

Diplomarbeit

**Analyse von Kopienzahlveränderungen in
Einzelzellen mit SC-mFAST-SeqS**

eingereicht von

Thomas Graier

zur Erlangung des akademischen Grades

**Doktor der gesamten Heilkunde
(Dr. med. univ.)**

an der

Medizinischen Universität Graz

ausgeführt am

Institut für Humangenetik

unter der Anleitung von

Assoz. Prof. Priv.-Doz. Dr.med.univ. Jochen Bernd Geigl

und

Assoz. Prof.ⁱⁿ Priv.-Doz.ⁱⁿ Mag.^a Dr.ⁱⁿ rer.nat. Ellen Heitzer

Graz, 23.04.2018

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 23.04.2018

Thomas Graier eh

Danksagungen

Ich möchte mich herzlich bei Herrn Assoz. Prof. Dr. Jochen Geigl bedanken, der mich zu dieser Diplomarbeit motivierte, mir die Möglichkeit gab, erste Einblicke in Klinik, Diagnostik und Erforschung genetischer Erkrankungen zu sammeln und von dem ich enorm viel lernen durfte.

Besonderer Dank geht auch an Frau Assoz. Prof.ⁱⁿ Dr.ⁱⁿ Ellen Heitzer, die meiner Arbeit durch innovative Ideen neue Impulse gab, mir immer mit Rat und Tat zur Seite stand und deren kostbare Zeit ich immer in Anspruch nehmen durfte.

Ein großes Dankeschön gebührt Martina Auer MSc., die mich mit viel Geduld in die Laborarbeit einführte, und deren großartige Hilfe diese Arbeit erst ermöglichte.

Bedanken möchte ich mich auch bei Peter Ulz MSc., der sich viel Zeit nahm, um mir die Grundlagen bioinformatischer Analysen zu erklären, stets für Fragen offen war und dessen Analysen mich immer wieder ins Staunen brachten.

Großer Dank gebührt auch Herrn Univ. Prof. Dr. Michael Speicher, der mir die Möglichkeit gab an seinem Institut mitzuwirken und dieses Projekt durch seine großartige Expertise förderte.

Ein herzliches Dankeschön auch an Julie, Jelena, Christine, Tina und Ricarda für die gute Zusammenarbeit und die abwechslungsreiche Zeit, die wir im Labor hatten und die Arbeit zum Vergnügen machten. Besondere Dank möchte ich auch Sam ausdrücken, mit der ich oft lachen durfte und die mich bei meinen Analysen oft unterstützte.

Ein herzliches Dankeschön gilt auch der Hans und Blanca Moser-Stiftung, die Teile dieser Diplomarbeit unterstützte.

Ein großer Dank gebührt auch meiner Familie, meinem Vater, der meine Neugierde für die Wissenschaft weckte und mich immer unterstützt und ermutigt, meiner Mutter, die in jeder Lebenslage für mich da ist und mir immer Vertrauen schenkt und meinem Bruder, der es versteht, mich in jeder Situation aufzumuntern.

Weiters will ich meiner Oma danken, die mich immer unterstützte und eifrig mitfieberte und mitfeierte, wenn es gut gelaufen war. Danke auch meiner Nonni, mit der ich oftmals philosophische Überlegungen anstellte, und meinem Nonno, der den Abschluss des Studiums leider nicht mehr miterleben konnte.

Ein großer Dank gilt auch der "Chillen Gruppe", die mich wohl besser kennt als ich mich selbst und ein wichtiges Fundament meines Lebens geworden ist. Ich danke jeder und jedem einzelnen von Euch.

Nicht zuletzt möchte ich mich auch bei meiner Kathi bedanken, die es verstand, mich beim Schreiben zu motivieren und die mein Leben täglich erheitert. Liebe Kizi, ich bin froh Dich in meinem Leben zu haben.

Vielen Dank

Zusammenfassung

Einleitung:

Genetische Untersuchungen in Einzelzellen sind Bestandteil intensiver Forschung im Bereich der Tumorbio­logie, aber auch in anderen medizinischen Bereichen wie z.B. der Prä­fertilisations- bzw. der Prä­implantationsdiagnostik. Besonders im Hinblick auf Krebserkrankungen bietet die Analyse von Einzelzellen Einblicke in die Evolution von Tumoren, deren genetischer Heterogenität und die Mechanismen der Metastasierung, die für eine detaillierte Prognose und ein optimiertes Therapieverfahren von enormer Bedeutung sind. Trotz großer Bemühungen mangelt es zurzeit noch an kostengünstigen und praktikablen Methoden, um eine Vielzahl an Zellen zu untersuchen. Daher wurde im Rahmen dieser Diplomarbeit untersucht, ob das „*Single-Cell modified Fast Aneuploidy Screening Test Sequencing System*“ (SC-mFAST-SeqS) eine geeignete Methode darstellt, um in Einzelzellen Kopienzahlveränderungen auf Chromosomenarmebene kostengünstig und einfach darzustellen. Damit könnte die Grundlage für eine breitere Anwendung von Einzelzelluntersuchungen in Forschung und klinischer Diagnostik geschaffen werden.

Methoden:

SC-mFAST-SeqS basiert auf der Analyse von repetitiven DNA-Sequenzen, die sich über das gesamte Genom erstrecken. Damit kann ein einzelnes Primer-Paar über 22.000 verschiedene *Loci* amplifizieren. Durch das Erkennen diskreter Unterschiede dieser Sequenzen können sie auf einzelne Chromosomenarme zurückverfolgt werden. Um die Reproduzierbarkeit der Methode zu testen, wurden Einzelzellen humaner Zelllinien eines Mammakarzinoms, eines Prostatakarzinoms, einer Leukämie sowie einer lymphoblastoiden Zelllinie untersucht. Anhand bioinformatischer Analysen wurde eine Aussage bezüglich Kopienzahlveränderungen getroffen. Die Ergebnisse wurden durch eine weitere Methode, dem "*Somatic Copy Number Aberration Sequencing*" (SCNA-Seq), das eine höherauflösende Methode darstellt, verifiziert.

