50 µl			

2.8 Etablierung und Optimierung des Ribosomen-Profilings für humanes Lebergewebe

Nach der erfolgreichen Etablierung des Ribosomen-Profilings für eine humane Zelllinie wurde das oben beschriebene Protokoll mit einigen Modifizierungen für humanes Lebergewebe optimiert. Hierfür wurden 50 mg Gewebe in flüssigem Stickstoff gefroren und mit einer Schwingmühle (Retsch MM301, Retsch GmbH, Haan, Deutschland) und unter Verwendung einer Stahlkugel (Retsch GmbH, Haan, Deutschland) physikalisch zerkleinert (30 s, 30 Schwingungen / s). Das pulverisierte Gewebe wurde anschließend in 1 ml Polysomenpuffer + 100 µg/ml Cycloheximid resuspendiert. Die Polysomen-RNA wurde isoliert und hydrolysiert wie bereits in Abschnitt 2.6 beschrieben.

Des Weiteren erfolgte die Footprint-Bibliotheken Präparation weitgehend wie sie im Abschnitt 2.7. beschrieben wurde. Abwandlungen der Methode werden im Folgenden beschrieben.

2.8.1 Subtraktive Hybridisierung von rRNAs

<u>Lösung A</u>		<u>Waschpuffer</u>	
0,1 M	NaOH	10 mM	Tris-HCl, pH 7,5
50 mM	NaCl	1 mM	EDTA
		2 M	NaCl

Um die rRNA-Kontaminationen in den Footprint-Banken zu verringern, wurde eine subtraktive Hybridisierung von rRNAs (Ingolia et al. 2011) durchgeführt. Hierfür wurde die polyadenylierte Footprint-RNA (Abschnitt 2.7.3) mit 30 µl 2 x SSC, 1 µl 250 mM EDTA und 250 pmol biotinylierten Oligonukleotiden (18S_rRNA_144, 18S_rRNA_101 und 18S_rRNA_284) versetzt. Dieser Reaktionsansatz wurde für 2 min bei 70°C denaturiert, auf 37°C abgekühlt und mit 1 µl SUPERaseIn behandelt. Die biotinylierten Oligonukleotide sind *reverse* komplementär zu Abschnitten der 18S rRNA und hybridisierten mit diesen während einer anschließenden Inkubation von 30 min bei 37°C. Dieser Ansatz wurde nach der Hybridisierung auf 200 µl mit DEPC-Wasser aufgefüllt.

Für die Eliminierung der biotinylierten Oligo-rRNA-Hybride aus den Footprint-RNA-Proben kamen die magnetischen *Beads* MyOne Streptavidin C1 DynaBeads (Life

Technologies, Darmstadt, Deutschland) zum Einsatz. Diese *Beads* präsentieren auf ihrer Oberfläche eine Streptavidinschicht, an der biotinylierte Moleküle binden können. Vor deren Verwendung wurden diese mittels mehrerer Waschschriffe vorbereitet. Pro Probe wurden 1 mg *Beads* in ein 1,5 ml Reaktionsgefäß überführt, welches dann für 2 min an einen Magneten gestellt wurde. Die magnetischen *Beads* wanderten zur Magnet zugewandten Gefäßseite, und der Überstand wurde abgenommen. Anschließend wurden die *Beads* 3 x in 200 µl Waschpuffer, 2 x in Lösung A und 2 x in 0,1 M NaCl resuspendiert. Zwischen jedem Waschschriffe wurde das Reaktionsgefäß wieder für 2 min an den Magneten gestellt und der Überstand wurde verworfen. Im Anschluss wurden die *Beads* in 200 µl Waschpuffer resuspendiert, mit der 200 µl Hybridisierungsreaktion vereinigt und 15 min bei Raumtemperatur leicht geschüttelt. Während dieser Inkubationszeit banden die biotinylierten Oligo-rRNA-Hybride an die *Beads*. Der gesamte Reaktionsansatz wurde wieder an den Magneten gestellt und der Überstand mit den Footprint-RNAs ohne rRNA-Kontaminationen wurde vorsichtig abgenommen und in ein neues Reaktionsgefäß überführt. Darauf folgend wurde die Footprint-RNA gefällt (Abschnitt 2.5.3) in 12 µl DEPC-Wasser gelöst und als Template für die cDNA-Synthese eingesetzt.

2.8.2 Abwandlung der cDNA-Synthese

Für die Footprint-cDNA-Synthese wurden 11,5 µl der rRNA reduzierten Poly(A)-RNA (Abschnitt 2.8.1) in einer 20 µl Reaktion eingesetzt. Hierbei kam die Superscript III Reverse Transkriptase (Life Technologies, Darmstadt, Deutschland) und der Oligo-dT-Primer RTF1 (Anhang 6.1) zum Einsatz.

Nach der Reversen Transkription wurde die Template-RNA durch Zugabe von 1 M NaOH und einer Inkubation von 15 min bei 98°C eliminiert. Durch das Hinzufügen von 1 M HCl wurde der Reaktionsansatz wieder neutralisiert. Die Zusammensetzung des Reaktionsansatzes und die Inkubationszeiten sind der Tabelle 4 zu entnehmen.

Tabelle 4: Zusammensetzung und Inkubationszeiten der Reversen Transkription. Die Konzentrationsangaben beziehen sich auf das Gesamtvolumen von 20 μ l.

Komponente	Konzentration / Volumen	Inkubation
Poly(A)-RNA	11,5 μ l	3 min bei 75°C Denaturierung
dNTPs	0,5 mM	
Primer RTF1	0,25 μ M	
SUPERaseIn	10 U	30 min bei 48°C Reverse Transkription
DTT	5 mM	
5 x Puffer	1 x	
SuperScript III RT	200 U	
Gesamtvolumen	20 μl	
NaOH	50 mM	15 min bei 98°C RNA-Eliminierung
HCl	50 mM	Neutralisation

2.8.3 Eluierung der Footprint-RNA und cDNA aus TBE-Urea-Gelen.

Für die Rückgewinnung von RNA und cDNA aus TBE-Urea-Gelen wurde der entsprechende Gelbereich ausgeschnitten, in flüssigem Stickstoff gefroren und mit einer Schwingmühle (Retsch MM301, Retsch GmbH, Haan, Deutschland) und unter Verwendung einer Stahlkugel (Retsch GmbH, Haan, Deutschland) physikalisch zerkleinert (30 s, 30 Schwingungen / s). Die pulverisierten Gelstücke wurden dann in 400 μ l Wasser resuspendiert und für 10 min bei 70°C in einem Thermomixer (Eppendorf AG, Köln, Deutschland) leicht geschüttelt. Anschließend wurde der gesamte Ansatz auf einen Spin-X Zentrifugenfilter (Corning, NY, USA) aufgetragen und für 10 min bei 10.000 x g zentrifugiert. Die gefilterten Nukleinsäuren wurden dann gefällt (Abschnitt 2.5.3), mit 70%igen (v/v) Ethanol gewaschen, luftgetrocknet und in den entsprechenden Puffern oder Wasser gelöst.

2.8.4 Amplifizierung der RPFs mittels PCR unter Verwendung von Multiplex-Primern

Für die Amplifizierung der RPFs aus Lebergeweben kamen der *forward*-Primer RTF2 sowie einer von sechs verschiedenen *reverse*-Primern (RPI1, RPI2, RPI3, RPI4, RPI5, RPI6) mit unterschiedlichen Barcodes zum Einsatz, was eine anschließende parallele Sequenzierung von drei Footprint-Bibliotheken auf einer einzigen *Lane* ermöglichte. Die verwendeten Komponenten des Reaktionsansatzes sowie das PCR-Programm entsprechen den Angaben in Abschnitt 2.7.8. Die Primersequenzen sind dem Anhang 6.1 zu entnehmen.

2.9 Illumina Next-Generation-Sequencing-System

Die Footprint-Fragmente wurden unter Verwendung der Primer oNTI200 und oNTI201 bzw. RTF2 und RPI1, RPI2, RPI3, RPI4, RPI5 oder RPI6 nach dem oben beschriebenen Protokoll amplifiziert (Abschnitt 2.7.8 und Abschnitt 2.8.4). Dadurch wurden DNA-Fragmente generiert, die an beiden Enden Illumina *Cluster-Generation*-Sequenzen sowie eine Bindestelle für den Sequenzierprimer aufwiesen.

Für die Sequenzierung der Footprint-Bibliotheken wurde das *Next-Generation HiSeq 2000 Sequencing-System* von Illumina (San Diego, USA) verwendet. Die Footprint-Bibliotheken wurden von der Genomanalysegruppe vom Fritz-Lipmann-Institut (Jena, Deutschland) nach dem vom Hersteller angegebenen Protokoll und unter Verwendung des Sequenzierprimers oNTI202 (5'-CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC-3') sequenziert.

2.10 Analyse der sequenzierten Fragmente

Die bioinformatischen Analysen wurden mit freundlicher Unterstützung von Alexander Herrmann (Genomische Gastroenterologie, Universitätsklinikum Schleswig-Holstein, Kiel, Deutschland) bewerkstelligt.

2.10.1 Sequenzanalyse

Die 50 nt langen Nukleotidsequenzen der RPFs wurden als FASTQ-Format gespeichert. Hierbei handelt es sich um ein Standardtextformat zur Speicherung von Sequenzen, die mit Hilfe von Hochdurchsatz-Sequenziermaschinen, wie z.B. dem Illumina *Genome-Analyzer*, generiert werden. Neben den eigentlichen Sequenzen werden unter anderem auch die korrespondierenden Qualitätswerte gespeichert.

Der erste Schritt während der Datenanalyse war die Entfernung des Poly(A)-Signal am 3'-Ende der Sequenzen, welches während der Bankenpräparation an die Footprint-RNA angefügt wurde.

Frühere Studien zeigten, dass der durch Ribosomen geschützte RNA-Bereich ungefähr 28 bis 30 nt lang ist (Ingolia et al. 2009, Guo et al. 2010). Aus diesem Grund wurden Fragmente die kleiner als 20 nt und größer als 40 nt sind bei weiteren Analysen nicht berücksichtigt und aus dem Datensatz entfernt.

Aus vorangegangenen Studien ist weiterhin bekannt, dass die Monosomenprobe eine signifikante Menge von ribosomaler RNA enthält, welche während der Hydrolyse mittels Ribonukleasen auch durch die Ribosomen geschützt wird. Demzufolge wurden die Fragmente als erstes mit den rRNA RefSeq Sequenzen NR_023371, NR_003287, NR_003286, NR_003285 und NR_023363 unter Verwendung der *Short-Read-Alignment-Software* Novoalign V2.07.13 (Novocraft, Selangor, Malaysia) abgeglichen und aus dem Datensatz für weitere Analysen entfernt. Zusätzlich wurden Kontaminationen wie Adapter- oder Primersequenzen entfernt. Die restlichen Fragmente wurden anschließend mit dem humanen Referenzgenom (UCSC GRCh37/hg19) abgeglichen.

Um die genaue Position der tatsächlich translatierten Codonen auf der Referenzsequenz zu kartieren, wurden 12 nt zum Anfangspunkt des kartierten Footprints in 3'-Richtung gezählt, da das translatierte Codon in der P-Stelle in 5'-Richtung von 12 nt flankiert ist (Ingolia et al. 2009). Wenn das erste Nukleotid des Footprints von der Referenzsequenz abwich, wurden 13 nt zugerechnet.

2.10.2 Vorhersage von Translationsinitiationsstellen mittels eines neuronalen Netzes

Die genomweite Identifizierung von neuen Initiationsstellen der Translation wurde mit Hilfe eines künstlichen neuronalen Netzes bewerkstelligt. Künstliche neuronale Netze

werden verwendet, um Probleme aus Bereichen wie z. B. in der Statistik zu lösen und sind nach einem Training anhand von Beispielen in der Lage in einer Menge von ähnlichen aber unbekannt Daten Regelmäßigkeiten, Wiederholungen oder Gesetzmäßigkeiten zu erkennen. Sie bestehen aus mehreren Neuronen, welche auch als *Units* bezeichnet werden und untereinander verbunden sind. Man unterscheidet zwischen *Input-Units*, die Signale wie Reize oder Muster empfangen, *Hidden-Units*, die diese Signale verarbeiten und weiter leiten und *Output-Units*. Die Neuronen einer Ebene werden zu Schichten zusammengefasst. Ein Neuron einer Schicht ist mit allen Neuronen der anderen Schichten verbunden (Abb. 6). Die Stärke der Verbindung zwischen zwei Neuronen wird durch ein Gewicht ausgedrückt. Ein positives Gewicht bedeutet, dass ein Neuron auf ein anderes Neuron einen erregenden Einfluss ausübt. Ein negatives Gewicht wiederum bewirkt eine Inhibition.

Um mit Hilfe von einem neuronalen Netz in einem unbekannt Datensatz Regelmäßigkeiten zu finden, muss dieses zuerst trainiert werden. Das Training erfolgt durch Eingabe von dem zu lernenden Muster (positives Signal) sowie einem Kontrollmuster (negatives Signal). Abhängig von den Eingabemustern ändert sich das neuronale Netz von selbst, indem die Gewichte zwischen den einzelnen Neuronen modifiziert werden. Nach der Trainingsphase erfolgt eine Testphase in der die Gewichte der *Units* nicht mehr verändert werden. Durch die erneute Präsentation der zu lernenden Ausgangsreize mit denen trainiert wurde (Trainingsmenge) kann man feststellen, ob das Netz das Trainingsmaterial erfasst hat. Durch die Präsentation von neuen Reizen (Validierungsmenge) wird geprüft, ob das Netz in der Lage ist, auf neue Reize zu reagieren bzw. die erforderlichen Aufgaben zu lösen (<http://www.neuronalesnetz.de>).

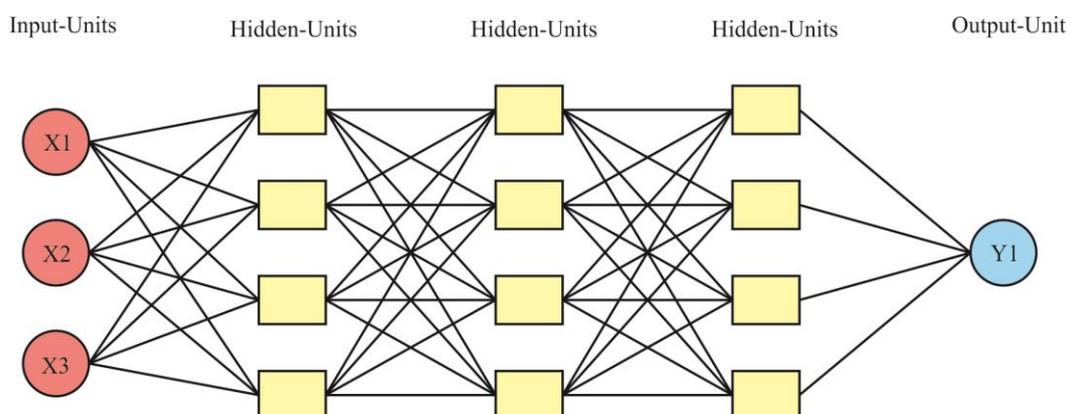


Abb. 6: Möglicher Aufbau eines neuronalen Netzes. Dieses Beispiel eines künstlichen neuronalen Netzes besteht aus drei *Input*- (rot), zwölf *Hidden*- (gelb) und einem *Output*-Neuron (blau). Die übereinander angeordneten Neuronen werden zu Schichten zusammengefasst. Die schwarzen Linien symbolisieren die erregenden oder hemmenden Verbindungen zwischen den einzelnen *Units*.

In der vorliegenden Arbeit sollten mit Hilfe eines künstlichen neuronalen Netzes Translationsinitiationsstellen identifiziert werden. Dafür wurden 10 Netze, aufgebaut aus 5 *Input*-Neuronen, 3 Schichten mit jeweils 11 *Hidden*-Neuronen und einem *Output*-Neuron, trainiert (Neuroph.sourceforge.net, V2.5). Für das Training wurden manuell 604 Transkripte von den mit Puromycin stimulierten Proben ausgewählt, welche eine deutliche Footprint-Abdeckung an den kanonischen Startcodonen aufwiesen. Dabei wurde auf eine variierende Sequenzabdeckung der selektierten Transkripte geachtet, indem 185 Transkripte mit einer 100 bis 8390-fachen, 250 Transkripte mit einer 30 bis 99-fachen und 196 Transkripte mit einer 10 bis 29-fachen Sequenzabdeckung am kanonischen AUG verwendet wurden.

Um darüber hinaus eine ausreichende Datenqualität zu gewährleisten, musste mindestens ein Nukleotid des ausgewählten Analyse-Bereiches eine 20-fache Fragment-Abdeckung aufweisen. Weiterhin wurde die Anzahl der kartierten Footprints pro Nukleotidposition für jedes Transkript auf den höchsten Sequenzpeak normalisiert.

Von den 604 Transkripten wurden 403 (Trainingsmenge) zufällig für das Training ausgewählt. Dabei dienten die kanonischen Startstellen dem Netz als positives Signal (= TIS), wohingegen die umgebenden Positionen -6, -3, +3, +9, +15, +21, +45, +90, +180, +300, +480, relativ zum kanonischen Startcodon, als negative Kontrollen (= keine TIS) fungierten. Die restlichen 201 Transkripte (Validierungsmenge) waren für das Netz unbekannt und konnten zur Bewertung des Lernerfolges bei der Erkennung von TISs verwendet werden. Für jedes Nukleotid ermittelt das Netz ein positives oder negatives Signal.

Für die Wahl eines optimalen Netzes wurde die Qualität der 10 verschiedenen neuronalen Netze mit Hilfe von Grenzwertoptimierungskurven (ROC-Kurven) analysiert. Hierfür musste der Anteil der richtig positiv erkannten Startstellen bezogen auf alle Nukleotidpositionen (Sensitivität des Netzes) gegen den Anteil der falsch positiv erkannten Nukleotidpositionen (Spezifität des Netzes) aufgetragen werden. Je steiler die ROC-Kurve senkrecht ansteigt (Sensitivität = 1, Spezifität = 0), desto besser kann das Netz zwischen den beiden Gruppen (TIS oder keine TIS) unterscheiden. Ein weiteres geeignetes Maß für die Qualitätsbestimmung ist die Fläche unter der Kurve (=ROC-AUC). Für die nachfolgende genomweite Identifizierung von TISs kam das neuronale Netz mit der medianen ROC-Kurve und einer AUC von 0,97 zum Einsatz.

Im Anschluss konnte der gesamte Ribosomen-Footprint-Datensatz mit dem ausgewählten Netz analysiert werden. Hierfür wurde eine Liste von Referenztranskripten mit einem

Transkript pro Gen zusammengestellt. Bei Genen mit mehreren Transkriptvarianten wurde für weitere Analysen nur das Transkript mit dem am weitesten 5' gelegenen kanonischen Start und der längsten 5'-UTR berücksichtigt. Da in dieser Arbeit die Identifizierung von neuen uORFs und N-terminalen Proteinextensionen von Interesse ist, war eine Begrenzung der weiteren Analysen auf die 5'-UTR und die ersten 30 nt der CDS der Referenzsequenzen sinnvoll.

2.10.3 Interpretation der vorhergesagten TISs

Die vom neuronalen Netz identifizierten TISs wurden in vier funktionelle Kategorien geteilt: kanonische TISs, uORFs, ovuORFs und N-terminale Proteinextensionen (Tab. 5). Mögliche AUG oder *near-cognate* Startcodonen wurden in einen ± 3 bp Bereich um die vom neuronalen Netz ermittelten positiven TIS-Signale gesucht. Als erstes wurden die zu den kanonischen Startcodonen gehörigen TIS-Signale klassifiziert. Darauf folgend wurden *upstream* gelegene AUG-Codonen und anschließend *near-cognate* Codonen definiert. Dabei wurde immer mit der TIS im Transkript begonnen, welche die höchste Sequenzabdeckung aufwies. Die genaue Vorgehensweise ist in Abb. 7 schematisiert. Für jedes Transkript wurde das Verfahren so oft wiederholt, bis jede vom Netz vorhergesagte TIS klassifiziert wurde.

Tabelle 5: Kategorien von Translationsinitiationsstellen

Kategorie	Beschreibung
kanonische TISs	vorhergesagte TISs liegt am kanonischen Start
N-terminale Extensionen	vorhergesagte TIS 5' vom kanonischen Start lokalisiert im gleichen Leserahmen wie kanonischer Start kein Stoppcodon zwischen 5'-Start und kanonischen Start
uORFs	vorhergesagte TIS 5' vom kanonischen Start lokalisiert alle Leserahmen möglich Stoppcodon im gleichen Leserahmen wie die 5'-TIS Stoppcodon zwischen 5'-TIS und kanonischen Start
ovuORF	vorhergesagte TIS 5' vom kanonischen Start lokalisiert nicht im gleichen Leserahmen wie kanonischer Start Stoppcodon 3' vom kanonischen Start lokalisiert

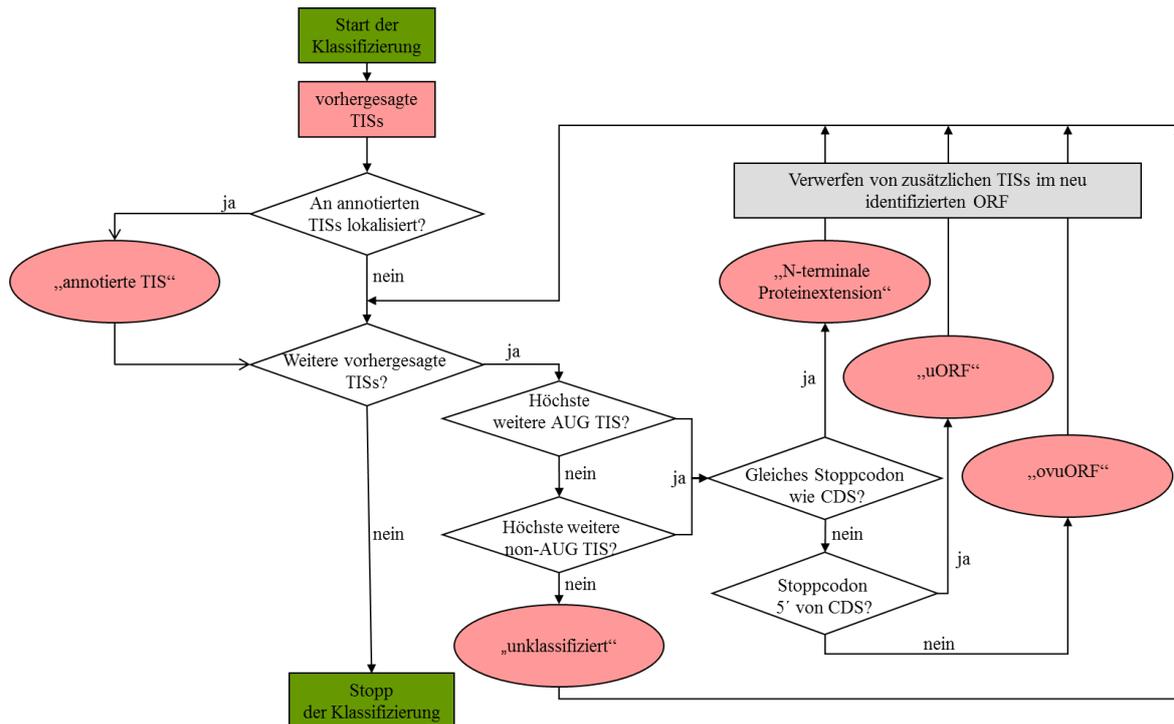


Abb. 7: Funktionelle Klassifizierung der vorhergesagten TISs.

2.10.4 Analyse der Sequenzkonservierung der identifizierten TICs

Um das Ausmaß der Konservierung der TIC-Sequenzen zu beurteilen, wurden die orthologen Positionen von neun nicht humanen Primatenspezies (Schimpanse, Gorilla, Orang-Utan, Rhesusaffe, Pavian, Krallenne, Koboldmaki, Lemur, Galagos) vom PhastCons46Primates Track vom UCSC Genome Browser (Pollard et al. 2010) extrahiert. Der mit Hilfe des PhastCons-Programms errechnete Grad der Konservierung wird von UCSC u.a. als Konservierungswert, der zwischen 0 und 1 liegt, angeboten. Die Konservierungswerte der neu identifizierten möglichen TICs wurden mit den Werten von Kontrollcodonen, aus der 5'-UTR von den analysierten RefSeq Sequenzen, verglichen.

