

Untersuchungen über den Einfluss des IGF2
mRNA-bindenden Proteins 1 auf
in-vitro-Tumorzellmigration mit Methoden
der digitalen Bildverarbeitung und
Genexpressionsanalysen

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät III
Agrar- und Ernährungswissenschaften,
Geowissenschaften und Informatik
der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Markus Glaß

geb. am 30.09.1981 in Wolfen

Gutachter:

Prof. Dr. Stefan Posch

Prof. Dr. Stefan Hüttelmaier

Prof. Dr. Michaela Müller-McNicoll

Verteidigung am 02.02.2017

Danksagung

Ich möchte hiermit allen danken, die mich während der Entstehung dieser Dissertation sowohl fachlich als auch moralisch unterstützt haben.

Besonderer Dank gilt meinen beiden Betreuern Prof. Dr. Stefan Hüttelmaier und Prof. Dr. Stefan Posch für die Möglichkeit ein sehr interessantes Forschungsthema zu bearbeiten. Sie standen mir stets mit Rat und Tat zur Seite und fanden darüber hinaus immer Zeit für hilfreiche Diskussionen.

Dr. Nadine Bley und Claudia Misiak danke ich dafür, dass sie die gesamten Laborarbeiten, welche für die vorgestellten Experimente notwendig waren, durchgeführt haben. Ohne ihre gewissenhaften Arbeiten wäre diese Dissertation nicht möglich gewesen.

Des Weiteren danke ich Dr. Kristin Wächter, Dr. Anne Zirkel und Britta Mühleck für die Erstellung zahlreicher Referenzsegmentierungen, welche ich für die Evaluierung der, für die Dissertation entwickelten, Bildverarbeitungsverfahren verwendet habe.

Zu guter Letzt bedanke ich mich bei der gesamten AG Hüttelmaier für eine überaus angenehme Arbeitsatmosphäre und zahlreiche wertvolle Diskussionen, welche oft auch über rein fachliche Belange hinausgingen.

Inhaltsverzeichnis

1	Einleitung	7
1.1	Zellmigration	7
1.1.1	Modell der invasiven Kriechbewegung	7
1.1.2	Einzelzellmigration	9
1.1.3	Kollektive Zellmigration	10
1.1.4	Zellmigration und Krebs	10
1.1.5	Methoden zur Untersuchung der <i>in-vitro</i> -Zellmigration	11
1.2	IGF2BP1	17
1.3	Methoden der Bildverarbeitung	22
1.3.1	Segmentierung	22
1.3.2	Level-Set-Verfahren	24
1.3.3	Objektverfolgung	26
1.3.4	Klassifikation mit Support Vector Machines	27
1.3.5	MiToBo	29
1.4	Hochdurchsatz-RNA-Sequenzierung	30
1.5	Zielstellung der Arbeit	34
2	Material und Methoden	35
2.1	Zellkultur und Transfektionen	35
2.2	Mikroskopie	36
2.3	Topologieerhaltende Level-Set-Segmentierung	37
2.4	CellMigrationAnalyzer	39
2.5	ScratchAssayAnalyzer	49
2.6	Auswertung von RNA-Seq Daten	55
2.6.1	Probenvorbereitung und Sequenzierung	55
2.6.2	Bestimmung differentieller Genexpression	55
2.6.3	Detektion von Mykoplasmakontaminationen	60
2.6.4	Genanreicherungsanalysen	61
2.6.5	Bestimmung putativer mRNA-microRNA-Bindungen	61
3	Ergebnisse	62
3.1	Algorithmenevaluation	62
3.1.1	Evaluation des CellMigrationAnalyzer	62

3.1.2	Evaluation des ScratchAssayAnalyzer	78
3.2	Anwendungen	90
3.2.1	Extrazelluläre Matrix	90
3.2.2	Chemotaxis	94
3.2.3	Pharmakologische Inhibitoren	97
3.3	Einfluss von IGF2BP1 auf Zellmigration	100
3.4	Einfluss von IGF2BP1 auf die Genexpression	106
3.4.1	Beeinflussung migrationsrelevanter Gene	112
3.4.2	Bestimmung potentieller neuer IGF2BP1-Zieltranskripte	122
4	Diskussion	131
5	Zusammenfassung	141
	Anhang	143
	Literaturverzeichnis	165

1 Einleitung

1.1 Zellmigration

Zellmigration bezeichnet die aktive Fortbewegung einer Zelle, von einem Ort an einen anderen [1]. Die Fähigkeit zur Migration ist sowohl für viele Einzeller, als auch für vielzellige Organismen ein Prozess von fundamentaler Bedeutung und ist einer komplexen Regulation unterworfen. Einzellige Organismen gelangen durch Zellmigration beispielsweise zu Nahrungsquellen oder Geschlechtspartnern und können Fressfeinden entkommen. Während der Embryonalentwicklung vielzelliger Tiere spielt Zellmigration zum Beispiel während der Gastrulation oder der Entwicklung des Nervensystems eine wichtige Rolle [70]. Im adulten Wirbeltierorganismus sind Wundheilung und Immunantwort Beispiele für Vorgänge, für die Zellmigration unerlässlich ist. Fehlerhafte oder fehlgesteuerte Zellmigration ist unter anderem mit Gefäßerkrankungen, Entwicklungsstörungen, Immunkrankheiten und Krebs assoziiert.

Zellmigration kann ungerichtet, aber auch von externen Faktoren geleitet sein. Wird die Bewegungsrichtung von einem Konzentrationsgradienten eines extrazellulären Signalstoffes beeinflusst, handelt es sich um Chemotaxis. Wird die Bewegung der Zellen durch Signalstoffe zwar induziert, jedoch nicht in eine bestimmte Richtung gelenkt, handelt es sich um Chemokinese. Ändern Zellen aufgrund von Kontakt zu physischen Hindernissen ihre Bewegungsrichtung, wird dies als Haptotaxis bezeichnet. Die Änderung der Bewegungsrichtung aufgrund der Elastizität des umgebenden Substrates wird als Durotaxis bezeichnet. Die Migration von Zellen kann über das Vorhandensein oder Fehlen von Kontakten zu anderen Zellen in die zwei Modi Einzelzellmigration und kollektive Zellmigration eingeteilt werden. Beide Modi werden für unterschiedliche Aufgaben, sowohl beim Aufbau und der Erhaltung von Geweben, als auch in pathologischen Prozessen, eingesetzt [37].

Die typische Umgebung, in der Zellmigration innerhalb eines vielzelligen Organismus stattfindet, ist die extrazelluläre Matrix. Diese bietet den Zellen eine dreidimensionale Umgebung, an der sie Halt finden, welche aber auch natürliche Barrieren bildet. Viele migrierende Zellen können die extrazelluläre Matrix aktiv umgestalten, um diese Barrieren zu überwinden.

1.1.1 Modell der invasiven Kriechbewegung

Im Gegensatz zu vielen Einzellern bewegen sich Zellen vielzelliger Organismen in der Regel nicht schwimmend, mit Hilfe von Geißeln oder Cilien, sondern kriechend fort. Diese Kriechbewegung basiert auf der Polarisierung der Zelle in Vorder- und Hinterteil, sowie auf Anheftung oder Verankerung an die Umgebung und anschließender Kontraktion des Zellkörpers.

Die Bewegung vieler Zelltypen, die von einer dreidimensionalen extrazellulären Matrix umgeben sind, kann durch zyklische Wiederholung der folgenden fünf Schritte beschrieben werden (nach [33]):

Polarisierung durch Membranausstülpungen. Aktinfilamente polymerisieren unterhalb der Zellmembran. Dadurch entstehen lokale Ausstülpungen der Zellmembran. Diese werden überwiegend in Bereichen der Zelle gebildet, welche in die Bewegungsrichtung zeigen. Dadurch entsteht eine Polarisierung der Zelle. Der Bereich mit den Membranausstülpungen definiert den Vorderteil, während der entgegengesetzte Teil der Zelle den Hinterteil bildet. Es existieren verschiedene Formen dieser Ausstülpungen. Zwei Formen, welche häufig für die Zellmigration gebildet werden, sind Lamellipodien und Filopodien. Lamellipodien sind flache, tuchartige Ausstülpungen mit verzweigtem Aktinnetzwerk. Filopodien hingegen sind dünne, fingerförmige Ausstülpungen mit parallel angeordneten Aktinbündeln. Abbildung 1 zeigt eine polarisierte Zelle und ihr Aktinfilamentnetzwerk.

Bildung von Verankerungen. Die Membranausstülpungen stellen Kontakt zur extrazellulären Matrix her. Extrazelluläre Domänen von Transmembranrezeptoren, welche in den Ausstülpungen lokalisiert sind, beispielsweise Integrine, binden an Liganden der extrazellulären Matrix. Über verschiedene Adapterproteine wird das Aktinzytoskelett mit den intrazellulären Domänen der Transmembranrezeptoren verbunden und dadurch eine Verbindung zwischen extrazellulärer Matrix und dem Aktinzytoskelett geschaffen. Dies induziert eine lokale Häufung weiterer Transmembranrezeptoren an der Zellmembran. Diese lokalen Anhäufungen werden Fokalkontakte genannt. Fokalkontakte sind dynamische Strukturen, welche die Geschwindigkeit der Zellmigration maßgeblich beeinflussen [36].

Proteolyse der extrazellulären Matrix. Proteasen, zum Beispiel Matrixmetalloproteinasen (MMPs), werden an die Zellmembranen rekrutiert oder sezerniert und spalten große Makromoleküle der extrazellulären Matrix in kürzere Fragmente. Dadurch entsteht freier Raum in dem sich die Zelle weiterbewegen kann.

Kontraktion. Durch Bindung von aktiviertem Myosin II an Aktinfilamente des Zytoskeletts bildet sich Aktomyosin. Diese Verbindung ist in der Lage zu kontrahieren. Durch die Verankerung des Aktinzytoskeletts mit der extrazellulären Matrix und der, durch Aktomyosin vermittelten, Kontraktion baut sich eine Zugspannung auf.

Lösen der Verankerungen. Fokalkontakte am hinteren Teil der Zelle werden gelöst. Dadurch wird der Zellkörper aufgrund der Zugspannung nach vorn gezogen.

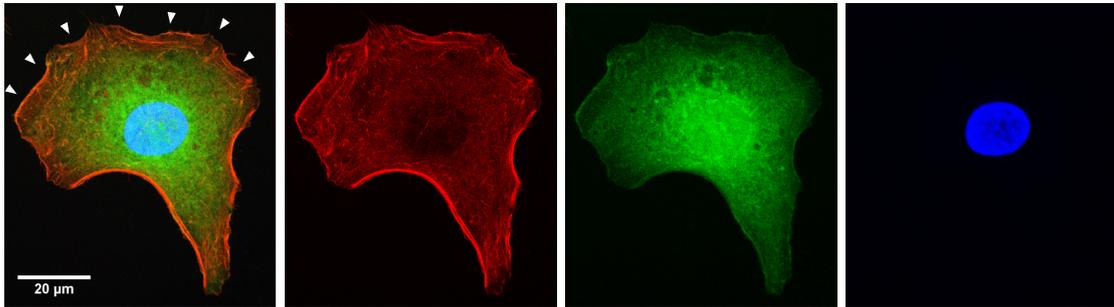


Abbildung 1: Beispiel einer polarisierten Zelle. Abgebildet ist eine GFP-exprimierende Zelle der Zelllinie SK-OV-3. Der obere Bereich der Zelle stellt den Vorderteil, der untere den Hinterteil dar. Pfeilspitzen markieren den Bereich des Lamellipodiums. Rot: Aktinfasern gefärbt mit an Phalloidin gekoppeltem Farbstoff TRITC. Grün: GFP. Blau: DNA-Färbung mit DAPI.

Die Polarisierung der Zelle ist dynamisch und kann spontan eintreten oder durch externe Stimuli, zum Beispiel Signalstoffe, induziert werden [99]. Eine weitere für Zellmigration genutzte Form von Membranausstülpungen stellen Bläschen dar. Im Gegensatz zu Lamellipodien und Filopodien entstehen Bläschen, wenn sich die Plasmamembran punktuell vom darunterliegenden Aktinkortex löst. Durch den entstehenden hydrostatischen Druck des Zytoplasmas bilden sich Membranausstülpungen. Die Rückbildung der Bläschen erfolgt durch Anlagerung von Aktinfilamenten und Myosin an der Zellmembran und anschließender Kontraktion des Aktomyosins [94].

1.1.2 Einzelzellmigration

Die individuelle Bewegung von Zellen, welche nicht oder nur kurzzeitig über Zell-Zellkontakte mit anderen Zellen verbunden sind, wird als Einzelzellmigration bezeichnet. Sie kann in die beiden Formen amöboid und mesenchymal unterteilt werden [37].

Die amöboide Variante der Einzelzellmigration wird von Amöben wie *Dictyostelium discoideum*, aber auch von Leukozyten, sowie einigen Tumorzellen angewendet. Dabei gehen die Zellen nur schwache und kurze Interaktionen mit ihrem Substrat ein und nutzen weitgehend proteaseunabhängige Mechanismen, um sich durch die extrazelluläre Matrix zu bewegen. Das bedeutet, sie bewegen sich an Barrieren vorbei anstatt sie zu degradieren. Zellen, welche diese Form der Migration anwenden, sind 10 bis 30 mal schneller als Zellen, welche sich nach der mesenchymalen Form fortbewegen [36].

Bewegung nach der mesenchymalen Form folgt dem oben beschriebenen Fünf-Schritte-Modell (s. Abschnitt 1.1.1). Die Zellen weisen eine spindelförmige Gestalt auf, welche von Integrin-vermittelter Adhäsion und hohen Traktionskräften an beiden Zellpolen abhängt ([36]). Der Übergang zwischen amöboider und mesenchymaler Form ist fließend und kann von der Beschaffenheit der extrazellulären Matrix beeinflusst werden [37].

Einzelne Zellen, welche hintereinander entlang eines gemeinsamen Pfades migrieren, können transiente Zell-Zellkontakte ausbilden. Dies wird als kettenförmige Migration bezeichnet [37].

1.1.3 Kollektive Zellmigration

Kollektive Zellmigration ist definiert als die Bewegung von Zellgruppen, in denen die Zellen sowohl physisch als auch funktionell miteinander verbunden bleiben. Innerhalb dieser verbundenen Zellkohorten bildet sich eine Polarität und suprazelluläre Organisation des Aktinzytoskeletts aus. Zellen an der Spitze der Kohorte leiten die Bewegung der übrigen Zellen [35]. Zell-Zellverbindungen ermöglichen die koordinierte Ausbildung der Membranaustülpungen und Zellkontraktionen der miteinander verbundenen Zellen ([33]).

Während der Embryogenese ist kollektive Zellmigration unerlässlich für die Entwicklung von Geweben und Organen. Im erwachsenen Organismus ist kollektive Zellmigration unter anderem an der Wundheilung und der Entstehung und Verzweigung von Blutgefäßen beteiligt. Darüber hinaus spielt kollektive Zellmigration auch bei der Invasion von Tumorzellen in andere Gewebe eine wichtige Rolle ([35, 98]).

Kollektiv migrierende Zellen können in kleinen Gruppen (Cluster) migrieren oder in großen Verbänden (Strand) [36]. Eine Übersicht der verschiedenen Zellmigrationsmodi ist in Abbildung 2 dargestellt.

1.1.4 Zellmigration und Krebs

Die Aktivierung der Fähigkeiten von Tumorzellen in fremde Gewebe einzudringen und Metastasen zu entwickeln, also sich vom primären Tumor zu lösen und an anderer Stelle im Organismus neue Kolonien zu bilden, gilt als ein grundlegendes Kennzeichen von Krebs [49, 33]. Für diese Vorgänge ist Zellmigration von großer Bedeutung. Da insbesondere Metastasenbildung mit einer schlechten Prognose einhergeht und für die meisten krebsassoziierten Todesfälle verantwortlich ist [64], ist die Untersuchung der Tumorzellmigration von hohem Interesse.

Die Mechanismen, welche Krebszellen für Migration einsetzen, entsprechen jenen, welche auch von gesunden Zellen verwendet werden [33]. Für die Invasion der Tumorzellen in fremde Gewebe wird vor allem kollektive Zellmigration, für Metastasenbildung hingegen überwiegend

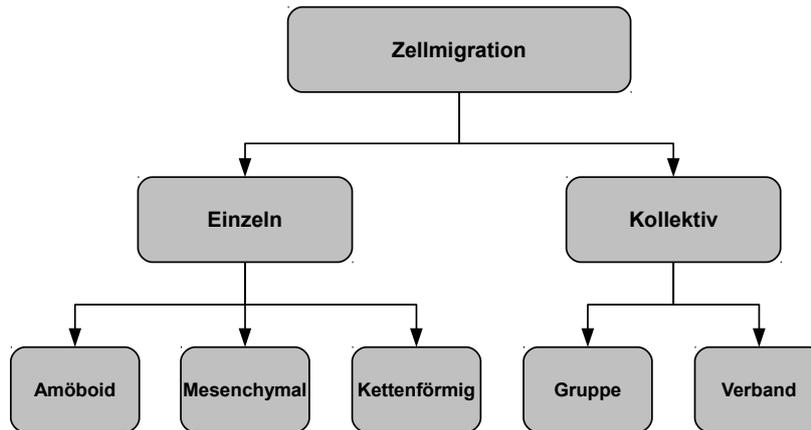


Abbildung 2: Zellmigrationsmodi. Zellmigration wird über das Fehlen oder Vorhandensein von persistenten Zell-Zellkontakten in die Modi einzeln oder kollektiv eingeteilt. Über die Stärke von Zell-Zellkontakten und der Adhäsion mit der umgebenden extrazellulären Matrix können diese Modi weiter unterteilt werden. Einteilung nach [36]

Einzelzellmigration verwendet [37]. Allerdings kann in vielen Tumortypen simultan sowohl kollektive, als auch Einzelzellmigration beobachtet werden, wobei generell die Wahrscheinlichkeit für Einzelzellmigration mit abnehmenden Differenzierungsgrad zunimmt [36].

Tumorzellen epithelialen Ursprungs können durch Reaktivierung des Entwicklungsprogramms, welches als epithelial-zu-mesenchymale Transition (EMT) bezeichnet wird, unter anderem die Fähigkeit zur Migration erwerben [49]. EMT und der entgegengesetzte Prozess MET (mesenchymal-zu-epitheliale Transition) werden während der Embryonalentwicklung für die Entstehung der meisten Gewebe und Organe benötigt [111]. Im adulten Organismus können die Prozesse, die dem Vorgang der EMT zugrunde liegen, für Wundheilung, aber auch bei der Entstehung von Fibrosen oder der Krebsentwicklung reaktiviert werden [66]. EMT ist gekennzeichnet durch den Verlust von Zell-Zell-Kontakten, Reorganisation des Zytoskeletts, Veränderungen der Zellform und erhöhte Motilität. Damit einhergehend wird oft die Fähigkeit die extrazelluläre Matrix zu degradieren aktiviert. Zellen, welche den EMT-Prozess durchlaufen haben, zeigen sich darüber hinaus häufig resistent gegenüber Apoptose und Seneszenz [66].

1.1.5 Methoden zur Untersuchung der *in-vitro*-Zellmigration

Für die Erforschung der Zellmigration sind zahlreiche Methoden entwickelt worden. Dabei wird die Migration der Zellen entweder innerhalb eines Organismus (*in vivo*) oder an isolier-

ten Zellen (*in vitro*) beobachtet. Im Gegensatz zu *in-vivo*-Methoden sind Experimente zur Untersuchung der Zellmigration *in vitro* in der Regel besser reproduzierbar, sowie einfacher und in höherer Zahl durchführbar. Des Weiteren spiegeln Experimente an lebenden Tieren physiologische Bedingungen zwar besser wider, jedoch sind sie mit hohen ethischen Vorbehalten verbunden.

In-vitro-Methoden zur Untersuchung der Zellmigration können in 2D- und 3D-Methoden unterteilt werden. Für 2D-Methoden werden Zellen auf die Oberfläche von Objektträgern oder Kulturschalen gesät. Diese Oberflächen bestehen in der Regel aus Glas oder Kunststoffen. Um physiologische Bedingungen besser zu approximieren, können diese Oberflächen mit Bestandteilen der extrazellulären Matrix (EZM) beschichtet werden. Für 3D-Methoden werden die Zellen stattdessen in ein Netzwerk aus EZM-Proteinen oder synthetisch hergestellten Polymeren eingebettet. Verschiedene Parameter dieser Netzwerke, wie beispielsweise Porengröße oder molekulare Zusammensetzung, ermöglichen es, eine Umgebung zu schaffen, die dem natürlichen Umfeld der untersuchten Zellen nahe kommt. Aufgrund ihrer Einfachheit sind viele 2D-Methoden für Hochdurchsatzanalysen geeignet. Das macht sie zu einem hervorragenden Mittel zum simultanen Testen des Migrationsverhaltens unter einer Vielzahl verschiedener Bedingungen. Da diese Methoden eine Anheftung der Zellen an den Untergrund voraussetzen, eignen sie sich jedoch nur für die Untersuchung adhärenter Zellen. 3D-Methoden sind aufwändiger in der Durchführung, jedoch simulieren sie die natürliche Umgebung der Zellen und sind daher besser geeignet den Migrationsprozess per se zu untersuchen [27]. Zudem erlauben sie die Untersuchung nicht-adhärenter Zellen, beispielsweise solcher Zellen, welche sich nach der amöboiden Form fortbewegen.

Viele der Methoden basieren auf der Beobachtung von Zellen mit Hilfe von Mikroskopen. Moderne Mikroskope verfügen häufig über digitale Kameras zur Datenakquisition. Dies ermöglicht die automatisierte Auswertung der Daten mittels Verfahren der digitalen Bildverarbeitung. Dadurch kann eine Vielzahl unterschiedlicher Bedingungen in geringer Zeit objektiv und reproduzierbar analysiert werden.

Die Färbung von Zellen mit Farbstoffen oder die Expression von fluoreszierenden Proteinen genetisch veränderter Zellen in Kombination mit Fluoreszenzmikroskopie, führt häufig zu einer verbesserten Detektierbarkeit der Zellen, da dadurch ein höherer Kontrast zum Hintergrund erreicht wird. Die Behandlung der Zellen mit Farbstoffen oder die Veränderung des genetischen Materials der Zellen birgt jedoch auch die Gefahr, einen Einfluss auf das migratorische Verhalten der Zellen auszuüben.

Im Folgenden werden einige häufig verwendete *in-vitro*-Methoden zur Untersuchung der Zellmigration beschrieben.

Transwell-Migration-Assay/Invasion-Assay Beim Transwell-Migration- oder Boyden-Chamber-Assay werden zwei mit Nährmedium gefüllte Kammern durch eine horizontal dazwischen liegende poröse Membran voneinander getrennt. In die obere Kammer werden Zellen gegeben, welche durch die Poren der Membran in die untere Kammer gelangen können. Um die Transmigration zu induzieren, werden in die untere Kammer oft Substanzen gegeben, welche auf die Zellen anziehend wirken. Durch die Membran hindurch entsteht ein Konzentrationsgradient dieser Lockstoffe, welcher jedoch nur kurzzeitig besteht. Dadurch ist es mit dieser Methode nicht möglich, Chemotaxis von Chemokinese, der ungerichteten Bewegung infolge einer chemischen Stimulation, zu unterscheiden [27]. Die Größe der Membranporen muss in Abhängigkeit des untersuchten Zelltyps so gewählt werden, dass die Zellen nur durch aktive Migration hindurch gelangen und nicht durchfallen können. Nach einer gewissen Inkubationszeit, deren Dauer ebenfalls vom Zelltyp abhängt, wird die Anzahl der Zellen, welche sich auf der unteren Membranseite befinden quantifiziert. Dies geschieht zum Beispiel durch Fixieren der Zellen und anschließender Betrachtung der Membran unter einem Mikroskop. Wird die Membranoberseite vor dem Einsäen der Zellen mit Proteinen der extrazellulären Matrix, beispielsweise Kollagenen, beschichtet, kann zusätzlich die Invasivität der untersuchten Zellen untersucht werden.

Transwell-Migration-/Invasion-Assays sind einfach durchführbar und eignen sich für eine Vielzahl von Zelltypen. Allerdings stellen die Ergebnisse nur Endpunktmessungen dar und in Abhängigkeit der Quantifizierungsmethode müssen Zellen, welche sich auf der Oberseite der Membran befinden, gegebenenfalls aufwändig entfernt werden [64]. Ein weiterer Nachteil dieser Methode ist, dass nur aktiv migrierende Zellen betrachtet werden. Demzufolge setzen präzise Aussagen über die Zellpopulation ein homogenes Migrationsverhalten aller Zellen voraus. Des Weiteren kann nicht bestimmt werden, welche der beobachteten Zellen erst nach der Transmigration durch Proliferation entstanden sind.

Scratch-Assay/Gap-Closure-Assay Scratch- beziehungsweise Gap-Closure-Assays sind beliebte und einfach durchzuführende Verfahren zur Untersuchung der Zellmigration. Beim Scratch-Assay wird in eine konfluente Schicht adhärenter Zellen mechanisch, beispielsweise durch Kratzen mit einer Pipettenspitze, eine dünne Lücke zwischen den Zellen erzeugt. Anschließend wird mit Hilfe mikroskopischer Aufnahmen zu diskreten Zeitpunkten die Fläche

der künstlich erzeugten Wunde dokumentiert. Durch Migration von Zellen in den Wundbereich, aber in gewissem Maß auch durch Proliferation, reduziert sich die Wundfläche im Laufe der Zeit. Das Ausmaß dieser Flächenreduktion dient dabei als Hinweis auf das migratorische Potential der untersuchten Zellpopulation. Eine weitere gebräuchliche Bezeichnung für diese Art von Experiment ist Wound-Healing-Assay. Beim Gap-Closure-Assay beschränkt eine physikalische Barriere das Absetzen und das Wachstum der zu untersuchenden Zellen auf einen Bereich um diese Barriere. Zu Beginn des Experimentes wird die Barriere entfernt, wodurch, wie beim Scratch-Assay, ein zellfreier Bereich entsteht, in den die Zellen migrieren können. Ein Vorteil des Gap-Closure-Assay gegenüber dem Scratch-Assay ist die höhere Reproduzierbarkeit der initialen Wundflächen. Sofern nicht die komplette Wunde aufgenommen wird, empfiehlt es sich ein Mikroskop zu verwenden, welches mit einem motorisierten und steuerbaren Objektisch ausgestattet ist. Dadurch kann sichergestellt werden, dass immer die selben Bereiche der Wunden aufgenommen werden. Für Experimente mit Zellen, welche bestimmte Anforderungen an die Temperatur und Gaszusammensetzung ihrer Umgebung stellen, empfiehlt sich zudem eine Inkubationskammer um das Mikroskop.

Scratch- und Gap-Closure-Assays sind einfach durchführbar und hochdurchsatzgeeignet. Nachteile sind der hohe Zeit- und Materialaufwand für die Durchführung [72], sowie die Tatsache, dass, insbesondere bei längeren Versuchen (> 24 h [64]), Zellproliferationsprozesse die Messung der Migration verzerren können.

Um Zellmigration mittels Scratch- oder Gap-Closure-Assays zu messen, werden üblicherweise die Ränder der Wunden zu verschiedenen Zeitpunkten bestimmt. Um diese Ränder eindeutig definieren zu können, sollten die Zellen nicht allzu weit von einander entfernt sein. Dadurch ist dieses Verfahren vor allem dazu geeignet, kollektiv migrierende Zellen zu untersuchen. Des Weiteren sollte die Dauer der Experimente an die verwendeten Zellen angepasst werden, um Proliferationseffekte zu minimieren

Die Markierung der Wundränder geschieht häufig manuell. Dies ist jedoch aufwändig, insbesondere für Experimente bei denen viele Bedingungen getestet werden. Häufig werden Scratch- und Gap-Closure-Assays nur für Endpunktmessungen verwendet. Um jedoch einen detaillierten Einblick in die Wundentwicklung zu erlangen, ist die Auswertung vieler Zeitpunkte nötig. Dies führt allerdings zu einer deutlichen Erhöhung der Anzahl auszuwertender Bilder. Deshalb wurden bereits einige Methoden entwickelt, um die Veränderung der Wunden automatisiert zu quantifizieren. In [81] wurden mit Hilfe der Bildverarbeitungssoftware ImageJ [104] manuell gewählte Schwellwerte zum Erkennen von fluoreszierenden Zellen verwendet. Die für die Bildverarbeitungssoftware CellProfiler [67] angebotene „Wound Healing Pipeline“

basiert ebenfalls auf der Anwendung von Intensitätsschwellwerten. Intensitätsbasierte Ansätze sind jedoch für Durchlichtmikroskopieaufnahmen ungefärbter Zellen wenig geeignet, da sich die Intensitäten der Zellen kaum von denen des Hintergrundes unterscheiden. Das speziell für die Analyse von Scratch-Assays entwickelte Programm TScratch [39] verwendet die sogenannte Curvelet-Transformation [12], um intensitätsunabhängig über Textureigenschaften Zellen und Wunde zu identifizieren. TScratch setzt jedoch Bilder mit hohem Kontrast und wenig Zelltrümmern voraus [42]. Im Rahmen der vorliegenden Arbeit wurde ein Verfahren entwickelt (s. Abschnitt 2.5), welches die Wundentwicklung von Scratch-/Gap-Closure-Assays automatisiert analysiert. Dieses Verfahren kann sowohl für fluoreszierende, als auch für nicht-fluoreszenzmarkierte Zellen verwendet werden. Darüber hinaus ist es durch die Verwendung nur weniger freier Parameter einfach zu bedienen.

Untersuchung der Einzelzellmigration mit kolloidalen Partikeln Mit Hilfe kolloidaler Partikel ist es möglich, die in einem definierten Zeitintervall zurückgelegten Pfade individueller Zellen darzustellen. Hierfür werden die Oberflächen von Zellkulturschalen beispielsweise mit winzigen Goldpartikeln beschichtet. Anschließend werden Zellen in geringer Dichte in diese Schalen eingesät. Die Zellen nehmen die Partikel durch Phagozytose auf. Die dadurch entstehenden Spuren können mikroskopisch aufgenommen und für die Quantifizierung verschiedener Migrationsparameter verwendet werden. Die Pfade können während ihrer Entstehung zu diskreten Zeitpunkten oder als Endpunktmessungen aufgenommen werden. Ein Vorteil dieser Methode ist die Tatsache, dass die vollständig zurückgelegten Pfade einzelner Zellen als Endpunktmessungen erfasst werden können, so dass eine Inkubationskammer um das Mikroskop nicht zwingend notwendig ist. Nachteile sind der hohe Vorbereitungsaufwand, sowie die geringe Anzahl beobachtbarer Zellen pro Experiment [64].

Einzelzell-Tracking Eine weitere Möglichkeit zur Untersuchung der Bewegungen einzelner Zellen stellen mikroskopiegestützte Zeitrafferaufnahmen von Zellen, welche zuvor auf geeignete Objektträger oder in Zellkulturschalen eingesät wurden, dar. Die Oberfläche, auf der sich die Zellen bewegen, kann dabei mit verschiedenen Proteinen beschichtet werden, um die Bewegung auf einer extrazellulären Matrix zu simulieren. Diese Art von Experiment erfordert vergleichsweise wenig experimentelle Vorbereitung. Für reproduzierbare Langzeituntersuchungen, beispielsweise an menschlichen Zellen, müssen jedoch physiologische Bedingungen geschaffen werden. Während des gesamten Aufnahmezeitraumes muss sichergestellt werden, dass die Temperatur, die Feuchtigkeit und die Gaszusammensetzung der Umgebung den Ansprüchen

der untersuchten Zellen gerecht wird. Mit Hilfe manueller oder computergestützter Erfassung der individuellen Zellen kann, neben der Geschwindigkeit, eine Vielzahl weiterer Migrationsparameter der Zellen bestimmt werden. Zusätzlich können durch geeignete Detektionsverfahren Veränderungen der Gestalt der Zellen ermittelt werden. Die automatisierte Auswertung dieser Zeitrafferaufnahmen ist allerdings komplizierter als für die zuvor genannten Methoden.

Durch Einbetten der Zellen in dreidimensionale Polymermatrizes können physiologische Bedingungen besser simuliert werden. In Kombination mit geeigneten Mikroskopie- und Auswertetechniken ist damit die Untersuchung von 3D-Zellmigration und -Invasion möglich. Einzelzell-Tracking-Experimente eignen sich hervorragend für die simultane Untersuchung vieler Zellen und unter verschiedenen Bedingungen. Jedoch sollte die Zelldichte nicht zu hoch sein, da dies die Detektion individueller Zellen erschwert.

Die Auswertung der Einzelzellmigration erfolgt in der Regel mit Hilfe spezieller Software. Eine Anwendung, welche bereits häufig zur semi-automatischen Analyse der Bewegungen von Zellen oder intrazellulären Partikeln verwendet wurde, zum Beispiel in [93, 38, 71, 51], ist das ImageJ-Plugin `ManualTracking` (rsbweb.nih.gov/ij/plugins/track/track.html). Dieses Plugin erfasst Bildkoordinaten, welche der Benutzer mit der Maus markiert. Anschließend setzt es Koordinaten, welche zu einem Objekt gehören, zu Trajektorien zusammen. Diese Art der Auswertung ist jedoch sehr aufwändig und nur schwer reproduzierbar. Deshalb wurden viele Ansätze entwickelt, um die Bewegung einzelner Zellen in Zeitraffervideos vollständig automatisiert auszuwerten. Diese Auswertung ist in der Regel gegliedert in die Identifizierung der Bildpunkte, welche Zellen darstellen (Segmentierung) und die Erkennung der individuellen Zellen innerhalb der Einzelbilder (Zuordnung). Die Wahl der Algorithmen zur erfolgreichen Erfüllung dieser Aufgaben hängt stark von der Art und dem Kontext der Aufnahmen ab. Dabei scheint es, dass für unterschiedliche Tracking-Aufgaben auch jeweils unterschiedliche Algorithmen benötigt werden, um optimale Ergebnisse zu erzielen [80]. Ein 2013 durchgeführter Vergleich verschiedener Verfahren mit Hilfe eines Wettbewerbes [76], unterstützt diese Hypothese. Im Rahmen dieses Wettbewerbes mussten die Programme der Teilnehmer Zellen oder Zellkerne in verschiedenen Videos erkennen und zuordnen. Die Charakteristiken der Videos unterschieden sich deutlich, und obwohl einige Programme generell besser geeignet zu sein schienen als andere, konnte keines der Programme überall die besten Ergebnisse erzielen. Im Rahmen der vorliegenden Arbeit wurde ein Verfahren entwickelt (s. Abschnitt 2.4), welches eine automatisierte Auswertung der Einzelzellmigration fluoreszenzmartierter Zellen aus 2D-Zeitrafferaufnahmen ermöglicht. Neben der Produktion verlässlicher Ergebnisse lag der Fokus der Entwicklung auf einfacher Bedienbarkeit durch die Verwendung weniger freier Parameter.

Chemotaxis-Assays Um Chemotaxis zu untersuchen, sind stabile Konzentrationsgradienten der eingesetzten Lockstoffe notwendig. Für diesen Zweck wurden verschiedene, auf voneinander getrennten Kammern basierende, Vorrichtungen entwickelt. Diese Kammern sind durch eine dünne Brückenregion miteinander verbunden. Werden in Nährmedium gelöste Lockstoffe in eine der Kammern gegeben, so bildet sich durch Diffusion ein Konzentrationsgradient dieses Stoffes entlang der Brückenregion aus. Zeigen Zellen, welche in die andere Kammer gegeben wurden, eine chemotaktische Reaktion, so migrieren sie über die Brückenregion in Richtung der steigenden Lockstoffkonzentration. Die Vorrichtungen sind so konzipiert, dass Zellen, welche die Brückenregion passieren, mikroskopisch beobachtet werden können. Mittels Methoden zum Erkennen der individuellen Zellen zu diskreten Aufnahmezeitpunkten können, analog zum Einzelzell-Tracking, Migrationsparameter individueller Zellen bestimmt werden. Über die Anzahl der Zellen, welche in die Brückenregion migrieren, können zusätzliche Rückschlüsse auf das chemotaktische Potential der untersuchten Zellpopulationen gezogen werden. Beispiele für Chemotaxis-Assays basierend auf getrennten Kammern sind Zigmond-Chamber-Assays [126] und Dunn-Chamber-Assays [125]. Die „ μ -Slides Chemotaxis“ der Firma Ibidi (Ibidi, Martinsried) stellen eine Variation dieser Methoden dar. Hierbei werden ebenfalls zwei Kammern verwendet, welche durch eine dünne Brücke verbunden sind, um einen Konzentrationsgradienten eines Lockstoffes zu erzeugen. Jedoch werden die zu untersuchenden Zellen nicht in eine der Kammern, sondern direkt in den Brückenbereich eingebracht. Auf einem „ μ -Slide Chemotaxis“ befinden sich drei unabhängige Kammersysteme, so dass simultan zum Chemotaxis-Versuch Kontrollbedingungen ohne Lockstoff beziehungsweise mit konstanter Lockstoffkonzentration aufgenommen werden können. Die Objektträger sind in zwei Varianten erhältlich, zum Aus säen adhärenter Zellen und zum Einbetten von Zellen in 3D-Matrizes. Der Konzentrationsgradient der verwendeten Lockstoffe bleibt nach Herstellerangaben über 48 h stabil.

Bei diesen Experimenten wird, analog zu Einzelzell-Tracking-Experimenten, die Bewegung einzelner Zellen unter dem Mikroskop beobachtet. Daher eignen sich Verfahren zur Analyse von Einzelzell-Tracking-Experimenten prinzipiell auch zur Quantifizierung von Chemotaxisexperimenten. Für die Ermittlung chemotaxisbezogener Parameter, sollten diese Verfahren jedoch auch Richtung und Stetigkeit der Zellbewegungen berücksichtigen.

1.2 IGF2BP1

Die Zellmigration unterliegt einer komplexen Regulation an der viele Faktoren beteiligt sind. Einer dieser Faktoren, dessen Einfluss auf die Migration von Tumorzellen im Rahmen der vorliegenden Arbeit genauer betrachtet wurde, ist das insulin-like growth factor-2 mRNA-binding

protein 1, kurz IGF2BP1. Es gehört zu einer hochkonservierten Proteinfamilie, welche in Säugtieren die drei Mitglieder IGF2BP1, IGF2BP2 und IGF2BP3 umfasst. Für alle bekannten Mitglieder der IGF2BP-Familie konnte eine RNA-bindende Funktion festgestellt werden [4]. Hierfür besitzen diese Proteine jeweils sechs RNA-Bindedomänen, zwei vom Typ RRM (RNA recognition motif) und vier KH-Domänen (hnRNP-K homology). Abbildung 3 zeigt die Gen- und Proteinstrukturen von IGF2BP1.

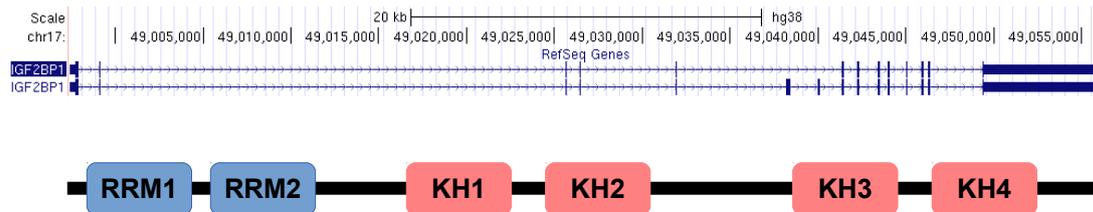


Abbildung 3: IGF2BP1-Gen- und -Proteinstruktur. Oben: genomische Struktur gemäß NCBI-Referenzsequenzen (RefSeq, GRCh38/hg38). Es befinden sich zwei proteinkodierende Transkriptvarianten des menschlichen IGF2BP1-Gens auf Chromosom 17. Abbildung aus UCSC Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgTracks>) Unten: Domänenstruktur des IGF2BP1-Proteins, welches aus der längeren Transkriptvariante resultiert. Das Protein besitzt sechs RNA-Bindedomänen, zwei vom Typ RRM (RNA recognition motif) und vier vom Typ KH (hnRNP-K homology domain). Nach [4].

IGF2BPs sind posttranskriptionelle Regulatoren der Genexpression. Die Bindung der IGF2BPs an ihre Zieltranskripte kann sich auf die Stabilität, die Translation und auf den Transport dieser Transkripte auswirken [29, 4]. Für den Großteil der Zieltranskripte von IGF2BP1 wurde eine stabilisierende Wirkung beschrieben. Lediglich für die leberspezifisch exprimierte, lange nicht-proteinkodierende RNA (lncRNA) HULC, wurde eine destabilisierende Wirkung durch Bindung an IGF2BP1 berichtet [52]. Abbildung 4 zeigt eine Übersicht bekannter IGF2BP1-Zieltranskripte und die nachgewiesenen Wirkungen der IGF2BP1-Bindung auf diese Transkripte.

IGF2BP1- und IGF2BP3-Expression konnte in embryonalen Geweben und einer Vielzahl verschiedener Tumore nachgewiesen werden. Mit Ausnahme der Keimdrüsen [47], ist die Expression dieser Gene in gesunden Geweben erwachsener Individuen jedoch weitgehend nicht vorhanden [4]. Daher kann das Expressionsmuster von IGF2BP1 und IGF2BP3 als onkofötal beschrieben werden. Die Expression von IGF2BP2 konnte hingegen auch in vielen gesunden Geweben erwachsener Individuen nachgewiesen werden. Für dieses Paralog wurden bisher nur

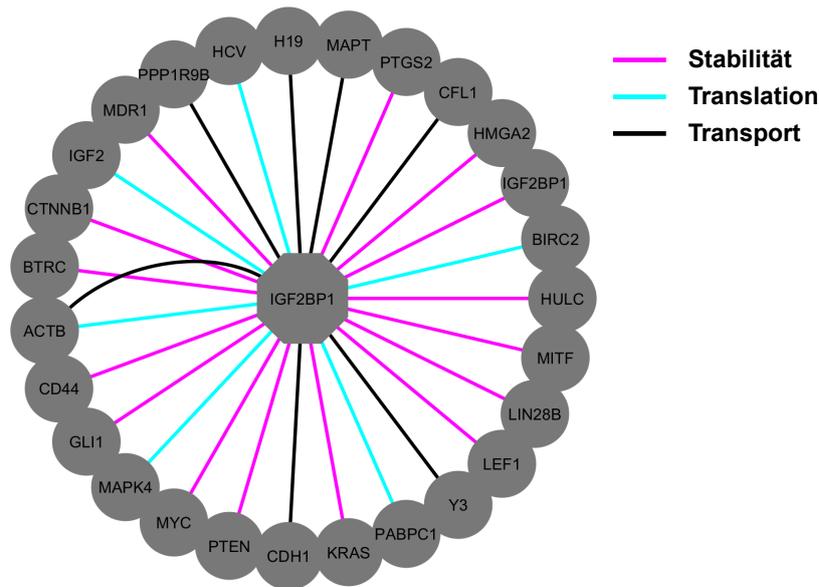


Abbildung 4: Gensymbole bekannter IGF2BP1-Zieltranskripte. Die Bindung einer RNA an das IGF2BP1-Protein kann sich auf die Stabilität, die Translation und/oder den Transport dieser RNA auswirken. Nach [4, 127, 52, 43, 31, 75] und [11]

wenige Hinweise auf eine onkogene Funktion gefunden, jedoch wurden Nachweise für eine Beteiligung des Proteins an Typ 2 Diabetes erbracht [4].

Das IGF2BP1-Protein wird in diversen menschlichen Tumoren de-novo synthetisiert. Es fördert die Expression der klassischen Onkogene MYC und KRAS, welche Ziele der tumorsuppressiven microRNA-Familie let-7 sind [4]. Es konnte gezeigt werden, dass die Bindung an IGF2BP1 den let-7-vermittelten Abbau der Transkripte der Onkogene LIN28B, HMGA2 und IGF2BP1 selbst reduziert, wodurch wiederum die LIN28B vermittelte Reduktion der Expression von let-7-microRNAs erhöht wird [11]. Darüberhinaus konnte gezeigt werden, dass IGF2BP1 die mRNA von β TrCP1 vor dem Abbau der microRNA miR-183 schützt [26]. In [43] wurde berichtet, dass IGF2BP1 durch Bindung an die mRNA des Transkriptionsfaktors MITF, diese vor dem Abbau durch miR-340 schützt, indem diese Bindung den Zugang der microRNA zu dem Transkript verhindert. Zusammen kann dies darauf hinweisen, dass IGF2BP1-Bindung einen generellen Mechanismus zum Schutz vor microRNA-vermitteltem Abbau darstellt.

Viele Beobachtungen deuten darauf hin, dass IGF2BP1 auch ein wichtiger Faktor für die gerichtete Bewegung von Tumorzellen ist. Die Bildung von Membranausstülpungen, wie Lamellipodien und Filopodien, erfordert die koordinierte Polymerisierung und Depolymerisierung von Aktinfasern an der Zellmembran. ZBP1, das Hühnerortholog von IGF2BP1, bindet

mRNAs von β -Actin und vermittelt deren Transport zu den Lamellipodien [29]. Weiterhin konnte gezeigt werden, dass die Bindung der β -Actin-mRNA an ZBP1 die Translation dieses Proteines verhindert. Die Translationsinhibierung kann jedoch an Lamellipodien durch Phosphorylierung des ZBP1-Proteins durch die Src-Kinase aufgehoben werden [54]. Ebenso konnte gezeigt werden, dass die mRNA von Cofilin, einem Protein, welches die Depolymerisierung von Aktinfasern vermittelt und somit freie Monomere für neu entstehende Fasern bereitstellt, durch Bindung an IGF2BP1 zu Lamellipodien transportiert wird [75]. Zellen, in denen die Bindung von IGF2BP1 an die mRNAs von β -Actin und Cofilin gestört wurden, zeigten ein reduziertes Migrationsvermögen ([29, 75]).

Ein weiterer Hinweis auf den Einfluss von IGF2BP1 auf das Aktinzytoskelett bildet der in [109] erbrachte Nachweis, dass die Bindung von IGF2BP1 an die mRNA von MAPK4 dessen Translation reduziert. Dadurch wird die Aktivierung von MK5 unterbunden, wodurch wiederum die Phosphorylierung von HSP27 unterdrückt wird. Phosphoryliertes HSP27 vermittelt den Abbau von Aktinmonomeren aus Aktinfasern. Es wurde weiterhin festgestellt, dass IGF2BP1 durch die Bindung an die mRNA von PTEN die Stabilität dieses Transkriptes erhöht und dadurch zu einer höheren Expression des PTEN-Proteins führt. Dies wiederum erhöht die Gerichtetheit der Zellmigration indem die vorhandene Zellpolarität erhalten wird [109]. Die Überexpression des GFP-fusionierten Hühner-Orthologen (GFP-chZBP1) in U-2OS-Zellen führte zu einer Erhöhung der mittleren Geschwindigkeit, sowie zu einer höheren Gerichtetheit. Durch transienten Knockdown von GFP-chZBP1/ IGF2BP1 in diesen Zellen wurde die Geschwindigkeit auf das Niveau von GFP-exprimierenden Kontrollzellen reduziert. Die Gerichtetheit der Zellen wurde durch den Knockdown jedoch nicht signifikant reduziert. Der Knockdown von PTEN führte hingegen zu einer signifikanten Reduzierung der Gerichtetheit in den GFP-chZBP1-überexprimierenden U-2OS-Zellen, nicht jedoch zu einer Reduktion der Geschwindigkeit. Die, für dieselbe Arbeit durchgeführte, Überexpression von GFP-chZBP1 in glioblastomabgeleiteten U-251-Zellen führte ebenfalls zu einer Erhöhung der Migrationsgeschwindigkeit, jedoch war keine Veränderung der Gerichtetheit feststellbar. Dies wurde auf eine fehlende Expression von PTEN in dieser Zelllinie zurückgeführt. Weiterhin führte der siRNA-vermittelte IGF2BP1-Knockdown in U-2OS und ovarialkarzinomabgeleiteten ES-2-Zellen in Scratch-Assay-Experimenten zu einem reduzierten Wundschluss. Die Überexpression von GFP-chZBP1 in U-2OS-Zellen führte hingegen zu einem beschleunigten Wundschluss [109]. Neben einem reduziertem Migrationsvermögen konnte in [127] durch transienten IGF2BP1-Knockdown eine reduzierte Expression des mesenchymalen Markergens Fibronectin in verschiedenen tumorabgeleiteten Zelllinien beobachtet werden. Zudem führ-

te die Reduktion der IGF2BP1-Expression zu einer verstärkten Lokalisation der epithelialen Markerproteine E-Cadherin und β -Catenin an Zell-Zellkontaktstellen. Daraus wurde geschlossen, dass IGF2BP1-Expression mesenchymale Eigenschaften aufrechterhält, beziehungsweise verstärkt [127].

Für Zervixkarzinomabgeleitete HeLa-Zellen wurde gezeigt, dass eine kombinierte Reduktion der IGF2BP1- und IGF2BP3-Expression zu einer stark beeinträchtigten Adhäsionsfähigkeit und einer reduzierten Bildung von Invadopodien führt [117]. Diese Ausstülpungen der Zellmembran aktivieren perizelluläre Proteasen, welche die extrazelluläre Matrix um sie herum degradieren können. Die Phänotypen gingen mit reduzierter Expression des IGF2BP1-Zieltranskriptes CD44, sowie weiterer Transkripte einher, welche für adhäsions- beziehungsweise invasionsassoziierte Proteine kodieren.

Aus Untersuchungen an mammarkarzinomabgeleiteten Zellen aus Ratten wurde geschlossen, dass IGF2BP1 die Polarität der Zellen so stark festlegt, dass die Zellen nicht mehr auf chemotaktische Signale reagieren können. Obwohl erhöhte Migration prinzipiell auf ein verstärktes Potential zur Metastasenbildung hinweist, könnte IGF2BP1 dadurch hemmend auf Metastasenbildung wirken [69]. Die für diese Untersuchungen verwendeten Zellen zeigten *in vitro* eine schnellere, sowie gerichtete Migration, wenn sie stabil IGF2BP1 exprimierten. Im Gegensatz dazu war das Migrationspotential der IGF2BP1-exprimierenden Zellen nach Injektion in Brustfettpolster von Ratten im Vergleich zu GFP-exprimierenden Zellen deutlich reduziert. Die Reduktion des chemotaktischen Potentials durch IGF2BP1 könnte erklären, warum eine reduzierte Expression des Proteins in metastatischen Brustkrebszelllinien, sowie -tumoren beobachtet wurde [45]. Andererseits wurden auch prometastatische Wirkungen von IGF2BP1 beschrieben. So entwickelte ein hoher Prozentsatz gentechnisch veränderter Mäuse, welche das murine IGF2BP-Ortholog CRD-BP spezifisch in Brustgewebe exprimierten, Brusttumore [110]. In einigen der Mäuse konnte Metastasierung in andere Gewebe nachgewiesen werden. Untersuchungen an Patienten mit kolorektalen Karzinomen zeigten, dass Tumore mit hoher IGF2BP1-Expression häufiger zu Metastasenbildung führten [21].

Trotz einiger gegenteiliger Ergebnisse, deutet der Großteil der Untersuchungen daraufhin, dass IGF2BP1 das Migrationspotential von Tumorzellen erhöht. Die Auswirkungen von IGF2BP1 auf das Metastasierungspotential scheinen jedoch vom Tumorkontext abhängig zu sein [4]. Die meisten Ergebnisse der Untersuchungen zur Wirkung von IGF2BP1 auf das Migrationspotential von Tumorzellen wurden durch manuelle Datenauswertung gewonnen.

In Abschnitt 3.3 werden Ergebnisse bezüglich der Wirkung von IGF2BP1 auf das Migrationspotential von Tumorzellen vorgestellt, welche mit den, für die vorliegende Arbeit entwickelten, Methoden zur automatisierten Auswertung von Zellmigration gewonnen wurden.

1.3 Methoden der Bildverarbeitung

Viele Methoden zur Untersuchung von Zellmigration verwenden Mikroskope oder andere optische Instrumente, um Bilddaten zu erzeugen. Die digitale Bildverarbeitung beschäftigt sich mit der Manipulation, sowie der Informationsgewinnung aus digitalen Bildern. Mit ihr können objektive und reproduzierbare Ergebnisse aus Bilddaten gewonnen werden. Die automatisierte Auswertung mittels Algorithmen der Bildverarbeitung ermöglicht es zudem, eine Vielzahl von Daten innerhalb kurzer Zeit zu bearbeiten. Somit stellt die digitale Bildverarbeitung ein geeignetes Hilfsmittel für die automatisierte Auswertung von Zellmigrationsexperimenten dar. In den folgenden Abschnitten werden einige Methoden und Konzepte der digitalen Bildverarbeitung vorgestellt, welche Grundlagen für die, im Rahmen der vorliegenden Arbeit entwickelten, Applikationen darstellen.

1.3.1 Segmentierung

Digitale Bilder setzen sich aus Elementen (picture elements, kurz Pixel) zusammen, welche rasterartig aneinandergesetzt sind. Segmentierung bezeichnet den Prozess, alle Pixel anhand bestimmter Eigenschaften so zu gruppieren, dass eine vollständige und überlappungsfreie Aufteilung dieser in Teilbereiche des Bildes entsteht. Diese Teilbereiche werden Regionen oder Zusammenhangskomponenten genannt. Diesen kann in Abhängigkeit vom Bildkontext häufig eine gewisse Bedeutung zugeordnet werden. So repräsentieren Regionen oft reale Objekte oder Teile davon. Alle Pixel einer Region sind paarweise verbunden. Zwei Pixel p und \tilde{p} eines Bildes I sind verbunden, wenn es eine Folge von Pixeln p_0, p_1, \dots, p_n mit $p = p_0$ und $p_n = \tilde{p}$ gibt, für die gilt:

$$\forall i \in \{1, \dots, n\} : p_i \text{ ist benachbart zu } p_{i-1}$$

Die Menge der Pixel einer Region R , welche zu mindestens einem Pixel einer anderen Region benachbart sind, bilden die Kontur(en) der Region R . Für zweidimensionale Bilder werden in der Regel zwei verschiedene Nachbarschaften definiert. Die 4-Nachbarschaft $N_4(p)$ eines Pixels p mit Koordinaten (x, y) besteht aus den vier Pixeln, die horizontal und vertikal an p angrenzen, also aus den Pixeln mit Koordinaten

$$(x, y - 1), (x + 1, y), (x, y + 1), (x - 1, y)$$

Zu der 8-Nachbarschaft $N_8(p)$ gehören zusätzlich noch die diagonal angrenzenden Pixel, also diejenigen mit den Koordinaten

$$(x - 1, y - 1), (x + 1, y - 1), (x + 1, y + 1), (x - 1, y + 1)$$

Abbildung 5 illustriert diese Nachbarschaftsdefinitionen.

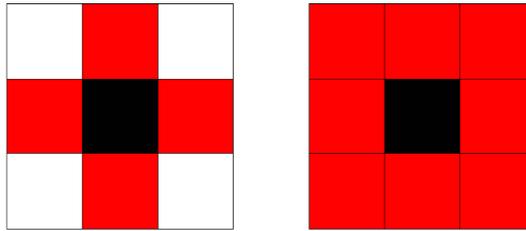


Abbildung 5: Nachbarschaften in zweidimensionalen Bildern. Links: 4-Nachbarschaft. Rechts: 8-Nachbarschaft. Nachbarpixel des schwarzen Pixels sind rot gekennzeichnet

Zu den am häufigsten verwendeten Verfahren für die Segmentierung von Bildern biologischer Zellen, gehört die Anwendung von Intensitätsschwellwerten [79]. Basierend auf der Verteilung von Intensitätswerten der Pixel des gesamten Bildes oder Teilbereichen, wird ein Schwellwert berechnet. Dieser legt fest, welche Pixel des Bildes oder Teilbereiches zu Objekten und welche zum Hintergrund gehören. Die zugrundeliegende Annahme ist, dass die Intensitätswerte von Pixeln, welche zu Objekten gehören, deutlich verschieden von Intensitätswerten der Pixel des Hintergrundes sind. Diese Annahme trifft für Mikroskopiebilder lebender Zellen jedoch oft nicht zu. Einen weiteren verbreiteten Ansatz zur Segmentierung von Zellbildern stellen sogenannte Region-Growing-Verfahren dar [79]. Ausgehend von initial gewählten Punkten werden Regionen iterativ entwickelt, indem benachbarte Pixel den bereits vorhandenen Regionen hinzugefügt werden, wenn sie gewisse Eigenschaften aufweisen, zum Beispiel ähnliche Intensitätswerte besitzen. Eine weitere häufig genutzte Klasse von Verfahren zur Segmentierung von Zellbildern stellen Aktive Konturen dar [79]. Bei diesen Verfahren werden die Konturen von Regionen, ausgehend von einer initial vorgegebenen Segmentierung, in iterativen Schritten entwickelt. Die Entwicklung basiert auf Minimierung eines zugrundeliegenden Energiefunktional. Durch die Wahl eines geeigneten Energiefunktional können sowohl Intensitäts-, als auch Formeigenschaften der zu segmentierenden Regionen modelliert werden. Aktive Konturen können entweder explizit als parametrisierte Funktionen oder implizit mit Hilfe einer

zusätzliche Funktion repräsentiert werden. Explizit repräsentierte aktive Konturen werden als Snakes [60], implizit repräsentierte als Level-Set-Verfahren [89] bezeichnet. Während die Topologie der mittels Snakes repräsentierten Regionen standardmäßig unveränderbar ist, sind Topologieänderungen bei einer Level-Set-Repräsentation inhärent möglich.

1.3.2 Level-Set-Verfahren

Level-Set-Verfahren sind numerische Algorithmen, mit denen die zeitabhängige Entwicklung von Grenzflächen beliebiger Dimensionalität modelliert werden kann. In der Bildverarbeitung werden sie für Segmentierungsaufgaben eingesetzt. Die zu entwickelnden Grenzflächen stellen dabei die Konturen von Regionen dar.

Level-Set-Verfahren repräsentieren Vordergrundregionen, deren Konturen, sowie den Hintergrund implizit durch die Werte einer Funktion ϕ . Diese Funktion ist auf dem gesamten Bildraum Ω definiert und wird als Level-Set-Funktion bezeichnet. Ein Level-Set ist die Menge aller Punkte von ϕ , welche den gleichen Funktionswert besitzen. Die Menge \mathcal{C} aller Konturpunkte wird in der Regel über die Menge der Punkte mit Wert null repräsentiert, das heißt:

$$\mathcal{C} = \{(x, y) | \phi(x, y) = 0\} \quad (1)$$

Weiterhin kann definiert werden, wie die Bereiche innerhalb und außerhalb der Vordergrundregionen des Bildes repräsentiert werden. Beispielsweise kann festgelegt werden, dass ϕ für alle Bildpunkte innerhalb von Vordergrundregionen größer als null und für alle Bildpunkte außerhalb dieser Regionen kleiner als null ist. Eine mögliche und übliche Funktion mit der diese Darstellung erzielt werden kann, ist die vorzeichenbehaftete Distanzfunktion. Zusammenfassend werden die Bestandteile des Bildes folgendermaßen dargestellt:

$$\begin{cases} \phi < 0 & : \text{außerhalb von Vordergrundregionen} \\ \phi > 0 & : \text{innerhalb von Vordergrundregionen} \\ \phi = 0 & : \text{auf der Kontur von Vordergrundregionen} \end{cases}$$

In Abbildung 6 werden diese Zusammenhänge schematisch dargestellt.

Durch die implizite Repräsentation der Konturpunkte ist die Anzahl der Regionen variabel. Konturen können im Zuge der Entwicklung der Level-Set-Funktion verschmelzen, aber auch auseinanderbrechen. Zudem können Löcher innerhalb der Regionen entstehen oder geschlossen werden.

Um Objekte in einem Bild mit Hilfe von Level-Sets zu detektieren, wird das Bild zu Beginn

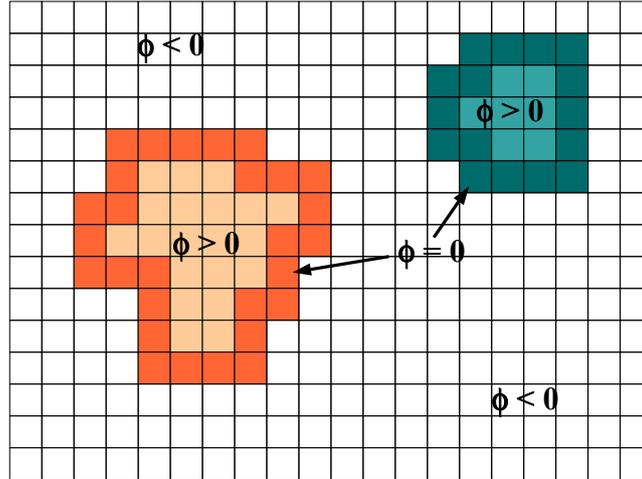


Abbildung 6: Repräsentation von Regionen und Konturen mit einer Level-Set-Funktion. Bildkoordinaten, an denen die Level-Set-Funktion ϕ Werte ≥ 0 annimmt gehören zu Regionen des Vordergrundes, wobei die Menge der Punkte mit $\phi = 0$ die Konturpunkte darstellt. Koordinaten mit $\phi < 0$ repräsentieren Hintergrundpixel.

mit einer beliebigen initialen Segmentierung in Hintergrund und Vordergrundregionen eingeteilt. Diese Initialsegmentierung wird anschließend in diskreten Schritten gemäß vorgegebener Bedingungen verändert. Dazu wird die Level-Set-Funktion ϕ zunächst um eine Dimension, die Zeit, erweitert:

$$\phi(x, y, t) : \Omega \times \mathbb{R}^+ \mapsto \mathbb{R} \quad (2)$$

Die initiale Segmentierung definiert die Level-Set-Funktion zum Zeitpunkt $t = 0$. Die Level-Set-Funktion wird anschließend iterativ durch Minimierung eines Energiefunktional in diskreten Zeitschritten entwickelt. Mit Hilfe des Energiefunktional können bestimmte Eigenschaften der zu segmentierenden Regionen modelliert werden. So werden Vorder- und Hintergrund mit der Energie von Chan und Vese [14] so modelliert, dass diese beiden Phasen einen konstanten, aber jeweils von der anderen Phase verschiedenen, Grauwert besitzen. Das Energiefunktional kann auch so konstruiert werden, dass die Entwicklung der Konturen in Bildbereichen mit hohen Intensitätsgradienten gestoppt wird, wie beispielsweise in [13]. Während der Entwicklung der Level-Set-Funktion muss für jeden Punkt (x, y) , welcher sich auf der Kontur befindet, zu jedem Zeitpunkt t gelten:

$$\phi(x, y, t) = 0 \quad (3)$$

Durch Differenzierung nach t erhält man daraus

$$\frac{\partial\phi}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial\phi}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial\phi}{\partial t} \cdot \frac{dt}{dt} = 0 \quad (4)$$

oder verkürzt

$$\frac{\partial\phi}{\partial t} + \nabla\phi \cdot \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = 0 \quad (5)$$

Die zeitliche Veränderung der Koordinaten der Konturpunkte, also deren Geschwindigkeit $F := (dx/dt, dy/dt)^T$ ergibt sich dabei aus dem vorgegebenen Energiefunktional, welches die der Bewegung zugrunde liegenden Eigenschaften modelliert. Zum Entwickeln der Kontur muss also das Anfangswertproblem

$$\begin{cases} \frac{\partial\phi}{\partial t} = -F \cdot \nabla\phi \\ \phi(x, y, 0) = \phi_0(x, y) \end{cases} \quad (6)$$

gelöst werden. Für einige Spezialfälle wurden Methoden entwickelt, welche auf das Lösen von Differentialgleichungen verzichten, zum Beispiel [108, 106].

1.3.3 Objektverfolgung

Um die Bewegung von Objekten in digitalen Videos zu charakterisieren, muss festgestellt werden, welche der segmentierten Regionen der Einzelbilder jeweils die selben realen Objekte repräsentieren. Die einfachste Methode der Zuordnung ist es, solche Regionen einander zuzuordnen, welche in aufeinanderfolgenden Bildern die geringste räumliche Distanz zu einander aufweisen [80]. Als Referenzpunkt für Distanzberechnungen wird in der Regel der Zentroid einer Region gewählt [82]. Der Zentroid (x_c, y_c) einer Region bestehend aus K Pixeln wird dabei wie folgt berechnet:

$$(x_c, y_c) = \frac{1}{K} \left(\sum_{k=1}^K x_k, \sum_{k=1}^K y_k \right), \quad (7)$$

wobei (x_k, y_k) die Koordinaten der zu der entsprechenden Region gehörenden Pixel sind.

Dieser Nächster-Nachbar-Ansatz kann jedoch bei einer hohen Objektdichte oder schnellen Bewegungen leicht zu falschen Zuordnungen führen. Die Zuordnungsgenauigkeit kann durch die Verwendung weiterer Informationen, wie beispielsweise Größe, Form oder Intensitätswerte der Objekte, erhöht werden.

Durch die Betrachtung der ermittelten Zentroiden zweier Einzelbilder als Knoten eines bipartiten Graphen, können Algorithmen, welche für die effiziente Zuweisung von Elementen zweier Mengen entwickelt wurden, für die Zuordnung von segmentierten Objekten verwendet werden. Hierfür werden Kosten zwischen Paaren, bestehend aus jeweils einem Element der einen Menge und einem Element der anderen Menge berechnet. Für die Zuordnung segmentierter Regionen können diese Kosten die Distanzen, sowie Form- und Intensitätswerte widerspiegeln. Die Zuordnung erfolgt durch Minimierung der summierten Gesamtkosten. Ein Beispiel für einen solchen Zuordnungsalgorithmus ist der Ungarische Algorithmus [65].

Für die Zuordnung schneller Objekte können auch statistische Analysen für die Berechnung der wahrscheinlichsten Korrespondenzen verwendet werden. Basierend auf zuvor beobachteten Werten, können, beispielsweise mit einem Kalman-Filter [57], Vorhersagen über wahrscheinliche Positionen von Objekten, basierend auf zuvor beobachteten Werten, getroffen werden. Detektierte Objekte, welche jeweils die geringste Distanz zu den berechneten Positionen aufweisen, können anschließend für die Zuordnung verwendet werden.

Werden aktive Konturen, wie Level-Set-Verfahren, für die Segmentierung verwendet, können Zuordnungen automatisch mit erstellt werden. Das Segmentierungsergebnis eines Bildes kann direkt als initiale Segmentierung des nächsten Bildes eines Videos genutzt werden. Dadurch entfällt ein separater Zuordnungsschritt. Diese Vorgehensweise eignet sich jedoch nur dann für die Zuordnung, wenn die segmentierten Regionen zweier aufeinanderfolgender Bilder eine gewisse Überlappung aufweisen.