Ergebnisse:

Für jede der vier Zelllinien wurden mindestens 48 Einzelzellen untersucht, wobei für jede Einzelzelle mindestens 100.000 *Reads* generiert wurden. Mittels hierarchischem *Clustering* und der *Principal Component Analysis* (PCA) konnte

gezeigt werden, dass SC-mFAST-SeqS in der Lage ist, anhand von Kopienzahlveränderungen der Zelllinien, die Populationen klar voneinander zu unterscheiden.

Zusammenfassung:

SC-mFAST-SeqS stellt Kopienzahlveränderungen lediglich auf Chromosomenarmebene dar und erlaubt daher keine Aussage bezüglich Punktmutationen, oder kleiner struktureller Veränderungen. Dennoch erlaubt die in dieser Arbeit getestete Methode mit einem vergleichsweise geringen labortechnischen und bioinformatischen Aufwand eine kostengünstige und Bewertung über eine Vielzahl an Zellen.

Abstract

Introduction:

Analysis of genetic alterations in single cells has become an intensively investigated topic in the field of tumor biology and other medical fields such as prefertilization and preimplantation diagnosis. Especially when it comes to tumors, single cell analysis delivers insight in evolution of tumors, as well as genetic heterogeneity and metastasis. These are crucial factors of influence for detailed prognosis and optimized treatment. Despite enormous effort in inventing new techniques for single cell analysis, methods that enable analysis of great amounts of single cells, are still lacking today. This limits outcome of such experiments.

Therefore the aim of this thesis is to evaluate “single cell modified fast aneuploidy screening test system” SC-mFAST-SeqS as a suitable and cost efficient method for detecting copy number changes on the level of chromosomal arms in single cells. This could be the basis for a broader application of single cell analysis in clinical diagnosis and science.

Methods:

Using SC-mFAST-SeqS single cells of human cell lines, including breast cancer, prostate cancer, leukemia and a lymphoblastoid cell line have been investigated.

SC-mFAST-SeqS is based on a repetitive DNA sequenz distributed all over the genome, yet still able to be traced back to chromosomal arms. This enables a single primer pair to amplify more than 22.000 *loci*. Afterwards bioinformatic analysis allows an estimation of copy number alterations. Results generated with SC-mFAST-SeqS have been verified using “Somatic copy number aberration sequencing” (SCNA-Seq), which provides a higher resolution of copy number changes.

Results:

For every cell line we investigated 48 single cells, generating at least 100.000 *Reads* per sample. Hierarchical clustering and principal component analysis revealed that SC-mFAST-SeqS ist able to distinguish between different cell lines, by analyzing copy number changes in different populations.

Summary:

Since SC-mFAST-SeqS provides information only on the level of chromosomal arms, it is not possible to detect point mutations or small structural aberrations.

Regardless of this fact, SC-mFAST-SeqS is able to analyze a multitude of single cells, by being a cost efficient and practical method.

Inhaltsverzeichnis

Danksagungen	3
Zusammenfassung	6
Abstract.....	8
Inhaltsverzeichnis	10
Glossar und Abkürzungen	12
Abbildungsverzeichnis	13
Tabellenverzeichnis	14
1 Einleitung	15
1.1 Tumorheterogenität.....	16
1.1.1 Entstehung der Tumorheterogenität	19
1.1.2 Klinisches Management von Tumorheterogenität.....	20
1.2 Einzelzellen als Schlüssel zur Erforschung der Tumorheterogenität	22
1.3 Untersuchung von Einzelzellen	22
1.3.1 Single cell sequencing	23
1.4 Kopienzahlveränderungen	25
1.5 Ziel der Diplomarbeit.....	28
2 Material und Methoden	29
2.1 Etablierung von SC-mFAST-SeqS	29
2.1.1 Zellkulturen	29
2.2 FAST-SeqS	29
2.2.1 Optimierung von mFAST-SeqS für Einzelzellen.....	30
2.2.2 Isolierung der Zellen.....	31
2.2.3 Proteinase-K-Verdau	31
2.2.4 Erste Amplifizierung	32
2.2.5 Erste Aufreinigung	33
2.2.6 Zweite Amplifikation	34
2.2.7 Zweite Aufreinigung	35
2.2.8 Lagerung.....	35
2.2.9 Quantifizierung der Amplikon-Library	35
2.2.10 Library Vorbereitung und Sequenzierung	36
2.3 Bioinformatische Auswertung von SC-mFAST-SeqS.....	36
2.4 FACS-Analyse der HL60 Zellen	37
3 Ergebnisse.....	38
3.1 Etablierung von SC-mFAST-SeqS	38
3.2 SC-mFAST-SeqS in Krebszelllinien	38
3.2.1 Untersuchung der humanen Prostatakarzinomzelllinie VCaP mit SC-mFAST-SeqS	38
3.2.2 Untersuchung der humanen Mammakarzinomzelllinie SKBr3 mit SC-mFAST-SeqS.....	40
3.2.3 Untersuchung der Leukämiezelllinie HL-60 mit SC-mFAST-SeqS	42
3.3 Weitere Anwendungen von SC-mFAST-SeqS.....	47
4 Diskussion	50
Literaturverzeichnis	54

Glossar und Abkürzungen

AML	Akute myeloische Leukämie
Array-CGH	Array-based Comparative Genomic Hybridization
ATCC	American Type Culture Collection
CD	Cluster of Differentiation
CNG	copy number gain
CNL	copy number loss
CNV	copy number variation
CPG	cancer predisposition gene
DNA	deoxyribonucleic acid
EDTA	Ethylendiamintetraessigsäure
FACS	fluorescence activated cell sorting
FISH	Fluoreszenz-in-situ-Hybridisierung
g	Erdbeschleunigung
NF-H ₂ O	Nukleasefreies Wasser
PCA	principal component analysis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PK-Verdau	Proteinase-K-Verdau
SC-mFAST-SeqS	single cell-modified fast aneuploidy screening test sequencing system
SCNA	somatic copy number alteration
SCNA-Seq	somatic copy number-sequencing
SCS	single cell sequencing
SNV	single nucleotide variation
TRIS	Tris(hydroxymethyl)aminomethan
WGA	whole genome amplification