Um statistisch zu überprüfen, ob die Konservierungswerte der verschiedenen TICs innerhalb der einzelnen TIC-Kategorien signifikant verschieden sind, wurde ein T-Test angewendet. Mit Hilfe einer Varianzanalyse wurden die Unterschiede der Konservierungswerte aller TICs über die TIC-Kategorien statistisch geprüft. Alle statistischen Analysen wurden mit dem Programm R Version 2.13.2 2 (www.r-project.org) (R Development Core Team 2011) durchgeführt.

Um die Konservierung der TIC-Verwendung zwischen Mensch und Maus zu analysieren, wurden TICs, die in dieser Arbeit definiert wurden, mit TICs verglichen, die in einer vorangegangenen Studie in embryonalen Mausstammzellen bestimmt wurden (Ingolia et al. 2011). Hierfür wurden multiple Sequenz-*Alignments* von Mensch- und Mausequenzen vom UCSC Genome Browser (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/maf/>) bezogen. Die humanen 5'-UTRs wurden gegen die entsprechenden Positionen im Mausgenom (mm9) kartiert. Anschließend wurde überprüft, ob die in den zwei Studien identifizierten TICs zwischen Mensch und Maus konserviert sind.

3. Ergebnisse

3.1 Etablierung des Ribosomen-Profilings

In dieser Arbeit wurde als erstes die für Hefe etablierte Ribosomen-Profilings-Methode (Ingolia et al. 2009) mit einigen Abwandlungen erfolgreich für die humane Zelllinie THP-1 adaptiert.

3.1.1 Hydrolyse von THP-1 Polysomen mit verschiedenen Ribonukleasen

Um eine optimale Hydrolyse von mRNA-Bereichen, welche nicht durch Ribosomen geschützt sind, zu gewährleisten, wurden THP-1 Polysomensuspensionen mit unterschiedlichen Konzentrationen von Ribonukleasen verschiedener Hersteller unter variablen Inkubationsbedingungen getestet. Da die Enzymaktivitätsangaben von den Herstellern unterschiedlich definiert sind, mussten die Unit-Angaben der verschiedenen Ribonukleasen aufeinander abgeglichen werden, um vergleichende Hydrolyseansätze durchzuführen. Ein Ansatz, der mit einem RNase Inhibitor versetzt wurde, diente als Kontrolle, bei dem die mRNA intakt blieb. Die Effizienz der Polysomenhydrolyse wurde mit Hilfe von Polysomenprofilen beurteilt (Abb. 8).

Die bestmögliche Hydrolyse der Polysomen wurde unter Verwendung von 2000 U RNaseI von Life Technologies (Abb. 8C) und 200 U RNaseI von Epicentre (Abb. 8D) während einer Inkubation von 60 min bei 30°C erzielt. Jedoch ist bei beiden Profilen noch ein kleiner Disomen-*Peak* erkennbar. Da aber nur die Gradientenfraktionen, welche die Monosomen enthielten, für die weiteren Experimente verwendet wurden, ist letztendlich die Hydrolyse mit diesen Ribonukleasen geeignet und die Effizienz der Hydrolyse ausreichend. In allen nachfolgenden Experimenten wurden die Polysomen für 60 min bei 30°C mit 2000 U der RNaseI von Life Technologies hydrolysiert.

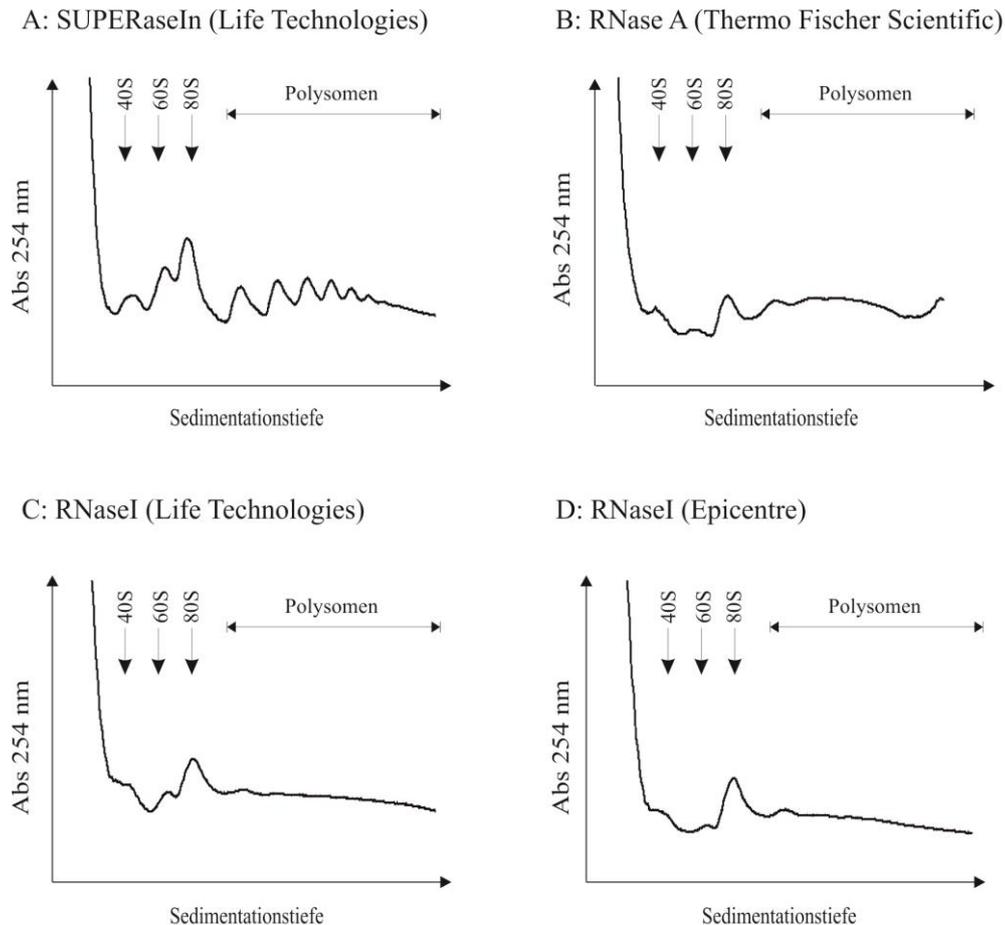


Abb. 8: THP-1 Polysomenprofile von verschiedenen Hydrolyseansätzen. Die Polysomen wurden unter Verwendung von Ribonukleasen verschiedener Hersteller für 1 h bei 30 °C hydrolysiert. **A:** Eine Kontrollprobe wurde mit 80 U SUPERaseIn behandelt. **B:** 1000 U RNaseA (Thermo Fischer Scientific), **C:** 2000 U RNaseI (Life Technologies), **D:** 200 U RNaseI (Epicentre). Die Unit-Angaben entsprechen den Angaben des Herstellers.

3.1.2 Etablierung der Footprint-Bibliotheken Präparation

Nach der Hydrolyse der Polysomen-RNA und Fraktionierung wurden kleine RNA-Moleküle der Monosomenfraktionen angereichert, aufgereinigt, dephosphoryliert und anschließend auf einem 15%igen Gel aufgetrennt (Abb. 9A). Neben den 30 nt großen RPFs enthielten die Proben unter anderem auch rRNA, die ebenfalls von den ribosomalen Untereinheiten vor der Hydrolyse geschützt war (Ingolia et al. 2012). Aus diesem Grund wurde nur der Gelbereich, welcher die RPFs enthielt, ausgeschnitten. Nach der Eluierung der RPFs aus dem Gel wurden diese als Voraussetzung für eine cDNA-Synthese mit Oligo-dT-Primern am 3'-Ende polyadenyliert. Die gesamte cDNA-Synthesereaktion wurde auf einem 10%igen Gel aufgetrennt (Abb. 9B), und der Gelbereich mit der cDNA-Bande

wurde ausgeschnitten. Die cDNA wurde aus dem Gel eluiert und zirkularisiert. Durch eine nachfolgende Relinearisierung mit Hilfe der Endonuklease APE1 entstanden wieder lineare DNA-Moleküle, bei denen sich nun die Footprint-Fragmente in der Mitte befanden und in 5' und 3'-Richtung von Nukleotiden flankiert waren, die als Primerbindestellen für die nachfolgende Amplifizierung dienten. Die gesamte Footprint-Bibliothek wurde auf einem 10%igen Gel aufgetrennt (Abb. 9C) und die 121 nt große PCR-Produkt-Bande wurde ausgeschnitten, aufgereinigt und sequenziert.

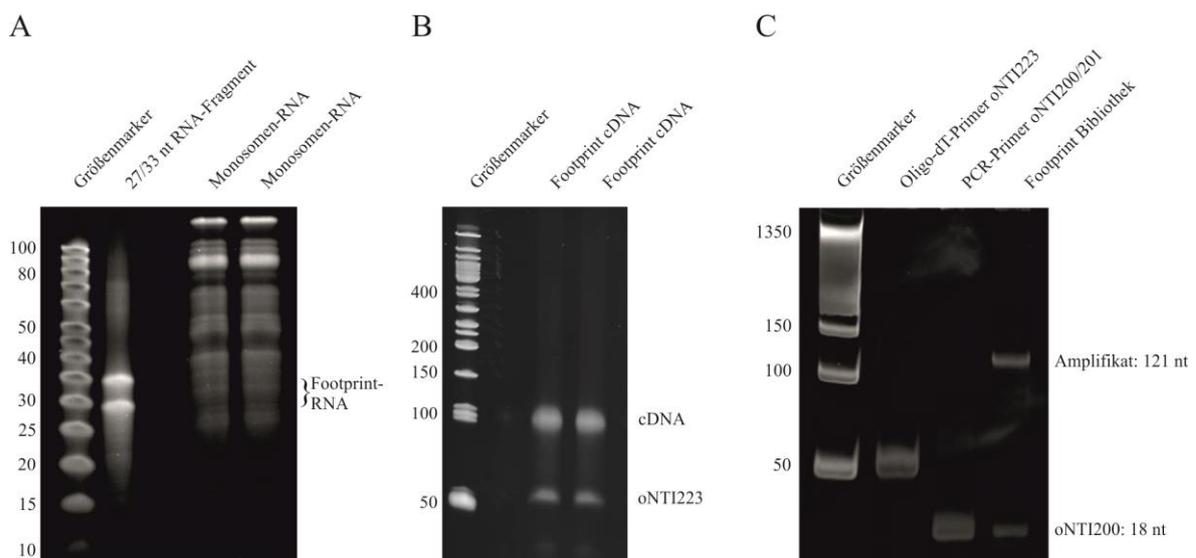


Abb. 9: Denaturierende TBE-Harnstoffgele bei der Footprint-Bibliotheken Präparation. **A:** RNA der Monosomenfraktionen von THP-1 Zellen. Die Klammer rechts kennzeichnet den Gelbereich mit der ca. 30 nt langen Footprint-RNA. **B:** cDNA **C:** Footprint-Bibliothek.

3.1.3 Analyse der sequenzierten Footprint-Fragmente

Nach der erfolgreichen Etablierung des Ribosomen-Profiling wurden drei THP-1 Footprint-Bibliotheken nach dem oben beschriebenen Standardprotokoll generiert und sequenziert. Daraus ergaben sich für die erste Bibliothek 74.797.236 für die zweite 62.993.913 und für die dritte 72.729.964 Sequenzen. Insgesamt konnten 210.518.113 Sequenzen erhalten werden (Tab. 6). Im Anschluss wurden die sequenzierten Fragmente mit den rRNA RefSeq Sequenzen NR_023371, NR_003287, NR_003286, NR_003285 und NR_023363 abgeglichen. Eine signifikante Anzahl von 145.343.373 Fragmenten (69%) repräsentieren nichtkodierende RNAs, die für spätere Analysen aus dem Datensatz entfernt wurden.

Nachdem aus früheren Studien bekannt ist, dass RPFs zwischen 28 und 30 nt lang sind (Ingolia et al. 2009, Guo et al. 2010), war eine Größenselektion der verbleibenden Fragmente notwendig. Sequenzen, die kleiner als 20 nt und größer als 40 nt lang waren, wurden aus dem Datensatz entfernt. Des Weiteren wurden Sequenzen, die Adapter-, oder Primersequenzen repräsentieren nicht berücksichtigt. Die verbleibenden 42.636.480 RPFs (20%) wurden dann mit dem humanen Referenzgenom (UCSC GRCh37/hg19) abgeglichen. Von diesen Sequenzen wurden 34.994.949 auf das humane Genom kartiert, wobei 17.557.659 eindeutig einer Position zuzuordnen waren.

Tabelle 6: Analyse der sequenzierten Fragmente

Sequenzen	Bank 1	Bank 2	Bank 3	Summe 1-3
gesamt	74.794.236	62.993.913	72.729.964	210.518.113
rRNA	51.763.831	45.483.797	48.095.745	145.343.373
Adapter / Primer, Sequenzen <20 <40	4.970.293	3.886.011	13.681.956	22.538.260
auf humanes Genom kartiert	16.311.507	10.699.670	7.983.772	34.994.949
eindeutig kartiert	6.837.143	6.048.993	4.671.523	17.557.659

3.1.4 Sequenzabdeckung der Transkripte

Insgesamt wurden 18.696 Transkripte analysiert, welche die definierten Analyse Kriterien erfüllten, und 7.243.072 RPFs konnten diesen Transkripten zugeordnet werden. Von diesen 7.243.072 RPFs wurden 6.548.506 (90%) in den proteinkodierenden Sequenzen und 627.553 (9%) in den 5'-UTRs kartiert. Ein repräsentatives Beispiel für ein Transkript mit einer deutlichen Sequenzabdeckung in der 5'-UTR ist *TUBB* (Abb. 10A), wohingegen z.B. bei *ACTB* RPFs nur in der CDS zu detektieren waren (Abb. 10B).

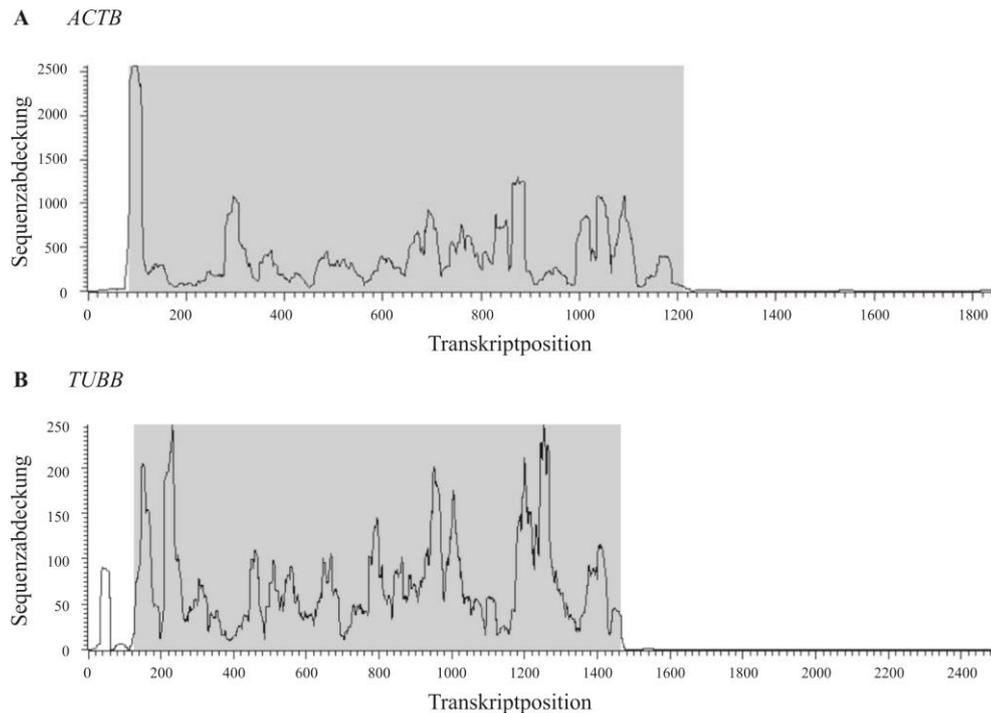


Abb. 10: Sequenzabdeckung von *ACTB* und *TUBB*. Die proteinkodierende Sequenz ist grau hinterlegt. Die weißen Bereiche repräsentieren die 5′- und 3′-UTRs der Transkripte.

Kartierte RPFs in der 5′-UTR spiegeln Translationsereignisse wider, die *upstream* von der eigentlichen annotierten CDS stattfinden. Es stellte sich die Frage, wo genau die Translation in der 5′-UTR initiiert. Aus diesem Grund wurden die 5′-UTRs von Transkripten, welche dort eine deutliche Sequenzabdeckung aufwiesen, nach möglichen alternativen Startcodonen (AUG, GUG, CUG, UUG, AAG, AGG, AUA, AUC, AUU und ACG) abgesucht.

Ein repräsentatives Beispiel hierfür ist die 5′-UTR von *LDHB*, in der über mehreren potentiellen Startcodonen eine deutliche Anreicherung von Ribosomen festgestellt wurde (Abbildung 11A). Eine Initiation an einen *upstream* gelegenen UUG Codon an Position 41 führt zur Bildung eines 23 AS langen Peptids. Demgegenüber führt die Initiation an einem ACG an Position 54, oder CUG-Codon an Position 90, welche im gleichen Leserahmen wie das kanonische Startcodon liegen, zur Expression von N-terminal verlängerten Proteinvarianten. Insgesamt finden sich in der 5′-UTR von *LDHB* neun mögliche alternative Startcodonen (Abb. 11B), von denen drei eine deutliche Anreicherung von Ribosomen aufweisen. Aufgrund der zahlreichen Initiationsmöglichkeiten war eine eindeutige Identifizierung der *upstream* gelegenen Startstellen in den Transkripten nicht möglich. Dies gab den Impuls die Ribosomen an den Translationsinitiationsstellen anzureichern, um diese genau nachweisen zu können.

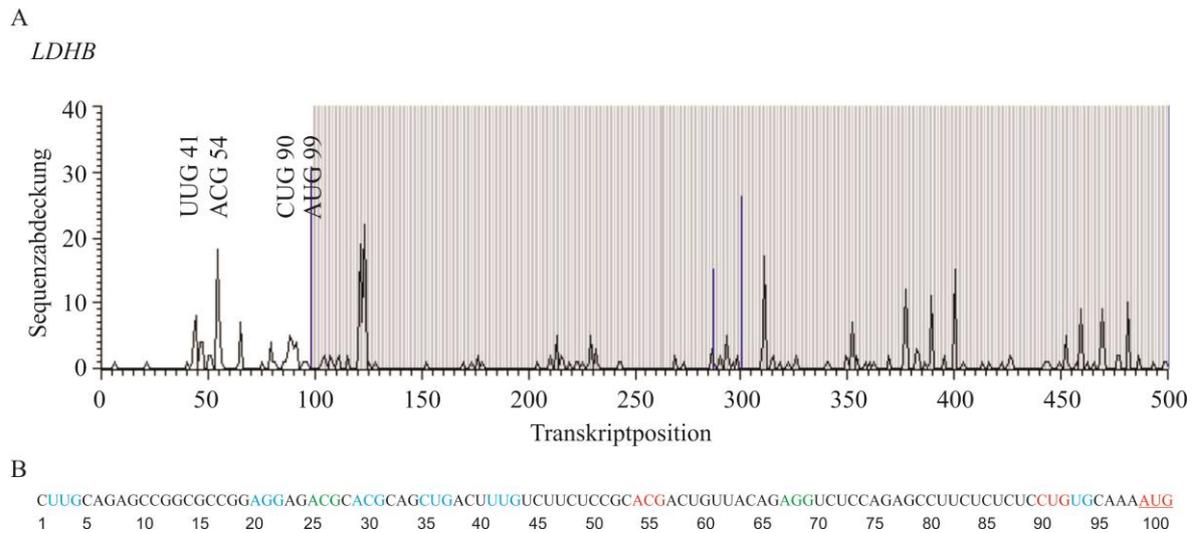


Abb. 11: Sequenzabdeckung von *LDHB*. **A:** Die Anzahl der kartierten RPFs wurde gegen die Transkriptposition aufgetragen. Die proteinkodierende Sequenz ist grau hinterlegt. Der weiße Bereich repräsentiert die 5'-UTR. Die Codonen über den Sequenzpeaks symbolisieren mögliche alternative Startcodonen an ihren entsprechenden Positionen. **B:** Sequenz der 5'-UTR von *LDHB*. Die farbigen Codonen markieren potentielle Startcodonen in den drei möglichen Leserahmen. Das AUG-Startcodon der kanonischen CDS ist unterstrichen.

3.2 Identifizierung und Klassifizierung von TISs

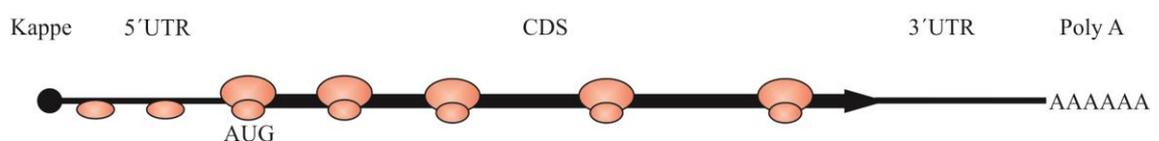
3.2.1 Anreicherung von RPFs an den TISs mit Hilfe von Puromycin

Um Ribosomen an den TISs anzureichern, wurden die Zellen zuerst mit dem Elongationsinhibitor Puromycin und nachfolgend mit Cycloheximid behandelt. Puromycin führte zur Dissoziation elongierender Ribosomen, wohingegen Cycloheximid neu initiiierende Ribosomen arretierte, welche somit an den TISs angereichert wurden (Abb. 12).

Die Wirksamkeit der Inhibitoren wurde mit Hilfe von Polysomenprofilen beurteilt. Im Vergleich zu den Kontrollen wurde bei den mit Puromycin versetzten Proben eine Verringerung der polysomalen Fraktionen festgestellt (Abb. 13). Diese Verringerung, verursacht durch die Dissoziation von elongierenden Ribosomen, bestätigt indirekt die Anreicherung von Ribosomen an den TISs.

Im Folgenden wurden im Triplikrat THP-1 Footprint-Bibliotheken nach dem oben beschriebenen Protokoll hergestellt, bei denen die Zellen zuvor mit Puromycin und Cycloheximid oder nur mit Cycloheximid versetzt wurden.

A: Cycloheximid



B: Puromycin und Cycloheximid

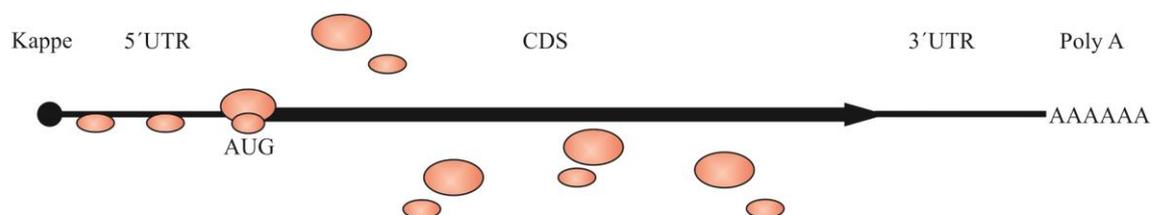


Abb. 12: Schematische Darstellung zur Theorie des Versuchs. Die farbigen Kreise symbolisieren die ribosomalen Untereinheiten. In schwarz ist ein Transkript mit 5'-Kappe, 5'-UTR, CDS, 3'-UTR und Poly(A)-Schwanz dargestellt. **A:** Cycloheximid arretiert elongierende Ribosomen auf der mRNA. **B:** Puromycin führt zur Dissoziation elongierender Ribosomen. Eine anschließende Zugabe von Cycloheximid führt zur Arretierung und Anreicherung von neu initiiierenden Ribosomen an den TISs.