Für die Zuordnung von Zellen in Mikroskopieaufnahmen muss in der Regel beachtet werden, dass Zellen im Verlauf des Videos den Betrachtungsbereich verlassen oder in ihm neu erscheinen können. Zellteilungsereignisse sind ein weiterer Faktor, welcher die Anzahl der zuzuordnenden Objekte variieren lässt. Dies muss bei der Auswahl eines geeigneten Zuordnungsalgorithmus beachtet oder durch Modifikationen ermöglicht werden.

1.3.4 Klassifikation mit Support Vector Machines

Support Vector Machines (SVMs) stellen eine Klasse von Techniken zur Lösung von Aufgaben aus dem Bereich des maschinellen Lernens dar. SVMs werden verwendet, um Klassifikations- oder Regressionsprobleme zu lösen, aber auch Anwendungen aus dem Bereich des unüberwachten Lernens können mit SVMs bearbeitet werden. In der Bildverarbeitung wurden SVMs beispielsweise für die Erkennung von Gesichtern oder handgeschriebenen Ziffern eingesetzt [6]. Für die Klassifikation von Objekten, also deren Einteilung in unterschiedliche Klassen, werden die Objekte durch bestimmte Eigenschaften repräsentiert. Diese Eigenschaften werden

zu Merkmalsvektoren der Objekte zusammengefasst. Werden n unterschiedliche Eigenschaften verwendet, bilden diese also einen n -dimensionalen Merkmalsvektor, welcher als Punkt in einem n -dimensionalen Raum betrachtet werden kann. Um einen Merkmalsvektor mit Hilfe einer SVM einer von zwei möglichen Klassen zuzuordnen, muss die SVM zunächst trainiert werden. Hierfür werden weitere Merkmalsvektoren benötigt, deren Klassenzugehörigkeiten im Vorfeld bekannt sein müssen. Die Grundidee von SVMs ist es, eine Hyperebene zu finden, welche die Merkmalsvektoren beider Klassen voneinander trennt. Diese Hyperebene wird zudem so konstruiert, dass ihr Abstand zu den Merkmalsvektoren der beiden Klassen maximal ist. Um eine solche separierende Hyperebene zu finden, werden zwei zueinander parallel verlaufende Hyperebenen, die Randebenen, gesucht, welche jeweils die nächsten Merkmalsvektoren der beiden Klassen tangieren. Diese Merkmalsvektoren werden als „support vectors“ bezeichnet. Die separierende Hyperebene befindet sich in der Mitte der beiden Randebenen. Abbildung 7 zeigt das Prinzip der Klassifikation mittels Hyperebenen für zweidimensionale Daten.

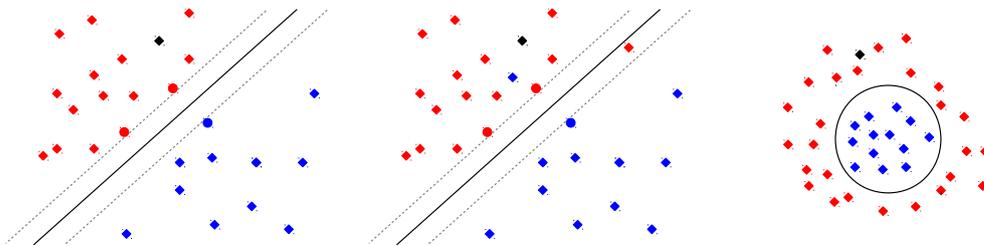


Abbildung 7: Prinzip der Klassifikation mit Hyperebenen am Beispiel zweidimensionaler Daten. Links: Die Punkte der beiden Klassen (rot und blau) können durch eine zweidimensionale Hyperebene, also eine Gerade (durchgezogene Linie), voneinander getrennt werden. Die Lage der Geraden maximiert den Abstand zu den beiden Randgeraden (gestrichelte Linien), welche die jeweils nächsten Punkte („support vectors“, Kreise) der beiden Klassen tangieren. Ein neuer Datenpunkt (schwarzes Karo) wird gemäß seiner Lage zur Geraden in eine der beiden Klassen eingeordnet. Mitte: Mit Hilfe eines Regularisierungsparameters können Ausreißer in den Daten zugelassen werden. Rechts: Beispiel für in 2D nicht linear separierbare Daten. Durch Projektion der Daten in eine höhere Dimension können die Daten mit Hilfe einer Hyperebene separiert werden. Die Rückprojektion dieser Hyperebene ergibt im 2-dimensionalen Raum jedoch nicht notwendigerweise eine Gerade.

Zwei Mengen von Merkmalsvektoren, welche mittels einer Hyperebene in zwei Klassen eingeteilt werden können, werden als linear separierbar bezeichnet. SVMs können jedoch auch auf nicht linear separierbare Daten angewendet werden. Hierfür verwenden sie Kernel-Funktionen, welche die gegebenen Merkmalsvektoren in einen höherdimensionalen Raum projizieren, in

welchem eine lineare Separierung möglich ist. Es kann gezeigt werden, dass für jede Menge von Merkmalsvektoren mit konsistenten Klassenmarkierungen eine Kernel-Funktion existiert, welche eine lineare Separierung ermöglicht [86]. Konsistent bedeutet hierbei, dass es in der Menge der Merkmalsvektoren keine zwei identischen Vektoren mit unterschiedlichen Klassenmarkierungen gibt.

Um mit Ausreißern in den Daten umgehen zu können, verfügen SVMs über einen Regulierungsparameter. Mit diesem kann die Anzahl und Entfernung von Merkmalsvektoren, welche sich nicht auf der korrekten Seite der Hyperebene befinden, grob reguliert werden [86].

Aufgrund ihrer Eigenschaften sind SVMs sehr flexibel und auf eine Vielzahl von Klassifikationsproblemen anwendbar.

1.3.5 MiToBo

MiToBo (Microscope Image Analysis Toolbox, (www.informatik.uni-halle.de/mitobo) ist eine Sammlung elementarer, sowie komplexerer Bildverarbeitungs- und Bildanalysealgorithmen. Das Hauptaugenmerk der Entwicklungen für MiToBo liegt dabei auf der Auswertung biomedizinischer Bilddaten [83]. MiToBo wird als Erweiterung für die weitverbreiteten Bildanalyseplattformen ImageJ [104] und Fiji [102] entwickelt. Dadurch steht MiToBo für eine Vielzahl von Betriebssystemen (Windows, Linux, Mac OS) zur Verfügung. Darüber hinaus erlaubt die Integration in ImageJ beziehungsweise Fiji die Verwendung einer Vielzahl unterschiedlicher Bildformate als Ein- und Ausgabe. Ausführbare Programmeinheiten von MiToBo werden als Operatoren bezeichnet. Diese können per Konsole oder über grafische Menüs ausgewählt und konfiguriert werden.

MiToBo baut auf der Java-Bibliothek Alida [84] auf und erbt grundlegende Konzepte dieser Bibliothek. Dazu gehören die Möglichkeit der automatischen Dokumentation aller Verarbeitungsschritte eines Operators, sowie einheitliche Schnittstellen zur Definition von Ein- und Ausgabeparametern. Diese Schnittstellen erlauben die automatisierte Generierung grafischer Benutzeroberflächen für die Konfiguration der Operatoren. MiToBo wird unter der GNU General Public License, Version 3, lizenziert und kann somit ohne Einschränkungen verwendet, verändert und verbreitet werden.

Die für die vorliegende Arbeit entwickelten Verfahren zur Analyse von Zellmigrationsvorgängen sind als MiToBo-Operatoren implementiert worden und können somit kostenlos benutzt und modifiziert werden.

1.4 Hochdurchsatz-RNA-Sequenzierung

Die in Abschnitt 1.1.5 vorgestellten Methoden zur Untersuchung der Zellmigration ermöglichen es, den Einfluss verschiedener Faktoren auf Zellmigration zu untersuchen. Mit Hilfe von Genexpressionsanalysen kann zudem die Wirkung dieser Faktoren auf die Synthese von Genprodukten betrachtet werden. Die Kombination dieser verschiedenen Daten kann dazu genutzt werden, den beobachteten Migrationsphänotypen Genexpressionsprofile zuzuordnen und damit zu einem besseren Verständnis der Wirkung der betrachteten Faktoren beitragen. Durch die Entwicklung von Verfahren, welche die parallele Sequenzierung von Millionen kurzer DNA-Moleküle erlauben (Next Generation Sequencing), wurde es möglich, komplexe eukaryotische Genome innerhalb weniger Tage vollständig zu sequenzieren. Die Kosten für die Sequenzierung eines Genoms in der Größe des menschlichen Genoms ($\approx 3 \cdot 10^9$ bp) betragen im Jahr 2015 nur noch ungefähr 1000 € (www.genome.gov/sequencingcosts/). Angewendet auf cDNA-Moleküle, welche durch reverse Transkription aus RNA-Transkripten einer Probe gewonnen wurden, kann ein Einblick auf das Transkriptom dieser Probe gewonnen werden. Diese Hochdurchsatz-RNA-Sequenzierungsverfahren, kurz RNA-Seq, erlauben die Identifikation der exprimierten Transkripte, sowie die Quantifizierung von Expressionsänderungen unter verschiedenen Bedingungen oder Entwicklungsstadien. Im Gegensatz zur Erstellung von Transkriptomprofilen mittels Micro-Arrays, können durch RNA-Seq Sequenzvariationen, wie beispielsweise Einzelnukleotidpolymorphismen (SNPs), aber auch unbekannte Transkripte gefunden werden. Darüber hinaus ist der dynamische Bereich für Transkriptquantifizierung bei RNA-Seq-Verfahren wesentlich größer als bei Micro-Arrays [121].

Probenvorbereitung Um das Transkriptom einer Probe mittels RNA-Seq zu bestimmen, wird eine cDNA-Bibliothek der Probe erstellt. Dazu wird zunächst die gesamte RNA aus den Zellen der Probe extrahiert. In einem Sequenzierdurchlauf kann nur eine begrenzte Anzahl von Sequenzen erzeugt werden. Daher ist es sinnvoll den Anteil solcher RNA-Moleküle zu erhöhen, welche für das konkrete Experiment von Interesse sind. So besteht der Großteil der RNA-Moleküle einer Zelle aus ribosomalen RNAs (rRNA, [116]). Diese sind jedoch für viele RNA-Seq-Experimente von untergeordneter Bedeutung und ihr Anteil wird häufig durch Bindung an spezifische Sonden stark reduziert. Eine weitere häufig verwendete Strategie besteht darin, Moleküle zu extrahieren, welche eine Poly-A-Sequenz tragen. Dadurch werden mRNAs angereichert, aber auch andere RNA-Moleküle mit einer poly-A-Sequenz, wie zum Beispiel viele long-noncoding RNAs (lncRNAs). Andererseits tragen beispielsweise Histon-mRNAs keine Poly-A-Sequenz an ihrem 3'-Ende. Für mRNAs von Prokaryoten ist

diese Methode generell ungeeignet, da diese Transkripte nicht polyadenyliert werden [116]. Während des Sequenziervorgangs werden nur kurze ($\approx 50 - 500$ bp) Bereiche der cDNA-Moleküle sequenziert. Dies geschieht, da die Zuverlässigkeit der für die Sequenzierung eingesetzten Polymerasen mit zunehmender Sequenzlänge abnimmt. Um möglichst vollständige Sequenzen langer RNA-Moleküle zu erhalten, werden diese deshalb häufig fragmentiert. Dadurch erhält man überlappende Teilsequenzen der kompletten Transkripte.

Die RNA-Fragmente werden mittels reverser Transkription in cDNA-Fragmente umgeschrieben und unterschiedliche Oligonukleotidadaptoren werden an die 5'- und 3'-Enden der Fragmente ligiert. Diese Adaptoren werden für die Immobilisierung der Fragmente für die Sequenzierungsschritte benötigt. Zudem können sie als Primer für die Sequenzierreaktionen, sowie für gegebenenfalls vorher stattfindende Amplifikationen der Transkripte mittels Polymerasekettenreaktion (PCR), dienen. Die Reihenfolge der Bearbeitungsschritte, die von einsträngigen RNA-Transkripten zu doppelsträngigen kurzen cDNA-Molekülen mit angehängten Adaptoren führen, hängt vom verwendeten Protokoll ab. Dabei kann jeder Schritt zu spezifischen Verzerrungen der Komposition der fertigen Fragmentbibliothek führen [116]. Einige Protokolle erlauben es, die Orientierungsinformation der Transkripte beizubehalten, indem verschiedene Adaptoren spezifisch an die 5'- oder 3'-Enden der Fragmente ligiert werden. Dadurch ist es möglich zu bestimmen, von welchem DNA-Strang die Transkripte abgelesen wurden. Sollen nur sehr kurze RNA-Moleküle, wie beispielsweise microRNAs, sequenziert werden, entfällt der Fragmentierungsschritt. Stattdessen ist eine Größenselektion der zu sequenzierenden Transkripte, zum Beispiel mittels Gelelektrophorese, nötig. Größenselektion kann auch verwendet werden, um die Bibliotheken auf Fragmente gewünschter Längen zu reduzieren. Dadurch können nichtligierte Adaptoren, sowie zu lange Fragmente von der Sequenzierung ausgeschlossen werden.

Sequenzierung Für die Hochdurchsatzsequenzierung der Bibliotheken existieren verschiedene Ansätze, wobei momentan vor allem Techniken beziehungsweise Sequenziermaschinen der Firmen Illumina, Life Technologies und Roche verwendet werden. Obwohl sich die technischen Details der Verfahren dieser Firmen unterscheiden, besitzen sie dennoch gemeinsame Prinzipien. Doppelsträngige cDNA-Moleküle der Bibliotheken werden denaturiert und immobilisiert, um die bei der Sequenzierung erzeugten Signale räumlich zuordnen zu können. Anschließend werden die immobilisierten Sequenzen amplifiziert, um eine höhere Signalstärke zu erreichen. Die amplifizierten einsträngigen DNA-Moleküle dienen als Templates für den Einbau komplementärer Nukleotide. Durch den Einbau werden nukleotidspezifische Fluoreszenz-

oder Biolumineszenzsignale erzeugt, welche optisch erfasst werden.

Dadurch können simultan Millionen kurzer Sequenzen, sogenannte Reads, erzeugt werden. Die Verfahren unterscheiden sich in der Anzahl und Länge der Reads, welche pro Durchlauf erzeugt werden können. Die produzierten Reads können außerdem danach unterschieden werden, ob sie nur von einem Ende der Fragmente erzeugt wurden, oder von beiden. Im ersten Fall spricht man von Single-End-Reads im zweiten Fall von Paired-End-Reads.

Verarbeitung Den erhaltenen Reads werden in der Regel für die weitere Verarbeitung als erstes Positionen eines Referenzgenoms zugeordnet. Dieser Vorgang wird als Mapping bezeichnet. Es existieren zahlreiche Algorithmen, welche das Mapping von Millionen von Reads, durch verschiedene heuristische Ansätze, in akzeptabler Zeit und Präzision ermöglichen. Für den Fall, dass kein Referenzgenom verfügbar ist, kann ein Transkriptom auch mit Hilfe überlappender Sequenzbereiche der Reads zusammengesetzt werden.

Ein großes Problem für die Zuordnung von Read-Sequenzen zu einem Referenzgenom stellen repetitive Sequenzen dar. Dabei handelt es sich um sehr ähnliche oder gar identische Sequenzen von unterschiedlichen Positionen des Genoms. Insbesondere wenn ihre Längen die der Reads überschreiten, kann es vorkommen, dass die Mappingalgorithmen Reads nicht mehr eindeutig zuordnen können. Aktuelle Schätzungen gehen davon aus, dass circa 50 % des menschlichen Genoms aus repetitiven Sequenzen bestehen [112]. Bei RNA-Seq-Experimenten reduziert sich das Problem auf Sequenzen, welche in RNA transkribiert werden. Allerdings kann es hierbei passieren, dass Reads Exon-Intron-Grenzen überspannen und deshalb vom Mappingalgorithmus aufgespalten werden. Die Teilsequenzen des Reads werden dann gegebenenfalls auf aufeinanderfolgende Exons gemappt. Die Wahrscheinlichkeit, dass Teilsequenzen der Reads nicht eindeutig zugeordnet werden können ist jedoch höher, da mit Abnahme der Sequenzlänge die Wahrscheinlichkeit für das Auftreten einer identischen Sequenz zunimmt. Die falsche Zuordnung von Reads eines RNA-Seq-Experimentes führt wiederum zu einer fehlerhaften Quantifizierung der betroffenen Gene. Mittels Paired-End-Sequenzierung können Mehrdeutigkeiten für das Mappen eines Reads durch die zusätzliche Sequenzinformation des zweiten Reads reduziert werden.

Durch Verwendung neuerer Sequenzierverfahren (3rd Generation Sequencing), welche deutlich längere Reads (> 1000 bp) produzieren und keine Amplifizierung der zu sequenzierenden Moleküle benötigen [85], können mehrdeutige Zuordnungen aufgrund repetitiver Sequenzbereiche reduziert werden.

Nachdem die Reads gemappt wurden, kann mit Hilfe bekannter Genannotationen die relative Abundanz eines Transkriptes durch die Anzahl der Reads, welche einen entsprechenden genomischen Bereich abdecken, abgeschätzt werden. Dabei ist jedoch zu beachten, dass nicht nur die Molekülanzahl einer bestimmten RNA, sondern auch deren Länge, sowie die Zusammensetzung der RNA-Population, einen Einfluss auf die Gesamtzahl der Reads des entsprechenden Gens haben [97]. Dies muss für Vergleiche durch Verwendung geeigneter Normalisierungsverfahren berücksichtigt werden.

Anwendungen Nach der Art der Anreicherung derjenigen RNA-Moleküle, welche sequenziert werden sollen, können verschieden RNA-Seq-Varianten unterschieden werden. Für ein Total-RNA-Seq-Experiment wird in der Regel lediglich die rRNA aus der Gesamt-RNA einer Probe abgereichert. Beim mRNA-Seq werden, wie oben beschrieben, spezifisch RNA-Moleküle angereichert, welche eine Poly-A-Sequenz enthalten. Für die Untersuchung der Expression von microRNAs werden per Größenselektion nur sehr kurze ($\approx 18 - 30$ nt) RNA-Moleküle für die Erstellung der Bibliotheken verwendet. Erfolgt die Größenselektion nach Adapterligation, werden entsprechend längere Moleküle extrahiert. Da die Selektion der RNAs ausschließlich über ihre Größe erfolgt ist jedoch davon auszugehen, dass sich neben microRNAs auch weitere kurze RNA-Moleküle in den so erstellten Bibliotheken befinden.

Eine spezielle Form des RNA-Seq zur spezifischen Sequenzierung von RNA-Molekülen, welche von RNA-bindenden Proteinen (RBPs) gebunden werden, ist CLIP-Seq. Hierbei wird eine Methode zur spezifischen Anreicherung von proteingebundenen RNA-Molekülen (CLIP - cross-linking and immunoprecipitation, [114]) mit einer Hochdurchsatzsequenzierung kombiniert. Dazu werden die zu untersuchenden Zellen zunächst mit UV-Licht bestrahlt, wodurch sich kovalente Bindungen zwischen Proteinen und RNA-Molekülen, welche sich in direktem Kontakt befinden, ausbilden. Nach dem Lysieren der Zellen werden die Proteine, deren RNA-Bindepartner untersucht werden sollen, mitsamt der an ihnen gebundenen RNA-Moleküle, per Immunpräzipitation herausgefiltert. Nach verschiedenen Aufreinigungsschritten werden die Proteine enzymatisch verdaut, wodurch nur noch ein kurzes Polypeptid an der Bindestelle verbleibt. Die RNA-Moleküle werden in cDNA-Moleküle umgeschrieben, welche anschließend sequenziert werden. Weiterentwicklungen wie PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation, [46]) und iCLIP (individual nucleotide resolution CLIP,[56]), erlauben eine höhere Effizienz in der Ausbildung der kovalenten Bindung zwischen Protein und RNA und eine präzisere Bestimmung der Bindestellen.

1.5 Zielstellung der Arbeit

Die Untersuchung der Migration von Tumorzellen leistet einen wichtigen Beitrag zur Erforschung von Krebserkrankungen. *In-vitro*-Experimente mit tumorabgeleiteten Zellen stellen dabei eine wichtige Grundlage dar. Mit ihrer Hilfe können Einflüsse verschiedener Faktoren auf die Zellmigration mit verhältnismäßig geringem Aufwand und reproduzierbar analysiert werden. Jedoch setzt dies eine präzise Auswertung der gewonnenen Daten voraus. Im Rahmen dieser Dissertation sollten Bildanalysemethoden entwickelt und implementiert werden, welche eine verlässliche und objektive automatisierte Auswertung von Mikroskopbildern aus Einzelzell-Tracking-Experimenten und Scratch-/Gap-Closure-Assays ermöglichen. Damit sollte insbesondere der Einfluss des RNA-bindenden Proteins IGF2BP1 auf die Migration von Tumorzellen untersucht werden. Für dieses onkofötal exprimierte Protein wurden bereits verschiedene Wirkungen auf Zellmigration berichtet. Die teilweise gegensätzlichen Ergebnisse dieser Studien basierten jedoch auf manueller Auswertung der erhobenen Daten. Die entwickelten Methoden zur automatisierten Auswertung von Zellmigrationsexperimenten sollten dazu verwendet werden, die Wirkung einer reduzierten IGF2BP1-Proteinexpression auf das migratorische Potential tumorabgeleiteter Zellen zu charakterisieren. Dies sollte der Validierung vorher publizierter Resultate dienen. Darüber hinaus sollten RNA-Seq-Daten verwendet werden, um die unter IGF2BP1-Knockdown beobachteten Migrationsphänotypen mit IGF2BP1-abhängiger Kontrolle der Genexpression zu korrelieren. Dabei sollte untersucht werden, ob die Genexpressionsdaten Rückschlüsse auf die von IGF2BP1 beeinflussten migrationsbezogenen zellulären Prozesse erlauben.

2 Material und Methoden

2.1 Zellkultur und Transfektionen

Zellen wurden in standardisiertem Nährmedium (DMEM) unter Zusatz von 10 % fötalem Rinderserum (FBS) bei 37°C und 5 % CO₂ kultiviert. Für IGF2BP1-Knockdown-Analysen wurden transiente Transfektionen mit, gegen die mRNA von IGF2BP1 gerichteten, siRNAs durchgeführt. Hierfür wurden die Zellen 24 h vor Durchführung anschließender Experimente mit Hilfe von Lipofectamine RNAiMax (Thermo Fisher, Waltham, USA) Transfektionsreagenz und je 75 nM siRNA transfiziert. Die Sequenzen der verwendeten siRNAs sind in Tabelle 1 aufgelistet.

Sämtliche Zellkulturarbeiten, sowie Transfektionen wurden von Dr. rer. nat. Nadine Bley und Claudia Misiak durchgeführt.

siRNA	Sequenz
Kontrolle (siC)	UUGUACUACACAAAAGUACUG
si1-1	CCGGGAGCAGACCAGGCAA
si1-2	UGAAUGGCCACCAGUUGGA
si1-P	CCGGGAGCAGACCAGGCAA UGAAUGGCCACCAGUUGGA CCAGGCAAGCCAUCAUGAAGCUGAA GGCUGCUCUCCUAUAGCUCCUUAUG GGGAAGAGCUGGAGGCCUA CCAUCCGCAACAUCACAAA AAGCUGAAUGGCCACCAGUUG AACACCUGACUCCAAAGUUCG GUAUGGUACAGUAGAGAAC CCUGAAGAAGGUAGAGCAA GUUCGUAUGGUUAUCAUCA GUGAACACCGAGAGUGAGA

Tabelle 1: Sequenzen der siRNAs, welche für transiente IGF2BP1-Knockdowns verwendet wurden.

2.2 Mikroskopie

Für Einzelzell-Tracking-Experimente wurden die Zellen in Ibidi 8 well μ -Slides (Ibidi, Martinsried) zu je circa 3000 Zellen pro Well eingesät. Die Böden der Wells wurden zuvor, sofern nicht anders angegeben, mit Collagen 1 beschichtet. Anschließend wurde den Zellen 24 Stunden Zeit gegeben sich abzusetzen. Die Aufnahmen erfolgten an einem Leica SP5 X Konfokalmikroskop (Leica Microsystems, Mannheim), welches mit einer Inkubationskammer und einem steuerbaren Objektisch ausgestattet war. Pro Well wurden jeweils zufällig zwei nicht-überlappende Aufnahmebereiche ausgewählt. Bei 37°C und 5% CO₂ wurden Bilder der Größe 1024 × 1024 mit 256 Graustufen in einem zeitlichen Abstand von fünf Minuten aufgenommen. Dabei wurden pro Aufnahmebereich jeweils neun bis 12 Z-Ebenen aufgenommen, da Unebenheiten der Beschichtungen dazu führten, dass sich die Zellen in unterschiedlichen Fokusebenen befanden. Aus den Aufnahmen der verschiedenen Z-Ebenen eines Aufnahmebereiches wurden vor der Weiterverarbeitung Maximalprojektionen mit Hilfe von ImageJ erstellt. Es wurde ein Objektiv mit 20-fach-Vergrößerung und einer numerischen Apertur von 0.5 verwendet.

Für Scratch-Assays wurden Zellen in 24-Well-Kulturplatten zu je circa $1 \cdot 10^5$ Zellen pro Well eingesät. Nach 24 Stunden wurden mit Pipettenspitzen Wunden in der konfluenten Zellschicht mittels Pipettenspitzen erzeugt. Nach drei Waschschritten zum Entfernen abgelöster Zellen und Zelltrümmern, wurde frisches Nährmedium hinzugefügt.

Gap-Closure-Assays wurden in Ibidi 8 well μ -Slides mit eingesetzten Ibidi Culture-Inserts durchgeführt. Diese aus Silikon gefertigten Einsätze erlauben das Einsähen zweier räumlich getrennter Zellpopulationen in ein Well. Pro Well wurden $4 \cdot 10^4$ Zellen eingesät. Nach 24 Stunden wurden die Einsätze entfernt und frisches Nährmedium hinzugefügt.

Durchlichtaufnahmen für Scratch-Assays erfolgten an einem Nikon Eclipse TS100-Mikroskop, welches mit einer Kamera vom Typ Nikon CoolPix 990 (Nikon Instruments, Tokyo) ausgestattet war. Pro Well wurde jeweils ein Aufnahmebereich ausgewählt. Die aufgenommenen 24-Bit-Farbbilder besaßen eine räumliche Auflösung von 2048 × 1536 Pixeln. Das gewählte Objektiv verfügte über eine 10-fach-Vergrößerung und eine numerische Apertur von 0.25. Da dieses Mikroskop nicht über eine Inkubationskammer verfügte, wurden die Kulturplatten für die zwischen den Aufnahmen liegende Zeit in einen Inkubator gestellt. Aufnahmen von Scratch- und Gap-Closure-Assays fluoreszenzmarkierter Zellen wurden genauso erstellt wie Aufnahmen für Einzelzell-Tracking-Experimente. Der Zeitraum zwischen zwei aufeinanderfolgenden Aufnahmen betrug jedoch jeweils 30 Minuten.

Die für die verschiedenen Fluoreszenzfarbstoffe beziehungsweise fluoreszierenden Proteine verwendeten Anregungs- und Detektionswellenlängen sind in Tabelle 2 aufgelistet.

Fluorochrom	Anregung (nm)	Detektionsbereich (nm)
GFP	489	505 - 560
Zs-Green1	493	500 - 550
RFP	584	600 - 650
E2-Crimson	611	630 - 750
CellMask Deep Red	658	670 - 750

Tabelle 2: Anregungs- und Detektionswellenlängen für die Aufnahmen der verwendeten Fluorochrome.

2.3 Topologieerhaltende Level-Set-Segmentierung

Energieminimierung Üblicherweise werden für die Entwicklung der Level-Set-Funktion ϕ partielle Differentialgleichungen (PDE) berechnet (s. Abschnitt 1.3.2). Dafür werden jedoch häufig numerische Stabilitätskriterien, sowie die Differenzierbarkeit der Level-Set-Funktion vorausgesetzt. Für Optimierungsprobleme, bei denen nicht die konkreten Werte von ϕ , sondern lediglich ihr Vorzeichen wichtig ist, wurde in [108] ein Algorithmus entwickelt, welcher auf das Lösen von Differentialgleichungen verzichtet. Mit diesem Algorithmus werden die Pixel eines Bildes in zwei Phasen eingeteilt, welche im Folgenden als Vorder- und Hintergrund bezeichnet werden. Bei diesem Verfahren wird das verwendete Energiefunktional direkt minimiert, indem überprüft wird, ob die Energie abnimmt, wenn ein Pixel des Vordergrundes dem Hintergrund zugeordnet wird oder umgekehrt. Ist dies der Fall, wird das entsprechende Pixel der anderen Phase zugeordnet. Dies wird für jedes Pixel des Bildes geprüft. Die Operation des Wechsels eines Bildpunktes in eine andere Phase wird als Flippen bezeichnet. Der Algorithmus ist auf Probleme anwendbar, die folgendermaßen definiert werden können:

$$\min_{\phi} E(H(\phi)) \quad (8)$$

Bei $H(\phi)$ handelt es sich um die Heaviside-Funktion von ϕ :

$$H(\phi) = \begin{cases} 0 & : \phi < 0 \\ 1 & : \phi > 0 \end{cases} \quad (9)$$

E ist ein beliebiges Funktional, das von der Heaviside-Funktion der Level-Set-Funktion abhängt und das, im Gegensatz zu PDE-basierten Methoden, nicht nach ϕ differenzierbar sein

muss. Der Entwicklungsschritt wird solange durchgeführt, bis sich die Energie von einer Iteration über alle Bildpunkte zur nächsten nicht mehr verändert. Der allgemeine Ablauf ist in Algorithmus 1 als Pseudocode dargestellt.

Algorithmus 1 non-PDE-LevelSetSolver(I, ϕ)

```
//  $I$ : Eingabebild
//  $\phi$ : initiale Level-Set-Funktion

// Initialisierung
berechne initiale Energie  $E$ 

// Entwicklung
 $E_{old} = 0$ 
while  $E_{old} \neq E$  do
   $E_{old} = E$ 
  for all Pixel  $p \in I$  do
    if  $E$  wird kleiner, wenn  $p$  in andere Phase geflippt würde then
      flippe  $p$  in diese Phase und aktualisiere  $E$ 
    end if
  end for
end while
```

Topologieerhaltung Gewöhnlich erlauben Level-Set-Verfahren Topologieveränderungen, also das Auseinanderbrechen oder Verschmelzen von Regionen, sowie die Entstehung oder das Schließen von Löchern innerhalb der Regionen während des Energieminimierungsprozesses. Diese Eigenschaft macht Level-Set-Verfahren flexibel in Bezug auf die Initialisierung, da im Vorfeld die genaue Anzahl der zu detektierenden Objekte nicht bekannt sein muss. Andererseits kann diese Eigenschaft aber auch unerwünscht sein, zum Beispiel wenn bereits mit der Initialisierung die Anzahl der zu detektierenden Objekte festgelegt werden soll. Um Topologieerhaltung zu ermöglichen wurde Algorithmus 1 um die Option erweitert, die Topologie der Vordergrundregion zu erhalten [41]. Hierfür wurde zunächst festgelegt, dass Vordergrundregionen 4-verbunden und der Hintergrund 8-verbunden ist. Um die initial vorgegebene Topologie zu erhalten, darf ein Pixel p nur dann in die jeweils andere Phase geflippt werden, wenn:

- p ein Konturpixel ist
- in der 8-Nachbarschaft von p genau eine Vordergrundregion vorkommt, welche in 4-Nachbarschaft zu p steht

Wäre p kein Konturpixel, würde ein Flippen vom Vorder- in den Hintergrund ein Loch in der Vordergrundregion erzeugen. Umgekehrt würde in diesem Fall das Flippen vom Hinter- in den

Vordergrund eine neue Vordergrundregion erzeugen. Gibt es keine Vordergrundregion in der 8-Nachbarschaft von p , welche zu diesem Pixel in 4-Nachbarschaft steht, würde das Flippen von p eine Vordergrundregion bestehend nur aus p auslöschen oder neu erzeugen. Wenn p ein Vordergrundpixel ist und es mehr als eine Vordergrundregion in der 8-Nachbarschaft von p gibt, die mit diesem Pixel jeweils in 4-Nachbarschaft stehen, dann ist p die einzige Verbindung zwischen diesen Regionen. Dadurch formen diese mit p eine gemeinsame Region. Das Flippen von p in den Hintergrund würde diese Region, welche in der 8-Nachbarschaft von p , sowie p selbst existiert, aufspalten. Dies könnte dazu führen, dass eine Vordergrundregion, welche im gesamten Bildraum existiert und p enthält, ebenfalls aufgespalten würde. Analog dazu würde das Flippen eines Hintergrundpixels p , dessen 8-Nachbarschaft mehr als eine Vordergrundregion, welche in 4-Nachbarschaft zu p stehen, möglicherweise separate Regionen zu einer verbinden.

Dieses Verfahren zur Topologieerhaltung ähnelt dem in [48] vorgestellten Verfahren, welches für die PDE-basierte Level-Set-Minimierung entwickelt wurde. Dabei ist eines der beiden Kriterien, welche dort verwendet wurden um Punkte zu bestimmen, welche der jeweils anderen Phase zugeordnet werden können ohne die Topologie der Phasen zu verändern (sogenannte „simple points“), identisch zu dem zweiten oben genannten Kriterium.

Das topologieerhaltende Level-Set-Verfahren mit Energieminimierung nach [108] wurde als Segmentierungsverfahren in den beiden im Folgenden vorgestellten Applikationen eingesetzt.

2.4 CellMigrationAnalyzer

Für die Auswertung der Bewegung einzelner Zellen in 2D-Einzelzell-Tracking-Experimenten wurde der MiToBo-Operator CellMigrationAnalyzer entwickelt. Ausgehend von Zeitraffervideos, werden fluoreszenzmarkierte Zellen detektiert und ihre Bewegungen, sowie Veränderungen der Zellformen und -intensitäten bestimmt.

Die Verarbeitung der Eingabevideos ist in verschiedene Schritte unterteilt. Abbildung 8 zeigt das Schema der Verarbeitungsschritte des CellMigrationAnalyzer. Nach Konvertierung der Bilder des Eingabevideos in 8-Bit Grauwertbilder, wird zunächst ein Gaußfilters auf jedes Einzelbild angewendet, um Hintergrundrauschen zu reduzieren. Im anschließenden Segmentierungsschritt werden die Bilder in Zellen und Hintergrund aufgeteilt. Darauf folgt die Berechnung von Zuordnungen für Objekte, welche in den Einzelbildern jeweils die selben Zellen repräsentieren. Diese Zuordnungen werden verwendet um Trajektorien der Zellen zu erstellen. Die Trajektorien ermöglichen es, das Migrationsverhalten der einzelnen Zellen zu charakterisieren.

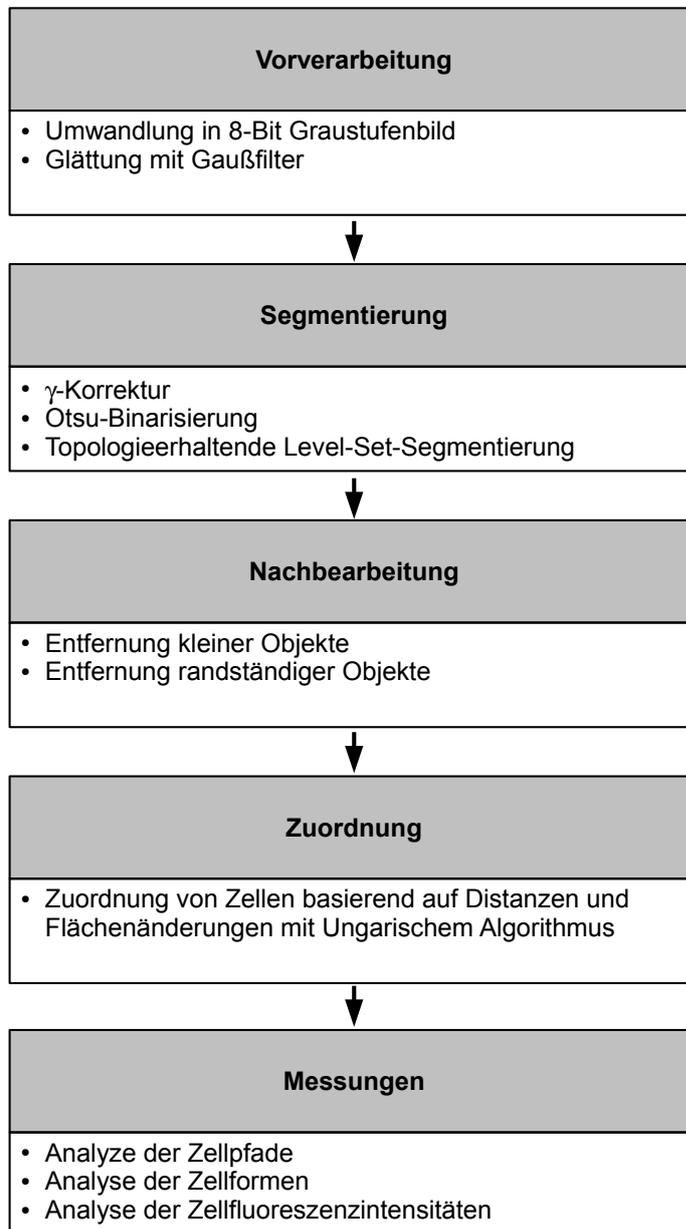


Abbildung 8: Verarbeitungsschritte des CellMigrationAnalyzer

Im Folgenden werden Segmentierung, Zuordnung und die Berechnung von Werten zur Charakterisierung des Migrationsverhaltens im Detail beschrieben.

Segmentierung Die Segmentierung basiert auf dem in Abschnitt 2.3 beschriebenen Level-Set-Verfahren, welches separat auf jedes Einzelbild des Eingabevideos angewendet wird. Die notwendige initiale Segmentierung für dieses Verfahren wird gewonnen, indem zunächst die Intensitätswerte des Bildes mit Hilfe einer Gammakorrektur ($\gamma < 1$) nichtlinear transformiert werden. Dadurch werden schwach fluoreszierende Zellen beziehungsweise Zellteile hervorgehoben. Je kleiner der Wert von γ gewählt wird, desto stärker ist die Hervorhebung. Jedoch kann es durch zu klein gewählte Werte von γ passieren, dass auch Hintergrundrauschen so sehr verstärkt wird, dass die anschließende Binarisierung unbrauchbare Ergebnisse liefert, da durch die Hervorhebung des Rauschens, dieses dem Vordergrund zugerechnet wird. Geeignete Werte von γ wurden im Rahmen der Parameteroptimierung bestimmt (s. Abschnitt 3.1.1). Auf dem transformierten Bild wird anschließend das globale, parameterfreie Schwellwertverfahren von Otsu [90] angewendet, wodurch ein Binärbild entsteht. Löcher in den entstandenen Vordergrundobjekten werden gefüllt und kleine Objekte, welche Zellfragmente darstellen könnten werden entfernt. Hierfür wird vor der Ausführung des Programms ein Schwellwert für die Mindestgröße festgelegt. Das Ergebnis ist eine grobe Segmentierung des Bildes in Zellen und Hintergrund (s. Abbildung 9), welche zur Initialisierung des topolgieerhaltenden Level-Set-Segmentierungsverfahrens verwendet wird. Das hierfür verwendete Energiefunktional wurde in [24] für die Segmentierung von fluoreszenzmarkierten Zellen vorgestellt und modelliert Vorder- und Hintergrund als zwei separate Normalverteilungen:

$$E(\phi(x, y)) = \int_{\Omega} e_1(x, y)H(\phi(x, y)) + e_2(x, y)(1 - H(\phi(x, y)))dxdy, \quad (10)$$

mit

$$e_i(x, y) = \log(\sigma_i^2) + \frac{(I(x, y) - \mu_i)^2}{\sigma_i^2}, i \in \{1, 2\} \quad (11)$$

Dabei bezeichnet I das Eingabebild, Ω die Bilddomäne und $H(\phi)$ die Heaviside-Funktion von ϕ (s. Gleichung 9). Die Symbole μ_i und σ_i^2 repräsentieren die Erwartungswerte beziehungsweise Varianzen der Normalverteilungen für Vordergrund ($i = 1$) und Hintergrund ($i = 2$):

$$\begin{aligned}
\mu_1 &= \frac{\int_{\Omega} I(x, y)H(\phi(x, y))dxdy}{\int_{\Omega} H(\phi(x, y))dxdy} \\
\mu_2 &= \frac{\int_{\Omega} I(x, y)(1 - H(\phi(x, y)))dxdy}{\int_{\Omega} (1 - H(\phi(x, y)))dxdy} \\
\sigma_1^2 &= \frac{\int_{\Omega} (I(x, y) - \mu_1)^2 H(\phi(x, y))dxdy}{\int_{\Omega} H(\phi(x, y))dxdy} \\
\sigma_2^2 &= \frac{\int_{\Omega} (I(x, y) - \mu_2)^2 (1 - H(\phi(x, y)))dxdy}{\int_{\Omega} (1 - H(\phi(x, y)))dxdy} \tag{12}
\end{aligned}$$

Die Erwartungswerte und Varianzen werden nach jedem Flippen neu berechnet. Im Gegensatz zu [24], wurden die Einschränkungen bezüglich Konturlänge und Fläche der Objekte hier nicht angewendet, wodurch das Energiefunktional keine freien Parameter besitzt. Das Energiefunktional aus Gleichung 10 erfüllt die Bedingung nach Gleichung 8 und ist daher geeignet, um mit dem in Abschnitt 2.3 beschriebenen Algorithmus minimiert zu werden.

An die Level-Set-basierte Segmentierung schließt sich ein Nachbearbeitungsschritt an. In diesem Schritt können Vordergrundregionen, deren Fläche eine definierte Anzahl Pixel unterschreitet, entfernt werden. Auf diese Weise werden Regionen, von denen vermutet wird, dass sie aufgrund ihrer geringen Größe keine Zellen darstellen, von der anschließenden Analyse ausgeschlossen. Weiterhin können Regionen entfernt werden, welche die Bildränder berühren. Dies ist in der Regel sinnvoll, da nicht festgestellt werden kann, welcher Teil einer Zelle über den Bildrand hinausragte, wodurch es zu einer fehlerhaften Bestimmung des Zentroiden der entsprechenden Region kommen würde.

Gelegentlich werden Zellen, welche im Bild sehr nah beieinander liegen, sich berühren oder überlappen, nicht als separate Regionen segmentiert. Aus Bildverarbeitungssicht handelt es sich dabei um Untersegmentierungen. In vielen solchen Fällen kann auch durch einen menschlichen Betrachter nicht eindeutig festgestellt werden, wo die Zellgrenzen verlaufen. Da solche Situationen zu fehlerhaften Messergebnissen führen, wurde ein Weg gesucht, mit solchen Untersegmentierungen umzugehen. Zwar existieren Methoden zur Detektion und Korrektur von Untersegmentierungen in Mikroskopieaufnahmen von Einzelzellen, zum Beispiel [119, 103], allerdings setzen diese Verfahren konvex geformte Zellflächen voraus. Da jedoch die Formen der hier untersuchten Zellen sehr komplex und variabel waren, hätten diese Verfahren das Problem nicht lösen können. Ein weiterer Lösungsansatz für das Problem der Untersegmentierung könnte darin bestehen, die Segmentierungsergebnisse eines Bildes direkt als Initialisierung für das darauffolgende zu verwenden. Dadurch würde zudem sowohl die Vorsegmentierung für alle bis auf das erste Bild eines Videos, als auch ein separater Zuordnungsschritt unnötig werden.

Aufgrund der Topologieerhaltung würden Zellen, sofern sie im ersten Bild des Eingabevideos als separate Regionen segmentiert wurden, auch wenn sie sich zu späteren Zeitpunkten berühren, als separate Regionen segmentiert. Jedoch würde dieses Vorgehen dazu führen, dass die Anzahl der detektierten Zellen über die Gesamtlänge des Videos konstant bliebe. Dadurch würden Zellen, welche im Verlauf des Videos neu im Bild auftauchen, nicht erfasst werden. Weiterhin würden Zellen, welche aus Zellteilungen hervorgehen, sowie solche, die aufgrund von Untersegmentierungen im ersten Bild fälschlicherweise zu einer Region zusammengefasst wurden, nicht detektiert werden. Zudem könnte dieses Vorgehen das Problem der Untersegmentierung sogar verstärken, indem Regionen durch die Topologieerhaltung weiterhin als einzelne Regionen segmentiert würden, obwohl sie bereits erkennbar voneinander getrennt sind.

Tatsächlich wirkt der hier verwendete Ansatz der Vorsegmentierung gekoppelt mit topologieerhaltender Level-Set-Segmentierung der Untersegmentierung entgegen, da die Fluoreszenzintensitäten der Zellen zu ihren Rändern hin in der Regel abnehmen. Dies führt dazu, dass bei der Vorsegmentierung die schwächer fluoreszierenden Ränder nicht erfasst werden, wodurch verhindert werden kann, dass naheliegende Zellen als eine Region segmentiert werden (s. Abbildung 9). Dennoch treten mit dem gewählten Segmentierungsverfahren Untersegmentierungen auf. Deshalb werden während des anschließenden Zuordnungsschrittes Größenveränderungen von Regionen in aufeinanderfolgenden Bildern eines Videos verwendet, um Untersegmentierungen zu detektieren.

Das Ergebnis des Segmentierungsschrittes ist ein Video, aufgebaut aus Binärbildern mit der gleichen räumlichen Auflösung wie das Eingabevideo und der gleichen Anzahl Einzelbilder. Die Vordergrundregionen des Videos repräsentieren dabei die detektierten Zellen.

Zuordnung Um Trajektorien für die individuellen Zellen zu erstellen, muss nach der Segmentierung bestimmt werden, welche der segmentierten Regionen in den einzelnen Bildern eines Videos jeweils dieselben Zellen repräsentieren. Dies geschieht mit Hilfe der Positionen und der Flächen der segmentierten Regionen. Als Erstes werden dazu die Vordergrundregionen des ersten Bildes mit individuellen Markierungen versehen. Daraufhin werden für alle Paare (I_t, I_{t+1}) von zeitlich aufeinanderfolgenden Bildern Zuordnungen der Vordergrundregionen des jeweils ersten Bildes I_t und des folgenden Bildes I_{t+1} gesucht. Dabei sind die Markierungen der Regionen von I_t bereits aus der vorhergehenden Iteration bekannt und nur Markierungen für I_{t+1} müssen neu bestimmt werden. Hierfür werden Kosten für alle Paare, bestehend aus

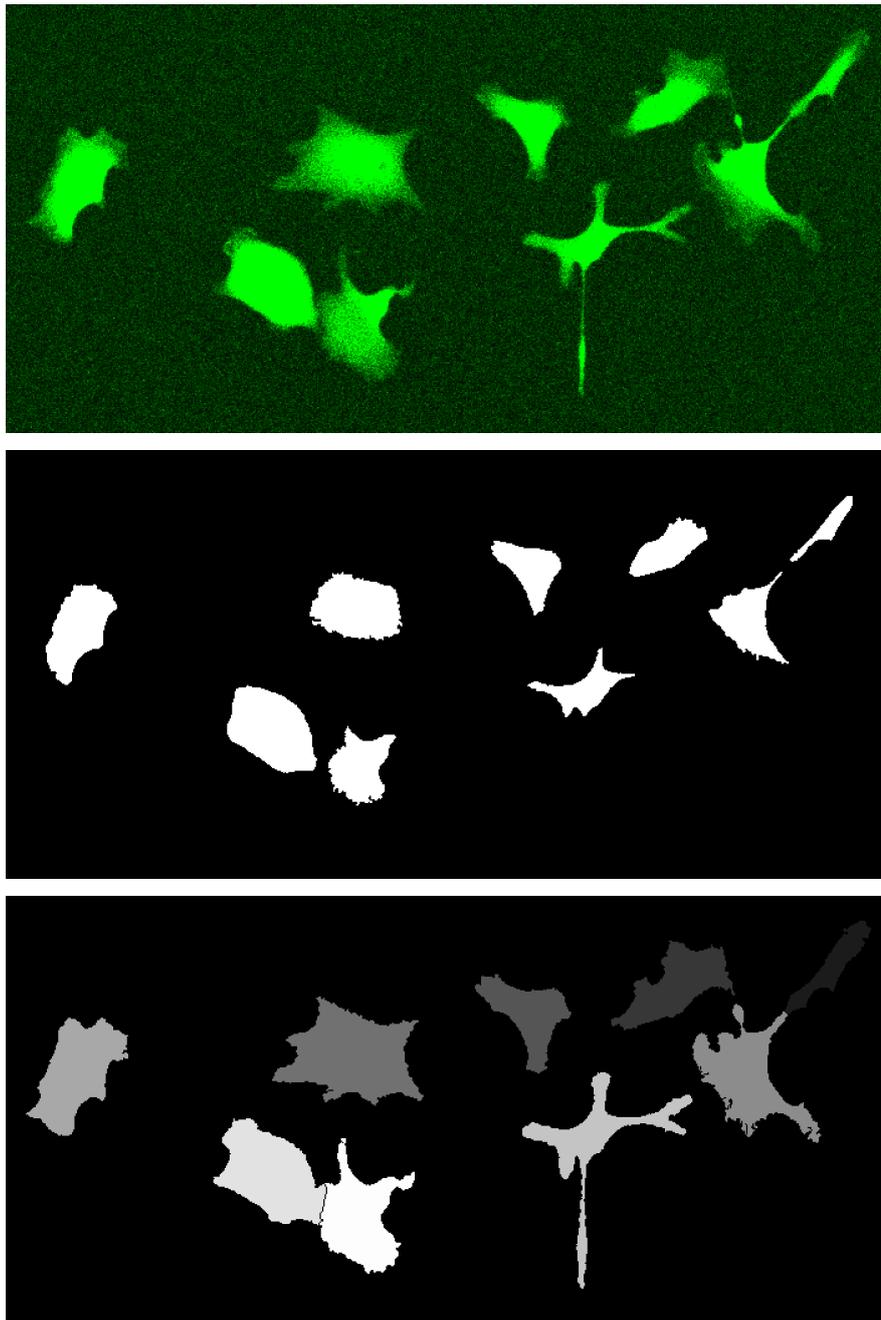


Abbildung 9: Beispielergebnisse des Segmentierungsschrittes des CellMigrationAnalyzer. Oben: Ausschnitt eines Bildes das GFP-exprimierende U-2OS-Zellen zeigt. Mitte: Ergebnis der Vorsegmentierung mittels Otsu-Binarisierung. Unten: Segmentierungsergebnis; die segmentierten Regionen sind in verschiedenen Graustufen dargestellt. Zur besseren Darstellung wurde der Kontrast der Bilder erhöht.

je einer Region i des Bildes I_t und einer Region j des darauffolgenden Bildes I_{t+1} berechnet. Diese Kosten $c(i, j)$ basieren auf den Euklid'schen Abständen $d(i, j)$ der Zentroiden der detektierten Zellen und dem Verhältnis ihrer Flächen A_i bzw. A_j und sind wie folgt definiert:

$$c(i, j) = \begin{cases} d(i, j), & \text{wenn } d(i, j) \leq T_d \wedge \frac{\max(A_i, A_j)}{\min(A_i, A_j)} \leq 1 + T_a \\ \infty, & \text{sonst} \end{cases} \quad (13)$$

Der Schwellwert T_d definiert die vermutete maximale Distanz, welche die Zellen in dem Zeitraum zurücklegen können, der zwischen der Aufnahme zweier Bilder vergangen ist. Diese Distanz hängt sowohl von der Geschwindigkeit der Zellen, als auch von der Bildaufnahme rate ab. In der Implementierung des CellMigrationAnalyzer wird T_d standardmäßig nach der Methode aus [58] automatisch berechnet. Dabei wird unter der Annahme, dass sowohl die Anzahl, als auch die räumliche Verteilung der Zellen im Eingabevideo keinen großen Schwankungen unterliegt, eine Distanz ermittelt, welche die zu erwartende Anzahl korrekt rekonstruierter Zuweisungen maximiert. Alternativ kann T_d auch vom Benutzer fest eingestellt werden. Der zweite Schwellwert T_a legt die vermutete maximale Flächenänderung der Zellen fest, welche aufgrund morphologischer Veränderungen der Zellen in der Zeit zwischen zwei Bildaufnahmen auftreten kann. Dieser Schwellwert dient der Behandlung von Untersegmentierungen, also dem Fall, dass mehrere Zellen fehlerhaft als eine Region segmentiert werden. Sollten Zellen, welche in einem Bild noch als separate Regionen segmentiert wurden, im darauffolgenden Bild, aufgrund zunehmender räumlicher Nähe, als eine Region segmentiert werden, dann wird die Fläche dieser zusammengesetzten Region in der Regel deutlich größer sein, als die Flächen der Einzelregionen. Wenn umgekehrt eine segmentierte Region aus mehreren Zellen besteht und sich im darauffolgenden Bild einzelne Zellen voneinander entfernt haben, wodurch diese als einzelne Regionen segmentiert werden, dann sollten deren Flächen deutlich kleiner sein. Treten Zellteilungen auf, dann sollten die Flächen der Tochterzellen kurz nach der Teilung ebenfalls kleiner sein, als die der entsprechenden Mutterzelle. Zusammengefasst bedeutet dies, dass deutliche Veränderungen der Flächen von detektierten Zellen als Hinweis für Untersegmentierung oder Zellteilungsereignisse betrachtet wurden. Der Schwellwert T_a definiert somit, welcher Flächenunterschied als deutlich gilt. Bei dessen Festlegung muss auch beachtet werden, in welchem Ausmaß die Fläche der betrachteten Zellen durch Formveränderungen im Zeitraum zwischen zwei Bildaufnahmen variieren kann. Damit hängt die Wahl von T_a , wie auch von T_d , von der Aufnahme rate der Bilder ab und beide müssen bei abnehmender Aufnahme rate tendenziell größere Werte bekommen. Ein Vergleich von tatsächlichen Flächenänderungen und Flächenänderungen aufgrund von Fehlsegmentierungen ist

in Abschnitt 3.1.1 gezeigt. Zwei Regionen aufeinanderfolgender Bilder, deren Distanzen oder Flächenunterschiede zu groß sind, sollen einander keinesfalls zugeordnet werden. Um dies zu kennzeichnen bekommen entsprechende Regionenpaare den Wert ∞ als Kosten zugewiesen. Die berechneten Kosten werden in einer Kostenmatrix zusammengefasst. Die Zeilen dieser Matrix repräsentieren dabei die Zellen des Bildes I_t , die Spalten die Zellen des Bildes I_{t+1} . Die eigentliche Zuordnung von Zellen erfolgt unter Verwendung des Ungarischen Algorithmus [65]. Dieser benötigt eine quadratische Kostenmatrix als Eingabe. Da die Anzahl der segmentierten Zellen zweier aufeinander folgender Bilder aber nicht notwendigerweise gleich ist, werden der Kostenmatrix zusätzlich Kosten für die Zuordnung detektierter Zellen zu Pseudoobjekten hinzugefügt. Die Anzahl der Pseudoobjekte wird dabei durch die Anzahl der detektierten Zellen des jeweils anderen Bildes bestimmt. Das bedeutet, wurden in Bild I_t N Zellen und in Bild I_{t+1} M Zellen detektiert, so wird eine quadratische Kostenmatrix der Größe $(N + M) \times (N + M)$ erzeugt. Zuordnungen zu Pseudoobjekten repräsentieren das Auftauchen oder Verschwinden von Zellen im Bild. Die Kosten für die Zuordnung zu Pseudoobjekten sind dabei jeweils gleich. Sie entsprechen eins plus dem Wert der größten Distanz zwischen zwei detektierten Zellen der Bilder I_t und I_{t+1} . Dadurch wird erreicht, dass Zellen, welche nach Gleichung (13) keiner anderen Zelle zugeordnet werden sollen, also zu allen detektierten Zellen des anderen Bildes jeweils die Kosten ∞ erhalten haben, stattdessen einem Pseudoobjekt zugeordnet werden. Zugeordnete Paare (i, j) , deren erstes Element i einem Pseudoobjekt aus Bild I_t und deren zweites Element j einer Zelle aus Bild I_{t+1} entspricht repräsentieren eine neue Zelle. Ist das erste Element eines zugeordneten Paares eine detektierte Zelle, das zweite Element jedoch ein Pseudoobjekt, bedeutet dies, dass eine Zelle verschwunden ist. Zwei einander zugeordnete Pseudoobjekte werden ignoriert. Die Ereignisse des Auftauchens neuer Zellen, sowie des Verschwindens von Zellen können eintreten, wenn Zellen tatsächlich neu im Bild auftauchen oder es verlassen. Allerdings führen auch Untersegmentierungen zu diesen Ereignissen. Nachdem die Zellen aus Bild I_t und I_{t+1} mit Hilfe des ungarischen Algorithmus einander zugeordnet worden sind, erhalten die Zellen aus Bild I_{t+1} Markierungen, die denen ihrer zugeordneten Zellen aus Bild I_t entsprechen. Wurde eine neue Zelle identifiziert, so bekommt sie eine noch nicht verwendete Markierung.

Nachdem Zuordnungen für alle Paare aufeinanderfolgender Bilder erstellt wurden, haben die individuellen Zellen Markierungen erhalten, welche sie im gesamten Video eindeutig kennzeichnen.

Migrationsanalyse Die Markierungen des Zuordnungsschrittes werden verwendet, um Trajektorien als Folge der Zentroidpositionen der Zellen zu generieren. Die Trajektorien, sowie die Segmentierungsergebnisse und Fluoreszenzsignale des Eingabevideos werden verwendet um Werte zum Beschreiben des Migrationsverhaltens, der Formveränderungen, sowie Veränderungen der Fluoreszenzintensitäten der detektierten Zellen zu berechnen.

Ausgehend von einer Folge von Zentroiden $((x_n(t), y_n(t)), t = 1, \dots, T)$ einer Zelle n werden folgende, die Bewegung beschreibende, Maße für diese Zelle berechnet:

Distanz:

$$dist_n(t, \tilde{t}) = \sqrt{(x_n(\tilde{t}) - x_n(t))^2 + (y_n(\tilde{t}) - y_n(t))^2} \quad (14)$$

Die Distanz beschreibt die Länge der Strecke, welche die Zelle n von Bild t zu Bild \tilde{t} zurückgelegt hat. Die Positionen der Zelle für die Zeit zwischen zwei aufeinanderfolgenden Bildern sind nicht bekannt. Deshalb ist zu erwarten, dass die tatsächlich zurückgelegte Strecke länger ist. Höhere Bildaufnahmezeiten helfen die Unterschätzung der Distanz zu reduzieren.

Gesamtdistanz:

$$dist_total_n = \sum_{t=2}^T dist_n(t-1, t) \quad (15)$$

Die Gesamtdistanz ist die Summe aller Distanzen, welche die Zelle n in allen Paaren aufeinanderfolgender Bilder zurückgelegt hat.

Nettodistanz:

$$dist_net_n = dist_n(1, T) \quad (16)$$

Die Nettodistanz ist die Länge der Strecke vom ersten zum letzten Zentroiden der Zelle n .

Gerichtetheit:

$$directionality_n = \frac{dist_net_n}{dist_total_n} \quad (17)$$

Die Gerichtetheit beschreibt wie geradlinig die Bewegung der Zelle war. Je größer dieser Wert ist, desto geradliniger war die Bewegung. Die Gerichtetheit somit kann als ein Maß für die Stetigkeit der Bewegungsrichtung einer Zelle betrachtet werden.

Geschwindigkeit:

$$speed_n(t, \tilde{t}) = \frac{dist(t, \tilde{t})}{(\tilde{t} - t) \cdot \Delta t} \quad (18)$$

Die Geschwindigkeit wird aus dem Quotienten der zurückgelegten Distanz der Zelle zwischen den Bildern t und \tilde{t} und der, zwischen diesen Aufnahmen vergangenen, Zeit berechnet.

Richtung:

$$dir_n(t, \tilde{t}) = \begin{cases} 360 - angle_n(t, \tilde{t}) & : (y_n(\tilde{t}) - y_n(t)) > 0 \\ angle_n(t, \tilde{t}) & : (y_n(\tilde{t}) - y_n(t)) \leq 0 \end{cases}, \quad (19)$$

$$dir_n \in [0, \dots, 360),$$

wobei

$$angle_n(t, \tilde{t}) = \frac{180}{\pi} \arccos \left(\frac{x_n(\tilde{t}) - x_n(t)}{\sqrt{(x_n(\tilde{t}) - x_n(t))^2 + (y_n(\tilde{t}) - y_n(t))^2}} \right), \quad (20)$$

dabei gilt $angle_n(t, t) = 0$ und $angle \in [0, \dots, 180]$

Die Richtung $dir_n(t, \tilde{t})$ beschreibt die Bewegungsrichtung der Zelle n von Bild t zu Bild \tilde{t} relativ zur horizontalen Bildachse.

Richtungsänderung:

$$dirdiff_n(t) = \begin{cases} 360 - diff_n(t) & : diff(t) > 180 \\ diff_n(t) & : diff(t) \leq 180 \end{cases}, \quad (21)$$

$$dirdiff_n \in [0, \dots, 180] \text{ und}$$

$$diff_n(t) = |dir_n(t, t+1) - dir_n(t+1, t+2)| \quad (22)$$

Die Richtungsänderung gibt die Differenz der Bewegungsrichtung zwischen zwei aufeinanderfolgenden Bildern an.

Persistenz:

$$persistence_n(t) = \frac{speed_n(t+1, t+2)}{1 + 100/360 \cdot dirdiff_n(t)} \quad (23)$$

Die Persistenz [30] beschreibt das Verhältnis von Geschwindigkeit und Richtungsänderung. Unter Verwendung der Segmentierungsergebnisse werden pro detektierter Zelle die folgenden Formparameter für jedes Bild berechnet:

Fläche: Die Anzahl der zu einer Zelle gehörenden Pixel, gegebenenfalls skaliert mit physikalischer Pixelgröße.

Umfang: Die Länge der äußeren Kontur einer Zelle, gegebenenfalls skaliert mit physikalischer Pixelgröße.

Länge: Die Länge der 1. Hauptachse der am besten an die Zellform anpassbaren Ellipse [92], gegebenenfalls skaliert mit physikalischer Pixelgröße.

Breite: Die Länge der 2. Hauptachse der am besten an die Zellform anpassbaren Ellipse [92], gegebenenfalls skaliert mit physikalischer Pixelgröße.

Rundheit:

$$\frac{4\pi \cdot \text{Fläche}}{\text{Umfang}^2} \quad (24)$$

Je kleiner der Wert der Rundheit ist, desto weniger ähnelt die Kontur der segmentierten Zelle einem Kreis.

Des Weiteren werden Intensitätswerte des Eingabevideos zusammen mit den Segmentierungsergebnissen verwendet, um pro Bild mittlere Fluoreszenzintensitäten der einzelnen Zellen zu bestimmen.

Die Ausgabe des Analyseschrittes sind Textdateien, welche sämtliche berechneten Maße als tab-separierte Werte enthalten. Durch dieses Format ist eine einfache Weiterverarbeitung der Daten gewährleistet.

Die Implementierung der einzelnen Schritte Segmentierung, Nachbearbeitung, Zuordnung und Migrationsanalyse erfolgte in jeweils separaten MiToBo-Operatoren (s. Abschnitt 1.3.5). Dadurch kann jeder Schritt gegebenenfalls durch einen alternativen Operator ausgetauscht werden.

2.5 ScratchAssayAnalyzer

Der zweite für die vorliegende Arbeit entwickelte MiToBo-Operator für die quantitative Auswertung von Zellmigrationsexperimenten ist der ScratchAssayAnalyzer [41, 42, 40]. Dieser Operator kann für die automatisierte Auswertung der Wundentwicklung von Scratch- oder Gap-Closure-Assays verwendet werden. Es können sowohl Durchlicht-, als auch Fluoreszenzaufnahmen analysiert werden. Aus Zeitrafferaufnahmen der Experimente werden die Wunden zu den verschiedenen Zeitpunkte detektiert und ihre Flächen berechnet. Die Bearbeitung eines Einzelbildes gliedert sich in die Schritte Vorverarbeitung, Segmentierung, Nachbearbeitung und Klassifikation. Diese Schritte wurden in einem weiteren MiToBo-Operator, dem ScratchAssaySegmenter, implementiert. Ein Schema dieses Operators ist in Abbildung 10 dargestellt.

Der ScratchAssayAnalyzer verwendet den ScratchAssaySegmenter um alle Bilder eines Eingab-

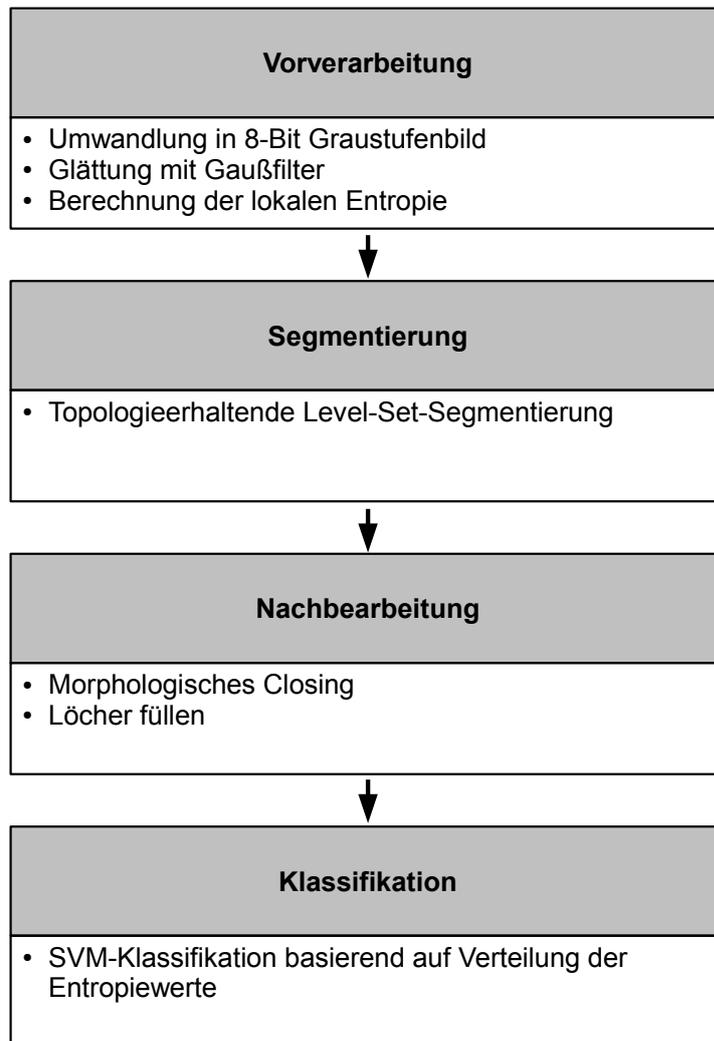


Abbildung 10: Verarbeitungsschritte des ScratchAssaySegmenter

bevideos zu analysieren und die Ergebnisse miteinander in Beziehung zu setzen. Im Folgenden werden die Verarbeitungsschritte näher erläutert.