Abbildungsverzeichnis

Abbildung 1: Krebsprävalenz, Inzidenz und Mortalität in Österreich (2) (Statistik Austria, 2018, online)

Abbildung 2: Modell zur Tumorevolution (adaptiert nach Alizadeh et al. (26) mit freundlicher Erlaubnis von Springer Nature, April 2018)

Abbildung 3: Schematische Darstellung des Arbeitsablaufes verändert nach Graier et. al (69)

Abbildung 4: Analyse der VCaP-Zellen

Abbildung 5: Heatmap der SKBr3-Einzelzellen

Abbildung 6: Direkter Vergleich des SCNA-SeqS und SC-mFAST-SeqS der SKBr3-Zellen

Abbildung 7: FACS-Gatingstrategie gesorteter Zellen

Abbildung 8: SCNA-Seq und SC-mFAST-SeqS aller HL-60 Zellen

Abbildung 9: Principal component analysis aller HL-60 Zellen

Abbildung 10: Hierarchisches Clustering aller HL-60 Zellen

Abbildung 11: Untersuchungsergebnisse der lymphoblastoiden Zelllinie

Abbildung 12: PCA der weibl. und männl. Kontrollen, sowie der SKBr3-, VCaP-, und lymphoblastoiden Zelllinien

Abbildung 13: Hierarchisches Clustering der männl. Kontrollzellen, sowie der SKBr3-, VCaP-, lymphoblastoiden Zelllinie

Tabellenverzeichnis

Tabelle 1: Thermocyclerprogramm des PK-Verdau

Tabelle 2: Reagenzien für FAST 1, für eine Reaktion

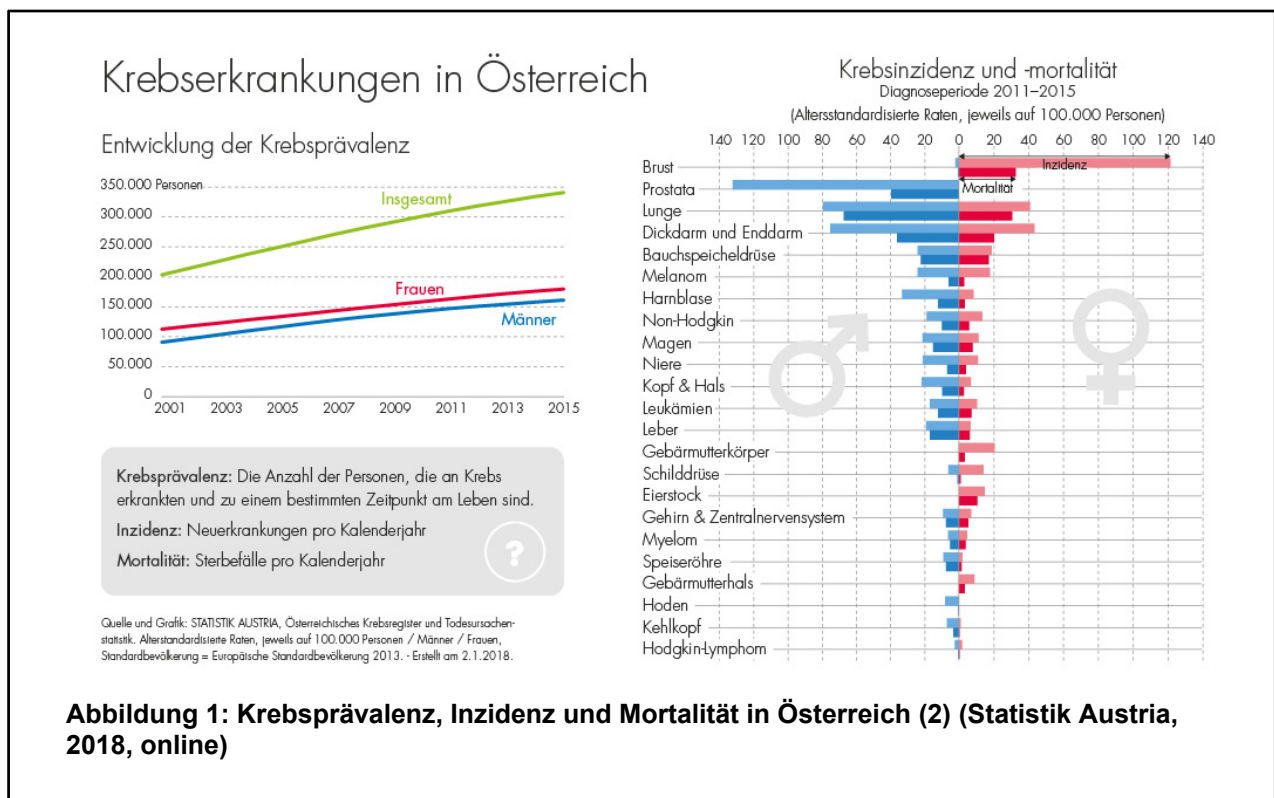
Tabelle 3: Thermocyclerprogramm für FAST 1

Tabelle 4: Reagenzien für FAST 2, für eine Reaktion

Tabelle 5: Thermocyclerprogramm für FAST 2

1 Einleitung

Krebserkrankungen sind mit 28% bei Männern und 23% bei Frauen die zweithäufigste Todesursache der österreichischen Bevölkerung und werden nur von Erkrankungen des Herz-Kreislauf-Systems übertroffen. Darunter gilt das Lungenkarzinom als häufigste krebserkrankungsbedingte Todesursache in beiden Geschlechtern, gefolgt von Mammakarzinomen bei Frauen und Prostatakarzinomen bei Männern (siehe **Abbildung 1**). Die Bedeutung dieser Zahlen wird einem noch mehr bewusst, wenn man bedenkt, dass Krebserkrankungen bei steigender Inzidenz 9,8% der Entlassungsdiagnosen ausmachen und krebsspezifische Krankenhausaufenthalte (ambulant und stationär) die häufigste Ursache eines Krankenhausaufenthaltes darstellen (1).