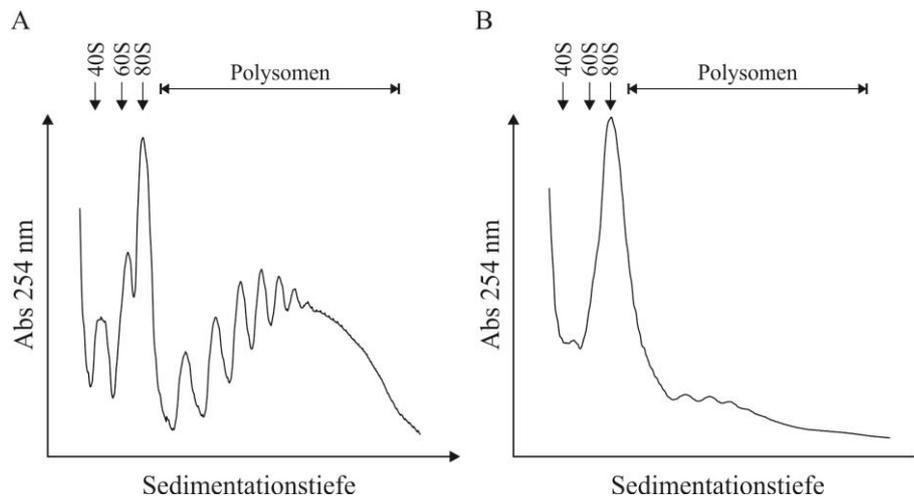


Abb. 13: Polysomenprofile von Kontroll- und mit Puromycin behandelten Proben. **A:** Die Zellen wurden für 2 min mit Cycloheximid behandelt. **B:** Die Zellen wurden für 10 min mit Puromycin und anschließend für 2 min mit Cycloheximid behandelt.

3.2.2 Analyse der RPFs

Insgesamt wurden für die Kontrollproben 296.350.271 und für die mit Puromycin behandelten Proben 283.288.579 Sequenzen erhalten (Anhang 6.2), von denen abermals ein Großteil ribosomale RNAs repräsentierten. Diese wurden genau wie Adapter- und Primersequenzen sowie Sequenzen, die kleiner als 20 nt und größer als 40 nt waren, aus den Analysedaten eliminiert. Die verbleibenden 74.961.420 RPFs der Kontrollproben und 34.698.371 RPFs der mit Puromycin behandelten Proben wurden dann gegen das humane Referenzgenom (UCSC GRCh37/hg19) kartiert, von denen 5-18 Millionen pro Bibliothek eindeutig einer Referenzsequenz zugeordnet werden konnten (Anhang 6.2). Eine Untersuchung der Längenverteilung zeigte, dass die meisten RPFs ca. 30 nt lang sind (Abb. 14).

Statistische Analysen sowie eine visuelle Einschätzung belegten, dass die Verteilung der RPFs über die Transkripte unter den Triplikaten hoch reproduzierbar ist (Abb. 15, Anhang 6.3). Der paarweise Korrelationskoeffizient der Sequenzabdeckung pro Nukleotid liegt zwischen 0,82 und 0,84 für die Puromycin-Proben und bei 0,98 für die Kontrollproben.

Um für weitere Analysen eine verstärkte Sequenzabdeckung der Transkripte zu erhalten, wurden die RPFs der Triplikate zusammengefasst.

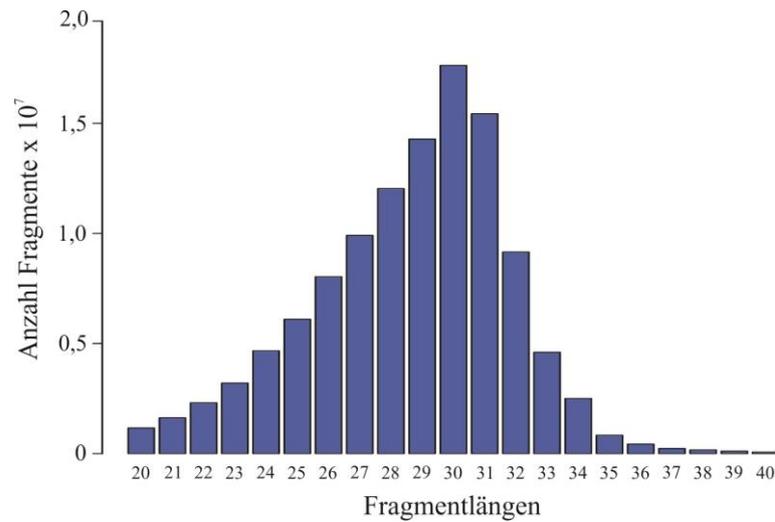


Abb. 14: Längenverteilung der sequenzierten Fragmente. Berücksichtigt wurden nur die 109.659.791 Sequenzen, die tatsächlich mit dem humanen Referenzgenom (UCSC GRCh37 /hg19) abgeglichen wurden.

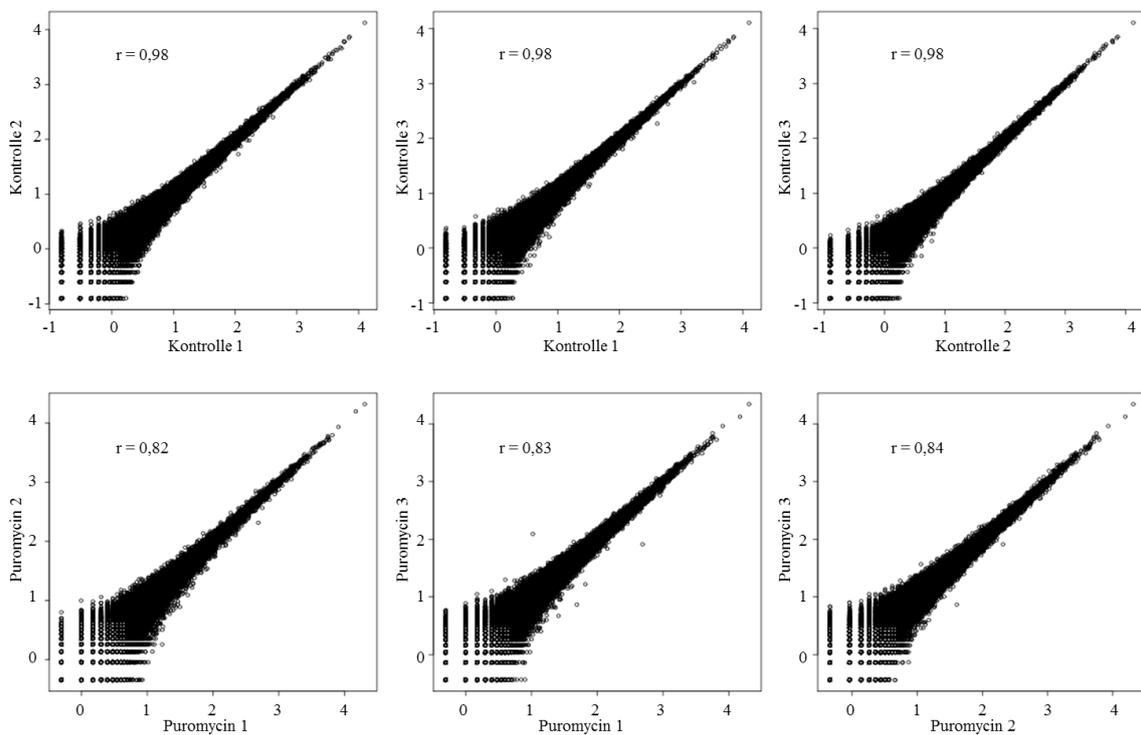


Abb. 15: Reproduzierbarkeit der Verteilung der RPFs über die Transkripte. Der paarweise Korrelationskoeffizient der Sequenzabdeckung pro Nukleotid liegt zwischen 0,82 und 0,84 für die Puromycin-Proben und bei 0,98 für die Kontrollproben.

3.2.3 Anreicherung von RPFs an TISs

Die erfolgreiche Anreicherung von RPFs an den TISs wurde überprüft, indem die Sequenzabdeckung an den annotierten TISs der untersuchten RefSeq Sequenzen analysiert wurde. Anhand des Transkriptbeispiels *TPPI* (Abb. 16) ist deutlich zu erkennen, dass die Puromycinbehandlung zu einer Anreicherung von Sequenzen an den TISs führt und die Sequenzabdeckung der CDS deutlich von 5' nach 3' abnimmt. Im Vergleich dazu sind bei der Kontrolle die RPFs über die gesamten CDS verteilt. Bei weiteren Analysen wurde die Sequenzverteilung der 500 höchst translatierten Gene zusammengefasst. Auch hier ist die Anreicherung der Sequenzen an den TISs deutlich erkennbar (Abb. 17A). Darüber hinaus wurde eine 3 bp Periodizität der Sequenzabdeckung am 5'-Ende der CDS von den analysierten Transkripten festgestellt (Abb. 17B). Diese Periodizität beruht auf der Translokation von Ribosomen auf der mRNA um genau 3 nt während jedes Elongationszyklus.

Die Beobachtung, dass RPFs nicht nur direkt am annotierten Startcodon, sondern auch *downstream* von den TISs kartiert sind, ist durch die Wirkungsweise von Puromycin zu erklären, welches mit den Aminoacyl-tRNAs um die gleiche Bindungsstelle im Ribosom konkurriert. Aus diesem Grund translatierten einige Ribosomen die mRNA weiter und wurden nach wenigen Elongationsrunden durch die Zugabe von Cycloheximid auf den Transkripten *downstream* von den TISs fixiert.

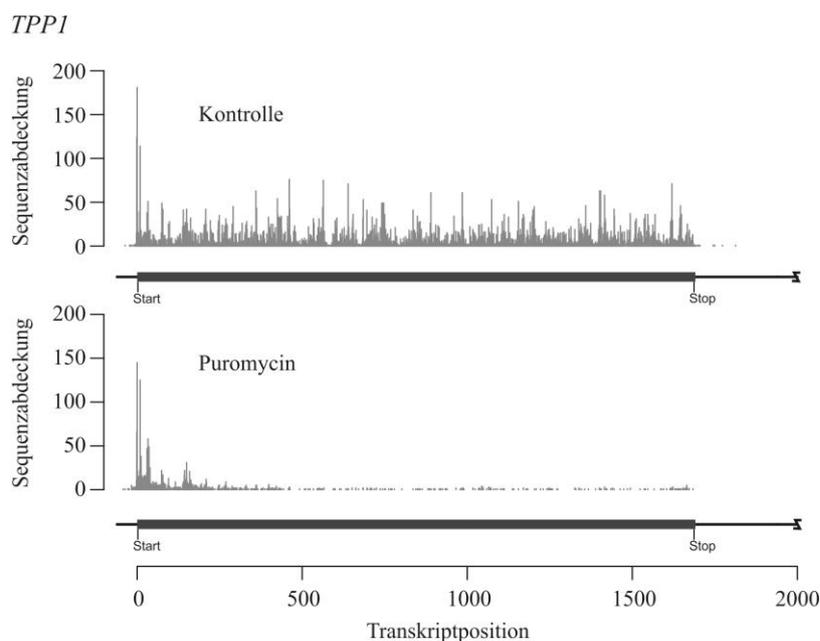


Abb. 16: Sequenzabdeckung von *TPPI* in Kontrollzellen und mit Puromycin behandelten Zellen. Der schwarze Balken symbolisiert den Aufbau des Transkripts mit CDS (dick), 5'-UTR und 3'-UTR (schmal).

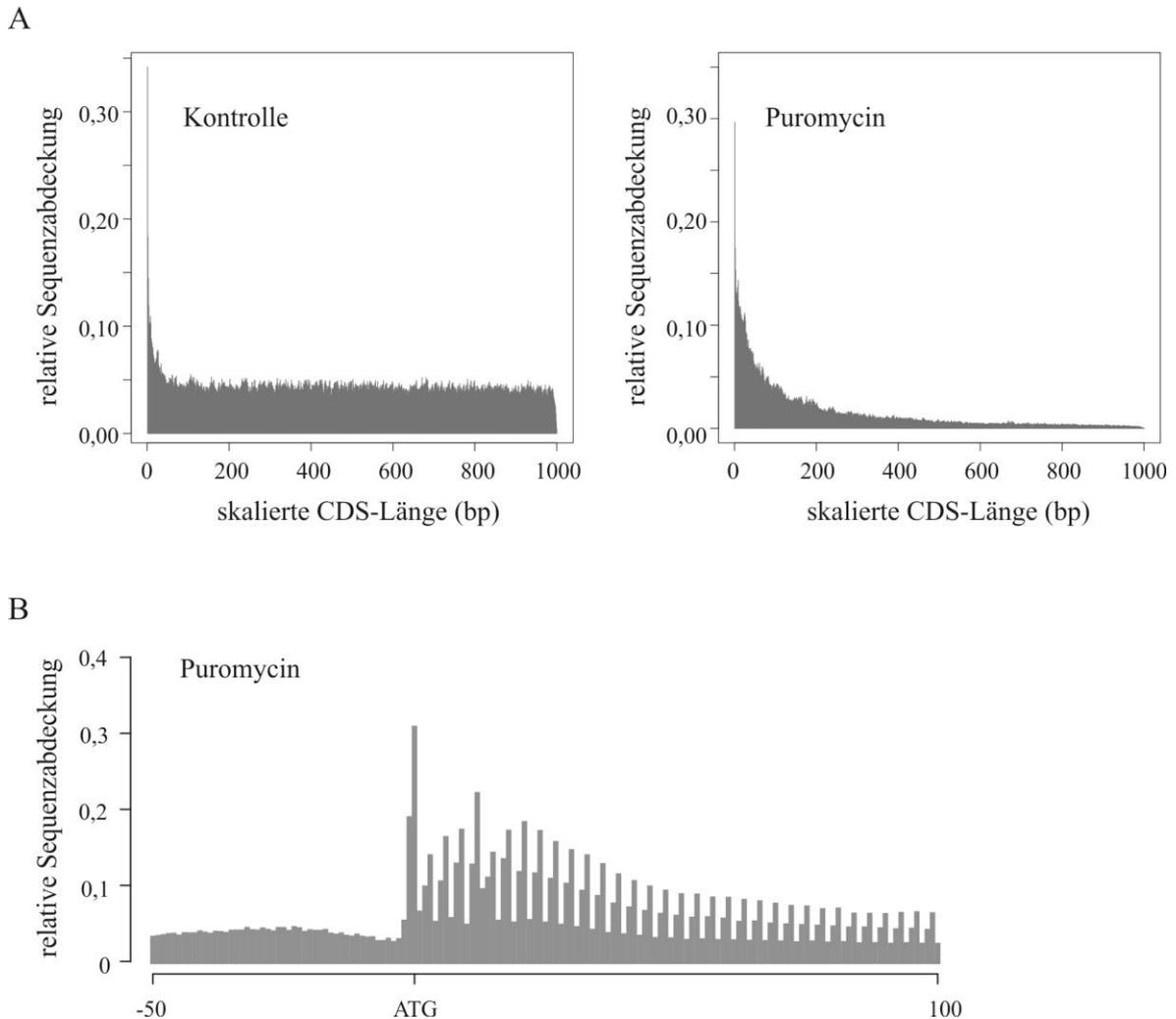


Abb. 17: Anreicherung von RPFs durch Puromycinbehandlung. **A:** Die Sequenzabdeckung der 500 höchst translatierten Gene wurde zusammengefasst. Die transkriptspezifische Sequenzabdeckung wurde auf die absolute Anzahl der kartierten RPFs des entsprechenden Gens normalisiert. Des Weiteren wurde die CDS-Länge der RefSeq Sequenzen auf 1000 bp skaliert. **B:** Sequenzabdeckung der 500 höchst translatierten Gene um das kanonische AUG-Startcodon von Puromycin behandelten Zellen.

3.2.4 Vorhersage von TISs mit Hilfe eines neuronalen Netzes

Für eine systematische Vorhersage von TISs wurde ein künstliches neuronales Netz verwendet. Diesbezüglich wurden zunächst 10 neuronale Netze trainiert. Hierfür wurden manuell 604 Transkripte ausgewählt von denen 403 zufällig gezogene für das Training verwendet wurden. Die restlichen 201 Transkripte (Validierungsmenge) wurden dem Netz nach dem Training präsentiert und dienten der Überprüfung, ob das Netz in der Lage ist, TISs korrekt zu identifizieren. Für jede getestete Nukleotidposition wurde dann vom Netz ein positives (TIS) oder negatives Signal (keine TIS) ausgegeben. Die Qualität der

getesteten Netze wurde dann mit Hilfe von Grenzwertoptimierungskurven bewertet (ROC-Kurven). Für diese Arbeit wurde das neuronale Netz mit der medianen ROC-Kurve und einer ROC-AUC von 0,97 gewählt (Abb. 18). Mit diesem Netz konnten in den 201 Transkripten der Validierungsmenge 93% der TISs korrekt identifiziert werden (Sensitivität = 93%), wohingegen von den Kontrollpositionen nur 1,4% falsch positiv bewertet wurden (Spezifität = 98,6%).

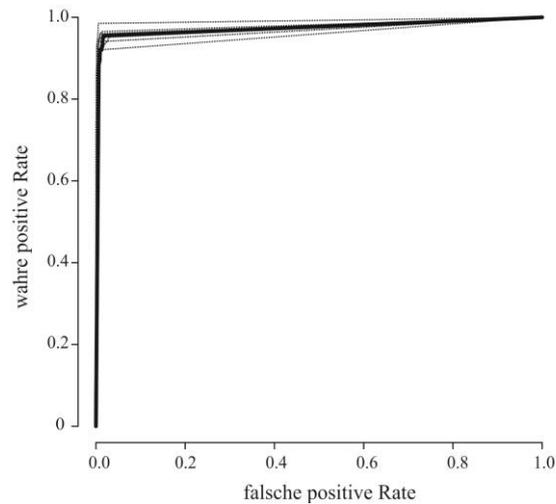


Abb. 18: ROC-Kurven der 10 neuronalen Netze: Mit Hilfe der Grenzwertoptimierungskurven wurde die Qualität der 10 neuronalen Netze bewertet. Der Anteil der richtig negativ erkannten und der Anteil der richtig positiv erkannten Startstellen bezogen auf alle Nukleotidpositionen wurden gegeneinander aufgetragen. Das neuronale Netz mit der medianen Grenzwertoptimierungskurve (fett) und einer ROC-AUC von 0,97 wurde für weitere Analysen verwendet.

Das ausgewählte neuronale Netz wurde für den kompletten Ribosomen-Footprint-Datensatz, welcher aus Puromycin behandelten Zellen generiert wurde, angewandt. Hierbei wurde die TIS-Identifizierung nur auf die 5'-UTRs und die ersten 30 nt der annotierten CDS der RefSeq Sequenzen begrenzt. Darüber hinaus wurden nur Transkripte analysiert, die mindestens eine Nukleotidposition mit einer zwanzigfachen Sequenzabdeckung in dieser definierten Region aufwiesen. Es ergaben sich 5062 Transkripte, die diesen Kriterien entsprachen und analysiert wurden. Insgesamt wurden in diesen 5062 Transkripten 14.464 positive TIS-Signale detektiert (Tab. 7). Um redundante positive Signale zu verringern, wurden Signale von benachbarten Nukleotiden zu einem Signal zusammengefasst. Da in dieser Arbeit nur Translationsinitiationsereignisse von Interesse sind, die stromaufwärts von der annotierten CDS stattfinden, wurden von den verbleibenden 10.386 zusammengefassten TIS-Signalen nur noch die stromaufwärts vom

annotierten TIC gelegenen weiter berücksichtigt. Daraus ergeben sich letztendlich 8710 zusammengefasste positive TIS-Signale, die weiter analysiert wurden.

Von den 5062 analysierten RefSeq Sequenzen konnten für 4364 Transkripte mindestens eine TIS nachgewiesen werden (Abb. 19). Bei diesen Transkripten wurden für 2399 eine, für 1200 zwei und für 510 drei TISs identifiziert. Zwischen vier und elf TISs konnten bei 255 Transkripten detektiert werden. Für 698 Transkripte wurden keine TISs in der analysierten Zielregion detektiert. Eine mögliche stromabwärts vom annotierten Start initiiierende Translation, die in dieser Arbeit nicht berücksichtigt wurde, könnte hierfür ausschlaggebend sein.

Tabelle 7: Zusammenfassung der vom neuronalen Netz ermittelten positiven TIS-Signale

analysierte Transkripte	5062
getestete Nukleotidpositionen	1.304.713
positive TIS-Signale	14.464
zusammengefasste positive TIS-Signale von benachbarten Nukleotiden	10.386
zusammengefasste positive TIS-Signale in der Zielregion (5'-UTR)	8710
klassifizierte TISs	7251
Transkripte mit mindestens einer vorhergesagten TIS in der Zielregion	4364
Transkripte ohne positives TIS-Signal in der Zielregion	698

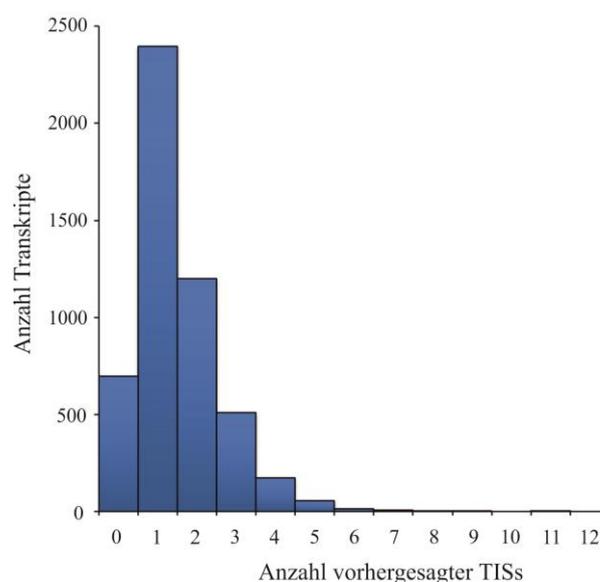


Abb. 19: Verteilung der TISs pro Transkript. Es wurden nur die 8710 zusammengefassten TIS-Signale berücksichtigt, die stromaufwärts von der annotierten CDS der 5062 RefSeq-Sequenzen detektiert wurden.

3.2.5 Funktionelle Klassifizierung der vorhergesagten TISs

Für die 4364 RefSeq Sequenzen, bei denen mindestens eine TIS in der definierten Zielregion identifiziert wurde, konnten insgesamt 8710 TIS-Signale detektiert werden. Diese wurden einer von vier funktionellen Kategorien (annotierte TIS, uORF, ovuORF, N-terminale Proteinextension) zugeordnet (Tab. 8). Hierfür wurde analysiert, ob ein AUG oder *near-cognate* Startcodon in der Nähe (± 3 bp) der vorhergesagten TISs lokalisiert ist. Anhand der Sequenzpositionen der entsprechenden Startcodonen konnten die TISs klassifiziert werden.

Bei 220 TIS-Signalen (2,5%) waren keine AUG oder *near-cognate* Codonen 3 bp vor oder 3 bp nach der identifizierten TIS-Transkriptposition lokalisiert. Da diesen TISs keine Startcodonen zugewiesen werden konnten, wurden diese nicht weiter analysiert. 2305 TISs repräsentierten annotierte Startstellen. 2994 TIS-Signale wurden stromaufwärts von den kanonischen Starts detektiert. Da die entsprechenden TICs im gleichen Leserahmen mit stromaufwärts gelegenen Stoppcodonen lokalisiert waren, konnten diese TISs als uORFs identifiziert werden. Des Weiteren spiegeln 1406 TISs ovuORFs wider, da diese stromaufwärts vom kanonischen Start in einem anderen Leserahmen als dieser lokalisiert waren und ihre entsprechenden Stoppcodonen in der kanonischen CDS festgestellt wurden. 546 TISs führen zur Bildung von N-terminal verlängerten Proteinen, da diese im gleichen Leserahmen wie die kanonischen Startstellen liegen. 1239 TISs, die innerhalb der neu annotierten uORFs oder N-terminalen Proteinextensionen liegen, wurden als Ergebnis einer unvollständigen Inhibition der Elongation durch Puromycin gedeutet und somit nicht weiter klassifiziert. Zusammenfassend wurden in 4364 Transkripten mit 8710 vorhergesagten TISs 7251 klassifiziert und neu annotiert (Tab. 8 und Abb. 20A).