Vorverarbeitung Als Erstes werden die Bilder des Eingabevideos unabhängig vom Bildtyp in 8-Bit Grauwertbilder konvertiert. Im Anschluss erfolgt eine Glättung der Bilder mit einem Gaußfilter, um Bildrauschen zu reduzieren. In Durchlichtmikroskopieaufnahmen sind die mittleren Grauwerte von Zellen und Wundfläche sehr ähnlich. Dadurch würde eine Detektion mittels Intensitätsschwellwerten keine zufriedenstellenden Ergebnisse liefern. Allerdings ist die Verteilung der Grauwerte in der Wunde deutlich homogener als in Bereichen, welche von Zellen besetzt sind (s. Abbildung 11).

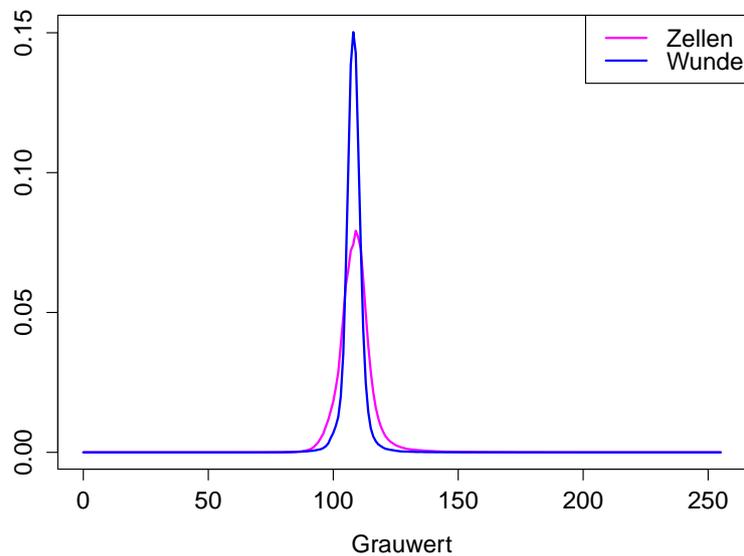


Abbildung 11: Relative Intensitätshistogramme von Zell- und Wundregion einer Durchlichtaufnahme eines Scratch-Assay-Experimentes. Der Mittelwert der Zellregion beträgt 108.8, der Mittelwert der Wundregion 108.2. Die Varianzen von Zell- und Wundregion betragen 46.5 beziehungsweise 18.0.

Ein Maß, welches diese Unterschiede reflektieren kann, ist die Shannon-Entropie \mathcal{H} :

$$\mathcal{H} = - \sum_{g=0}^{G-1} P(g) \cdot \log_2 P(g) \quad (25)$$

Hierbei bezeichnet $P(g)$ die Wahrscheinlichkeit für das Auftreten des Intensitätswerts g und G stellt die Anzahl unterschiedlicher Grauwerte dar.

Die Entropie wird für jedes Pixel (x, y) berechnet, indem Werte in einem Fenster um dieses

Pixel verwendet werden, um die Wahrscheinlichkeit des Grauwertes durch relative Häufigkeiten zu approximieren. Wenn Bereiche des Berechnungsfensters über die Bildgrenzen hinaus reichen, werden nur die innerhalb der Bildgrenzen liegenden Positionen für die Berechnung der relativen Häufigkeiten verwendet. Der so berechnete Entropiewert $\mathcal{H}(x, y)$ für ein Pixel (x, y) wird anschließend in ein neues Bild an Position (x, y) eingefügt. Als Ergebnis erhält man ein Bild, welches dunkle Bildpunkte an den Positionen aufweist, die im Eingabebild in Bereichen mit homogenen Grauwerten lagen, sowie helle Bildpunkte an den Positionen besitzt, die Bereichen mit hoher Variabilität der Grauwerte im Eingabebild entsprechen (s. Abbildung 12).

Segmentierung Das so entstandene Entropiebild dient als Eingabe für eine topologieerhaltende Zwei-Phasen-Level-Set-Segmentierung (s. Abschnitt 2.3). Als Energie wurde eine vereinfachte Variante des Energiefunktional von Chan und Vese [14] verwendet. Beim Chan-Vese-Modell wird davon ausgegangen, dass das Eingabebild aus zwei, nicht notwendigerweise zusammenhängenden, Phasen besteht, dem Vorder- und dem Hintergrund. Diese beiden Phasen sind jeweils durch einen konstanten, aber von dem der anderen Phase verschiedenen, Grauwert gekennzeichnet. Im Vergleich zum Funktional in [14], wurde auf die Terme verzichtet, welche die Länge, sowie die Fläche der Phasen einschränken. Das vereinfachte Energiefunktional ist folgendermaßen definiert:

$$\begin{aligned}
E(\phi(x, y)) &= \lambda_1 \int_{\Omega} |I(x, y) - \mu_1|^2 H(\phi(x, y)) dx dy \\
&+ \lambda_2 \int_{\Omega} |I(x, y) - \mu_2|^2 (1 - H(\phi(x, y))) dx dy
\end{aligned} \tag{26}$$

Die beiden Terme integrieren über den gesamten Bildraum Ω , wobei durch Multiplikation mit der Heaviside-Funktion $H(\phi)$, beziehungsweise $1 - H(\phi)$ nur jeweils die Bildpunkte Einfluss nehmen, welche zum Vorder- beziehungsweise zum Hintergrund gehören. Die mittleren Grauwerte μ_1, μ_2 von Vorder- und Hintergrund werden dabei wie folgt berechnet:

$$\mu_1 = \frac{\int_{\Omega} I(x, y) H(\phi(x, y)) dx dy}{\int_{\Omega} H(\phi(x, y)) dx dy}, \quad \mu_2 = \frac{\int_{\Omega} I(x, y) (1 - H(\phi(x, y))) dx dy}{\int_{\Omega} (1 - H(\phi(x, y))) dx dy} \tag{27}$$

Die beiden Gewichtungparameter für den Vorder- und Hintergrundterm (λ_1 beziehungsweise λ_2) wurden konstant auf eins gesetzt. Das Energiefunktional des Chan-Vese-Modells hat die Form $E(H(\phi))$ und kann daher mit dem in Abschnitt 2.3 genannten Algorithmus 1 ohne das Lösen von Differentialgleichungen minimiert werden. Darüber hinaus benötigt dieser Algorithmus bei Verwendung des Chan-Vese-Energiefunktional deutlich weniger Rechenzeit als

PDE-basierte Verfahren [108]. Als Initialisierung der Level-Set-Segmentierung dient ein horizontaler oder vertikaler Streifen, der sich über die gesamte Bildbreite oder -höhe erstreckt und jeweils 10 % der Bildhöhe beziehungsweise -breite einnimmt. Die Ausrichtung des Streifens ist abhängig von der Ausrichtung der Wunde und wird vor Ausführung des Programms vom Nutzer festgelegt. Während Zellen die Wundfläche besiedeln kann es vorkommen, dass zellfreie Bereiche hinter der eigentlichen Zellfront entstehen. Diese Bereiche sollten allerdings nicht als Vordergrund, das heißt als Wundfläche, detektiert werden. Aus diesem Grund wurde das in Abschnitt 2.3 beschriebene Verfahren zur Topologieerhaltung verwendet. Dadurch bleibt der initial vorgegebene Vordergrund während der gesamten Level-Set-Entwicklung eine Region.

Nachverarbeitung Während der Erstellung der Wunde können Zelltrümmer im Wundbereich entstehen. Dies kann dazu führen, dass Bereiche, welche sich innerhalb der Wunde befinden als Hintergrund segmentiert werden. Das Ergebnis der Segmentierung ist eine Region für die Wunde, innerhalb derer sich Hintergrundbereiche befinden, die aufgrund der topologieerhaltenden Eigenschaft des Segmentierungsalgorithmus über dünne Verbindungen mit dem restlichen Hintergrund verbunden sind (s. Abbildung 12). Um die Fehlsegmentierung dieser Zelltrümmer zu korrigieren, werden die dünnen Verbindungen im Anschluss an die Segmentierung durch die morphologische Closing-Operation entfernt. Dadurch entstehen jedoch Löcher in der Wundregion, welche anschließend geschlossen werden.

Zellen, welche sich vom Verband der übrigen Zellen gelöst haben können allein in die Wundfläche einwandern. Diese Zellen werden dann durch die Nachverarbeitungsschritte wie auch Zelltrümmer zur Wunde gerechnet. Dies könnte als Segmentierungsfehler betrachtet werden, da Zellen eigentlich zum Hintergrund segmentiert werden sollten. Jedoch sollten mit dem ScratchAssayAnalyzer die Grenzen der kollektiven Zellverbände erfasst werden und Zellen, welche sich vollständig vom Verband gelöst haben, migrieren definitionsgemäß nicht mehr kollektiv.

Abbildung 12 zeigt die Zwischenergebnisse des ScratchAssayAnalyzer auf einem exemplarischen Bildausschnitt einer Durchlichtaufnahme.

Klassifikation Das beschriebene Segmentierungsverfahren liefert als Ergebnis stets eine Region für die Wundfläche, auch wenn im Eingabebild keine zellfreie Fläche vorhanden ist. Dies ist aber der Fall, wenn sich die Wunde im Bild bereits vollständig geschlossen hat. Das Segmentierungsergebnis eines solchen Eingabebildes ist ein Binärbild, welches einige Zellen als Vordergrund und andere als Hintergrund darstellt. Eine solche willkürliche Segmentierung

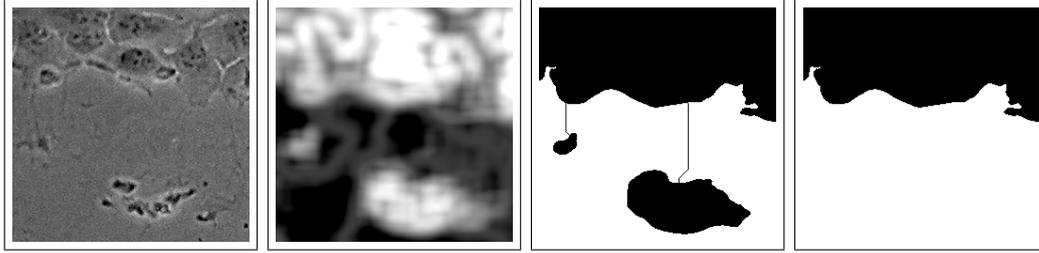


Abbildung 12: Beispielsegmentierung des ScratchAssayAnalyzer. Von links nach rechts: Ausschnitt einer Durchlichtaufnahme eines Scratch-Assay-Experimentes mit U-2OS-Zellen; Transformation durch Gaussfilter und anschließender Berechnung der lokalen Entropie für jedes Pixel; Segmentierungsergebnis des Bildausschnittes; schwarz: Zellen, weiß: Wundfläche; Ergebnis nach Nachverarbeitungsschritt

würde jedoch das Messergebnis verfälschen. Deshalb wurde eine Methode implementiert, welche Bilder, in denen die Wundfläche vollständig geschlossen ist, automatisch erkennen kann. Mithilfe einer Support Vector Machine (SVM) werden die Eingabebilder in zwei Klassen eingeteilt. Solche, die eine Wunde zeigen und solche, die dies nicht tun. Die Grundannahme für die Klassifikation ist, dass sich die Verteilungen der ermittelten Entropiewerte innerhalb der segmentierten Wund- und Zellregion deutlich unterscheiden, wenn diese tatsächlich jeweils Wunde beziehungsweise Zellen darstellen. Sollte die detektierte Wundregion aber genau wie die detektierte Zellregion in Wahrheit zum Großteil Zellen beinhalten, dann ähneln sich die Verteilungen der Entropiewerte innerhalb der jeweiligen Regionen. Für jedes Bild werden drei Maße berechnet, welche die Ähnlichkeit der Entropieverteilungen repräsentieren und welche als Eingabe für das Training beziehungsweise die Klassifikation mit der SVM dienen. Diese Maße sind:

1. Differenz der durchschnittlichen Entropiewerte
2. Bhattacharyya-Koeffizient (BC) [7]

$$BC(p_s, p_c) = \sum_{i=1}^L \sqrt{p_s(i) \cdot p_c(i)}, \quad (28)$$

wobei p_s und p_c die normalisierten und in L Bins diskretisierten Histogramme der Entropiewerte in Wund- beziehungsweise Zellregion sind. Der Bhattacharyya-Koeffizient kann Werte zwischen null und eins annehmen, wobei das Maximum bei zwei identischen Verteilungen erreicht wird.

3. Kolmogorov-Smirnov-Statistik (KS) [78]

$$KS(F_s, F_c) = \max_{i \in \{1, \dots, L\}} |F_s(i) - F_c(i)|, \quad (29)$$

wobei F_s und F_c die empirischen kumulativen Verteilungsfunktionen der Wund- beziehungsweise Zellregion bezeichnen, welche aus den normalisierten Histogrammen der Entropiewerte abgeleitet werden. Die Kolmogorov-Smirnov-Statistik kann ebenfalls Werte zwischen eins und null annehmen, wobei sie bei identischen Verteilungen den Wert null bekommt.

Implementiert wurde die SVM-basierte Klassifikation mit Hilfe der Java-Bibliothek libSVM [15]. Des Weiteren wurde der MiToBo-Operator ScratchAssaySVMTrainer entwickelt, um SVMs für Scratch-Assay-Bilder zu trainieren und deren Leistung mittels Kreuzvalidierung zu bestimmen (s. Abschnitt 3.1.2).

2.6 Auswertung von RNA-Seq Daten

2.6.1 Probenvorbereitung und Sequenzierung

Für die Erstellung von cDNA-Bibliotheken für Hochdurchsatzsequenzierungen wurde Total-RNA mittels Trizol aus Zellen extrahiert. Die Erstellung der Bibliotheken, sowie die Sequenzierungen erfolgten am interdisziplinären Zentrum für klinische Forschung (IZKF) der Universität Leipzig. Für Total-RNA-Seq wurde ribosomale RNA abgereichert und anschließend wurden die verbleibenden RNA-Moleküle chemisch fragmentiert. Mit Hilfe randomisierter Hexamere als Primer wurde cDNA synthetisiert. Für microRNA-Seq wurden, nach Adapter- und Barcode-Ligation, Moleküle mit einer Länge zwischen 140 und 165 Basenpaare per Gelelektrophorese herausgesucht und für die Bibliotheken verwendet. Sequenzierungen erfolgten in einer Sequenziermaschine vom Typ Illumina HighScan-SQ (Illumina, San Diego, USA). Für Total-RNA-Seq erfolgte die Sequenzierung im Paired-End-, für microRNA-Seq im Single-End-Verfahren. Die entstandenen Reads bestanden aus 101 Basen im Total-RNA-Seq und 51 Basen im microRNA-Seq.

2.6.2 Bestimmung differentieller Genexpression

Die Bestimmung differentieller Genexpression ausgehend von sequenzierten Reads kann in die Schritte Adapterentfernung/Qualitätskürzung, Zuordnung, Zusammenfassung und Tests auf differentielle Expression unterteilt werden. Im Folgenden werden diese Schritte näher erläutert. Abbildung 13 zeigt das Schema der Bestimmung differentieller Genexpression aus Total-RNA-Seq- und microRNA-Seq-Experimenten, sowie die dafür verwendeten Programme.

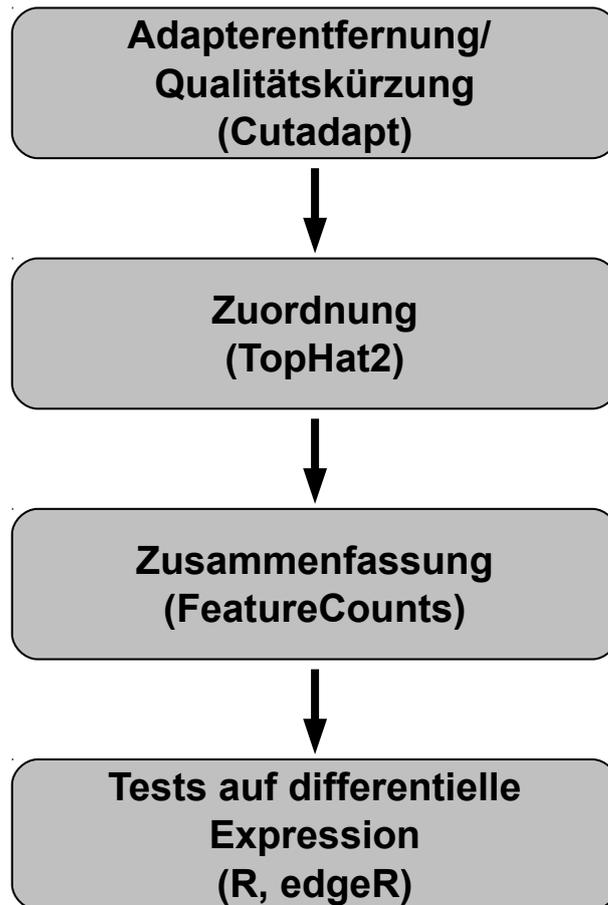


Abbildung 13: Bearbeitungsschritte zum Bestimmen differentiell exprimierter Gene und die dafür verwendeten Programme

Adapterentfernung/Qualitätskürzung Wenn die Anzahl der Basen von zu sequenzierenden cDNA-Fragmenten kürzer als der Anzahl der Sequenzierzyklen ist, in diesem Fall also 101 für Total-RNA-Seq und 51 für microRNA-Seq, dann werden auch Nucleotide des, an das 3'-Ende des Fragmentes ligierten, Adapters sequenziert. Diese zusätzlichen Basen können bewirken, dass die entsprechenden Fragmente dem Genom nicht oder an falscher Stelle zugeordnet werden. Des Weiteren nimmt die Qualität der sequenzierten Basen bei Sequenziermaschinen der Firma Illumina mit zunehmender Länge der Sequenz ab [74]. Das bedeutet, die Wahrscheinlichkeit für fehlerhafte Basen nimmt insbesondere zum Ende der Sequenzen zu. Das Entfernen von Basen mit niedrigen Qualitätswerten kann den Anteil der dem Genom zuordenbaren Sequenzen deutlich erhöhen. Zu rigoroses Kürzen kann allerdings zu deutlichen Informationsverlusten führen [28, 74]. Putative Adaptersequenzen, sowie Sequenzen mit zu niedrigen Qualitätswerten wurden mit Hilfe des Programms Cutadapt [77] in der Version 1.6 entfernt. Als Qualitätsschwellwert (Parameter -q) wurde 20 gewählt. Sequenzen des Total-RNA-Seq, welche nach dem Kürzen aus weniger als 20 Basen bestanden wurden komplett entfernt. Für microRNA-Seq wurde eine Mindestbasenanzahl von 16 festgelegt. Abbildung 14 zeigt ein Beispiel für die Verteilungen der Qualitätswerte der Reads einer Probe vor und nach Adapterentfernung und Qualitätskürzung. Tabelle 3 listet die verwendeten Bearbeitungsparameter für Cutadapt auf.

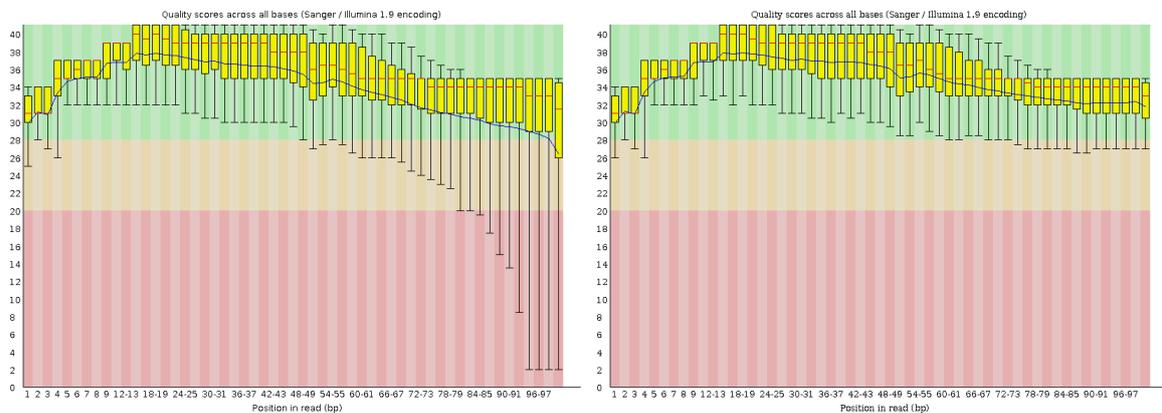


Abbildung 14: Basenqualitäten vor (links) und nach (rechts) Adapterentfernung und Qualitätskürzung einer Beispielprobe eines Total-RNA-Seq. Die Boxplots zeigen die Verteilung der Qualitätswerte pro Basenposition innerhalb der Reads. Die blaue Linie zeigt die Mittelwertverteilung über die Basenpositionen. Die Abbildungen wurden erstellt mit FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/)

Parameter	Total-RNA-Seq	microRNA-Seq	Bedeutung
-a	s. Tabelle 4	s. Tabelle 4	3'-Adaptersequenz
-O	7	7	minimale Anzahl an Basen, die mit Adaptersequenz überlappen
-q	20	20	Qualitätsschwellwert
-m	20	16	minimale Sequenzlänge nach Kürzen

Tabelle 3: *Cutadapt-Parameter und die für Total-RNA-Seq, sowie microRNA-Seq verwendeten Werte*

3'-Adapter Total-RNA-Seq Anfangs-Reads	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
3'-Adapter Total-RNA-Seq End-Reads	AGATCGGAAGAGCGTCTGTAGGGAAAGAGTGTA
3'-Adapter microRNA-Seq	TGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC

Tabelle 4: *Zum Adapterentfernen verwendete 3'-Adaptersequenzen*

Zuordnung Im Anschluss an Adapterentfernung und Qualitätskürzungen wurden die verbleibenden Reads auf das menschliche Genom (UCSC GRCh37/hg19) [100] gemappt. Für Total-RNA-Seq wurde das Programm TopHat2 [61] in Version 2.0.13 verwendet, für microRNA-Seq das Programm Bowtie2 [68] in Version 2.2.4. TopHat2 benutzt intern ebenfalls Bowtie2 für das Mappen. Zusätzlich kann TopHat2 Exon-Exon-Grenzen innerhalb von Reads erkennen und die entsprechenden Teilsequenzen auf die passenden Bereiche des Referenzgenomes mappen. Es wurden maximal 2 unterschiedliche Basen pro Read im Vergleich zum Referenzgenom toleriert. Dabei durfte maximal ein Unterschied in den sogenannten Seed-Sequenzen vorkommen. Seed-Sequenzen sind überlappende Teilsequenzen einer Read-Sequenz, welche von Bowtie2 verwendet werden, um die Zuordnung zu beschleunigen [68]. Reads, welche diese Kriterien nicht erfüllten wurden verworfen. Die verwendeten zuordnungsrelevanten Parameter von TopHat2 und Bowtie2 sind in den Tabellen 5 und 6 aufgelistet.

Parameter	Wert	Bedeutung
-r	50	erwarteter innerer Abstand zwischen Anfangs- und End-Read eines Fragmentes
-b2-N	1	max. Anzahl Fehler in Seed-Sequenzen
-g	20	max. Anzahl Zuordnungen pro Read

Tabelle 5: *TopHat2-Parameter*

Parameter	Wert	Bedeutung
-N	1	max. Anzahl Fehler in Seed-Sequenzen
-k	20	max. Anzahl Zuordnungen pro Read

Tabelle 6: *Bowtie2-Parameter*

Zusammenfassung der Reads Nach dem Mappen wurde für jede Probe bestimmt, wieviele Reads jeweils annotierten Genen des menschlichen Genoms zugeordnet wurden. Dazu wurde das Programm featureCounts [73] in Version 1.4.6 verwendet. Die Ensembl-Annotation GRCh37.75 [18] des menschlichen Genoms wurde für das Annotieren der Reads des Total-RNA-Seq verwendet. Da es sich dabei um Paired-End-Sequenzierungsdaten handelte, wurden nicht einzelne Reads, sondern Fragmente gezählt. Ein Fragment überspannt dabei jeweils den genomischen Bereich, der mit Beginn des ersten Reads eines Paares anfängt und mit dem Ende des zweiten Reads endet. Gezählt wurden nur solche Fragmente, die mindestens mit einer Base eines annotierten Exon überlappten. Fragmente, welche mehrere Exons des selben Gens überspannten, wurden nur einmal gezählt. Für die Annotation der microRNA-Seq-Daten wurde miRBase [63], Version 20 verwendet. Die für das Zusammenfassen mit featureCounts verwendeten Parameter sind in Tabelle 7 aufgelistet.

Parameter	Total-RNA-Seq	microRNA-Seq	Bedeutung
-p	-p		paired-end Daten
-M	-M	-M	Fragmente/ Reads, welche auf mehrere Positionen gemappt wurden, werden mitgezählt
-t	exon	miRNA	zu zählende Merkmale
-g	gene_id	ID	Metamerkmale zu dem zugehörige Merkmale zusammengefasst werden

Tabelle 7: *featureCounts-Parameter*

Tests auf differentielle Expression Die ermittelten Fragment- beziehungsweise Read-Zahlen wurden zur Bestimmung differentieller Genexpression verwendet. Mit Hilfe der TMM-Normalisierung [97] wurde für jede Probe ein Normalisierungsfaktor berechnet, um Unterschiede in Bibliotheksgrößen, sowie unterschiedliche RNA-Kompositionen der Proben, auszu-

gleichen. Dies geschah unter der Annahme, dass ein Großteil der untersuchten Gene nicht differentiell exprimiert war. Anschließend wurden für jedes Gen Unterschiede der Expression bestimmt. Diese wurden als zur Basis zwei logarithmierte Quotienten aus den mittleren Expressionswerten (\log_2FC) der unterschiedlichen Versuchsbedingungen berechnet. Mit Hilfe statistischer Tests wurde zudem für jedes Gen die Wahrscheinlichkeit bestimmt, dass dieses nicht differentiell exprimiert wurde. Mit Hilfe der Methode von Benjamini und Hochberg [5] wurden aus den P-Werten der Tests, die für multiples Testen korrigierten, FDR-Werte (False Discovery Rate) berechnet. Gene mit einem FDR-Wert kleiner als 0.05 wurden als differentiell exprimiert betrachtet. TMM-Normalisierung, Test auf differentielle Expression, sowie FDR-Wertbestimmung wurden mit Hilfe der R-Bibliothek edgeR [96] in Version 3.12 durchgeführt.

2.6.3 Detektion von Mykoplasmakontaminationen

Mykoplasmen sind parasitäre Bakterien, welche menschliche Atemwege oder den Urogenitaltrakt befallen können. Einige Arten sind pathogen. So kann eine Infektion mit *Mycoplasma genitalium* beispielsweise zu Unterleibsentzündungen führen. Auch Zellkulturen sind häufig mit Mykoplasmen kontaminiert [88]. Menschliche Zellen, welche mit Mykoplasmen infiziert wurden, zeigen Unterschiede in der Genexpression im Vergleich zu nicht-infizierten Zellen [53]. Mykoplasmakontamination kann mit Hilfe PCR-basierter Methoden erkannt werden. Allerdings sind diese Methoden fehleranfällig und können sowohl falsch-positive, als auch falsch-negative Ergebnisse produzieren [115]. In [88] wurde postuliert, dass Reads aus RNA-Seq-Daten, welche auf Mykoplasmen Genome gemappt werden können, auf eine Mykoplasma-kontamination hinweisen können.

Reads des Total-RNA-Seq, welche mit TopHat2 nicht dem menschlichen Genom zugeordnet werden konnten, wurden verwendet um sie auf Mykoplasmen Genome zu mappen. Hierfür wurden die Referenzgenome von *Mycoplasma arginini*, *Mycoplasma fermentans*, *Mycoplasma hominis*, *Mycoplasma hyorhinitis*, *Mycoplasma genitalium* und *Acholeplasma laidlawii* benutzt. Zum Mappen wurde TopHat2 verwendet. Entsprechend [88] wurde von einer Mykoplasma-kontamination ausgegangen, wenn die Anzahl der auf Mykoplasmen Genome gemappten Reads mehr als 0.01% der Gesamtanzahl der Reads dieser Probe entsprachen. Als Gesamtanzahl Reads einer Probe wurde die Menge der Reads betrachtet, welche nach Adapter- und Qualitätskürzungen übrig blieb.

2.6.4 Genanreicherungsanalysen

Mit Hilfe der Liste differentiell exprimierter Gene wurde untersucht, ob diese Gene Eigenschaften besaßen, welche häufiger auftraten, als zufällig zu erwarten wären. Hierfür wurden Genanreicherungsanalysen mit dem Programm Cytoscape ([105]) in Version 3.2.1 und dem für diesen Zweck entwickelten Cytoscape-Plugin ClueGO ([8]) in Version 2.3.4 durchgeführt. Dieses Plugin prüft mittels Hypergeometrischen Tests, ob bestimmte Eigenschaften einer Genmenge gegenüber einer Referenzmenge statistisch signifikant angereichert sind. Die aus den Hypergeometrischen Tests ermittelten P-Werte wurden mit der Methode von Benjamini und Hochberg [5] für multiples Testen korrigiert. Die so berechneten False Discovery Rates (FDR) wurden als Maß für die Signifikanz verwendet. FDR-Werte kleiner als 0.05 wurden als Maß für eine signifikante Anreicherung betrachtet. Als Eigenschaften wurden Annotationen der Gene Ontology ([17]) in der Version vom 16.02.2016 und der KEGG (Kyoto Encyclopedia of Genes and Genomes) [59] in der Version vom 18.02.2016 verwendet. Das Gene Ontology Projekt stellt einheitliche Beschreibungen für Genprodukte zu Verfügung. Mit Hilfe dreier hierarchisch strukturierter Ontologien stellt es artunabhängig Beschreibungen für Genprodukte bezüglich biologischer Prozesse an denen sie beteiligt sind, ihrer molekularen Funktionen, sowie ihrer zellulären Lokalisation bereit. Von signifikant angereicherten ($FDR < 0.05$) Gene Ontology-Kategorien, welche zueinander in Eltern-Kind-Beziehung standen und sich in maximal einem Gen unterschieden wurde jeweils die Kategorie mit der größeren FDR entfernt um Redundanzen zu reduzieren (Option „Use GO Term Fusion“). KEGG annotiert Gene danach, ob sie an bestimmten Geninteraktionsnetzwerken beteiligt sind. Diese Netzwerke enthalten Informationen zu Stoffwechselwegen, regulatorischen Netzwerken oder Krankheiten.

2.6.5 Bestimmung putativer mRNA-microRNA-Bindungen

Für die Bestimmung möglicher mRNA-microRNA-Bindungen wurde das R-Paket multiMiR [101] in Version 1.0.1 verwendet. Mit diesem Paket können verschiedene Datenbanken, welche entweder Informationen zu experimentell validierten mRNA-microRNA-Bindungen oder zu *in-silico*-Vorhersagen enthalten, abgefragt werden. Für die Ermittlung putativer mRNA-microRNA-Bindungen wurden jeweils alle acht vorhandenen Datenbanken, welche *in-silico*-Vorhersagen bereitstellen, abgefragt. Des Weiteren wurden die Standardparameter von multiMiR verwendet. Nur microRNA-mRNA-Paare, welche von mindestens zwei der acht abgefragten Datenbanken als putative Bindungspartner bestätigt wurden, wurden für weitere Betrachtungen verwendet.

3 Ergebnisse

3.1 Algorithmenevaluation

Programme für die automatisierte Auswertung von Migrationsexperimenten müssen verlässliche Ergebnisse produzieren. Damit sie eine breite Anwendung finden müssen sie darüber hinaus auch für eine Vielzahl unterschiedlicher Computerplattformen zur Verfügung stehen und leicht bedienbar sein.

In diesem Abschnitt werden Bewertungen für die beiden Bildverarbeitungsapplikationen beschrieben, welche für die vorliegende Arbeit entwickelt wurden. Sowohl die Anpassung der Parameter der Applikationen, als auch die Beurteilung ihrer Leistungen, erfolgte mit Hilfe manuell generierter Vergleichsdaten. Diese Daten wurden von Experten erstellt und dienten als Referenzdaten. Beide Applikationen wurden zudem mit Programmen verglichen, welche für dieselben Zwecke entwickelt wurden.

3.1.1 Evaluation des CellMigrationAnalyzer

Für die Beurteilung des CellMigrationAnalyzer wurden die Schritte Segmentierung und Zuordnung separat betrachtet und evaluiert. Zusätzlich wurden vom CellMigrationAnalyzer berechnete Zelltrajektorien mit manuell erstellten Trajektorien verglichen. Dies geschah um beurteilen zu können, wie hoch die Übereinstimmung von manuell und automatisch ermittelten Werten für das Migrationsverhalten von Zellen ist. Weiterhin wurde der CellMigrationAnalyzer mit Programmen verglichen, welche ebenfalls für die automatisierte Analyse von Einzelzell-Tracking-Experimenten entwickelt wurden.

Segmentierung Zur Optimierung und Evaluierung des Segmentierungsschrittes wurden drei Videos verwendet. Diese Videos bestanden aus jeweils 25 Einzelbildern und zeigten entweder GFP-exprimierende Zellen der Zelllinien ES-2 oder U-2OS, oder E2-Crimson-exprimierende Zellen der Linie HT-144. Für die Beurteilung der Segmentierungsqualität des CellMigrationAnalyzer wurden die Konturen der Zellen von einer Biologin (Dr. rer. nat. Kristin Wächter) manuell bestimmt. Die von den Konturen begrenzten Regionen repräsentierten die Zellen. Die Pixel dieser Regionen wurden mit Werten gekennzeichnet, welche die individuellen Zellen im gesamten Film eindeutig identifizierten. Für die Optimierung der Segmentierungsparameter und die anschließende Bewertung der Ergebnisse wurden die Bilder der Videos in Trainings- und Testdatensätze aufgeteilt. Ein Trainingsdatensatz bestand aus jeweils den ersten 15 Bildern, ein Testdatensatz aus jeweils den letzten 10 Bildern eines Videos.

Die Werte für die Standardabweichung des Gaußfilters (σ) und für den Exponenten der Gammakorrektur (γ) wurden systematisch variiert, um eine optimale Kombination dieser Werte für die Segmentierung der gegebenen Videos zu finden. Für jede getestete Kombination aus σ und γ wurden in allen Einzelbildern die manuell segmentierten Referenzregionen mit den automatisch segmentierten Regionen verglichen. Regionen beider Segmentierungen mit überlappenden Positionen wurden dabei einander zugeordnet. Für den Fall, dass mehrere Regionen der automatischen Segmentierung mit einer Referenzregion überlappten, wurden die Regionen mit der größten Überlappung einander zugeordnet. Für jede Zuordnung einer Referenzregion i und einer Region j der automatischen Segmentierung wurde bestimmt, wie viel Pixel jeweils sowohl i als auch j angehörten. Diese Zahl wird im Folgenden als TP (true positives) bezeichnet. Weiterhin wurde die Anzahl der Pixel bestimmt, welche zu i aber nicht zu j gehörten, sowie die Anzahl der Pixel welche zu j aber nicht zu i gehörten. Diese bildeten jeweils die Werte FN (false negatives) und FP (false positives). Mithilfe dieser Werte wurden die Evaluierungsmaße Precision (\mathcal{P}), Recall (\mathcal{R}) und daraus das F-Maß (\mathcal{F}) wie folgt berechnet:

$$\mathcal{P} = \frac{TP}{TP + FP}, \quad \mathcal{R} = \frac{TP}{TP + FN}, \quad \mathcal{F} = \frac{2 \cdot \mathcal{P} \cdot \mathcal{R}}{\mathcal{P} + \mathcal{R}} \quad (30)$$

Precision beschreibt somit für jede detektierte Zelle den Anteil der Pixel, welche gemäß der Referenzsegmentierung tatsächlich zu der entsprechenden Zelle gehören. Recall hingegen beschreibt den Anteil der Pixel, welche zu einer Zelle gehören und auch als solche detektiert wurden. Das F-Maß als gewichtetes Mittel dieser beiden Maße kondensiert Precision und Recall zu einer Maßzahl, deren Maximalwert von eins erreicht würde, wenn genau die Pixel einer Region der automatischen Segmentierung auch in der Referenzsegmentierung alle Pixel der zugeordneten Region darstellen. Insgesamt wurden 160 Parameterkombinationen getestet, in denen σ von 0 bis 7 in Schritten der Größe 1 und γ von 0.1 bis 2.0 in Schritten der Größe 0.1 variiert wurde. Zunächst wurden die Trainingsdatensätze der drei Zelllinien unabhängig voneinander betrachtet. Precision, Recall und F-Maß wurden für alle Regionen berechnet und der Parametersatz, welcher im Mittel das höchste F-Maß erzielte wurde als optimaler Parametersatz gewählt. Tabelle 8 listet die optimalen Parameterkombinationen, sowie die damit ermittelten durchschnittlichen Werte für Precision, Recall und F-Maß, auf.

Der konsistent niedrigere Precision-Wert deutet daraufhin, dass die detektierten Zellflächen unter Verwendung der gewählten Parameter tendenziell etwas zu groß waren. Für alle Trainingsdatensätze außer dem für die U-2OS-Zellen, erzielte die Kombination aus $\sigma = 1$ und

Datensatz	Anz. Zellen	beste Kombination	$\varnothing \mathcal{P}$	$\varnothing \mathcal{R}$	$\varnothing \mathcal{F}$
ES-2	285	$\sigma: 1, \gamma: 0.1$	0.727	0.899	0.804
HT-144	230	$\sigma: 1, \gamma: 0.1$	0.789	0.959	0.866
U-2OS	261	$\sigma: 1, \gamma: 0.3$	0.868	0.922	0.894
Pool	776	$\sigma: 1, \gamma: 0.1$	0.790	0.927	0.852

Tabelle 8: Parameterkombinationen (σ, γ) , welche auf den Trainingsdaten das jeweils höchste durchschnittliche F-Maß ($\varnothing \mathcal{F}$) erzielten und die dazugehörigen durchschnittlichen Precision- ($\varnothing \mathcal{P}$) und Recall- ($\varnothing \mathcal{R}$) Werte. Die zweite Spalte enthält die Summe der gemäß der Referenzsegmentierungen abgebildeten Zellen aller Bilder eines Datensatzes. Der Datensatz Pool stellt die Zusammenfassung der Trainingsdaten aller drei Zelllinien dar.

$\gamma = 0.1$ die besten durchschnittlichen F-Maße. Allerdings war das mit diesen Parametern erzielte durchschnittliche F-Maß von 0.893 auf dem U-2OS-Trainingsdatensatz nur minimal geringer als das für diesen Datensatz maximal erreichte F-Maß von 0.894. Deshalb wurde zum Testen für alle Testdatensätze die Kombination $\sigma = 1, \gamma = 0.1$ verwendet. Tabelle 9 zeigt die erhaltenen durchschnittlichen Werte für Precision-, Recall- und F-Maß auf den Testdaten.

Datensatz	Anz. Zellen	$\varnothing \mathcal{P}$	$\varnothing \mathcal{R}$	$\varnothing \mathcal{F}$
ES-2	213	0.698	0.891	0.783
HT-144	144	0.826	0.941	0.880
U-2OS	131	0.854	0.914	0.883
Pool	488	0.783	0.912	0.842

Tabelle 9: Durchschnittliche Precision ($\varnothing \mathcal{P}$), Recall ($\varnothing \mathcal{R}$), und F-Maße ($\varnothing \mathcal{F}$), welche auf den Testdaten der Einzelzellsegmentierung mittels der Parameterkombination $(\sigma = 1, \gamma = 0.1)$ erzielt wurden. Die zweite Spalte enthält die Summe der gemäß der Referenzsegmentierungen abgebildeten Zellen aller Bilder eines Datensatzes. Der Datensatz Pool stellt die Zusammenfassung der Testdaten aller drei Zelllinien dar.

Auch in den Ergebnissen der Testdaten deuteten niedrigere Precision-Werte auf eine Überschätzung der Zellflächen hin. Dies wurde auch durch die Betrachtung der Flächen der segmentierten Zellen bestätigt (s. Abbildung 15). Abbildung 16 zeigt exemplarische Segmentierungsergebnisse für die drei Testdatensätze und die dazugehörigen manuellen Segmentierungen. Insgesamt konnten für alle getesteten Datensätze hohe Übereinstimmungen mit den Referenzsegmentierungen erreicht werden.

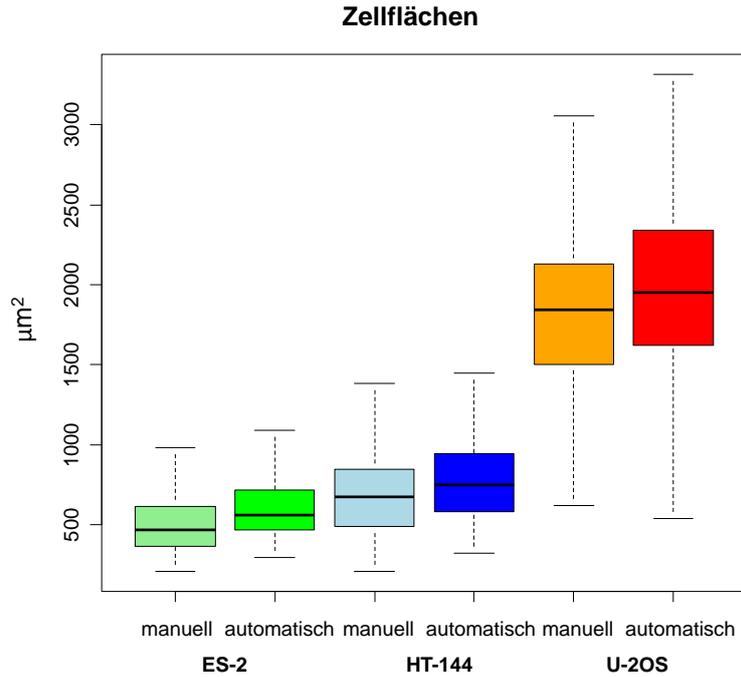


Abbildung 15: Zellflächen der manuellen und automatischen Segmentierungen der Testdatensätze.

Neben den für das Optimieren variierten Parametern, wurden die übrigen Segmentierungsparameter mit folgenden konstanten Werten verwendet:

- minimale Größe der vorsegmentierten Regionen: 300 Pixel
- minimale Größe der detektierten Regionen: 500 Pixel
- maximale Anzahl Level-Set-Iterationen: 1000

Bei der hier verwendeten Methode zur Evaluierung der Segmentierungsqualität wird ausgehend von den Referenzregionen geprüft, wie hoch die Übereinstimmung dieser mit den automatisch segmentierten Vordergrundregionen ist. Das bedeutet allerdings, dass gegebenenfalls zusätzlich segmentierte Regionen des Vordergrundes, welche keiner Referenzregion entsprechen, nicht bestraft werden. Solche zusätzlichen Regionen würden jedoch das Ergebnis der Beurteilung auch nicht fälschlicherweise verbessern.

Zuordnung Um die Zuordnung der segmentierten Zellen in den Einzelbildern der Videos zu optimieren wurde nach einem geeigneten Schwellwert T_a für die Erkennung von Untersegmentierungen gesucht. Hierfür wurde betrachtet, in welchem Maß sich die Flächen segmentierter

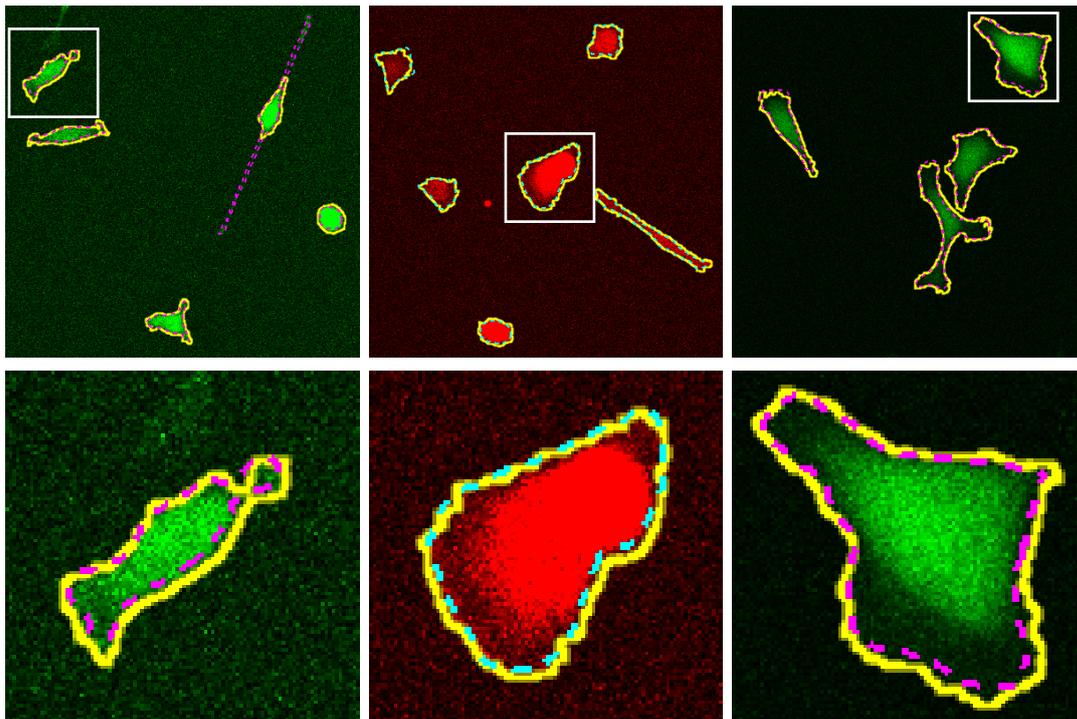


Abbildung 16: Exemplarische Segmentierungsergebnisse auf den Einzelzelltestdatensätzen. Links: ES-2-Zellen, GFP-exprimierend. Mitte: HT-144-Zellen, E2-Crimson exprimierend. Rechts: U-2OS-Zellen, GFP-exprimierend. Durchgezogene Linien (gelb) kennzeichnen die Konturen der automatisch detektierten Zellen, unterbrochene Linien (magenta beziehungsweise cyan) kennzeichnen die Konturen der manuell bestimmten Zellen. Die untere Reihe zeigt vergrößerte Ausschnitte der Bilder in der oberen Reihe. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.

Regionen veränderten, wenn sie im vorhergehenden oder folgenden Bild zusammen mit anderen Regionen fälschlicherweise als eine Region segmentiert wurden. Demgegenüber wurden die Flächenänderungen von Regionen gestellt, welche die selben Zellen in der Referenzsegmentierung in aufeinanderfolgenden Bildern repräsentierten. Dies diente der Abschätzung tatsächlicher Flächenänderungen aufgrund von Formveränderungen der Zellen. Als Daten hierfür wurden die, für die Optimierung der Segmentierungsparameter verwendeten, Trainings- und Testdatensätze vereinigt, so dass für die drei Zelllinien ES-2, U-2OS und HT-144 jeweils ein Film bestehend aus 25 Bildern zur Verfügung stand.

Abbildung 17 zeigt, dass es für die gewählten Filme keinen Schwellwert gibt, der Flächenänderungen aufgrund von tatsächliche Formveränderungen und Untersegmentierungen fehlerfrei trennen kann.

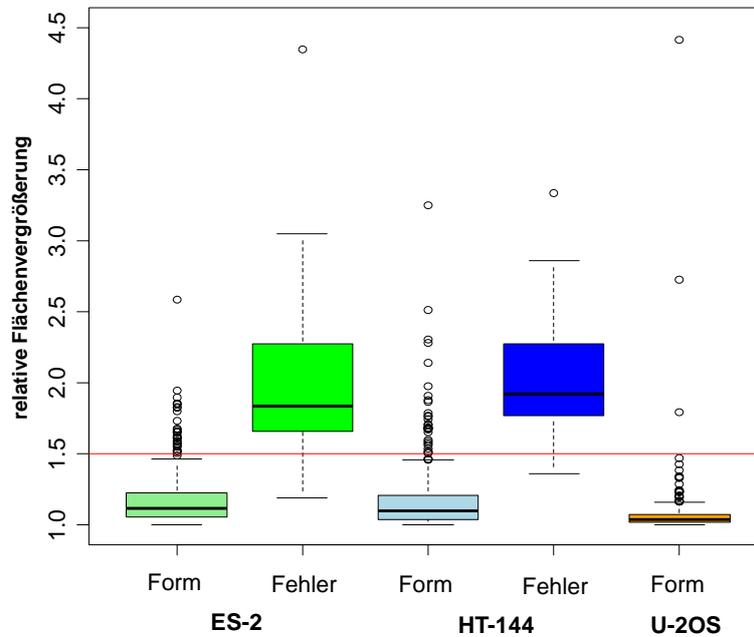


Abbildung 17: Flächenvergrößerungen detektierter Zellen aufgrund von Formveränderungen (Form) und Untersegmentierungen (Fehler). Die Boxplots zeigen die Verteilungen der Flächenänderungen so, dass jeweils die Veränderung der Flächen der kleineren zu den größeren Regionen zweier zeitlich aufeinanderfolgender Bilder gezeigt ist. In dem Datensatz der U-2OS-Zellen traten keine Untersegmentierungen auf, weshalb es hierfür keine Werte gab. Die rote horizontale Linie markiert eine relative Flächenzunahme auf 150 %.

Jedoch wird mit einer maximal zulässigen Flächenvergrößerung auf 150 %, das heißt $T_a = 0.5$, der Großteil der Untersegmentierungen erkannt und nur wenige Flächenveränderungen wer-

den fälschlicherweise als Untersegmentierungen betrachtet. Deshalb wurde für die folgenden Betrachtungen ein Wert von 0.5 für T_a gewählt. Ein kleinerer Schwellwert würde zu einer besseren Erkennung von Untersegmentierungen führen, aber gleichzeitig würden auch Formveränderungen häufiger fälschlicherweise als Untersegmentierung betrachtet. Für den Fall, dass die korrekte Rekonstruktion einzelner Zellpfade eine hohe Priorität hat, sollte T_a eher groß gewählt werden, damit auch größere Formveränderungen einzelner Zellen nicht dazu führen, dass ihre Trajektorien beendet werden. Wenn hingegen das Migrationsverhalten der ganzen Zellpopulation im Vordergrund steht oder Untersegmentierungen aufgrund einer hohen Zelldichte häufig auftreten, dann sollte ein kleinerer Wert für T_a gewählt werden. Eine Erhöhung der Bildaufnahme rate führt im Allgemeinen dazu, dass Formveränderungen zwischen aufeinanderfolgenden Aufnahmen kleiner und damit auch kleinere Werte für T_a verwendet werden können.

Zur Validierung des Zuordnungsoperators wurden aus den Referenzsegmentierungen der 25 Bilder umfassenden Videos der drei Zelllinien neue Videos erstellt, in denen jeweils die individuellen Markierungen der Zellen entfernt wurden. Die Ergebnisse der automatischen Zuordnungen für diese Videos wurden mit denen der Referenzsegmentierungen, in denen die Markierungen für individuelle Zellen vorhanden waren, verglichen. Diese Daten wurden verwendet, um Segmentierungsfehler von der Beurteilung auszuschließen.

Ähnlich wie in [58] wurde für die Bestimmung der Zuordnungsleistung die Anzahl sogenannter Austauschfehler berechnet. Ein Austauschfehler liegt vor, wenn zwei Regionen aus aufeinanderfolgenden Bildern, welche dieselbe Zelle repräsentieren, unterschiedliche Markierungen bekommen. Im Gegensatz zu [58] wurde es zusätzlich auch als Austauschfehler betrachtet, wenn eine Region, welche in einem Bild neu erscheint, die Markierung einer bereits vorhandenen Zelle erhielt. Die Anzahl der Austauschfehler wurde zur Bestimmung der Zuordnungsleistung ins Verhältnis zur Summe aller Referenzzuordnungen zwischen jeweils zwei aufeinanderfolgenden Bildern gesetzt. Formal ist die Zuordnungsleistung P_{links} wie folgt definiert:

$$P_{links} = 1 - \frac{\text{Anzahl Austauschfehler}}{\text{Anzahl Referenzzuordnungen}} \quad (31)$$

Die Parameter, welche die Zuordnungsleistung beeinflussen, sind der Schwellwert T_d für die maximale Distanz, welche die Zellen in der Zeit zwischen zwei Aufnahmen migrieren und der Schwellwert T_a , welcher die maximale Flächenänderung einer Zelle in der Zeit zwischen zwei Aufnahmen festlegt. T_d wurde nach der Methode aus [58] für jedes Eingabevideo neu berechnet. Für T_a wurde der Wert 0.5 gewählt. Da die Referenzsegmentierungen als fehlerfrei angesehen wurden, wurde die Zuordnungsleistung zusätzlich mit praktisch deaktivierter

Erkennung von Untersegmentierung betrachtet. Hierfür wurde $T_a = 100$ verwendet. Tabelle 10 listet die Ergebnisse der Zuordnungsleistung auf den drei Vergleichsvideos auf.

	P_{links}	
Datensatz	$T_a = 0.5$	$T_a = 100$
ES-2	0.920	0.969
HT-144	0.905	0.978
U-2OS	0.986	0.993

Tabelle 10: Zuordnungsleistung (P_{links}) auf manuell segmentierten Daten. Gestestet wurde die Zuordnungsleistung P_{links} mit einem Wert für T_a von 0.5, sowie von 100.

Erwartungsgemäß wurden unter Verwendung von $T_a = 100$ bessere Ergebnisse als mit $T_a = 0.5$ erzielt. Dennoch waren, auch mit dem in der Praxis sinnvolleren Wert 0.5, in jedem Video mehr als 90 % der automatischen Zuordnungen korrekt. Insgesamt zeigt dies, dass der entwickelte Zuordnungsoperator auf den verwendeten Daten detektierte Zellen zuverlässig zuordnet.

Vergleich mit manuell erstellten Trajektorien Um zu beurteilen, wie gut der CellMigrationAnalyzer das Migrationsverhalten von Zellpopulationen bestimmen kann, wurden Trajektorien von Zellen von einer Expertin (Dr. rer. nat. Kristin Wächter) manuell ermittelt. Dazu wurden drei weitere Videos verwendet. Auf diesen waren ebenfalls fluoreszenzmarkierte Zellen der Linien ES-2, HT-144 oder U-2OS zu sehen. Diese Videos bestanden aus jeweils 121 Bildern, welche die Bewegungen der Zellen über einen Zeitraum von 10 Stunden abbildeten. Das manuelle Erstellen der Trajektorien geschah mit Hilfe des ImageJ-Plugins ManualTracking (rsbweb.nih.gov/ij/plugins/track/track.html). Dabei wurden optisch geschätzte Zellzentroiden als Referenzpunkte gewählt. Wenn eine Zelle vor dem Ende eines Videos den Bildrand berührte oder sich teilte, dann wurde die entsprechende Trajektorie an dieser Stelle beendet. Im Fall einer Zellteilung wurden anschließend zwei neue Trajektorien für die Tochterzellen erstellt. Die drei Videos wurden ebenfalls mit dem CellMigrationAnalyzer ausgewertet. Die Werte der verwendeten Parameter entsprachen den oben beschriebenen. Detektierte Zellen, welche die Bildränder berührten, wurden entfernt. Mit Hilfe der manuell und automatisch bestimmten Trajektorien wurden Geschwindigkeiten, sowie Gerichtetheiten der Zellen berechnet und miteinander verglichen. Dabei wurden allerdings nur Trajektorien deren Länge mindestens 24 Punkte betrug, welche also eine Zeit von mindestens zwei Stunden abdeckten, in die Berechnungen aufgenommen. Abbildung 18 zeigt die Verteilungen der mittleren Geschwin-

digkeiten, sowie der Gerichtetheiten der Zellen in den drei Videos. Es ist erkennbar, dass die jeweils zu einem Video gehörenden Werte der automatischen und manuellen Auswertung nur geringfügig voneinander abweichen. Tabelle 11 listet die manuell beziehungsweise automatisch ermittelten Werte für die mittleren Geschwindigkeiten, sowie die Gerichtetheiten der drei Zelllinien auf.

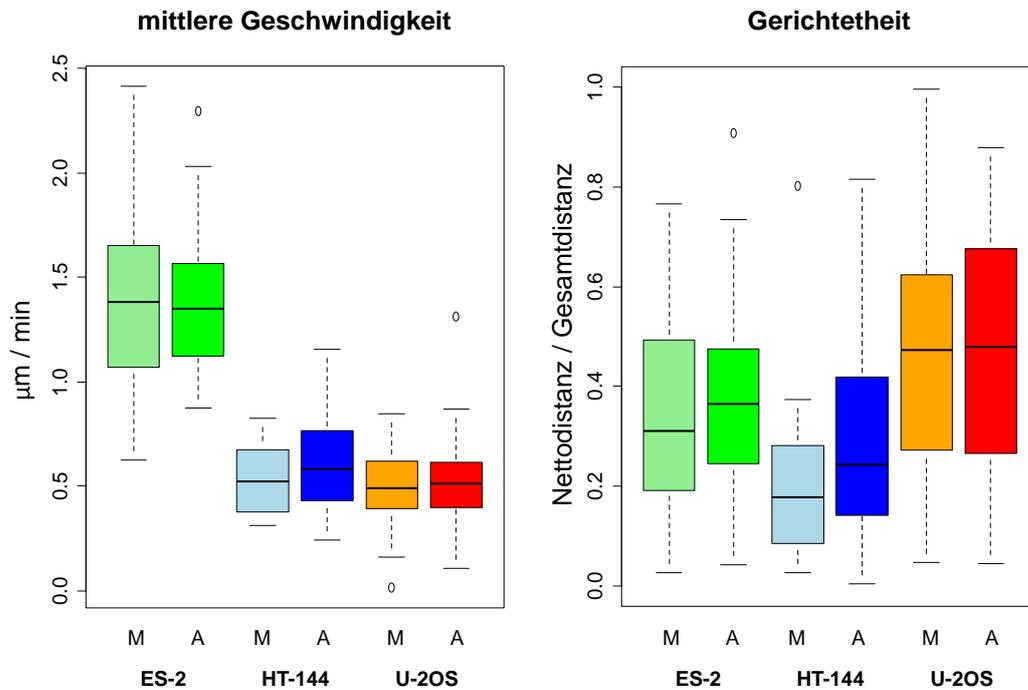


Abbildung 18: Vergleich von Geschwindigkeit und Gerichtetheit des CellMigrationAnalyzer mit manuell erstellten Trajektorien. Die Boxplots zeigen die mittleren Geschwindigkeiten und die Gerichtetheiten aller detektierter Zellen der manuellen Auswertung (M), sowie der automatischen Auswertung (A). Ausgewertet wurden nur Trajektorien, welche einen Zeitraum von mindestens 2h abdeckten.

Mit Hilfe von zweiseitigen Mann-Whitney-Tests wurden die Wahrscheinlichkeiten dafür berechnet, dass die ermittelten Werte für die mittleren Geschwindigkeiten und Gerichtetheiten der jeweiligen Zellpopulation aus der selben Grundgesamtheit stammten. P-Werte kleiner als 0.05 führten zur Ablehnung dieser Hypothese. Die Tabellen 12 und 13 zeigen die P-Werte der statistischen Tests für die jeweiligen Zellpopulationen. Vergleiche der Daten der manuellen und automatischen Auswertung derselben Population führten in keinem Fall zur Ablehnung der Nullhypothesen. Sowohl in den manuell, als auch in den automatisch ermittelten Daten zeigte sich, dass die mittleren Geschwindigkeiten der ES-2-Zellen signifikant höher als die der

Linie	Geschwindigkeit		Gerichtetheit	
	manuell	automatisch	manuell	automatisch
ES-2	1.40 ± 0.4	1.37 ± 0.32	0.34 ± 0.20	0.37 ± 0.2
HT-144	0.54 ± 0.18	0.61 ± 0.22	0.22 ± 0.19	0.29 ± 0.22
U-2OS	0.49 ± 0.21	0.51 ± 0.24	0.48 ± 0.26	0.46 ± 0.25

Tabelle 11: Durchschnittliche mittlere Geschwindigkeiten ($\mu\text{m}/\text{min}$) und durchschnittliche Gerichtetheiten \pm Standardabweichungen der manuellen, sowie automatischen Auswertungen der 10-stündigen Aufnahmen von Zellen der Linien ES-2, HT-144 und U-2OS.

HT-144- und U-2OS-Zellen waren. HT-144- und U-2OS-Zellen, zeigten untereinander jedoch keine signifikanten Geschwindigkeitsunterschiede. In der manuellen Auswertung zeigte sich ein signifikanter Unterschied in der Gerichtetheit der ES-2-Zellen zu den HT-144, sowie zu den U-2OS-Zellen. Weiterhin zwischen den HT-144- und den U-2OS-Zellen. Ein signifikanter Unterschied in der Gerichtetheit der U-2OS-Zellen und den HT-144-Zellen wurde auch mit dem CellMigrationAnalyzer festgestellt. Die signifikanten Unterschiede in der Gerichtetheit der ES-2-Zellen zu den anderen beiden Linien konnten mit dem CellMigrationAnalyzer jedoch nicht erkannt werden.

	ES-2 M	ES-2 A	HT-144 M	HT-144 A	U-2OS M	U-2OS A
ES-2 M	1.00	$6.7 \cdot 10^{-1}$	$1.4 \cdot 10^{-12}$	$4.6 \cdot 10^{-13}$	$1.2 \cdot 10^{-16}$	$1.4 \cdot 10^{-16}$
ES-2 A		1.00	$6.2 \cdot 10^{-14}$	$1.8 \cdot 10^{-13}$	$2.3 \cdot 10^{-17}$	$3.3 \cdot 10^{-16}$
HT-144 M			1.00	$2.9 \cdot 10^{-1}$	$6.1 \cdot 10^{-1}$	$6.9 \cdot 10^{-1}$
HT-144 A				1.00	$1.0 \cdot 10^{-1}$	$1.4 \cdot 10^{-1}$
U-2OS M					1.00	$8.3 \cdot 10^{-1}$
U-2OS A						1.00

Tabelle 12: P-Werte der Mann-Whitney-Tests der mittleren Geschwindigkeiten von manuell (M), sowie automatisch (A) ermittelten Zelltrajektorien.

Zusammenfassend zeigte sich, dass die mit Hilfe des CellMigrationAnalyzer gewonnenen Daten, bezüglich Geschwindigkeiten und Gerichtetheiten der drei untersuchten Zellpopulationen, eine hohe Übereinstimmung mit den manuell ermittelten Werten aufwiesen.

Vergleich mit anderer Software Um die Segmentierungs- und Zuordnungsleistung des CellMigrationAnalyzer im Vergleich zu anderen Programmen einzuschätzen, wurden Bildse-

	ES-2 M	ES-2 A	HT-144 M	HT-144 A	U-2OS M	U-2OS A
ES-2 M	1.00	$5.5 \cdot 10^{-1}$	$2.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-1}$	$2.9 \cdot 10^{-2}$	$4.7 \cdot 10^{-2}$
ES-2 A		1.00	$2.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
HT-144 M			1.00	$2.9 \cdot 10^{-1}$	$7.6 \cdot 10^{-4}$	$9.0 \cdot 10^{-4}$
HT-144 A				1.00	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
U-2OS M					1.00	$8.2 \cdot 10^{-1}$
U-2OS A						1.00

Tabelle 13: *P-Werte der Mann-Whitney-Tests der Gerichtetheiten von manuell (M), sowie automatisch (A) ermittelten Zelltrajektorien.*

quenzen der Cell Tracking Challenge von 2013 [76] ausgewertet und die Ergebnisse mit denen der Programme der Teilnehmer dieses Wettbewerbes verglichen. Im Rahmen des Wettbewerbes wurden verschiedene Datensätze vorgestellt, welche Zellen oder Zellkerne in 2D- beziehungsweise 3D-Umgebungen zeigten. Der Datensatz „C2DL-MS“^c, welcher GFP-exprimierende mesenchymale Stammzellen von Ratten zeigt, wies die höchste Ähnlichkeit zu den für diese Arbeit verwendeten Bildern auf. Deshalb wurde dieser Datensatz für die anschließenden Vergleiche verwendet. Aufgrund des schlechten Signal-zu-Rauschverhältnisses und der variablen Morphologie der Zellen wurde die Komplexität der Auswertung dieses Datensatzes in [76] als hoch eingestuft.

Der Datensatz bestand aus zwei Paaren von Bildsequenzen. Diese Paare bestanden wiederum aus jeweils einer Trainings- und einer Testsequenz. Mit den Trainingssequenzen konnten die Teilnehmer des Wettbewerbes die Parameter ihrer Algorithmen anpassen. Die Testsequenzen dienten der Evaluierung der eingereichten Programme im Rahmen des Wettbewerbes. Für das Optimieren der Parameter waren zusätzlich manuell erstellte Segmentierungs- und Zuordnungsergebnisse für die Trainingssequenzen verfügbar.

Für den Vergleich des CellMigrationAnalyzer mit den Ergebnissen des Wettbewerbes wurden die Parameter σ und γ unabhängig auf den beiden Trainingssequenzen, wie oben beschrieben, systematisch variiert. Die Parameterkombinationen, welche jeweils im Vergleich zu den Referenzsegmentierungen das beste mittlere F-Maß erzielten wurden für das Auswerten der Testsequenzen verwendet. Die Resultate der Testsequenzen wurden anschließend an Dr. Martin Maška geschickt. Dieser analysierte die Daten mit den für den Wettbewerb entwickelten Programmen zum Auswerten der Ergebnisse. Eine Übersicht aller ergebnisrelevanter Parameter ist in Tabelle 14 gezeigt.

Parameter	Wert	
	Sequenz 1	Sequenz 2
Standardabweichung Gaußfilter (σ)	2	3
Exponent Gammakorrektur (γ)	0.7	0.1
minimale Pixelzahl der vorsegmentierten Regionen	300	300
minimale Pixelzahl der detektierten Regionen	500	1000
maximale Anzahl Level-Set-Iterationen	1000	1000
minimale Trajektorienlänge	2	2
maximal zulässige relative Flächenänderung (T_a)	0.5	0.5

Tabelle 14: Programmparameter des *CellMigrationAnalyzer*, welche für die Bearbeitung der beiden Testsequenzen des *Cell Tracking Challenge 2013-Datensatzes* verwendet wurden.