So ist es nicht verwunderlich, dass die Medizin intensiv an Behandlungsmöglichkeiten forscht und es heute bereits viele Möglichkeiten der Krebstherapien gibt: die Radiotherapie, chirurgische Intervention (2), die heute bereits für manche Tumore roboterassistiert erfolgt (3), sowie die Chemotherapie (4). In den letzten Jahren wurden zusehend neue Therapeutika gefunden, die speziell für gewisse Tumore entwickelt wurden und daher als gezielte Mittel eingesetzt werden (5).

Aufgrund dieser Bemühungen und den ständig besserwerdenden Konzepten der modernen Krebstherapie sinkt die Tumormortalität in Österreich stetig (6), wobei metastasierte Tumore nach wie vor mit einer sehr hohen Mortalität vergesellschaftet sind (7,8).

1.1 Tumorheterogenität

Tumorheterogenität stellt eine der größten Herausforderungen der modernen Onkologie dar, da sie mit Resistenzen gegen Therapien und folglich mit schlechterer Prognose und Tumorprogression einhergeht. Der Begriff „klinische medikamentöse Resistenz“ bezeichnet dabei die Resistenz eines Tumors gegen eine Vielzahl an Chemotherapeutika unterschiedlicher Wirkungsklassen (9).

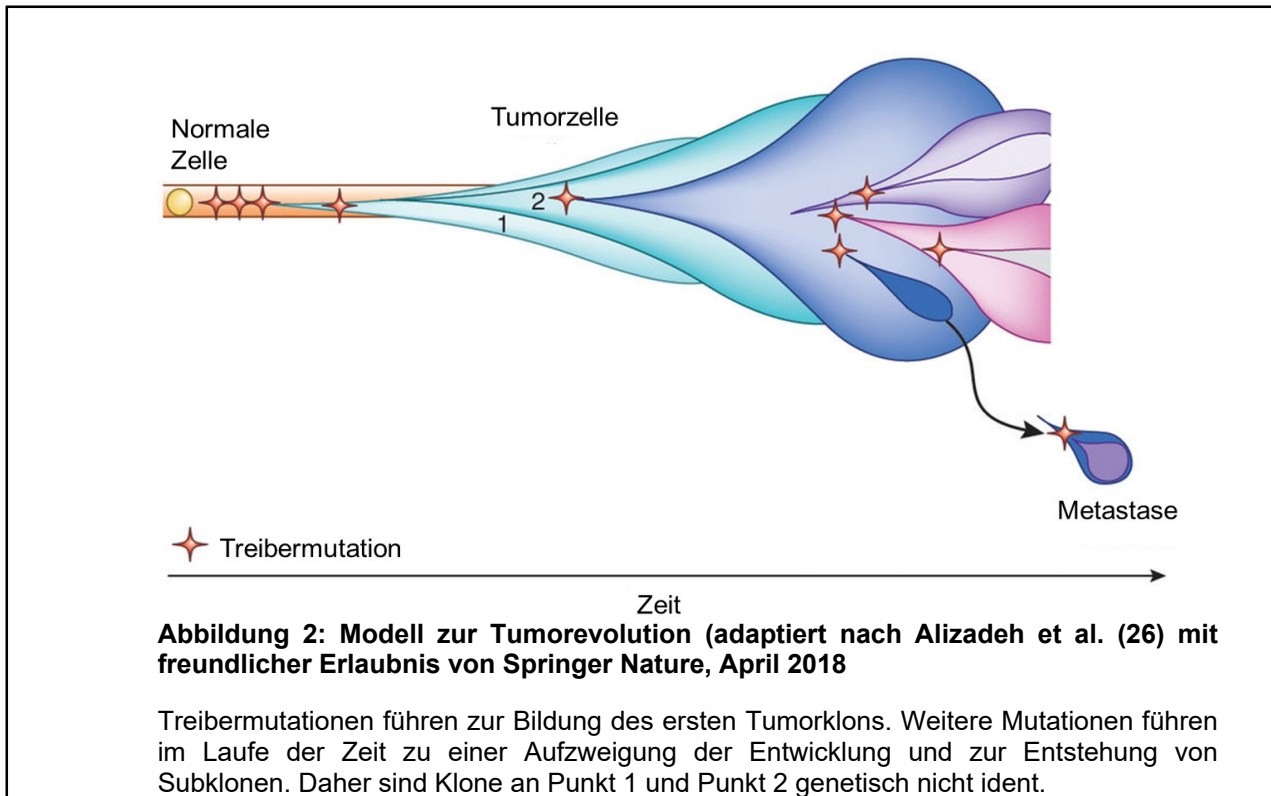
Fälschlicherweise wird Tumorheterogenität oft nur auf die Ebene des Genoms reduziert, obwohl neben den genetischen Veränderungen auch ein verändertes Epigenom, eine veränderte Mikroumgebung des Tumor und ein verändertes Differenzierungsverhalten im Vergleich mit normalen Zellen zu finden sind (10).