Tabelle 8: Funktionelle Klassifizierung der identifizierten TISs

TISs	Anzahl	in %
vorhergesagt	8710	100
klassifiziert	7251	83,3
annotierte TISs	2305	26,5
uORFs	2994	34,4
ovuORFs	1406	16,1
N-terminale Proteinextensionen	546	6,3
nicht klassifiziert	220	2,5
in neuen ORF lokalisiert	1239	14,2

Tabelle 9: Verteilung der klassifizierten TISs über die Transkripte. aTIS = annotierte TIS, uORF = *Upstream-open-Reading-Frame*, ovuORF = CDS überlappender uORF, N-term = N-terminale Proteinextension, *near-cognate* = AUG, GUG, CUG, UUG, AAG, AGG, AUA, AUC, AUU, ACG.

TIS Kategorie	Transkripte		aTIS		uORF		ovuORF		N-terminale Extension	
	Anzahl	in %	AUG	<i>near-cognate</i>	AUG	<i>near-cognate</i>	AUG	<i>near-cognate</i>	AUG	<i>near-cognate</i>
aTIS	1320	30,2	1317	3	0	0	0	0	0	0
uORF	976	22,4	0	0	601	1014	0	0	0	0
ovuORF	454	10,4	0	0	0	0	56	471	0	0
N-term	152	3,5	0	0	0	0	0	0	4	159
aTIS, uORF	383	8,8	383	0	177	415	0	0	0	0
aTIS, ovuORF	305	7,0	303	2	0	0	13	337	0	0
aTIS, N-term	106	2,4	106	0	0	0	0	0	1	114
aTIS, uORF, ovuORF	116	2,7	115	1	24	140	4	123	0	0
aTIS, uORF, N-term	41	0,9	41	0	14	40	0	0	0	44
aTIS, ovuORF, N-term	30	0,7	30	0	0	0	0	34	0	32
aTIS, uORF, ovuORF, N-term	4	0,1	4	0	0	6	0	4	0	4
uORF, ovuORF	264	6,0	0	0	78	288	42	254	0	0
uORF, N-term	108	2,5	0	0	27	142	0	0	1	117
ovuORF, N-term	45	1,0	0	0	0	0	0	49	0	51
uORF, ovuORF, N-term	18	0,4	0	0	3	25	1	18	0	19
kein AUG o. <i>near-cognate</i> Codon	42	1,0	-	-	-	-	-	-	-	-
Summe	4364	100	2299	6	924	2070	116	1290	6	540

58

3.2.6 Häufigkeit und Verteilung der TICs

Insgesamt konnten 7251 TISs klassifiziert werden. Es wurde als nächstes analysiert, welche TICs als Startstellen für die Translation verwendet werden (Abb. 21A, Anhang 6.4). 6363 (85%) TISs nutzen den Konsensus NUG, wobei AUG-Codonen, wie erwartet, am häufigsten vertreten sind (47%). Zusätzlich initiiert die Translation sehr oft an CUGs und GUGs. Demgegenüber macht die Initiation an AAG- und AUA-Codonen weniger als 1% aus. Eine Analyse der TICs der einzelnen Kategorien zeigte, dass AUG ein häufiges Startcodon für uORFs ist (924, 30%). Im Vergleich dazu nutzen CDS überlappende uORFs (8%) und N-terminale Proteinextensionen (1%) nur selten AUG-Codonen als Initiationsstart (Abb. 21A, oben). Betrachtet man nur die *near-cognate* Codonen ohne AUG kann man eine sehr ähnliche TIC-Verteilung unter allen Kategorien feststellen. Sowohl die Translation von uORFs, CDS überlappenden uORFs und N-terminalen Proteinextensionen initiieren hauptsächlich an CUG-, GUG- und UUG-Codonen (Abb. 21A, unten).

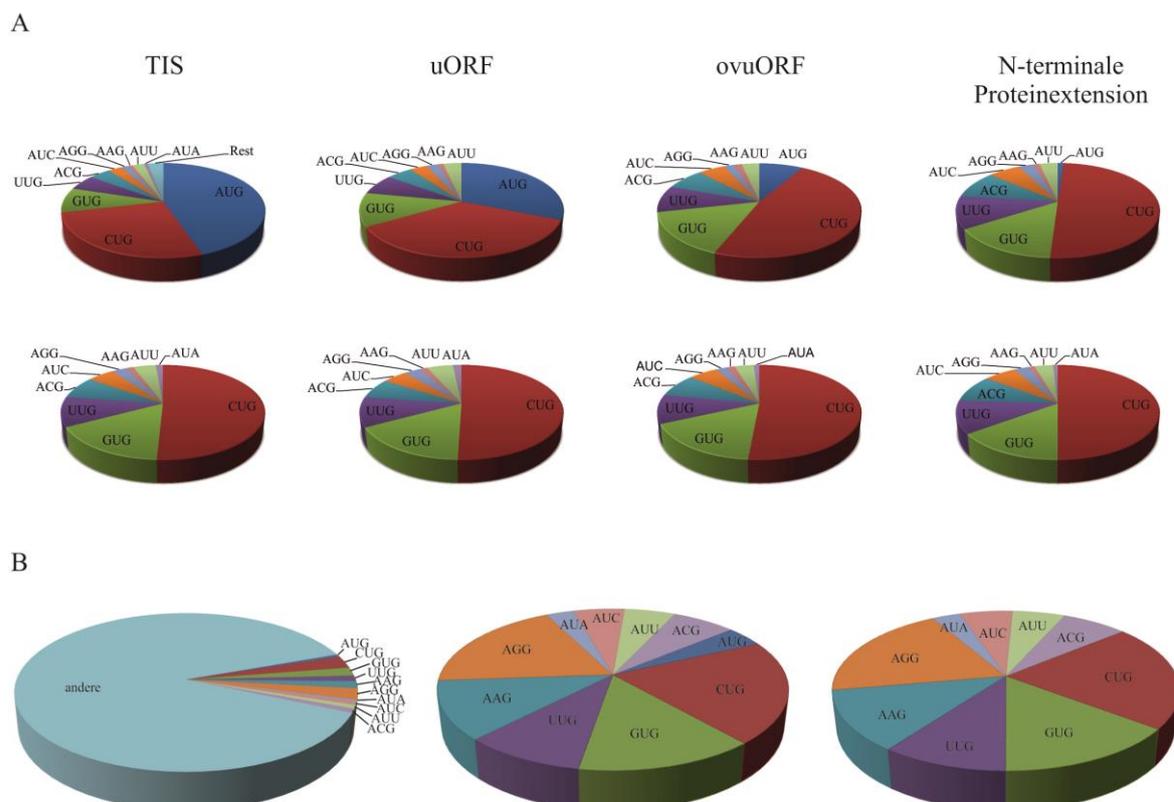


Abb. 21: Häufigkeit und Verteilung der TICs: **A:** TICs der identifizierten TISs. Oben: einschließlich der AUG-Codonen, unten: nur *near-cognate* Codonen wurden berücksichtigt. **B:** Codonhäufigkeit in den analysierten 5'-UTRs. Berücksichtigt wurden alle Codonen der verschiedenen Leserahmen. Links: alle möglichen Codonen, Mitte: die 10 möglichen Startcodonen, rechts: nur *near-cognate* Codonen.

Bei weiteren Analysen wurde die Häufigkeit aller möglichen Codons in den 5'-UTRs der analysierten Transkripte untersucht (Abb. 21B). Hierfür wurden alle Codonen der verschiedenen Leserahmen berücksichtigt. Bei 12% aller Codonen in der 5'-UTR handelt es sich um Codonen, die als mögliche Startstellen in Frage kommen.

3.2.7 Validierung der Identifizierten TISs

Um die hier beschriebenen Ergebnisse zu validieren, wurde in der Literatur nach experimentell nachgewiesenen Beispielen für non-AUG Starts und uORFs gesucht. Diese zuvor beschriebenen alternativen Translationsstarts wurden mit den vom neuronalen Netz detektierten TISs verglichen. 28 humane Gene mit non-AUG initiiierenden N-terminalen Proteinextensionen sind in der Literatur beschrieben und auf der *Database of mRNA Sequences with non-AUG Start Codons* (<http://bioinfo.iitk.ac.in/>) zusammengefasst. Von diesen 28 Genen sind 18 (*CACNG8*, *FGF2*, *HMHB1*, *LUZP6*, *MDFIC*, *MRVII*, *NPTXR*, *NRII2*, *OAZ3*, *PEG10*, *PIGX*, *POUF5F1*, *PRPS1L1*, *STIM2*, *TEAD1*, *TEAD3*, *TXNRD3* und *VEGFA*) in der hier untersuchten THP-1 Zelllinie nicht ausreichend translatiert. Aufgrund der geringen Sequenzabdeckung der entsprechenden RefSeq-Sequenzen konnten diese Beispiele nicht überprüft werden (Anhang 6.5). *SP3* wurde ebenfalls aus der Analyse ausgeschlossen, da in der Literatur ein non-AUG Start einer Isoform beschrieben ist, welche in diesem Datensatz nicht berücksichtigt wurde.

Insgesamt wurden die alternativen non-AUG Startstellen von neun Genen analysiert. Bei sechs Genen (*BAG1*, *DDX17*, *EIF4G2*, *GTF3A*, *MYC* und *NPW*) konnten die zuvor beschriebenen non-AUG N-terminalen Proteinextensionen vom neuronalen Netz korrekt ermittelt werden (Tab. 10). Die non-AUG-Starts von *TEAD4* und *HCK* konnten durch das neuronale Netz nicht bestätigt werden. Jedoch wurden bei diesen Genen uORFs vorhergesagt, welche die Initiation an den beschriebenen alternativen Starts möglicherweise inhibieren. Unter anderen zellulären Bedingungen, bei denen die uORFs inaktiviert sind, hätten diese alternativen TISs möglicherweise detektiert werden können. Des Weiteren wurde in dieser Arbeit an Stelle eines in früheren Studien identifizierten CUG-Startcodon von *WT1* (Bruening & J Pelletier 1996) ein stromaufwärts gelegenes ACG ermittelt.

Die anschließende Analyse bezog sich auf die in der Literatur beschriebenen uORFs. Insgesamt wurden in früheren Studien 50 uORFs in Säugetieren experimentell bestätigt,

welche in dem Artikel von Calvo et al. (2009) zusammengefasst sind. Von diesen 50 Beispielen konnten hier 25 (*ABCC2*, *AGTR1*, *AVPR1B*, *BACE1*, *BIRC2*, *CAMP*, *CD36*, *CYP27A1*, *ERBB2*, *ESR1*, *GATA6*, *GDNF*, *GRAP2*, *LPA*, *MOS*, *MYEOV*, *NOD2*, *OPRM1*, *RARB*, *SOCS1*, *TAL1*, *TEK*, *THPO*, *VEGFA* und *WNT2B*) wegen einer zu geringen Sequenzabdeckung der entsprechenden Transkripte nicht weiter analysiert werden (Anhang 6.5). Die uORFs von fünf weiteren Genen (*ASS1*, *CDKN1B*, *CEBPB*, *NR3C1* und *SEPT9*) konnten nicht überprüft werden, da sie in Transkriptisofomen identifiziert wurden, welche in dieser Arbeit nicht berücksichtigt wurden. Von den verbleibenden 20 Genen (Tab. 11) konnten bei fünf Genen (*AMD1*, *CITED2*, *HDLBP*, *HTT* und *ODC1*) sowohl die beschriebenen uORFs als auch die annotierten Startstellen der hauptkodierenden Sequenz korrekt identifiziert werden. Bei neun Genen (*ATF4*, *ATF5*, *BCKDK*, *CEBPA*, *DDIT3*, *MDM2*, *PPP1R15A*, *SP3* und *UCP2*) wurden nur die uORFs jedoch nicht die kanonischen TISs detektiert. Bei den verbleibenden sechs Genen (*ADH5*, *BCL2*, *FLI1*, *MTR*, *MVP* und *SLC7A1*) scheinen die uORFs in der hier analysierten Zelllinie inaktiv zu sein. Bei diesen Beispielen wurde eine Translationsinitiation nur an den annotierten Startstellen vorhergesagt. Zusätzlich wurden vom neuronalen Netz 12 neue uORFs in den 20 analysierten Genen identifiziert.

Ein vom neuronalen Netz detektierter uORF ist am Beispiel von *SP3* schematisiert (Abb. 22). In diesem Fall sind die RPFs deutlich an der uORF-TIS, jedoch nicht am kanonischen Start angereichert und nur die uORF-TIS wurde identifiziert. Dieses Ergebnis ist vereinbar mit einer älteren Studie in der gezeigt wurde, dass die Translation des *SP3* uORFs zu einer Inhibition der Translation der hauptkodierenden Sequenz führt (Sapetschnig et al. 2004).

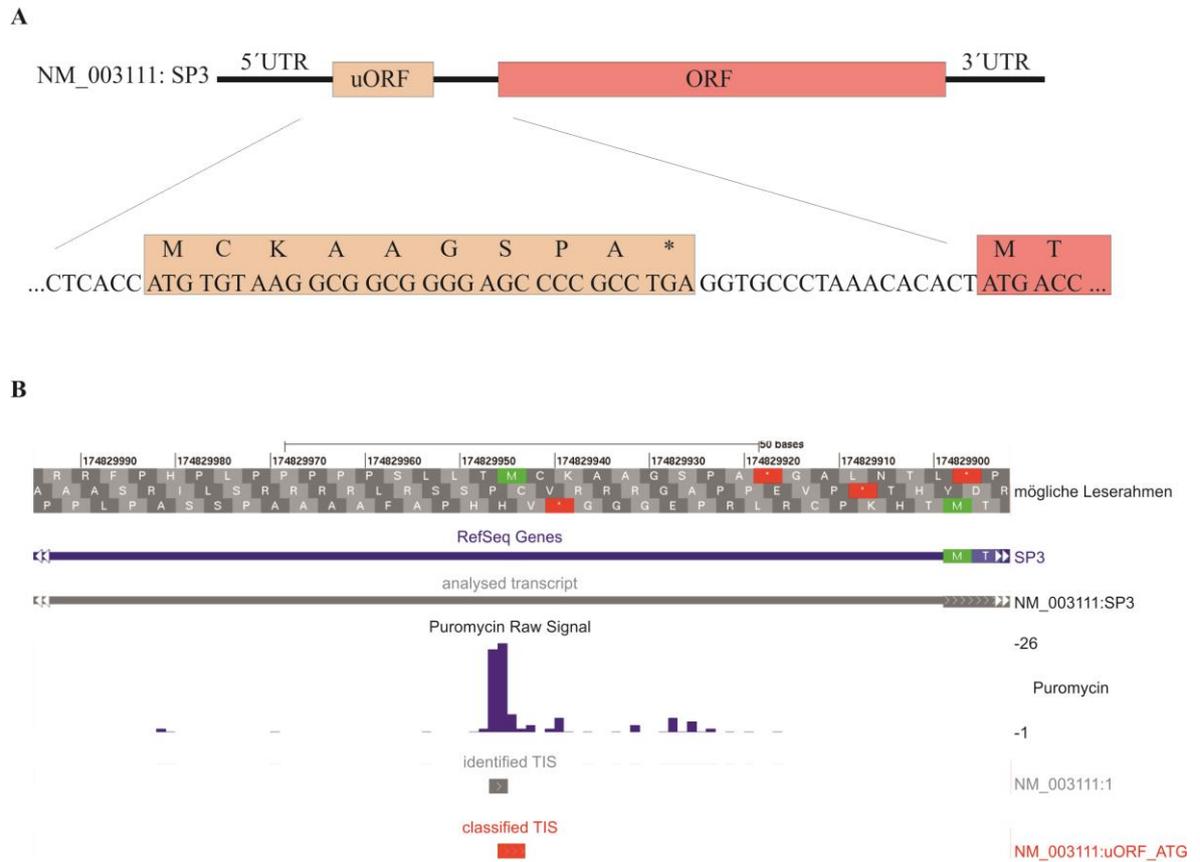


Abb. 22: Schematische Abbildung vom *SP3* uORF. **A:** Schematische Abbildung des *SP3* Transkripts (NM_003111) mit uORF (orange) und der hauptkodierenden Sequenz (rot). **B:** Sequenzabdeckung an der TIS vom *SP3* uORF in der Darstellungsweise des UCSC Genome Browsers. Die Startcodonen der verschiedenen Leserahmen sind in grün, Stoppcodonen in rot gekennzeichnet. Darunter ist in blau die *SP3* RefSeq Sequenz und in grau die analysierte Transkriptisoform schematisch dargestellt. Die vom neuronalen Netz vorhergesagte TIS ist mit einem grauen Balken, das klassifizierte TIC mit einem roten Balken markiert.

Tabelle 10: Validierung der identifizierten non-AUG initiierten N-terminalen Proteinextensionen unter Verwendung von Literaturdaten. Die in der Literatur beschriebenen Gene mit non-AUG Startcodonen sind in der Datenbank “*Database of mRNA Sequences with non-AUG Start Codons*” (<http://bioinfo.iitk.ac.in/>) zusammengestellt und wurden aus dieser zusammengetragen.

RefSeq ID	Gen	Alternative beschriebene TICs	Länge Protein-extension [AS]	Detektion der alternativen TIS	Detektion zusätzlicher alternativer TISs	Detektion zusätzlicher uORFs	Referenz
NM_004323	<i>BAG1</i>	CUG	71	ja	nein	nein	(Packham et al. 1997)
NM_001098504	<i>DDX17</i>	CUG	79	ja	nein	1 (CUG)	(Uhlmann-Schiffler et al. 2002)
NM_001172705	<i>EIF4G2</i>	GUG	206	ja	nein	nein	(Imataka et al. 1997)
NM_002097	<i>GTF3A</i>	CUG	235	ja	nein	nein	(Ivanov et al. 2011)
NM_002110	<i>HCK</i>	CUG	21	nein	nein	1 (CUG)	(Lock et al. 1991)
NM_002467	<i>MYC</i>	CUG	15	ja	nein	2 (CUG) 1 (UUG)	(Hann et al. 1988)
NM_001099456	<i>NPW</i>	CUG	52	ja	nein	2 (CUG)	(H. Tanaka et al. 2003)
NM_003213	<i>TEAD4</i>	UUG	73	nein	nein	1 (CUG) 1 (ACG)	(Stewart et al. 1996)
NM_024426	<i>WT1</i>	CUG	68	nein	1 (uACG)	nein	(Bruening & J Pelletier 1996)

Tabelle 11: Validierung der identifizierten uORFs in humanen Genen unter Verwendung von Literaturdaten. Die in der Literatur beschriebenen Gene mit funktionellen uORFs wurden der Arbeit von Calvo et al. (2009) entnommen.

RefSeq ID	Gen	Beschriebene uORF TICs	uORF Position	uORF Länge [AS]	Detektion des uORF	Detektion der annotierten TIS	Detektion von N-terminalen Protein- extensionen	Detektion zusätzlicher uORFs	Referenz
NM_000671	<i>ADH5</i>	AUG AUG	-65 -35	20 10	nein nein	ja	nein	nein	(Kwon et al. 2001)
NM_001634	<i>AMD1</i>	AUG	-310	6	ja	ja	nein	nein	(Mize et al. 1998)
NM_001675	<i>ATF4</i>	AUG AUG AUG	-860 -798 -699	1 3 59	ja ja nein	nein	nein	nein	(Harding et al. 2000)
NM_012068	<i>ATF5</i>	AUG AUG	-246 -124	3 59	ja nein	nein	nein	nein	(Watatani et al. 2008)
NM_001122957	<i>BCKDK</i>	AUG	-177	19	ja	nein	nein	1 (CUG)	(Muller & Danner 2004)
NM000633	<i>BCL2</i>	AUG	-119	11	nein	ja	nein	2 (AUG) 1 (UUG) 1 (AUC)	(Harigai et al. 1996)
NM_004364	<i>CEBPA</i>	AUG	-25	5	ja	nein	nein	nein	(Lincoln et al. 1998)
NM_006079	<i>CITED2</i>	AUG	-82	3	ja	ja	nein	1 (AUC)	(Beucken van den et al. 2007)
NM001195056	<i>DDIT3</i>	AUG	-341	34	ja	nein	nein	nein	(Jousse et al. 2001)
NM_002017	<i>FLI1</i>	AUG GUG	-41 -37	25 17	nein nein	ja	nein	1 (CUG)	(Sarrazin et al. 2000)
NM_005336	<i>HDLBP</i>	AUG	-178	13	ja	ja	nein	nein	(Rohwedel et al. 2003)
NM_002111	<i>HTT</i>	AUG	-125	21	ja	ja	nein	nein	(J. Lee et al. 2002)
NM_002392	<i>MDM2</i>	AUG AUG	-231 -97	14 14	ja nein	nein	nein	nein	(Brown et al. 1999) (Jin et al. 2003)

RefSeq ID	Gen	Beschriebene uORF TICs	uORF Position	uORF Länge [AS]	Detektion des uORF	Detektion der annotierten TIS	Detektion von N-terminalen Protein- extensionen	Detektion zusätzlicher uORFs	Referenz
NM_000254	<i>MTR</i>	AUG AUG	-383 -360	141 30	nein nein	ja	nein	1 (AAG)	(Col et al. 2007)
NM_017458	<i>MVP</i>	AUG	-78	18	nein	ja	nein	1 (GUG)	(Holzmann et al. 2001)
NM_002539	<i>ODC1</i>	AUG	-161	10	ja	ja	nein	nein	(Ivanov et al. 2008) (Pegg 2006)
NM_014330	<i>PPP1R1 5A</i>	AUG AUG	-206 -107	22 26	ja nein	nein	nein	nein	(Lee et al. 2009)
NM_003045	<i>SLC7A1</i>	AUG	-378	110	nein	ja	1 (CUG)	1 (CUG) 1 (UUG)	(Fernandez et al. 2002)
NM_003111	<i>SP3</i>	AUG	-47	9	ja	nein	nein	nein	(Sapetschnig et al. 2004)
NM_003355	<i>UCP2</i>	AUG	-246	36	ja	nein	nein	1 (CUG)	(Hurtaud et al. 2006)

3.2.8 Ermittlung einer Konsensussequenz der vorhergesagten TISs

Für eine effiziente Translationsinitiation ist der Nukleotidkontext des Startcodons von großer Bedeutung. Bei Säugetieren ist die Kozaksequenz $CGCC(A/G)CCAUGG$ optimal für die Initiation. Vor allem die Nukleotide an der Position -3 und +4 (relativ zum A vom AUG) sind hierbei relevant (Kozak 1986). Demgemäß wurde in dieser Studie untersucht, ob die neu identifizierten TISs von einer Kozaksequenz oder einer anderen Konsensussequenz flankiert sind. Um die TIS-Sequenzen bildlich darstellen und auswerten zu können, wurde das Programm WebLogo 3.3 verwendet (Crooks et al. 2004).

Als erstes wurde der Sequenzkontext von den funktionellen TIS-Kategorien analysiert. Wie erwartet entspricht der Konsensus der kanonischen TICs einer optimalen Kozaksequenz. Diese Konsensussequenz konnte bei uORFs, CDS überlappenden uORFs und bei N-terminalen Proteinextensionen nicht festgestellt werden (Abb. 23).

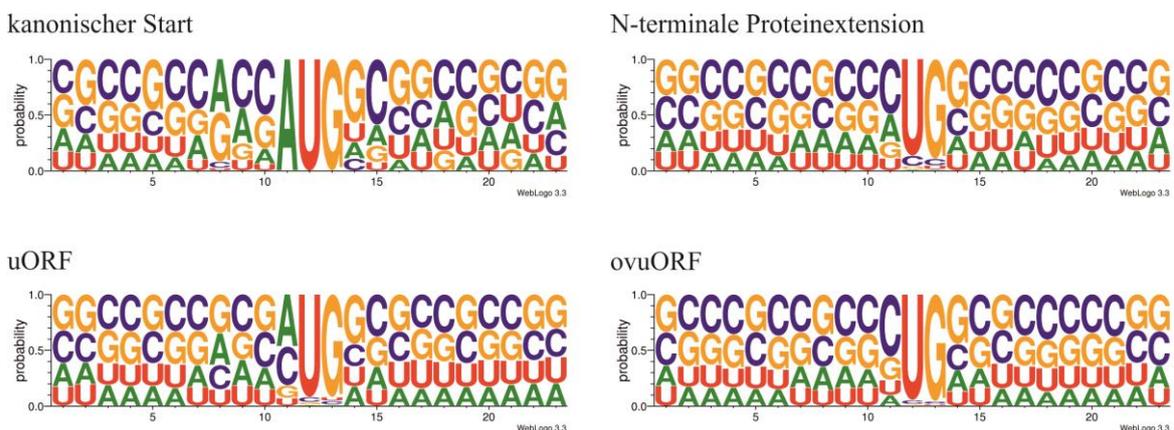


Abb. 23: Sequenzkontexte der TISs verschiedener funktioneller Kategorien. Die Abbildung wurde mit Hilfe des Programms WebLogo 3.3 erstellt (Crooks et al. 2004). Die Größe der Symbole spiegelt die relative Häufigkeit der Nukleotide an dieser Position wider.