Die Evaluierung der Ergebnisse erfolgte getrennt für die Segmentierungs- und die Zuordnungsergebnisse. Die Segmentierungsqualität wurde mit Hilfe des Jaccard-Ähnlichkeitsindex (J) bestimmt:

$$J(R, S) = \frac{|R \cap S|}{|R \cup S|} \quad (32)$$

Hierbei stellt R die Menge der Pixel einer manuell bestimmten Referenzregion und S die Menge der Pixel einer Region, welche mit dem zu bewertenden Segmentierungsalgorithmus bestimmt wurde, dar. Dabei wurde festgelegt, dass eine automatisch segmentierte Region dann mit einer Referenzregion korrespondiert, wenn die Schnittmenge der Pixelkoordinaten beider Regionen mehr als die Hälfte der Pixelkoordinaten der Referenzregion abdeckt, also wenn gilt:

$$|R \cap S| > 0.5 \cdot |R| \quad (33)$$

Das Maß für die Gesamtsegmentierungsqualität wurde als SEG bezeichnet und aus den Mittelwerten der Jaccard-Ähnlichkeitsindizes aller Referenzregionen berechnet. Der Wertebereich von SEG liegt im Intervall $[0, 1]$, wobei das Maximum erreicht würde, wenn alle Referenzregionen exakt mit automatisch segmentierten Regionen übereinstimmen. Ähnlich wie in der oben beschriebenen Beurteilung der Segmentierungsqualität mittels des F-Maßes, fließen auch bei dieser Art der Auswertung, gegebenenfalls auftretende automatisch segmentierte Regionen, welche zu keiner Referenzregion korrespondieren, nicht in die Bewertung ein.

Um die Zuordnungsgenauigkeit zu bestimmen, wurden die berechneten Trajektorien als gerichtete azyklische Graphen betrachtet. Dabei stellten die zu bestimmten Zeitpunkten detek-

tierten Zellen die Knoten des Graphen dar. Die Kanten des Graphen repräsentierten die Zuordnung der detektierten Zellen zwischen den Bildern. Dabei wurde zwischen zwei Arten von Kanten unterschieden. Kanten, welche Zellen mit gleichen Markierungen verbinden wurden als „track links“ bezeichnet. Sie definierten die Bewegung einer Zelle. Kanten zwischen Zellen mit verschiedenen Markierungen wurden hingegen als „parent links“ bezeichnet. Diese Kanten repräsentierten Zellteilungsereignisse. Das als TRA bezeichnete Maß für die Zuordnungsleistung spiegelte den Aufwand wider, den Ergebnisgraphen eines Zuordnungsalgorithmus in den Graphen der manuellen Referenzzuordnung zu überführen. Die verschiedenen Operationen (Hinzufügen/Spalten/Löschen von Knoten, sowie Hinzufügen/Löschen von Kanten und Ändern der Semantik von Kanten) wurden dabei unterschiedlich gewichtet und das Ergebnis war die kleinste gewichtete Summe der benötigten Änderungsoperationen normalisiert auf die Anzahl der Referenzknoten. Das Programm zum Berechnen des TRA-Maßes benötigte für jede detektierte Zelle die Information darüber, welche Markierung deren jeweilige Mutterzelle besaß. Für Zellen deren Mutterzellmarkierungen nicht bekannt waren, wurde der Wert Null verwendet. Da der CellMigrationAnalyzer solche Abstammungsinformationen nicht auswertet, wurde als Wert für die jeweiligen Mutterzellen immer Null verwendet.

Im Gegensatz zu den vorher beschriebenen Auswertungen, wurden für die Bildsequenzen des Wettbewerbes detektierte Zellen, welche die Bildränder berührten, nicht entfernt, da diese auch in den Referenzsegmentierungen nicht ausgeschlossen wurden. Weiterhin wurde die Farbtiefe von 16 Bit in allen Bildsequenzen vor der Bearbeitung mit den CellMigrationAnalyzer auf 8 Bit reduziert, da dies bessere Ergebnisse erzielte.

Die Ergebnisse für SEG und TRA der Cell Tracking Challenge 2013 auf den Testbildern des Datensatzes C2DL-MSD für die Programme der teilnehmenden Gruppen, sowie die Ergebnisse des CellMigrationAnalyzer sind in Tabelle 15 aufgelistet. Die Werte der teilnehmenden Gruppen stammen von der Website des Wettbewerbes (www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Results_First_CTC.html).

Die Ergebnisse des CellMigrationAnalyzer für die Testsequenzen waren sehr verschieden. Für die erste Sequenz erhielten sowohl SEG als auch TRA schlechtere Werte als für je vier der fünf Programme der Teilnehmer des Wettbewerbes. Für die zweite Sequenz hingegen waren beide Maße lediglich geringfügig schlechter als die des Programms, welches die besten Ergebnisse auf dieser Sequenz erzielt hatte.

Das schlechtere Ergebnis für die erste Testsequenz lässt sich auf die geringe Anzahl detektierter Zellen durch die Binarisierung zurückführen. Dies wiederum ist auf die Verwendung des vergleichsweise großen Wertes von 0.7 für den Parameter γ zurückzuführen. Dieser Wert

Programm	SEG		TRA		# Parameter
	Sequenz 1	Sequenz 2	Sequenz 1	Sequenz 2	
COM-US	0.06	0.15	11.06	11.16	16
HEID-GE	0.34	0.43	5.06	4.38	9
KTH-SE	0.49	0.66	2.52	1.91	10
PRAG-CZ	0.15	0.21	8.70	7.77	8
UPM-ES	0.41	0.19	4.26	8.15	10
CMA	0.12	0.63	9.94	2.52	7

Tabelle 15: Segmentierungs- (SEG) und Zuordnungsergebnisse (TRA) für die beiden Testsequenzen des Cell Tracking Challenge 2013-Datensatzes C2DL-MSD der Teilnehmer des Wettbewerbes und des CellMigrationAnalyzer (CMA). Die Spalte # Parameter beziffert die Anzahl der zu variierenden Parameter.

erzielte auf dem Trainingsdatensatz zusammen mit $\sigma = 2$ das beste Segmentierungsergebnis in Bezug auf F-Maß. Allerdings war, im Vergleich zur Trainingssequenz, in der Testsequenz eine Zelle vorhanden, deren Intensitätswerte deutlich höher als die der übrigen Zellen war. Durch die geringe Hervorhebung schwächerer Fluoreszenzsignale mittels γ wurden bei der Binarisierung der Bilder der Testsequenz mittels Otsu-Algorithmus nur wenige, besonders helle Pixel als Vordergrund detektiert. Durch die Verwendung eines kleineren Wertes für γ konnten erwartungsgemäß deutlich mehr Zellen detektiert werden. Das bedeutet, das schlechte Ergebnis kann auf die unterschiedlichen Charakteristika von Trainings- und Testsequenz zurückgeführt werden. Für einen fairen Vergleich der Programme wurden dennoch die auf den Trainingsdaten optimierten Parameter verwendet. Für die zweite Testsequenz wurde $\sigma = 3$ und $\gamma = 0.1$ als bester Parametersatz ermittelt, welcher auf der entsprechenden Testsequenz das zweitbeste Ergebnis im Vergleich zu den Wettbewerbsteilnehmern erzielte.

Die Abbildungen 19 und 20 zeigen jeweils ein Beispielfeld und die dazugehörigen Segmentierungsergebnisse der beiden Testsequenzen des C2DL-MSD-Datensatzes.

Die Ergebnisse des CellMigrationAnalyzer hängen von der Einstellung von sieben Parametern ab. Diese Zahl umfasst alle Parameter, welche die Bildverarbeitung beeinflussen, nicht jedoch Kalibrierparameter, wie beispielsweise physikalische Pixelgrößen oder das auszuwertende Eingabevideo. Damit erfordert die Verwendung des CellMigrationAnalyzer weniger Parameteranpassungen als die der anderen Programme.

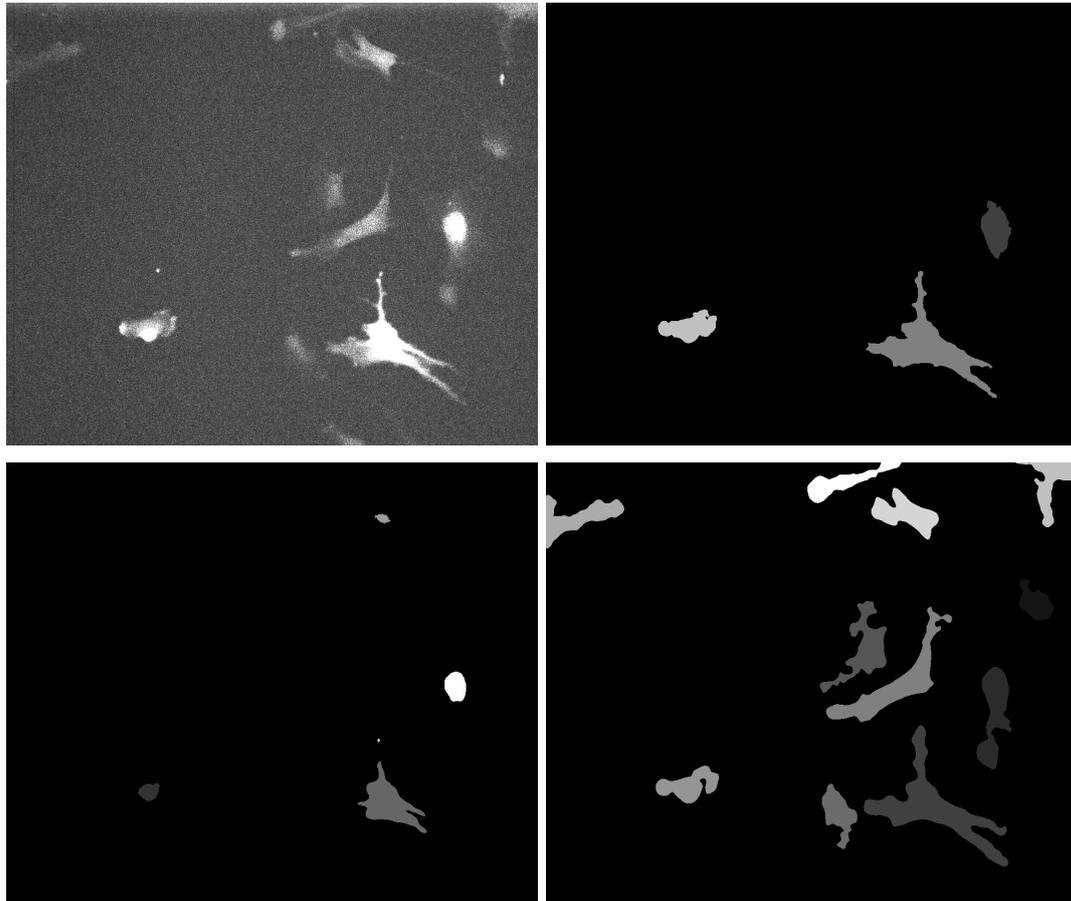


Abbildung 19: Beispielbild der ersten zum Testen verwendeten Bildsequenz des „Cell Tracking Challenge“ Datensatzes „C2DL-MS“ (links oben) und entsprechende Segmentierungsergebnisse des CellMigrationAnalyzer (rechts oben), sowie der für diesen Datensatz letzt- (links unten) und erstplatzierten (rechts unten) Programme. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.

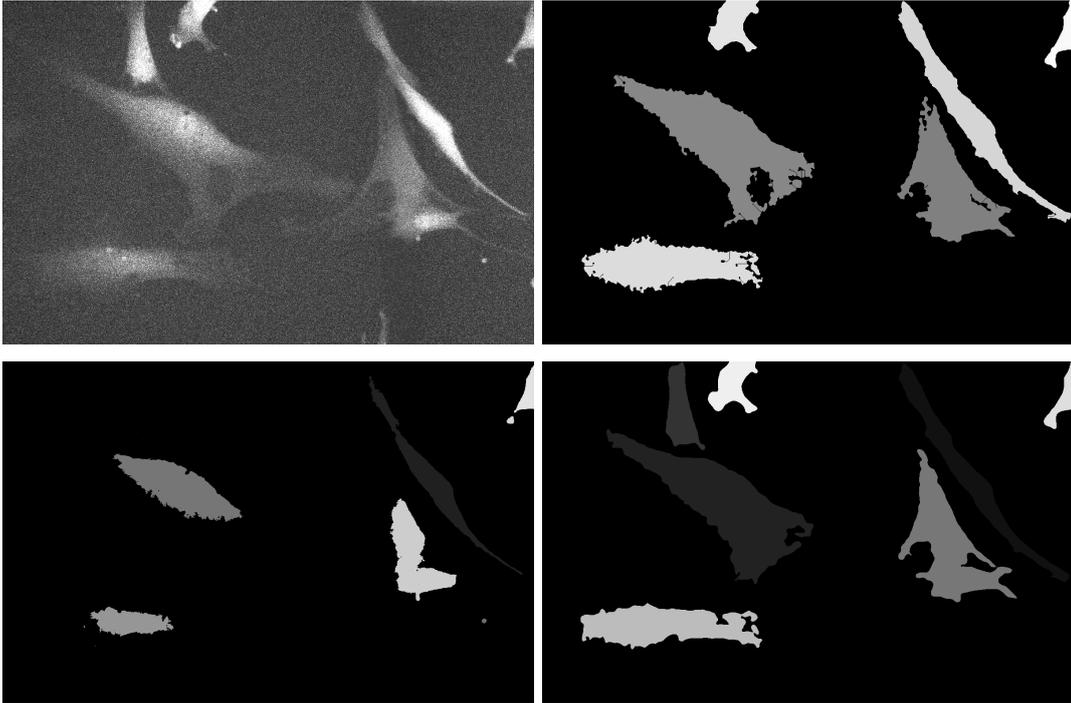


Abbildung 20: Beispielbild der zweiten zum Testen verwendeten Bildsequenz des „Cell Tracking Challenge“ Datensatzes „C2DL-MS-C“ (links oben) und entsprechende Segmentierungsergebnisse des CellMigrationAnalyzer (rechts oben), sowie der für diesen Datensatz letzt- (links unten) und erstplatzierten (rechts unten) Programme. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.

Diese Ergebnisse zeigen, dass der CellMigrationAnalyzer nicht nur auf die Bildsequenzen anwendbar ist, für die er entwickelt wurde. Die Ergebnisse sind mit denen anderer aktueller Entwicklungen vergleichbar.

Um die Segmentierungs- und Zuordnungsqualität der Programme des Wettbewerbes auf den Videos, welche für die Entwicklung des CellMigrationAnalyzer verwendet wurden, fair beurteilen zu können, müssten geeignete Parameterkombinationen verwendet werden. Aufgrund der Vielzahl der verwendeten Parameter würde dies einen sehr hohen Aufwand bereiten. Zudem waren nur drei der fünf Teilnehmerprogramme, welche auch Ergebnisse für den Datensatz C2DL-MSD eingereicht hatten, verfügbar. Davon konnten nur zwei (Gruppen COM-US und KTH-SE) ausgeführt werden, da das dritte dieser Programme (Gruppe PRAG-CZ) eine Matlab-Laufzeitumgebung (The MathWorks Inc., Natick, Massachusetts, USA) in Version R2009a benötigte, welche nicht mehr vom Hersteller erhältlich ist. Daher wurde auf Vergleiche dieser Art verzichtet.

Die geringere Anzahl Parameter, welche für die Verwendung des CellMigrationAnalyzer anzupassen ist, erleichtert dessen Bedienbarkeit im Vergleich zu den Programmen des Wettbewerbes. Weiterhin profitiert der CellMigrationAnalyzer als Bestandteil von MiToBo von der breiten Plattformverfügbarkeit von ImageJ beziehungsweise Fiji. Die meisten Teilnehmerprogramme der Cell Tracking Challenge waren hingegen nur für Microsoft Windows Betriebssysteme erhältlich.

3.1.2 Evaluation des ScratchAssayAnalyzer

Um die Wunddetektionsqualität des ScratchAssayAnalyzer zu optimieren und zu bewerten, wurden die Segmentierungsergebnisse mit Ergebnissen verglichen, welche von Experten manuell erstellt wurden. Zur Optimierung und Bewertung der Fähigkeit des Programms, Bilder ohne erkennbare Wunde zu detektieren, wurden Bilder mit und ohne Wunde verwendet, um SVMs zu trainieren. Die Klassifikationsgüte dieser SVMs wurde anschließend mittels Kreuzvalidierung bestimmt. Dies wurde separat für Durchlichtaufnahmen, sowie für Fluoreszenzaufnahmen durchgeführt. Dazu wurden aus Bildern dieser beiden Aufnahmearten je drei nicht-überlappende Datensätze erstellt. Diese werden im Folgenden als Trainings-, Test- und Klassifikationsdatensätze bezeichnet.

Der Trainingsdatensatz wurde jeweils zum Optimieren der Segmentierung verwendet. Dazu wurden die Werte für die Standardabweichung des Gaußfilters (σ) und die Seitenlänge des Fensters, welches für die Berechnung der lokalen Entropie verwendet wurde (e), systematisch verändert und die Segmentierungsergebnisse mit den manuell gekennzeichneten Wundflächen

verglichen. Hierfür wurde für jedes Pixel eines Bildes verglichen, welcher Region (Wunde oder Zellen) es nach automatischer beziehungsweise manueller Referenzsegmentierung zugewiesen wurde. Die Anzahl der Pixel, welche in beiden Segmentierungen der Wunde zugeordnet wurden, wird im Folgenden als TP (true positives) bezeichnet. Wurde ein Pixel, welches gemäß der Referenzsegmentierung zur Wunde gehörte, vom Programm der Zellregion zugewiesen, so wurde dieses Pixel zu den FN (false negatives) gezählt. Pixel, welche vom Programm als Wundpixel segmentiert wurden, in der Referenzsegmentierung aber als Zellpixel markiert waren, bildeten die Zahl FP (false positives). Daraus wurden die Evaluierungsmaße Precision (\mathcal{P}), Recall (\mathcal{R}) und das F-Maß (\mathcal{F}) wie in Gleichung 30 berechnet.

Unter der Annahme, dass die Referenzsegmentierung der wahren Einteilung der Pixel entspricht, beschreibt Precision, welcher Anteil der vom Programm als Wundpixel detektierten Pixel tatsächlich zur Wunde gehören. Recall definiert, welcher Anteil der wahren Wundpixel als solche detektiert wurden. Das F-Maß als gewichtetes Mittel dieser beiden Maße kondensiert diese zu einer Zahl, deren Maximalwert von eins erreicht würde, wenn alle Pixel der Vergleichssegmentierungen, jeweils den gleichen Regionen zugeordnet wurden.

Anhand der Ergebnisse der Parametervariation wurde für Durchlicht-, sowie Fluoreszenzaufnahmen jeweils der Parametersatz ausgewählt, welcher das mittlere F-Maß maximierte. Dieser Parametersatz wurde anschließend für die Evaluierung der Testdaten verwendet. Für das Training der SVMs wurde jeweils der Klassifikationsdatensatz verwendet. Die Parameteranpassung der SVMs erfolgte hierbei über Leave-One-Out-Kreuzvalidierungen. Hierfür wurden pro SVM-Parameterkombination so viele SVMs trainiert wie Bilder im Datensatz vorhanden waren. Für das Validieren einer SVM wurde jeweils ein Bild des Klassifikationsdatensatzes herausgesucht. Die übrigen Bilder wurden für das Training dieser SVM verwendet. Im Anschluss wurde das zuvor ausgesuchte Bild mit dieser SVM klassifiziert. Pro Parametersatz wurde jedes Bild des Klassifikationsdatensatzes genau einmal zum Klassifizieren verwendet. Die Anzahl der korrekt klassifizierten Bilder ergab die Klassifikationsrate, die als Gütemaß für die Leistung der SVM mit dem gewählten Parametersatz verwendet wurde.

Durchlicht Zur Abschätzung der Qualität des ScratchAssayAnalyzer angewendet auf Durchlichtbilder, wurden Scratch-Assays mit Zellen der Osteosarkomzelllinie U-2OS durchgeführt. Mikroskopaufnahmen erfolgten dabei direkt nach Erzeugung der Wunden (0 h), sowie nach 2 h, 4 h, 6 h, 8 h, 12 h und nach 24 h.

Um die Segmentierung des ScratchAssayAnalyzer für Durchlichtbilder zu optimieren, wurde ein Trainingsdatensatz bestehend aus sieben Bildern erstellt. Diese Bilder zeigen die Entwick-

lung einer Wunde über einen Zeitraum von 24 h. Die Konturen der Wunden in diesen Bildern wurden von zwei Expertinnen (Dr. rer. nat. Anne Zirkel und Dr. rer. nat. Kristin Wächter) manuell markiert. Abbildung 21 zeigt die Konturen der manuell bestimmten Wunden für einen Ausschnitt eines Bildes des Trainingsdatensatzes. Um die Ähnlichkeit der manuellen Segmentierungen zueinander zu quantifizieren, wurden Precision, Recall und F-Maße berechnet. Die Verteilungen dieser Werte für die Bilder des Trainingsdatensatzes sind ebenfalls in Abbildung 21 dargestellt. Es zeigte sich, dass die Segmentierungen der Experten zueinander eine hohe Übereinstimmung aufwiesen.

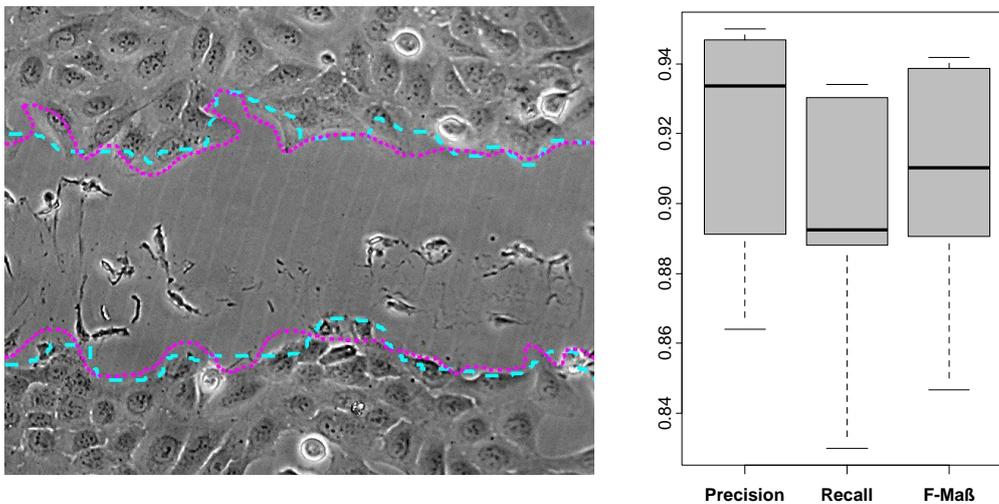


Abbildung 21: Vergleich manueller Segmentierungen von Durchlichtaufnahmen von Scratch-Assays. Links: manuelle Segmentierungen für einen Ausschnitt eines Bildes des Trainingsdatensatzes. Die Konturen der manuell bestimmten Wundflächen sind mit unterbrochenen Linien (magenta beziehungsweise cyan) gekennzeichnet. Zur besseren Darstellung des Bildes wurde der Kontrast nachträglich erhöht. Rechts: Verteilung der Precision-, Recall- und F-Maße, welche auf dem Trainingsdatensatz für die beiden manuellen Segmentierungen berechnet wurden.

Aus den manuellen Segmentierungen wurde für jedes Bild eine Referenzsegmentierung erstellt, deren Pixel dann zur Wunde gehörten, wenn sie von beiden Expertinnen als zur Wunde gehörig markiert worden waren. Diese Referenzsegmentierungen wurden anschließend für die Vergleiche mit der automatischen Segmentierung verwendet. Für die Parameteroptimierung wurde der Parameter σ von 0 bis 10, bei einer Schrittgröße von 1 und der Parameter e wurde von 3 bis 69, bei einer Schrittgröße von 2 variiert. Insgesamt wurden somit 374 verschiedene Parameterkombinationen getestet. Die Kombination aus $\sigma = 7$ und $e = 27$ lieferte, gemittelt über alle Bilder des Trainingsdatensatzes, das höchste durchschnittliche F-Maß von 0.820. Die

durchschnittlichen Werte für Precision und Recall lagen bei diesem Parametersatz bei 0.870 beziehungsweise 0.779.

Der Datensatz zum Testen der mithilfe des Trainingsdatensatzes bestimmten Segmentierungsparameter, bestand aus 70 Bildern, welche die Entwicklung von 10 Wunden zu jeweils sieben verschiedenen Zeitpunkten zeigten. Bei fünf dieser 10 Experimente wurde dem Medium der Zellen fötales Rinderserum (FBS) zugesetzt, bei den restlichen fünf Experimenten nicht. Die Referenzsegmentierungen des Testdatensatzes wurden wie für den Trainingsdatensatz erstellt. Abbildung 22 zeigt die Verteilung der Precision-, Recall- und F-Maße für alle 70 Bilder aufgetrennt nach Zeitpunkten als Boxplots. Für die Ermittlung dieser Werte wurde der Erken-

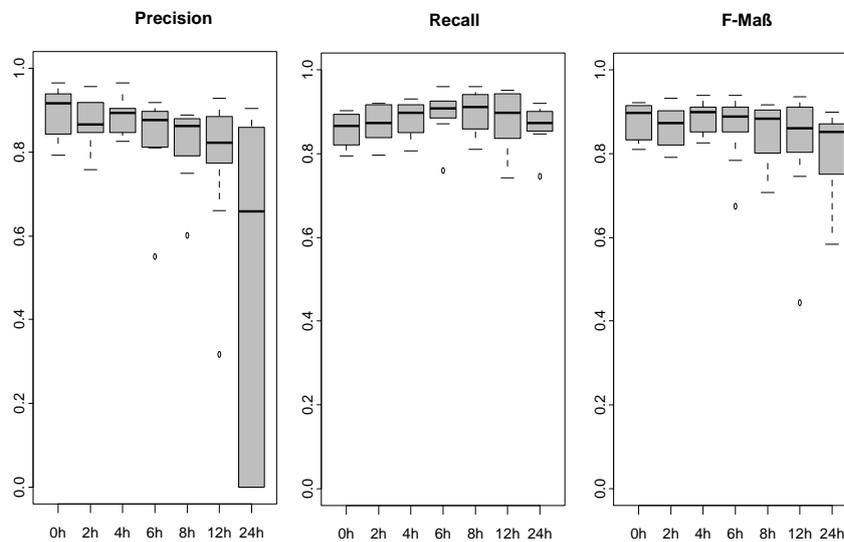


Abbildung 22: Verteilung der Precision-, Recall- und F-Maße für die 70 Scratch Assay Bilder des Testdatensatzes für Durchlichtaufnahmen. Für die Ermittlung der Segmentierungsqualität wurde der Erkennungsmechanismus für geschlossene Wunden nicht verwendet.

nungsmechanismus für geschlossene Wunden nicht verwendet. Da jedoch einige der Wunden nach 12 h beziehungsweise 24 h bereits geschlossen waren, stellen die Segmentierungsergebnisse auf diesen Bildern keine Einteilung in Zellen und Hintergrund dar. Dies erklärt die teilweise sehr niedrigen Precision-Werte für die letzten beiden Aufnahmezeitpunkte. Abbildung 23 zeigt zwei Beispielbilder, auf denen die mittels des optimierten Parametersatzes automatisch detektierten und manuell bestimmten Wundflächen zu sehen sind.

Zum Trainieren und Evaluieren der SVM für das Erkennen von Bildern ohne Wunde, wurde ein weiterer Datensatz, bestehend aus 30 Bildern, erstellt. Fünf dieser Bilder enthielten

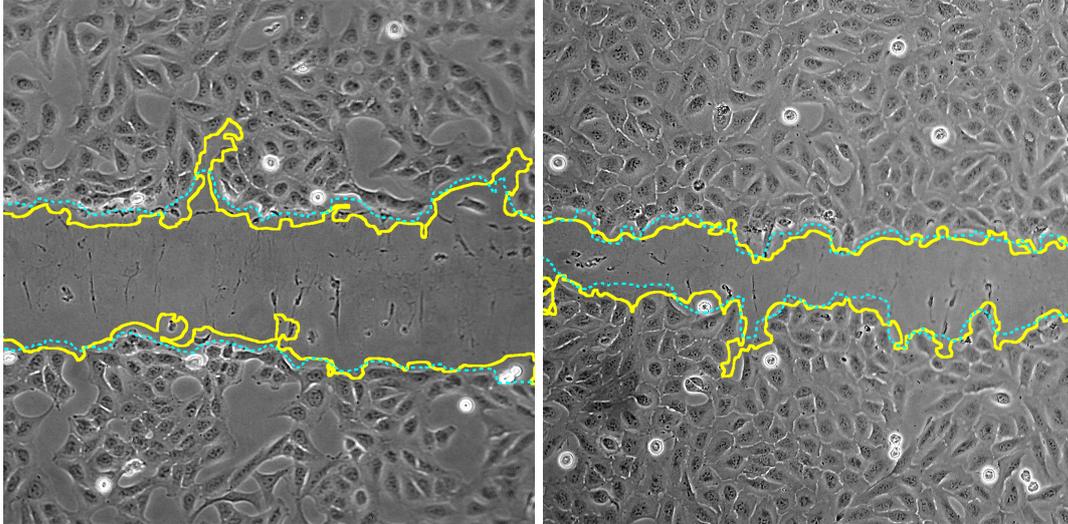


Abbildung 23: *Scratch-Assay-Segmentierungsergebnisse von Durchlichtaufnahmen im Vergleich zu manuell bestimmten Wundkonturen. Die Konturen der automatisch detektierten Wunden sind mit durchgezogenen Linien (gelb) gekennzeichnet, die manuell bestimmten Referenzkonturen mit unterbrochenen Linien (cyan). Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.*

keine Wunde, die restlichen 25 Bilder zeigten Wunden unterschiedlicher Größe. Diese Aufteilung entsprach ungefähr der Verteilung solcher Bilder in den betrachteten Scratch-Assay-Experimenten. Mit Hilfe dieses Datensatzes wurden SVMs mit linearem Kernel und variierendem Regularisierungsparameter (C) trainiert und mittels Leave-One-Out-Kreuzvalidierung ausgewertet. Die Ergebnisse der Validierungen sind in Tabelle 16 aufgelistet. Für die Klassifikation der Durchlichtbilder wurde eine SVM gewählt, deren Regularisierungsparameter 50 betrug. Diese klassifizierte lediglich ein Bild falsch und erreichte damit eine Klassifikationsrate von ungefähr 97%. Dies zeigt, dass eine SVM mit linearem Kernel, unter Verwendung der ausgesuchten Histogrammeigenschaften (s. Abschnitt 2.5), sehr gut geeignet ist, Bilder ohne erkennbare Wunde zu detektieren.

In Abbildung 24 sind die mittleren pro Zeitpunkt geschlossenen Wundflächen der beiden Populationen des Testdatensatzes mit aktivierter Wunderkennung dargestellt. Es ist deutlich zu erkennen, dass der Wundschluss der Zellen, denen Serum vorenthalten wurde, reduziert war. Dies entsprach den Erwartungen, da das Serum Faktoren enthält, welche auf die Proliferation und Migration der Zellen stimulierend wirken. Der reduzierte Wundschluss wurde durch die Ergebnisse des ScratchAssayAnalyzer widerspiegelt. Die im Vergleich zu den Referenzsegmentierungen konsistent geringer bestimmten Flächenabnahmen des ScratchAssayAnalyzer zeigen jedoch, dass dieser im Allgemeinen größere Wundflächen berechnete. Dennoch waren

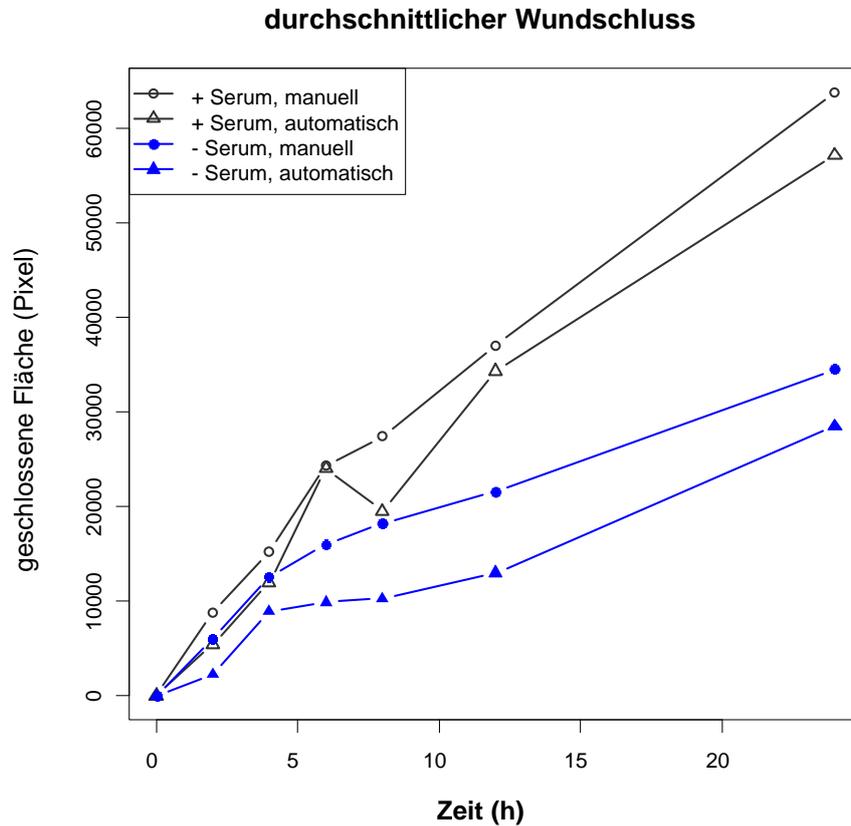


Abbildung 24: Vergleich der manuell und automatisch bestimmten Wundflächenentwicklung aus Durchlichtaufnahmen von Scratch-Assay-Experimenten mit U-2OS-Zellen. Gezeigt sind die durchschnittlich pro Zeitpunkt geschlossenen Wundflächen der manuellen und automatischen Auswertungen in Bezug zum Beginn der Experimente (0 h). Es wurden zwei Populationen von Zellen betrachtet. Dem Medium der einen Population wurde fötales Rinderserum zugesetzt (+ Serum), dem Medium der anderen Population nicht (- Serum). Je Zeitpunkt erfolgten fünf unabhängige Messungen pro Population.

C	TP	TN	FP	FN
1	25	0	5	0
5	25	3	2	0
10	25	3	2	0
25	25	4	1	0
50	25	4	1	0
75	25	3	2	0
100	25	3	2	0
500	25	0	5	0

Tabelle 16: Validierungsergebnisse der Klassifikation von Durchlicht-Scratch-Assay-Bildern mit SVMs unter Verwendung eines linearen Kerns und verschiedenen Werten des Regularisierungsparameters C . Aufgelistet sind die Anzahlen der Bilder, welche unter Verwendung des entsprechenden Wertes von C jeweils korrekt als Bilder mit enthaltener Wunde (TP) oder ohne Wunde (TN), sowie fälschlicherweise als Bilder mit (FP) oder ohne (FN) Wunde klassifiziert wurden.

die Abweichungen der manuellen und automatischen Auswertungen im Allgemeinen nur gering. Insgesamt zwei Bilder wurden falsch klassifiziert. Dabei handelte es sich um Bilder des selben Experimentes mit Zellen, welche Serum erhalten hatten. Die Bilder, welche die Wundentwicklung dieses Experimentes nach sechs beziehungsweise 12 Stunden zeigten, wurden fälschlicherweise als Bilder ohne Wunde klassifiziert. Dies erklärt die Abnahme der automatisch bestimmten, durchschnittlich geschlossenen Wundfläche nach acht Stunden im Vergleich zum vorherigen Zeitpunkt. Insgesamt zeigen die Ergebnisse jedoch, dass der ScratchAssay-Analyser gut geeignet ist, die Entwicklung der Wundflächen von Scratch-Assay-Experimenten in Durchlichtmikroskopieaufnahmen zu bestimmen.

Fluoreszenz Um die Segmentierungs- und Klassifizierungsqualität des ScratchAssayAnalyser für Aufnahmen von fluoreszierenden Zellen bewerten zu können, wurden weitere Scratch-Assay-Experimente mit Zellen der Zelllinie U-2OS durchgeführt. Diese Experimente zeigten entweder Zellen, welche das grün-fluoreszierende Protein (GFP) exprimierten oder Zellen, welche das rot-fluoreszierende Protein (RFP) exprimierten. Die Bilder dieser Experimente wurden verwendet, um überlappungsfreie Datensätze zum Evaluieren des ScratchAssayAnalyser zu erstellen.

Die beiden Datensätze für das Optimieren und Testen der Segmentierungsparameter bestanden aus jeweils zehn Bildern von denen je fünf GFP-exprimierende Zellen und die restlichen fünf RFP-exprimierende Zellen zeigten. Die Konturen der Wunden in diesen insgesamt 20 Bildern wurden von einer Expertin (Dipl.-Biochem. Britta Mühleck) manuell markiert. Für die Parameteroptimierung wurde σ von 0 bis 10, bei einer Schrittgröße von 1 und e wurde von 3 bis 69, bei einer Schrittgröße von 2 variiert. Insgesamt wurden somit 374 verschiedene Parameterkombinationen getestet. Durchschnittliche Precision-, Recall- und F-Maße wurden sowohl für die GFP- beziehungsweise RFP-exprimierenden Zellen separat, als auch für alle Bilder der Datensätze zusammen bestimmt. Die jeweiligen Parameterkombinationen, welche die höchsten durchschnittlichen F-Maße erzielten sind in Tabelle 17 aufgelistet.

	Kombination	$\varnothing \mathcal{P}$	$\varnothing \mathcal{R}$	$\varnothing \mathcal{F}$
GFP	$\sigma = 7, e = 39$	0.967	0.957	0.962
RFP	$\sigma = 6, e = 37$	0.975	0.958	0.966
GFP + RFP	$\sigma = 6, e = 37$	0.968	0.960	0.964

Tabelle 17: Kombinationen der Parameter σ und e , welche auf den Trainingsdaten GFP- und RFP-exprimierender Zellen, sowie aller Bilder zusammen (GFP + RFP), das jeweils höchste durchschnittliche F-Maß ($\varnothing \mathcal{F}$) erzielten und die dazugehörigen durchschnittlichen Precision- ($\varnothing \mathcal{P}$) und Recall- ($\varnothing \mathcal{R}$) Werte

Da die Parameterkombination $\sigma = 6$ und $e = 37$ sowohl für die RFP-, als auch für die Kombination aus GFP- und RFP-exprimierenden Zellen das beste durchschnittliche F-Maß ergab und diese Kombination für die GFP-exprimierenden Zellen mit 0.961 ein nur unwesentlich niedrigeres durchschnittliches F-Maß als das optimale ergab, wurde diese Parameterkombination sowohl für GFP- als auch für RFP-exprimierende Zellen verwendet. Tabelle 18 zeigt die mittleren Precision-, Recall-, und F-Maße für die Bilder des Testdatensatzes. Bilder der GFP-beziehungweise RFP-exprimierenden Zellen wurden sowohl separat, als auch zusammen analysiert. Die hohen Werte des F-Maßes von jeweils 0.975 zeigen, dass die Referenzwundränder mit Hilfe des ScratchAssayAnalyzer sehr gut bestimmt wurden.

Abbildung 25 zeigt die Konturen der manuell und automatisch ermittelten Wunden an je einem Beispielbild für GFP- und RFP-exprimierende Zellen.

Um von den Fluoreszenzbildern solche zu erkennen, welche keine Wunde zeigten, wurde ein weiterer Datensatz erstellt. Dieser bestand aus 24 Bildern mit erkennbarer Wunde, sowie 12 Bildern, in denen keine Wunde erkennbar war. Diese Aufteilung entsprach ungefähr der Vertei-

	$\varnothing \mathcal{P}$	$\varnothing \mathcal{R}$	$\varnothing \mathcal{F}$
GFP	0.999	0.953	0.975
RFP	0.996	0.954	0.975
GFP + RFP	0.998	0.954	0.975

Table 18: Durchschnittliche Precision- ($\varnothing \mathcal{P}$), Recall- ($\varnothing \mathcal{R}$) und F-Maßwerte ($\varnothing \mathcal{F}$) für die Segmentierung der Bilder von GFP- beziehungsweise RFP-exprimierenden Zellen, sowie aller Bilder (GFP + RFP) des Testdatensatzes unter Verwendung der Parameterkombinationen ($\sigma = 6$, $e = 37$)

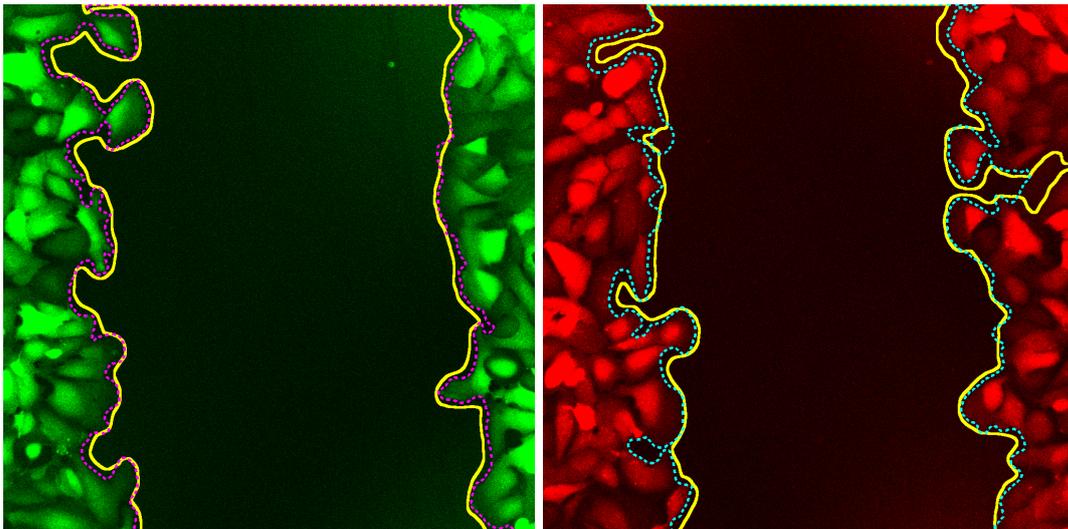


Abbildung 25: Segmentierungsergebnisse des ScratchAssayAnalyzer im Vergleich zu manuell bestimmten Wundkonturen. Die Konturen der automatisch detektierten Wunden sind mit durchgezogenen Linien (gelb) gekennzeichnet, die manuell bestimmten Konturen mit unterbrochenen Linien (magenta beziehungsweise cyan). Links: GFP-exprimierende Zellen; rechts: RFP-exprimierende Zellen. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht

lung solcher Bilder in den betrachteten Scratch-Assay-Experimenten, welche jeweils über einen Zeitraum von 16 Stunden aufgenommen worden waren. Die Hälfte der Bilder mit beziehungsweise ohne Wunde zeigte GFP-exprimierende Zellen, die andere Hälfte RFP-exprimierende Zellen. Eine SVM mit linearem Kernel und einem Regulierungsparameter (C) von 100 wurde mit diesen Daten trainiert und mittels Leave-One-Out-Kreuzvalidierung validiert. Alle 36 Bilder des Klassifikationsdatensatzes wurden mit dieser SVM korrekt klassifiziert. Die daraus resultierende Klassifikationsrate von 100 % zeigt, dass die Klassifikation mit Hilfe einer SVM und den gewählten Eigenschaften (s. Abschnitt 2.5) sehr gut geeignet ist, in Aufnahmen von Scratch-Assay-Experimenten fluoreszenzmarkierter Zellen Bilder zu identifizieren, in denen keine Wunde vorhanden ist.

Vergleich mit anderer Software Neben dem hier vorgestellten ScratchAssayAnalyzer gibt es einige weitere Programme, deren Zweck die automatisierte Auswertung von Scratch-/Gap-Closure-Assays ist. Zu diesen Programmen gehören CellProfiler [67] und TScratch [39]. CellProfiler ist ein universelles Bildverarbeitungsprogramm, welches speziell für die Analyse von Bildern biologischer Zellen entwickelt wurde. Über sogenannte Pipelines können Basisfunktionalitäten des Programmes zu komplexeren Anwendungen kombiniert werden. Die Vorverarbeitung der für CellProfiler erhältlichen „Wound Healing Pipeline“ ähnelt dem Vorverarbeitungsschritt des ScratchAssayAnalyzer. Zunächst werden die Eingabebilder zu 8-Bit-Graustufenbildern konvertiert, dann werden die Bilder mittels Gaußfilter geglättet. Die Segmentierung der Bilder in Zellen und Wundfläche basiert jedoch lediglich auf der Anwendung von Intensitätsschwellwerten. TScratch wurde speziell für die Analyse von Scratch-Assays entwickelt und verwendet die sogenannte Curvelet-Transformation [12], um intensitätsunabhängig über Textureigenschaften Zellen und Wunde zu identifizieren. Der Versuch, die oben vorgestellten Durchlichtbilder zum Testen der Segmentierungsqualität mit beiden Programmen auszuwerten, führte mit keiner der gewählten Parameterkombinationen zu zufriedenstellenden Ergebnissen. Nach Kontaktaufnahme mit den Entwicklern von TScratch, wiesen diese daraufhin, dass ihr Programm Bilder mit höherem Kontrast und weniger Zelltrümmern benötigt. Deswegen wurden zum Vergleichen Beispielbilder verwendet, welche jeweils für CellProfiler und TScratch zum Download angeboten wurden. Da es jedoch für diese Bilder keine als korrekt segmentiert anzusehenden Referenzergebnisse gab, wurden lediglich qualitative Vergleiche durchgeführt.

Abbildung 26 zeigt die Segmentierungsergebnisse von CellProfiler und dem ScratchAssayAnalyzer auf den beiden für die „Wound Healing Pipeline“ erhältlichen Beispielbildern. Diese

Bilder zeigen die Wundfläche eines Scratch-Assay-Experimentes direkt nach Erzeugung der Wunde, sowie 24 Stunden danach.

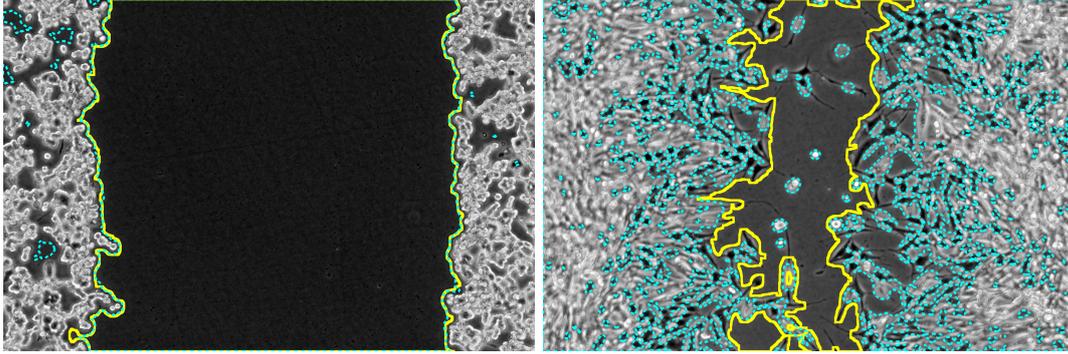


Abbildung 26: Qualitativer Vergleich des ScratchAssayAnalyzer und der CellProfiler Wound Healing Pipeline. Links: Wundfläche zu Beginn des Experimentes. Rechts: Wundfläche 24 Stunden nach Beginn des Experimentes. Die Konturen der mit dem ScratchAssayAnalyzer detektierten Wunden sind mit durchgezogenen Linien (gelb) gekennzeichnet, die Konturen der Wundflächen, welche mit CellProfiler bestimmt wurden, mit unterbrochenen Linien (cyan). Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.

Für CellProfiler wurden die Voreinstellungen der Pipeline verwendet, da davon ausgegangen wurde, dass diese für die angebotenen Beispielbilder optimiert wurden. Für den ScratchAssayAnalyzer wurden einige Parameterkombinationen getestet und $\sigma = 3$ und $e = 31$ gewählt, da diese Kombination ein, gemäß visueller Betrachtung, gutes Ergebnis für beide Bilder erzielte. Die segmentierten Wundränder für den Zeitpunkt Null waren bei beiden Programmen nahezu identisch, jedoch lieferte CellProfiler auch Teile von Löchern in den Zellbereichen als Wundregion. Allerdings wurden nicht alle vorhandenen Löcher als Wundfläche segmentiert. Der ScratchAssayAnalyzer bezog, aufgrund seiner topologieerhaltenden Segmentierungsstrategie, keines dieser Löcher in die Wundfläche ein. Die Ergebnisse für das zweite Bild unterschieden sich deutlich. Während der ScratchAssayAnalyzer eine einzige zusammenhängende Region als Wunde zurücklieferte, bestand das Ergebnis des CellProfiler aus vielen nicht-zusammenhängenden Bereichen. Dabei wurden auch Teile von Zellen, welche individuell in den Wundbereich migriert waren, als Teil der Zellfläche segmentiert. Die vom ScratchAssayAnalyzer als Wunde bestimmte Fläche betrug 66.4 % der Gesamtfläche des ersten und 18.2 % der Gesamtfläche des zweiten Bildes. Die mit CellProfiler erhaltenen Flächenanteile betragen 66.9 % für das erste und 42.6 % für das zweite Bild. Die Motivation Scratch-Assay-Bilder topologieerhaltend zu segmentieren bestand darin, Löcher im Zellrasen, welche bereits zu Beginn des Experimentes vorhanden waren oder während der Wundentwicklung entstehen, nicht der

Wundfläche zuzuordnen. Solche Löcher entstehen, wenn Zellen die Kontakte zu einigen ihrer Nachbarzellen verlieren. Die Fläche dieser Löcher sollte nicht der Wundfläche zugerechnet werden, da das Migrationsverhalten mittels Scratch-Assay-Experimenten durch die Entwicklung der Wundränder abgeschätzt wird. Kollektiv migrierende Zellen, für welche diese Art von Experiment verwendet wird, erhalten dabei weitestgehend zusammenhängende Wundränder. Die hohe Anzahl Zellen, welche sich im zweiten Beispielbild vom Verband gelöst haben, weist daraufhin, dass die verwendeten Zellen für ein Scratch-Assay-Experiment nicht gut geeignet waren.

Für TScratch waren insgesamt 24 Beispielbilder verfügbar, von denen 12 die Wundflächen zu Beginn von Scratch-Assay-Experimenten und die restlichen 12 Bilder die Wundflächen derselben Experimente nach 24 Stunden zeigten. Auch TScratch wurde mit Standardeinstellungen gestartet, da angenommen wurde, dass diese für die Beispielbilder optimiert worden waren. Für den ScratchAssayAnalyzer wurden verschiedene Werte für die Parameter σ und e auf einigen der Beispielbilder getestet. Die Kombination aus $\sigma = 2$ und $e = 27$ erzielte dabei optisch gute Ergebnisse, so dass diese für die Auswertung der Bilder gewählt wurde. Abbildung 27 zeigt die Segmentierungsergebnisse beider Programme für die beiden Zeitpunkte eines der Beispielexperimente.

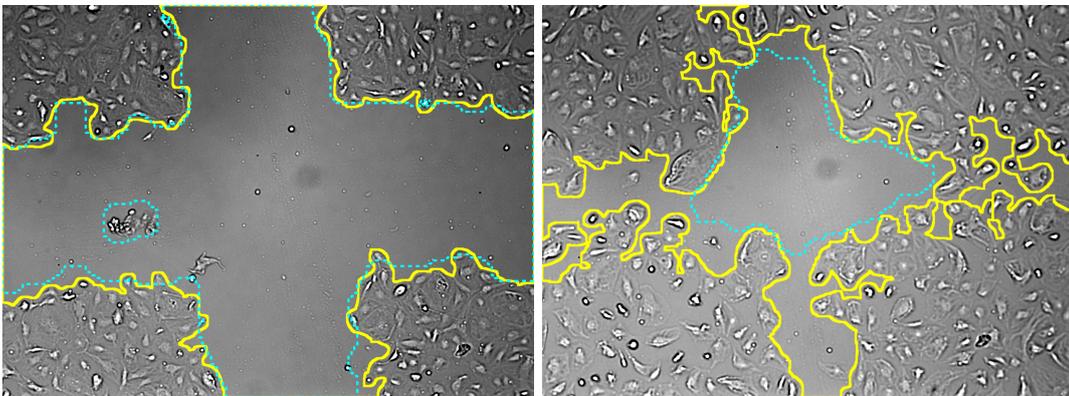


Abbildung 27: Qualitativer Vergleich des ScratchAssayAnalyzer und TScratch. Links: Wundfläche zu Beginn des Experimentes. Rechts: Wundfläche 24 Stunden nach Beginn des Experimentes. Die Konturen der mit dem ScratchAssayAnalyzer detektierten Wunden sind mit durchgezogenen Linien (gelb) gekennzeichnet, die Konturen der Wundflächen, welche mit TScratch bestimmt wurden mit unterbrochenen Linien (cyan).

Für die Bilder, welche jeweils den Beginn eines Experimentes darstellten, zeigten sich im Allgemeinen hohe Übereinstimmungen der Segmentierungsergebnisse beider Programme. Die Abweichungen der, von beiden Programmen als Wunde segmentierten, Gesamtbildfläche be-

trug im Mittel 2.2 %. Die maximale Abweichung betrug 7.2 %. Die Segmentierungen für die Endpunkte der Experimente zeigten jedoch größere Unterschiede. Die mittlere Abweichung der als Wunde segmentierten Gesamtbildfläche betrug für diese Bilder 8.8 % und die maximale Abweichung lag bei 23.8 %. Vereinzelt in die Wunde migrierte Zellen, sowie Zelltrümmer wurden von TScratch der Zellfläche zugeordnet. Im Allgemeinen schienen die Ergebnisse des ScratchAssayAnalyzer die Wundkonturen besser wiederzugeben. Neben der Auswahl von Werten für die Parameter σ und e , kann die Wahl der zu verwendenden SVM, sowie die Bestimmung der maximalen Iterationszahl für die Level-Set-Segmentierung die Segmentierungsergebnisse des ScratchAssayAnalyzer beeinflussen. TScratch und CellProfiler erfordern hingegen die Anpassung einer Vielzahl von Parametern, welche einen Einfluss auf die Segmentierungsergebnisse haben. Darüberhinaus erlaubt die CellProfiler-Pipeline auch die Auswahl verschiedener Algorithmen für die einzelnen Teilschritte der Verarbeitungskette. Eine hohe Parameterzahl macht die Anwendungen flexibler, jedoch wird ihre Anwendung dadurch komplizierter. Der ScratchAssayAnalyzer hingegen kann trotz weniger freier Parameter eine Vielzahl unterschiedlicher Scratch-/Gap-Closure-Assay-Bilder mit hoher Genauigkeit auswerten. Zudem bietet weder TScratch noch CellProfiler die Möglichkeit, vollständig geschlossene Wunden als solche zu identifizieren.

3.2 Anwendungen

In diesem Abschnitt werden beispielhafte Anwendungen des CellMigrationAnalyzer- und des ScratchAssayAnalyzer-Operators und deren Ergebnisse vorgestellt. Diese Beispiele zeigen unterschiedliche Anwendungsszenarien für die Operatoren auf.

3.2.1 Extrazelluläre Matrix

In dieser Beispielstudie sollte der Einfluss verschiedener Proteine, welche als Bestandteile der extrazellulären Matrix in tierischen Geweben vorkommen, auf die Geschwindigkeit und Adhäsion unterschiedlicher Zellen untersucht werden. Hierfür wurden Einzelzell-Tracking-Experimente mit Zellen der Linien ES-2, HT-144, PANC-1 und U-2OS durchgeführt und mit dem CellMigrationAnalyzer ausgewertet. ES-2-, PANC-1-, sowie U-2OS-Zellen waren stabil GFP-exprimierend, HT-144-Zellen exprimierten stabil das infrarot fluoreszierende Protein E2-Crimson. Die Zellen wurden auf Objektträgern vom Typ Ibidi 8 well μ -slide ausplattiert, welche entweder mit Collagen 1, Collagen 4 oder Fibronectin beschichtet wurden. Zellmigration wurde, wie in Abschnitt 2.2 beschrieben, über einen Zeitraum von 10 Stunden gefilmt.

Abbildung 28 zeigt die Verteilungen der ermittelten Durchschnittsgeschwindigkeiten der Zellen aufgeteilt nach Zelllinie und Beschichtungsprotein.

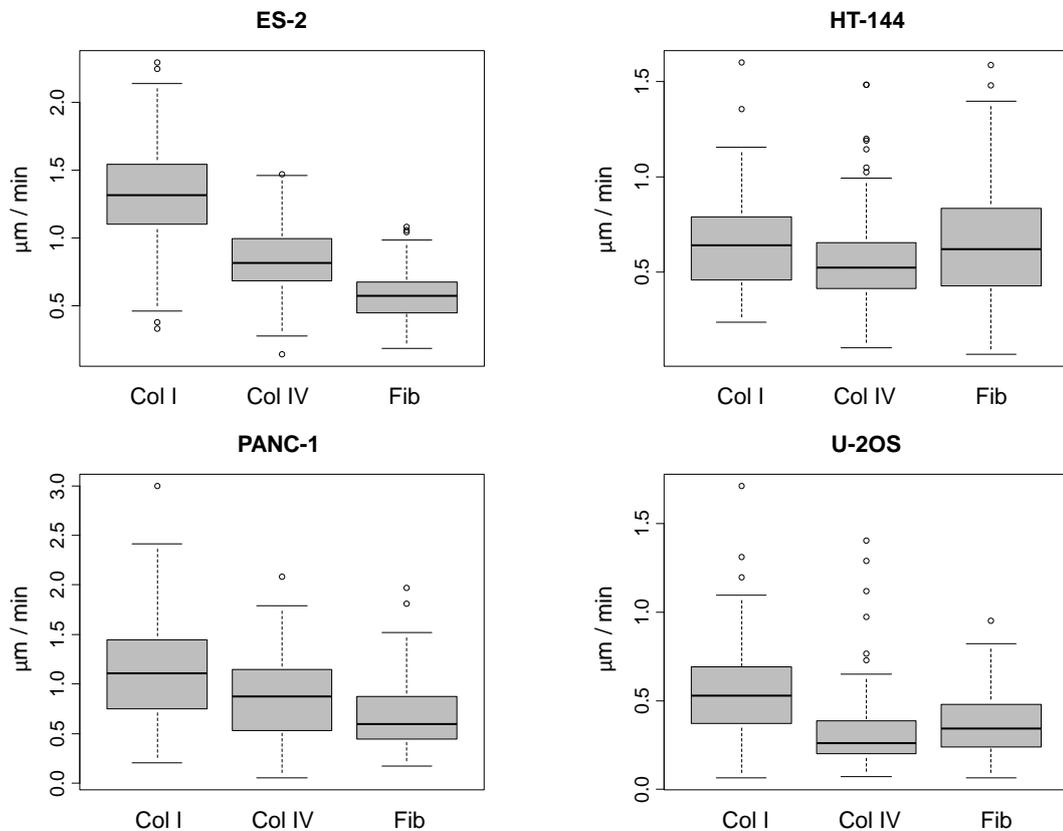


Abbildung 28: Durchschnittsgeschwindigkeiten von detektierten Zellen der Linien ES-2, HT-144, PANC-1 und U-2OS, ausgesät auf verschiedenen Beschichtungsproteinen: Collagen 1 (Col I), Collagen 4 (Col IV) und Fibronectin (Fib). Pro Zelllinie und Beschichtung wurden vier unabhängige Aufnahmebereiche betrachtet.

Zellen der Linien ES-2 und PANC-1 zeigten die höchsten Durchschnittsgeschwindigkeiten auf mit Collagen 1 beschichteten Untergründen und die niedrigsten Durchschnittsgeschwindigkeiten auf Fibronectin. Die Zellen der Linien HT-144 und U-2OS zeigten hingegen die niedrigste Durchschnittsgeschwindigkeit auf Untergründen beschichtet mit Collagen 4. Während U-2OS-Zellen jedoch eine signifikant höhere Durchschnittsgeschwindigkeit auf Collagen 1 im Vergleich zu Fibronectin zeigten, war bei HT-144-Zellen kein signifikanter Unterschied zwischen diesen beiden Beschichtungen zu beobachten.

Über die auf 2D-Aufnahmen detektierten Flächen adhärenter Zellen können indirekt Aussagen zu deren Adhäsionsvermögen getroffen werden. Je stärker die Adhäsion auf einem bestimmten Untergrund ist, desto mehr breiten sich die Zellen darauf aus und zeigen dementsprechend eine

größere Fläche. Dies wird beispielsweise in sogenannten Spreading Assays [55] genutzt, um das Adhäsionsvermögen von Zellen zu untersuchen. Abbildung 29 zeigt die Verteilung der detektierten Durchschnittsflächen der Zellen aufgeteilt nach Zelllinie und Beschichtungsprotein.

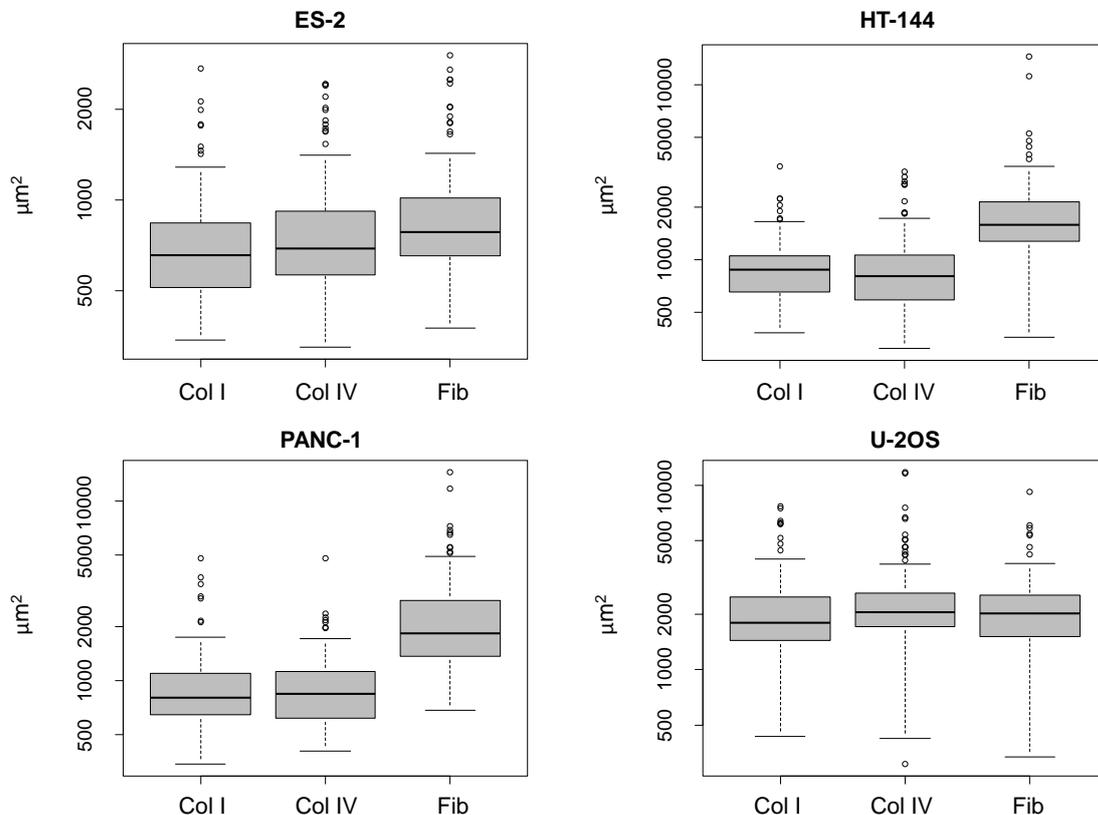


Abbildung 29: Durchschnittliche Zellflächen detektierter Zellen der Linien ES-2, HT-144, PANC-1 und U-2OS, ausgesät auf verschiedenen Beschichtungsproteinen: Collagen 1 (Col I), Collagen 4 (Col IV) und Fibronectin (Fib). Pro Zelllinie und Beschichtung wurden vier unabhängige Aufnahmebereiche betrachtet.

Zellen der Linien ES-2, HT-144 und PANC-1 zeigten, im Vergleich zu Collagen 1 und Collagen 4, signifikant größere Zellflächen, wenn sie auf Fibronectin ausgesät wurden. Die mittlere Durchschnittsfläche der U-2OS-Zellen war auf Collagen 4 am größten ($2521 \mu\text{m}^2$), jedoch waren die Unterschiede zu den Flächen der U-2OS-Zellen, welche auf Fibronectin ausgesät worden waren (Mittelwert: $2216 \mu\text{m}^2$) nicht signifikant. Allerdings zeigte diese Linie als einzige einen signifikanten Unterschied in den mittleren Flächen von Zellen ausgesät auf Collagen 1 zu Zellen ausgesät auf Collagen 4. Untersuchungen zu statistischer Signifikanz der Geschwindigkeits- und Flächenunterschiede in den Populationen wurden mittels zweiseitigen

Mann-Whitney-Tests durchgeführt. P-Werte ≥ 0.05 führten dabei jeweils zur Ablehnung der Nullhypothese, welche besagte, dass die Werte der zu vergleichenden Populationen aus der gleichen Grundgesamtheit stammten.

Abbildung 30 zeigt die mittleren Durchschnittsgeschwindigkeiten der Zelllinien auf den verschiedenen Beschichtungsproteinen, aufgetragen gegen die entsprechenden mittleren Durchschnittsflächen der Zellen dieser Populationen.

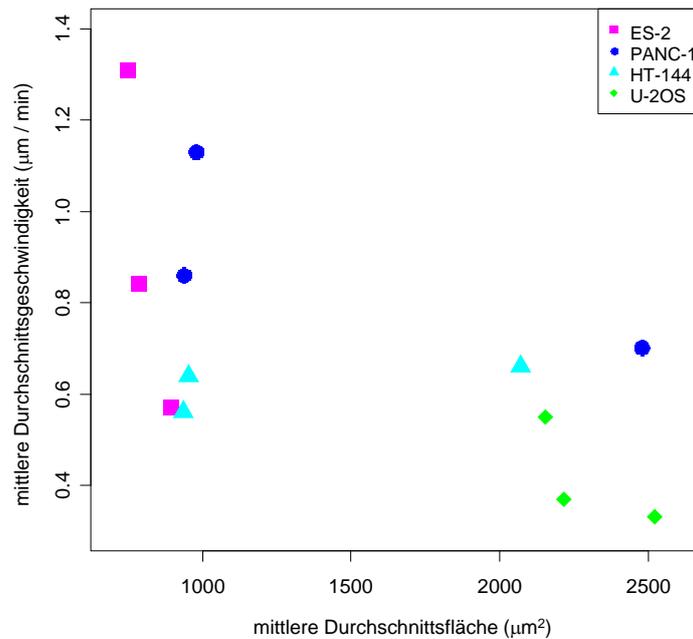


Abbildung 30: Zusammenhang zwischen Geschwindigkeit und Zellfläche auf verschiedenen Beschichtungsproteinen.