Dabei beschreibt der Begriff „intertumorale Heterogenität“ Unterschiede zwischen Tumorerkrankungen einzelner Individuen, während mit „intratumoraler Heterogenität“ Unterschiede in einem einzelnen Individuum gemeint sind. Die intratumorale Heterogenität wird wiederum in eine „spatiale Heterogenität“ (= Unterschiede zwischen zwei Tumorlokalisationen, beispielsweise Karzinom und Metastase), sowie in eine „temporale Heterogenität“ unterteilt (= Änderung von Charakteristika eines Tumors im zeitlichen Verlauf, beispielsweise Verlauf einer Vorläuferläsion zu einem invasiven Karzinom). Diese Unterscheidung bezeichnet jedoch in der Praxis keine strikten Grenzen (11) sondern erleichtert lediglich das Verständnis der Tumor-Situation.

Während Datenbanken wie „*The Cancer Genome Atlas (TCGA)*“ die Forschung an der intertumoralen Heterogenität erlauben, können in derartigen Tumorsammlungen damit nur unbefriedigend die intratumorale Heterogenität untersucht werden (11). Allerdings gibt es andere Ansätze, die dies möglich machen und später noch erläutert werden (12).

Das Lehrbuchmodell der Tumorgenese beschreibt eine Ansammlung von Mutationen, die zur Aktivierung von Onkogenen und Deaktivierung von Tumorsuppressorgenen führen. Durch die vielen Zellzyklen und die genomische Instabilität (13) entwickeln sich Tumore nicht linear, sondern baumartig verzweigt

siehe **Abbildung 2**, wodurch es zur Entstehung mehrerer Klone und



Subpopulationen kommt, die genetisch nicht homogen sind (10). Therapeutische Entscheidungen basieren aber oft nur auf den dominanten Klon, der den Großteil der vorhandenen Tumormasse ausmacht. Gegen diese Therapie können Subpopulationen allerdings resistent sein, was bei der therapeutischen Eradikation des dominanten Klons den Aufstieg der früher unterdrückten Klone erlaubt (14). Interessanterweise konnte man bereits in mehreren unterschiedlichen Krebsarten nachweisen, dass die Mortalität bei einer Vielzahl an von Subpopulationen höher ist, als bei weniger heterogenen Tumoren (15–17). Weiters gibt es Hinweise, dass unterschiedliche Tumor-Subpopulationen sich sogar gegenseitig unterstützen und im Wachstum fördern können (18).

Ähnlich verhält es sich auch bei der Behandlung von Metastasen, die den Hauptgrund der krebsspezifischen Mortalität darstellen, da ihre Therapie oft vom dominanten Klon des Primärtumors direkt auf sie übertragen wird (10). Dabei können Metastasen tatsächlich vom primären Tumorklon stammen (19,20), müssen es aber nicht zwangsläufig (10,21), da Primärtumoren und Metastasen, z.B. bei frühzeitig metastasierenden Tumoren, eine parallele Entwicklung erfahren können (22).

Folglich stellt die genetische Abstammung von Metastasen und die Heterogenität von Metastasen im Vergleich zu Primärtumoren einen sehr interessanten Forschungsansatz dar. Hierzu fanden bereits einige Studien statt, die von Marusyk et al. (10) in einem Review zusammengefasst wurden und die Bedeutung der Untersuchung der Tumorheterogenität klar darstellt. Allerdings entdeckt man durch diese Zusammenfassung auch das breite Spektrum an Ergebnissen, die teilweise eine eindeutige Meinung nicht zulassen. So wurden z.B. bei einer Untersuchung von Pankreastumoren gemeinsame Veränderungen in Primärtumor und Metastase entdeckt, aber auch Metastasen-eigene, wobei Klone der Metastase bereits im Primärtumor nachzuweisen waren (23).

Zwei Untersuchungen von Mammakarzinomen entdeckten gegensätzliche Ergebnisse. So waren sich Primärtumor und Metastase bei einem zeitlichen Abstand zwischen Diagnose des Primärtumors und Auftreten der Metastase von 8 Monaten einmal sehr ähnlich (24). Hingegen waren einer anderen Studie zufolge Primärtumor und Metastase, allerdings bei zeitlichem Abstand von 9 Jahren sehr unterschiedlich (25), wobei Veränderungen des dominanten Klons zumindest deutlich weniger häufig in der Metastase gefunden wurden als man vermutet hätte(10). Anzumerken ist allerdings ebenfalls, dass nicht erwartet werden kann, jede Veränderung in jeder Zelle eines Tumors zu finden (13) und dass das Ausmaß der intratumoralen Heterogenität direkt proportional mit der dem zeitlichen Verlauf ist (11).

Diese Arbeiten lassen eine einfache Erklärung bezüglich der genetischen Variabilität von Metastasen und Rezidiven nicht zu. Jedoch ist es möglich ist, dass Subpopulationen eines Tumors bereits vor Beginn einer Therapie resistent sind und es während der Therapie zu einer Selektion von resistenten Populationen kommen kann (26).

Anderenfalls könnte auch die Therapie Einfluss auf die Entstehung eines Rezidivs haben, wie McGranahan in einem Review zusammenfasst (13). Er sieht die Therapie als möglichen Destabilisierungsfaktor der genomischen Stabilität, da ein AML-Rezidiv mehr Mutationen aufwies als bei Erstdiagnose (27) und selbiges für Gliome, die als Glioblastome rezidierten, nachgewiesen werden konnte (28). Ausgehend von der von ihm bearbeiteten Sachlage schlussfolgert McGranahan, dass die Therapie nicht kausal Mutationen verursacht, jedoch die Selektion eines Klons mit erhöhter genomischer Instabilität triggert (13) bzw. zumindest den

Selektionsdruck dazu erhöht (11). Diese Überlegung führt auch zu einer unkonventionellen Überlegung, der Tumorheterogenität Herr zu werden, die als adaptive Therapie bekannt ist und in Kapitel 1.1.2. erörtert wird (11,13).