Darüber hinaus wurde der Sequenzkontext von den TISs in Abhängigkeit von den verschiedenen alternativen Startcodonen analysiert (Abb. 24). TISs mit einem AUG-Codon weisen häufig eine optimale Kozaksequenz auf. Eine weitere TIS mit dem Startcodon AUA hat keinen optimalen Konsensus, jedoch ist hier sehr häufig ein A- oder G-Nukleotid an der -3 Position und ein G-Nukleotid an der +4 Position zu vermerken. Demgegenüber wurde bei AUU-TICs nur ein A oder G an der -3 Position ermittelt. Bei TISs mit anderen Startcodonen weicht die Sequenz sehr stark von der optimalen Kozaksequenz ab und es wurde auch kein anderer Konsensus nachgewiesen.

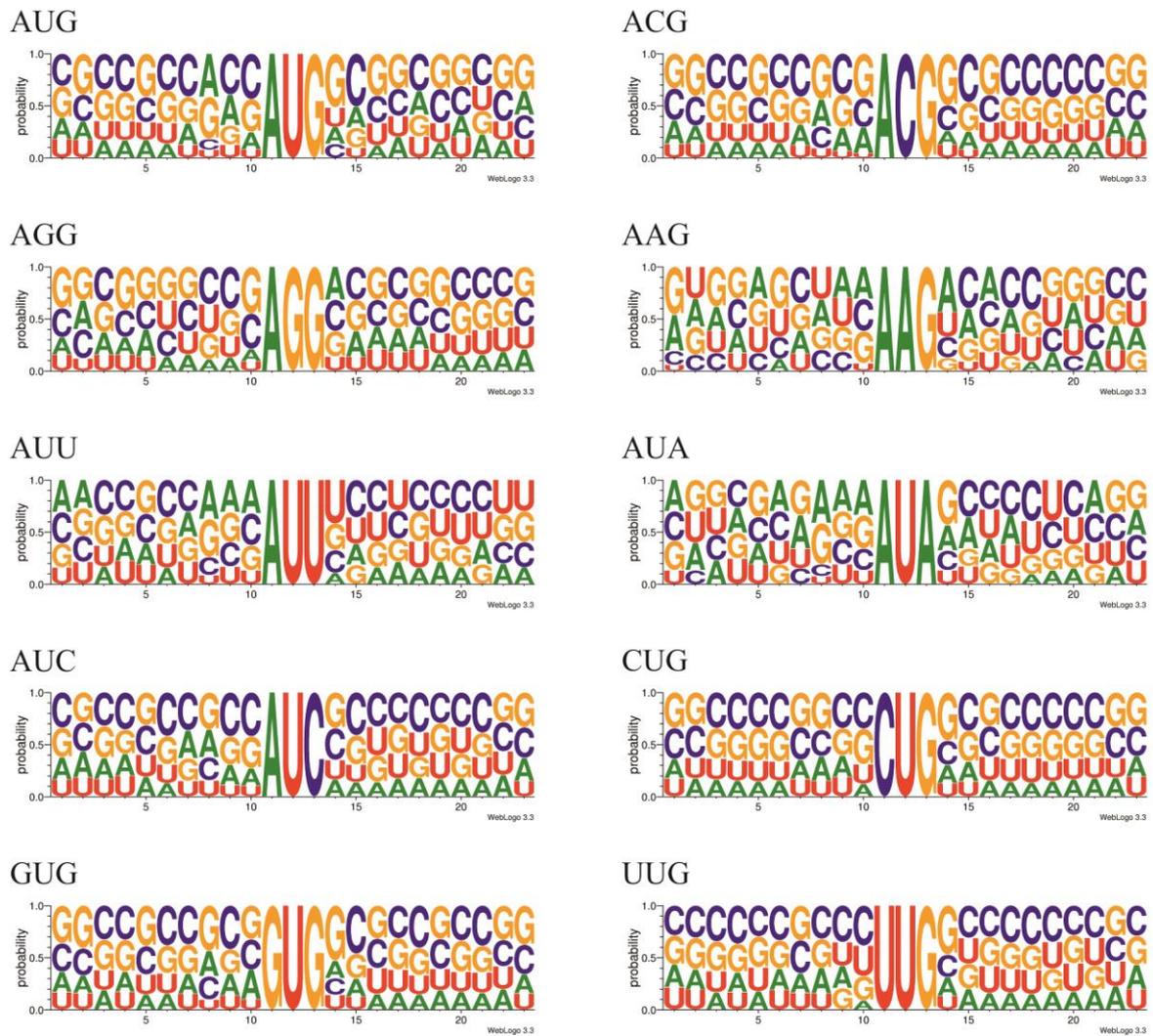


Abb. 24: Sequenzkontexte der TISs in Abhängigkeit von den verschiedenen Startcodonen. Die Abbildung wurde mit Hilfe des Programms WebLogo 3.3 erstellt (Crooks et al. 2004). Die Größe der Symbole spiegelt die relative Häufigkeit der Nukleotide an dieser Position wider.

3.2.9 Konservierung der TIC-Sequenzen zwischen Primaten

Die evolutionäre Konservierung von TICs ist ein Indiz für eine mögliche biologische Relevanz und Funktionalität dieser Startstellen. Aus diesem Grund wurde der Konservierungsgrad der hier identifizierten TICs zwischen neun nichthumanen Primaten (Schimpanse, Gorilla, Orang-Utan, Rhesusaffe, Pavian, Krallenaffe, Koboldmaki, Lemur, Galagos) analysiert. Hierfür wurden Konservierungswerte von Kontrollcodonen und von den ermittelten TICs miteinander verglichen. Diese Konservierungswerte wurden von UCSC bezogen, wobei für jede Positionscoordinate ein Konservierungswert geliefert wird, welcher zwischen 0 und 1 liegt.

Wie zu erwarten sind die AUG-Codonen der kanonischen Startstellen hoch konserviert. Für diese Codonen konnte ein Konservierungswert von 0,58 ermittelt werden. Im Vergleich dazu liegt dieser Wert bei den Kontrollcodonen nur bei 0,02 (T-Test, $p < 10^{-10}$), (Abb. 25). Fasst man alle TICs der vom neuronalen Netz identifizierten TISs zusammen, sind diese höher konserviert als die Kontrollcodonen (Anova, $p < 10^{-10}$). Auch für die *near-cognate* Codonen, ohne AUG, konnte ein höherer Konservierungsgrad nachgewiesen werden (Anova, $p = 3,9 \times 10^{-3}$).

Die Konservierung der Codonen variiert zwischen den verschiedenen Kategorien von Startstellen. Bei N-terminalen Proteinextensionen ist nur das AAG-Codon konserviert. Demgegenüber konnten bei uORFs sechs konservierte Codonen (ACG, AUA, AUC, AUG, AUU und UUG) und bei den CDS überlappenden uORFs zwei konservierte Codonen (AUA, AUG) nachgewiesen werden. Des Weiteren variiert die Konservierung sehr stark zwischen den verschiedenen Codontypen. Die höchst konservierten Codonen sind AUGs und AUAs, in der uORF-Kategorie und bei den CDS überlappenden uORFs. AGG-, CUG- und GUG-Codonen sind in keiner der Kategorien konserviert. Die genauen Konservierungswerte sowie die Signifikanzwerte für alle Codonen sind im Anhang 6.6 aufgelistet.

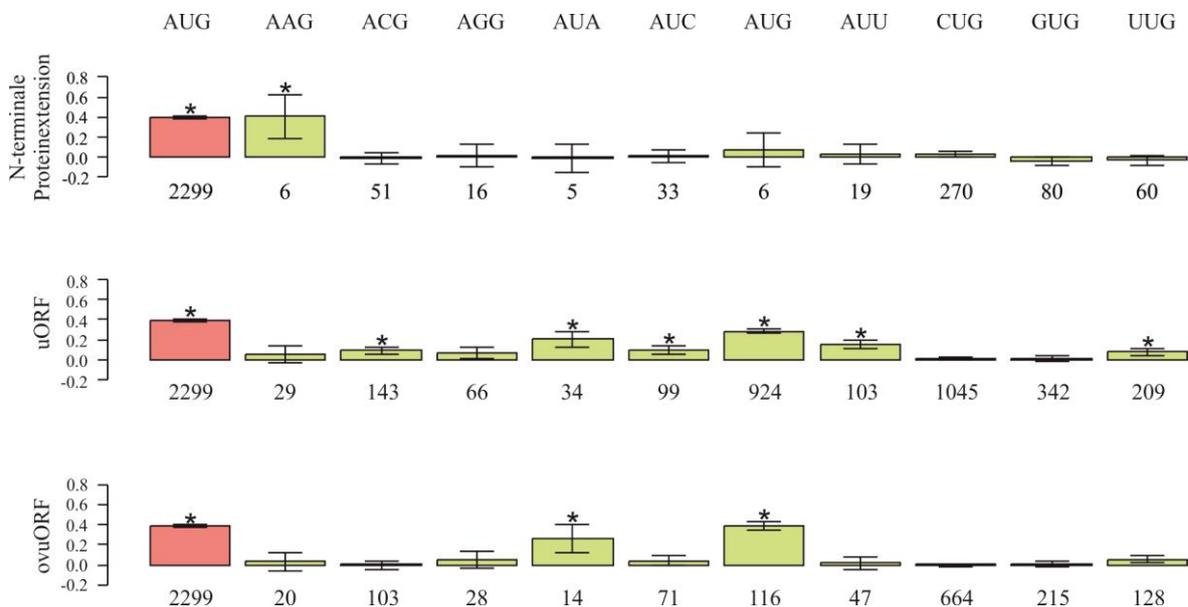


Abb. 25: Konservierung der Sequenzen identifizierter TICs zwischen Primaten. Die Unterschiede der gemittelten Konservierungswerte zwischen Kontrollcodonen und identifizierten TICs sind hier für die verschiedenen Codontypen und für die funktionellen Kategorien grafisch dargestellt. Der rote Balken zeigt den Unterschied der Konservierungswerte zwischen AUG-Codonen der kanonischen Startstellen und Kontrollcodonen. Die Zahlen unter den Balken geben die Anzahl der analysierten TICs an. TICs, die eine statistisch signifikante Konservierung aufweisen wurden mit Sternen markiert ($p < 0.0016$).

3.2.10 Konservierung der TIC-Sequenzen zwischen Mensch und Maus

Weiterhin wurde die Konservierung der TIC-Sequenzen zwischen Mensch und Maus untersucht. Hierfür wurden die in dieser Arbeit definierten TICs mit TICs verglichen, die in einer vorangegangenen Studie in embryonalen Mausstammzellen bestimmt wurden (Ingolia et al. 2011). Von den in THP-1 Zellen identifizierten 7251 TICs konnten 3294 TICs auf das Mausgenom kartiert werden. Von diesen sind 2141 zwischen Mensch und Maus konserviert. Von den TICs, welche in der Studie von Ingolia et al. 2011 in Mäusen definiert wurden, konnten 7391 mit dem humanen Genom verglichen werden, wobei 4458 konserviert sind (Tab. 12).

Die kanonischen Startstellen sind zwischen Mensch und Maus hoch konserviert. Demgegenüber konnte nur bei 38% und 46% der TICs von uORFs, ovuORFs und N-terminalen Proteinextensionen eine Sequenzkonservierung vermerkt werden.

Tabelle 12: Konservierung der TIC-Verwendung zwischen Mensch und Maus. Die humanen 5'-UTRs wurden gegen die entsprechenden Positionen im Mausgenom (mm9) kartiert. Die Maus TICs wurden aus der Studie von Ingolia et al. 2011 bezogen.

	identifizierte humane TICs	bei Maus konserviert	identifizierte Maus TICs	bei Mensch konserviert
TIC gesamt	3294	2141 (65%)	7391	4458 (60,3%)
kanonische CDS	1321	1293 (97,8%)	2350	2312 (98,4%)
N-terminale Proteinextension	243	98 (40,3%)	404	157 (38,9%)
uORF	970	455 (46,9%)	1748	833 (47,6)
ovuORF	681	263 (38,6%)	1041	427 (41%)

Des Weiteren wurden die TICs analysiert und verglichen, die sowohl beim Menschen als auch bei der Maus als TISs genutzt werden. Bei diesen 1490 gemeinsam genutzten TISs konnte ein sehr ähnliches Muster bei der Codonverteilung innerhalb der verschiedenen Kategorien festgestellt werden (Abb. 26 und Anhang 6.7).

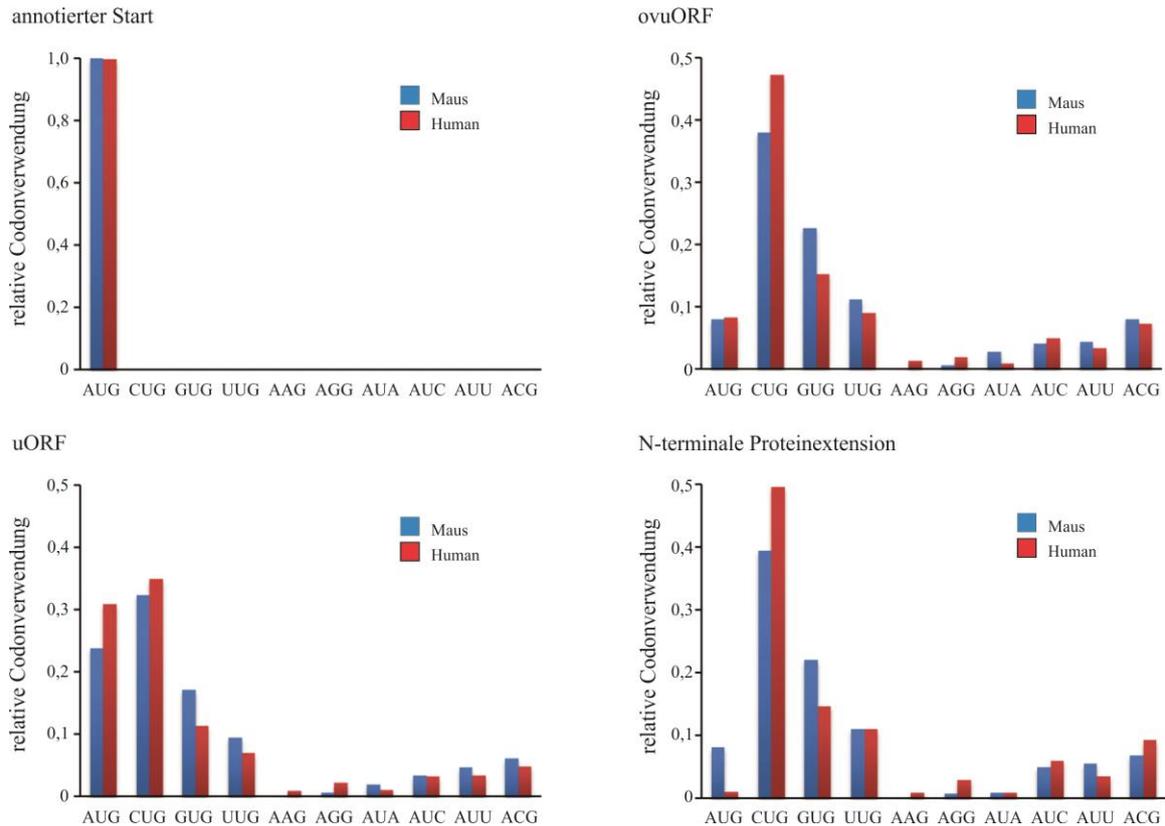


Abb. 26: Vergleich der Codonverteilung zwischen Mensch und Maus. Die Ergebnisse der Mausexperimente stammen aus der Studie von Ingolia et al. (2011).

3.2.11 Annotationstrack der TISs

Die in dieser Arbeit neu identifizierten TISs, die klassifizierten TISs und die Sequenzabdeckung der Transkripte wurden von Alexander Herrmann (Genomische Gastroenterologie, Universitätsklinikum Schleswig-Holstein, Kiel) zu einem Annotationstrack zusammengestellt, welcher online in der Darstellungsweise des UCSC Genome Browser zur Verfügung steht (<http://gengastro.1med.uni-kiel.de/suppl/footprint/>). Für jede in dieser Studie analysierte Transkriptvariante können die identifizierten TISs, die klassifizierten TISs, die entsprechenden TICs sowie die kartierten Footprint-Sequenzen erfasst werden. Ein Beispiel für einen Annotationstrack ist in Abb. 27 für *AMD1* dargestellt.

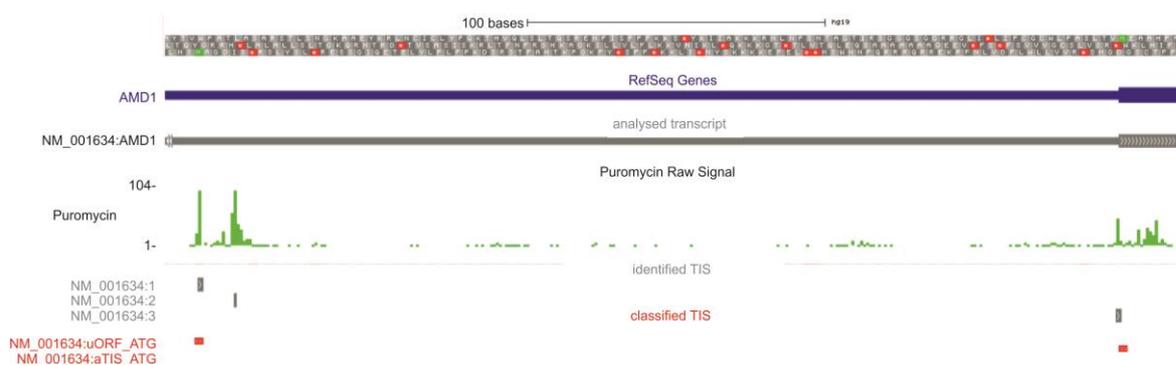


Abb. 27: Screenshot eines Annotationstracks am Beispiel von *AMD1*: (NM_001634). Oben: mögliche Leserahmen. Mit Rot sind die Stoppcodonen mit Grün die Startcodonen gekennzeichnet. Der blaue und graue Balken symbolisiert die analysierte Transkriptvariante, wobei der schmale Bereich die 5'-UTR und der dicke Bereich die CDS repräsentiert. Darunter sind in grün die kartierten RPFs abgebildet. Die vom neuronalen Netz vorhergesagten TISs sind mit grauen Balken, die klassifizierten TICs mit roten Balken markiert. Bei dem dargestellten Beispiel konnte der bereits in der Literatur beschriebene uORF, sowie die kanonische Startstelle von *AMD1* detektiert werden (Mize et al. 1998).

3.3 Etablierung des Ribosomen-Profilings für humanes Lebergewebe

Nachdem die Ribosomen-Profilings-Methode erfolgreich für eine humane Zelllinie etabliert wurde, sollte diese auch für humanes Gewebe angepasst werden. Hierfür wurden Footprint-Bibliotheken aus Lebergeweben prozessiert. Für jeden Ansatz wurden 50 mg Gewebe homogenisiert und in Polysomenpuffer, welcher mit Cycloheximid versetzt war resuspendiert. Unter Verwendung von 2000 U RNaseI (Life Technologies, Darmstadt, Deutschland) wurden die Polysomen hydrolysiert, was mit Hilfe von Polysomenprofilen beurteilt wurde.

Für die Präparation der Leber-Footprint-Bibliotheken wurde die bei den THP1-Zellen durchgeführte Methode optimiert. Die THP-1 Footprint-Bibliotheken enthielten nur einen geringen Anteil von informativen Sequenzen und 70 bis 80% der sequenzierten Fragmente repräsentierten ribosomale RNA Kontaminationen.

Aus diesem Grund wurden die polyadenylierten RNA-Fragmente der Leberproben mit biotinylierten Oligonukleotiden hybridisiert, die zu bestimmten Abschnitten der 18S rRNA *reverse* komplementär waren. Mit Hilfe von magnetischen, streptavidinbeschichteten Kügelchen wurden die Oligo-rRNA-Hybride aus den Proben entfernt. Die prozessierten Bibliotheken wurden dann sequenziert, wobei zwischen 57.933.731 und 213.212.067 sequenzierte Fragmente pro Bibliothek resultierten, von denen nur noch 19-50% rRNAs

repräsentierten (Tab. 13). Diese restlichen rRNA-Sequenzen sowie Sequenzen die Adapter oder Primerkontaminationen repräsentierten wurden aus den Analysedaten entfernt. Die verbleibenden RPFs wurden anschließend auf das humane Referenzgenom (UCSC GRCh37/hg19) kartiert.

Tabelle 13: Analyse der Leber-Footprint-Sequenzen

Leberprobe	Sequenzen total	rRNA Kontaminationen	rRNA in %	RPFs	RPFs in %
1	137.343.003	69.042.557	50,3	65.467.181	47,7
2	57.933.731	16.218.819	28,0	40.454.307	69,8
3	112.253.929	41.658.060	37,1	66.855.002	59,6
4	99.591.673	29.207.218	29,3	68.996.951	69,3
5	168.574.306	41.153.227	24,4	126.275.672	74,9
6	76.426.172	26.833.876	35,1	48.273.881	63,2
7	148.980.444	68.767.851	46,2	78.447.281	52,7
8	111.395.032	30.257.288	27,2	80.257.841	72,0
9	77.150.953	35.052.297	45,4	41.243.820	53,5
10	110.538.568	38.624.185	34,9	71.272.168	64,5
11	165.354.041	57.112.280	34,5	107.209.430	64,8
12	211.870.983	104.414.691	49,3	103.037.753	48,6
13	105.132.430	20.537.538	19,5	82.957.132	78,9
14	213.212.067	52.326.523	24,5	158.688.801	74,4
15	196.750.429	63.167.872	32,1	131.961.748	67,1
16	205.331.112	63.652.247	31,0	140.139.077	68,3
17	168.932.731	58.861.145	34,8	108.458.716	64,2
18	190.315.692	80.189.005	42,1	101.199.436	53,2

4. Diskussion

Die Regulation der Translation ist aufgrund ihrer Bedeutung für die Kontrolle der Genexpression und Steigerung der Proteindiversität zunehmend von Interesse. Immer mehr Studien belegen, dass die Information einer mRNA für die Bildung verschiedener Proteinisoformen genutzt werden kann, indem die Translation an mehreren alternativen TISs eines Transkripts initiiert (Lee et al. 2012). Darüber hinaus werden immer mehr regulatorische uORFs beschrieben, welche die Translation der hauptkodierenden Sequenz kontrollieren können (Morris & Geballe 2000, Calvo et al. 2009a). Neue Arbeiten zeigen zudem, dass die Initiation der Translation nicht zwingend an einem AUG, sondern auch an *near-cognate* Codonen erfolgen kann (Ingolia et al. 2009, Ingolia et al. 2011, Fritsch et al. 2012, Lee et al. 2012). Diese alternativen Initiationsereignisse sind einerseits für die Regulation der Genexpression und andererseits für die Erhöhung der Proteinviefalt von großer Relevanz. Eine zuverlässige Vorhersage dieser TISs mit *in silico* Methoden, wie sie zur Identifizierung von AUG-Startcodonen verwendet werden, war nicht möglich.