Korrelationstests ergaben für Zellen der Linien ES-2, PANC-1 und U-2OS eine negative Korrelation der mittleren Durchschnittsflächen und mittleren Durchschnittsgeschwindigkeiten der Zellen. Die Pearson-Korrelationskoeffizienten betragen -0.90, -0.77 und -0.75. Der Korrelationskoeffizient für die Zellen der Linie HT-144 war hingegen positiv (0.66). Allerdings wurden für diese Tests nur jeweils drei Werte miteinander verglichen und in keinem Test konnte eine statistisch signifikante positive oder negative Korrelation festgestellt werden. Ausgehend von der Annahme, dass größere Zellflächen auf eine stärkere Adhäsion schließen lassen, geben diese Ergebnisse dennoch einen Hinweis darauf, dass Beschichtungen, welche eine geringere Adhäsion verursachen, für Zellen der Linien ES-2, PANC-1 und U-2OS, eine höhere Migrationsgeschwindigkeit erlauben. Umgekehrt verhält es sich bei HT-144-Zellen, welche auf Beschichtungen, die eine stärkere Adhäsion zur Folge hatten, tendenziell schneller migrierten. Diese Versuche demonstrieren, dass die verwendete Beschichtung einen starken Einfluss auf die

Migration von Zellen in 2D-Einzelzell-Tracking-Experimenten ausüben kann. Dieser Einfluss ist sowohl vom verwendeten Zelltyp, als auch von der Beschichtung abhängig. Es empfiehlt sich daher, vor der Ausführung von Migrationsexperimenten mit neuen Zelllinien, in Vorversuchen zu prüfen, auf welchen Beschichtungen diese Zellen eine ausgeprägte Migration zeigen.

3.2.2 Chemotaxis

In dieser Studie sollte eine Methode zur Untersuchung der Chemotaxis von Tumorzellen etabliert werden. Hierfür wurden Chemotaxis μ -Slides der Firma Ibidi verwendet. Diese Objektträger besitzen jeweils drei unabhängige Abschnitte zur Untersuchung von Chemotaxis. Jeder dieser Abschnitte enthält einen 1 mm breiten Bereich in den Zellen eingesät werden können. Dieser Bereich dient auch der Betrachtung der Zellen mit Hilfe eines Mikroskops. Ober- und unterhalb dieses Betrachtungsbereiches befindet sich jeweils ein Reservoir, in welches chemische Lockstoffe gegeben werden können. Von einem Reservoir aus bildet sich nach Zugabe eines Lockstoffes ein Konzentrationsgradient dieses Stoffes aus, der sich über den Betrachtungsbereich erstreckt und laut Herstellerangaben mehr als 48 Stunden stabil bleibt. In Abbildung 31 ist der Aufbau eines Chemotaxis μ -Slide schematisch dargestellt.

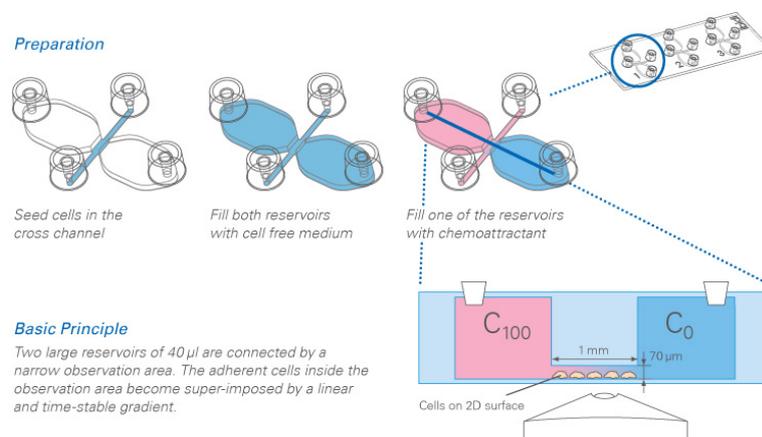


Abbildung 31: Schematischer Aufbau eines Chemotaxis μ -Slide der Firma Ibidi. Quelle: <http://ibidi.com>

Für das Experiment wurden stabil GFP-exprimierende Zellen der Linie PANC-1, welche zuvor 24 Stunden in serumfreiem Medium gehalten wurden, in ein mit Collagen 4 vorbeschichtetes Chemotaxis μ -Slide eingesät. Nachdem die Zellen sich abgesetzt hatten, wurde in die Reservoirs des ersten Abschnittes serumfreies Medium gegeben. In beide Reservoirs des zweiten Abschnittes wurde Medium welches 20 % fötales Rinderserum (FBS) enthielt, hinzugegeben. In das obere Reservoir des dritten Abschnittes wurde ebenfalls mit 20 % FBS versetztes Me-

dium, in das untere Reservoir jedoch serumfreies Medium gegeben. Demnach sollte sich im dritten Abschnitt ein definierter FBS-Konzentrationsgradient von oben nach unten ausgebildet haben. Im zweiten Abschnitt sollten sich zwei Gradienten ausgebildet haben, welche sich gegenseitig abschwächten und im ersten Abschnitt war kein Serum vorhanden. Anschließend wurden die Zellen der drei Abschnitte über einen Zeitraum von 10 Stunden gefilmt. Die Videos wurden anschließend mit dem CellMigrationAnalyzer ausgewertet.

Abbildung 32 zeigt die Verteilungen der mittleren Geschwindigkeiten und der Gerichtetheiten. Die Zellen aus den Abschnitten mit Serum wiesen geringfügig erhöhte mittlere Geschwindig-

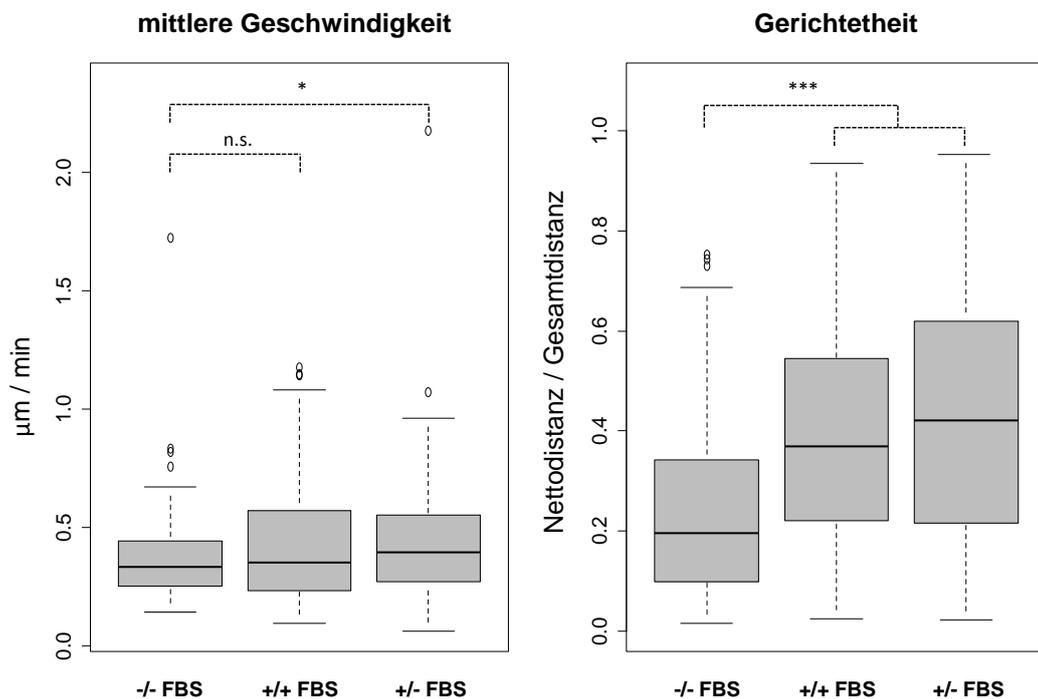


Abbildung 32: Mittlere Geschwindigkeiten und Gerichtetheiten von Zellen mit gerichteten Serumkonzentrationen. Stabil GFP-exprimierende PANC-1-Zellen in mit Kollagen 4 vorbeschichtetem Ibidi Chemotaxis μ -Slide. -/- FBS: kein FBS wurde zugesetzt; +/+ FBS: 20 % FBS wurde in beide Reservoirs gegeben; +/- FBS: 20 % FBS wurde nur in das obere Reservoir gegeben. Pro Bedingung wurden zwei unabhängige Aufnahmebereiche betrachtet. Gleichheit der Mittelwerte der Populationen wurde mittels zweiseitigem Mann-Whitney-Test geprüft; n.s.: nicht signifikant; * p -Wert < 0.05 ; ***: p -Wert < 0.001 .

keiten, jedoch deutlich höhere Gerichtetheiten auf. Der Serumentzug vor dem Experiment war vermutlich hauptsächlich für die deutlich geringeren Geschwindigkeiten der PANC-1-Zellen im Vergleich zu dem Beispiel aus Abschnitt 3.2.1 verantwortlich.

In Abbildung 33 sind Nettomigrationsrichtungen der Zellen dargestellt. Diese stellen die Richtung des Vektors aufgespannt zwischen der jeweils ersten und letzten Zentroidposition einer detektierten Zelle dar. Die Verteilung der Nettomigrationsrichtungen auf jeweils vier Qua-

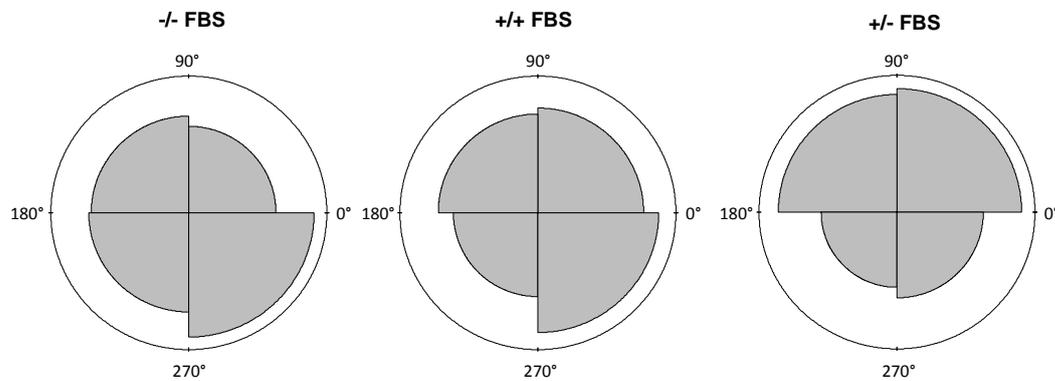


Abbildung 33: Verteilung der Nettomigrationsrichtungen von Zellen mit gerichteten Serumkonzentrationen. Gezeigt sind die Aufteilungen der Nettomigrationsrichtungen der detektierten PANC-1-Zellen auf vier Quadranten. Die Flächen der Kreisausschnitte geben den relativen Anteil der Anzahl an Zellen mit Nettorichtung in dem entsprechenden Quadranten wieder. -/- FBS: kein FBS wurde zugesetzt; +/+ FBS: 20 % FBS wurde in beide Reservoirs gegeben (90° und (270°)); +/- FBS: 20 % FBS wurde nur in das obere Reservoir gegeben (90°).

dranten zeigte, dass die Zellen, bei denen sich kein Serum im Medium befand, eine leichte Tendenz zur Migration nach unten (270°) zeigten. Diese Tendenz wurde von den Zellen, welche einem FBS-Gradienten von oben und unten gleichzeitig ausgesetzt waren, etwas reduziert. Zellen, welche lediglich einem FBS-Gradienten von oben ausgesetzt waren, migrierten größtenteils in Richtung der steigenden FBS-Konzentration (90°). In allen drei Bedingungen war zudem ein Trend der Migration nach rechts (0°) zu beobachten. Dieser deutet möglicherweise auf einen systematischen Verzerrungseffekt hin. Mögliche Ursachen hierfür könnten weitere Einflussfaktoren für die gerichtete Migration sein, wie zum Beispiel ein Temperaturgradient oder eine unebene Lage des Objektträgers auf dem Objektisch.

Dieses Beispiel zeigt, dass Ibidi Chemotaxis μ -Slides ein geeignetes Instrument zur Untersuchung der Chemotaxis von Tumorzellen darstellen. Mit Hilfe des CellMigrationAnalyzer konnte eine gerichtete Migration der verwendeten PANC-1-Zellen entlang eines FBS-Gradienten festgestellt werden.

3.2.3 Pharmakologische Inhibitoren

In dieser Studie sollte der Einfluss des Stoffes Silibinin auf die Migration verschiedener, aus Ovariakarzinomen abgeleiteter, Zelllinien untersucht werden. Das Flavonoid Silibinin ist der bioaktive Hauptbestandteil von Silymarin, einem Gemisch, welches aus den Samen der Milchdistel (*Silybum marianum*) gewonnen und bereits seit Jahrhunderten in der traditionellen Medizin gegen verschiedene Erkrankungen eingesetzt wird. Aufgrund seiner hepatoprotektiven Eigenschaften wird es beispielsweise bei Leberschäden infolge viraler Infektion, Alkoholmissbrauch oder Vergiftungen eingesetzt [20]. Es häufen sich jedoch auch Berichte über positive Auswirkungen von Silibinin auf Tumorerkrankungen, zum Beispiel [20, 16, 10]. Insbesondere wurden inhibitorische Effekte auf das Migrations- und Invasionsvermögen von Tumorzellen beschrieben [19, 120, 123, 87].

Zum Vergleich des Einflusses von Silibinin auf die Zellmigration wurde der für seine inhibitorische Wirkung auf Zellmigration bekannte Rac1-Inhibitor NSC23766 (Merck Millipore, Billerica, Massachusetts, USA) verwendet. Dieser Stoff verhindert, dass das GTP-bindende Protein Rac1 in seine aktive Form (Rac-GTP) überführt wird [23]. Rac1 wird unter anderem für die Ausbildung von Lamellipodien benötigt [94]. Festes Silibinin (Sigma Aldrich, St. Louis, Missouri, USA) wurde aufgrund seiner geringen Wasserlöslichkeit in Dimethylsulfoxid (DMSO) gelöst. Da ein Einfluss dieses Lösungsmittel auf die Migration der Zellen nicht ausgeschlossen werden konnte, wurde DMSO auch in das Medium der anderen Bedingungen gegeben, so dass die DMSO-Gesamtkonzentration im Medium in allen Bedingungen konstant bei 0.1 % lag.

Die Zellen wurden 24 Stunden vor Beginn der Mikroskopieaufnahmen mit jeweils 120 μM Silibinin oder 50 μM (Einzelzell-Tracking) beziehungsweise 60 μM (Gap-Closure-Assay) Rac1-Inhibitor inkubiert. Für Einzelzellmigrationsexperimente wurden Zellen der Linien ES-2, SK-OV-3 und NIH:OVCAR-3 verwendet, welche stabil das grün fluoreszierende Protein ZS-Green1 exprimierten. Als Beschichtung wurde Collagen 1 verwendet. Gap-Closure-Experimente wurden mit Zellen der Linien SK-OV-3 und NIH:OVCAR-3 durchgeführt. ES-2-Zellen bildeten keine konfluenten Zellschichten aus und wurden deshalb nicht für diese Art von Experiment verwendet. Abbildung 34 zeigt die Wirkung von Silibinin, sowie des Rac1-Inhibitors auf die mittlere Geschwindigkeit der Einzelzellmigration. Während der Rac1-Inhibitor in allen drei untersuchten Zelllinien die Geschwindigkeit deutlich reduzierte, war nur für Zellen der Linie NIH:OVCAR-3 auch eine signifikante Reduktion der Geschwindigkeit durch Silibinin erkennbar.

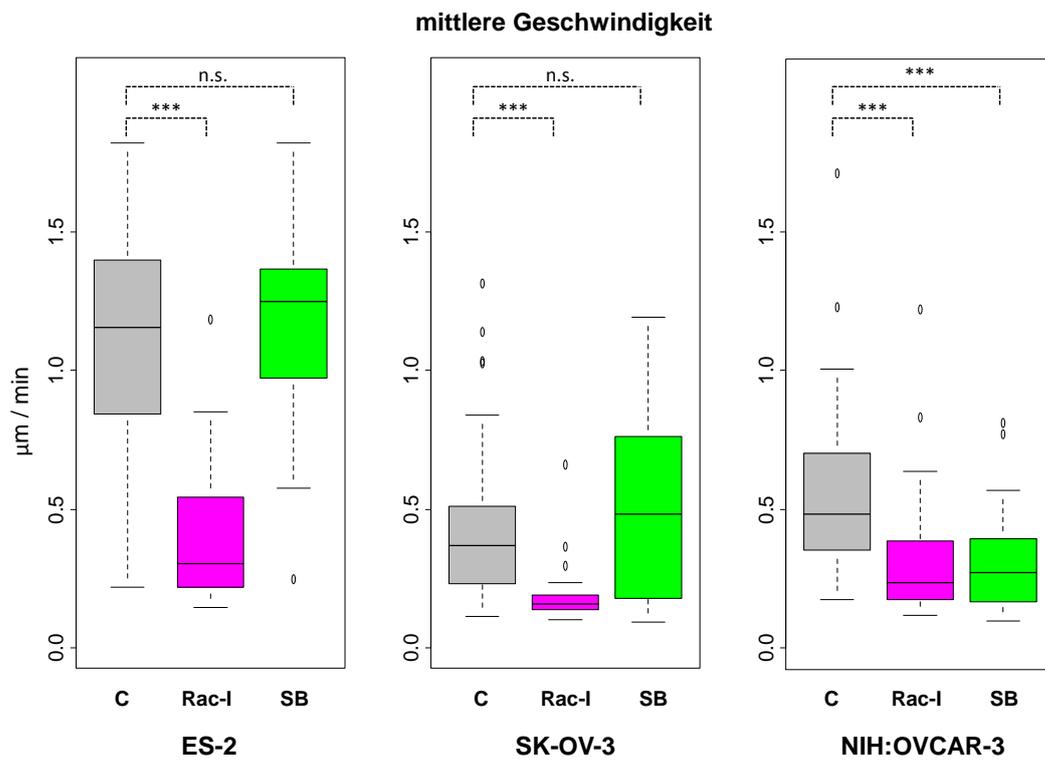


Abbildung 34: Einfluss von Silibinin auf die Geschwindigkeit von ovarialkarzinomabgeleiteten Zellen. Gezeigt ist der Einfluss von Silibinin (SB, 120 µM) und des Rac1-Inhibitors NSC2376 (Rac-I, 50 µM) auf die mittlere Migrationsgeschwindigkeit von stabil ZS-Green1-exprimierenden Zellen der Linien ES-2, SK-OV-3 und NIH:OVCAR-3 im Vergleich zu Kontrollzellen (C). Pro Zelllinie und Inhibitor wurden zwei unabhängige Aufnahmebereiche betrachtet. Statistische Signifikanz wurde mittels zweiseitigen Mann-Whitney-Tests geprüft; n.s.: nicht signifikant; ***: p-Wert < 0.001

Gap-Closure-Experimente mit Zellen der Linien SK-OV-3 und NIH:OVCAR-3 zeigten eine Verzögerung des Wundschlusses in beiden Zelllinien nach Behandlung mit dem Rac1-Inhibitor, sowie mit Silibinin. Während der Wundschluss in SK-OV-3-Zellen durch den Rac1-Inhibitor stärker verzögert wurde als durch Silibinin, war für NIH:OVCAR-3-Zellen das Gegenteil zu beobachten (s. Abbildung 35).

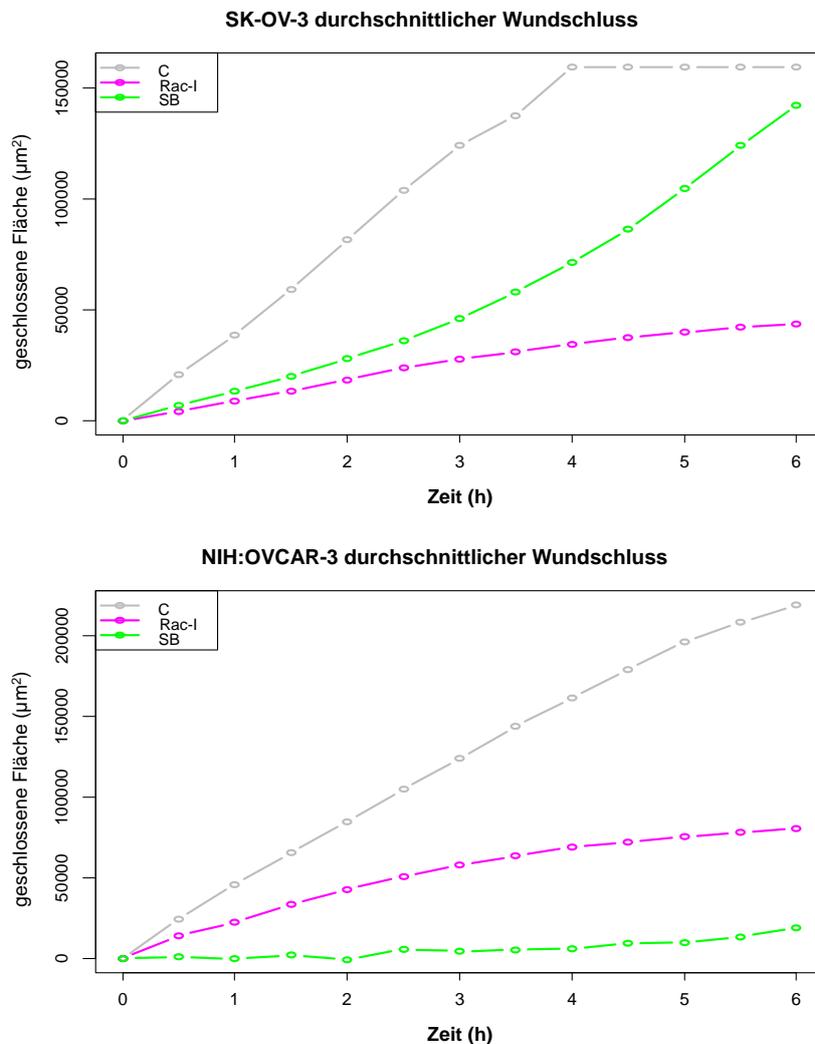


Abbildung 35: Wirkung von Silibinin auf die Entwicklung des Wundschlusses von stabil ZS-Green1-exprimierenden SK-OV-3 und NIH:OVCAR-3-Zellen. Die betrachteten Wunden der SK-OV-3-Kontrollzellen waren nach vier Stunden bereits vollständig geschlossen, weshalb sich der Wert für die geschlossene Fläche dieser Population nach vier Stunden nicht mehr veränderte. Pro Zelllinie und Inhibitor wurden zwei unabhängige Aufnahmebereiche betrachtet.

Die beschriebenen Einzelzell-Tracking- und Gap-Closure-Experimente mit drei verschiedenen ovarialkarzinomabgeleiteten Zelllinien zeigten unterschiedliche Auswirkungen auf die Zellmi-

gration nach Behandlung mit dem pflanzlichen Wirkstoff Silibinin. Um ein besseres Verständnis für die Wirkung von Silibinin auf das Migrationsvermögen tumorabgeleiteter Zellen zu erlangen, müssen weitere Experimente durchgeführt werden. So könnten beispielsweise Migrationsexperimente mit Vielzahl unterschiedlicher Zelllinien durchgeführt werden. Next Generation Sequencing-Analysen, welche beispielsweise Informationen zu Genexpression, genomischer Variation oder DNA-Methylierung liefern, könnten zudem helfen, Resultate der Migrationsexperimente zu interpretieren. So könnte zum Beispiel untersucht werden, ob Zelllinien, welche ähnliche Effekte durch Silibinin-Behandlung zeigen, auch Gemeinsamkeiten in der Genexpression zeigen und ob diese sie von Zelllinien unterscheiden, in denen sich andere/keine Effekte zeigen. Durch groß angelegte Sequenzierprojekte, zum Beispiel „The Cancer Cell Line Encyclopedia“ (CCLE) [3], sind verschiedenartige Next Generation Sequencing-Daten öffentlich und kostenlos zugänglich. Dadurch können verhältnismäßig günstige und schnelle Analysen mit diesen Daten durchgeführt werden.

3.3 Einfluss von IGF2BP1 auf Zellmigration

Für das RNA-bindende onkofötal-exprimierte Protein IGF2BP1 wurden bereits Einflüsse auf die Migration tumorabgeleiteter Zellen beschrieben (s. Abschnitt 1.2). Diese Beobachtungen basierten größtenteils auf manueller Auswertung von Migrationsexperimenten und zeigten zum Teil gegensätzliche Ergebnisse. Für die vorliegende Arbeit wurden Einzelzell-Tracking-, sowie Scratch-Assay-Experimente mit Zellen durchgeführt, in welchen die endogene IGF2BP1-Proteinmenge durch siRNA-vermittelte, transiente Transfektionen reduziert worden war.

Dieser Abschnitt stellt die Resultate vor, welche mit Hilfe des CellMigrationAnalyzer und des ScratchAssayAnalyzer aus diesen Experimenten gewonnen worden. Die Ergebnisse werden mit manuell erhobenen Daten aus vorangegangenen Experimenten verglichen.

Einfluss von IGF2BP1 auf Einzelzellmigration In [109] wurde durch manuelle Auswertung der Einzelzellmigration von Zellen der osteosarkomabgeleiteten Zelllinie U-2OS gezeigt, dass die Überexpression eines GFP-fusionierten IGF2BP1-Orthologen aus dem Huhn (GFP-chZBP1) zu einer Steigerung der Geschwindigkeit, sowie zu einer höheren Gerichtetheit der Bewegung führt. Der siRNA-vermittelte Knockdown von chZBP1/IGF2BP1 reduzierte die Geschwindigkeit der GFP-chZBP1-überexprimierenden Zellen signifikant. Eine signifikante Veränderung der Gerichtetheit dieser Zellen konnte jedoch nicht festgestellt werden.

Um zu untersuchen, ob der Knockdown von IGF2BP1 in stabil GFP-exprimierenden U-2OS-Zellen ebenfalls zu einer Reduktion der Geschwindigkeit führt, wurden diese Zellen transient

mit einer von zwei verschiedenen, gegen IGF2BP1 gerichteten, siRNAs transfiziert (si1-1 beziehungsweise si1-2, s. Abschnitt 2.1). Kontrollzellen wurden mit einer siRNA transfiziert, deren Sequenz zu keiner menschlichen mRNA-Sequenz komplementär war (siC). Aus dem Western-Blot in Abbildung 36 ist ersichtlich, dass die Transfektionen zu einer effizienten Reduktion der IGF2BP1-Proteinmenge führten.

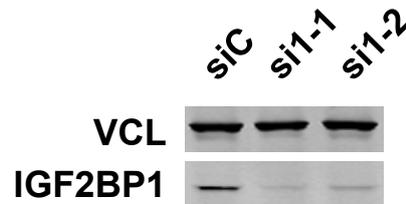


Abbildung 36: Western Blot, welcher die Proteinmenge von IGF2BP1 in U-2OS-Zellen nach Transfektion mit einer Kontroll-siRNA (siC), sowie zwei verschiedenen gegen IGF2BP1-gerichteten siRNAs (si1-1 und si1-2), zeigt. Vinculin (VCL) diente als Ladekontrolle.

Abbildung 37 zeigt die ermittelten Verteilungen für Geschwindigkeiten, Gerichtetheiten und Flächen der U-2OS-Zellen unter IGF2BP1-Knockdown. Die Auswertung der jeweils 10 Stunden umfassenden Filme ergab, dass die Geschwindigkeit der Knockdown-Zellen im Vergleich zu den Kontrollzellen signifikant reduziert war. Dabei zeigten Zellen, welche mit der siRNA si1-2 transfiziert worden waren, eine deutlichere Geschwindigkeitsreduktion. Diese Zellen zeigten zudem eine signifikante Reduktion der Gerichtetheit ihrer Bewegungen. Die Population der Zellen, welche mit der si1-1 transfiziert worden waren, zeigte hingegen keine signifikante Reduktion und eine vergleichsweise hohe Streuung der Werte für die Gerichtetheit. Diese Unterschiede könnten auf Kreuzreaktionen der siRNAs mit verschiedenen Transkripten hinweisen. Die mittleren Zellflächen der Knockdown- und Kontrollzellen zeigten keine signifikanten Unterschiede zueinander.

Um den Einfluss möglicher Kreuzreaktionen einzelner siRNAs auf andere mRNA-Transkripte und damit unspezifische Wirkungen zu reduzieren, wurde eine Kombination aus 12 verschiedenen, gegen IGF2BP1 gerichteten, siRNAs (si1-P) verwendet (s. Abschnitt 2.1) und deren Wirkung auf das Migrationsverhalten von Zellen untersucht. Dadurch sollte die Wirkung potentieller unspezifischer Kreuzreaktionen reduziert und die Wirkung der spezifischen Unterdrückung der IGF2BP1-Expression stärker hervorgehoben werden. Für diese Versuche wurden Zellen der ovarialkarzinomabgeleiteten Linie ES-2 verwendet. Der Western-Blot in Abbildung 38 zeigt, dass die Transfektion zu einer effizienten Reduktion der IGF2BP1-Proteinmenge führte.

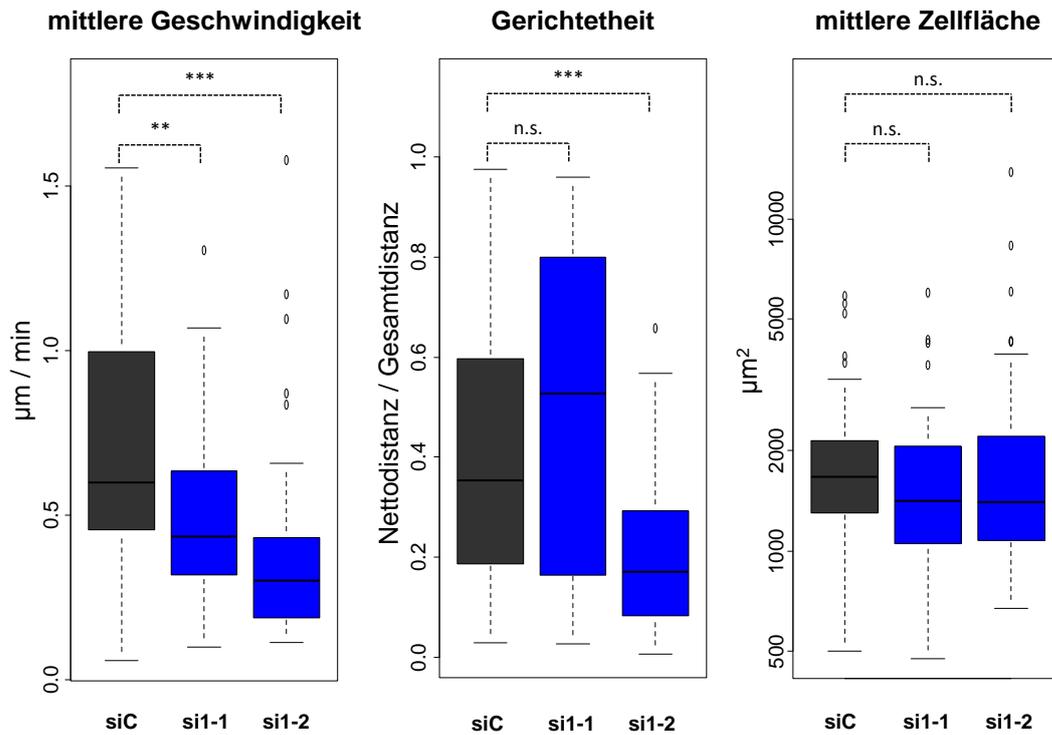


Abbildung 37: Mittlere Geschwindigkeiten, Gerichtetheiten und mittlere Zellflächen von U-2OS-Zellen unter siRNA-vermitteltem IGF2BP1-Knockdown. Die stabil GFP-exprimierenden Zellen wurden entweder mit einer von zwei unterschiedlichen, gegen IGF2BP1 gerichteten, siRNAs (si1-1 und si1-2) oder mit einer Kontroll-siRNA (siC) transfiziert. Pro Zellpopulation wurden zwei unabhängige Aufnahmebereiche betrachtet. Unterschiede in den Verteilungen wurden mittels zweiseitigen Mann-Whitney-Tests auf statistische Signifikanz geprüft; n.s.: nicht signifikant; **: p-Wert < 0.01; ***: p-Wert < 0.001

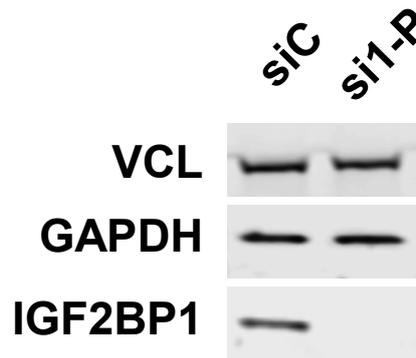


Abbildung 38: Western Blot, welcher die Proteinmenge von IGF2BP1 in ES-2-Zellen nach Transfektion mit einer Kontroll-siRNA (siC), sowie einer Kombination aus 12 verschiedenen, gegen IGF2BP1-gerichteten siRNAs (si1-P), zeigt. Vinculin (VCL) und Glycerinaldehyd-3-phosphat-Dehydrogenase (GAPDH) dienten als Ladekontrollen.

Die Zellen wurden durch Zugabe des Farbstoffes CellMask Deep Red (ThermoFisher Scientific, Waltham, Massachusetts, USA) fluoreszierend markiert. In Abbildung 39 sind die Verteilungen der mittleren Geschwindigkeiten, der Gerichtetheiten und der mittleren Zellflächen als Boxplots dargestellt. Der IGF2BP1-Knockdown führte zu einer mäßigen, jedoch statistisch signifikanten Reduktion der mittleren Geschwindigkeiten von durchschnittlich $0.93 \mu\text{m}/\text{min}$ auf $0.79 \mu\text{m}/\text{min}$. In der Gerichtetheit der Zellen konnte kein signifikanter Unterschied zur Kontrollpopulation (siC) festgestellt werden. Die Zellflächen waren hingegen signifikant verändert. Die ermittelten mittleren Flächen der Kontrollzellen betragen durchschnittlich $1267 \mu\text{m}^2$, die der Knockdown-Zellen $1627 \mu\text{m}^2$. Abbildung 40 zeigt Ausschnitte von Beispielaufnahmen und die korrespondierenden Segmentierungsergebnisse von Knockdown- und Kontrollzellen. Eine mögliche Erklärung für die Flächenzunahme ist eine gesteigerte Adhäsion der Knockdown-Zellen an die Kollagenbeschichtung. Diese könnte, zumindest teilweise, für die Reduktion der Geschwindigkeit verantwortlich sein.

Zusammenfassend konnte gezeigt werden, dass die Reduktion der IGF2BP1-Expression in U-2OS, sowie ES-2-Zellen zu einer Verringerung der Migrationsgeschwindigkeit dieser Zellen führt. Ob der Knockdown die Gerichtetheit der Zellen beeinflusst hat konnte jedoch nicht eindeutig festgestellt werden.

Einfluss von IGF2BP1 auf kollektive Zellmigration Um die Wirkung des IGF2BP1-Knockdowns auf kollektive Zellmigration zu untersuchen, wurden Scratch-Assay-Experimente

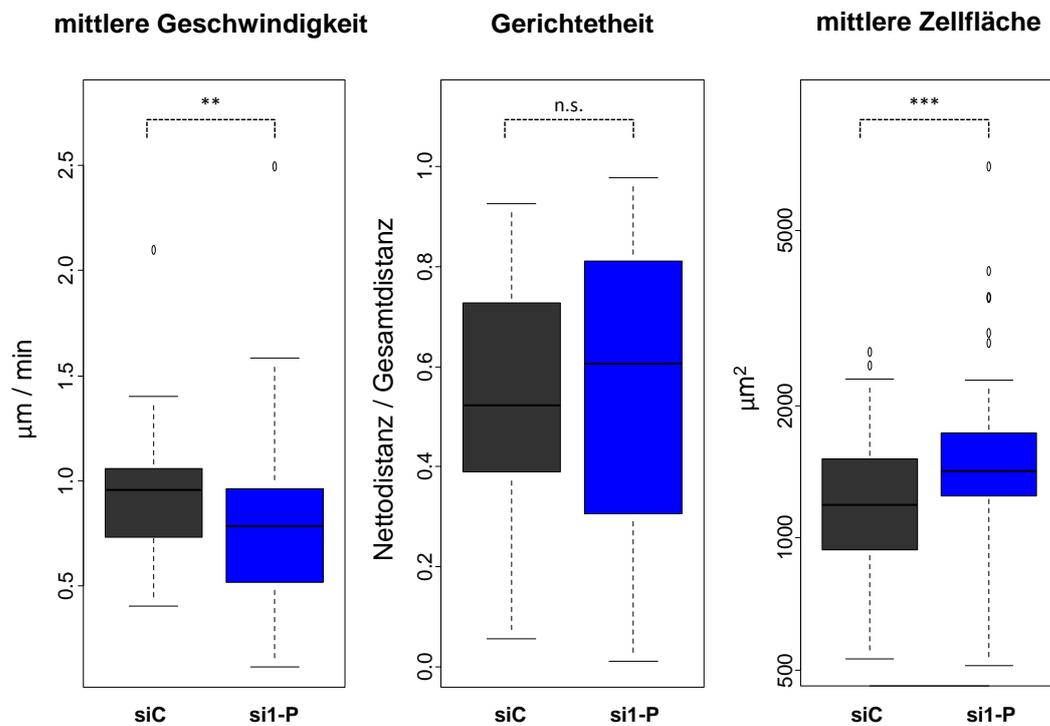


Abbildung 39: Mittlere Geschwindigkeiten, Gerichtetheiten und mittlere Zellflächen von ES-2-Zellen unter siRNA-Pool-vermitteltem IGF2BP1-Knockdown. Zellen wurden mit einer Mischung aus 12 verschiedenen gegen IGF2BP1 gerichteten siRNAs (si1-P) transfiziert und mit Zellen, welche mit einer Kontroll-siRNA (siC) transfiziert wurden, verglichen. Die Zellen wurden mit dem Farbstoff CellMask Deep Red angefärbt. Für jede der beiden Zellpopulationen wurden sechs unabhängige Aufnahmebereiche betrachtet. Unterschiede in den Verteilungen wurden mittels zweiseitigen Mann-Whitney-Tests auf statistische Signifikanz geprüft; n.s.: nicht signifikant; **: p-Wert < 0.01; ***: p-Wert < 0.001

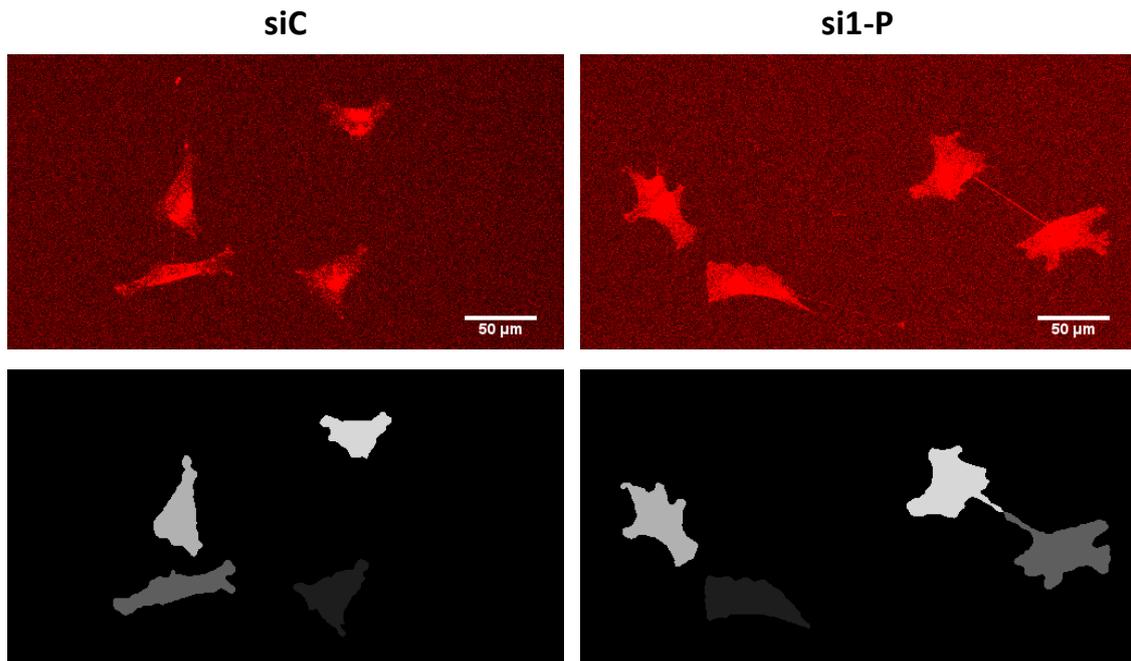


Abbildung 40: *Aufnahmen von Kontrollzellen (siC) und IGF2BP1-Knockdownzellen (si1-P) der Linie ES-2 und korrespondierende Segmentierungsergebnisse. Die obere Reihe zeigt Zellen gefärbt mit dem Farbstoff CellMask Deep Red. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht. Die untere Reihe zeigt die dazugehörigen Segmentierungsergebnisse des CellMigrationAnalyzer.*

durchgeführt und mit dem ScratchAssayAnalyzer ausgewertet. Hierfür wurden U-2OS-Zellen, welche stabil GFP exprimierten, transient mit verschiedenen, gegen die IGF2BP1-mRNA gerichteten siRNAs transfiziert. Die Zellen wurden entweder mit je einer (siI1-1, beziehungsweise siI1-2) oder mit einer Mischung aus 12 verschiedenen siRNAs (si1-P), wie in Abschnitt 2.1 beschrieben, transfiziert. Als Kontrolle wurden Zellen mit einer siRNA transfiziert, deren Sequenz zu keiner menschlichen mRNA-Sequenz komplementär war (siC). Zellen transfiziert mit einem Vektor ohne siRNA (Mock) und untransfizierte Zellen (ut) dienten als weitere Kontrollen. In Abbildung 41 ist die Wirkung der Transfektionen auf die Expression des IGF2BP1-Proteins mit Hilfe eines Western-Blots dargestellt.

Nach Transfektion wurden Scratch-Assays, wie in 2.1 beschrieben, durchgeführt. Die Aufnahmen erfolgten in Abständen von 30 Minuten über einen Gesamtzeitraum von 16 Stunden.

Die Wunden der Kontrollzellen waren teilweise bereits nach sieben Stunden geschlossen, weshalb Auswertungen nur auf die Zeitpunkte bis einschließlich sechs Stunden nach Erzeugung der Wunde erfolgten. Durch die Beschränkung der Analysen auf diesen kurzen Zeitraum konnte



Abbildung 41: Western Blot, welcher die Proteinmenge von IGF2BP1 in U-2OS-Zellen in untransfizierten Zellen (*ut*), sowie nach Transfektion mit einem Vektor ohne *siRNA* (*Mock*), einer Kontroll-*siRNA* (*siC*), zwei verschiedenen, gegen IGF2BP1 gerichteten, *siRNAs* (*si1-1* beziehungsweise *si1-2*) und einer Kombination aus 12 verschiedenen, gegen IGF2BP1-gerichteten *siRNAs* (*si1-P*), zeigt. Vinculin (*VCL*) diente als Ladekontrolle.

zudem der Einfluss von Zellproliferation auf den Wundschluss reduziert werden. Abbildung 42 zeigt beispielhaft die Wunden einer Kontrollbedingung (*siC*) und einer IGF2BP1-Knockdown-Bedingung (*si1-2*) zu Beginn des Experimentes und sechs Stunden danach.

Abbildung 43 zeigt, dass sich die Wunden der Knockdown-Bedingungen deutlich langsamer schlossen, als die der Kontrollbedingungen. Der Mix aus *siRNAs* reduzierte hierbei den Wundschluss im Vergleich zu den Kontrollen am geringsten aber dennoch signifikant. Dies steht in Einklang zu manuell ausgewerteten Ergebnissen aus [109]. Dort führte der IGF2BP1-Knockdown in U-2OS-Zellen mit Hilfe der *siRNA* *si1-1* ebenfalls zu einer deutlichen Reduktion des Wundschlusses.

Zusammenfassend konnte mit Hilfe des ScratchAssayAnalyzer gezeigt werden, dass IGF2BP1-Knockdown in Scratch-Assay-Experimenten mit U-2OS-Zellen zu einem verzögerten Wundschluss führt.

3.4 Einfluss von IGF2BP1 auf die Genexpression

Um die beobachteten Auswirkungen der Reduktion der endogenen IGF2BP1-Proteinexpression auf die Migration tumorabgeleiteter Zellen mit veränderter Genexpression in Verbindung zu bringen, wurden RNA-Seq-Daten von IGF2BP1-Knockdown-Experimenten analysiert. Dafür wurde RNA aus ES-2-Zellen isoliert, welche mit der in Abschnitt 3.3 beschriebenen Kombination aus 12 verschiedenen gegen IGF2BP1 gerichteten *siRNAs* transfiziert worden waren. Als Kontrolle dienten ES-2-Zellen, welche mit einer *siRNA* transfiziert worden waren, deren Sequenz zu keiner menschlichen mRNA-Sequenz komplementär war.

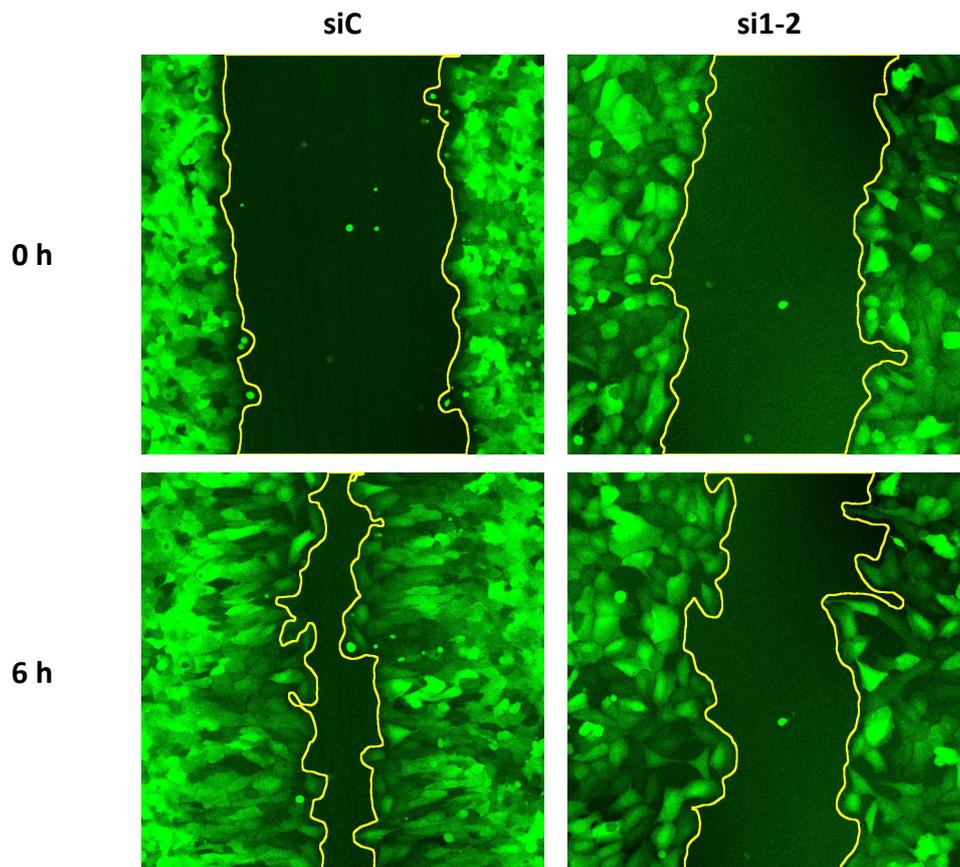


Abbildung 42: Exemplarische Bilder GFP-exprimierender U-2OS Zellen, welche mit einer gegen IGF2BP1-gerichteten siRNA (si1-2) beziehungsweise mit einer Kontroll-siRNA (siC) transfiziert wurden. Obere Reihe: Wunden direkt nach ihrer Erzeugung; untere Reihe: Wunden sechs Stunden später. Die Konturen der mit dem ScratchAssayAnalyzer detektierten Wunden sind mit durchgezogenen Linien (gelb) gekennzeichnet. Zur besseren Darstellung der Bilder wurde der Kontrast nachträglich erhöht.

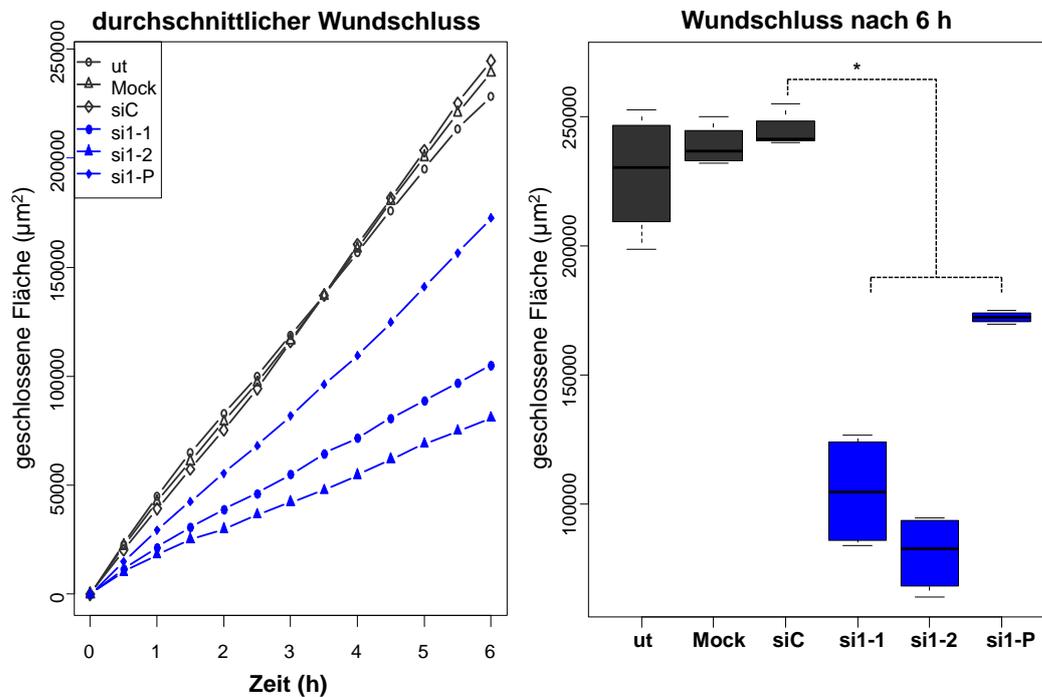


Abbildung 43: Wundschluss durch *IGF2BP1*-Knockdown in U-2OS-Zellen. GFP-exprimierende Zellen wurden transient mit je einer (*siI1-1*, *siI1-2*) oder einem Mix aus 12 verschiedenen (*siI1-P*) siRNAs, welche gegen *IGF2BP1*-mRNAs gerichtet sind, sowie mit einer Kontroll-siRNA (*siC*) und mit einem Vektor ohne siRNA (*Mock*) transfiziert. Untransfizierte (*ut*) Zellen dienten als zusätzliche Kontrolle. Pro Bedingung wurden vier unabhängige Aufnahmebereiche betrachtet, davon jeweils zwei im gleichen Well. Links: Zeitlich aufgelöst dargestellte Zunahme der durchschnittlich geschlossenen Wundflächen. Rechts: Verteilungen der nach sechs Stunden geschlossenen Wundflächen. Unterschiede in der geschlossenen Flächen wurden mittels zweiseitigen Mann-Whitney-Tests auf statistische Signifikanz geprüft; *: p -Wert < 0.05 .

Wie in Abschnitt 3.3 beschrieben, führte dieser Knockdown zu einer Reduktion der Geschwindigkeit der Zellen, sowie zu größeren detektierten Zellflächen.

Gesamt-RNA aus je drei Proben der Knockdown- und Kontrollzellen wurde isoliert, um daraus Bibliotheken für Total-RNA-Seq und microRNA-Seq zu erzeugen. Die Erstellung der Bibliotheken, sowie die Bestimmung der differentiellen Expression zwischen Knockdown- und Kontrollzellen erfolgte wie in Abschnitt 2.6.2 beschrieben.

Total-RNA-Seq Die durchschnittliche Anzahl sequenzierter Reads je Probe betrug für das im Paired-End-Verfahren sequenzierte Total-RNA-Seq-Experiment ungefähr $18.4 \cdot 10^6$. Nach Adapterentfernung und Qualitätskürzungen wurden je Probe im Durchschnitt 8.5 % der Basen entfernt. Infolgedessen wurden je Probe durchschnittlich 5.4 % der Reads entfernt, da deren Längen die Mindestlänge von 20 Basen unterschritten. Von den übrigen Reads konnten im Mittel je Probe 82.0 % auf das menschliche Genom gemappt werden. Obwohl in einem vorangegangenen Kontrollschritt mittels PCR keine Mykoplasma kontamination festgestellt wurde, konnten durchschnittlich 28.4 % der Reads, welche nicht auf das menschliche Genom gemappt wurden, genomischen Mykoplasma sequenzen zugeordnet werden. Der höchste Anteil war für Kontrollprobe 1 zu beobachten. 46.9 % der Reads dieser Probe, welche nicht dem menschlichen Genom zugeordnet werden konnten, wurden auf Mykoplasma-Sequenzen gemappt. Dies entsprach 8.9 % der Gesamtanzahl der Reads dieser Probe. Ausgehend von einer unteren Grenze von 0.01 % der Reads einer Probe, wie in [88] vorgeschlagen, deutet dies darauf hin, dass Zellen aus allen Proben mit Mykoplasmen infiziert waren. Dies könnte zu veränderter Expression von Genen geführt haben, wodurch die Erkennung von durch IGF2BP1-Knockdown verursachten Veränderungen der Genexpression möglicherweise beeinträchtigt wurde. Da jedoch alle Proben in einem ähnlichen Ausmaß kontaminiert zu sein schienen, sollten die beobachteten Veränderungen der Genexpression zwischen Kontroll- und Knockdown-Zellen überwiegend auf den IGF2BP1-Knockdown zurückzuführen sein.

Nach Zusammenfassung der auf das menschliche Genom gemappten Read-Paare zu Fragmenten, konnten im Mittel für 69.5 % der Fragmente einer Probe Genannotationen gefunden werden. Die übrigen Fragmente waren dementsprechend auf nicht-annotierte Bereiche des menschlichen Genoms gemappt worden. Die Zahlen der Fragmente, welche auf annotierte Genombereiche gemappt waren, wurden zur Bestimmung der differentiellen Genexpression verwendet. Mit einer False Discovery Rate (FDR) von weniger als 0.05 als Kriterium für differentielle Expression, konnten insgesamt 466 Gene als differentiell exprimiert bestimmt werden. Abbildung 44 stellt die Expression dieser Gene in den Kontroll- und IGF2BP1-Knockdown-

Proben als Heatmap dar. Eine Übersicht über die Aufteilung der verschiedenen differentiell exprimierten RNA-Spezies ist in Abbildung 45 dargestellt.

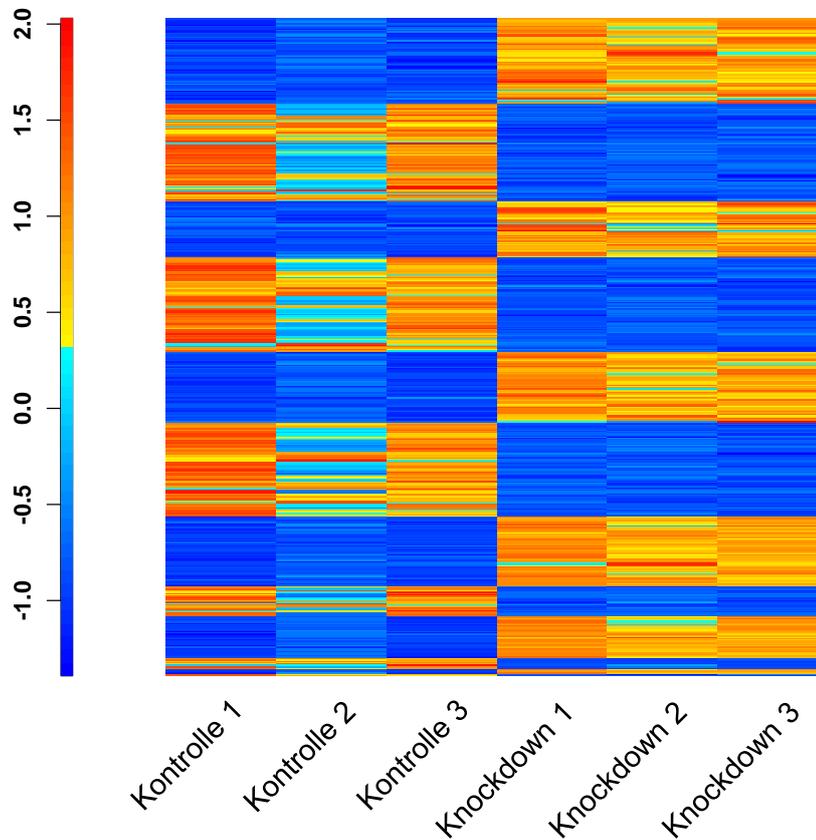


Abbildung 44: Heatmap der Expressionen der 466 unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene. Für die Darstellung wurden die FPKM-Werte pro Gen so skaliert, dass die Mittelwerte in allen Proben 0 und die Standardabweichungen 1 ergaben.

Ungefähr die Hälfte der differentiell exprimierten Gene war hoch-, die andere Hälfte herunterreguliert. Mit 418 Genen bildeten die proteinkodierenden Gene die größte Gruppe der differentiell exprimierten Gene. Von den restlichen differentiell exprimierten Genen waren 19 als Pseudogene, 25 als lange nichtkodierende RNAs (lncRNA) und vier als kurze nichtkodierende RNAs annotiert. Die lncRNAs konnten weiter unterteilt werden in 13 lincRNAs (long interspersed non-coding RNA), sechs antisense RNAs, eine sense overlapping RNA, sowie fünf nicht näher bestimmte Transkripte ohne offenen Leserahmen (processed transcript). Zu den differentiell exprimierten kurzen nichtkodierenden RNAs gehörten eine herunterregulierte snRNA (small nuclear RNA), sowie drei ebenfalls herunterregulierte und nach Ensembl-Annotation funktionell nicht näher bestimmte RNAs (miscRNA). Die Expression

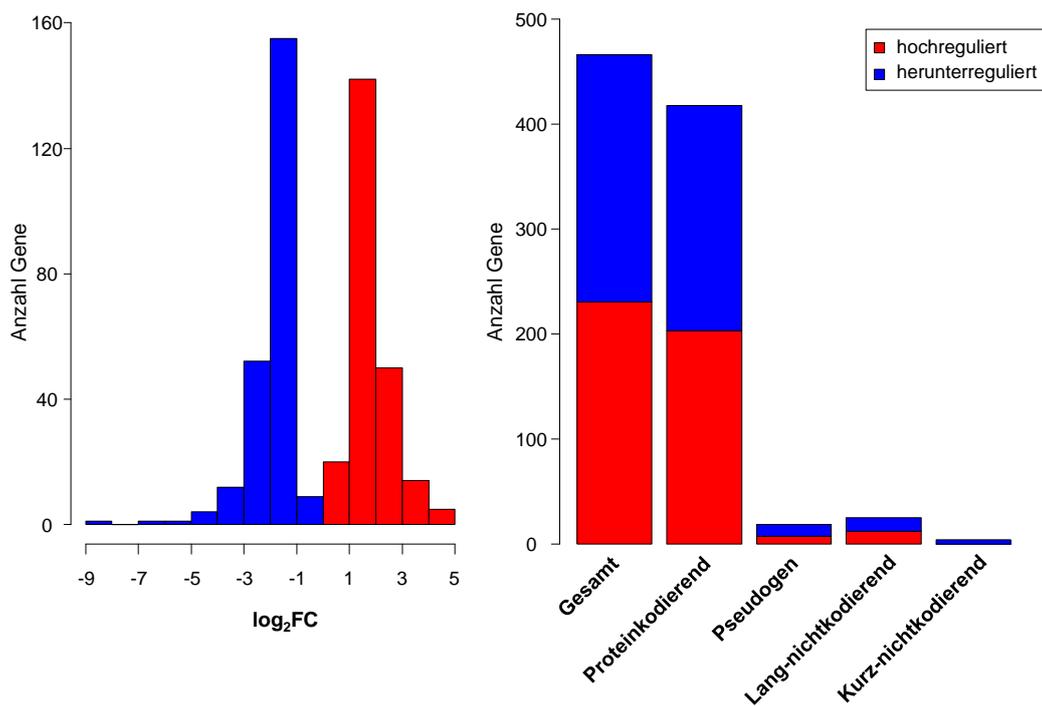


Abbildung 45: Aufteilung der differentiell exprimierten Gene des Total-RNA-Seq-Experimentes in ES-2-Zellen nach IGF2BP1-Knockdown im Vergleich zu Kontrollzellen. Links: Histogramm der Verteilung der zur Basis zwei logarithmierten Quotienten der mittleren Expressionswerte (\log_2FC) von Kontrollen und IGF2BP1-Knockdown. Rechts: Aufteilung der differentiell exprimierten Gene auf Transkripts-kategorien.

der IGF2BP1-mRNA war erwartungsgemäß reduziert ($\log_2FC \approx -2.03$). Die mRNAs der beiden Paralogen IGF2BP2 und IGF2BP3 wurden jedoch nicht differentiell exprimiert. Die vollständige Liste der differentiell exprimierten Gene befindet sich im Anhang (Tabelle A1). Um zu untersuchen, ob sich unter den differentiell exprimierten Genen sogenannte Haushaltsgene befanden, also solche, welche die grundlegenden Funktionen aller Zellen eines Organismus steuern, wurde die Liste der differentiell exprimierten Gene mit der in [25] vorgestellten Liste von Haushaltsgenen verglichen. Diese Liste wurde aus RNA-Seq Daten 16 verschiedener menschlicher Gewebe erstellt. Sie umfasst 3804 Gene welche in allen untersuchten Geweben exprimiert wurden und deren Expressionsvarianz sehr niedrig war. Keines der 466 unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene war in der Liste der Haushaltsgene enthalten.

MicroRNA Die microRNA-Seq-Daten umfassten im Durchschnitt $12.4 \cdot 10^6$ sequenzierter Reads je Probe. Nach Adapterentfernung und Qualitätskürzungen wurden je Probe im Durchschnitt 25.4 % der Basen entfernt. Dies führte zur Entfernung von durchschnittlich 16.2 % der Reads einer Probe, da die Längen dieser Reads die Mindestlänge von 16 Basen unterschritten. Durchschnittlich konnten je Probe 95.9 % der übrigen Reads auf das menschliche Genom gemappt werden. Im Mittel $2.2 \cdot 10^6$ der gemappten Reads konnten mit Hilfe der miRBase, Version 20 [63], microRNA-Annotationen zugeordnet werden. Die Sequenzen der restlichen gemappten Reads entsprachen demnach nicht denen annotierter microRNA-Sequenzen. Bei den Tests auf differentielle Expression zeigte sich, dass sich die normalisierten Expressionswerte der Kontrollprobe 1 und Knockdown-Probe 3 deutlich von denen der anderen Proben unterschieden. Deshalb wurden diese Proben von den weiteren Auswertungen ausgeschlossen. Die Auswertung der übrigen microRNA-Seq-Daten des IGF2BP1-Knockdown ergab, dass 35 verschiedene reife microRNAs in den ES-2-Zellen differentiell exprimiert ($FDR < 0.05$) waren. Von diesen waren 25 im Vergleich zu den Kontrollzellen in ihrer Expression gesteigert und 10 reduziert. Abbildung 46 zeigt die Expression der differentiell exprimierten microRNAs in einer Heatmap, sowie die Verteilung der Expressionsveränderungen. Eine Auflistung der differentiell exprimierten microRNAs befindet sich im Anhang (Tabelle A2).

3.4.1 Beeinflussung migrationsrelevanter Gene

Viele der differentiell exprimierten Gene kodieren Proteine, welche an der Interaktion mit oder an der Umgestaltung der extrazellulären Matrix beteiligt und damit wichtig für Zell-Matrix-Adhäsion und Zellmigration sind.

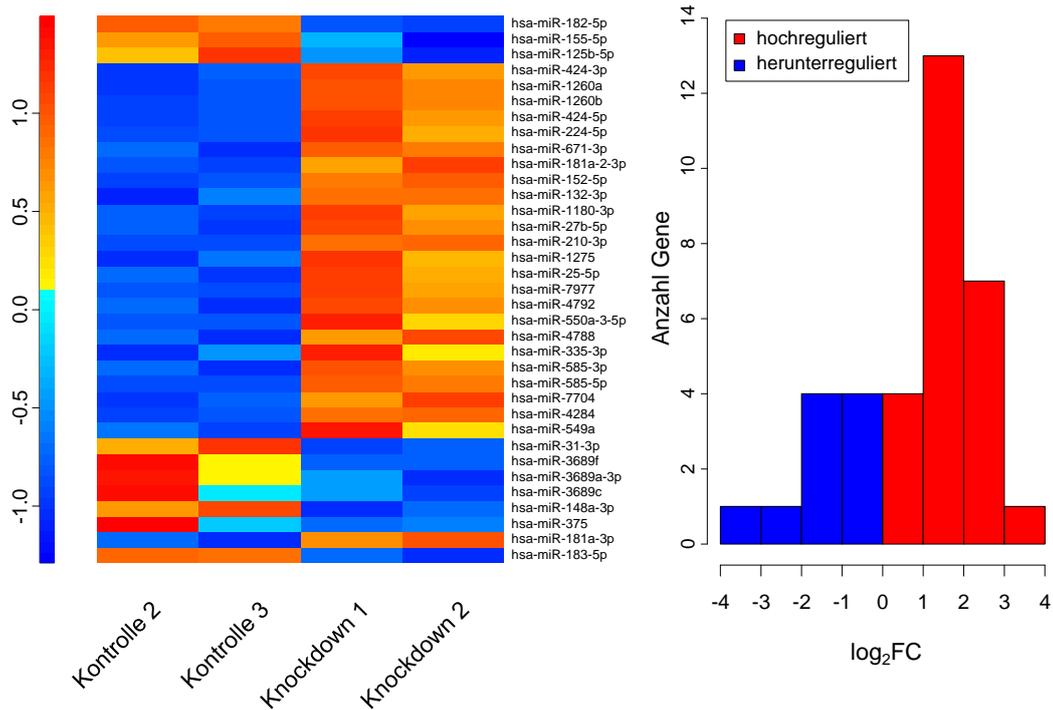


Abbildung 46: Expression der differentiell exprimierten microRNAs. Links: Heatmap der Expressionen der unter IGF2BP1-Knockdown differentiell exprimierten microRNAs. Für die Darstellung wurden die RPM-Werte pro Gen so skaliert, dass die Mittelwerte in allen Proben 0 und die Standardabweichungen 1 ergaben. Rechts: Histogramm der Verteilung der zur Basis zwei logarithmierten Quotienten der mittleren Expressionswerte (\log_2FC) von Kontrollen und IGF2BP1-Knockdown.