1.1.1 Entstehung der Tumorheterogenität

Prinzipiell gibt es heute drei Theorien zur Entwicklung der Tumorheterogenität (11) die im Folgenden kurz beschrieben werden:

- *Klonales Evolutionsmodell:* Postuliert 1976 von Peter Nowell (29). Dabei entstehen aggressivere Klone durch genomische Instabilität und den Selektionsdruck, sobald eine Zelle entartet ist (11,30). Krebs ist daher kein linearer Prozess, sondern eine sich baumartig verzweigende Entwicklung, die in der Ausbildung der Tumorheterogenität gipfelt (11,18).
- *Krebsstammzellmodell:* In diesem von Meacham und Morrison (31) erstmals formulierten Modell stehen ein paar wenige Krebsstammzellen (cancer stem cells, CSC) im Vordergrund, die in der Lage sind zu heterogenen Zellen zu differenzieren (11,30). Haynes et al. (18) glauben, dass Krebsstammzellen das fehlende Puzzlestück sind, um zu erklären wieso sich die Morphologie von Tumoren kaum ändert, obwohl es Subklonen erlaubt ist, ständig neue Überlebensvorteile zu akquirieren. Krebsstammzellen sollen dabei stabilisierend wirken und Tumore zwingen sich in einen speziellen histologischen Typ zu differenzieren. Tumorauslösende Mutationen werden an ihre Nachkommen weitergegeben und der Tumor kann proliferieren, wobei dies kein unidirektionaler Prozess ist und bereits differenzierte Zellen wieder zu Krebsstammzellen werden können. Hinter der Tumorresistenz vermuten sie daher eine Subpopulation an Krebsstammzellen, die therapieresistent sind und folglich halten sie ihre Eradikation für unausweichlich für eine erfolgreiche Therapie (11,18).
- *Tumormikroumgebungsmodell:* Der Einfluss der Tumormikroumgebung auf die Biologie von Tumoren wurde bereits 1982 beschrieben (32) und erlebt nun eine Renaissance, da der wichtige Einfluss der Mikroumgebung des Tumors, auf dessen Charakteristika und Verhalten neuerlich nachgewiesen werden konnte (33,34). Die spezielle Mikroumgebung verhilft dem Tumor zu einem Überlebensvorteil und kann sein Expressionsmuster beeinflussen,

wobei es sich selbst während der Tumorprogression verändert. Die Mikroumgebung selbst ist ebenfalls heterogen und übt um den Tumor verschiedene Selektionsdrücke aus (30). Im besseren Verständnis der Eigenschaften und Funktion der Mikroumgebung von Tumoren erhofft man sich, neue therapeutische Angriffspunkte in der Antikrebstherapie zu finden (11).

1.1.2 Klinisches Management von Tumorheterogenität

Trotz der vielen Probleme, die die Tumorheterogenität für die gezielte Behandlung mit sich bringt, hat die genaue Charakterisierung der inter- und intratumoralen Heterogenität zu potentiellen klinisch verwertbaren Ergebnissen geführt.

So ist es heute möglich, Brusttumore einer Einteilung zuzuführen, die sowohl genetische als auch nicht genetische Merkmale beinhaltet: histopathologische Merkmale, Differenzierungstyp, Hormonrezeptorstatus und Expressionsprofile. Diese Einteilung dient sowohl der therapeutischen Entscheidungshilfe und lässt außerdem prognostische Aussagen zu (26).

Derartige Einteilungen unterliegen freilich einem Wandel durch neue Erkenntnisse und können in immer mehr Subtypen unterteilt werden, was wiederum Potential für personalisierte Medizin schafft (35).

Zardavas et al. (11) haben in ihrem Übersichtsartikel fünf Möglichkeiten postuliert, wie man der inter- und intratumoralen Heterogenität möglicherweise Herr werden kann:

- *„target all of them“*: Kombinationstherapien, die auf Treibermutationen zielen und die Resistenzbildung womöglich vermindern; d.h. die Therapie wird so gewählt, dass möglichst viele Populationen getroffen werden, und nicht nur der dominante Klon, der einen Großteil der Tumormasse ausmacht. Auch „liquid biopsies“ könnten hier einen äußerst interessanten Ansatz darstellen, um die akute Situation von Tumoren verfolgen zu können.
- *„Wenn die Stärke zur Schwäche wird“*: Hochgradige genomische Instabilität ist mit einer Reduktion der Tumorgroße vergesellschaftet (36). Dies soll durch Mutationen geschehen, die keinen Überlebensvorteil bieten und

vorherige Vorteile verschwinden lassen, was Krebszellen zwingt, auf andere Signalwege zu wechseln, die therapeutisch gehemmt werden können. Zusätzlich könnten auch s.g. Neoantigene entstehen, die von Immunzellen als fremd erkannt werden und daher eine Immunreaktion auslösen (37).

- „*Search and destroy*“: Für manche Krebsarten ist bekannt, dass der Klon, der kurz vor dem Tod dominant wird, bereits zu Beginn der Erkrankung als Subpopulation vertreten war (11,38,39). Seine Entdeckung in früheren Stadien könnte eine Therapieadaption erwirken und somit dessen Ausbreitung verhindern. Dies ist diagnostisch sehr aufwendig und auch nicht immer durchführbar (11).
- „*treat to stabilize*“: In diesem Ansatz, der als „adaptive Therapie“ bezeichnet wird, sollen therapiesensitive Zellen erhalten werden, um mit anderen Populationen, die möglicherweise Resistenzen aufweisen, zu konkurrieren und deren Wachstum zu unterdrücken (11). Zitat: „*Elimination of vulnerable cell populations while reducing heterogeneity could revive otherwise dormant or minor subpopulations that restore heterogeneity and transform the tumor into a more resistant and aggressive type. This raises the question if it would be better to maintain stable disease or preserve heterogeneity by not disturbing the tumor ecosystem (preserving homeostasis*“ aus “Breast cancer complexity: implications of intratumoral heterogeneity in clinical management. Cancer Metastasis Rev.2017” von Haynes et al. (18).
- „*target tumor’s home*“: Der Mikroumgebung von Tumoren wird ein großer Einfluss auf die Tumorerhaltung und Progression nachgesagt, durch Angiogenese, Immunzellen der Mikroumgebung, Hypoxie. Dieser Bereich bietet eigene therapeutische Möglichkeiten, die nicht auf den Tumor selbst sondern auf ihn fördernde Prozesse im Körper abzielen und diese unterbinden sollen (11).