In der vorliegenden Arbeit wurde die erste transkriptomweite Übersicht von alternativen TISs für eine humane Zelllinie erstellt, indem das zuvor beschriebene Ribosomen-Profiling (Ingolia et al. 2009, Guo et al. 2010) leicht modifiziert an mit Puromycin behandelten Zellen angewandt wurde. Diese Methode basiert auf der transkriptomweiten Sequenzierung von durch Ribosomen geschützten mRNA-Fragmenten. Durch den Abgleich der RPF-Sequenzen mit dem humanen Transkriptom werden die genauen Positionen der Ribosomen auf den Transkripten und somit die tatsächlich translatierten Bereiche identifiziert.

Bei den Pilotexperimenten wurden die Zellen nur mit dem Elongationsinhibitor Cycloheximid behandelt, was zur Fixierung elongierender Ribosomen auf der mRNA führte. Jedoch ist diese Standardmethode des Ribosomen-Profiling für die Identifizierung von TISs nicht optimal, da zahlreiche RPFs in überlappenden kodierenden Bereichen, wie z.B. ovuORFs kartiert wurden, was keine eindeutige Zuordnung zu einer spezifischen TIS bzw. zu einem spezifischen offenen Leserahmen ermöglicht. Infolgedessen wurden die Zellen mit dem Elongationsinhibitor Puromycin behandelt, was zur Dissoziation elongierender und zur Anreicherung von initiiierenden Ribosomen an den TISs führte.

Insgesamt konnten somit in 4364 analysierten Transkripten 7251 TISs identifiziert und klassifiziert werden, von denen 68% in den 5'-UTRs der Transkripte lokalisiert sind. Diese *upstream* gelegenen TISs spiegeln uORFs, ovuORFs und N-terminale Proteinextensionen

wider, deren Translation sowohl an *upstream* gelegenen AUG- als auch an non-AUG-Codonen initiiert.

Zusammenfassend ist das Ribosomen-Profilierung hervorragend geeignet, um das Kodierungspotential von Genomen zu erfassen und zu analysieren (Ingolia et al. 2011).

4.1 Anreicherung von RPFs an den TISs mit Hilfe von Puromycin

Um eine akkurate Identifizierung von TISs zu gewährleisten, wurden die Ribosomen mit Hilfe des Elongationsinhibitors Puromycin angereichert. Puromycin ist ein Aminoacyl-tRNA Analogon und konkurriert mit der tRNA um die A-Stelle im Ribosom. Die wachsende Polypeptidkette wird auf Puromycin übertragen, was zu einer verfrühten Termination der Translation führt und die Dissoziation der elongierenden Ribosomen vom Transkript bewirkt (Allen & Zamecnik 1962, Nathans 1964). Da dieses Agens keinen Einfluss auf die Initiation der Ribosomen hat, ist von einer fortlaufenden Initiation auszugehen. Neu initiierte Ribosomen wurden dann durch Zugabe von Cycloheximid an den TISs arretiert.

Bei der Analyse der kartierten Footprint-Sequenzen zeigte sich, dass einige Ribosomen als Folge der konkurrierenden Wirkungsweise von Puromycin und Cycloheximid erst *downstream* von den TISs arretiert wurden. Wenn zuerst Cycloheximid an das elongierende Ribosom bindet, wird dieses fixiert, wodurch die Bindung von Puromycin und folglich eine Dissoziation des Ribosoms inhibiert wird. Infolgedessen konnte zwar eine deutliche Sequenzabdeckung an den TISs der Transkripte erreicht werden, die aber erst *downstream* von den Initiationsstellen langsam ausläuft.

In anderen Publikationen wurden die Ribosomen mit Hilfe des Elongationsinhibitors Harringtonin (Ingolia et al. 2011) oder mit Hilfe des Translationsinhibitors Lactimidomycin (Lee et al. 2012) an den TISs angereichert. Harringtonin kann nur an die freie ribosomale 60S Untereinheit binden. Nach der Assemblierung eines 80S Ribosoms an einer TIS aus der Harringtonin gebundenen 60S und der 40S Untereinheit wird die Elongation inhibiert (Fresno et al. 1977, Robert et al. 2009). Da Ribosomen, die Harringtonin gebunden haben keine Elongationsrunde vollziehen, erhält man mit diesem Agens deutliche Signale an den TISs. Allerdings ist dieser Inhibitor unter bestimmten Bedingungen resistent gegen non-AUG Startcodonen (Ingolia et al. 2011). Im Gegensatz zu Harringtonin bindet Lactimidomycin an die E-Stelle des bereits am Startcodon

assemblierten 80S Komplexes und blockiert die Ribosomentranslokation (Schneider-Poetsch et al. 2010). Da die Bindestelle für dieses Molekül nur erreichbar ist, wenn sich noch keine tRNA in der E-Stelle befindet, werden nur initiiierende Ribosomen arretiert.

4.2 Etablierung und Optimierung des Ribosomen-Profilings für humanes Lebergewebe

Das Ribosomen-Profilings konnte erfolgreich für humanes Lebergewebe eingesetzt werden. Um hierbei ein umfassendes Spektrum von Genen, die unter verschiedenen physiologischen Bedingungen exprimiert sind, abzudecken, wurde Lebergewebe mit unterschiedlich schweren Verfettungs- und Entzündungsgraden sowie gesundes Gewebe verwendet.

Bei dem THP-1 Zelllinienexperiment wurden Footprint-Bibliotheken generiert, die neben den vergleichsweise wenigen RPFs auch viele rRNA-Fragmente enthielten (Abschnitt 3.1.3), da ein intakter Ribosomenkomplex nicht nur die 30 nt lange Footprint-RNA sondern auch ribosomale RNA einer Gesamtlänge von ~6,9 kb enthält und während der Hydrolyse teilweise schützt (Ingolia et al. 2012). Hierbei wurde festgestellt, dass bestimmte rRNA-Abschnitte einen sehr großen Anteil der Kontaminationen ausmachen, vermutlich weil die RNA-Hydrolyse an Ribosomen zu reproduzierbaren Schnitten an spezifischen Positionen der rRNA führt (Fritsch et al. 2012, Ingolia et al. 2012) .

Indem biotinylierte Oligonukleotide eingesetzt wurden, die *reverse* komplementär zu den häufigsten rRNA-Fragmenten waren, konnten die Kontaminationen während der Leberbankenpräparation erfolgreich reduziert werden, wodurch eine verbesserte Ausbeute von informativen RPFs erreicht wurde. Die verbleibenden rRNA-Verunreinigungen mussten, wie zuvor auch mit bioinformatischen Methoden aus den Sequenzdaten gefiltert werden.

Des Weiteren wurde die Methode optimiert, indem für die Amplifizierung der RPFs aus Lebergeweben PCR-Primer mit verschiedenen Barcodes zum Einsatz kamen, was eine parallele Sequenzierung von mehreren Footprint-Bibliotheken auf einer einzigen *Lane* ermöglichte.

In dieser Arbeit wurde das Ribosomen-Profilings zum ersten Mal für ein humanes Gewebe etabliert und optimiert. Indem die rRNA-Kontaminationen in den Bibliotheken verringert

und drei Bibliotheken parallel sequenziert wurden, konnte die Ausbeute an informativen Sequenzen verbessert und die Kosten für das Experiment verringert werden.

4.3 Diversität und Komplexität der neu annotierten TISs

Von den 7251 klassifizierten TISs in 5062 analysierten Transkripten wurde der Großteil in den 5'-UTRs detektiert und nur 32% der identifizierten TISs repräsentieren die kanonischen Startstellen.

Für 4364 Transkripte konnte mindestens eine TIS annotiert werden. Bei 2305 von diesen 4364 Transkripten (53%) initiiert die Translation an der CDS. In den restlichen Transkripten wurden nur *upstream* gelegene Initiationsereignisse nachgewiesen. So ist möglicherweise die Translation an den kanonischen TISs von 1694 Transkripten durch die Aktivierung von uORFs inhibiert, denn bei diesen Transkripten wurden RPFs nur an Startstellen von uORFs und ovuORFs nachgewiesen. Diese Beobachtung verdeutlicht die möglichen regulatorischen Funktionen von uORFs in Bezug auf die Translationsinitiation an den kanonischen CDS (Calvo et al. 2009).

Für 698 analysierte Transkripte konnte keine TIS in der definierten Zielregion detektiert werden. Dies ist eventuell darauf zurückzuführen, dass die Translation bei diesen Transkripten nur *downstream* von den kanonischen Startstellen initiiert, was bei der Analyse nicht berücksichtigt wurde. Dieses Ergebnis ist vereinbar mit den Resultaten einer vorangegangenen Studie an embryonalen Mausstammzellen, bei der 25% der detektierten TISs *downstream* vom kanonischen Start lokalisiert sind und N-terminal verkürzte Proteine oder alternative ORFs repräsentieren (Ingolia et al. 2011).

Insgesamt konnten für 44% der analysierten RefSeq Sequenzen uORFs und für 28% ovuORFs vorhergesagt werden. Dieses Ergebnis wird ebenfalls durch die Studie von Ingolia et al. 2011 bestärkt, bei der für 31% der Transkripte uORFs und für 25% ovuORFs identifiziert wurden (Ingolia et al. 2011). Darüber hinaus wurden in einer aktuellen Publikation nachgewiesen, dass 54% der Transkripte von HEK-293 Zellen eine oder mehrere TISs aufweisen, die *upstream* von der kanonischen Startstelle lokalisiert sind und die zum Großteil uORFs und ovuORFs repräsentieren (Lee et al. 2012). Zudem besagen bioinformatische Analysen, dass über 50% der humanen RefSeq Sequenzen uORFs enthalten (Matsui et al. 2007, Calvo et al. 2009, Iacono et al. 2005).

Eine Analyse der TICs der vorhergesagten TISs zeigte, dass das häufigste verwendete Startcodon wie zu erwarten AUG ist (47%). Die Translation an *upstream* gelegenen TISs initiiert hauptsächlich an *near-cognate* Codonen, wobei CUG (26%) und GUG (9%) die am meisten genutzten TICs sind. Ein ähnliches Muster der TIC-Verwendung wurde auch für HEK-293 Zellen ermittelt (Lee et al. 2012). Hier initiiert die Translation ebenfalls am häufigsten an AUG- (51%) und CUG-Codonen (16%), und vergleichbare Ergebnisse wurden auch für embryonale Mausstammzellen festgestellt (Ingolia et al. 2011).

4.4 Validierung der neu annotierten TISs

Um die Ergebnisse dieser Arbeit zu validieren, wurden die hier annotierten TISs mit zuvor in der Literatur beschriebenen und experimentell nachgewiesenen non-AUG Startstellen und uORFs verglichen. Infolge einer Literatur- und Datenbankrecherche konnten 28 humane Gene mit beschriebenen non-AUG-N-terminalen Proteinextensionen und 50 Säugetiergene mit beschriebenen uORFs ausgemacht werden.

Die sehr limitierte Anzahl von beschriebenen alternativen Translationsereignissen verdeutlicht das Fehlen zuverlässiger *in silico*-Methoden, um diese alternativen Starts korrekt vorherzusagen. Erst mit der Etablierung genomweiter experimenteller Nachweismethoden konnte veranschaulicht werden, wie unerwartet häufig uORFs und *near-cognate* TICs sind (Ingolia et al. 2009, Ingolia et al. 2011, Fritsch et al. 2012, S. Lee et al. 2012).

Da viele der beschriebenen Gene mit alternativen TISs in den hier verwendeten Zellen nicht ausreichend exprimiert waren, verringerte sich die Anzahl der vergleichbaren alternativen Initiationsereignisse zusätzlich. Darüber hinaus sind in der Literatur Transkriptvarianten beschrieben, welche bei der hier durchgeführten genomweiten Analyse nicht berücksichtigt wurden.

Letztendlich waren die Ergebnisse der hier angewandten Analyse mit acht von neun beschriebenen non-AUG-N-terminalen Proteinextensionen vergleichbar, und nur bei *WT1* wurde anstelle des beschriebenen CUG- (Bruening & J Pelletier 1996), ein *upstream* lokalisiertes ACG-Startcodon ermittelt. Eine Initiation an diesem ACG führt zur Bildung einer um 20 Aminosäuren verlängerten Proteinvariante. Hierbei ist zu berücksichtigen, dass die am CUG initiiierende Proteinvariante in einer Maus Testiszelllinie sowie in Mausgonaden detektiert wurde (Bruening & J Pelletier 1996), wohingegen die am ACG

initiierende Proteinvariante in einer humanen Monozyten-Zelllinie nachzuweisen war. Möglicherweise wird die Translation dieser Proteine gewebe- oder artspezifisch reguliert. Weiterhin konnten 14 von 20 humanen uORFs mit dem Ribosomen-Profiling bestätigt werden. Für die restlichen sechs Gene (*ADH5*, *BCL2*, *FLII*, *MTR*, *MVP*, *SLC7A1*) wurden nur die kanonischen Startstellen, jedoch nicht die beschriebenen uORFs detektiert. Dieses Ergebnis ist möglicherweise mit der Verwendung einer anderen Zelllinie zu begründen, die in den vorangegangenen Studien analysiert wurde, denn die Aktivität der uORFs von *MTR* wurde in einer nicht humanen Primatenzelllinie (Col et al. 2007), von *SLC7A1* in einer Rattenzelllinie (Fernandez et al. 2002), von *MVP* in anderen humanen Zelllinien (Holzmann et al. 2001) und von *FLII* in einer Mauszelllinie analysiert (Sarrazin et al. 2000). Für *FLII* wurde hinzukommend in dieser Arbeit ein weiterer *upstream* gelegener uORF vorhergesagt, dessen Translation möglicherweise auch zu einer Inhibition der beschriebenen uORFs führt.

Weitere Publikationen dokumentieren, dass die Aktivität der uORFs von *BCL2* und *ADH5* und deren regulatorischer Einfluss auf die Translation der CDS ebenfalls gewebe- und zelltypspezifisch reguliert wird (Harigai et al. 1996, Kwon et al. 2001).

4.5 Konservierung der TIC-Sequenzen

In dieser Arbeit wurde gezeigt, dass sowohl die AUG- als auch die non-AUG-Startcodonen der neu ermittelten TISs zwischen Primaten signifikant konserviert sind. Zusätzlich zeigte ein Vergleich der hier identifizierten TICs mit TICs, die zuvor in der Maus bestimmt wurden (Ingolia et al. 2011), dass ein Großteil der alternativen TISs auch zwischen Menschen und Nagetieren konserviert sind. Sequenzen, die während der Evolution konserviert werden, haben eine wichtige physiologische Bedeutung und sind von biologischer Relevanz. Es ist anzunehmen, dass TISs, die zwischen Mensch und Maus konserviert sind, für alle Säugetiere sehr wichtig sind.

Für die verschiedenen TICs und TIS-Kategorien wurde ein unterschiedlicher Konservierungsgrad festgestellt. Während z.B. AUG- und AUA-Codonen bei uORFs und ovuORFs und AAG-Codonen bei N-terminalen Proteinextensionen sehr hoch konserviert sind, sind AGG-, CUG- und GUG-TICs in keiner der Kategorien konserviert.

Weiterhin sehr interessant ist die einheitliche Verwendung von non-AUG-Codonen innerhalb der verschiedenen TIS-Kategorien, wohingegen die Häufigkeit der

Initiationsereignisse an AUG-Codonen bei den verschiedenen TIS-Kategorien sehr unterschiedlich ist. Vor allem für N-terminale Proteinextensionen wurde im Vergleich zu uORFs und ovuORFs sehr selten ein AUG als Initiationsstart detektiert. Die Identifizierung von nur wenigen N-terminalen Proteinextensionen, die an einem AUG-Codon initiieren, beruht möglicherweise auf den in den letzten Jahren verwendeten *in silico* Methoden, mit denen proteinkodierende Transkriptbereiche bzw. TISs hauptsächlich vorhergesagt und annotiert wurden. Diese Methoden basieren auf relativ einfachen bioinformatischen Regeln, nach denen häufig das am weitesten 5'-gelegene AUG-Codon in einem guten Sequenzkontext und mit einem langen nachfolgenden offenen Leserahmen, schon als Translationsstart definiert wurde.

Zusätzlich konnte nachgewiesen werden, dass die in humanen Transkripten detektierten uORFs in 57% der Fälle auch in den Transkripten der Maus vorhanden waren. Dieses Ergebnis wird unterstützt durch frühere Studien, die darauf hinweisen, dass funktionelle, an AUG-Codonen initiiierende uORFs zwischen verschiedenen Spezies konserviert sind (Göpfert et al. 2003, Zimmer et al. 1994).

Zusammenfassend verdeutlicht die in dieser Arbeit belegte Konservierung der TICs, die Wichtigkeit und biologische Bedeutung von alternativen Translationsereignissen für die Regulation der Genexpression.

4.6 Biologische Bedeutung von alternativen TISs

In aktuellen Publikationen konnte gezeigt werden, dass die Translation sehr oft an alternativen AUG- und *near-cognate* Codonen initiiert (Ingolia et al. 2009, Ingolia et al. 2011, Fritsch et al. 2012, Lee et al. 2012). Eine bedeutende Konsequenz der alternativen Translationsinitiation ist vor allem die vergrößerte Komplexität des Proteoms (Lee et al. 2012). Durch die Initiation an mehreren Startcodonen im gleichen Leserahmen einer mRNA entstehen Proteinisoformen, die sich an ihrem N-terminalen Ende unterscheiden. N-terminale Extensionen dienen dabei oft als Lokalisierungssignale, welche den Transport der Proteinvarianten in unterschiedliche zelluläre Kompartimente regulieren (Porras et al. 2006). Ein Beispiel hierfür ist *Bag1*, dessen verkürzte am AUG initiiierende Variante im Zytoplasma und dessen N-terminal verlängerte, an einem CUG-Codon initiiierende Isoform im Kern lokalisiert ist (Packham et al. 1997).

Darüber hinaus können die Proteinvarianten aber auch komplett andere oder antagonistische Funktionen aufweisen. Durch die Initiation an zwei verschiedenen AUG-Codonen im gleichen Leserahmen auf einer mRNA wird z.B. ein Transkriptionsaktivator LAP und ein Transkriptionsinhibitor LIP translatiert (Descombes & Schibler 1991).

Des Weiteren haben die alternativen Startstellen häufig regulatorische Funktionen. Durch die Translation von uORFs kann die Translation der kodierenden Hauptsequenz kontrolliert werden (Calvo et al. 2009) als Antwort auf globale (Sonenberg & Hinnebusch 2009) oder genspezifische regulatorische Signale (Medenbach et al. 2011).

Eines der am besten untersuchten Gene mit mehreren regulatorischen uORFs ist *Gcn4*. *Gcn4* kodiert für einen Transkriptionsfaktor, der die Expression von ca. 50 Genen für die Aminosäurebiosynthese in Hefe aktiviert. Bei Aminosäuremangel wird die globale Proteinsynthese inhibiert, wohingegen die Translation von *Gcn4* verstärkt wird. Die verstärkte Translation wird durch die Interaktion und kombinierte Aktivität von vier verschiedenen uORFs in der 5'-UTR der mRNA vermittelt (Hinnebusch 1997).

Aufgrund der Häufigkeit von alternativen TISs scheint ihre Relevanz für die Regulation der Genexpression von größerer Wichtigkeit zu sein als bisher angenommen wurde.

4.7 Perspektive

Die Erstellung umfassender genomweiter Karten von TISs und den entsprechenden ORFs mit Hilfe des Ribosomen-Profiling bietet eine hervorragende Möglichkeit das Proteom von verschiedenen Organismen zu definieren und die Synthese zu quantifizieren (Ingolia et al. 2011). Das Aufzeigen von neuen TISs ist aber erst der Anfang bei der Erforschung der Relevanz von translationalen Kontrollmechanismen für die Genexpression (Lee et al. 2012). Es sind noch weitere Untersuchungen notwendig, um die Funktionen und die regulatorischen Mechanismen der neu detektierten uORFs und N-terminalen Proteinextensionen aufzuklären.

In dieser Arbeit konnten interessante Gene identifiziert werden, die multiple TISs in der 5'-UTR besitzen und somit vermutlich unter einer starken translationalen Kontrolle stehen. Die intensivere Untersuchung dieser Kandidaten lässt erwarten, dass die darunter liegenden Mechanismen besser verstanden werden.

Des Weiteren können die hier generierten Daten verwendet werden, um bereits annotierte TISs und Leserahmen zu korrigieren und/oder zu ergänzen, denn anhand von RPF-Sequenzen, die z.B. in Introns kartiert sind, können alternative Exons identifiziert werden. Darüber hinaus kann mit Hilfe des Ribosomen-Profiling die TIS-Verwendung bei verschiedenen Geweben oder Zellen analysiert werden, wodurch spezifische Unterschiede der translationalen Kontrolle der Genexpression bestimmbar sind.

Nicht zuletzt ist eine Analyse der Umprogrammierung der Translation während der Entwicklung oder der Entstehung von Krankheiten und Krebs möglich. Ein relevantes Beispiel hierfür sind somatische Mutationen, die zur Bildung oder zur Zerstörung von regulatorischen uORFs führen und dadurch zu einer veränderten Verwendung von TISs beitragen.

Weiterhin ist zu klären, wie sich die Initiationsmechanismen an AUG- und non-AUG-Startcodonen unterscheiden und welche Initiationsfaktoren die Wahl der TISs regulieren. Hier stellt sich außerdem die Frage, ob es Gruppen von Translationsstarts gibt, die gemeinsam auf globale oder genspezifische regulatorische Signale aktiviert oder inhibiert werden.

Hinzukommend ist interessant, ob translatierte uORFs möglicherweise als Peptide in den Zellen vorliegen. Diese mögliche neue Gruppe von Peptiden könnte wichtige und noch unbekannt biologische Funktionen haben.

5. Literaturverzeichnis

- Allen, D.W. & Zamecnik, P.C., 1962. The effect of puromycin on rabbit reticulocyte ribosomes. *Biochim Biophys Acta*, 55, pp.865–874.
- de Almeida, R.A. et al., 2006. Control of MYEOV protein synthesis by upstream open reading frames. *J Biol Chem*, 281(2), pp.695–704.
- Arnaud, E. et al., 1999. A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. *Mol Cell Biol*, 19(1), pp.505–514
- Ashida, H. et al., 2005. Mammalian PIG-X and yeast Pbn1p are the essential components of glycosylphosphatidylinositol-mannosyltransferase I. *Mol Biol Cell*, 16(3), pp.1439–1448.
- Balagopal, V., Fluch, L. & Nissan, T., 2012. Ways and means of eukaryotic mRNA decay. *Biochim Biophys Acta*, 1819(6), pp.593–603.
- Bastide, A. et al., 2008. An upstream open reading frame within an IRES controls expression of a specific VEGF-A isoform. *Nucleic Acids Res*, 36(7), pp.2434–2445.
- Bazykin, G.A. & Kochetov, A.V., 2011. Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic Acids Res*, 39(2), pp.567–577.
- Beucken van den, T. et al., 2007. Regulation of Cited2 expression provides a functional link between translational and transcriptional responses during hypoxia. *Radiother Oncol*, 83(3), pp.346–352.
- Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 72, pp.291–336.
- Brown, C.Y. et al., 1999. Role of two upstream open reading frames in the translational control of oncogene mdm2. *Oncogene*, 18(41), pp.5631–5637.
- Bruening, W. & Pelletier, J., 1996. A non-AUG translational initiation event generates novel WT1 isoforms. *J Biol Chem*, 271(15), pp.8646–8654.
- Buchan, J.R. & Parker, R., 2009. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell*, 36(6), pp.9329–41.
- Calkhoven, C.F. et al., 2003. Translational control of SCL-isoform expression in hematopoietic lineage choice. *Genes Dev*, 17(8), pp.959–964.
- Calkhoven, C.F., Müller, C. & Leutz, A., 2002. Translational control of gene expression and disease. *Trends Mol Med*, 8(12), pp.577–83..