Dazu gehörten insgesamt fünf Serpine. Die Mitglieder dieser Proteinfamilie sind überwiegend Proteaseinhibitoren und für einige wurden Funktionen in der Umgestaltung der extrazellulären Matrix beschrieben ([107]). Die Expression von Serpin B2 war unter IGF2BP1-Knockdown um circa 90 % reduziert ($\log_2 FC \approx -3.23$). Für dieses Serpin wurde jedoch beschrieben, dass es Tumorzellmigration inhibiert [32]. Zwei Mitglieder der Serpinunterfamilie der Ov-Serpine (Serpine B7 und Serpin B8), waren ebenfalls herunterreguliert. Serpin E1, welches sowohl inhibitorisch als auch stimulatorisch auf die Zellmigration wirken kann [32] war, ebenso wie Serpin H1, ein Kollagenbindeprotein [107], circa dreifach hochreguliert. Weiterhin befanden sich unter den differentiell exprimierten Genen solche, welche für Untereinheiten von Proteinen der Integrinfamilie kodieren. Diese sind an der Verankerung von Zellen an die extrazelluläre Matrix beteiligt. Einige (ITGB5, ITGB2, ITGA11) waren unter IGF2BP1-Knockdown hochreguliert, andere (ITGA6, ITGA10) waren jedoch herunterreguliert. Weiterhin war die Expression von verschiedenen Proteasen, welche den proteolytischen Abbau von Bestandteilen der extrazellulären Matrix bewirken, verändert. Dazu gehörte die Matrix-Metalloproteinase 1 (MMP1), deren Expression in verschiedenen Krebserkrankungen verstärkt ist [91]. Ihre Expression war nach IGF2BP1-Knockdown reduziert. Dies wurde auch in [117] in HeLa-Zellen unter simultanem IGF2BP1- und IGF2BP3-Knockdown beobachtet. Die Expression der Matrix-Metalloproteinase 2 (MMP2) war hingegen verstärkt. MMP1 ist eine Kollagenase, MMP2 gehört zu den Gelatinasen, welche denaturierte Kollagene (Gelatine) spalten [118]. Die Transkripte von vier ADAMTS-Genen, welche ebenfalls für Metalloproteinasen kodieren, waren differentiell exprimiert. Die Expression der Transkripte der Gene ADAMTS15 und ADAMTS2 war unter IGF2BP1-Knockdown verstärkt, während die Expression der Transkripte von ADAMTS6 und ADAMTS18 reduziert war. Auch Transkripte von Proteinen, welche selbst Bestandteil der extrazellulären Matrix sein können, waren differentiell exprimiert. Dazu gehörten Transkripte von fünf Kollagenen, deren Expression jeweils erhöht war. Darüberhinaus war die Expression von Fibronectin herunterreguliert und die Expression einer Laminin-Untereinheit (LAMB2) war hochreguliert.

Genanreicherungen Die differentielle Expression der oben genannten Gene wies bereits darauf hin, dass IGF2BP1 einen Einfluss auf Prozesse ausübt, welche der Zellmigration zugrunde liegen. Um zu testen, ob die Anzahl der differentiell exprimierten Gene, welche an diesen Prozessen beteiligt sind, statistisch signifikant größer war als zufällig zu erwarten wäre, wurden Genanreicherungsanalysen (s. Abschnitt 2.6.4) durchgeführt. Für die drei Ontologien der Gene Ontology [17] für biologische Prozesse, zelluläre Komponenten und molekulare Funk-

tionen, sowie für KEGG PATHWAY [59], wurden jeweils unabhängige Analysen durchgeführt. Für die Analysen wurden die 466 Gene verwendet, welche durch Auswertung der Total-RNA-Seq-Daten als differentiell exprimiert bestimmt worden waren.

In der Ontologie für biologische Prozesse waren Annotationen für 382 der 466 Gene vorhanden. Nach Entfernung hochredunder Kategorien und Korrektur der p-Werte lieferten die hypergeometrischen Anreicherungstests 332 signifikant ($FDR < 0.05$) angereicherte Kategorien für biologische Prozesse. Abbildung 47 zeigt die 20 am signifikantesten angereicherten Kategorien für biologische Prozesse und die Anzahl der darin enthaltenen hoch- beziehungsweise herunterregulierten Gene.

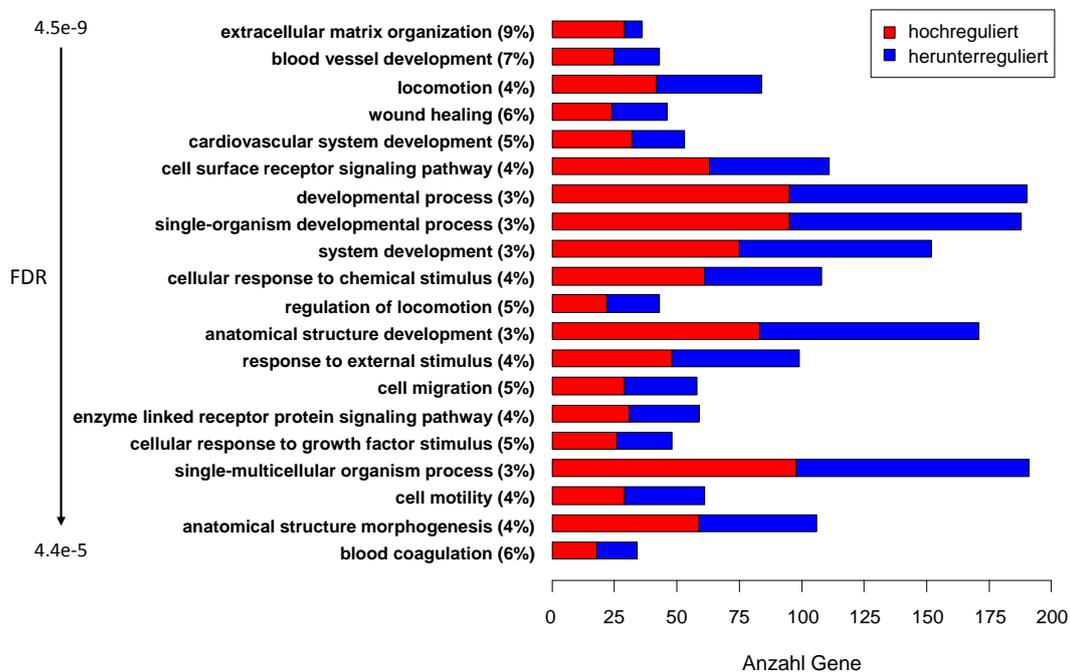


Abbildung 47: Top-20 der signifikant angereicherten Kategorien der Ontologie für biologische Prozesse unter Verwendung der 466 unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene. Die Zahlen hinter den Kategorienamen geben den prozentualen Anteil der zugeordneten differentiell exprimierten Gene an der Gesamtzahl der Gene an, welche Annotationen für die entsprechende Kategorie aufwiesen.

Die als am wahrscheinlichsten als angereichert betrachtete Kategorie war „extracellular matrix organisation“, welcher 36 der differentiell exprimierten Gene zugeordnet waren und von denen der Großteil unter IGF2BP1-Knockdown hochreguliert war. Damit war die Expression von ungefähr 9 % aller Gene welche in der Ontologie für biologische Prozesse mit dieser Kategorie

annotiert waren, verändert. Zudem war ein Großteil der im vorigen Paragraphen beschriebenen Gene in dieser Kategorie enthalten. Nach „blood vessel development“ folgte „locomotion“ an dritter Stelle der signifikant angereicherten Kategorien mit 84 der 466 differentiell exprimierten Gene. Die direkte Unterkategorie „cell motility“ folgte auf Platz 18 und „cell migration“, welche wiederum eine direkte Unterkategorie von „cell motility“ darstellt, erreichte Platz 14 der am signifikantesten angereicherten Kategorien für biologische Prozesse. Insgesamt befanden sich unter den signifikant angereicherten Kategorien 27, welche die Begriffe „migration“, „extracellular matrix“ oder „adhesion“ in ihren Namen trugen. Diese Kategorien enthielten 103 der 466 differentiell exprimierten Gene ($\approx 22\%$), wobei 63 davon hoch- und 40 herunterreguliert waren. Abbildung 48 zeigt die Verknüpfungen dieser Kategorien innerhalb der Gene Ontology.

Gene, deren Produkte die extrazelluläre Matrix und die Adhäsion zu anderen Zellen, sowie zur extrazellulären Matrix verändern können, beeinflussen dadurch auch die Zellmigration. Dies kann dadurch veranschaulicht werden, dass von den insgesamt 58 Genen, welche mit „cell migration“ annotiert waren, 17 ebenfalls die Annotation „extracellular matrix organization“ und 27 die Annotation „cell adhesion“ trugen. Insgesamt 12 dieser Gene trugen Annotationen für alle drei Kategorien. Zu diesen gehörten unter anderem Fibronectin, sowie verschiedene für Kollagene und Integrine kodierende Gene. Abbildung 49 zeigt die Überlappungen der drei Kategorien in einem Venn-Diagramm.

Viele der weiteren signifikant angereicherten Kategorien für biologische Prozesse deuteten darauf hin, dass IGF2BP1 auch einen Einfluss auf Gene ausübt, deren Produkte an der Entwicklung verschiedener Organe und Organsysteme beteiligt sind.

Die Ontologie für zelluläre Komponenten, welche Genprodukte nach ihrer Lokalisation inner- und außerhalb von Zellen annotiert, enthielt Annotationen für 402 der 466 differentiell exprimierten Gene. Mittels der Genanreicherungsanalyse für diese Ontologie konnten insgesamt 24 signifikant angereicherte Kategorien bestimmt werden. Die signifikanteste Kategorie war „extracellular matrix“. Die Kategorien auf dem zweiten und dritten Platz stellten Unterkategorien von „extracellular matrix“ dar. Ein großer Teil der differentiell exprimierten Gene, welche diesen drei Kategorien zugeordnet waren, war unter IGF2BP1-Knockdown hochreguliert. Darüber hinaus befanden sich auch „focal adhesion“ und „cell junction“ unter den signifikant angereicherten Kategorien. Die diesen Kategorien zugeordneten Gene waren in ihrer Expression ebenfalls überwiegend hochreguliert.

Die Anreicherungsanalyse für die Ontologie für molekulare Funktionen führte zu insgesamt 12 signifikant angereicherten Kategorien, denen 380 der 466 differentiell exprimierten Gene

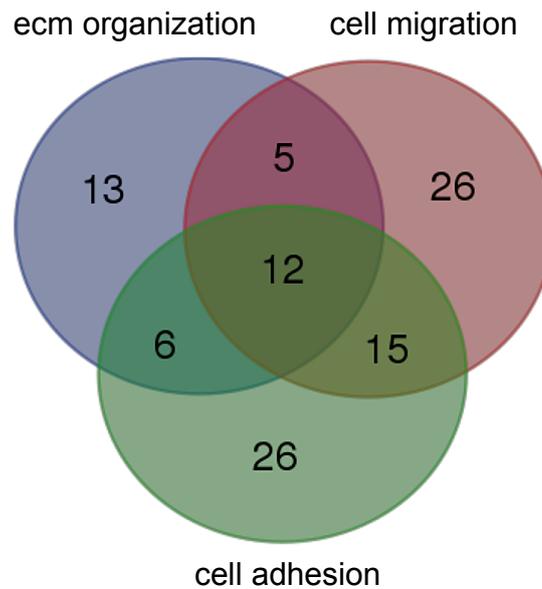


Abbildung 49: Überlappungen der differentiell exprimierten Gene mit Annotationen aus den Kategorien „cell migration“, „extracellular matrix organization“ und „cell adhesion“ der Ontologie für biologische Prozesse.

zugeordnet werden konnten. Auch unter diesen Kategorien befanden sich solche, welche eine Beteiligung der, durch IGF2BP1-Knockdown in ihrer Expression beeinträchtigten, Gene an Zellmigration vermuten lassen. Dazu zählten beispielsweise „integrin binding“, „cell adhesion molecule binding“ und „metalloendopeptidase activity“. Auffällig war, dass 273 der 380 Gene ($\approx 72\%$), welche Kategorien zugeordnet werden konnten, als proteinbindend kategorisiert wurden. Jedoch beinhaltet diese Kategorie ungefähr 11000 und damit etwa 64% aller Gene, für die Annotationen in dieser Ontologie vorhanden waren.

Neben Kategorien der Gene Ontology wurde auch nach angereicherten Proteininteraktionsnetzwerken der Kyoto Encyclopedia of Genes and Genomes (KEGG PATHWAY, [59]) gesucht. Diese Netzwerke enthalten Informationen über die Interaktion von Proteinen in bestimmten Stoffwechselwegen, regulatorischen Netzwerken oder Krankheiten. KEGG PATHWAY besitzt Annotationen für deutlich weniger Gene als Gene Ontology (6989 in der Version vom 18.02.2016), wobei nur 175 der 466 differentiell exprimierten Gene zu diesen gehörten. Dennoch befanden sich unter den sechs signifikant angereicherten Kategorien mit „PI3K-Akt signaling pathway“, „ECM-receptor interaction“ und „Focal adhesion“ solche, die einen Zusammenhang mit Zellmigration vermuten lassen. Vollständige Auflistungen der signifikant angereicherten Kategorien für Gene Ontology und KEGG befinden sich im Anhang (Tabellen A3 - A6).

Zusammengefasst zeigte sich, dass mehr als 20 % der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene an signifikant angereicherten biologischen Prozessen beteiligt sind, welche mit Zellmigration in Verbindung stehen. Unter den am signifikantesten angereicherten Kategorien aller Ontologien und KEGG befanden sich jeweils solche, welche die extrazelluläre Matrix betreffen. Zu diesen zählten Kategorien für die Organisation, die Adhäsion an, sowie die Interaktion mit der extrazellulären Matrix. Es ist bekannt, dass IGF2BP1 über verschiedene Faktoren einen Einfluss auf das Aktinzytoskelett ausübt. Unter den signifikant angereicherten Kategorien der unter IGF2BP1-Knockdown differentiell exprimierten Gene befand sich jedoch keine, welche den Begriff „cytoskeleton“ enthielt.

Epithelial-zu-mesenchymale Transition Die hier untersuchten Zellen der Linie ES-2 stammen aus einem Ovarialkarzinom und zeigen eine hohe Motilität, welche durch IGF2BP1-Knockdown reduziert wird (s. Abschnitt 3.3). Die beobachtete Veränderung der Morphologie der Zellen, sowie die Ergebnisse der Genanreicherungsanalysen deuten darauf hin, dass diese Geschwindigkeitsreduktion mit veränderter Adhäsion und Umstrukturierung der extrazellulären Matrix einhergeht. Veränderungen am Aufbau und an der Interaktion mit der extrazellulären Matrix, sowie veränderter Aufbau von Zell-Zell-Kontakten und als Folge daraus auch eine Änderung der Zellmigration sind ebenfalls beim Vorgang der epithelial-zu-mesenchymalen Transition (EMT) zu beobachten. Durch Reaktivierung dieses Entwicklungsprogramms können Tumorzellen epithelialen Ursprungs die Fähigkeit zur Migration erwerben [49].

In [127] wurde für HEK-293A-Zellen beschrieben, dass IGF2BP1-Expression den Vorgang der EMT zwar nicht induziert, aber mesenchymale Eigenschaften von Tumorzellen aufrechterhält, indem es beispielsweise die Expression der EMT-induzierenden Faktoren LEF-1 und SNAI2 fördert, wodurch wiederum die Expression des mesenchymalen Markers Fibronectin erhöht wird. Neben diesen drei Faktoren zählen auch E-cadherin und β -Catenin, deren Expression nach [127] ebenfalls in HEK-293A-Zellen von IGF2BP1 beeinflusst wird, zu einer Liste von Proteinen, welche nach [124] Biomarker für EMT darstellen.

In den vorliegenden Total-RNA-Seq-Daten des IGF2BP1-Knockdown in ES-2-Zellen konnte keine Reduktion der mRNA-Expression von LEF-1 und SNAI2 beobachtet werden. Die Expression von SNAI2 war stattdessen leicht erhöht ($\log_2FC \approx 1.18$). Allerdings war die Expression der Fibronectin-mRNA auch hier reduziert ($\log_2FC \approx -1.42$). Eine Erhöhung der E-Cadherin-Transkriptmenge, wie in [127] für HEK-293A-Zellen unter IGF2BP1-Kockdown beobachtet, konnte ebenfalls nicht festgestellt werden, jedoch lagen die ermittelten durchschnittlichen Expressionswerte für dieses Gen in beiden Populationen unter 0.04 FPKM, was

auf eine fehlende Expression dieses epithelialen Markergens in ES-2-Zellen hinweisen kann. Für die mRNA von β -Catenin konnte ebenfalls keine signifikante Reduktion der Expression festgestellt werden.

Gemäß den oben beschriebenen Tests auf angereicherte biologische Prozesse, zählten von den insgesamt 113 Genen, welche nach Gene Ontology-Annotation mit EMT („epithelial to mesenchymal transition“, GO:0001837) assoziiert waren, nur sechs zu den unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Genen. Nach Korrektur für multiples Testen wurde diese Kategorie nicht als signifikant angereichert betrachtet (p-Wert ≈ 0.04 , FDR ≈ 0.15). Jedoch werden Gene Ontology-Annotationen aus den Ergebnissen verschiedenartiger Experimente oder Analysen bezogen. Für einen relevanteren Vergleich sollten die differentiell exprimierten Gene daher mit EMT-bezogenen Genexpressionsdaten verglichen werden. In [44] wurde eine Liste von EMT-bezogenen Genen erstellt. Dies geschah, indem aus einer Vielzahl von Genexpressionsexperimenten in denen EMT induziert wurde, diejenigen Gene extrahiert wurden, welche in mindestens 10 Experimenten differentiell exprimiert waren. Diese Liste besteht aus 365 proteinkodierenden Genen. Davon waren 35 auch im hier beschriebenen IGF2BP1-Knockdown-Experiment der ES-2-Zellen differentiell exprimiert. Ausgehend von der 365 Gene umfassenden Liste wurde in [44] eine weitere, reduzierte Liste generiert, welche nur die Gene enthält, welche in mindestens 10 der untersuchten Experimente konsistent hoch- beziehungsweise herunterreguliert waren. Diese reduzierte Liste besteht aus 131 Genen, von denen neun im vorliegenden Total-RNA-Seq differentiell exprimiert waren.

Es wurde getestet, ob die Anzahl der unter IGF2BP1-Knockdown differentiell exprimierten Gene, welche sich in den beiden EMT-Listen befanden, signifikant höher war als zufällig zu erwarten wäre. Dazu wurde, wie bei den Genanreicherungsanalysen, die Verteilung der differentiell exprimierten Gene, welche sich auch in den EMT-Listen befanden, als hypergeometrische Verteilung modelliert. Anschließend wurde die Wahrscheinlichkeit dafür berechnet, mindestens die beobachtete Anzahl differentiell exprimierter Gene in diesen Listen zu erhalten. Als Gesamtzahl der Gene wurde die Anzahl der Gene verwendet, welche laut Ensembl-Annotation GRCh37.75 ([18]) als proteinkodierend annotiert waren. Dies waren 20327 Gene. Als Stichprobengröße wurde die Anzahl der differentiell exprimierten proteinkodierenden Gene verwendet, also 418.

Das bedeutet, es wurde die Wahrscheinlichkeit berechnet, aus einer Gesamtmenge von 20327 Genen, von denen 365 (131) am EMT-Prozess beteiligt sind, in 418 Versuchen zufällig mindestens 35 (9) dieser EMT-assoziierten Gene auszuwählen. Tabelle 19 zeigt die Ergebnisse der Anreicherungstests. In beiden Listen war der Anteil der unter IGF2BP1-Knockdown differen-

Liste	Anzahl Gene	davon differentiell exprimiert	P-Wert
EMT_dExp	365	35	$4.00 \cdot 10^{-14}$
EMT_dExp core	131	9	$1.57 \cdot 10^{-3}$

Tabelle 19: Ergebnisse der hypergeometrischen Tests auf Anreicherung von EMT-bezogenen Genen innerhalb der Menge der in ES-2-Zellen unter IGF2BP1-Knockdown differentiell exprimierten proteinkodierenden Gene. Der P-Wert gibt die Wahrscheinlichkeit dafür an, zufällig mindestens die beobachtete Anzahl differentiell exprimierter Gene aus den entsprechenden Listen zu erhalten. EMT_dExp: Gene welche in mindestens 10 Genexpressionsexperimenten zu EMT differentiell exprimiert waren. EMT_dExp core: Gene welche in mindestens 10 Genexpressionsexperimenten zu EMT konsistent hoch- oder herunterreguliert waren (aus [44]).

tiell exprimierten Gene signifikant höher (P-Wert < 0.05) als zufällig zu erwarten wäre. Tabelle 20 zeigt eine Übersicht über die Expressionsveränderungen der neun differentiell exprimierten Gene, welche sich auch in der reduzierten Liste (EMT_dExp core) befanden.

Unter der Annahme, dass IGF2BP1 EMT aufrecht erhält, würde man durch die Reduzierung der Expression dieses Proteins einen gegenteiligen Effekt auf diese Gene erwarten. Für sechs der neun Gene galt dies jedoch nicht. Die Expression des EMT-Markers Fibronectin (FN1) [124], sowie der Gene TMEM158 und PLAT zeigten hingegen den erwarteten Trend.

In [44] wurden Genanreicherungstests für KEGG-Netzwerke, sowie biologische Prozesse und molekulare Funktionen der Gene Ontology mit den Genen aus den dort generierten EMT-Listen durchgeführt. Dabei wurden solche Kategorien ausgewählt, welche für beide EMT-Listen die dort verwendeten Kriterien für Signifikanz erfüllten. Von den so ermittelten 20 Kategorien für biologische Prozesse waren 17 auch in den oben beschriebenen angereicherten Kategorien der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene vorhanden. Sieben der 15 ermittelten Kategorien für molekulare Funktionen waren ebenfalls im IGF2BP1-Knockdown signifikant angereichert. Von den sechs angereicherten KEGG-Netzwerken waren „ECM-receptor interaction“, „focal adhesion“ und „pathways in cancer“ auch unter IGF2BP1-Knockdown angereichert.

Insgesamt deuten diese Ergebnisse auf eine Beteiligung der differentiell exprimierten Gene an der epithelial-zu-mesenchymalen Transition und unterstützen damit die Theorie, dass die Expression von IGF2BP1 in Verbindung mit diesem Vorgang steht. Für viele der biologischen Prozesse und Reaktionswege, in denen, gemäß der Daten aus [44], die Expression signifikant

Gensymbol	EMT_dExp core	IGF2BP1-KD
MMP2	+	+
ADAM12	+	+
TMEM158	+	-
COL1A1	+	+
PLAT	+	-
FN1	+	-
SERPINE1	+	+
COL5A1	+	+
FBN1	+	+

Tabelle 20: Expressionsveränderungen von Genen welche in mindestens 10 Genexpressionsexperimenten zu EMT (EMT_dExp core) konsistent hoch- (+) oder herunterreguliert (-) waren ([44]) und nach IGF2BP1-Knockdown in ES-2-Zellen (IGF2BP1-KD) ebenfalls differentiell exprimiert waren.

vieler Gene durch EMT beeinflusst wird, sind auch signifikant viele Gene nach IGF2BP1-Knockdown in ES-2-Zellen in ihrer Expression verändert. Der Vergleich mit Genen, welche nach EMT-Induktion konsistent hoch- oder herunterreguliert waren ergab lediglich für drei Gene einen gegenteiligen Expressionseffekt durch IGF2BP1-Knockdown. Jedoch waren die Experimente, welche als Grundlage für die Erstellung der Liste der konsistent exprimierten Gene dienten, in verschiedenen Zellsystemen und mit unterschiedlichen EMT-induzierenden Faktoren durchgeführt worden.

3.4.2 Bestimmung potentieller neuer IGF2BP1-Zieltranskripte

Die Bindung von IGF2BP1 an Zieltranskripte kann die Stabilität, die Translation und den Transport dieser Transkripte beeinflussen. Für den Großteil der bekannten Zieltranskripte wurde eine Erhöhung der Stabilität nachgewiesen. Die Bindung des IGF2BP1-Proteins an mRNAs erfolgt häufig in deren nichttranslatierten Bereich des 3'-Endes (3'-UTR) [4]. RNA-Seq-Daten können helfen, indirekt Einflüsse auf die Stabilität von Transkripten zu untersuchen. Die Expression von IGF2BP1-Zieltranskripten, deren Stabilität durch Bindung an das Protein erhöht wird, sollte in IGF2BP1-Knockdown-Populationen reduziert sein. Umgekehrt sollte die Expression von Transkripten, welche durch IGF2BP1-Bindung destabilisiert werden, unter IGF2BP1-Knockdown verstärkt sein. Die 3'-UTR-Bereiche vieler mRNAs enthalten oft

Bindemotive für Faktoren, welche die posttranskriptionelle Genexpression regulieren. Bei diesen Faktoren kann es sich um RNA-bindende Proteine (RBPs) oder microRNAs handeln. Längere 3'-UTR-Sequenzen erhöhen dabei die Wahrscheinlichkeit für microRNA- oder RBP-Bindestellen.

Mit Hilfe des R-Paketes biomaRt [22] wurden die nach Ensembl (GRCh37.75) [18] längsten bekannten 3'-UTR-Sequenzen proteinkodierender Gene bestimmt. Abbildung 50 zeigt die Verteilungen der Längen dieser längsten bekannten 3'-UTR-Sequenzen von in ES-2-Zellen unter IGF2BP1-Knockdown herunterregulierten, hochregulierten, sowie zufällig ausgewählten proteinkodierenden Genen, welche mit hoher Wahrscheinlichkeit nicht-differenziell exprimiert (FDR = 1) waren.

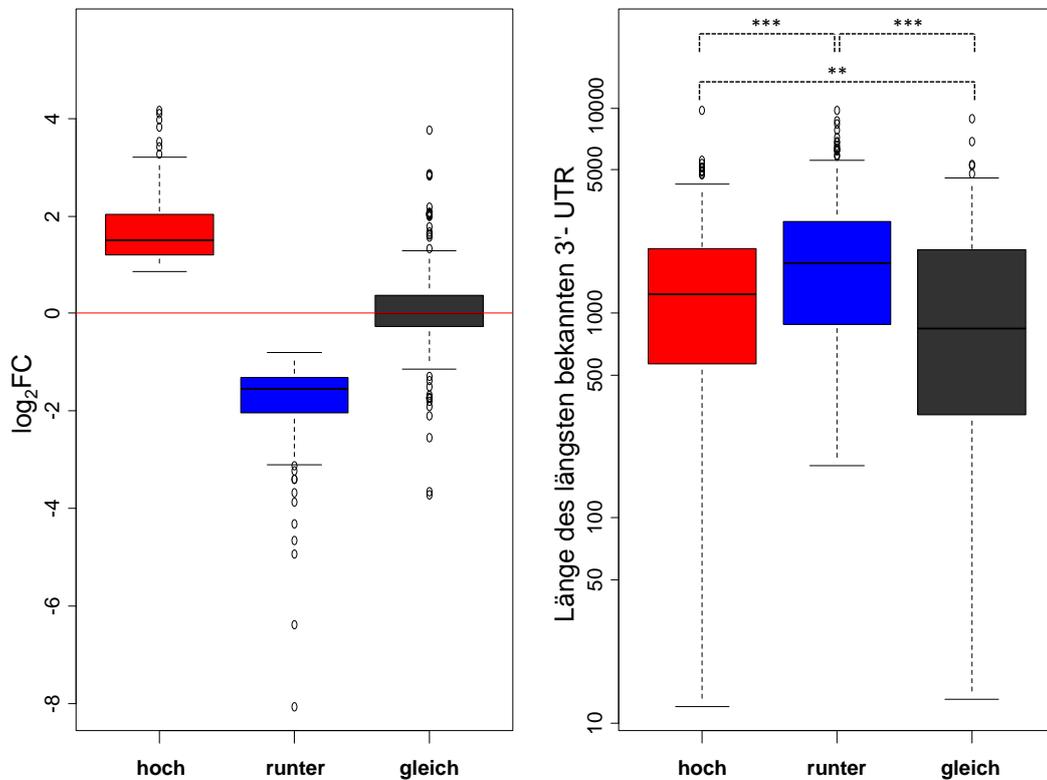


Abbildung 50: Längste bekannte 3'-UTR-Längen von unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Genen. Links: Zur Basis zwei logarithmierte Quotienten der mittleren Expressionswerte (\log_2FC) von Kontrollen und IGF2BP1-Knockdown der 203 hochregulierten (hoch), 215 herunterregulierten (runter) und 209 zufällig ausgewählten, nicht differentiell exprimierten (gleich, FDR = 1), proteinkodierenden Gene. Rechts: Längen der nach Ensembl (GRCh37.75) [18] längsten bekannten 3'-UTR-Sequenzen der Gene. Mittels einseitigen Mann-Whitney-Tests wurde auf statistisch signifikant längere 3'-UTR-Längen geprüft; **: p-Wert < 0.01; ***: p-Wert < 0.001.

Die längsten bekannten 3'-UTR-Sequenzen der unter IGF2BP1-Knockdown herunterregulierten proteinkodierenden Gene waren signifikant länger als diejenigen der hochregulierten, sowie nicht differentiell exprimierten, Genen. Des Weiteren waren die längsten bekannten 3'-UTR-Sequenzen der hochregulierten Gene ebenfalls signifikant länger als die der nicht-differentiell exprimierten Gene.

Obwohl nicht untersucht wurde, ob die theoretisch längsten 3'-UTR-Sequenzen der Transkripte den Varianten entsprachen, welche tatsächlich in ES-2-Zellen exprimiert wurden, kann dies als Hinweis dafür betrachtet werden, dass sich unter den differentiell exprimierten Genen viele Zieltranskripte von IGF2BP1, anderen RNA-bindenden Proteinen oder microRNAs befinden.

Putative IGF2BP1-Zieltranskripte Um zu untersuchen, welche Transkripte der differentiell exprimierten Gene möglicherweise direkt von IGF2BP1 gebunden werden, wurden die Ergebnisse des Total-RNA-Seq-Experimentes mit bereits veröffentlichten CLIP-Daten verglichen. Hierfür wurde ein PAR-CLIP-Datensatz für die drei Mitglieder der humanen IGF2BP-Familie aus der RNA-Interaktionsdatenbank doRiNA [2] verwendet. In den ursprünglichen Experimenten wurden die Mitglieder der IGF2BP-Familie in Zellen der menschlichen Nierenzelllinie HEK293 überexprimiert [46]. Für die Erstellung des doRiNA-Datensatzes wurden die ermittelten genomischen Positionen der sequenzierten RNA-Fragmente mit Hilfe der Referenzannotation des humanen Genoms (hg19) annotiert. Weiterhin wurden die Fragmente entsprechend ihres genomischen Kontextes dahingehend klassifiziert, ob sie im 3'-UTR, im 5'-UTR, in der kodierenden Sequenz oder in intronischen Bereichen auftraten. Gemäß dieser Klassifizierung wurden vier Listen (5UTR, 3UTR, CDS, Intron) aus der doRiNA-Datenbank heruntergeladen und für die anschließenden Analysen verwendet. Zum Vergleich mit den RNA-Seq-Daten wurden den aus doRiNA stammenden RefSeq-Transkriptbezeichnungen (NM_/NR_) mit Hilfe des R-Paketes BioMart [22] Ensembl-Gene-IDs (ENSG) zugewiesen. Duplikate, also Gennamen mit mehreren Bindestellen dergleichen Kategorie, wurden entfernt. Die so ermittelten 7659 Gene enthielten putative IGF2BP-Bindestellen vor allem in ihren exonischen Bereichen und dort hauptsächlich in den 3'-UTR-Bereichen. Allerdings enthielten 2020 Gene putative Bindestellen in intronischen Bereichen. Da die Mitglieder der IGF2BP-Familie vorwiegend im Cytoplasma vorkommen und keine Beteiligung der Proteine an Spleißprozessen bekannt ist, deutet dies auf eine hohe Falsch-Positiv-Rate der CLIP-Stellen hin. Eine mögliche Ursache für eine hohe Zahl falsch bestimmter Bindestellen ist die Überexpression der Proteine. Durch die daraus resultierende größere Menge an IGF2BP-Proteinen wurde die

Wahrscheinlichkeit erhöht, dass IGF2BP-Proteine und RNA-Moleküle innerhalb der Zelle in unmittelbare Nähe gelangen, wodurch unspezifische Kontaktstellen durch den Cross-Linking-Prozess konserviert wurden.

Ein Großteil der durch weitere Methoden validierten IGF2BP1-Zielgene tauchte in den PAR-CLIP-Daten auf. Allerdings ist die Wahrscheinlichkeit für ein zufälliges Auftauchen dieser Gene aufgrund der großen Menge an Kandidaten hoch und die Validität der restlichen Gene kann damit nicht unterstützt werden. Das Vorhandensein eines Genes in den PAR-CLIP-Listen wurde daher lediglich als Hinweis für eine mögliche Bindung des entsprechenden Transkripts an ein IGF2BP-Protein betrachtet.

Von den 418 proteinkodierenden Genen, welche in ES-2-Zellen unter IGF2BP1-Knockdown als differentiell exprimiert bestimmt worden waren, entsprachen 201 ($\approx 43\%$) putativen Bindungspartnern von IGF2BPs in HEK293-Zellen. Von diesen waren 144 ($\approx 72\%$) in ES-2-Zellen unter IGF2BP1-Knockdown herunterreguliert. Die Stabilität der entsprechenden Transkripte könnte durch Bindung an IGF2BP1 erhöht werden. Abbildung 51 zeigt die Aufteilung der differentiell exprimierten Gene nach dem Vorhandensein von PAR-CLIP-Bindestellen. In Ab-

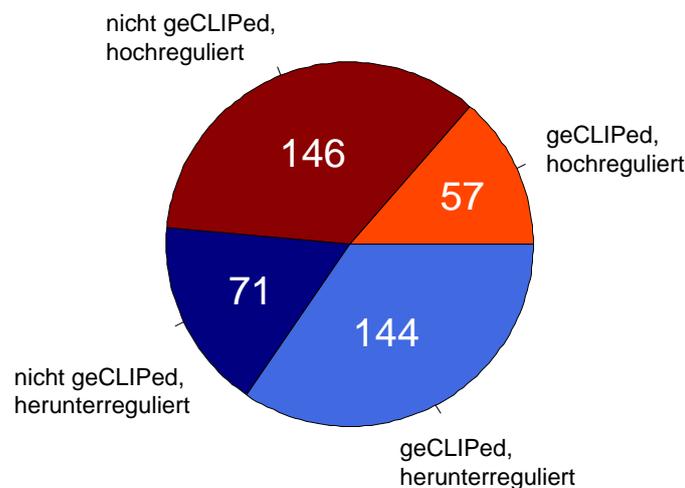


Abbildung 51: Aufteilung der unter IGF2BP1-Knockdown in ES-2-Zellen hoch- beziehungsweise herunterregulierten proteinkodierenden Gene auf solche, welche in HEK-293-Zellen PAR-CLIP-Bindestellen aufwiesen (geCLIPed) oder nicht (nicht geCLIPed).

Abbildung 52 zeigt die Aufteilung der unterschiedlichen Bindestellenkategorien der differentiell exprimierten Gene mit PAR-CLIP-Bindestellen. Für die überwiegende Mehrheit ($\approx 92\%$) der differentiell exprimierten proteinkodierenden Gene mit PAR-CLIP-Bindestellen waren solche in den 3'-UTR-Bereichen der entsprechenden mRNAs zu finden. Aber auch innerhalb der

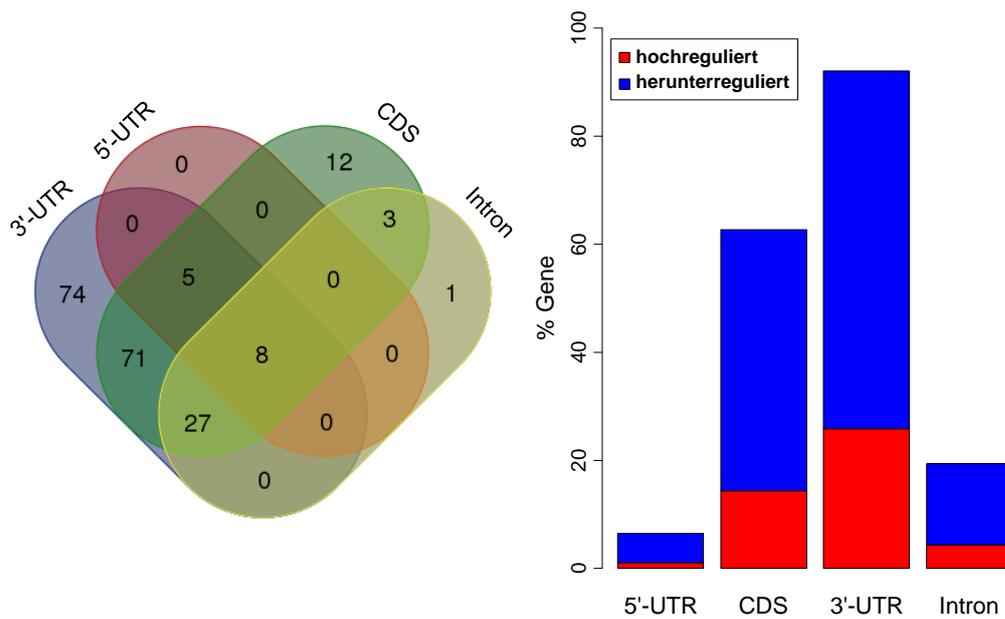


Abbildung 52: Verteilung von PAR-CLIP-Bindestellen der Gene, deren Transkripte in ES-2-Zellen nach IGF2BP1-Knockdown differentiell exprimiert wurden. Links: Überlappungen der Bindestellenkategorien der insgesamt 201 differentiell exprimierten Gene, welche PAR-CLIP-Bindestellen für Proteine der IGF2BP-Familie in HEK293-Zellen aufwiesen. Rechts: Kategorien der CLIP-Bindestellen und der Prozentsatz der jeweils hoch- beziehungsweise herunterregulierten Gene, welche die entsprechenden Bindestellen aufwiesen.

kodierenden Bereiche waren häufig ($\approx 62\%$) Bindestellen vorhanden. Und für etwa jedes fünfte Gen waren Bindestellen innerhalb von intronischen Bereichen bestimmt worden. Für die Mehrzahl der Gene gab es in jeweils mehreren der genannten Bereiche Bindestellen. Die mRNAs von acht Genen wiesen Bindestellen aus allen vier Kategorien auf, darunter auch die mRNA des IGF2BP1-Gens.

Zusammengefasst ergab der Vergleich der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene mit PAR-CLIP-Daten für IGF2BP1-Proteine, dass 201 der differentiell exprimierten proteinkodierenden Gene in HEK-293-Zellen CLIP-Bindestellen aufwiesen. Die Expression des Großteils dieser Gene war unter IGF2BP1-Knockdown in ES-2-Zellen herunterreguliert. Obwohl die PAR-CLIP-Experimente in einem anderen Zellsystem und unter Überexpression der IGF2BP1-Proteine durchgeführt worden waren, lieferten diese Daten Hinweise auf mögliche IGF2BP1-Zielgene, deren Stabilität durch Bindung an IGF2BP1 in ES-2-Zellen modifiziert werden kann.

Putative microRNA-Ziele Für einige IGF2BP1-Zieltranskripte ist bekannt, dass sie durch Bindung an das IGF2BP1-Protein vor microRNA-vermitteltem Abbau geschützt werden [26, 43, 11]. Um zusätzliche potentielle Zieltranskripte, welche durch IGF2BP1-Bindung vor dem Abbau durch microRNA-Bindung geschützt werden, zu identifizieren, wurden die Resultate des Total-RNA-Seq mit denen des microRNA-Seq verglichen. Mit Hilfe des R-Paketes multi-MiR [101], wurden putative mRNA-microRNA-Bindungen bestimmt (s. Abschnitt 2.6.5).

Hierfür wurden nur solche microRNAs verwendet, welche gemäß der microRNA-Seq-Daten zwei Bedingungen erfüllten. Zum einen sollten die microRNAs sowohl in den Kontroll-, als auch in den Knockdown-Proben eine hohe Abundanz aufweisen. Daher wurden alle microRNAs verworfen, denen im Durchschnitt über alle vier betrachteten Proben weniger als 100 Reads je Million gemappter Reads (RPM) zugeordnet worden waren. Von den übrigen microRNAs wurden nur solche betrachtet, deren Expression mit hoher Wahrscheinlichkeit nicht durch den IGF2BP1-Knockdown beeinträchtigt wurde. Deshalb wurden von den hochabundanten microRNAs diejenigen herausgesucht, deren FDR für den Test auf differentielle Expression 1 betrug. Insgesamt 92 verschiedene microRNAs erfüllten diese Bedingungen. Eine Auflistung dieser microRNAs befindet sich im Anhang (Tabelle A7).

Für die Bestimmung potentieller mRNA-Bindungspartner für diese microRNAs wurden die 203 unter IGF2BP1-Knockdown hochregulierten, sowie die 215 herunterregulierten proteinkodierenden Gene verwendet. Um diese mit einer ähnlich großen Menge nicht-differentiell exprimierter mRNAs zu vergleichen, wurden zusätzlich zufällig 209 proteinkodierende Gene

ausgewählt, deren FDR für den Test auf differentielle Expression 1 betrug. Abbildung 53 zeigt welcher Anteil dieser drei mRNA-Populationen (hochreguliert, herunterreguliert, nicht-differentiell exprimiert), jeweils von den, in ES-2-Zellen am stärksten und nicht-differentiell exprimierten, microRNAs gebunden werden könnte. Zudem sind die Verteilungen der Anzahl der pro mRNA putativ bindenden verschiedenen microRNAs aufgezeigt.

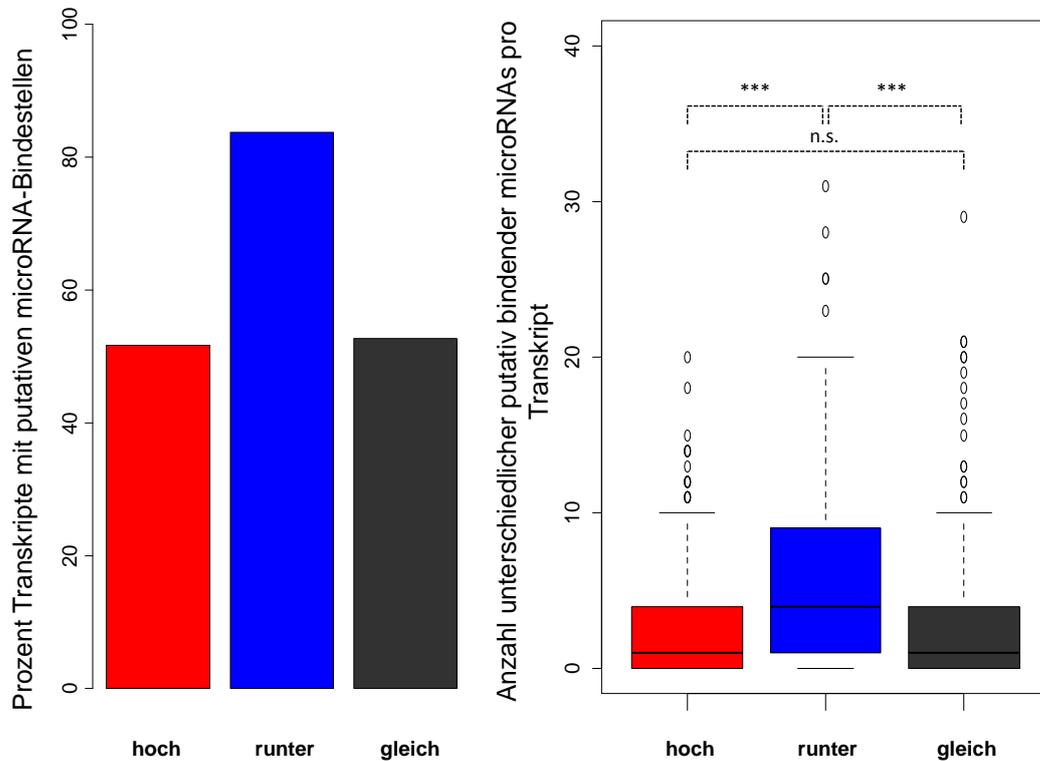


Abbildung 53: Anteil und Verteilung vorhergesagter microRNA-Bindungen von, in ES-2-Zellen unter IGF2BP1-Knockdown, differentiell und nicht-differentiell exprimierten proteinkodierenden Genen. Links: Anteil der hochregulierten (hoch), herunterregulierten (runter) und nicht-differentiell exprimierten (gleich) proteinkodierenden Gene, welche als Ziele der, in ES-2-Zellen am stärksten ($RPM \geq 100$) und nicht-differentiell exprimierten, microRNAs vorhergesagt wurden. Rechts: Verteilungen der pro mRNA putativ bindenden nicht-differentiell exprimierten microRNAs, denen im Durchschnitt mindestens 100 RPM zugeordnet wurden. Die Anzahlen putativ bindender unterschiedlicher microRNAs pro mRNA wurden mittels zweiseitigen Mann-Whitney-Tests auf statistische Signifikanz geprüft; n.s.: nicht signifikant; ***: p -Wert < 0.001 .

Während für circa die Hälfte der hochregulierten beziehungsweise nicht-differentiell exprimierten Gene putativ bindende microRNAs gefunden wurden, war dies für ungefähr 84 % der herunterregulierten Gene der Fall. Zudem war die Anzahl der pro herunterregulierter mRNA

putativ bindenden, verschiedenen microRNAs signifikant höher, als die Anzahl putativ bindender microRNAs für hochregulierte beziehungsweise nicht-differentiell exprimierte Gene. Dies könnte darauf hindeuten, dass viele der unter IGF2BP1-Knockdown in ihrer Expression reduzierten Gene durch Bindung an IGF2BP1 in ihrer Stabilität erhöht werden, indem sie dadurch vor microRNA-Bindung-vermitteltem Abbau geschützt werden.

Um aus der Menge der unter IGF2BP1-Knockdown herunterregulierten proteinkodierenden Gene Kandidaten zu finden, welche möglicherweise durch Bindung an IGF2BP1 vor Abbau, induziert durch microRNA-Bindung, geschützt werden, wurden diejenigen mRNAs herausgesucht, welche vorhergesagte microRNA-Bindungen und zusätzlich PAR-CLIP-Bindestellen aufwiesen. Unter den so bestimmten 124 Genen befand sich mit IGF2BP1 nur ein bereits validiertes Zielgen des IGF2BP1-Proteins. Die übrigen 123 Gene stellen somit putative neue IGF2BP1-Zielgene dar. Insgesamt 17 dieser Gene gehören zu jenen signifikant angereicherten biologischen Prozessen der Gene Ontology, deren Bezeichnungen die Begriffe „migration“, „adhesion“ oder „extracellular matrix“ enthalten. Für diese putativen IGF2BP1-Zielgene kann demnach eine Beteiligung an Zellmigrationsprozessen vermutet werden.

Ob die entsprechenden Transkripte der so bestimmten putativen IGF2BP1-Zielgene tatsächlich von IGF2BP1 gebunden werden und ob diese darüber hinaus durch die Bindung vor microRNA-vermitteltem Abbau geschützt werden, muss jedoch noch durch weitere Experimente validiert werden. Tabelle 21 listet die zehn Kandidaten auf, für welche die höchste Anzahl verschiedener putativ bindender microRNAs vorhergesagt wurde. Die vollständige Liste der potentiellen IGF2BP1-Zielgene befindet sich im Anhang (Tabelle A8).

Gensymbol	\log_2FC	FDR	Anzahl unterschiedlicher putativ bindender microRNAs
ABL2	-1.20	$5.93 \cdot 10^{-3}$	31
SLC7A11	-2.03	$6.77 \cdot 10^{-3}$	28
ARPP19	-1.34	$2.76 \cdot 10^{-2}$	25
CCDC6	-3.11	$3.15 \cdot 10^{-22}$	23
IGF2BP1	-2.03	$1.57 \cdot 10^{-14}$	20
DCBLD2	-1.32	$7.94 \cdot 10^{-3}$	20
UHMK1	-1.28	$1.99 \cdot 10^{-2}$	19
BZW1	-1.41	$3.94 \cdot 10^{-2}$	19
BEND4	-1.19	$5.83 \cdot 10^{-3}$	18
MTDH	-1.51	$4.02 \cdot 10^{-2}$	18

Tabelle 21: Potentielle IGF2BP1-Zielgene mit der höchsten Anzahl vorhergesagter microRNA-Bindungen. Durch IGF2BP1-Knockdown in ES-2-Zellen in ihrer Expression reduzierte proteinkodierende Gene, für die in HEK-293-Zellen PAR-CLIP-Bindestellen für IGF2BP-Proteine gefunden wurden und welche putative Ziele für in ES-2-Zellen stark exprimierte ($RPM \geq 100$) microRNAs darstellen. \log_2FC : Zur Basis zwei logarithmierte Quotienten der mittleren Expressionswerte von Kontroll- und IGF2BP1-Knockdown-Proben. FDR: False Discovery Rate; Für multiples Testen auf differentielle Expression korrigierte P-Werte. Grau untermalte Gensymbole kennzeichnen Gene, welche an signifikant angereicherten biologischen Prozessen (Gene Ontology) beteiligt sind, deren Bezeichnungen die Begriffe „migration“, „adhesion“ oder „extracellular matrix“ enthalten.

4 Diskussion

Applikationen zur automatisierten Quantifizierung von Zellmigration Im Rahmen der vorliegenden Arbeit wurden zwei Methoden entwickelt, mit denen *in-vitro*-Zellmigrationsexperimente automatisiert ausgewertet werden können. Beide Methoden wurden mit dem Ziel der einfachen Benutzbarkeit entwickelt. Sie wurden als Bestandteil des MiToBo-Projektes [83] implementiert, wodurch sie kostenlos und für eine Vielzahl von Computerplattformen verfügbar sind.

Mit dem CellMigrationAnalyzer kann die Migration fluoreszenzmarkierter Zellen in 2D-Einzelzell-Tracking-Experimenten quantifiziert werden. Dabei wird für die detektierten Zellen eine Vielzahl migrations- und formbezogener Merkmale berechnet. Vergleiche mit manuell erhobenen Referenzdaten verschiedener Zelltypen zeigten eine große Übereinstimmung der ermittelten Werte. Der CellMigrationAnalyzer wurde verwendet, um eine Vielzahl unterschiedlicher Zelllinien zu untersuchen. Die getesteten Zellen exprimierten entweder fluoreszierende Proteine oder waren mit fluoreszierenden Farbstoffen eingefärbt worden. Durch die Ermittlung von Migrationsparametern, wie der Gerichtetheit oder Migrationsrichtung, ist die Applikation auch geeignet, Chemotaxis einzelner Zellen zu untersuchen. Die Ausgabe formbezogener Merkmale erlaubt zudem die Untersuchung von Formveränderungen. Dadurch ist es möglich, Veränderungen der Gestalt mit dem Migrationsverhalten zu korrelieren. Der Vergleich mit Ergebnissen anderer aktueller Programme, welche für die automatisierte Auswertung von Einzelzellmigration entwickelt wurden, zeigte, dass der CellMigrationAnalyzer ähnlich gute Ergebnisse produzieren kann, wie die besten der getesteten Programme. Diese Ergebnisse zeigten jedoch auch, dass die Qualität der Ergebnisse des CellMigrationAnalyzer stark von der Auswahl geeigneter Programmparameter abhängt. Die Anzahl der Parameter, welche einen Einfluss auf die Segmentierungs- und Zuordnungsergebnisse ausüben ist jedoch geringer als die der anderen Programme. Dies vereinfacht die Bedienung der Applikation. Eine Möglichkeit, aneinandergrenzende oder überlappende Zellen mit komplexer Morphologie zuverlässig als separate Zellen zu segmentieren, konnte nicht gefunden werden. Jedoch können solche Ereignisse durch Analyse der Veränderungen detektierter Zellflächen erkannt werden. Detektierte Objekte, welche demnach mehr als eine Zelle darstellen, werden von der Auswertung ausgeschlossen. Dies kann jedoch dazu führen, dass für dieselben Zellen unter Umständen mehrere Trajektorien erstellt werden. Deshalb sollte die Zelldichte in Videos, welche mit dem CellMigrationAnalyzer ausgewertet werden, nicht zu hoch sein. Die in Abschnitt 2.2 beschriebenen Werte stellen eine Orientierung für Zellzahlen dar, mit denen gute Ergebnisse erzielt

wurden. Weiterhin sollten Zelltypen vermieden werden, welche dazu neigen, auch unter geringer Zelldichte Verbände zu bilden.

Zusammenfassend zeichnet sich der CellMigrationAnalyzer durch verlässliche Ergebnisse, sowie vielfältige Anwendungsmöglichkeiten und eine einfache Bedienbarkeit, als geeignetes Werkzeug für die automatisierte Auswertung von 2D-Einzelzelltracking-Experimenten mit fluoreszenzmarkierten Zellen aus. Die verwendeten Algorithmen lassen sich mit geringem Aufwand dahingehend modifizieren, dass 3D-Bilddaten fluoreszenzmarkierter Zellen bearbeitet werden können. Dadurch wäre es möglich 3D-Einzelzelltracking-Experimente auszuwerten.

Ein Vorteil der Fluoreszenzmikroskopie im Vergleich zu Durchlichtmikroskopie ist der höhere Kontrast der fluoreszierenden Objekte, welcher die Detektion dieser Objekte erleichtert. Zudem ist es möglich, durch gezielte Anregung oder Detektion von Fluoreszenz in definierten Fokusebenen, hochauflösende 3D-Rekonstruktionen der fluoreszierenden Objekte zu erstellen. Dies wird zum Beispiel in der Konfokal- oder Multi-Photon-Mikroskopie ausgenutzt. Für die Erstellung fluoreszenzmikroskopischer Aufnahmen lebender Zellen, werden diese mit fluoreszierenden Farbstoffen gefärbt oder gentechnisch dahingehend verändert, dass sie fluoreszierende Proteine exprimieren. Dies kann jedoch zu unvorhersagbaren Veränderungen im Stoffwechsel der Zellen führen, welche sich wiederum auf beobachtbare Eigenschaften, wie beispielsweise das migratorische Verhalten der Zellen, auswirken können. Hinzu kommen mögliche phototoxische Effekte, welche durch die Bestrahlung der Zellen mit intensiven Lichtquellen, zum Zweck der Fluoreszenzanregung, induziert werden können. Diese Probleme können durch die Verwendung von Durchlichtmikroskopie vermieden werden. Die Detektion von Zellen in Durchlichtmikroskopaufnahmen ist jedoch auch unter Verwendung kontraststeigernder Methoden, wie beispielsweise Phasenkontrastmikroskopie oder Differential-Interferenz-Kontrast, aufwändiger als in Fluoreszenzmikroskopaufnahmen. Dennoch wurden bereits Verfahren entwickelt, welche einzelne Zellen erfolgreich in Durchlichtmikroskopaufnahmen detektieren können, beispielsweise [113]. Der CellMigrationAnalyzer würde von einer zusätzlichen Option zur zuverlässigen Segmentierung von Zellen in Durchlichtmikroskopaufnahmen profitieren, da die Anwendbarkeit der Applikation dadurch größer würde. Zudem könnte damit ein Einfluss verschiedener Fluoreszenzmarkierungen im Vergleich zu unmarkierten Zellen automatisiert untersucht werden.

Die zweite für diese Arbeit entwickelte Applikation, der ScratchAssayAnalyzer, dient der automatisierten Auswertung von Scratch- und Gap-Closure-Assays. Diese Experimentarten stellen eine günstige und einfach durchführbare Möglichkeit dar, das Migrationspotential kollektiv migrierender Zellen zu untersuchen [72]. Der ScratchAssayAnalyzer kann die Wundflächen sowohl

in Durchlichtaufnahmen, als auch in fluoreszenzmikroskopischen Aufnahmen detektieren. Vergleiche mit manuell segmentierten Referenzwunden ergaben hohe Übereinstimmungen mit den Ergebnissen des ScratchAssayAnalyzer. Diese Übereinstimmungen waren für Bilder fluoreszenzmarkierter Zellen größer als für Durchlichtbilder. Dennoch zeigten Vergleiche mit anderen Programmen zur automatisierten Quantifizierung von Wundflächen in Scratch-Assay-Bildern, dass mit dem ScratchAssayAnalyzer unterschiedliche Arten von Durchlichtaufnahmen zuverlässiger ausgewertet werden können. Es stellte sich heraus, dass die beiden für die Vergleiche verwendeten Programme, CellProfiler [67] und TScratch [39], höhere Anforderungen an den Kontrast der zu bearbeitenden Bilder stellen. Darüberhinaus erfordert die Bedienung dieser Programme das Einstellen einer weitaus höheren Parameteranzahl. Die SVM-basierte Klassifizierung erwies sich als geeignetes Mittel zur Identifizierung von Bildern mit geschlossenen Wundflächen. Dabei stellte sich heraus, dass auch die Klassifikationsrate auf Fluoreszenzmikroskopiebildern etwas höher war.

Dies zeigt, dass der ScratchAssayAnalyzer ein geeignetes Werkzeug für die quantitative Analyse von Scratch- beziehungsweise Gap-Closure-Assays darstellt. Die Tatsache, dass die Applikation bereits von verschiedenen Arbeitsgruppen unterschiedlicher Universitäten verwendet wurde, um Scratch-Assay-Experimente zu analysieren, unterstützt diese Aussage. Beispiele finden sich in [122], [50] und [62].

Migrationspotential von Zellen nach IGF2BP1-Knockdown CellMigrationAnalyzer und ScratchAssayAnalyzer wurden verwendet, um das Migrationsverhalten tumorabgeleiteter Zellen, in denen die Expression des onkofötal exprimierten RNA-bindenden Proteins IGF2BP1 reduziert wurde, zu untersuchen.

In Einzelzell-Tracking-Experimenten konnte für osteosarkomabgeleitete U-2OS-Zellen gezeigt werden, dass der transiente Knockdown von IGF2BP1 mit zwei verschiedenen siRNAs die Geschwindigkeit der Zellen in beiden Fällen signifikant reduzierte. Während jedoch der Knockdown mit der siRNA, welche die Geschwindigkeit der Zellen deutlicher verringert hatte, auch zu einer signifikanten Reduktion der Gerichtetheit führte, war nach dem Knockdown mit der zweiten siRNA keine signifikante Veränderung der Gerichtetheit beobachtet worden. Diese Diskrepanz ist möglicherweise auf Kreuzreaktionen einer oder beider siRNAs mit weiteren Transkripten zurückzuführen. Um die Wirkung solcher Kreuzreaktionen zu vermindern, wurden Knockdown-Experimente in ovarialkarzinomabgeleiteten ES-2-Zellen mit einer Kombination aus 12 verschiedenen, gegen IGF2BP1-gerichteten, siRNAs durchgeführt. Dies führte zu einer signifikant reduzierten Geschwindigkeit im Vergleich zu Kontrollzellen, jedoch wurde

kein signifikanter Unterschied in der Gerichtetheit festgestellt.

Ein weiterer Versuch zur Untersuchung der Wirkung einer Reduktion der endogenen IGF2BP1-Menge auf die Geschwindigkeit von ES-2-Zellen wurde in [11] durchgeführt. Dort wurde der CellMigrationAnalyzer verwendet um zu untersuchen, wie sich der siRNA-vermittelte Knockdown von IGF2BP1 auf die Migrationsgeschwindigkeit von ES-2-Zellen, in denen die Aktivität von microRNAs der let-7-Familie reduziert worden war, auswirkt. Die Versuche wurden sowohl mit Zellen, welche auf kollagenbeschichteten Oberflächen ausplattiert wurden, als auch mit Zellen, welche in 3D-Kollagenmatrizen eingebettet wurden, durchgeführt. Für die Auswertung der Zellen in der 3D-Matrix wurden Maximalprojektionen der aufgenommenen 3D-Volumen verwendet, das heißt, die axialen Komponenten der Zellbewegungen wurden nicht betrachtet. Die Reduktion der let-7-Aktivität führte sowohl in den 2D-, als auch in der 3D-Umgebungen zu einer signifikanten Erhöhung der Geschwindigkeit, welche durch IGF2BP1-Knockdown jedoch wieder deutlich reduziert werden konnte.

In [109] führte die Überexpression eines GFP-fusionierten IGF2BP1-Orthologen aus dem Huhn (GFP-chZBP1) in U-2OS-Zellen zu einer signifikanten Erhöhung der mittleren Geschwindigkeit und Gerichtetheit im Vergleich zu GFP-exprimierenden Zellen. Durch siRNA-vermittelten GFP-chZBP1/IGF2BP1-Knockdown konnte anschließend die mittlere Geschwindigkeit der GFP-chZBP1-überexprimierenden Zellen soweit reduziert werden, dass sie etwa dem Wert der GFP-exprimierenden Zellen entsprach. Dieser Knockdown führte jedoch zu keiner signifikanten Reduktion der Gerichtetheit der U-2OS-Zellen.

Auch in [69] führte IGF2BP1-Überexpression *in vitro* zu höheren Geschwindigkeiten, sowie erhöhter Gerichtetheit. Die dort verwendeten mammarkarzinomabgeleiteten MTLn3-Zellen zeigten nach IGF2BP1-Überexpression zudem eine verringerte Fähigkeit sich einem Lockstoffgradienten (EGF) zuzuwenden. Das verringerte chemotaktische Potential der Zellen wurde einer durch IGF2BP1-vermittelten erhöhten Polarität zugeschrieben, welche die Fähigkeit der Zellen, auf chemotaktische Signale zu reagieren, verminderte. Allerdings wurden für dieses Chemotaxisexperiment nur sehr wenige Zellen betrachtet, so dass die Zuverlässigkeit dieser Schlussfolgerung bezweifelt werden kann. In einem weiteren Versuch dieser Studie wurden GFP- und GFP-IGF2BP1-überexprimierende MTLn3-Zellen in die Brustfettpolster von Ratten injiziert und mittels Zwei-Photon-Mikroskopie beobachtet. Dabei zeigte sich, dass die GFP-IGF2BP1-überexprimierenden Zellen, im Gegensatz zu den GFP-exprimierenden Zellen, nicht in Richtung der nächstgelegenen Blutgefäße orientiert waren. Zudem zeigten sie eine geringere Motilität. Diese Werte wurden jedoch lediglich durch das Zählen der beobachteten optischen Felder, in denen Zellbewegungen gesehen wurden, gewonnen. Das reduzierte

chemotaktische Potential wurde als mögliche Ursache für eine metastasierungshemmende Wirkung von IGF2BP1 vorgeschlagen. Auf eine antimetastatische Wirkung von IGF2BP1 kann auch durch die Resultate aus [45] geschlossen werden. Dort wurde gezeigt, dass IGF2BP1-Expression in menschlichen Brustkrebsmetastasen im Vergleich zu den entsprechenden Primärtumoren deutlich reduziert war. Weiterhin war die Expression von IGF2BP1 in metastatischen brustkrebsabgeleiteten Zelllinien verschiedener Säugetierspezies im Vergleich zu nicht-metastatischen Zelllinien reduziert. ShRNA-vermittelter IGF2BP1-Knockdown in der nicht-metastatischen menschlichen Zelllinie T47D führte zu einem deutlichen Anstieg der Zellanzahl, welche in einem Transwell-Invasion-Assay mit Heregulin-Gradienten die Filtermembran passierten. Daraus wurde auf eine suppressive Wirkung von IGF2BP1 auf das Migrationspotential von Brustkrebszellen geschlossen. Allerdings kann Zellmigration mit Transwell-Invasion-Assays nicht direkt gemessen werden. Eine verringerte Fähigkeit der IGF2BP1-exprimierenden Zellen auf chemotaktische Signale zu reagieren, wie in [69] vorgeschlagen, könnte die nach IGF2BP1-Knockdown erhöhte Anzahl beobachteter Zellen ebenfalls erklären. Den Beobachtungen metastasierungshemmender Wirkung von IGF2BP1 stehen jedoch Erkenntnisse aus [110] und [21] gegenüber. In diesen Arbeiten wurde gezeigt, dass Tumore mit hoher IGF2BP1-Expression zu Metastasenbildung neigten.

Die für die vorliegende Arbeit ausgewerteten Scratch-Assay-Experimente bestätigten ein vermindertes Migrationspotential von Zellen mit reduzierter IGF2BP1-Expression. Der siRNA-vermittelte Knockdown in GFP-exprimierenden U-2OS-Zellen führte, wie auch in [109], zu einem verzögerten Wundschluss im Vergleich zu Kontrollzellen. Zusätzlich zu Knockdown-Experimenten mit einzelnen siRNAs, wurde für die Experimente der vorliegenden Arbeit ein IGF2BP1-Knockdown mit 12 verschiedenen siRNAs durchgeführt. Dieser führte, im Vergleich zu IGF2BP1-Knockdowns mit einzelnen siRNAs, zu einer geringeren, aber dennoch signifikanten Verzögerung des Wundschlusses.

In [127] konnte mit Hilfe des ScratchAssayAnalyzer gezeigt werden, dass IGF2BP1-Knockdown mittels lentiviral transduzierten shRNAs in melanomabgeleiteten HT-144-Zellen ebenfalls eine Reduktion des Wundschlusses bewirkt.

Zusammengefasst zeigen diese Ergebnisse, dass die Reduktion der endogenen IGF2BP1-Expression zu einem verminderten Migrationspotential in verschiedenen tumorabgeleiteten Zellen führte. Dies äußerte sich in reduzierter Geschwindigkeit von Zellen in Einzelzell-Tracking-Experimenten, sowie verzögertem Wundschluss in Scratch-Assay-Experimenten. Diese Ergebnisse bestätigen Beobachtungen, dass IGF2BP1-Expression das Migrationspotential von Tumorzellen *in vitro* erhöht.