1.2 Einzelzellen als Schlüssel zur Erforschung der Tumorheterogenität

Obwohl es bereits große Datenbanken wie beispielsweise TCGA gibt, die es sich zur Aufgabe machten, Tumorgenome umfassend zu charakterisieren, so werden sie heute dennoch als nicht ausreichend empfunden, um als Basis einer intensiven Analyse der intratumoralen Heterogenität dienen zu können. Dies liegt vor allem daran, dass derartige Datensammlungen möglicherweise lediglich eine Art Mittelwert über alle Klone der untersuchten Tumore erfassen (37) und damit folglich in erster Linie den dominanten Klon widerspiegeln und nicht die intratumorale Heterogenität (30). Dies ist auch ein gängiges Problem in der Diagnostik von Tumoren mittels Biopsien, da damit nur schwer das ganze Ausmaß der vorhandenen Tumorheterogenität aufgefunden wird und es folglich oft zu einer Unterschätzung der wirklichen Tumorheterogenität kommt (40).

Deshalb stellt das sogenannte „*single cell sequencing*“ eine sehr vielversprechende Alternative dar, die beim Vergleich mehrerer Einzelzellen einzigartige genetische Veränderungen aufzeigen und zusätzlich zum Verstehen der Phylogenetik der Zellen beitragen kann. Weiters besteht bei „*single cell sequencing*“ die Möglichkeit, ausgewählte Subpopulationen wie zirkulierende Tumorzellen (circulating tumor cells, CTC) (37), oder makroskopisch und mikroskopisch (12,41), bzw. mittels FACS ausgewählte Populationen (42) zu analysieren.

Die Erforschung der Genetik von Einzelzellen ist bereits in viele Bereiche der Medizin vorgedrungen und findet z.B. in der Pränataldiagnostik, in der prädiktiv genetischen Diagnostik (43) und in der Erforschung der Tumorevolution (42,44) Verwendung.

1.3 Untersuchung von Einzelzellen

Für die Untersuchung der DNA in Einzelzellen steht eine Vielzahl an Methoden zur Verfügung. Begonnen bei Karyogrammen und der Bänderung einzelner Chromosomen, über die Einführung der molekularen Zytogenetik mittels FISH, bis hin zur modernen Array-CGH und *Next Generation Sequencing* (NGS) Methoden. Diese Entwicklung ging auch mit einer Veränderung der Größenordnung, mit der genetische Veränderungen in Einzelzellen untersucht wurden, einher. War dies

zunächst noch auf die sichere Identifikation jedes einzelnen Chromosoms und die Entdeckung von Aneuploidien in Karyogrammen beschränkt, konnte die molekulare Zytogenetik bereits komplexere Veränderungen nachweisen und erlaubte Studien von Zellen in Interphase, sowie die Färbung einzelner Chromosomen mittels spektraler Karyotypisierung (43).

Die Array-CGH stellte hier bereits den nächsten großen Entwicklungssprung dar, der eine viel höhere Auflösung mit sich brachte und eine gleichzeitige Erfassung mehrerer Zellen erlaubte (43). Die derzeit höchst auflösende Methode stellt allerdings das NGS dar, welches bereits in der Einzelzellforschung Anwendung gefunden hat (12,41).

1.3.1 Single cell sequencing

Der momentan modernste und hochauflösendste Ansatz zum Nachweis der intratumoralen Heterogenität findet sich heutzutage im Bereich des *single cell sequencing* (SCS).

Kuipers et al. (45) schlussfolgern, dass derzeitige NGS Methoden für die Untersuchung von Bulk-DNA oder von Plasma, zwar Mutationen mit geringer Häufigkeit entdecken und quantifizieren können und man damit auf Tumorerheterogenität aufmerksam wird, der genaue Rückschluss auf die Klonalität des Tumors allein damit aber schwierig sei. Daher würde die Kombination mit SCS durchaus Sinn machen. Angesichts des Problems, inwiefern Einzelzellen die exakte Tumorbilogie wiedergeben, wie es auch bei diagnostischen Biopsien auftritt, könnte man diesem Problem durch die Kombination von NGS und SCS begegnen und die Kosten durch Einschränkung der Anzahl der zu untersuchenden Zellen senken (45).

Das große Potential an SCS wird anhand von CTCs deutlich. CTCs weisen starke genetische Unterschiede untereinander auf und bilden daher das Ausmaß der Heterogenität ab, die aufgrund einer ständigen Änderung, die auch während der Therapie stattfindet, ein dynamisches Geschehen darstellt, weshalb es Sinn macht, CTCs auch während der Therapie zu untersuchen. Aufgrund dieser Unterschiede der Zellen untereinander ist es ratsam, Zellen in Clustern zu ordnen, was wiederum die Analyse einer Vielzahl von Zellen voraussetzt (30,40).

Dies sollte nicht zuletzt auch deshalb erfolgen, da einzelne Zellen besonders in der Untersuchung von SNVs falsch positive und falsch negative Ergebnisse aufweisen können (45).

SCS findet aber auch in soliden Tumoren seine Anwendung (12,46,47), wie vor allem von Nicholas E. Navin mehrfach bewiesen wurde, was ihm auch die Bezeichnung als „Vater des single cell sequencing“ verlieh (9). So untersuchte Navin 2011 mittels SCS die intratumorale Heterogenität in zwei dreifach (triple) negativen Mammakarzinomen, wovon eines als heterogen, das andere als homogen galt.