- Calvo, S.E., Pagliarini, D.J. & Mootha, V.K., 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci*, 106(18), pp.7507–7512.
- Cheng, H. et al., 2006. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, 127(7), pp.1389–1400.
- Cheung, Y.-N. et al., 2007. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes Dev*, 21(10), pp.1217–1230.
- Child, S.J., Miller, M.K. & Geballe, A P, 1999. Translational control by an upstream open reading frame in the HER-2/neu transcript. *J Biol Chem*, 274(34), pp.24335–24341.
- Cigan, A.M., Feng, L. & Donahue, T.F., 1988. tRNAⁱ(met) functions in directing the scanning ribosome to the start site of translation. *Science*, 242(4875), pp.93–7.
- Col, B., Oltean, S. & Banerjee, R., 2007. Translational regulation of human methionine synthase by upstream open reading frames. *Biochim Biophys Acta*, 1769(9-10), pp.532–540.
- Crooks, G.E. et al., 2004. WebLogo: a sequence logo generator. *Genome Res*, 14(6), pp.1188–1190.
- Das, S., Ghosh, R. & Maitra, U., 2001. Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J Biol Chem*, 276(9), pp.6720–6726.
- Descombes, P. & Schibler, U., 1991. A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell*, 67(3), pp.569–579.
- Diba, F., Watson, C.S. & Gametchu, B., 2001. 5'UTR sequences of the glucocorticoid receptor 1A transcript encode a peptide associated with translational regulation of the glucocorticoid receptor. *J Cell Biochem*, 81(1), pp.149–161.
- Dodds, D.C. et al., 1997. Neuronal pentraxin receptor, a novel putative integral membrane pentraxin that interacts with neuronal pentraxin 1 and 2 and taipoxin-associated calcium-binding protein 49. *J Biol Chem*, 272(34), pp.21488–21494.
- Dolstra, H. et al., 1999. A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. *J Exp Med*, 189(2), pp.301–308.
- Fekete, C.A. et al., 2007. N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. *EMBO J*, 26(6), pp.1602–1614.
- Fernandez, J. et al., 2002. Regulation of internal ribosome entry site-mediated translation by eukaryotic initiation factor-2 α phosphorylation and translation of a small upstream open reading frame. *J Biol Chem*, 277(3), pp.2050–2058.
- Fitzgerald, K.D. & Semler, B.L., 2009. Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochim Biophys Acta*, 1789(949), pp.518–528.

- Flick, K. & Kaiser, P., 2012. Protein degradation and the stress response. *Semin Cell Dev Biol*, 23(5), pp.515–22.
- Fresno, M., Jiménez, A. & Vázquez, D., 1977. Inhibition of translation in eukaryotic systems by harringtonine. *Eur J Biochem*, 72(2), pp.323–330.
- Fritsch, C. et al., 2012. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res*, 22(11), pp.2208–2218.
- Gerashchenko, M.V., Su, D. & Gladyshev, V.N., 2010. CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes. *J Biol Chem*, 285(7), pp.4595–4602.
- Goossen, B. et al., 1990. Translational repression by a complex between the iron-responsive element of ferritin mRNA and its specific cytoplasmic binding protein is position-dependent in vivo. *EMBO J*, 9(12), pp.4127–4133.
- Griffin, E. et al., 2001. A link between diabetes and atherosclerosis: Glucose regulates expression of CD36 at the level of translation. *Nat Med*, 7(7), pp.840–846.
- Guo, H. et al., 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308), pp.835–840.
- Guyot, B. et al., 2002. Upstream open reading frames regulate translation of Mona/Gads adapter mRNA in the megakaryocytic lineage. *Platelets*, 13(8), pp.459–464.
- Göpfert, U., Kullmann, M. & Hengst, L., 2003. Cell cycle-dependent translation of p27 involves a responsive element in its 5'-UTR that overlaps with a uORF. *Hum Mol Genet*, 12(14), pp.1767–1779.
- Hann, S.R. et al., 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell*, 52(2), pp.185–195.
- Harding, H P et al., 2000. Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell*, 6(5), pp.1099–1108.
- Harding, Heather P et al., 2003. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell*, 11(3), pp.619–633.
- Harigai, M. et al., 1996. A cis-acting element in the BCL-2 gene controls expression through translational mechanisms. *Oncogene*, 12(6), pp.1369–1374.
- Hay, N. & Sonenberg, Nahum, 2004. Upstream and downstream of mTOR. *Genes Dev*, 18(16), pp.1926–1945.
- Hernandez, E.M. et al., 2002. AUA as a translation initiation site in vitro for the human transcription factor Sp3. *J Biochem Mol Biol*, 35(3), pp.273–282.
- Hershey, J.W., 1991. Translational control in mammalian cells. *Annu Rev Biochem*, 60, pp.717–755.

- Hinnebusch, A G, 1997. Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *J Biol Chem*, 272(35), pp.21661–21664.
- Hinnebusch, A. G., 2011. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol Mol Biol Rev*, 75(3), pp.434–467.
- Hinnebusch, Alan G & Lorsch, J.R., 2012. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol*, 4(10).
- Hocine, S., Singer, R.H. & Grünwald, D., 2010. RNA processing and export. *Cold Spring Harb Perspect Biol*, 2(12), p.a000752.
- Holzmann, K. et al., 2001. A small upstream open reading frame causes inhibition of human major vault protein expression from a ubiquitous mRNA splice variant. *FEBS Lett*, 494(1-2), pp.99–104.
- Huez, I. et al., 2001. New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. *Mol Endocrinol*, 15(12), pp.2197–2210.
- Hurtaud, C. et al., 2006. Translation control of UCP2 synthesis by the upstream open reading frame. *Cell Mol Life Sci*, 63(15), pp.1780–1789.
- Iacono, M., Mignone, F. & Pesole, G., 2005. uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene*, 349(11), pp.97–105.
- Imataka, H., Olsen, H.S. & Sonenberg, N, 1997. A new translational regulator with homology to eukaryotic translation initiation factor 4G. *EMBO J*, 16(4), pp.817–825.
- Ingolia, N.T. et al., 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), pp.218–23.
- Ingolia, N.T. et al., 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc*, 7(8), pp.1534–50.
- Ingolia, N.T., Lareau, L.F. & Weissman, J.S., 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*, 147(4), pp.789–802.
- Ito, T., 2007. Role of histone modification in chromatin dynamics. *J Biochem*, 141(5), pp.609–614.
- Ivanov, I.P. et al., 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res*, 39(10), pp.4220–4234.
- Ivanov, I.P. et al., 2010. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc Natl Acad Sci U S A*, 107(42), pp.18056–60.

- Ivanov, I.P., Loughran, G. & Atkins, J.F., 2008. uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc Natl Acad Sci*, 105(29), pp.10079–10084.
- Jackson, R.J., Hellen, C.U.T. & Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, 11(2), pp.113–27.
- Jacquemin, P., Martial, J.A. & Davidson, I., 1997. Human TEF-5 is preferentially expressed in placenta and binds to multiple functional elements of the human chorionic somatomammotropin-B gene enhancer. *J Biol Chem*, 272(20), pp.12928–12937.
- Jiang, H.-Y., Jiang, L. & Wek, R.C., 2007. The eukaryotic initiation factor-2 kinase pathway facilitates differential GADD45a expression in response to environmental stress. *J Biol Chem*, 282(6), pp.3755–3765.
- Jin, X. et al., 2003. The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. *J Biol Chem*, 278(28), pp.25716–25721.
- Jousse, C. et al., 2001. Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR. *Nucleic Acids Res*, 29(21), pp.4341–4351.
- Keene, J.D. & Tenenbaum, S. a, 2002. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell*, 9(6), pp.1161–1167.
- Kochetov, A.V. et al., 2007. AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics*, 8, p.318.
- Kos, M. et al., 2002. Upstream open reading frames regulate the translation of the multiple mRNA variants of the estrogen receptor alpha. *J Biol Chem*, 277(40), pp.37131–37138.
- Kozak, M., 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*, 15(20), pp.8125–48.
- Kozak, M., 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol*, 115(4), pp.887–903.
- Kozak, M., 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol*, 9(11), pp.5073–5080.
- Kozak, M., 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*, 87(21), pp.8301–5.
- Kozak, M., 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*, 15(4), pp.1109–1123.
- Kozak, Marilyn, 1980. Evaluation of the “scanning model” for initiation of protein synthesis in eucaryotes. *Cell*, 22(1), pp.7–8.

- Kozak, Marilyn, 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2), pp.283–292.
- Kozak, Marilyn, 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, 299(1-2), pp.1–34.
- Kwon, H.S. et al., 2001. Posttranscriptional regulation of human ADH5/FDH and Myf6 gene expression by upstream AUG codons. *Arch Biochem Biophys*, 386(2), pp.163–171.
- Lee, J. et al., 2002. An upstream open reading frame impedes translation of the huntingtin gene. *Nucleic Acids Res*, 30(23), pp.5110–5119.
- Lee, S. et al., 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A*, 109, pp.2424–2432.
- Lee, Y.-Y., Cevallos, R.C. & Jan, E., 2009. An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation. *J Biol Chem*, 284(11), pp.6661–6673.
- Lehmann, J.M. et al., 1998. The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J Clin Invest*, 102(5), pp.1016–1023.
- Leissring, M.A. et al., 2004. Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria. *Biochem J*, 383(Pt. 3), pp.439–446.
- Lincoln, A.J. et al., 1998. Inhibition of CCAAT/enhancer-binding protein alpha and beta translation by upstream open reading frames. *J Biol Chem*, 273(16), pp.9552–9560.
- Liu, L. et al., 1999. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat Genet*, 21(1), pp.128–32.
- Lock, P. et al., 1991. Two isoforms of murine hek, generated by utilization of alternative translational initiation codons, exhibit different patterns of subcellular localization. *Mol Cell Biol*, 11(9), pp.4363–4370.
- Lodhi, K.M., Ozdener, M.H. & Shayiq, R.M., 2003. The upstream open reading frame mediates constitutive effects on translation of cytochrome p-450c27 from the seventh in-frame AUG codon in rat liver. *J Biol Chem*, 278(42), pp.40647–40657.
- Lomakin, I.B. et al., 2003. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev*, 17(22), pp.2786–2797.
- Lorsch, J.R. & Dever, T.E., 2010. Molecular view of 43 S complex formation and start site selection in eukaryotic translation initiation. *J Biol Chem*, 285(28), pp.21203–21207.
- Loughran, G. et al., 2011. Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res*.

- Lutz, C.S. & Moreira, A., 2010. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA*, 2(1), pp.22–31.
- Lux, H. et al., 2010. Genetic and molecular analyses of PEG10 reveal new aspects of genomic organization, transcription and translation. *PLoS One*, 5(1), p.e8686.
- Ma, X.M. & Blenis, J., 2009. Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol*, 10(5), pp.307–18.
- Maag, D. et al., 2005. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell*, 17(2), pp.265–275.
- Mangus, D.A., Evans, M.C. & Jacobson, A., 2003. Protein family review Poly (A) - binding proteins : multifunctional scaffolds for the post- transcriptional control of gene expression. *Genome Biol*, 4(7), pp.1–14.
- Martin, M.M. et al., 2006. TGF-beta(1) regulation of human AT(1) receptor mRNA splice variants harboring exon 2. *Mol Cell Endocrinol*, 249(1-2), pp.21–31.
- Maston, G. a, Evans, S.K. & Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7, pp.29–59.
- Matsui, M. et al., 2007. Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse. *FEBS Lett*, 581(22), pp.4184–4188.
- McDade, S.S., Hall, P.A. & Russell, S.E.H., 2007. Translational control of SEPT9 isoforms is perturbed in disease. *Hum Mol Genet*, 16(7), pp.742–752.
- Medenbach, J., Seiler, M. & Hentze, Matthias W, 2011. Translational Control via Protein-Regulated Upstream Open Reading Frames. *Cell*, 145(6), pp.902–913.
- Mehta, A., Trotta, C.R. & Peltz, S.W., 2006. Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes Dev*, 20(8), pp.939–953.
- Meiron, M. et al., 2001. New isoforms of VEGF are translated from alternative initiation CUG codons located in its 5'UTR. *Biochem Biophys Res Commun*, 282(4), pp.1053–1060.
- Mignone, F. et al., 2002. Untranslated regions of mRNAs. *Genome Biol*, 3(3), p.REVIEWS0004.
- Mize, G J et al., 1998. The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J Biol Chem*, 273(49), pp.32500–32505.
- Morris, D R & Geballe, A P, 2000. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol*, 20(23), pp.8635–8642.

- Muller, E.A. & Danner, D.J., 2004. Tissue-specific translation of murine branched-chain alpha-ketoacid dehydrogenase kinase mRNA is dependent upon an upstream open reading frame in the 5'-untranslated region. *J Biol Chem*, 279(43), pp.44645–44655.
- Nanda, J.S. et al., 2009. eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *J Mol Biol*, 394(2), pp.268–285.
- Nathans, D., 1964. Puromycin Inhibition of Protein Synthesis: Incorporation of Puromycin into Peptide Chains. *Proc Natl Acad Sci*, 51, pp.585–592.
- Nomura, A. et al., 2001. Involvement of upstream open reading frames in regulation of rat V(1b) vasopressin receptor expression. *Am J Physiol Endocrinol Metab*, 280(5), pp.E780–787.
- Packham, G., Brimmell, M. & Cleveland, J.L., 1997. Mammalian cells express two differently localized Bag-1 isoforms generated by alternative translation initiation. *Biochem J*, 328 (Pt 3, pp.807–813.
- Park, E.-H., Lee, J.M. & Pelletier, Jerry, 2006. The Tie2 5' untranslated region is inhibitory to 5' end-mediated translation initiation. *FEBS Lett*, 580(5), pp.1309–1319.
- Passmore, L.A. et al., 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell*, 26(1), pp.41–50.
- Peabody, D.S., 1989. Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem*, 264(9), pp.5031–5035.
- Pegg, A.E., 2006. Regulation of ornithine decarboxylase. *J Biol Chem*, 281(21), pp.14529–14532.
- Pendleton, L.C. et al., 2005. Regulation of endothelial argininosuccinate synthase expression and NO production by an upstream open reading frame. *J Biol Chem*, 280(25), pp.24252–24260.
- Pentecost, B.T. et al., 2005. Upstream regions of the estrogen receptor alpha proximal promoter transcript regulate ER protein expression through a translational mechanism. *Mol Cell Endocrinol*, 229(1-2), pp.83–94.
- Pestova, T.V. & Kolupaeva, V.G., 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev*, 16(22), pp.2906–2922.
- Porras, P. et al., 2006. One single in-frame AUG codon is responsible for a diversity of subcellular localizations of glutaredoxin 2 in *Saccharomyces cerevisiae*. *J Biol Chem*, 281(24), pp.16551–16562.
- Prats, H et al., 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. *Proc Natl Acad Sci*, 86(6), pp.1836–1840.
- Pópulo, H., Lopes, J.M. & Soares, P., 2012. The mTOR Signalling Pathway in Human Cancer. *Int J Mol Sci*, 13(2), pp.1886–1918.

- Rabadan-Diehl, C. et al., 2007. Inhibition of vasopressin V1b receptor translation by upstream open reading frames in the 5'-untranslated region. *J Neuroendocrinol*, 19(4), pp.309–319.
- Reynolds, K., Zimmer, A.M. & Zimmer, A., 1996. Regulation of RAR beta 2 mRNA expression: evidence for an inhibitory peptide encoded in the 5'-untranslated region. *J Cell Biol*, 134(4), pp.827–835.
- Robert, F. et al., 2009. Altering chemosensitivity by modulating translation elongation. *PLoS One*, 4(5), p.e5428.
- Rogers, G.W., Edelman, G.M. & Mauro, V.P., 2004. Differential utilization of upstream AUGs in the beta-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc Natl Acad Sci*, 101(9), pp.2794–2799.
- Rohwedel, J. et al., 2003. Evidence for posttranscriptional regulation of the multi K homology domain protein vigilin by a small peptide encoded in the 5' leader sequence. *Cell Mol Life Sci*, 60(8), pp.1705–1715.
- Rosenstiel, P. et al., 2007. Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. *BMC Genomics*, 8(472).
- Sapetschnig, A. et al., 2004. Complexity of translationally controlled transcription factor Sp3 isoform expression. *J Biol Chem*, 279(40), pp.42095–42105.
- Sarrazin, S. et al., 2000. Negative and translation termination-dependent positive control of FLI-1 protein synthesis by conserved overlapping 5' upstream open reading frames in Fli-1 mRNA. *Mol Cell Biol*, 20(9), pp.2959–2969.
- Schlüter, G., Boinska, D. & Nieman-Seyde, S.C., 2000. Evidence for translational repression of the SOCS-1 major open reading frame by an upstream open reading frame. *Biochem Biophys Res Commun*, 268(2), pp.255–261.
- Schneider-Poetsch, T. et al., 2010. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol*, 6(3), pp.209–217.
- Shaughnessy, J.D. et al., 1999. Mrvi1, a common MRV integration site in BXH2 myeloid leukemias, encodes a protein with homology to a lymphoid-restricted membrane protein Jaw1. *Oncogene*, 18(12), pp.2069–2084.
- Sonenberg, N & Hinnebusch, Alan G, 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4), pp.731–745.
- Song, K.Y. et al., 2007. Translational repression of mouse mu opioid receptor expression via leaky scanning. *Nucleic Acids Res*, 35(5), pp.1501–1513.
- Souciet, G. et al., 1999. Characterization of two bifunctional *Arabidopsis thaliana* genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. *Eur J Biochem*, 266(3), pp.848–854.

- Spevak, C.C. et al., 2006. her-2 upstream open reading frame effects on the use of downstream initiation codons. *Biochem Biophys Res Commun*, 350(4), pp.834–841.
- Spotts, G.D. et al., 1997. Identification of downstream-initiated c-Myc proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. *Mol Cell Biol*, 17(3), pp.1459–68.
- Spriggs, K. a, Bushell, M. & Willis, A.E., 2010. Translational regulation of gene expression during conditions of cell stress. *Mol Cell*, 40(2), pp.228–237.
- Steel, L.F. et al., 1996. Elements in the murine c-mos messenger RNA 5'-untranslated region repress translation of downstream coding sequences. *Cell Death Differ*, 7(10), pp.1415–1424.
- Stewart, A.F. et al., 1996. Cloning of human RTEF-1, a transcriptional enhancer factor-1-related gene preferentially expressed in skeletal muscle: evidence for an ancient multigene family. *Genomics*, 37(1), pp.68–76.
- Stockklausner, C. et al., 2006. The uORF-containing thrombopoietin mRNA escapes nonsense-mediated decay (NMD). *Nucleic Acids Res*, 34(8), pp.2355–2363.
- Taira, M. et al., 1990. A human testis-specific mRNA for phosphoribosylpyrophosphate synthetase that initiates from a non-AUG codon. *J Biol Chem*, 265(27), pp.16491–16497.
- Takeda, M. et al., 2004. A unique role of an amino terminal 16-residue region of long-type GATA-6. *J Biochem*, 135(5), pp.639–650.
- Tanaka, H. et al., 2003. Characterization of a family of endogenous neuropeptide ligands for the G protein-coupled receptors GPR7 and GPR8. *Proc Natl Acad Sci*, 100(10), pp.6251–6256.
- Tanaka, M., Ito, S. & Kiuchi, K., 2001. The 5'-untranslated region of the mouse glial cell line-derived neurotrophic factor gene regulates expression at both the transcriptional and translational levels. *Brain Res Mol Brain Res*, 91(1-2), pp.81–95.
- Tang, T. et al., 2008. Upstream open reading frames regulate the expression of the nuclear Wnt13 isoforms. *Biochem Biophys Res Commun*, 366(4), pp.1081–1088.
- Tee, M.K. & Jaffe, R.B., 2001. A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *Biochem J*, 359(Pt 1), pp.219–226.
- Thomas, M.G. et al., 2011. RNA granules: the good, the bad and the ugly. *Cell Signal*, 23(2), pp.324–34.
- Thébault, S. et al., 2000. Molecular cloning of a novel human I-mfa domain-containing protein that differently regulates human T-cell leukemia virus type I and HIV-1 expression. *J Biol Chem*, 275(7), pp.4848–4857.

- Tikole, S. & Sankararamakrishnan, R., 2006. A survey of mRNA sequences with a non-AUG start codon in RefSeq database. *J Biomol Struct Dyn*, 24(1), pp.33–42.
- Uhlmann-Schiffler, H., Rössler, O.G. & Stahl, H., 2002. The mRNA of DEAD box protein p72 is alternatively translated into an 82-kDa RNA helicase. *J Biol Chem*, 277(2), pp.1066–1075.
- Valasek, L.S., 2012. “Ribozoomin” - Translation Initiation from the Perspective of the ribosome-bound Eukaryotic Initiation Factors (eIFs). *Curr Protein Pept Sci*, 13(4), pp.305–330.
- Vattem, K.M. & Wek, R.C., 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A*, 101(31), pp.11269–74.
- Volkova, O.A. & Kochetov, A.V., 2010. Interrelations between the nucleotide context of human start AUG codon, N-end amino acids of the encoded protein and initiation of translation. *J Biomol Struct Dyn*, 27(5), pp.611–618.
- Wang, X. et al., 2009. Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. *Stem Cells*, 27(6), pp.1265–1275.
- Warnakulasuriyarachchi, D., Ungureanu, N.H. & Holcík, M., 2003. The translation of an antiapoptotic protein HIAP2 is regulated by an upstream open reading frame. *Cell Death Differ*, 10(8), pp.899–904.
- Watatani, Y. et al., 2008. Stress-induced translation of ATF5 mRNA is regulated by the 5'-untranslated region. *J Biol Chem*, 283(5), pp.2543–2553.
- Wen, Y. et al., 2009. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat Genet*, 41(2), pp.228–33.
- Wethmar, K., Smink, J.J. & Leutz, A., 2010. Upstream open reading frames: molecular switches in (patho)physiology. *Bioessays*, 32(10), pp.885–893.
- Wiestner, A. et al., 1998. An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. *Nat Genet*, 18(1), pp.49–52.
- Williams, R.T. et al., 2001. Identification and characterization of the STIM (stromal interaction molecule) gene family: coding for a novel class of transmembrane proteins. *Biochem J*, 357(Pt 3), pp.673–685.
- Wu, H., Ross, C.R. & Blecha, F., 2002. Characterization of an upstream open reading frame in the 5' untranslated region of PR-39, a cathelicidin antimicrobial peptide. *Mol Immunol*, 39(1-2), pp.9–18.
- Xiao, J.H. et al., 1991. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell*, 65(4), pp.551–568.

-
- Xiong, Z. et al., 2006. An unconventional antigen translated by a novel internal ribosome entry site elicits antitumor humoral immune reactions. *J Immunol*, 177(7), pp.4907–4916.
- Yu, Y. et al., 2009. Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res*, 37(15), pp.5167–82.
- Zhang, M.Q., 2002. Computational prediction of eukaryotic protein-coding genes. *Nat Rev Genet*, 3(9), pp.698–709.
- Zhang, Yuanyuan, Li, W. & Vore, M., 2007. Translational regulation of rat multidrug resistance-associated protein 2 expression is mediated by upstream open reading frames in the 5' untranslated region. *Mol Pharmacol*, 71(1), pp.377–383.
- Zhou, W. & Song, W., 2006. Leaky Scanning and Reinitiation Regulate BACE1 Gene Expression. *Mol Cell Biol*, 26(9), pp.3353–3364.
- Zimmer, A., Zimmer, A.M. & Reynolds, K., 1994. Tissue specific expression of the retinoic acid receptor-beta 2: regulation by short open reading frames in the 5'-noncoding region. *J Cell Biol*, 127(4), pp.1111–1119.
- Zysow, B.R. et al., 1995. C/T polymorphism in the 5' untranslated region of the apolipoprotein(a) gene introduces an upstream ATG and reduces in vitro translation. *Arterioscler Thromb Vasc Biol*, 15(1), pp.58–64.