Um ein vollständigeres Bild der Wirkung von IGF2BP1 auf die Migration von Tumorzellen zu erhalten, könnten Einzelzell-Tracking- und Scratch-/Gap-Closure-Experimente mit einer Vielzahl verschiedener Tumorzelllinien durchgeführt werden. Chemotaxis-Assays könnten zudem helfen, die kontroversen Erkenntnisse bezüglich der Wirkung von IGF2BP1 auf das chemotaktische Potential von Tumorzellen, zu verstehen. Mit Hilfe der für die vorliegende Arbeit entwickelten Applikationen zur automatisierten Analyse solcher Experimente, könnten diese in kurzer Zeit ausgewertet werden. Dabei sollten sowohl IGF2BP1-Überexpressions-, als auch -Knockdown-Experimente in Betracht gezogen werden. Aussagen über Gerichtetheit der Zellen sollten zudem in Zusammenhang mit Expressionswerten des Proteins PTEN betrachtet werden, da für dieses ein IGF2BP1-abhängiger Einfluss auf die gerichtete Zellmigration gezeigt wurde [109].

Auf zweidimensionalen Oberflächen ausgesäte Zellen zeigen, im Vergleich zu Zellen, welche sich in dreidimensionalen Gelen befinden, deutliche Unterschiede in Bezug auf Form, sowie Organisation des Zytoskeletts und Aufbau von Zell-Matrix-Kontakten [34, 27]. Die Einbettung von Tumorzellen in Gele, deren Zusammensetzung einer extrazellulären Matrix ähnelt, würde die natürliche Umgebung der Zellen besser simulieren. Dadurch könnte ein natürlicheres Verhalten der Zellen stimuliert werden. In Kombination mit geeigneten Aufnahmetechniken und Algorithmen, welche für die Auswertung der Zellmigration in 3D-Umgebungen angepasst sind, könnten so noch realistischere Daten aus *in-vitro*-Zellmigrationsexperimenten gewonnen werden.

Um die Relevanz der mittels *in-vitro*-Experimenten gewonnenen Erkenntnisse für pathophysiologische Vorgänge zu überprüfen, müssen zusätzlich *in-vivo*-Experimente durchgeführt werden. Hierfür könnten transgene Mausmodelle verwendet werden. Die Tiere könnten beispielsweise so verändert werden, dass sie ubiquitär oder in spezifischen Geweben ein Fusionsprotein, bestehend aus IGF2BP1 und einem fluoreszierenden Protein, exprimieren. Die Entstehung von Primärtumoren und Metastasen könnte dann mit Bildaufnahmesystemen, welche Fluoreszenz innerhalb lebender Tiere anregen und detektieren können (Small Animal Imaging), untersucht werden. Die Auswirkungen von Xenotransplantationen, bei denen IGF2BP1-exprimierende Zellen in Versuchstiere implantiert werden, könnten ebenfalls auf diese Weise analysiert werden.

Expression migrationsrelevanter Gene Um den Einfluss von IGF2BP1 auf die Expression migrationsrelevanter Gene zu untersuchen, wurden Total-RNA-Seq-, sowie microRNA-Seq-Experimente mit ES-2-Zellen durchgeführt, in denen die endogene IGF2BP1-Expression mit-

tels siRNA-vermitteltem Knockdown reduziert worden war. Der IGF2BP1-Knockdown wurde mit der gleichen Kombination aus verschiedenen siRNAs durchgeführt, mit der auch die ES-2-Zellen für die Einzelzell-Tracking-Experimente transfiziert worden waren. Obwohl vorangegangene PCR-basierte Tests auf Mykoplasma kontamination kein positives Ergebnis erzielten, konnte in allen Proben eine hohe Anzahl Reads Mykoplasmaengenomen zugeordnet werden. Als Konsequenz konnte eine Infektion der Zellen mit diesen parasitären Bakterien nicht ausgeschlossen werden. Die Ursache für die Diskrepanz der PCR- und RNA-Seq-basierten Detektionsmethode war nicht bekannt. Dies sollte jedoch vor der Durchführung zukünftiger Experimente dieser Art genauer untersucht werden. Genexpressionsanalysen von chronisch mit *Mycoplasma hominis* infizierten HeLa-Zellen zeigten eine veränderte Genexpression von mehr als 1700 Genen im Vergleich zu nichtinfizierten Zellen [53]. Da gemäß RNA-Seq-basiertem Test sowohl Knockdown- als auch Kontrollzellen die Infektion zeigten, sollte die differentielle Expression von Genen dennoch überwiegend durch den IGF2BP1-Knockdown verursacht worden sein. Jedoch ist nicht auszuschließen, dass das Ausmaß der differentiellen Expression einzelner Gene durch eine Infektion mit Mykoplasmen beeinflusst wurde.

Bereits bekannte Ursachen für einen Einfluss von IGF2BP1 auf die Zellmigration beinhalten die Umgestaltung des Zytoskeletts durch die Beeinflussung von Faktoren, welche den Auf- und Abbau von Aktinfasern koordinieren [29, 54, 109, 75]. Des Weiteren ist ein Einfluss von IGF2BP1 auf Zelladhäsion berichtet worden. In [117] führte der siRNA-vermittelte gleichzeitige Knockdown von IGF2BP1 und IGF2BP3 in HeLa-Zellen zu geringerer Adhäsion und kleineren Zellflächen auf mit Laminin-1 beschichteten Oberflächen. In der vorliegenden Arbeit wurde nach siRNA-vermitteltem Knockdown von IGF2BP1 in U-2OS-Zellen keine signifikante Veränderung der Zellflächen festgestellt. Jedoch führte der IGF2BP1-Knockdown in ES-2-Zellen zu einer deutlichen Zunahme der mittleren Zellflächen. Die größeren Flächen deuten dabei auf eine verstärkte Adhäsion der Zellen an die verwendete Collagen 1-Beschichtung. Eine Zunahme der Zellflächen nach IGF2BP1-Knockdown wurde in [127] für HEK293-Zellen ebenfalls beobachtet. Zusätzlich durchgeführte durchflusszytometrische Analysen dieser Zellen zeigten, dass die Volumina der Zellen unverändert waren. Die größeren Flächen resultierten daher aus einer größeren Ausbreitung der Zellen, welche, zumindest partiell, durch eine stärkere Adhäsion verursacht wurde. Der zusätzliche IGF2BP3-Knockdown, sowie die verwendete Laminin-1-Beschichtung, könnten Ursachen für die in [117] berichteten gegenteiligen Effekte auf Adhäsion sein. Weitere Experimente zur Adhäsion von verschiedenen Zellmodellen unter separaten IGF2BP1- und IGF2BP3-Knockdown-Bedingungen, sowie mit verschiedenen Beschichtungsproteinen, könnten helfen die gegensätzlichen Resultate aufzuklären.

Genanreicherungsanalysen der differentiell exprimierten Gene des in der vorliegenden Arbeit beschriebenen Total-RNA-Seq-Experimentes legen nahe, dass IGF2BP1 unter anderem die Expression von Genen beeinflusst, welche an Zellmigration im Allgemeinen, aber auch an Zelladhäsion und an der Umgestaltung der extrazellulären Matrix beteiligt sind. Beispielsweise wurden Gene, deren Produkte Bestandteile der extrazellulären Matrix spalten können, differentiell exprimiert. Zu diesen gehörten Mitglieder der MMP- und ADAMTS-Proteinfamilie. Aber auch die Expression von Genen deren Produkte selbst Bestandteil der extrazellulären Matrix sein können, war nach IGF2BP1-Knockdown verändert. Dazu zählten verschiedene Kollagenene, sowie das Gen für Fibronectin (FN1). Die Fähigkeit zur Umgestaltung der sie umgebenden extrazellulären Matrix ist Voraussetzung dafür, dass Tumorzellen in fremde Gewebe eindringen können. Die Expression der Kollagenase MMP1 war unter IGF2BP1-Knockdown reduziert. Zudem konnte beobachtet werden, dass in Collagen 1-Gele eingebettete ES-2-Zellen, in denen die Expression von IGF2BP1 reduziert worden war, ein reduziertes Migrationspotential aufwiesen [11]. Dies legt die Vermutung nahe, dass eine Reduktion der IGF2BP1-Expression die Invasivität dieser Zellen reduzieren kann. Um den Einfluss von IGF2BP1 auf Tumorzellinvasivität genauer zu betrachten, sollten Experimente dazu an weiteren Zelllinien durchgeführt werden.

Die beobachteten Phänotypen, sowie Anreicherungsanalysen der unter IGF2BP1-Knockdown differentiell exprimierten Gene, deuten zudem auf eine Beteiligung von IGF2BP1 am Prozess der epithelial-zu-mesenchymalen Transition (EMT) hin. Die mittels der Migrationsanalysen beobachtete Reduktion der Geschwindigkeiten verschiedener Zelllinien und die Tatsache, dass der mesenchymale Marker Fibronectin nach Reduktion der IGF2BP1-Proteinmenge in seiner Expression ebenfalls reduziert war, lässt vermuten, dass IGF2BP1 mesenchymale Eigenschaften der Zellen aufrechterhält oder auch verstärkt. Dies wurde bereits in [127] postuliert.

Insgesamt deuten die hier gezeigten, sowie der Großteil bereits publizierter Ergebnisse daraufhin, dass verstärkte IGF2BP1-Expression das Migrationspotential von Tumorzellen erhöht. Dies geschieht durch Beeinflussung verschiedener zellulärer Prozesse. Der koordinierte Auf- und Abbau von Aktinfasern an Lamellipodien, aber auch die Verankerung an die extrazelluläre Matrix, sowie deren aktive Umgestaltung wird von IGF2BP1 beeinflusst. Die Wirkung des Proteins auf die beiden letztgenannten Prozesse bedarf jedoch weiterer Analysen.

Neben Prozessen, welche die Migration von Zellen beeinflussen, zeigten die Genanreicherungsanalysen, dass auch die Expression von Genen, welche an diversen Entwicklungsprozessen beteiligt sind, durch IGF2BP1 beeinflusst wird. Dies steht in Einklang mit der Tatsache, dass IGF2BP1-Knockout zu Fehlentwicklungen in Mäusen führt [4].

Weitere RNA-Seq-Experimente können helfen, ein vollständigeres Bild der Auswirkungen von IGF2BP1 auf die Genexpression in Tumorzellen zu erhalten. Dabei sollten sowohl IGF2BP1-Knockdown-, als auch -Überexpressionsmodelle verwendet werden. Die Auswirkungen auf Transkripte, deren Translation oder Transport durch IGF2BP1 beeinflusst wird, kann mittels RNA-Seq-Experimenten jedoch nicht untersucht werden. Proteomanalysen, beispielsweise basierend auf massenspektrometrischen Methoden, können dabei helfen, einen Einblick auf die Wirkungen von IGF2BP1 auf die Proteinexpression zu erhalten. In Verbindung mit Ergebnissen aus RNA-Seq- und CLIP-Experimenten könnten weitere Transkripte identifiziert werden, deren Translation durch Bindung an IGF2BP1 modifiziert wird. Die Ergebnisse der Hochdurchsatzverfahren sollten zudem, zumindest stichprobenartig, mit Hilfe anderer Verfahren validiert werden.

Potentielle neue IGF2BP1-Zieltranskripte Um die Wirkung von IGF2BP1 im Detail besser zu verstehen, ist es sinnvoll Gene zu betrachten, auf welche das Protein eine direkte Wirkung ausübt. IGF2BP1 ist ein RNA-bindendes Protein und diverse Ziel-RNAs konnten bereits nachgewiesen werden. Darüber hinaus wurden mittels PAR-CLIP-Experimenten ungefähr 8000 putative Ziel-RNAs gefunden [46]. Da für diese Experimente die Menge des IGF2BP1-Proteins in den Zellen jedoch gesteigert wurde und sich ein großer Teil der identifizierten CLIP-Bindestellen in intronischen Bereichen von mRNAs befand, liegt die Vermutung nahe, dass viele der putativen Ziel-RNAs *in vivo* nicht von IGF2BP1 gebunden werden. In der vorliegenden Arbeit wurden die Ergebnisse von Total-RNA-Seq- und microRNA-Seq-Experimenten verwendet, um aus der Menge der Transkripte mit PAR-CLIP-Bindestellen, solche zu identifizieren, welche wahrscheinliche Kandidaten für IGF2BP1-Zieltranskripte darstellen. Bei der Mehrheit der validierten IGF2BP1-Zieltranskripte handelt es sich um mRNAs deren Stabilität durch Bindung an das Protein erhöht wird. Deshalb wurden von den putativen Ziel-Transkripten zunächst solche mRNAs herausgesucht, deren Expression durch IGF2BP1-Knockdown reduziert wurde. Für einige validierte IGF2BP1-Zieltranskripte wurde gezeigt, dass die Bindung dieser Transkripte an IGF2BP1, deren microRNA-vermittelten Abbau inhibiert [26, 43, 11]. Deshalb wurden *in silico*-Vorhersagen zu microRNA-mRNA-Bindungen verwendet, um die Anzahl putativer IGF2BP1-Zieltranskripte auf solche Transkripte einzugrenzen, welche wahrscheinliche Ziele von microRNAs darstellen. Für die Vorhersagen wurden microRNAs verwendet, welche gemäß microRNA-Seq in ES-2-Zellen unter IGF2BP1-Knockdown am höchsten, aber nicht differentiell, exprimiert waren. Für mehr als 80 % der unter IGF2BP1-Knockdown in ES-2-Zellen herunterregulierten proteinkodierenden

Gene konnten putativ bindende microRNAs gefunden werden. Insgesamt wurden 124 Transkripte identifiziert, deren Expression unter IGF2BP1-Knockdown in ES-2-Zellen herunterreguliert war und welche zusätzlich PAR-CLIP-Bindestellen aufwiesen und als Ziele von microRNAs vorhergesagt wurden. Abbildung 54 zeigt eine Übersicht über die Zahlen der, gemäß Total-RNA-Seq, differentiell exprimierten proteinkodierenden Gene, sowie deren Unterteilung nach dem Vorhandensein von PAR-CLIP-Bindestellen und ob sie putative Bindepartner der in ES-2-Zellen am stärksten exprimierten microRNAs darstellen.

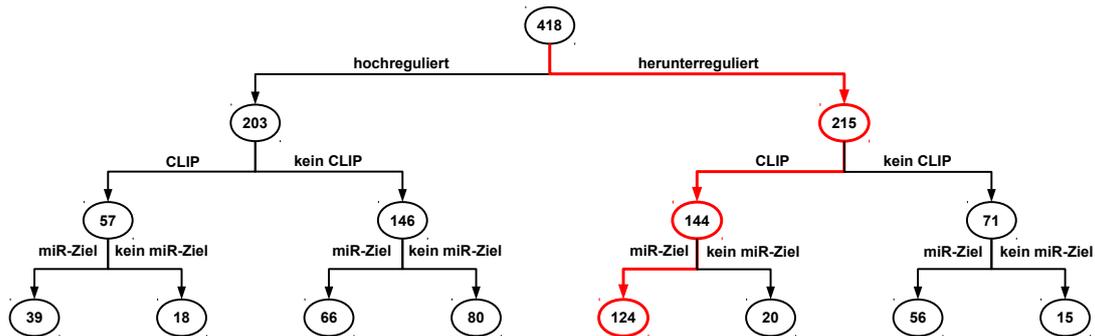


Abbildung 54: Aufteilung der 418 differentiell exprimierten proteinkodierenden Gene des Total-RNA-Seq-Datensatzes in Bezug auf (i) ihre Expression unter IGF2BP1-Knockdown relativ zu den Kontrollen (hochreguliert/herunterreguliert); (ii) im PAR-CLIP-Datensatz gefundene Bindestellen für Proteine der IGF2BP-Familie (CLIP/kein CLIP); (iii) multiMir-Vorhersagen für mögliche Ziele der in ES-2-Zellen am stärksten exprimierten (mehr als 100 RPM im Durchschnitt) microRNAs (miR-Ziel/kein miR-Ziel). Der rot markierte Pfad kennzeichnet die Auswahlkriterien für Transkripte, welche als putative IGF2BP1-Zieltranskripte betrachtet wurden.

Eine detaillierte Betrachtung der Funktionen, welche die so bestimmten putativen IGF2BP1-Zieltranskripte ausüben, beziehungsweise der regulatorischen Netzwerke, an welchen diese beteiligt sind, kann helfen die Rolle von IGF2BP1 in Krebserkrankungen besser zu verstehen. Jedoch muss die Bindung der so bestimmten putativen Zieltranskripte noch durch weitere Analysen validiert werden. Weiterhin sollte überprüft werden, ob die beobachtete Korrelation der Expression von IGF2BP1 und den putativen Zieltranskripten auch in anderen Zelllinien, sowie in primären Tumorentitäten beobachtet werden kann. Expressionsdaten aus öffentlich zugänglichen Datenbanken können helfen solche Korrelationen zu finden.

5 Zusammenfassung

Die Fähigkeit menschlicher Zellen sich aktiv fortzubewegen ist essentiell für die Entstehung von Geweben und Organen während der Embryogenese. Aber auch im adulten Organismus ist Zellmigration von fundamentaler Bedeutung. Sie wird unter anderem für Wundheilungsprozesse und Immunreaktionen benötigt. Jedoch ist Zellmigration auch an zahlreichen Krankheitsprozessen beteiligt. So spielt sie beispielsweise bei der Metastasierung im Verlauf der Progression maligner Tumorerkrankungen eine zentrale Rolle.

Im Rahmen der vorliegenden Arbeit wurden zwei Bildanalyseapplikationen für die automatisierte Auswertung von Zellmigrationsexperimenten entwickelt und als Teil einer Erweiterung für die populäre Bildanalyseplattform ImageJ implementiert. Diese Applikationen dienen der Quantifizierung der Zellmigration mit Hilfe von Mikroskopaufnahmen aus Einzelzell-Tracking-beziehungswise Scratch-Assay-Experimenten. Um die Ergebnisse der Applikationen zu evaluieren, wurden diese mit manuell ausgewerteten Daten, sowie den Ergebnissen ähnlicher Programme verglichen. Dabei zeigte sich, dass sich die hier vorgestellten Verfahren durch verlässliche Ergebnisse, Robustheit und besonders einfache Bedienbarkeit auszeichnen.

Die Applikationen wurden unter anderem dazu verwendet, die Wirkungen unterschiedlicher Proteinbeschichtungen und pharmakologischer Inhibitoren auf die 2D-Migration von Tumorzellen zu untersuchen. Zudem wurde ein Konzept für die Analyse von Chemotaxisexperimenten etabliert.

Weiterhin wurde mit Hilfe der Applikationen der Einfluss des onkofötal exprimierten RNA-bindenden Proteins IGF2BP1 auf die Tumorzellmigration untersucht. Hierfür wurde die Expression endogenen IGF2BP1-Proteins in tumorabgeleiteten Zellen mit Hilfe von siRNA-vermittelten transienten Knockdowns reduziert. Die Auswertung von Einzelzell-Tracking-Experimenten mit Zellen der aus einem Ovarialkarzinom abgeleiteten Zelllinie ES-2 ergab, dass eine verminderte Expression von IGF2BP1 zu einer reduzierten Geschwindigkeit der betrachteten Tumorzellen führte. Zudem zeigten die detektierten Zellen unter IGF2BP1-Knockdown eine signifikant vergrößerte Zellfläche. Dies deutet auf eine höhere Zell-Matrix-Adhäsion durch den IGF2BP1-Knockdown hin. Durch Analyse von Gap-Closure-Assays mit osteosarkomabgeleiteten Zellen der Linie U-2OS konnte zudem gezeigt werden, dass IGF2BP1-Knockdown zu einem verzögerten Wundschluss führt. Damit konnte bestätigt werden, dass eine verminderte Expression von IGF2BP1 das migratorische Potential von Tumorzellen reduziert.

Die Genexpression von Zellen der ovarialkarzinomabgeleiteten Zelllinie ES-2, für die zuvor gezeigt werden konnte, dass transienter IGF2BP1-Knockdown deren Migrationsgeschwindig-

keit reduziert, wurde mittels RNA-Seq-Daten ausgewertet. Die Analyse dieser Daten zeigte, dass IGF2BP1 einen Einfluss auf die Expression vieler migrationsrelevanter Gene ausübt. Insbesondere die Expression von Genen, deren Produkte an der zellulären Verankerung an die extrazelluläre Matrix, sowie an deren Aufbau und Umgestaltung beteiligt sind, wird von IGF2BP1 beeinflusst. Die beobachtete Geschwindigkeitsreduktion, sowie die Hinweise auf veränderte Adhäsion und Interaktion mit der extrazellulären Matrix, unterstützen Befunde, wonach IGF2BP1 am Prozess der epithelial-zu-mesenchymalen Transition (EMT) beteiligt ist. Der Vergleich mit Genen, welche nach EMT-Induktion eine veränderte Expression zeigten ergab, dass mehr von diesen Genen als zufällig zu erwarten wäre, auch unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimiert waren. Zudem befand sich unter diesen Genen das Gen FN1, welches für das Protein Fibronectin kodiert. Die Expression von Fibronectin gilt als ein Hinweis auf EMT und die Expression der entsprechenden mRNA war unter IGF2BP1-Knockdown reduziert. Dass IGF2BP1-Knockdown auch zu einer Reduktion der Fibronectin-Proteinmenge führt, wurde bereits für andere Zellsysteme beschrieben [127]. Die Stabilität einiger Zieltranskripte von IGF2BP1 wird durch die Bindung an das Protein erhöht, indem dadurch ein durch microRNA-Bindung vermittelter Abbau verhindert wird. *In-silico*-Analysen legten nahe, dass IGF2BP1 insbesondere mRNAs mit langen 3'-UTRs und diversen potentiellen microRNA-Zielmotiven bindet und diese reguliert. Dies wurde genutzt, um unter Verwendung von Total-RNA-Seq-, microRNA-Seq- und öffentlich verfügbaren PAR-CLIP-Daten potentielle neue IGF2BP1-Zieltranskripte zu identifizieren. Hierfür wurden solche proteinkodierenden Gene herausgesucht, deren Expression durch IGF2BP1-Knockdown reduziert wurde und welche sowohl gemäß PAR-CLIP-Daten von IGF2BP1 gebunden werden, als auch wahrscheinliche Ziele von in ES-2-Zellen abundanten microRNAs darstellen. Die Validität der so identifizierten putativen IGF2BP1-Zieltranskripte muss jedoch noch experimentell überprüft werden.

Durch die kombinierte Analyse von Mikroskopie-, sowie RNA-Seq-Daten, konnte somit eine Korrelation zwischen beobachtetem Migrationsverhalten (Phänotyp) und der IGF2BP1-vermittelten Regulation der Zielgenexpression (Genotyp) gezeigt werden. Die etablierten Verfahren stellen eine essentielle Grundlage für die funktionelle Charakterisierung der Rolle von IGF2BP1 und anderen Faktoren in der Tumorzellmigration dar.

Anhang

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000197632	SERPINB2	protein coding	-3.23	3.01E-032
ENSG00000184347	SLIT3	protein coding	2.83	5.65E-023
ENSG00000108091	CCDC6	protein coding	-3.11	3.15E-022
ENSG00000115677	HDLBP	protein coding	-2.19	1.34E-019
ENSG00000159217	IGF2BP1	protein coding	-2.03	1.57E-014
ENSG00000087245	MMP2	protein coding	2.03	2.60E-013
ENSG00000164283	ESM1	protein coding	-4.94	2.69E-013
ENSG00000166106	ADAMTS15	protein coding	3.53	3.47E-013
ENSG00000108821	COL1A1	protein coding	2.96	3.91E-013
ENSG00000135074	ADAM19	protein coding	1.83	1.21E-012
ENSG00000154096	THY1	protein coding	2.23	1.27E-012
ENSG00000107738	C10orf54	protein coding	4.17	4.15E-012
ENSG00000173706	HEG1	protein coding	2.43	6.07E-012
ENSG00000172379	ARNT2	protein coding	2.59	8.01E-012
ENSG00000113140	SPARC	protein coding	1.84	8.01E-012
ENSG00000158525	CPA5	protein coding	2.71	9.96E-012
ENSG00000091879	ANGPT2	protein coding	-3.87	3.63E-011
ENSG00000130635	COL5A1	protein coding	1.85	2.39E-010
ENSG00000143367	TUFT1	protein coding	2.00	5.20E-010
ENSG00000113721	PDGFRB	protein coding	3.02	3.89E-009
ENSG00000141526	SLC16A3	protein coding	2.26	5.23E-009
ENSG00000146966	DENND2A	protein coding	1.77	6.00E-009
ENSG00000090530	LEPREL1	protein coding	2.24	6.40E-009
ENSG00000115380	EFEMP1	protein coding	2.89	1.28E-008
ENSG00000137834	SMAD6	protein coding	2.36	1.36E-008
ENSG00000114812	VIPR1	protein coding	2.99	4.76E-008
ENSG00000175352	NRIP3	protein coding	-2.54	3.91E-007
ENSG00000137700	SLC37A4	protein coding	-1.82	3.91E-007
ENSG00000129038	LOXL1	protein coding	1.86	5.83E-007
ENSG00000128512	DOCK4	protein coding	-1.54	6.12E-007
ENSG00000165390	ANXA8	protein coding	2.56	8.83E-007
ENSG00000134030	CTIF	protein coding	1.71	1.00E-006
ENSG00000165388	ZNF488	protein coding	1.87	1.17E-006
ENSG00000133466	C1QTNF6	protein coding	1.66	1.78E-006
ENSG00000168487	BMP1	protein coding	1.88	2.16E-006
ENSG00000186567	CEACAM19	protein coding	2.28	3.00E-006
ENSG00000136261	BZW2	protein coding	-2.08	3.10E-006
ENSG00000186807	ANXA8L2	protein coding	2.59	3.34E-006
ENSG00000143819	EPHX1	protein coding	1.75	3.54E-006
ENSG00000103485	QPRT	protein coding	1.86	3.54E-006
ENSG00000149257	SERPINH1	protein coding	1.53	4.23E-006
ENSG00000213626	LBH	protein coding	3.82	4.24E-006
ENSG00000123352	SPATS2	protein coding	-1.47	4.24E-006
ENSG00000170421	KRT8	protein coding	1.70	4.24E-006
ENSG00000115159	GPD2	protein coding	-2.02	4.24E-006
ENSG00000163735	CXCL5	protein coding	-1.95	4.43E-006
ENSG00000185885	IFITM1	protein coding	2.72	4.83E-006
ENSG00000111305	GSG1	protein coding	2.76	5.05E-006
ENSG00000081923	ATP8B1	protein coding	2.76	5.05E-006
ENSG00000185022	MAFF	protein coding	-1.57	5.23E-006
ENSG00000173221	GLRX	protein coding	-1.58	5.85E-006
ENSG00000154127	UBASH3B	protein coding	-1.30	5.90E-006
ENSG00000148848	ADAM12	protein coding	1.23	6.20E-006
ENSG00000166741	NNMT	protein coding	1.74	7.42E-006
ENSG00000171951	SCG2	protein coding	2.97	7.98E-006
ENSG00000143641	GALNT2	protein coding	1.44	9.13E-006
ENSG00000122694	GLIPR2	protein coding	2.14	9.29E-006

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000136732	GYPC	protein coding	1.55	9.46E-006
ENSG00000137801	THBS1	protein coding	1.20	1.03E-005
ENSG00000157087	ATP2B2	protein coding	-1.65	1.54E-005
ENSG00000125726	CD70	protein coding	1.69	2.39E-005
ENSG00000155066	PROM2	protein coding	1.94	2.44E-005
ENSG00000204174	NPY4R	protein coding	2.87	2.62E-005
ENSG00000157557	ETS2	protein coding	-1.49	3.24E-005
ENSG00000115414	FN1	protein coding	-1.42	3.62E-005
ENSG00000070614	NDST1	protein coding	1.43	3.62E-005
ENSG00000142623	PADI1	protein coding	3.20	4.45E-005
ENSG00000167693	NXN	protein coding	1.35	4.53E-005
ENSG00000165832	TRUB1	protein coding	-2.47	4.94E-005
ENSG00000050165	DKK3	protein coding	1.13	4.94E-005
ENSG00000136158	SPRY2	protein coding	-1.28	4.94E-005
ENSG00000179776	CDH5	protein coding	1.18	5.05E-005
ENSG00000182809	CRIP2	protein coding	2.07	5.19E-005
ENSG00000011105	TSPAN9	protein coding	1.48	5.63E-005
ENSG00000169126	ARMC4	protein coding	-2.31	7.31E-005
ENSG00000187498	COL4A1	protein coding	1.38	8.01E-005
ENSG00000212724	KRTAP2-3	protein coding	-2.13	8.08E-005
ENSG00000108448	TRIM16L	protein coding	1.39	9.53E-005
ENSG00000082781	ITGB5	protein coding	1.36	9.53E-005
ENSG00000167107	ACSF2	protein coding	1.38	1.17E-004
ENSG00000100577	GSTZ1	protein coding	1.38	1.34E-004
ENSG00000147160	AWAT2	protein coding	1.98	1.38E-004
ENSG00000168874	ATOH8	protein coding	2.81	1.48E-004
ENSG000000071553	ATP6AP1	protein coding	1.67	1.48E-004
ENSG00000148541	FAM13C	protein coding	2.54	1.50E-004
ENSG00000106366	SERPINE1	protein coding	1.54	1.50E-004
ENSG00000111266	DUSP16	protein coding	-1.51	1.62E-004
ENSG00000166401	SERPINB8	protein coding	-1.52	1.62E-004
ENSG00000185483	ROR1	protein coding	2.02	1.64E-004
ENSG00000166147	FBN1	protein coding	1.51	1.67E-004
ENSG00000073008	PVR	protein coding	1.21	1.79E-004
ENSG00000143515	ATP8B2	protein coding	-1.16	1.91E-004
ENSG00000029725	RABEP1	protein coding	-2.09	1.93E-004
ENSG00000006756	ARSD	protein coding	1.49	2.16E-004
ENSG00000160255	ITGB2	protein coding	1.42	2.18E-004
ENSG00000108846	ABCC3	protein coding	2.04	2.20E-004
ENSG00000170485	NPAS2	protein coding	-1.50	2.20E-004
ENSG00000124762	CDKN1A	protein coding	-2.02	2.20E-004
ENSG00000180340	FZD2	protein coding	1.65	2.25E-004
ENSG00000144834	TAGLN3	protein coding	-1.80	3.19E-004
ENSG00000134259	NGF	protein coding	1.76	3.27E-004
ENSG00000114850	SSR3	protein coding	-1.23	3.27E-004
ENSG00000105971	CAV2	protein coding	1.45	3.57E-004
ENSG00000129474	AJUBA	protein coding	1.23	3.61E-004
ENSG00000166881	TMEM194A	protein coding	-1.54	3.64E-004
ENSG00000069020	MAST4	protein coding	-1.46	3.86E-004
ENSG00000197694	SPTAN1	protein coding	1.05	4.27E-004
ENSG00000172992	DCAKD	protein coding	1.64	4.39E-004
ENSG00000166897	ELFN2	protein coding	1.46	4.41E-004
ENSG00000123892	RAB38	protein coding	-3.40	4.41E-004
ENSG00000116299	KIAA1324	protein coding	2.31	4.68E-004
ENSG00000112584	FAM120B	protein coding	-1.55	4.68E-004
ENSG00000115758	ODC1	protein coding	-1.26	4.68E-004
ENSG00000064666	CNN2	protein coding	1.40	4.70E-004
ENSG00000100311	PDGFB	protein coding	1.61	5.76E-004
ENSG00000134013	LOXL2	protein coding	1.37	5.92E-004
ENSG00000157613	CREB3L1	protein coding	1.87	7.16E-004
ENSG00000144749	LRIG1	protein coding	1.26	7.94E-004
ENSG00000099282	TSPAN15	protein coding	1.42	8.24E-004

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000169129	AFAP1L2	protein coding	1.26	8.32E-004
ENSG00000103196	CRISPLD2	protein coding	1.68	8.39E-004
ENSG00000099998	GGT5	protein coding	1.76	9.18E-004
ENSG00000104332	SFRP1	protein coding	-0.99	9.84E-004
ENSG00000057252	SOAT1	protein coding	-1.63	9.98E-004
ENSG00000102048	ASB9	protein coding	1.71	1.03E-003
ENSG00000100664	EIF5	protein coding	-1.43	1.15E-003
ENSG00000221869	CEBPD	protein coding	1.98	1.15E-003
ENSG00000134871	COL4A2	protein coding	1.32	1.15E-003
ENSG00000147100	SLC16A2	protein coding	1.32	1.23E-003
ENSG00000011201	KAL1	protein coding	-1.87	1.23E-003
ENSG00000123453	SARDH	protein coding	1.62	1.24E-003
ENSG00000095303	PTGS1	protein coding	1.64	1.24E-003
ENSG00000119408	NEK6	protein coding	1.14	1.29E-003
ENSG00000145390	USP53	protein coding	-2.59	1.29E-003
ENSG00000183250	C21orf67	protein coding	1.21	1.33E-003
ENSG00000096063	SRPK1	protein coding	-1.45	1.33E-003
ENSG00000071575	TRIB2	protein coding	1.30	1.36E-003
ENSG00000095383	TBC1D2	protein coding	1.39	1.60E-003
ENSG00000225485	ARHGAP23	protein coding	1.29	1.60E-003
ENSG00000160207	HSF2BP	protein coding	2.28	1.62E-003
ENSG00000180861	C12orf36	protein coding	-8.06	1.66E-003
ENSG00000156475	PPP2R2B	protein coding	2.55	1.66E-003
ENSG00000126838	PZP	protein coding	4.09	1.66E-003
ENSG00000006534	ALDH3B1	protein coding	1.49	1.77E-003
ENSG00000171055	FEZ2	protein coding	-2.09	1.77E-003
ENSG00000180917	CMTR2	protein coding	-2.03	1.82E-003
ENSG00000067798	NAV3	protein coding	-1.26	1.83E-003
ENSG00000049192	ADAMTS6	protein coding	-1.81	1.97E-003
ENSG00000166848	TERF2IP	protein coding	-2.08	2.04E-003
ENSG00000064692	SNCAIP	protein coding	-2.19	2.25E-003
ENSG00000073711	PPP2R3A	protein coding	-1.87	2.27E-003
ENSG00000091986	CCDC80	protein coding	1.21	2.30E-003
ENSG00000103888	KIAA1199	protein coding	2.13	2.32E-003
ENSG00000162892	IL24	protein coding	-1.86	2.35E-003
ENSG00000136802	LRRC8A	protein coding	-1.15	2.35E-003
ENSG00000242265	PEG10	protein coding	1.22	2.69E-003
ENSG00000249992	TMEM158	protein coding	-1.20	2.74E-003
ENSG00000105926	MPP6	protein coding	-2.02	2.77E-003
ENSG00000221926	TRIM16	protein coding	1.36	2.84E-003
ENSG00000213417	KRTAP2-4	protein coding	-2.13	2.96E-003
ENSG00000171310	CHST11	protein coding	-1.14	3.04E-003
ENSG00000198894	CIPC	protein coding	-1.25	3.04E-003
ENSG00000163659	TIPARP	protein coding	-1.61	3.19E-003
ENSG00000070669	ASNS	protein coding	-1.97	3.20E-003
ENSG00000138166	DUSP5	protein coding	-1.03	3.24E-003
ENSG00000007923	DNAJC11	protein coding	-1.14	3.39E-003
ENSG00000139278	GLIPR1	protein coding	-2.13	3.48E-003
ENSG00000091409	ITGA6	protein coding	-1.49	3.52E-003
ENSG00000164933	SLC25A32	protein coding	-1.45	3.52E-003
ENSG00000169429	IL8	protein coding	-3.40	3.53E-003
ENSG00000204381	LAYN	protein coding	-1.16	3.65E-003
ENSG00000149294	NCAM1	protein coding	1.13	3.67E-003
ENSG00000128294	TPST2	protein coding	1.11	3.73E-003
ENSG00000144959	NCEH1	protein coding	-1.54	4.11E-003
ENSG00000163900	TMEM41A	protein coding	-1.49	4.17E-003
ENSG00000140873	ADAMTS18	protein coding	-2.05	4.21E-003
ENSG00000114120	SLC25A36	protein coding	-1.57	4.29E-003
ENSG00000182179	UBA7	protein coding	1.89	4.42E-003
ENSG00000057704	TMCC3	protein coding	-1.50	4.46E-003
ENSG00000182492	BGN	protein coding	2.21	4.53E-003
ENSG00000099957	P2RX6	protein coding	2.73	4.75E-003

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000151136	BTBD11	protein coding	-1.27	4.75E-003
ENSG00000106571	GLI3	protein coding	-1.43	4.80E-003
ENSG00000139722	VPS37B	protein coding	-1.18	4.88E-003
ENSG00000233276	GPX1	protein coding	1.41	4.97E-003
ENSG00000185567	AHNAK2	protein coding	1.01	5.14E-003
ENSG00000163297	ANTXR2	protein coding	-1.78	5.15E-003
ENSG00000149506	ZP1	protein coding	2.82	5.21E-003
ENSG00000184557	SOCS3	protein coding	1.31	5.37E-003
ENSG00000163545	NUAK2	protein coding	1.74	5.52E-003
ENSG00000142089	IFITM3	protein coding	1.08	5.52E-003
ENSG00000095794	CREM	protein coding	-1.49	5.52E-003
ENSG00000167889	MGAT5B	protein coding	1.33	5.62E-003
ENSG00000188848	BEND4	protein coding	-1.19	5.83E-003
ENSG00000131016	AKAP12	protein coding	-1.33	5.83E-003
ENSG00000171517	LPAR3	protein coding	-1.33	5.83E-003
ENSG00000173818	ENDOV	protein coding	1.32	5.92E-003
ENSG00000141150	RASL10B	protein coding	1.64	5.93E-003
ENSG00000143322	ABL2	protein coding	-1.20	5.93E-003
ENSG00000134602	MST4	protein coding	-1.80	5.93E-003
ENSG00000183853	KIRREL	protein coding	0.92	6.09E-003
ENSG00000059758	CDK17	protein coding	-1.52	6.53E-003
ENSG00000108861	DUSP3	protein coding	-1.01	6.53E-003
ENSG00000049130	KITLG	protein coding	-6.38	6.60E-003
ENSG00000165449	SLC16A9	protein coding	-2.01	6.66E-003
ENSG00000133321	RARRES3	protein coding	3.00	6.70E-003
ENSG00000075213	SEMA3A	protein coding	-2.45	6.74E-003
ENSG00000133433	GSTT2B	protein coding	1.67	6.77E-003
ENSG00000137193	PIM1	protein coding	1.05	6.77E-003
ENSG00000151012	SLC7A11	protein coding	-2.03	6.77E-003
ENSG00000196611	MMP1	protein coding	-2.41	6.82E-003
ENSG00000197471	SPN	protein coding	1.50	6.82E-003
ENSG00000033178	UBA6	protein coding	-2.03	7.14E-003
ENSG00000112893	MAN2A1	protein coding	-1.57	7.24E-003
ENSG00000164647	STEAP1	protein coding	-2.00	7.86E-003
ENSG00000057019	DCBLD2	protein coding	-1.32	7.94E-003
ENSG00000115419	GLS	protein coding	-2.55	8.03E-003
ENSG00000114573	ATP6V1A	protein coding	-1.64	8.08E-003
ENSG00000162441	LZIC	protein coding	-1.73	8.26E-003
ENSG00000104368	PLAT	protein coding	-0.95	8.98E-003
ENSG00000020181	GPR124	protein coding	1.44	8.98E-003
ENSG00000113356	POLR3G	protein coding	-1.45	9.00E-003
ENSG00000150165	ANXA8L1	protein coding	2.52	9.03E-003
ENSG00000125841	NRSN2	protein coding	1.10	9.20E-003
ENSG00000105968	H2AFV	protein coding	-1.48	9.33E-003
ENSG00000159164	SV2A	protein coding	1.32	9.33E-003
ENSG00000113070	HBEGF	protein coding	-1.39	9.39E-003
ENSG00000152492	CCDC50	protein coding	-1.35	9.89E-003
ENSG00000004399	PLXND1	protein coding	1.13	9.93E-003
ENSG00000100490	CDKL1	protein coding	-2.18	9.93E-003
ENSG00000105889	STEAP1B	protein coding	-1.45	1.00E-002
ENSG00000153879	CEBPG	protein coding	-1.83	1.00E-002
ENSG00000137809	ITGA11	protein coding	2.10	1.05E-002
ENSG00000178882	FAM101A	protein coding	1.31	1.05E-002
ENSG00000109133	TMEM33	protein coding	-1.08	1.10E-002
ENSG00000115539	PDCL3	protein coding	-2.03	1.11E-002
ENSG00000125347	IRF1	protein coding	1.14	1.12E-002
ENSG00000102760	RGCC	protein coding	-2.57	1.14E-002
ENSG00000134824	FADS2	protein coding	1.10	1.15E-002
ENSG00000228672	PROB1	protein coding	2.26	1.15E-002
ENSG00000197150	ABCB8	protein coding	-1.35	1.17E-002
ENSG00000166396	SERPINB7	protein coding	-2.55	1.20E-002
ENSG00000156395	SORCS3	protein coding	-1.73	1.26E-002

Tabelle A1: Differenziell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000110237	ARHGEF17	protein coding	1.20	1.29E-002
ENSG00000181061	HIGD1A	protein coding	-2.50	1.33E-002
ENSG00000138311	ZNF365	protein coding	-1.76	1.34E-002
ENSG00000033170	FUT8	protein coding	-1.33	1.34E-002
ENSG00000111845	PAK1IP1	protein coding	-1.78	1.34E-002
ENSG00000137876	RSL24D1	protein coding	-1.94	1.34E-002
ENSG00000099984	GSTT2	protein coding	1.62	1.34E-002
ENSG00000163734	CXCL3	protein coding	-1.68	1.34E-002
ENSG00000100632	ERH	protein coding	-1.42	1.35E-002
ENSG00000139438	FAM222A	protein coding	1.95	1.42E-002
ENSG00000160991	ORAI2	protein coding	1.08	1.42E-002
ENSG00000166037	CEP57	protein coding	-1.41	1.42E-002
ENSG00000135643	KCNMB4	protein coding	3.28	1.43E-002
ENSG00000063601	MTMR1	protein coding	-1.08	1.44E-002
ENSG00000161395	PGAP3	protein coding	1.47	1.46E-002
ENSG00000158125	XDH	protein coding	-1.29	1.46E-002
ENSG00000176236	C10orf111	protein coding	3.11	1.54E-002
ENSG00000198133	TMEM229B	protein coding	1.48	1.54E-002
ENSG00000123689	G0S2	protein coding	-1.04	1.57E-002
ENSG00000111371	SLC38A1	protein coding	-1.18	1.58E-002
ENSG00000093167	LRRFIP2	protein coding	-1.01	1.60E-002
ENSG00000141012	GALNS	protein coding	1.17	1.60E-002
ENSG00000185306	C12orf56	protein coding	-3.09	1.61E-002
ENSG00000013588	GPRC5A	protein coding	0.91	1.61E-002
ENSG00000064547	LPAR2	protein coding	1.96	1.63E-002
ENSG00000174606	ANGEL2	protein coding	-1.21	1.63E-002
ENSG00000136068	FLNB	protein coding	0.89	1.64E-002
ENSG00000019549	SNAI2	protein coding	1.17	1.64E-002
ENSG00000169504	CLIC4	protein coding	-1.33	1.68E-002
ENSG00000005175	RPAP3	protein coding	-2.51	1.68E-002
ENSG00000114686	MRPL3	protein coding	-1.79	1.75E-002
ENSG00000075239	ACAT1	protein coding	-2.00	1.77E-002
ENSG00000174059	CD34	protein coding	2.65	1.77E-002
ENSG00000174013	FBXO45	protein coding	-1.36	1.81E-002
ENSG00000146278	PNRC1	protein coding	1.21	1.82E-002
ENSG00000125657	TNFSF9	protein coding	1.48	1.84E-002
ENSG00000110619	CARS	protein coding	-0.92	1.90E-002
ENSG00000173535	TNFRSF10C	protein coding	1.98	1.93E-002
ENSG00000100997	ABHD12	protein coding	1.09	1.94E-002
ENSG00000144381	HSPD1	protein coding	-1.59	1.94E-002
ENSG00000176171	BNIP3	protein coding	0.97	1.95E-002
ENSG00000119616	FCF1	protein coding	-1.94	1.95E-002
ENSG00000068724	TTC7A	protein coding	1.05	1.95E-002
ENSG00000134516	DOCK2	protein coding	1.21	1.95E-002
ENSG00000120885	CLU	protein coding	0.99	1.98E-002
ENSG00000152332	UHMK1	protein coding	-1.28	1.99E-002
ENSG00000116489	CAPZA1	protein coding	-2.49	2.00E-002
ENSG00000203805	PPAPDC1A	protein coding	0.92	2.01E-002
ENSG00000147526	TACC1	protein coding	-1.13	2.03E-002
ENSG00000155254	MARVELD1	protein coding	0.99	2.06E-002
ENSG00000103647	CORO2B	protein coding	-1.23	2.06E-002
ENSG00000139192	TAPBPL	protein coding	1.32	2.07E-002
ENSG00000151835	SACS	protein coding	-1.56	2.07E-002
ENSG00000078018	MAP2	protein coding	1.49	2.10E-002
ENSG00000087116	ADAMTS2	protein coding	1.00	2.10E-002
ENSG00000128965	CHAC1	protein coding	-1.53	2.12E-002
ENSG00000107438	PDLIM1	protein coding	1.01	2.18E-002
ENSG00000185201	IFITM2	protein coding	1.13	2.18E-002
ENSG00000137040	RANBP6	protein coding	-2.49	2.18E-002
ENSG00000107643	MAPK8	protein coding	-1.72	2.18E-002
ENSG00000138768	USO1	protein coding	-1.94	2.20E-002
ENSG00000216588	IGSF23	protein coding	2.60	2.24E-002

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000174574	AKIRIN1	protein coding	-1.42	2.25E-002
ENSG00000181751	C5orf30	protein coding	-1.39	2.25E-002
ENSG00000212722	KRTAP4-9	protein coding	-3.68	2.25E-002
ENSG00000121644	DESI2	protein coding	-1.11	2.25E-002
ENSG00000064102	ASUN	protein coding	-1.43	2.27E-002
ENSG00000186318	BACE1	protein coding	0.95	2.31E-002
ENSG00000110011	DNAJC4	protein coding	1.04	2.34E-002
ENSG00000136518	ACTL6A	protein coding	-1.74	2.35E-002
ENSG00000180332	KCTD4	protein coding	-2.38	2.35E-002
ENSG00000105662	CRTC1	protein coding	1.27	2.35E-002
ENSG00000180884	ZNF792	protein coding	-1.21	2.35E-002
ENSG00000122952	ZWINT	protein coding	-0.94	2.35E-002
ENSG00000165156	ZHX1	protein coding	-2.52	2.36E-002
ENSG00000162694	EXTL2	protein coding	-1.51	2.37E-002
ENSG00000119977	TCTN3	protein coding	-1.28	2.37E-002
ENSG00000115946	PNO1	protein coding	-1.41	2.38E-002
ENSG00000123505	AMD1	protein coding	-1.54	2.40E-002
ENSG00000050130	JKAMP	protein coding	-2.69	2.40E-002
ENSG00000180537	RNF182	protein coding	-1.82	2.43E-002
ENSG00000011028	MRC2	protein coding	1.16	2.45E-002
ENSG00000198911	SREBF2	protein coding	1.08	2.48E-002
ENSG00000166250	CLMP	protein coding	-0.94	2.48E-002
ENSG00000156869	FRRS1	protein coding	-4.65	2.49E-002
ENSG00000153922	CHD1	protein coding	-2.19	2.52E-002
ENSG00000168077	SCARA3	protein coding	1.21	2.53E-002
ENSG00000112658	SRF	protein coding	-1.08	2.53E-002
ENSG00000115963	RND3	protein coding	-2.11	2.53E-002
ENSG00000130309	COLGALT1	protein coding	0.97	2.53E-002
ENSG00000115295	CLIP4	protein coding	-1.67	2.53E-002
ENSG00000118785	SPP1	protein coding	-4.32	2.63E-002
ENSG00000101187	SLCO4A1	protein coding	-1.23	2.64E-002
ENSG00000085117	CD82	protein coding	1.29	2.68E-002
ENSG00000063180	CA11	protein coding	1.23	2.74E-002
ENSG00000087111	PIGS	protein coding	0.95	2.75E-002
ENSG00000151287	TEX30	protein coding	-2.32	2.75E-002
ENSG00000204956	PCDHGA1	protein coding	2.37	2.76E-002
ENSG00000128989	ARPP19	protein coding	-1.34	2.76E-002
ENSG00000167601	AXL	protein coding	0.97	2.76E-002
ENSG00000173334	TRIB1	protein coding	-1.04	2.78E-002
ENSG00000148426	PROSER2	protein coding	-1.28	2.79E-002
ENSG00000159055	MIS18A	protein coding	-1.67	2.83E-002
ENSG00000124702	KLHDC3	protein coding	0.91	2.84E-002
ENSG00000183978	COA3	protein coding	-1.04	2.90E-002
ENSG00000197965	MPZL1	protein coding	-0.80	2.92E-002
ENSG00000095203	EPB41L4B	protein coding	-1.79	2.94E-002
ENSG00000105173	CCNE1	protein coding	-1.15	2.94E-002
ENSG00000106692	FKTN	protein coding	-2.18	3.01E-002
ENSG00000104635	SLC39A14	protein coding	-0.86	3.01E-002
ENSG00000188279	FAM25C	protein coding	3.97	3.01E-002
ENSG00000102316	MAGED2	protein coding	0.95	3.18E-002
ENSG00000115233	PSMD14	protein coding	-1.37	3.18E-002
ENSG00000155100	OTUD6B	protein coding	-2.64	3.18E-002
ENSG00000169251	NMD3	protein coding	-1.94	3.22E-002
ENSG00000153162	BMP6	protein coding	-1.91	3.32E-002
ENSG00000143127	ITGA10	protein coding	-1.64	3.33E-002
ENSG00000109107	ALDOC	protein coding	1.51	3.35E-002
ENSG00000213516	RBMXL1	protein coding	-1.49	3.37E-002
ENSG00000215033	AL603965.1	protein coding	2.04	3.41E-002
ENSG00000083720	OXCT1	protein coding	-2.37	3.45E-002
ENSG00000184117	NIPSNAP1	protein coding	0.86	3.49E-002
ENSG00000100030	MAPK1	protein coding	-1.12	3.49E-002
ENSG00000090238	YPEL3	protein coding	1.37	3.49E-002

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000074935	TUBE1	protein coding	-1.40	3.65E-002
ENSG00000077800	FKBP6	protein coding	2.48	3.79E-002
ENSG00000058729	RIOK2	protein coding	-2.63	3.79E-002
ENSG00000153006	SREK11P1	protein coding	-1.75	3.80E-002
ENSG00000204941	PSG5	protein coding	1.65	3.81E-002
ENSG00000179546	HTR1D	protein coding	2.69	3.84E-002
ENSG00000141510	TP53	protein coding	0.88	3.84E-002
ENSG00000105290	APLP1	protein coding	3.41	3.84E-002
ENSG00000123933	MXD4	protein coding	1.24	3.84E-002
ENSG00000184985	SORCS2	protein coding	1.64	3.89E-002
ENSG00000008323	PLEKHG6	protein coding	2.32	3.91E-002
ENSG00000082153	BZW1	protein coding	-1.41	3.94E-002
ENSG00000162496	DHRS3	protein coding	1.24	3.94E-002
ENSG00000198892	SHISA4	protein coding	1.27	3.98E-002
ENSG00000149196	C11orf73	protein coding	-1.40	3.98E-002
ENSG00000166908	PIP4K2C	protein coding	-1.11	3.98E-002
ENSG00000073849	ST6GAL1	protein coding	-1.10	4.00E-002
ENSG00000145220	LYAR	protein coding	-1.57	4.02E-002
ENSG00000147649	MTDH	protein coding	-1.51	4.02E-002
ENSG00000122126	OCRL	protein coding	-1.55	4.07E-002
ENSG00000096717	SIRT1	protein coding	-1.72	4.08E-002
ENSG00000136827	TOR1A	protein coding	-0.95	4.11E-002
ENSG00000168685	IL7R	protein coding	1.35	4.11E-002
ENSG00000104369	JPH1	protein coding	-1.47	4.11E-002
ENSG00000010278	CD9	protein coding	-1.26	4.11E-002
ENSG00000130508	PXDN	protein coding	1.00	4.12E-002
ENSG00000172380	GNG12	protein coding	-1.27	4.12E-002
ENSG00000117411	B4GALT2	protein coding	1.08	4.12E-002
ENSG00000244486	SCARF2	protein coding	1.68	4.12E-002
ENSG00000126458	RRAS	protein coding	1.32	4.12E-002
ENSG00000082212	ME2	protein coding	-1.88	4.12E-002
ENSG00000241973	PI4KA	protein coding	0.85	4.12E-002
ENSG00000054392	HHAT	protein coding	1.13	4.16E-002
ENSG00000091490	SEL1L3	protein coding	-1.96	4.29E-002
ENSG00000197265	GTF2E2	protein coding	-1.78	4.29E-002
ENSG00000168389	MFSD2A	protein coding	-1.22	4.33E-002
ENSG00000163946	FAM208A	protein coding	-1.28	4.36E-002
ENSG00000115368	WDR75	protein coding	-2.22	4.41E-002
ENSG00000115183	TANC1	protein coding	0.91	4.45E-002
ENSG00000138180	CEP55	protein coding	-1.87	4.53E-002
ENSG00000183527	PSMG1	protein coding	-2.08	4.53E-002
ENSG00000111412	C12orf49	protein coding	-1.14	4.68E-002
ENSG00000203668	CHML	protein coding	-2.13	4.70E-002
ENSG00000151465	CDC123	protein coding	-1.25	4.75E-002
ENSG00000100612	DHRS7	protein coding	-1.55	4.75E-002
ENSG00000073712	FERMT2	protein coding	-1.54	4.78E-002
ENSG00000151376	ME3	protein coding	1.89	4.80E-002
ENSG00000132429	POPDC3	protein coding	-2.05	4.89E-002
ENSG00000143333	RGS16	protein coding	-1.43	4.91E-002
ENSG00000163291	PAQR3	protein coding	-1.40	4.98E-002
ENSG00000172037	LAMB2	protein coding	0.95	4.98E-002
ENSG00000226044	RP11-463P17.1	lincRNA	3.89	8.99E-007
ENSG00000238266	LINC00707	lincRNA	-2.11	3.13E-004
ENSG00000223749	MIR503HG	lincRNA	1.73	4.04E-004
ENSG00000249743	RP11-60A8.1	lincRNA	-2.40	5.58E-004
ENSG00000233521	RP5-1172A22.1	lincRNA	1.60	6.33E-004
ENSG00000245522	RP11-540A21.2	lincRNA	3.00	6.85E-004
ENSG00000260604	RP1-140K8.5	lincRNA	3.11	1.74E-003
ENSG00000247095	MIR210HG	lincRNA	2.82	6.77E-003
ENSG00000223477	LINC00842	lincRNA	1.69	1.25E-002
ENSG00000229563	RP11-245M24.1	lincRNA	-1.95	1.37E-002
ENSG00000226380	MIR29A	lincRNA	-1.54	1.54E-002

Tabelle A1: Differentiell exprimierte Gene des Total-RNA-Seq-Experimentes - Fortsetzung

Ensembl Gene ID	Gene symbol	Bio type	Log ₂ FC	FDR
ENSG00000232325	AC093627.7	lincRNA	1.95	3.02E-002
ENSG00000270069	RP6-99M1.2	lincRNA	-1.57	3.79E-002
ENSG00000006062	MAP3K14	processed transcript	1.90	1.38E-008
ENSG00000224459	RP11-169K16.4	antisense	4.41	1.19E-004
ENSG00000250303	RP11-356J5.12	processed transcript	2.90	2.43E-003
ENSG00000230712	AP000354.4	pseudogene	4.35	5.10E-003
ENSG00000204054	LINC00963	processed transcript	1.50	5.21E-003
ENSG00000232442	CTD-3184A7.4	antisense	2.17	6.56E-003
ENSG00000224321	RP11-169K16.6	pseudogene	1.26	9.03E-003
ENSG00000197847	SLC22A20	pseudogene	2.27	9.10E-003
ENSG00000236972	FABP5P1	pseudogene	-3.50	1.05E-002
ENSG00000268049	CTD-2619J13.9	antisense	2.05	1.14E-002
ENSG00000207294	Y RNA.501	misc RNA	3.51	1.28E-002
ENSG00000255248	RP11-166D19.1	sense overlapping	-1.79	1.36E-002
ENSG00000215447	BX322557.10	processed transcript	1.23	1.37E-002
ENSG00000250899	RP11-253E3.3	processed transcript	-0.98	1.63E-002
ENSG00000088340	FER1L4	pseudogene	2.48	1.64E-002
ENSG00000220793	RPL21P119	pseudogene	-3.52	2.25E-002
ENSG00000234964	FABP5P7	pseudogene	-1.94	2.36E-002
ENSG00000251196	RP11-54F2.1	pseudogene	1.42	2.53E-002
ENSG00000240409	MTATP8P1	pseudogene	-1.41	2.78E-002
ENSG00000213860	RPL21P75	pseudogene	-4.01	2.81E-002
ENSG00000143429	AC027612.6	pseudogene	0.87	2.83E-002
ENSG00000203497	PDCD4-AS1	antisense	1.70	2.92E-002
ENSG00000126005	MMP24-AS1	antisense	1.14	3.01E-002
ENSG00000244112	RN7SL508P	misc RNA	3.48	3.11E-002
ENSG00000206706	Y RNA.522	misc RNA	3.11	3.24E-002
ENSG00000201699	RNU1-59P	snRNA	-2.91	3.49E-002
ENSG00000249353	NPM1P27	pseudogene	-2.38	3.73E-002
ENSG00000197582	GPX1P1	pseudogene	1.37	3.98E-002
ENSG00000237686	RP5-1120P11.1	antisense	4.25	4.07E-002
ENSG00000244021	RP11-50D9.1	pseudogene	-3.24	4.28E-002
ENSG00000267185	PTP4A2P1	pseudogene	-3.51	4.36E-002
ENSG00000228002	DHX9P1	pseudogene	-3.41	4.39E-002
ENSG00000248278	SUMO2P17	pseudogene	-2.42	4.56E-002
ENSG00000237676	RPL30P4	pseudogene	-5.09	4.62E-002
ENSG00000231313	CLIC1P1	pseudogene	-1.30	4.64E-002

Tabelle A2: Differentiell exprimierte microRNAs des microRNA-Seq-Experimentes

MirBase ID	Gene symbol	Log ₂ FC	FDR
MIMAT0000267	hsa-miR-210-3p	2.57	8.99E-014
MIMAT0004749	hsa-miR-424-3p	1.76	6.65E-010
MIMAT0005911	hsa-miR-1260a	1.40	1.26E-009
MIMAT0015041	hsa-miR-1260b	1.29	5.30E-009
MIMAT0019964	hsa-miR-4792	2.82	5.30E-009
MIMAT0000243	hsa-miR-148a-3p	-2.13	1.38E-006
MIMAT0004558	hsa-miR-181a-2-3p	1.16	6.84E-006
MIMAT0001341	hsa-miR-424-5p	1.38	6.97E-006
MIMAT0026479	hsa-miR-152-5p	1.49	4.83E-005
MIMAT0020925	hsa-miR-550a-3-5p	2.09	9.72E-005
MIMAT0003250	hsa-miR-585-3p	2.33	1.26E-004
MIMAT0000261	hsa-miR-183-5p	-0.77	2.54E-004
MIMAT0019958	hsa-miR-4788	1.91	2.67E-004
MIMAT0000728	hsa-miR-375	-3.43	3.98E-004
MIMAT0004588	hsa-miR-27b-5p	1.65	6.50E-004
MIMAT0031180	hsa-miR-7977	1.33	6.54E-004
MIMAT0000281	hsa-miR-224-5p	1.07	7.63E-004
MIMAT0000270	hsa-miR-181a-3p	0.93	8.68E-004
MIMAT0000259	hsa-miR-182-5p	-0.65	1.48E-003
MIMAT0005929	hsa-miR-1275	1.17	2.87E-003
MIMAT0026618	hsa-miR-585-5p	2.18	4.01E-003
MIMAT0030019	hsa-miR-7704	2.70	4.06E-003
MIMAT0016915	hsa-miR-4284	2.91	1.10E-002
MIMAT0004703	hsa-miR-335-3p	1.94	1.10E-002
MIMAT0019007	hsa-miR-3689c	-1.39	1.19E-002
MIMAT0018118	hsa-miR-3689a-3p	-1.23	1.75E-002
MIMAT0003333	hsa-miR-549a	3.79	1.90E-002
MIMAT0019010	hsa-miR-3689f	-1.15	2.03E-002
MIMAT0005825	hsa-miR-1180-3p	0.99	2.20E-002
MIMAT0000426	hsa-miR-132-3p	1.29	3.21E-002
MIMAT0004819	hsa-miR-671-3p	0.70	3.47E-002
MIMAT0004504	hsa-miR-31-3p	-1.75	3.60E-002
MIMAT0000423	hsa-miR-125b-5p	-0.68	3.60E-002
MIMAT0000646	hsa-miR-155-5p	-0.56	4.28E-002
MIMAT0004498	hsa-miR-25-5p	0.86	4.47E-002

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene

GOID	GO Term	FDR	% Associated Genes	No. of Genes
GO:0030198	extracellular matrix organization	4.50E-009	8.78	36
GO:0001568	blood vessel development	1.50E-007	6.65	43
GO:0040011	locomotion	9.40E-007	4.31	84
GO:0042060	wound healing	1.10E-006	5.91	46
GO:0072358	cardiovascular system development	6.70E-006	5.05	53
GO:0007166	cell surface receptor signaling pathway	7.60E-006	3.69	111
GO:0032502	developmental process	1.00E-005	3.07	190
GO:0044767	single-organism developmental process	1.10E-005	3.08	188
GO:0070887	cellular response to chemical stimulus	2.00E-005	3.61	108
GO:0048731	system development	2.00E-005	3.25	152
GO:0040012	regulation of locomotion	2.60E-005	5.27	43
GO:0048856	anatomical structure development	3.10E-005	3.11	171
GO:0009605	response to external stimulus	3.10E-005	3.67	99
GO:0016477	cell migration	3.20E-005	4.52	58
GO:0007167	enzyme linked receptor protein signaling pathway	3.30E-005	4.44	59
GO:0071363	cellular response to growth factor stimulus	3.40E-005	4.89	48
GO:0044707	single-multicellular organism process	3.40E-005	3	191
GO:0048870	cell motility	3.70E-005	4.35	61
GO:0009653	anatomical structure morphogenesis	3.80E-005	3.55	106
GO:0007596	blood coagulation	4.40E-005	5.74	34
GO:0007275	multicellular organismal development	4.70E-005	3.11	163
GO:0048513	animal organ development	6.10E-005	3.42	115
GO:0010033	response to organic substance	7.30E-005	3.46	109
GO:0061448	connective tissue development	1.10E-004	7.72	20
GO:0048583	regulation of response to stimulus	1.10E-004	3.26	128
GO:0001525	angiogenesis	1.10E-004	6.07	28
GO:0071560	cellular response to transforming growth factor beta stimulus	1.40E-004	7.52	20
GO:0048514	blood vessel morphogenesis	1.40E-004	5.66	31
GO:0006928	movement of cell or subcellular component	1.40E-004	3.77	78
GO:0009611	response to wounding	1.80E-004	4.42	50
GO:1901342	regulation of vasculature development	1.90E-004	7.6	19
GO:0071496	cellular response to external stimulus	2.10E-004	7	21
GO:0048519	negative regulation of biological process	2.50E-004	3.08	148
GO:0044236	multicellular organismal metabolic process	2.50E-004	9.52	14
GO:0032963	collagen metabolic process	2.60E-004	10.24	13
GO:0001935	endothelial cell proliferation	2.80E-004	10.08	13
GO:0051270	regulation of cellular component movement	2.90E-004	4.8	39
GO:0048523	negative regulation of cellular process	2.90E-004	3.12	138
GO:0031589	cell-substrate adhesion	2.90E-004	6.57	22
GO:0071310	cellular response to organic substance	3.20E-004	3.53	87
GO:0006468	protein phosphorylation	3.30E-004	3.74	72
GO:0010646	regulation of cell communication	3.40E-004	3.33	106
GO:0001937	negative regulation of endothelial cell proliferation	3.70E-004	20	7
GO:0019220	regulation of phosphate metabolic process	3.90E-004	3.77	69
GO:0006935	chemotaxis	3.90E-004	4.55	42
GO:0030334	regulation of cell migration	4.00E-004	4.92	35
GO:0042221	response to chemical	4.10E-004	3.07	142
GO:0050878	regulation of body fluid levels	4.20E-004	4.78	37
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	4.20E-004	6.15	23
GO:0032501	multicellular organismal process	4.80E-004	2.8	207
GO:0000904	cell morphogenesis involved in differentiation	5.30E-004	4.23	48
GO:0030154	cell differentiation	6.10E-004	3.13	126
GO:0048585	negative regulation of response to stimulus	7.20E-004	3.9	58
GO:0007155	cell adhesion	8.90E-004	3.84	59
GO:0001936	regulation of endothelial cell proliferation	9.10E-004	10.09	11
GO:0006950	response to stress	9.30E-004	3.07	131
GO:0048869	cellular developmental process	9.50E-004	3.06	132
GO:0007179	transforming growth factor beta receptor signaling pathway	9.70E-004	7.31	16

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene - Fortsetzung

GOID	GO Term	FDR	% Associated Genes	No. of Genes
GO:0009719	response to endogenous stimulus	1.00E-003	3.66	67
GO:0070482	response to oxygen levels	1.20E-003	5.97	21
GO:0032102	negative regulation of response to external stimulus	1.20E-003	6.33	19
GO:0009888	tissue development	1.30E-003	3.59	69
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	1.30E-003	4.25	42
GO:0045765	regulation of angiogenesis	1.40E-003	7.02	16
GO:0042325	regulation of phosphorylation	1.40E-003	3.75	59
GO:0051246	regulation of protein metabolic process	1.50E-003	3.3	91
GO:0050896	response to stimulus	1.50E-003	2.68	235
GO:1901700	response to oxygen-containing compound	1.80E-003	3.62	64
GO:1904018	positive regulation of vasculature development	2.10E-003	8.33	12
GO:0090288	negative regulation of cellular response to growth factor stimulus	2.20E-003	8.28	12
GO:0048646	anatomical structure formation involved in morphogenesis	2.20E-003	3.89	50
GO:0046952	ketone body catabolic process	2.20E-003	60	3
GO:0007160	cell-matrix adhesion	2.20E-003	7.01	15
GO:0051094	positive regulation of developmental process	2.40E-003	3.91	48
GO:0043542	endothelial cell migration	2.40E-003	7.65	13
GO:0031668	cellular response to extracellular stimulus	2.40E-003	6.91	15
GO:0051239	regulation of multicellular organismal process	2.50E-003	3.27	89
GO:0042127	regulation of cell proliferation	2.70E-003	3.63	60
GO:0016310	phosphorylation	2.90E-003	3.36	78
GO:0023051	regulation of signaling	3.10E-003	3.16	100
GO:0007229	integrin-mediated signaling pathway	3.10E-003	9.17	10
GO:0006793	phosphorus metabolic process	3.20E-003	3.13	103
GO:0060324	face development	3.30E-003	12.96	7
GO:0044699	single-organism process	3.30E-003	2.44	338
GO:2000026	regulation of multicellular organismal development	3.40E-003	3.57	61
GO:0051240	positive regulation of multicellular organismal process	3.40E-003	3.66	56
GO:0032268	regulation of cellular protein metabolic process	3.40E-003	3.28	84
GO:0032101	regulation of response to external stimulus	3.40E-003	4.12	39
GO:0008283	cell proliferation	3.40E-003	3.4	72
GO:0000902	cell morphogenesis	3.50E-003	3.61	58
GO:0040013	negative regulation of locomotion	3.80E-003	6.03	17
GO:0048468	cell development	3.90E-003	3.31	78
GO:0032989	cellular component morphogenesis	3.90E-003	3.55	60
GO:0032964	collagen biosynthetic process	3.90E-003	15	6
GO:0000165	MAPK cascade	3.90E-003	4.12	38
GO:0007165	signal transduction	4.40E-003	2.8	169
GO:0007154	cell communication	4.40E-003	2.76	182
GO:0010648	negative regulation of cell communication	4.50E-003	3.74	49
GO:0009966	regulation of signal transduction	4.60E-003	3.19	89
GO:0051716	cellular response to stimulus	4.70E-003	2.7	200
GO:0031399	regulation of protein modification process	4.70E-003	3.44	65
GO:0033673	negative regulation of kinase activity	4.80E-003	6.06	16
GO:0009968	negative regulation of signal transduction	4.80E-003	3.83	45
GO:0031669	cellular response to nutrient levels	4.90E-003	6.91	13
GO:0043408	regulation of MAPK cascade	5.30E-003	4.15	35
GO:0001706	endoderm formation	5.30E-003	11.67	7
GO:0050793	regulation of developmental process	5.40E-003	3.28	76
GO:0001666	response to hypoxia	5.50E-003	5.57	18
GO:0044344	cellular response to fibroblast growth factor stimulus	5.60E-003	5.4	19
GO:0071495	cellular response to endogenous stimulus	5.90E-003	3.65	50
GO:0060055	angiogenesis involved in wound healing	5.90E-003	23.53	4
GO:0050900	leukocyte migration	5.90E-003	5.21	20
GO:0048584	positive regulation of response to stimulus	5.90E-003	3.3	73
GO:0045859	regulation of protein kinase activity	5.90E-003	4.03	37
GO:0030574	collagen catabolic process	5.90E-003	9.88	8

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene - Fortsetzung

GOID	GOTerm	FDR	% Associated Genes	No. of Genes
GO:0030336	negative regulation of cell migration	5.90E-003	6.42	14
GO:0023057	negative regulation of signaling	5.90E-003	3.7	48
GO:0009991	response to extracellular stimulus	5.90E-003	4.55	27
GO:0048705	skeletal system morphogenesis	6.10E-003	6.33	14
GO:0044710	single-organism metabolic process	6.10E-003	2.79	163
GO:0018277	protein deamination	6.10E-003	100	2
GO:0030335	positive regulation of cell migration	6.40E-003	5.04	21
GO:0040017	positive regulation of locomotion	6.50E-003	4.91	22
GO:0019538	protein metabolic process	6.60E-003	2.78	164
GO:0001704	formation of primary germ layer	6.60E-003	7.94	10
GO:0060976	coronary vasculature development	7.10E-003	12.77	6
GO:0044700	single organism signaling	7.10E-003	2.73	178
GO:0034329	cell junction assembly	7.70E-003	5.91	15
GO:1903035	negative regulation of response to wounding	7.80E-003	6.78	12
GO:0061035	regulation of cartilage development	7.80E-003	10.61	7
GO:0072001	renal system development	8.40E-003	5.45	17
GO:0045766	positive regulation of angiogenesis	8.50E-003	7.63	10
GO:0071900	regulation of protein serine/threonine kinase activity	9.00E-003	4.31	28
GO:0060840	artery development	9.00E-003	9.09	8
GO:0044763	single-organism cellular process	9.00E-003	2.46	313
GO:0090400	stress-induced premature senescence	9.20E-003	33.33	3
GO:0032107	regulation of response to nutrient levels	9.20E-003	7.52	10
GO:0009628	response to abiotic stimulus	9.20E-003	3.62	47
GO:0072224	metanephric glomerulus development	9.60E-003	20	4
GO:0060325	face morphogenesis	9.70E-003	14.71	5
GO:0032269	negative regulation of cellular protein metabolic process	9.70E-003	3.75	41
GO:0006929	substrate-dependent cell migration	9.70E-003	14.71	5
GO:0080135	regulation of cellular response to stress	1.00E-002	4.17	29
GO:0051896	regulation of protein kinase B signaling	1.00E-002	7.35	10
GO:0051248	negative regulation of protein metabolic process	1.00E-002	3.69	43
GO:0043405	regulation of MAP kinase activity	1.00E-002	4.61	23
GO:0042326	negative regulation of phosphorylation	1.00E-002	4.76	21
GO:0035966	response to topologically incorrect protein	1.00E-002	6.52	12
GO:0035455	response to interferon-alpha	1.00E-002	19.05	4
GO:0031670	cellular response to nutrient	1.00E-002	11.54	6
GO:0031667	response to nutrient levels	1.00E-002	4.43	25
GO:0006749	glutathione metabolic process	1.00E-002	9.86	7
GO:0001667	ameboidal-type cell migration	1.00E-002	5.14	18
GO:0080134	regulation of response to stress	1.10E-002	3.43	53
GO:0061041	regulation of wound healing	1.10E-002	7.14	10
GO:0048812	neuron projection morphogenesis	1.10E-002	3.86	36
GO:0048667	cell morphogenesis involved in neuron differentiation	1.10E-002	3.87	35
GO:0045936	negative regulation of phosphate metabolic process	1.10E-002	4.38	25
GO:0043066	negative regulation of apoptotic process	1.10E-002	3.86	36
GO:0007162	negative regulation of cell adhesion	1.10E-002	5.81	14
GO:0006979	response to oxidative stress	1.10E-002	4.74	21
GO:0035456	response to interferon-beta	1.20E-002	18.18	4
GO:0006915	apoptotic process	1.20E-002	3.26	65
GO:2000379	positive regulation of reactive oxygen species metabolic process	1.30E-002	8.33	8
GO:1901420	negative regulation of response to alcohol	1.30E-002	66.67	2
GO:0072229	metanephric proximal convoluted tubule development	1.30E-002	66.67	2
GO:0060548	negative regulation of cell death	1.30E-002	3.73	38
GO:0050795	regulation of behavior	1.30E-002	5.47	15
GO:0035793	positive regulation of metanephric mesenchymal cell migration by platelet-derived growth factor receptor-beta signaling pathway	1.30E-002	66.67	2
GO:0033993	response to lipid	1.30E-002	3.81	36
GO:0032496	response to lipopolysaccharide	1.30E-002	4.97	18

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene - Fortsetzung

GOID	GO Term	FDR	% Associated Genes	No. of Genes
GO:0030168	platelet activation	1.30E-002	5.42	15
GO:0022617	extracellular matrix disassembly	1.30E-002	6.9	10
GO:0022408	negative regulation of cell-cell adhesion	1.30E-002	6.94	10
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	1.30E-002	7.56	9
GO:0016043	cellular component organization	1.30E-002	2.7	172
GO:0007423	sensory organ development	1.30E-002	4.33	25
GO:0007411	axon guidance	1.30E-002	4.23	26
GO:0007369	gastrulation	1.30E-002	6.15	12
GO:0007044	cell-substrate junction assembly	1.30E-002	8.25	8
GO:0051216	cartilage development	1.40E-002	6.09	12
GO:0042594	response to starvation	1.40E-002	6.4	11
GO:0030030	cell projection organization	1.40E-002	3.35	55
GO:0022603	regulation of anatomical structure morphogenesis	1.40E-002	3.7	38
GO:0009887	organ morphogenesis	1.40E-002	3.7	38
GO:0008285	negative regulation of cell proliferation	1.40E-002	4.04	29
GO:0001817	regulation of cytokine production	1.40E-002	4.21	26
GO:0048518	positive regulation of biological process	1.50E-002	2.73	157
GO:0045061	thymic T cell selection	1.50E-002	16.67	4
GO:0010647	positive regulation of cell communication	1.50E-002	3.31	57
GO:0008219	cell death	1.50E-002	3.18	68
GO:0001656	metanephros development	1.50E-002	8	8
GO:0001503	ossification	1.50E-002	4.75	19
GO:2000181	negative regulation of blood vessel morphogenesis	1.60E-002	8.75	7
GO:0061383	trabecula morphogenesis	1.60E-002	12.2	5
GO:0071840	cellular component organization or biogenesis	1.70E-002	2.67	174
GO:0051179	localization	1.70E-002	2.71	161
GO:0014909	smooth muscle cell migration	1.70E-002	10	6
GO:0010632	regulation of epithelial cell migration	1.70E-002	6.21	11
GO:0002376	immune system process	1.70E-002	3.02	85
GO:0030199	collagen fibril organization	1.80E-002	11.9	5
GO:0030155	regulation of cell adhesion	1.80E-002	4	28
GO:0007173	epidermal growth factor receptor signaling pathway	1.80E-002	4.75	18
GO:0090287	regulation of cellular response to growth factor stimulus	1.90E-002	5.32	14
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	1.90E-002	5.53	13
GO:0050680	negative regulation of epithelial cell proliferation	1.90E-002	6.98	9
GO:0044267	cellular protein metabolic process	1.90E-002	2.74	144
GO:0009790	embryo development	1.90E-002	3.59	39
GO:0002062	chondrocyte differentiation	1.90E-002	7.55	8
GO:1900138	negative regulation of phospholipase A2 activity	2.00E-002	50	2
GO:0051348	negative regulation of transferase activity	2.00E-002	4.69	18
GO:0051097	negative regulation of helicase activity	2.00E-002	50	2
GO:0044319	wound healing, spreading of cells	2.00E-002	14.81	4
GO:0001822	kidney development	2.00E-002	5.1	15
GO:0001654	eye development	2.00E-002	4.68	18
GO:0032835	glomerulus development	2.10E-002	9.38	6
GO:2000351	regulation of endothelial cell apoptotic process	2.20E-002	11.11	5
GO:0060429	epithelium development	2.20E-002	3.49	42
GO:0050921	positive regulation of chemotaxis	2.20E-002	6.72	9
GO:0050673	epithelial cell proliferation	2.20E-002	4.63	18
GO:0071260	cellular response to mechanical stimulus	2.30E-002	8.05	7
GO:0060322	head development	2.30E-002	3.78	31
GO:0045937	positive regulation of phosphate metabolic process	2.30E-002	3.45	43
GO:0045860	positive regulation of protein kinase activity	2.40E-002	3.99	26
GO:0042035	regulation of cytokine biosynthetic process	2.40E-002	7.21	8
GO:0010634	positive regulation of epithelial cell migration	2.40E-002	7.21	8
GO:0042327	positive regulation of phosphorylation	2.50E-002	3.51	39
GO:0023056	positive regulation of signaling	2.50E-002	3.23	55

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene - Fortsetzung

GOID	GOTerm	FDR	% Associated Genes	No. of Genes
GO:0018212	peptidyl-tyrosine modification	2.50E-002	4.67	17
GO:0009636	response to toxic substance	2.50E-002	5.09	14
GO:0001501	skeletal system development	2.60E-002	4.19	22
GO:2000772	regulation of cellular senescence	2.80E-002	13.33	4
GO:0045601	regulation of endothelial cell differentiation	2.80E-002	13.33	4
GO:1901701	cellular response to oxygen-containing compound	2.90E-002	3.45	40
GO:0051336	regulation of hydrolase activity	2.90E-002	3.36	45
GO:0006024	glycosaminoglycan biosynthetic process	2.90E-002	6.96	8
GO:0090270	regulation of fibroblast growth factor production	3.00E-002	40	2
GO:0045216	cell-cell junction organization	3.00E-002	5.16	13
GO:0038001	paracrine signaling	3.00E-002	40	2
GO:0031175	neuron projection development	3.00E-002	3.43	41
GO:0055098	response to low-density lipoprotein particle	3.10E-002	18.75	3
GO:0055093	response to hyperoxia	3.10E-002	12.9	4
GO:0048522	positive regulation of cellular process	3.10E-002	2.73	135
GO:0046718	viral entry into host cell	3.10E-002	6.84	8
GO:0044711	single-organism biosynthetic process	3.10E-002	3.14	59
GO:0042730	fibrinolysis	3.10E-002	12.9	4
GO:0032103	positive regulation of response to external stimulus	3.10E-002	4.79	15
GO:0002710	negative regulation of T cell mediated immunity	3.10E-002	18.75	3
GO:0072006	nephron development	3.20E-002	6.25	9
GO:0006898	receptor-mediated endocytosis	3.20E-002	4.64	16
GO:2000377	regulation of reactive oxygen species metabolic process	3.30E-002	5.81	10
GO:0097191	extrinsic apoptotic signaling pathway	3.30E-002	5.08	13
GO:0061036	positive regulation of cartilage development	3.30E-002	12.5	4
GO:0050866	negative regulation of cell activation	3.30E-002	5.81	10
GO:0048858	cell projection morphogenesis	3.30E-002	3.4	41
GO:0007265	Ras protein signal transduction	3.30E-002	4.14	21
GO:0006986	response to unfolded protein	3.30E-002	5.81	10
GO:0006955	immune response	3.30E-002	3.14	58
GO:0001649	osteoblast differentiation	3.30E-002	5.29	12
GO:0051128	regulation of cellular component organization	3.40E-002	3.02	71
GO:0090083	regulation of inclusion body assembly	3.50E-002	17.65	3
GO:0071902	positive regulation of protein serine/threonine kinase activity	3.50E-002	4.7	15
GO:0071636	positive regulation of transforming growth factor beta production	3.50E-002	17.65	3
GO:0071222	cellular response to lipopolysaccharide	3.50E-002	5.75	10
GO:0071216	cellular response to biotic stimulus	3.50E-002	5.47	11
GO:0051788	response to misfolded protein	3.50E-002	17.65	3
GO:0051338	regulation of transferase activity	3.50E-002	3.45	38
GO:0045597	positive regulation of cell differentiation	3.50E-002	3.6	32
GO:0009725	response to hormone	3.50E-002	3.42	39
GO:0002011	morphogenesis of an epithelial sheet	3.50E-002	9.62	5
GO:0060326	cell chemotaxis	3.60E-002	4.98	13
GO:0090101	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	3.70E-002	6.5	8
GO:0050678	regulation of epithelial cell proliferation	3.70E-002	4.66	15
GO:0030203	glycosaminoglycan metabolic process	3.70E-002	5.68	10
GO:0009967	positive regulation of signal transduction	3.70E-002	3.22	50
GO:0008361	regulation of cell size	3.70E-002	5.68	10
GO:0006464	cellular protein modification process	3.70E-002	2.78	113
GO:0000060	protein import into nucleus, translocation	3.70E-002	9.43	5
GO:2000774	positive regulation of cellular senescence	3.80E-002	33.33	2
GO:1990144	intrinsic apoptotic signaling pathway in response to hypoxia	3.80E-002	33.33	2
GO:1900221	regulation of beta-amyloid clearance	3.80E-002	33.33	2
GO:0070278	extracellular matrix constituent secretion	3.80E-002	33.33	2
GO:0060560	developmental growth involved in morphogenesis	3.80E-002	5.13	12
GO:0051241	negative regulation of multicellular organismal process	3.80E-002	3.43	37
GO:0048840	otolith development	3.80E-002	33.33	2

Tabelle A3: Signifikant angereicherte biologische Prozesse der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene - Fortsetzung

GOID	GO Term	FDR	% Associated Genes	No. of Genes
GO:0045595	regulation of cell differentiation	3.80E-002	3.19	51
GO:0021885	forebrain cell migration	3.80E-002	8	6
GO:0008543	fibroblast growth factor receptor signaling pathway	3.80E-002	4.63	15
GO:0002578	negative regulation of antigen processing and presentation	3.80E-002	33.33	2
GO:0002576	platelet degranulation	3.80E-002	7.07	7
GO:0001957	intramembranous ossification	3.80E-002	33.33	2
GO:0065008	regulation of biological quality	3.90E-002	2.8	106
GO:0048008	platelet-derived growth factor receptor signaling pathway	3.90E-002	9.26	5
GO:0042542	response to hydrogen peroxide	3.90E-002	5.96	9
GO:0032655	regulation of interleukin-12 production	3.90E-002	9.26	5
GO:0003158	endothelium development	3.90E-002	6.4	8
GO:0097305	response to alcohol	4.00E-002	4.26	18
GO:0048545	response to steroid hormone	4.00E-002	4.1	20
GO:0032879	regulation of localization	4.00E-002	2.95	75
GO:0001819	positive regulation of cytokine production	4.00E-002	4.26	18
GO:0002709	regulation of T cell mediated immunity	4.10E-002	9.09	5
GO:1902531	regulation of intracellular signal transduction	4.20E-002	3.12	55
GO:0098602	single organism cell adhesion	4.30E-002	3.59	30
GO:0071850	mitotic cell cycle arrest	4.30E-002	15.79	3
GO:0061298	retina vasculature development in camera-type eye	4.30E-002	15.79	3
GO:0048666	neuron development	4.30E-002	3.26	44
GO:0043534	blood vessel endothelial cell migration	4.30E-002	7.69	6
GO:0032331	negative regulation of chondrocyte differentiation	4.30E-002	15.79	3
GO:0010544	negative regulation of platelet activation	4.30E-002	15.79	3
GO:0010506	regulation of autophagy	4.30E-002	5.24	11
GO:0009267	cellular response to starvation	4.30E-002	6.25	8
GO:0006907	pinocytosis	4.30E-002	15.79	3
GO:2000278	regulation of DNA biosynthetic process	4.40E-002	7.59	6
GO:0048041	focal adhesion assembly	4.40E-002	7.59	6
GO:0021675	nerve development	4.40E-002	7.59	6
GO:0007569	cell aging	4.40E-002	6.8	7
GO:0060348	bone development	4.50E-002	5.43	10
GO:0044281	small molecule metabolic process	4.50E-002	2.89	82
GO:0050790	regulation of catalytic activity	4.60E-002	2.94	73
GO:0043412	macromolecule modification	4.60E-002	2.74	116
GO:0034111	negative regulation of homotypic cell-cell adhesion	4.60E-002	6.73	7
GO:0007399	nervous system development	4.60E-002	2.93	75
GO:0000186	activation of MAPKK activity	4.70E-002	4.92	12
GO:0051549	positive regulation of keratinocyte migration	4.80E-002	28.57	2
GO:0043086	negative regulation of catalytic activity	4.80E-002	3.49	32
GO:0014070	response to organic cyclic compound	4.80E-002	3.46	33
GO:0010668	ectodermal cell differentiation	4.80E-002	28.57	2
GO:0001775	cell activation	4.80E-002	3.38	36
GO:0071901	negative regulation of protein serine/threonine kinase activity	4.90E-002	6.06	8
GO:0071295	cellular response to vitamin	4.90E-002	10.53	4
GO:0051247	positive regulation of protein metabolic process	4.90E-002	3.12	52
GO:0003205	cardiac chamber development	4.90E-002	5.66	9

Tabelle A4: Signifikant angereicherte molekulare Funktionen der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene

GOID	GOTerm	FDR	% Associated Genes	No. of Genes
GO:0048407	platelet-derived growth factor binding	7.10E-004	45.45	5
GO:0005201	extracellular matrix structural constituent	5.50E-003	12.33	9
GO:0005178	integrin binding	6.10E-003	9.48	11
GO:0050839	cell adhesion molecule binding	1.00E-002	7.00	14
GO:0005518	collagen binding	1.10E-002	11.59	8
GO:0005539	glycosaminoglycan binding	1.70E-002	6.54	14
GO:0032403	protein complex binding	2.30E-002	3.89	39
GO:0019838	growth factor binding	2.50E-002	7.75	10
GO:0005515	protein binding	2.80E-002	2.49	273
GO:0004222	metalloendopeptidase activity	3.00E-002	8.11	9
GO:0008201	heparin binding	3.70E-002	6.63	11
GO:0016723	oxidoreductase activity, oxidizing metal ions, NAD or NADP as acceptor	4.20E-002	27.27	3

Tabelle A5: Signifikant angereicherte zelluläre Komponenten der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene

GOID	GOTerm	FDR	% Associated Genes	No. of Genes
GO:0031012	extracellular matrix	3.30E-008	7.80	35
GO:0005578	proteinaceous extracellular matrix	3.10E-007	7.85	30
GO:0044420	extracellular matrix component	4.50E-006	11.68	16
GO:0044421	extracellular region part	5.30E-006	3.39	133
GO:0009986	cell surface	5.20E-005	5.02	41
GO:0005576	extracellular region	1.30E-004	3.10	145
GO:0005925	focal adhesion	3.10E-004	5.97	25
GO:0031988	membrane-bounded vesicle	7.90E-004	3.15	116
GO:0005912	adherens junction	8.30E-004	5.33	27
GO:0005783	endoplasmic reticulum	2.60E-003	3.60	61
GO:0005737	cytoplasm	2.70E-003	2.56	275
GO:0070062	extracellular exosome	4.50E-003	3.17	91
GO:0016020	membrane	5.70E-003	2.58	246
GO:0071944	cell periphery	1.40E-002	2.78	146
GO:0031091	platelet alpha granule	1.40E-002	10.29	7
GO:0030054	cell junction	1.40E-002	3.63	45
GO:0044444	cytoplasmic part	1.60E-002	2.60	213
GO:0012505	endomembrane system	1.90E-002	2.88	112
GO:0044459	plasma membrane part	2.10E-002	3.05	81
GO:0098644	complex of collagen trimers	2.20E-002	17.39	4
GO:0005615	extracellular space	2.80E-002	3.37	49
GO:0005788	endoplasmic reticulum lumen	3.50E-002	5.66	12
GO:0044425	membrane part	3.60E-002	2.60	187
GO:0031224	intrinsic component of membrane	4.80E-002	2.64	161

Tabelle A6: Signifikant angereicherte KEGG-Interaktionsnetzwerke der unter IGF2BP1-Knockdown in ES-2-Zellen differentiell exprimierten Gene

Pathway	Name	FDR	% Associated Genes	No. of Genes
hsa04151	PI3K-Akt signaling pathway	1.20E-005	7.92	27
hsa04510	Focal adhesion	4.20E-004	8.42	17
hsa04512	ECM-receptor interaction	6.00E-005	14.63	12
hsa04933	AGE-RAGE signaling pathway in diabetic complications	5.30E-003	9.90	10
hsa05200	Pathways in cancer	5.10E-003	5.79	23
hsa05219	Bladder cancer	3.30E-004	19.51	8

Tabelle A7: Unter IGF2BP1-Knockdown in ES-2-Zellen hochabundante und nicht-differentiell exprimierte microRNAs

MirBase ID	Gene symbol	Log ₂ FC	FDR
MIMAT0000692	hsa-miR-30e-5p	-0.30	1
MIMAT0004494	hsa-miR-21-3p	0.77	1
MIMAT0000254	hsa-miR-10b-5p	0.33	1
MIMAT0000084	hsa-miR-27a-3p	0.33	1
MIMAT0004810	hsa-miR-629-5p	0.40	1
MIMAT0002821	hsa-miR-181d-5p	0.31	1
MIMAT0002809	hsa-miR-146b-5p	0.42	1
MIMAT0000087	hsa-miR-30a-5p	-0.27	1
MIMAT0000097	hsa-miR-99a-5p	-0.57	1
MIMAT0004502	hsa-miR-28-3p	0.34	1
MIMAT0000098	hsa-miR-100-5p	-0.57	1
MIMAT0000257	hsa-miR-181b-5p	0.30	1
MIMAT0000449	hsa-miR-146a-5p	0.30	1
MIMAT0000443	hsa-miR-125a-5p	0.27	1
MIMAT0000089	hsa-miR-31-5p	-0.29	1
MIMAT0031177	hsa-miR-7974	0.39	1
MIMAT0010214	hsa-miR-151b	0.34	1
MIMAT0004550	hsa-miR-30c-2-3p	-0.35	1
MIMAT0000446	hsa-miR-127-3p	-0.90	1
MIMAT0000085	hsa-miR-28-5p	0.31	1
MIMAT0000691	hsa-miR-130b-3p	-0.27	1
MIMAT0000065	hsa-let-7d-5p	0.35	1
MIMAT0000759	hsa-miR-148b-3p	-0.30	1
MIMAT0000417	hsa-miR-15b-5p	-0.30	1
MIMAT0000076	hsa-miR-21-5p	0.21	1
MIMAT0000081	hsa-miR-25-3p	0.25	1
MIMAT0003218	hsa-miR-92b-3p	0.26	1
MIMAT0000070	hsa-miR-17-5p	-0.25	1
MIMAT0000063	hsa-let-7b-5p	0.25	1
MIMAT0000103	hsa-miR-106a-5p	-0.25	1
MIMAT0005951	hsa-miR-1307-3p	-0.25	1
MIMAT0000430	hsa-miR-138-5p	0.20	1
MIMAT0000458	hsa-miR-190a-5p	0.37	1
MIMAT0000104	hsa-miR-107	-0.18	1
MIMAT0000088	hsa-miR-30a-3p	-0.20	1
MIMAT0000078	hsa-miR-23a-3p	-0.23	1
MIMAT0004512	hsa-miR-100-3p	-0.31	1
MIMAT0003294	hsa-miR-625-5p	0.17	1
MIMAT0004808	hsa-miR-625-3p	0.17	1
MIMAT0000510	hsa-miR-320a	0.16	1
MIMAT0000414	hsa-let-7g-5p	0.13	1
MIMAT0000075	hsa-miR-20a-5p	-0.24	1
MIMAT0000693	hsa-miR-30e-3p	-0.16	1
MIMAT0000093	hsa-miR-93-5p	-0.14	1
MIMAT0005792	hsa-miR-320b	0.15	1
MIMAT0000092	hsa-miR-92a-3p	-0.14	1

Tabelle A7: Unter IGF2BP1-Knockdown in ES-2-Zellen hochabundante und nicht-differenziell exprimierte microRNAs - Fortsetzung

MirBase ID	Gene symbol	Log ₂ FC	FDR
MIMAT0000736	hsa-miR-381-3p	-0.40	1
MIMAT0000425	hsa-miR-130a-3p	-0.15	1
MIMAT0000435	hsa-miR-143-3p	-0.32	1
MIMAT0000688	hsa-miR-301a-3p	-0.28	1
MIMAT0029782	hsa-miR-7641	-0.36	1
MIMAT0004984	hsa-miR-941	0.14	1
MIMAT0000083	hsa-miR-26b-5p	-0.13	1
MIMAT0000256	hsa-miR-181a-5p	0.13	1
MIMAT0000418	hsa-miR-23b-3p	-0.13	1
MIMAT0005881	hsa-miR-1291	-0.83	1
MIMAT0000232	hsa-miR-199a-3p	0.09	1
MIMAT0004563	hsa-miR-199b-3p	0.09	1
MIMAT0000069	hsa-miR-16-5p	0.09	1
MIMAT0000440	hsa-miR-191-5p	-0.08	1
MIMAT0000231	hsa-miR-199a-5p	0.12	1
MIMAT0003886	hsa-miR-769-5p	-0.09	1
MIMAT0001080	hsa-miR-196b-5p	-0.09	1
MIMAT0000252	hsa-miR-7-5p	-0.15	1
MIMAT0002174	hsa-miR-484	0.10	1
MIMAT0000433	hsa-miR-142-5p	-0.18	1
MIMAT0003339	hsa-miR-421	-0.08	1
MIMAT0000226	hsa-miR-196a-5p	-0.08	1
MIMAT0000415	hsa-let-7i-5p	-0.06	1
MIMAT0004672	hsa-miR-106b-3p	0.07	1
MIMAT0000279	hsa-miR-222-3p	-0.06	1
MIMAT0004945	hsa-miR-744-5p	-0.07	1
MIMAT0000064	hsa-let-7c-5p	-0.06	1
MIMAT0000703	hsa-miR-361-5p	0.05	1
MIMAT0000420	hsa-miR-30b-5p	0.04	1
MIMAT0000096	hsa-miR-98-5p	-0.04	1
MIMAT0004597	hsa-miR-140-3p	0.05	1
MIMAT0000067	hsa-let-7f-5p	-0.04	1
MIMAT0000222	hsa-miR-192-5p	-0.04	1
MIMAT0000066	hsa-let-7e-5p	-0.05	1
MIMAT0000062	hsa-let-7a-5p	-0.04	1
MIMAT0000424	hsa-miR-128-3p	0.04	1
MIMAT0003393	hsa-miR-425-5p	0.04	1
MIMAT0004692	hsa-miR-340-5p	0.06	1
MIMAT0000689	hsa-miR-99b-5p	0.04	1
MIMAT0000073	hsa-miR-19a-3p	0.03	1
MIMAT0019074	hsa-miR-378i	-0.02	1
MIMAT0000074	hsa-miR-19b-3p	0.01	1
MIMAT0005882	hsa-miR-548k	0.01	1
MIMAT0000099	hsa-miR-101-3p	-0.01	1
MIMAT0016847	hsa-miR-378c	0.00	1
MIMAT0022727	hsa-miR-1307-5p	0.00	1

Tabelle A8: Putative IGF2BP1-Ziel-RNAs

Ensembl Gene ID	Gene symbol	Log ₂ FC	FDR	No. of predicted binding miRs	Part of significantly enriched biological process related to cell migration
ENSG00000108091	CCDC6	-3.11	3.15E-022	23	FALSE
ENSG00000115677	HDLBP	-2.19	1.34E-019	13	FALSE
ENSG00000159217	IGF2BP1	-2.03	1.57E-014	20	FALSE
ENSG00000136261	BZW2	-2.08	3.10E-006	9	FALSE
ENSG00000123352	SPATS2	-1.47	4.24E-006	1	FALSE
ENSG00000115159	GPD2	-2.02	4.24E-006	14	FALSE
ENSG00000157557	ETS2	-1.49	3.24E-005	3	FALSE
ENSG00000165832	TRUB1	-2.47	4.94E-005	1	FALSE
ENSG00000136158	SPRY2	-1.28	4.94E-005	7	FALSE
ENSG00000111266	DUSP16	-1.51	1.62E-004	10	FALSE
ENSG00000143515	ATP8B2	-1.16	1.91E-004	4	FALSE
ENSG00000029725	RABEP1	-2.09	1.93E-004	5	FALSE
ENSG00000124762	CDKN1A	-2.02	2.20E-004	16	FALSE
ENSG00000114850	SSR3	-1.23	3.27E-004	5	FALSE
ENSG00000166881	TMEM194A	-1.54	3.64E-004	16	FALSE
ENSG00000069020	MAST4	-1.46	3.86E-004	10	FALSE
ENSG00000123892	RAB38	-3.40	4.41E-004	3	FALSE
ENSG00000104332	SFRP1	-0.99	9.84E-004	4	TRUE
ENSG00000057252	SOAT1	-1.63	9.98E-004	1	FALSE
ENSG00000100664	EIF5	-1.43	1.15E-003	6	FALSE
ENSG00000011201	KAL1	-1.87	1.23E-003	4	TRUE
ENSG00000145390	USP53	-2.59	1.29E-003	9	FALSE
ENSG00000096063	SRPK1	-1.45	1.33E-003	9	FALSE
ENSG00000171055	FEZ2	-2.09	1.77E-003	2	FALSE
ENSG00000180917	CMTR2	-2.03	1.82E-003	6	FALSE
ENSG00000136802	LRRRC8A	-1.15	2.35E-003	1	FALSE
ENSG00000105926	MPP6	-2.02	2.77E-003	2	FALSE
ENSG00000171310	CHST11	-1.14	3.04E-003	6	FALSE
ENSG00000198894	CIPC	-1.25	3.04E-003	7	FALSE
ENSG00000163659	TIPARP	-1.61	3.19E-003	5	FALSE
ENSG00000139278	GLIPR1	-2.13	3.48E-003	3	FALSE
ENSG00000091409	ITGA6	-1.49	3.52E-003	11	TRUE
ENSG00000164933	SLC25A32	-1.45	3.52E-003	16	FALSE
ENSG00000163900	TMEM41A	-1.49	4.17E-003	3	FALSE
ENSG00000114120	SLC25A36	-1.57	4.29E-003	15	FALSE
ENSG00000057704	TMCC3	-1.50	4.46E-003	4	FALSE
ENSG00000106571	GLI3	-1.43	4.80E-003	3	TRUE
ENSG00000139722	VPS37B	-1.18	4.88E-003	1	FALSE
ENSG00000095794	CREM	-1.49	5.52E-003	15	FALSE
ENSG00000188848	BEND4	-1.19	5.83E-003	18	FALSE
ENSG00000131016	AKAP12	-1.33	5.83E-003	6	FALSE
ENSG00000171517	LPAR3	-1.33	5.83E-003	2	FALSE
ENSG00000143322	ABL2	-1.20	5.93E-003	31	TRUE
ENSG00000134602	MST4	-1.80	5.93E-003	7	TRUE
ENSG00000059758	CDK17	-1.52	6.53E-003	14	FALSE
ENSG00000108861	DUSP3	-1.01	6.53E-003	11	TRUE
ENSG00000049130	KITLG	-6.38	6.60E-003	13	TRUE
ENSG00000165449	SLC16A9	-2.01	6.66E-003	15	FALSE
ENSG00000151012	SLC7A11	-2.03	6.77E-003	28	TRUE
ENSG00000112893	MAN2A1	-1.57	7.24E-003	5	FALSE
ENSG00000057019	DCBLD2	-1.32	7.94E-003	20	FALSE
ENSG00000115419	GLS	-2.55	8.03E-003	7	FALSE
ENSG00000114573	ATP6V1A	-1.64	8.08E-003	11	FALSE
ENSG00000113356	POLR3G	-1.45	9.00E-003	11	FALSE
ENSG00000105968	H2AFV	-1.48	9.33E-003	2	FALSE
ENSG00000152492	CCDC50	-1.35	9.89E-003	11	FALSE
ENSG00000153879	CEBPG	-1.83	1.00E-002	4	FALSE
ENSG00000109133	TMEM33	-1.08	1.10E-002	15	FALSE
ENSG00000181061	HIGD1A	-2.50	1.33E-002	3	FALSE
ENSG00000137876	RSL24D1	-1.94	1.34E-002	1	FALSE

Tabelle A8: Putative IGF2BP1-Ziel-RNAs - Fortsetzung

Ensembl Gene ID	Gene symbol	Log ₂ FC	FDR	No. of predicted binding miRs	Part of significantly enriched biological process related to cell migration
ENSG00000100632	ERH	-1.42	1.35E-002	2	FALSE
ENSG00000166037	CEP57	-1.41	1.42E-002	6	FALSE
ENSG00000063601	MTMR1	-1.08	1.44E-002	5	FALSE
ENSG00000111371	SLC38A1	-1.18	1.58E-002	3	FALSE
ENSG00000093167	LRRFIP2	-1.01	1.60E-002	3	FALSE
ENSG00000174606	ANGEL2	-1.21	1.63E-002	5	FALSE
ENSG00000169504	CLIC4	-1.33	1.68E-002	4	TRUE
ENSG00000174013	FBXO45	-1.36	1.81E-002	16	TRUE
ENSG00000110619	CARS	-0.92	1.90E-002	3	FALSE
ENSG00000152332	UHMK1	-1.28	1.99E-002	19	FALSE
ENSG00000116489	CAPZA1	-2.49	2.00E-002	9	FALSE
ENSG00000147526	TACC1	-1.13	2.03E-002	12	FALSE
ENSG00000151835	SACS	-1.56	2.07E-002	10	FALSE
ENSG00000128965	CHAC1	-1.53	2.12E-002	3	FALSE
ENSG00000137040	RANBP6	-2.49	2.18E-002	8	FALSE
ENSG00000174574	AKIRIN1	-1.42	2.25E-002	15	FALSE
ENSG00000181751	C5orf30.1	-1.39	2.25E-002	15	FALSE
ENSG00000121644	DESI2	-1.11	2.25E-002	14	FALSE
ENSG00000064102	ASUN	-1.43	2.27E-002	2	FALSE
ENSG00000136518	ACTL6A	-1.74	2.35E-002	2	FALSE
ENSG00000180884	ZNF792	-1.21	2.35E-002	1	FALSE
ENSG00000165156	ZHX1	-2.52	2.36E-002	11	FALSE
ENSG00000162694	EXTL2	-1.51	2.37E-002	3	FALSE
ENSG00000119977	TCTN3	-1.28	2.37E-002	1	FALSE
ENSG00000115946	PNO1	-1.41	2.38E-002	1	FALSE
ENSG00000123505	AMD1	-1.54	2.40E-002	3	FALSE
ENSG00000050130	JKAMP	-2.69	2.40E-002	2	FALSE
ENSG00000153922	CHD1	-2.19	2.52E-002	8	FALSE
ENSG00000112658	SRF	-1.08	2.53E-002	3	TRUE
ENSG00000115963	RND3	-2.11	2.53E-002	10	TRUE
ENSG00000101187	SLCO4A1	-1.23	2.64E-002	1	FALSE
ENSG00000128989	ARPP19	-1.34	2.76E-002	25	FALSE
ENSG00000173334	TRIB1	-1.04	2.78E-002	13	TRUE
ENSG00000148426	PROSER2	-1.28	2.79E-002	6	FALSE
ENSG00000197965	MPZL1	-0.80	2.92E-002	3	FALSE
ENSG00000106692	FKTN	-2.18	3.01E-002	1	FALSE
ENSG00000104635	SLC39A14	-0.86	3.01E-002	5	FALSE
ENSG00000115233	PSMD14	-1.37	3.18E-002	2	FALSE
ENSG00000155100	OTUD6B	-2.64	3.18E-002	9	FALSE
ENSG00000169251	NMD3	-1.94	3.22E-002	3	FALSE
ENSG00000153162	BMP6	-1.91	3.32E-002	6	FALSE
ENSG00000213516	RBMXL1	-1.49	3.37E-002	7	FALSE
ENSG00000083720	OXCT1	-2.37	3.45E-002	2	FALSE
ENSG00000100030	MAPK1	-1.12	3.49E-002	14	TRUE
ENSG00000074935	TUBE1	-1.40	3.65E-002	1	FALSE
ENSG00000153006	SREK1IP1	-1.75	3.80E-002	17	FALSE
ENSG00000082153	BZW1	-1.41	3.94E-002	19	FALSE
ENSG00000166908	PIP4K2C	-1.11	3.98E-002	5	FALSE
ENSG00000147649	MTDH	-1.51	4.02E-002	18	FALSE
ENSG00000122126	OCRL	-1.55	4.07E-002	5	FALSE
ENSG00000096717	SIRT1	-1.72	4.08E-002	13	FALSE
ENSG00000136827	TOR1A	-0.95	4.11E-002	2	TRUE
ENSG00000104369	JPH1	-1.47	4.11E-002	9	FALSE
ENSG00000172380	GNG12	-1.27	4.12E-002	6	FALSE
ENSG00000197265	GTF2E2	-1.78	4.29E-002	4	FALSE
ENSG00000163946	FAM208A	-1.28	4.36E-002	4	FALSE
ENSG00000138180	CEP55	-1.87	4.53E-002	6	FALSE
ENSG00000111412	C12orf49	-1.14	4.68E-002	2	FALSE
ENSG00000203668	CHML	-2.13	4.70E-002	1	FALSE
ENSG00000151465	CDC123	-1.25	4.75E-002	1	FALSE

Tabelle A8: Putative IGF2BP1-Ziel-RNAs - Fortsetzung

Ensembl Gene ID	Gene symbol	Log ₂ FC	FDR	No. of predicted binding miRs	Part of significantly enriched biological process related to cell migration
ENSG00000073712	FERMT2	-1.54	4.78E-002	9	TRUE
ENSG00000132429	POPDC3	-2.05	4.89E-002	1	FALSE
ENSG00000143333	RGS16	-1.43	4.91E-002	9	FALSE
ENSG00000163291	PAQR3	-1.40	4.98E-002	6	FALSE

Literaturverzeichnis

- [1] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Lehrbuch der Molekularen Zellbiologie*. WILEY-VCH Verlag GmbH & Co. KGaA, 2012.
- [2] G. Anders, S. D. Mackowiak, M. Jens, J. Maaskola, A. Kuntzagk, N. Rajewsky, M. Landthaler, and C. Dieterich. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Research*, 40(D1):D180–D186, 2012.
- [3] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Pallescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, and L. A. Garraway. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–307, 2012.
- [4] J. L. Bell, K. Wächter, B. Mühleck, N. Pazaitis, M. Köhn, M. Lederer, and S. Hüttelmaier. Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cellular and Molecular Life Sciences*, 70:2657–2675, 2013.
- [5] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [6] K. P. Bennett and C. Campbell. Support vector machines: Hype or hallelujah? *SIGKDD Explorations Newsletter*, 2(2):1–13, 2000.
- [7] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.
- [8] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H. Fridman, F. Pagès, Z. Trajanoski, and J. Galon. ClueGO: a Cytoscape plug-in to decipher

- functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*, 25(8):1091–1093, 2009.
- [9] N. Bley, M. Lederer, B. Pfalz, C. Reinke, T. Fuchs, M. Glaß, B. Moller, and S. Hüttelmaier. Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic acids research*, 43:e26, 2015.
- [10] J. Bosch-Barrera and J. A. Menendez. Silibinin and STAT3: A natural way of targeting transcription factors for cancer therapy. *Cancer treatment reviews*, 41(6):540–546, 2015.
- [11] B. Busch, N. Bley, S. Müller, M. Glaß, D. Misiak, M. Lederer, M. Vetter, H. Strauß, C. Thomssen, and S. Hüttelmaier. The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. *Nucleic acids research*, 44(8):3845–3864, 2016.
- [12] E. Candès, L. Demanet, D. Donoho, and L. Ying. Fast Discrete Curvelet Transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.
- [13] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
- [14] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [15] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [16] H. J. Cho, D. S. Suh, S. H. Moon, Y. J. Song, M. S. Yoon, D. Y. Park, K. U. Choi, Y. K. Kim, and K. H. Kim. Silibinin inhibits tumor growth through downregulation of extracellular signal-regulated kinase and Akt in vitro and in vivo in human ovarian cancer cells. *Journal of agricultural and food chemistry*, 61(17):4089–4096, 2013.
- [17] Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(Database issue):D258–D261, 2004.
- [18] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo,

- S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, and P. Flicek. Ensembl 2015. *Nucleic Acids Research*, 43(Database issue):D662–D669, 2015.
- [19] M. Dastpeyman, N. Motamed, K. Azadmanesh, E. Mostafavi, K. Vahid, A. Jahanian-Najafabadi, and M. A. Shokrgozar. Inhibition of silibinin on migration and adhesion capacity of human highly metastatic breast cancer cell line, MDA-MB-231, by evaluation of β 1-integrin and downstream molecules, Cdc42, Raf-1 and D4GDI. *Medical oncology*, 29(4):2512–2518, 2012.
- [20] G. Deep and R. Agarwal. Antimetastatic efficacy of silibinin: molecular mechanisms and therapeutic potential against cancer. *Cancer metastasis reviews*, 29(3):447–463, 2010.
- [21] E. Dimitriadis, T. Trangas, S. Milatos, P. G. Foukas, I. Gioulbasanis, N. Courtis, F. C. Nielsen, N. Pandis, U. Dafni, G. Bardi, and P. Ioannidis. Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. *International journal of cancer*, 121(3):486–494, 2007.
- [22] S. Durinck, P. T. Spellman, E. Birney, and W. Huber. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8):1184–1191, 2009.
- [23] S. Dwivedi, A. L. Khandoga, R. Brandl, and W. Siess. Rac1-mediated signaling plays a central role in secretion-dependent platelet aggregation in human blood stimulated by atherosclerotic plaque. *Journal of translational medicine*, 8:128, 2010.
- [24] O. Dzyubachyk, W. J. Niessen, and E. H. W. Meijering. A Variational Model for Level-set Based Cell Tracking in Time-lapse Fluorescence Microscopy Images. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 97–100. IEEE, 2007.
- [25] E. Eisenberg and E. Y. Levanon. Human housekeeping genes, revisited. *Trends in Genetics*, 29(10):569–574, 2013.
- [26] I. Elcheva, S. Goswami, F. K. Noubissi, and V. S. Spiegelman. CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. *Molecular cell*, 35(2):240–246, 2009.

- [27] F. Entschladen, T. L. Drell, K. Lang, K. Masur, D. Palm, P. Bastian, B. Niggemann, and K. S. Zaenker. Analysis methods of human cell migration. *Experimental cell research*, 307(2):418–426, 2005.
- [28] C. Del Fabbro, S. Scalabrin, M. Morgante, and F. M. Giorgi. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PloS one*, 8:e85024, 2013.
- [29] K. L. Farina, S. Hüttelmaier, K. Musunuru, R. Darnell, and R. H. Singer. Two ZBP1 KH domains facilitate beta-actin mRNA localization, granule formation, and cytoskeletal attachment. *The Journal of cell biology*, 160(1):77–87, 2003.
- [30] K. L. Farina, J. B. Wyckoff, J. Rivera, H. Lee, J. E. Segall, J. S. Condeelis, and J. G. Jones. Cell motility of tumor cells visualized in living intact primary tumors using green fluorescent protein. *Cancer Research*, 58:2528–2532, 1998.
- [31] M. D. Faye, S. T. Beug, T. E. Graber, N. Earl, X. Xiang, B. Wild, S. Langlois, J. Michaud, K. N. Cowan, R. G. Korneluk, and M. Holcik. IGF2BP1 controls cell death and drug resistance in rhabdomyosarcomas by regulating translation of cIAP1. *Oncogene*, 34(12):1532–1541, 2015.
- [32] Y. Fortenberry. The role of serpins in tumor cell migration. *Biological chemistry*, 396(3):205–213, 2015.
- [33] P. Friedl and S. Alexander. Cancer Invasion and the Microenvironment: Plasticity and Reciprocity. *Cell*, 147:992–1009, 2011.
- [34] P. Friedl and E. B. Bröcker. The biology of cell locomotion within three-dimensional extracellular matrix. *Cellular and molecular life sciences*, 57(1):41–64, 2000.
- [35] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews. Molecular cell biology*, 10(7):445–457, 2009.
- [36] P. Friedl and K. Wolf. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nature Reviews Cancer*, 3:362–374, 2003.
- [37] P. Friedl and K. Wolf. Plasticity of cell migration: a multiscale tuning model. *The Journal of Cell Biology*, 188:11–19, 2010.
- [38] C. A. Garrett, M. Barri, A. Kuta, V. Soura, W. Deng, E. M. C. Fisher, G. Schiavo, and M. Hafezparast. DYNC1H1 mutation alters transport kinetics and ERK1/2-cFos

- signalling in a mouse model of distal spinal muscular atrophy. *Brain*, 137(Pt 7):1883–1893, 2014.
- [39] T. Gebäck, M. M. Schulz, P. Koumoutsakos, and M. Detmar. TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays. *Bio-Techniques*, 46(4):265–274, 2009.
- [40] M. Glaß, B. Möller, and S. Posch. Scratch Assay Analysis in ImageJ. In *Proceedings of the ImageJ User and Developer Conference*, volume 1, pages 211 – 214, 2012.
- [41] M. Glaß, B. Möller, A. Zirkel, K. Wächter, S. Hüttelmaier, and S. Posch. Scratch Assay Analysis with Topology-preserving Level Sets and Texture Measures. In *Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA '11)*, number 6669 in LNCS, pages 100–108, 2011.
- [42] M. Glaß, B. Möller, A. Zirkel, K. Wächter, S. Hüttelmaier, and S. Posch. Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines. *Pattern Recognition*, 45(9):3154 – 3165, 2012.
- [43] S. Goswami, R. S. Tarapore, A. M. Poenitzsch Strong, J. J. TeSlaa, Y. Grinblat, V. Setaluri, and V. S. Spiegelman. MicroRNA-340-mediated degradation of microphthalmia-associated transcription factor (MITF) mRNA is inhibited by coding region determinant-binding protein (CRD-BP). *The Journal of biological chemistry*, 290(1):384–395, 2015.
- [44] C. J. Gröger, M. Grubinger, T. Waldhör, K. Vierlinger, and W. Mikulits. Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. *PloS one*, 7(12):e51136, 2012.
- [45] W. Gu, F. Pan, and R. H. Singer. Blocking beta-catenin binding to the ZBP1 promoter represses ZBP1 expression, leading to increased proliferation and migration of metastatic breast-cancer cells. *Journal of cell science*, 122(Pt 11):1895–1905, 2009.
- [46] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, 141(1):129–141, 2014.
- [47] N. A. Hammer, T. v O Hansen, A. G. Byskov, E. Rajpert-De Meyts, M. L. Grøndahl, H. E. Bredekjaer, U. M. Wewer, J. Christiansen, and F. C. Nielsen. Expression of

- IGF-II mRNA-binding proteins (IMPs) in gonads and testicular cancer. *Reproduction*, 130(2):203–212, 2005.
- [48] X. Han, C. Xu, and J. L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):755–768, 2003.
- [49] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.
- [50] F. He and P. Soriano. A critical role for PDGFRalpha signaling in medial nasal process development. *PLoS genetics*, 9(9):e1003851, 2013.
- [51] L. E. Hind, D. Cox, and D. A. Hammer. Two-dimensional motility of a macrophage cell line on microcontact-printed fibronectin. *Cytoskeleton*, 71(9):542–554, 2014.
- [52] M. Hämmerle, T. Gutschner, H. Uckelmann, S. Ozgur, E. Fiskin, M. Gross, B. Skawran, R. Geffers, T. Longerich, K. Breuhahn, P. Schirmacher, G. Stoecklin, and S. Diederichs. Posttranscriptional destabilization of the liver-specific long noncoding RNA HULC by the IGF2 mRNA-binding protein 1 (IGF2BP1). *Hepatology*, 58(5):1703–1712, 2013.
- [53] M. Hopfe, R. Deenen, D. Degrandi, K. Köhrer, and B. Henrich. Host cell responses to persistent mycoplasmas-different stages in infection of HeLa cells with *Mycoplasma hominis*. *PloS one*, 8(1):e54219, 2013.
- [54] S. Hüttelmaier, D. Zenklusen, M. Lederer, J. Dichtenberg, M. Lorenz, X. Meng, G. J. Bassel, J. Condeelis, and R. H. Singer. Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature*, 438(7067):512–515, 2005.
- [55] M. J. Humphries. Cell adhesion assays. *Molecular biotechnology*, 18(1):57–61, 2001.
- [56] I. Huppertz, J. Attig, A. D’Ambrogio, L. E. Easton, C. R. Sibley, Y. Sugimoto, M. Tadjnik, J. König, and J. Ule. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods*, 65(3):274–287, 2014.
- [57] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- [58] A. Kan, R. Chakravorty, J. Bailey, C. Leckie, J. Markham, and M. R. Dowling. Automated and semi-automated cell tracking: addressing portability challenges. *Journal of microscopy*, 244:194–213, 2011.

- [59] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. The KEGG resource for deciphering the genome. *Nucleic acids research*, 32(Database issue):D277–D280, 2004.
- [60] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [61] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36, 2013.
- [62] R. K. Komaravolu, C. Adam, J. R. Moonen, M. C. Harmsen, M. Goebeler, and M. Schmidt. Erk5 inhibits endothelial migration via KLF2-dependent down-regulation of PAK1. *Cardiovascular research*, 105(1):86–95, 2015.
- [63] A. Kozomara and S. Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, 42(Database issue):D68–73, 2014.
- [64] N. Kramer, A. Walzl, C. Unger, M. Rosner, G. Krupitza, M. Hengstschlager, and H. Dolznig. In vitro cell migration and invasion assays. *Mutation research*, 752:10–24, 2013.
- [65] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistic Quarterly*, 2:83–97, 1955.
- [66] S. Lamouille, J. Xu, and R. Derynck. Molecular mechanisms of epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology*, 15(3):178–196, 2014.
- [67] M. R. Lamprecht, D. M. Sabatini, and A. E. Carpenter. CellProfiler: free, versatile software for automated biological image analysis. *BioTechniques*, 42(1):71–75, 2007.
- [68] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, 2012.
- [69] K. Lapidus2007, J. Wyckoff, G. Mouneimne, M. Lorenz, L. Soon, and R. H. Singer. ZBP1 enhances cell polarity and reduces chemotaxis. *Journal of cell science*, 120(Pt 18):3173–3178, 2007.
- [70] D. A. Lauffenburger and A. F. Horwitz. Cell migration: a physically integrated molecular process. *Cell*, 84(3):359–369, 1996.

- [71] J. Lee, S. L. Veatch, B. Baird, and D. Holowka. Molecular mechanisms of spontaneous and directed mast cell motility. *Journal of leukocyte biology*, 92(5):1029–1041, 2012.
- [72] Chun-Chi Liang, Ann Y Park, and Jun-Lin Guan. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nature Protocols*, 2:329–333, 2007.
- [73] Y. Liao, G. K. Smyth, and W. Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 2014.
- [74] M.D. MacManes. On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in genetics*, 5(13):1–7, 2014.
- [75] Y. Maizels, F. Oberman, R. Miloslavski, N. Ginzach, M. Berma, and J. K. Yisraeli. Localization of cofilin mRNA to the leading edge of migrating cells promotes directed cell migration. *Journal of cell science*, 128(10):1922–1933, 2015.
- [76] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. W. Balak, P. Karas, T. Bolcková, M. Štreitová, C. Carthel, S. Coraluppi, N. Harder, K. Rohr, K. E. G. Magnusson, J. Jaldén, H. M. Blau, O. Dzyubachyk, P. Křížek, G. M. Hagen, D. Pastor-Escuredo, D. Jimenez-Carretero, M. J. Ledesma-Carbayo, A. Muñoz-Barrutia, E. Meijering, M. Kozubek, and C. Ortiz de Solorzano. A benchmark for comparison of cell tracking algorithms. *Bioinformatics*, 30(11):1609–1617, 2014.
- [77] M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, 2011.
- [78] Massey, F. J., Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.
- [79] E. Meijering. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, 29(5):140–145, 2012.
- [80] E. Meijering, O. Dzyubachyk, I. Smal, and W. A. van Cappellen. Tracking in cell and developmental biology. *Seminars in cell and developmental biology*, 20(8):894–902, 2009.
- [81] M. B. Menon, N. Ronkina, J. Schwermann, A. Kotlyarov, and M. Gaestel. Fluorescence-based quantitative scratch wound healing assay demonstrating the role of MAPKAPK-2/3 in fibroblast migration. *Cell Motility and the Cytoskeleton*, 66:1041–1047, 2009.

- [82] K. Miura. Tracking movement in cell biology. *Advances in biochemical engineering/biotechnology*, 95:267–295, 2005.
- [83] B. Möller, M. Glaß, D. Misiak, and S. Posch. MiToBo - A Toolbox for Image Processing and Analysis. *Journal of Open Research Software*, 4(1):e17, 2016.
- [84] B. Möller and S. Posch. A Framework Unifying the Development of Image Analysis Algorithms and Associated User Interfaces. In *Proceedings of the 13th IAPR International Conference on Machine Vision Applications (MVA)*, pages 447–450, Kyoto, Japan, May 2013.
- [85] D. J. Munroe and T. J. Harris. Third-generation sequencing fireworks at Marco Island. *Nature biotechnology*, 28(5):426–428, 2010.
- [86] W. S. Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.
- [87] S. Oh, S. P. Jung, J. Han, S. Kim, J. S. Kim, S. J. Nam, J. E. Lee, and J. Kim and. Silibinin inhibits TPA-induced cell migration and MMP-9 expression in thyroid and breast cancer cells. *Oncology reports*, 29(4):1343–1348, 2013.
- [88] A. O. Olarerin-George and J. B. Hogenesch. Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI’s RNA-seq archiv. *Nucleic acids research*, 43(5):2535–2542, 2015.
- [89] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.
- [90] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.
- [91] I. Poola, R. L. DeWitty, J. J. Marshalleck, R. Bhatnagar, J. Abraham, and L .D. Leffall. Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nature medicine*, 11(5):481–483, 2005.
- [92] R. J. Prokop and A. P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing*, 54(5):438–460, 1992.

- [93] B. G. Ricart, B. John, D. Lee, and D. A. Hammer. Dendritic cells distinguish individual chemokine signals through CCR7 and CXCR4. *Journal of immunology*, 186(1):53–61, 2011.
- [94] A. J. Ridley. Life at the leading edge. *Cell*, 145(7):1012–1022, 2011.
- [95] K. Rietscher, A. Wolf, G. Hause, A. Rother, R. Keil, T. M. Magin, M. Glaß, C. M. Niesen, and M. Hatzfeld. Growth retardation, loss of desmosomal adhesion and impaired tight junction function identify a unique role of plakophilin 1 in vivo. *The Journal of investigative dermatology*, 136(7):1471–1478, 2016.
- [96] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.
- [97] M. D. Robinson and A. Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3):1–9, 2010.
- [98] P. Rørth. Collective Cell Migration. *Annual Review of Cell and Developmental Biology*, 25(1):407–429, 2009.
- [99] P. Rørth. Whence directionality: guidance mechanisms in solitary and collective cell migration. *Developmental Cell*, 20(1):9–18, 2011.
- [100] K. R. Rosenbloom, J. Armstrong, Galt P. G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, and W. J. Kent. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, 43(Database issue):D670–D681, 2015.
- [101] Y. Ru, K. J. Kechris, B. Tabakoff, P. Hoffman, R. A. Radcliffe, R. Bowler, S. Mahaffey, S. Rossi, G. A. Calin, L. Bemis, and D. Theodorescu. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic acids research*, 42(17):e133, 2014.
- [102] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair T., Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.

- [103] O. Schmitt and S. Reetz. On the Decomposition of Cell Clusters. *Journal of Mathematical Imaging and Vision*, 33(1):85–103, 2009.
- [104] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, 2012.
- [105] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.
- [106] Y. Shi and W. C. Karl. A Real-Time Algorithm for the Approximation of Level-Set-Based Curve Evolution. *IEEE Transactions on Image Processing*, 17(5):645–656, 2008.
- [107] G. A. Silverman, F. C. Church, P. B. Coughlin, P. G. Gettins, J. A. Irving, D. A. Lomas, C. J. Luke, R. W. Moyer, P. A. Pemberton, E. Remold-O’Donnell, G. S. Salvesen, J. Travis, and J. C. Whisstock. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *The Journal of biological chemistry*, 276(36):33293–33296, 2001.
- [108] B. Song and T. F. Chan. A fast algorithm for level set based optimization. *CAM-UCLA*, 68:02–68, 2002.
- [109] N. Stöhr, M. Köhn, M. Lederer, M. Glaß, C. Reinke, R. H. Singer, and S. Hüttelmaier. IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes and Development*, 26:176–189, 2012.
- [110] C. R. Tessier, G. A. Doyle, B. A. Clark, H. C. Pitot, and J. Ross. Mammary tumor induction in transgenic mice expressing an RNA-binding protein. *Cancer research*, 64(1):209–214, 2004.
- [111] J. P. Thiery, H. Acloque, R. Y. Huan, and M. A. Nieto. Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5):871–890, 2009.
- [112] T. J. Treangen and S. L. Salzberg. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1):36–46, 2011.
- [113] S. Tse, L. Bradbury, J. W. L. Wan, H. Djambazian, R. Sladek, and T. Hudson. A Combined Watershed and Level Set Method for Segmentation of Brightfield Cell Images. volume 7259, pages 72593G–72593G–10, 2009.

- [114] J. Ule, K. B. Jensen, M. Ruggiu, A. Mele, A. Ule, and R. B. Darnell. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science*, 302(5648):1212–1215, 2003.
- [115] C. C. Uphoff and H. G. Drexler. Detecting mycoplasma contamination in cell cultures by polymerase chain reaction. *Methods in molecular biology*, 731:93–103, 2011.
- [116] E. L. van Dijk, Y. Jaszczyszyn, and C. Thermes. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental cell research*, 322(1):12–20, 2014.
- [117] J. Vikesaa, T. V. Hansen, L. Jønson, R. Borup, U. M. Wewer, J. Christiansen, and F. C. Nielsen. RNA-binding IMPs promote cell adhesion and invadopodia formation. *The EMBO journal*, 25(7):1456–1468, 2006.
- [118] R. Visse and H. Nagase. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circulation research*, 92(8):827–839, 2003.
- [119] W. Wang. Study on Cell Cluster Splitting . In *Fourth International Conference on Image and Graphics*, pages 427–431. IEEE, 2007.
- [120] Y. Wang, W. Liang, W. Pan, W. Law, J. Hu, D. T. Ip, M. M. Waye, T. Ng, and D. C. Wang. Silibinin, a novel chemokine receptor type 4 antagonist, inhibits chemokine ligand 12-induced migration in breast cancer cells. *Phytomedicine*, 21(11):1310–1317, 2014.
- [121] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Reviews Genetics*, 10:57–63, 2009.
- [122] A. Wolf, K. Rietscher, M. Glaß, S. Hüttelmaier, M. Schutkowski, C. Ihling, A. Sinz, A. Wingenfeld, A. Mun, and M. Hatzfeld. Insulin signaling via Akt2 switches plakophilin 1 function from stabilizing cell adhesion to promoting cell proliferation. *Journal of cell science*, 126(Pt 8):1832–1844, 2013.
- [123] K. Wu, J. Zeng, G. Zhu, L. Zhang, L. Li, J. Fan, X. Wang, and D. He. Silibinin inhibits prostate cancer invasion, motility and migration by suppressing vimentin and MMP-2 expression. *Acta pharmacologica Sinica*, 30(8):1162–1168, 2009.
- [124] M. Zeisberg and E. G. Neilson. Biomarkers for epithelial-mesenchymal transitions. *The Journal of clinical investigation*, 119(6):1429–1437, 2009.

- [125] D. Zicha, G. A. Dunn, and A. F. Brown. A new direct-viewing chemotaxis chamber. *Journal of cell science*, 99(Pt 4):769–775, 1991.
- [126] S. H. Zigmond. Ability of polymorphonuclear leukocytes to orient in gradients of chemotactic factors. *The Journal of cell biology*, 75(2 Pt 1):606–616, 1977.
- [127] A. Zirkel, M. Lederer, N. Stöhr, N. Pazaitis, and S. Hüttelmaier. IGF2BP1 promotes mesenchymal cell properties and migration of tumor-derived cells by enhancing the expression of LEF1 and SNAI2 (SLUG). *Nucleic acids research*, 41(13):6618–6636, 2013.

Lebenslauf

Persönliche Daten

Name Markus Glaß
geboren am 30.09.1981
in Wolfen
Staatsangehörigkeit deutsch
Anschrift Osendorfer Str. 7, 06112 Halle (Saale)

Schulbildung

Sep. 1987 - Aug. 1990 Diesterweg-Grundschule Bitterfeld
Sep. 1990 - Aug. 1991 Pestalozzi-Grundschule Bitterfeld
Sep. 1991 - Aug. 1992 Gymnasium-Bitterfeld
Aug. 1992 - Jul. 2000 Anne-Frank-Gymnasium Sandersdorf

Zivildienst

Sep. 2000 - Jul. 2001 Diakonie-Sozialstation Sandersdorf

Berufsausbildung

Aug. 2001 - Jul. 2004 Ausbildung zum IT-Systemelektroniker
bei der LMBV mbH, Bitterfeld

Universitätsausbildung

Okt. 2004 - Mär. 2010 Studium der Bioinformatik an der
Martin-Luther-Universität Halle-Wittenberg,
Abschluss Diplom Bioinformatiker (Dipl.-Bioinf.)
seit Apr. 2010 wissenschaftlicher Mitarbeiter an der
Martin-Luther-Universität Halle-Wittenberg,
Institut für Molekulare Medizin, Abteilung
molekulare Zellbiologie

Publikationen

- M. Glaß, B. Möller, A. Zirkel, K. Wächter, S. Hüttelmaier, and S. Posch. Scratch Assay Analysis with Topology-preserving Level Sets and Texture Measures. In *Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA '11)*, number 6669 in LNCS, pages 100–108, 2011
- M. Glaß, B. Möller, A. Zirkel, K. Wächter, S. Hüttelmaier, and S. Posch. Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines. *Pattern Recognition*, 45(9):3154 – 3165, 2012
- M. Glaß, B. Möller, and S. Posch. Scratch Assay Analysis in ImageJ. In *Proceedings of the ImageJ User and Developer Conference*, volume 1, pages 211 – 214, 2012
- N. Stöhr, M. Köhn, M. Lederer, M. Glaß, C. Reinke, R. H. Singer, and S. Hüttelmaier. IGF2BP1 promotes cell migration by regulating MK5 and PTEN signaling. *Genes and Development*, 26:176–189, 2012
- A. Wolf, K. Rietscher, M. Glaß, S. Hüttelmaier, M. Schutkowski, C. Ihling, A. Sinz, A. Wingenfeld, A. Mun, and M. Hatzfeld. Insulin signaling via Akt2 switches plakophilin 1 function from stabilizing cell adhesion to promoting cell proliferation. *Journal of cell science*, 126(Pt 8):1832–1844, 2013
- N. Bley, M. Lederer, B. Pfalz, C. Reinke, T. Fuchs, M. Glaß, B. Moller, and S. Hüttelmaier. Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic acids research*, 43:e26, 2015
- B. Busch, N. Bley, S. Müller, M. Glaß, D. Misiak, M. Lederer, M. Vetter, H. Strauß, C. Thomssen, and S. Hüttelmaier. The oncogenic triangle of HMGA2, LIN28B and IGF2BP1 antagonizes tumor-suppressive actions of the let-7 family. *Nucleic acids research*, 44(8):3845–3864, 2016
- K. Rietscher, A. Wolf, G. Hause, A. Rother, R. Keil, T. M. Magin, M. Glaß, C. M. Niesen, and M. Hatzfeld. Growth retardation, loss of desmosomal adhesion and impaired tight junction function identify a unique role of plakophilin 1 in vivo. *The Journal of investigative dermatology*, 136(7):1471–1478, 2016
- B. Möller, M. Glaß, D. Misiak, and S. Posch. MiToBo - A Toolbox for Image Processing and Analysis. *Journal of Open Research Software*, 4(1):e17, 2016

Eidesstattliche Erklärung / Declaration under Oath

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Datum / Date

Unterschrift des Antragstellers / Signature of the applicant