6. Anhang

6.1 Oligonukleotide

Tabelle 14: Verwendete Oligonukleotide

Bezeichnung	Sequenz (5'-3')	Verwendung	Info
oNTI199	AUG UAC ACG GAG UCG ACC CGC AAC GCG A	RNA-Größenstandard	
RNA_ 27nt	AUG UAC ACG GAG UCG ACC CGC AAC GCG	RNA-Größenstandard	basiert auf oNTI199
RNA_ 33nt	AUG UAC ACG GAG UCG ACC CGC AAC GCG AAG CGC	RNA-Größenstandard	basiert auf oNTI199
oNTI223	p-GAT CGT CGG ACT GTA GAA CTC TØC AAG CAG AAG ACG GCA TAC GAT TTT TTT TTT TTT TTT TTT TVN	RT-oligo-d(T)-Primer	5'-Phosphat, dSpacer Furan
oNTI200	CAA GCA GAA GAC GGC ATA	RT-PCR	
oNTI201	AAT GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GAC G	RT-PCR	
oNTI202	CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC	Sequenzierung	
RTF1	p-AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT ØGT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTT TTT TTT TTT TTT TTT VN	RT-oligo-d(T)-Primer	5'-Phosphat, dSpacer Furan
RTF2	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T	RT-PCR	
RPI1	CAA GCA GAA GAC GGC ATA CGA GAT CGC ACT AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
RPI2	CAA GCA GAA GAC GGC ATA CGA GAT CTG TAG CGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
RPI3	CAA GCA GAA GAC GGC ATA CGA GAT CCG GAT TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
RPI4	CAA GCA GAA GAC GGC ATA CGA GAT CAC CAG TGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
RPI5	CAA GCA GAA GAC GGC ATA CGA GAT CGT GAC AGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
RPI6	CAA GCA GAA GAC GGC ATA CGA GAT CTA ACC GGT GAC TGG AGT TCA GAC GTG TGC TCT TCC	RT-PCR	
18S_rRNA_144	TAG CTC TAG AAT TAC CAC AGT TAT CCA	rRNA-Hybridisierung	5'-Biotin
18S_rRNA_101	CGA CCA AAG GAA CCA TAA CTG ATT TA	rRNA-Hybridisierung	5'-Biotin
18S_rRNA_284	TCG GCC CGA GGT TAT CTA GAG TCA CCA AAG	rRNA-Hybridisierung	5'-Biotin

6.2 Analyse der RPFs

Tabelle 15: Analyse der sequenzierten Fragmente

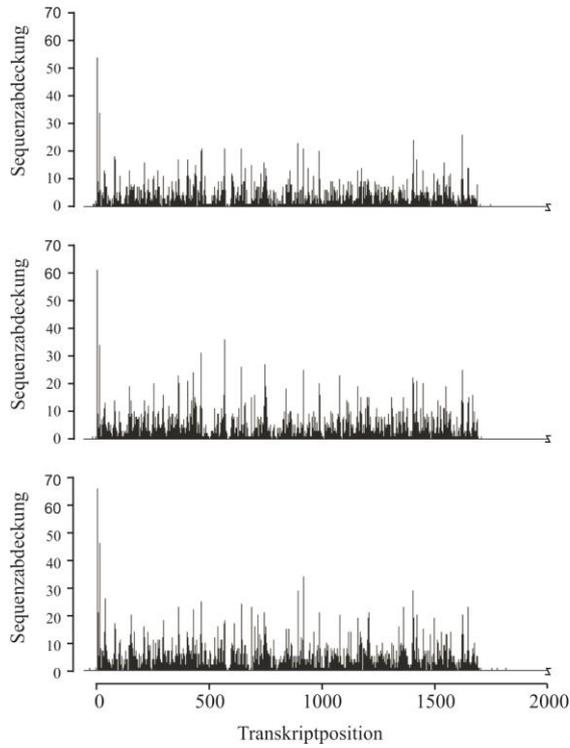
Reads	Cycloheximid I	Cycloheximid II	Cycloheximid III	Cycloheximid Summe I-III
gesamt	96.712.346	101.960.552	97.677.373	296.350.271
> 20	91.866.869	97.725.855	95.206.779	284.799.503
Adapter 1	35074	18082	25753	55731
Adapter 2	746	4723	5097	10566
oNTI223	40076	20815	29155	90046
oNTI199	4074	31256	25465	60795
oNTI200	16	4	5	25
oNTI202	0	0	0	0
∑ Adapter / Primer	79.986	74.880	85.475	240.341
NR_003286	59.560.754	6218065	58425434	180.171.253
NR_003287	9402720	8948760	9304840	27.656.320
NR_003285	50887	62287	79960	193.134
NR_003371	51537	62663	77086	191.286
NR_023363	55824	65300	80302	201.426
∑ rRNA	69.121.722	71.324.075	67.967.622	208.413.419
∑ Adapter / Primer / rRNA	69.201.708	71.398.955	68.053.097	208.653.760
Rest	22.665.161	26.326.900	26.080.236	75.072.297
Rest >20 / >40	22.618.145	26.295.121	26.048.154	74.961.420
eindeutig einer Referenzsequenz zugeordnet	14.762.803	18.087.352	18.556.057	51.406.212

Reads	Puromycin I	Puromycin II	Puromycin III	Puromycin Summe I-III
gesamt	88.213.584	95.946.401	99.128.594	283.288.579
> 20	84.215.702	92.748.540	93.228.013	270.192.255
Adapter 1	14967	20888	15821	51676
Adapter 2	774	275	479	1528
oNTI223	16897	23581	18351	58829
oNTI199	4861	1758	2945	9564
oNTI200	16	20	21	57
oNTI202	0	0	1	1
∑ Adapter / Primer	37.515	46.522	37.618	121.655
NR_003286	63335349	70890917	69.074.725	203.300.991
NR_003287	9898430	10425901	10948830	31.273.161
NR_003285	76915	108765	102809	288.489
NR_003371	65713	64785	54801	185.299
NR_023363	71478	67377	56945	195.800
∑ rRNA	73.447.885	81.557.745	80.238.110	235.243.740
∑ Adapter / Primer / rRNA	73.485.400	81.604.267	80.275.728	235.365.395
Rest	10.730.302	11.144.273	12.952.285	34.826.860
Rest >20 / >40	10.683.387	11.119.499	12.895.485	34.698.371
eindeutig einer Referenzsequenz zugeordnet	5.585.473	5.930.977	7.1897.73	18.706223

6.3 Reproduzierbarkeit der Ribosomen-Profilierung-Methode

TPPI

Kontrolle



Puromycin

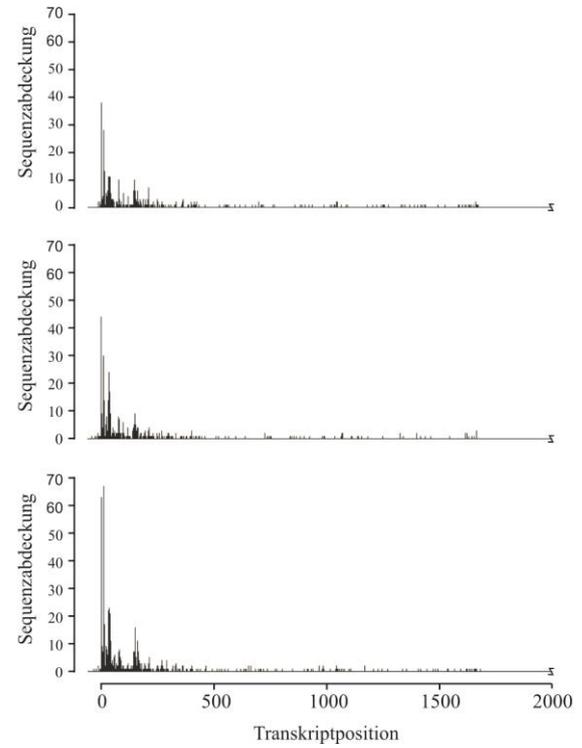


Abb. 28: Reproduzierbarkeit der Verteilung der RPFs über die Transkripte am Beispiel von *TPPI*. Die Experimente wurden im Triplikate durchgeführt.

6.4 Häufigkeit und Verteilung der TICs

Tabelle 16: TICs der identifizierten TISs

Codon	AUG	CUG	GUG	UUG	ACG	AUC	AGG	AAG	AUU	AUA	Summe
aTIS	2299	5	1	0	0	0	0	0	0	0	2305
uORF	924	1045	342	209	143	99	66	29	103	34	2994
ovuORF	116	664	215	128	103	71	28	20	47	14	1406
N-terminale Proteinextension	6	270	80	60	51	33	16	6	19	5	546
Gesamtverwendung	3345	1984	638	397	297	203	110	55	169	53	7251

6.5 Validierung der identifizierten TICs

Tabelle 17: Gene mit experimentell nachgewiesenen N-terminalen Proteinextensionen und uORFs. Bei diesen Genen wurden in zuvor durchgeführten Studien N-terminale Proteinextensionen oder uORFs nachgewiesen. Jedoch erfüllten sie nicht die für diese Studie definierten Sequenzabdeckungen oder repräsentierten nicht die hier analysierten Transkriptisoformen. Gene mit N-terminalen Proteinextensionen wurden aus der “*Database of mRNA Sequences with non-AUG Start Codons*” (<http://bioinfo.iitk.ac.in/>) und Gene mit uORFs aus der Studie von Calvo et al. (2009) bezogen.

Ref Seq ID	Gen	Klassifizierung	geringe Sequenzabdeckung	falsche Transkriptisoform	Referenz
NM_031895	<i>CACNG8</i>	non-AUG Start	ja		Ivanov et al. 2011
NM_002006	<i>FGF2</i>	non-AUG Start	ja		Arnaud et al. 1999, Prats et al. 1989
NM_021182	<i>HMHB1</i>	non-AUG Start	ja		Dolstra et al. 1999
NM_001128619	<i>LUZP6</i>	non-AUG Start	ja		Xiong et al. 2006
NM_199072	<i>MDFIC</i>	non-AUG Start	ja		Thébault et al. 2000
NM_001098579	<i>MRV11</i>	non-AUG Start	ja		Shaughnessy et al. 1999
NM_014293	<i>NPTXR</i>	non-AUG Start	ja		Dodds et al. 1997
NM_022002	<i>NR112</i>	non-AUG Start	ja		Lehmann et al. 1998
NM_001134939	<i>OAZ3</i>	non-AUG Start	ja		Ivanov et al. 2011
NM_001172438	<i>PEG10</i>	non-AUG Start	ja		Lux et al. 2010
NM_001166304	<i>PIGX</i>	non-AUG Start	ja		Ashida et al. 2005
NM_002701	<i>POUF1</i>	non-AUG Start	ja		Wang et al. 2009
NM_175886	<i>PRPS1L1</i>	non-AUG Start	ja		Taira et al. 1990
NM_001169117	<i>STIM2</i>	non-AUG Start	ja		Williams et al. 2001
NM_021961	<i>TEAD1</i>	non-AUG Start	ja		Xiao et al. 1991
NM_003214	<i>TEAD3</i>	non-AUG Start	ja		Jacquemin et al. 1997
NM_052883	<i>TXNRD3</i>	non-AUG Start	ja		Gerashchenko et al. 2010
NM_001025366	<i>VEGFA</i>	non-AUG Start	ja		Meiron et al. 2001, Tee & Jaffe 2001, Huez et al. 2001
NM_003111	<i>SP3</i>	non-AUG Start		ja	Hernandez et al. 2002
NM_000392	<i>ABCC2</i>	uORF	ja		Zhang et al. 2007
NM_031850	<i>AGTR1</i>	uORF	ja		Martin et al. 2006
NM_000707	<i>AVPR1B</i>	uORF	ja		Nomura et al. 2001, Rabadan-Diehl et al. 2007
NM_012104	<i>BACE1</i>	uORF	ja		Rogers et al. 2004
NM_001166	<i>BIRC2</i>	uORF	ja		Warnakulasuriyarachchi et al. 2003
NM_004345	<i>CAMP</i>	uORF	ja		Wu et al. 2002
NM_001001548	<i>CD36</i>	uORF	ja		Griffin et al. 2001
NM_000784	<i>CYP27A1</i>	uORF	ja		Lodhi et al. 2003
NM_004448	<i>ERBB2</i>	uORF	ja		Child, Miller, and Geballe 1999, Mehta et al. 2006, Spevak et al. 2006

NM_001122742	<i>ESR1</i>	uORF	ja		Kos et al. 2002, Pentecost et al. 2005
NM_005257	<i>GATA6</i>	uORF	ja		Takeda et al. 2004
NM_001190468	<i>GDNF</i>	uORF	ja		Tanaka et al. 2001
NM_004810	<i>GRAP2</i>	uORF	ja		Guyot et al. 2002
NM_005577	<i>LPA</i>	uORF	ja		Zysow et al. 1995
NM_005372	<i>MOS</i>	uORF	ja		Steel et al. 1996
NM_138768	<i>MYEOV</i>	uORF	ja		de Almeida et al. 2006
NM_022162	<i>NOD2</i>	uORF	ja		Rosenstiel et al. 2007
NM_001145281	<i>OPRM1</i>	uORF	ja		Song et al. 2007
NM_000965	<i>RARB</i>	uORF	ja		Reynolds et al. 1996
NM_003745	<i>SOCS1</i>	uORF	ja		Schlüter et al. 2000
NM_003189	<i>TAL1</i>	uORF	ja		Calkhoven et al. 2003
NM_000459	<i>TEK</i>	uORF	ja		Park et al. 2006
NM_000460	<i>THPO</i>	uORF	ja		Stockklausner et al. 2006
NM_001025366	<i>VEGFA</i>	uORF	ja		Bastide et al. 2008
NM_004185	<i>WNT2B</i>	uORF	ja		Tang et al. 2008
NM_000050	<i>ASS1</i>	uORF		ja	Pendleton et al. 2005
NM_004064	<i>CDKN1B</i>	uORF		ja	Göpfert et al. 2003
NM_005194	<i>CEBPB</i>	uORF		ja	Lincoln et al. 1998
NM_001018077	<i>NR3C1</i>	uORF		ja	Diba et al. 2001
NM_001113491	<i>SEPT9</i>	uORF		ja	McDade et al. 2007

6.6 Konservierung der TIC-Sequenzen zwischen Primaten

Tabelle 18: Konservierung der identifizierten TIC-Sequenzen zwischen Primaten. Extension: N-terminale Proteinextension, uORF: *Upstream-open-Reading-Frame*, ovuORF: CDS überlappender uORF, KW: Konservierungswert, Differenz: zwischen den gemittelten Konservierungswerten der TICs und Kontrollcodonen, KI: Konfidenzintervall (95%), p: Signifikanzwert, bezieht sich auf einen T-Test, wobei die gemittelten Konservierungswerte der TICs und Kontrollcodonen verglichen wurden, signifikante p-Werte ($p < 0.0016$) sind kursiv und fett gekennzeichnet, N: Anzahl analysierter TICs

Kategorie	Codon	Mittelwert KW TIC	Mittelwert KW Kontrollen	Differenz	KI	P	N TIC	N Kontroll- Codonen
CDS	ATG	0.580	0.202	0.378	0.358 - 0.397	$<10^{-10}$	2299	403
Extension	AAG	0.661	0.254	0.407	0.183 - 0.631	<i>1.32×10^{-3}</i>	6	793
Extension	ACG	0.253	0.259	-0.006	-0.063 - 0.050	0.754	51	894
Extension	AGG	0.300	0.292	0.008	-0.101 - 0.117	0.882	16	2727
Extension	ATA	0.179	0.194	-0.015	-0.162 - 0.132	0.832	5	152
Extension	ATC	0.239	0.236	0.003	-0.064 - 0.070	0.901	33	494
Extension	ATG	0.265	0.205	0.060	-0.111 - 0.230	0.473	6	184
Extension	ATT	0.272	0.222	0.051	-0.049 - 0.151	0.316	19	328
Extension	CTG	0.267	0.240	0.027	0.001 - 0.054	0.045	270	1887
Extension	GTG	0.218	0.268	-0.050	-0.095 - -0.005	0.030	80	1613
Extension	TTG	0.225	0.256	-0.031	-0.083 - 0.021	0.240	60	976
uORF	AAG	0.314	0.278	0.036	-0.047 - 0.120	0.389	29	1558
uORF	ACG	0.351	0.263	0.087	0.049 - 0.126	<i>1.03×10^{-5}</i>	143	898
uORF	AGG	0.353	0.291	0.061	0.007 - 0.116	0.028	66	2800
uORF	ATA	0.393	0.213	0.180	0.098 - 0.262	<i>3.08×10^{-5}</i>	34	308
uORF	ATC	0.323	0.235	0.088	0.041 - 0.136	<i>2.79×10^{-4}</i>	99	531
uORF	ATG	0.469	0.206	0.263	0.239 - 0.286	$<10^{-10}$	924	397
uORF	ATT	0.387	0.248	0.140	0.093 - 0.186	<i>1.17×10^{-8}</i>	103	634
uORF	CTG	0.254	0.244	0.011	-0.004 - 0.026	0.150	1045	2039
uORF	GTG	0.273	0.270	0.003	-0.022 - 0.027	0.753	342	1645
uORF	TTG	0.335	0.257	0.078	0.044 - 0.111	<i>7.95×10^{-6}</i>	209	889
ovuORF	AAG	0.287	0.268	0.019	-0.070 - 0.109	0.667	20	1472
ovuORF	ACG	0.252	0.260	-0.008	-0.047 - 0.031	0.695	103	912
ovuORF	AGG	0.328	0.287	0.041	-0.041 - 0.123	0.326	28	2304
ovuORF	ATA	0.449	0.181	0.267	0.131 - 0.404	<i>2.88×10^{-4}</i>	14	158
ovuORF	ATC	0.266	0.236	0.030	-0.022 - 0.082	0.258	71	546
ovuORF	ATG	0.573	0.208	0.365	0.317 - 0.413	$<10^{-10}$	116	338
ovuORF	ATT	0.253	0.247	0.006	-0.058 - 0.071	0.846	47	540
ovuORF	CTG	0.224	0.243	-0.019	-0.036 - -0.003	0.023	664	1988
ovuORF	GTG	0.271	0.267	0.005	-0.024 - 0.033	0.675	215	1557
ovuORF	TTG	0.309	0.258	0.051	0.011 - 0.091	0.012	128	992

6.7 Vergleich der TIC-Verwendung zwischen Mensch und Maus

Tabelle 19: Vergleich der TIC-Verwendung zwischen Mensch und Maus. Die in der Mausstudie identifizierten TISs wurden gegen das humane Genom kartiert. 1490 TICs, die sowohl beim Menschen als auch bei der Maus als TISs verwendet werden, wurden miteinander verglichen.

		Maus										
Human		ATG	CTG	GTG	TTG	AAG	AGG	ATA	ATC	ATT	ACG	andere
	ATG	1279	1	3	0	0	0	1	0	0	2	0
	CTG	1	94	8	3	0	0	0	0	0	1	0
	GTG	0	4	33	2	0	0	0	0	0	2	0
	TTG	0	2	0	11	0	0	1	1	0	0	0
	AAG	0	0	0	0	1	0	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0	0
	ATA	0	0	0	0	0	0	3	0	0	0	0
	ATC	0	0	0	0	0	0	0	5	1	0	0
	ATT	0	0	0	0	0	0	0	1	6	0	0
	ACG	0	2	1	0	0	0	0	0	0	19	0
andere	0	0	0	0	0	0	0	1	0	1	0	

Tabelle 20: Vergleich der TIC-Verwendung von uORFs zwischen Mensch und Maus. Die in der Mausstudie identifizierten TISs wurden gegen das humane Genom kartiert. 191 uORF-TICs, die sowohl beim Menschen als auch bei der Maus als TISs verwendet werden, wurden miteinander verglichen.

		Maus										
Human		ATG	CTG	GTG	TTG	AAG	AGG	ATA	ATC	ATT	ACG	andere
	ATG	115	0	2	0	0	0	1	0	0	2	0
	CTG	1	25	3	0	0	0	0	0	0	1	0
	GTG	0	1	14	1	0	0	0	0	0	1	0
	TTG	0	0	0	6	0	0	1	0	0	0	0
	AAG	0	0	0	0	1	0	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0	0
	ATA	0	0	0	0	0	0	3	0	0	0	0
	ATC	0	0	0	0	0	0	0	1	1	0	0
	ATT	0	0	0	0	0	0	1	5	0	0	0
	ACG	0	2	0	0	0	0	0	0	3	0	0
andere	0	0	0	0	0	0	0	0	0	0	0	

Tabelle 21: Vergleich der TIC-Verwendung von ovuORFs zwischen Mensch und Maus. Die in der Mausstudie identifizierten TISs wurden gegen das humane Genom kartiert. 61 TICs von ovuORFs, die sowohl beim Menschen als auch bei der Maus als TISs verwendet werden, wurden miteinander verglichen.

		Maus										
Human		ATG	CTG	GTG	TTG	AAG	AGG	ATA	ATC	ATT	ACG	andere
	ATG	16	0	0	0	0	0	0	0	0	0	0
	CTG	0	21	1	1	0	0	0	0	0	0	0
	GTG	0	3	6	0	0	0	0	0	0	1	0
	TTG	0	0	0	3	0	0	0	0	0	0	0
	AAG	0	0	0	0	0	0	0	0	0	0	0
	AGG	0	0	0	0	0	0	0	0	0	0	0
	ATA	0	0	0	0	0	0	0	0	0	0	0
	ATC	0	0	0	0	0	0	0	2	0	0	0
	ATT	0	0	0	0	0	0	0	0	1	0	0
	ACG	0	0	0	0	0	0	0	0	0	6	0
	andere	0	0	0	0	0	0	0	0	0	0	0

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit nach Inhalt und Form, abgesehen von den Beratungen durch meinen akademischen Betreuer oder den angegebenen Hilfsmitteln und Quellen, meine eigene Arbeit ist.

Die Arbeit wurde bis jetzt weder vollständig noch in Teilen einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt und ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden.

Ferner erkläre ich, dass ich noch keine früheren Promotionsversuche unternommen habe.

Kiel, den

(Claudia Fritsch)

Lebenslauf

Persönliche Daten:

Name: Claudia Fritsch
Geburtstag: 22.12.1983
Geburtsort: Erfurt
Eltern: Birgit Fritsch, Frank Fritsch
Staatsangehörigkeit: deutsch
Familienstand: ledig
derzeitige Anschrift: Hohenrade 7, 24106 Kiel

Schulischer Werdegang:

06/2002 Abitur am Johann-Gutenberg-Gymnasium Erfurt

Studium:

10/2002 bis 03/2010 Studium der Biologie (Diplom) an der Friedrich-Schiller-Universität Jena
Fächer im Hauptstudium: Genetik, Neurobiologie, Anthropologie

03/2010 Diplom
Diplomarbeit: Der Einfluss physiologischer Bedingungen auf das alternative Spleißen an Tandem-3'-Spleißstellen
Gesamtnote: sehr gut

Promotion:

02/2010 Doktorand in der Arbeitsgruppe Genomische Gastroenterologie
bis voraussichtlich der Klinik für Allgemeine Innere Medizin des
06/2013 Universitätsklinikums S-H, Campus Kiel, unter der Leitung von Prof. Dr. Stefan Schreiber

Publikationen

Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M. (2012)

„Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting.”

Genome Res. 2012 Nov;22(11):2208-18

Danksagung

Mein ganz besonderer Dank gilt Herrn Prof. Dr. Jochen Hampe für seine Förderung und Unterstützung. Er ermöglichte mir in seiner Arbeitsgruppe die interessante Themenstellung der vorliegenden Arbeit zu bearbeiten.

Ich möchte mich bei Frau Prof. Dr. Karin Krupinska für die Betreuung dieser Arbeit bedanken. Sie hat durch konstruktive Kritik und gute und hilfreiche Vorschläge zur Verbesserung dieser Arbeit beigetragen.

Mein aufrichtiger Dank geht an Dr. Mario Brosch für seine engagierte Betreuung und seine große Begeisterung an meinem Thema. Ich möchte ihm außerdem für seinen unschätzbaren wissenschaftlichen Rat und seine Korrekturhilfen danken.

Ein großes Dankeschön geht an Alexander Herrmann für seine unersetzbare Hilfe bei den bioinformatischen Analysen.

Ich bedanke mich bei den Mitarbeitern der Genomanalyse-Gruppe des Fritz-Lipmann-Instituts in Jena, für ihren wissenschaftlichen Rat und die Hilfe bei den Sequenzierungen.

Weiterhin gilt ein großes Dankeschön all meinen Kollegen, die mich von Anfang an in ihrer Arbeitsgruppe so freundschaftlich aufgenommen haben und mir stets mit Rat und Tat zur Seite standen. Bei dieser Gelegenheit möchte ich besonders Caro, Julia, Sabine, Frank und Stephan erwähnen.

Abschließend gilt ein ganz besonderer Dank meiner Familie und meinen Freunden, dass sie immer an mich geglaubt haben und bei der Verwirklichung meiner Ziele an meiner Seite standen.

Vielen Dank!