Für den heterogenen Tumor entdeckte er drei Subpopulationen, die er sehr genau den Regionen des unterteilten Karzinoms zuweisen konnte: Ferner wies er nach, dass diese Klone bereits früh in der Tumorgenese entstanden waren. Für das homogene Karzinom war auch eine Lebermetastase der gleichen Patientin vorhanden, wobei es bei einer Tumorphilung blieb. Dennoch wiesen Tumor und Metastase eigene Veränderungen auf, weshalb sie im hierarchischen Clustering deutlich voneinander differenziert werden konnten (12).

Eine kürzlich veröffentlichte Arbeit von Casasent et al. beleuchtet auf eindrucksvolle Art und Weise die spatiale Korrelation zwischen ductalem Carcinoma in situ (DCIS) und invasivem ductalem Karzinom (IDC) in ein und denselben Patientinnen als direkten Übergang ohne vorherige therapeutische Intervention (47). Sie zeigen die direkte genomische Abstammung der invasiven Karzinome von *in situ* Karzinomen und entdeckten, dass die meisten genetischen Veränderungen bereits vor der Invasion geschehen waren. Dabei wies die Arbeitsgruppe um Casasent eine bis fünf Subpopulationen für jeden Tumor nach und konnte alle Subklone auf einen gemeinsamen Vorfahren zurückführen. Alle gefundenen Subpopulationen waren dabei sowohl im *in situ* Karzinom zu finden, als auch im invasiven Karzinom, was sie dazu veranlasste, eine Entstehung der Subpopulationen bereits vor der Invasion anzunehmen (47).

Was sich jedoch änderte waren die Relationen, mit der sie die jeweiligen Klone vorfanden, was sie auf genetische Veränderungen, die mit Migration assoziieren, zurückführen konnten. Dies lasse nämlich manche Populationen invasiver als andere werden. Dadurch kamen sie zu dem Schluss, dass entgegen der gängigen Theorie, wonach nur ein Klon unter Selektionsdruck invasiv wird (48–50), die Invasion ein multiklonales Geschehen darstellt (47).

Da eine Einzelzelle nur einen DNA-Gehalt von ca. 6,6 Picogramm aufweist, was für eine direkte NGS Analyse nicht ausreichen würde, muss im Vorfeld eine WGA durchgeführt werden. Daraus schlussfolgern Ellsworth et al., dass die Notwendigkeit einer WGA und der dadurch erhöhte bioinformatische Aufwand, sowie die große Anzahl an zu untersuchenden Zellen für aussagekräftige Studien, für die derzeit noch sehr hohen Kosten und die noch fehlende Anwendung in klinischer Routine verantwortlich sind (40).

1.4 Kopienzahlveränderungen

Da Kopienzahlveränderungen die Basis der Analyse der Einzelzellen mittels SC-mFAST-SeqS darstellen, soll an dieser Stelle noch einmal auf ihre allgemeine Bedeutung eingegangen werden.

Der Begriff „Kopienzahlveränderung“ steht für Veränderungen des Genoms bei der eine variable Anzahl an Kopien eines speziellen DNA-Segmentes vorliegen. Diese entsprechen einer Größe von mindestens 50 Basenpaaren bis hin zu mehreren Megabasen (51). So kann es sich um einen „*gain*“ im Sinne eines Zugewinns und Vermehrung der DNA oder auch um einen „*loss*“, was einem Verlust von Erbgut entspricht, handeln. Derartige Veränderungen können bis zu 9,5% des humanen Genoms ausmachen und müssen nicht zwangsläufig phänotypische Veränderungen mit sich bringen (52).

Im Gegenteil sogar, ob und in welchem Ausmaß CNVs für die Prädisposition und Entstehung diverser Krankheiten verantwortlich sind, ist heftig umstritten (52,53). Die Kontroverse bezieht sich einerseits auf die vorhandenen technischen Limitationen (52,54), sowie auf das Vorhandensein von CNVs in gesunden, phänotypisch unauffälligen Individuen (55) und die Schwierigkeiten in der Interpretation derartiger Veränderungen (51).

Mittlerweile wurden aber bereits Zusammenhänge zwischen CNVs und einer Reihe von Erkrankungen gefunden, wie beispielsweise bei Mb. Crohn (56), Typ I Diabetes (53), Dyslipidämie (57), neurologischen Entwicklungsstörungen, neuropsychiatrischen Erkrankungen (58), sowie Tumoren (59). Gerade im Zusammenhang mit Tumoren spielen Kopienzahlveränderungen eine interessante Rolle, da sie einerseits die Prädisposition für die Entstehung von Tumoren erhöhen können (59), andererseits die Tumorevolution und dessen Heterogenität

(12) beeinflussen können, sowie damit indirekten Einfluss auf Progression, Metastasierung und Therapiewirksamkeit ausüben (37).

In der englischen Literatur werden für Kopienzahlveränderungen daher zwei Begriffe unterschieden. Während „CNV (engl. copy number variation)“ meist für die Benennung von genomischen Veränderungen der Keimzellen verwendet wird, werden somatische Kopienzahlveränderungen, wie sie hauptsächlich in Tumoren zu finden sind als „SCNA“ (engl. somatic copy number alteration) bezeichnet (60,61). Trotz dieser verschwommenen Grenze, erleichtert die Trennung des Begriffs Kopienzahlveränderung in SCNA und CNV das prinzipielle Verständnis entsprechender Literatur.

Die genetisch erhöhte Prädisposition zur Entstehung von Tumoren erklärt man sich mit Hilfe der 2-Treffer-Hypothese nach Knudson (62,63). Diese Hypothese geht davon aus, dass innerhalb eines Tumorsuppressorgens Veränderungen stattfinden, die zum Verlust des Genabschnittes oder zu einem nicht funktionierenden Protein führen. Dieses Ereignis entspricht demnach dem „ersten Treffer“. Kommt es daneben zu einer somatischen Mutation des zweiten Allels, gilt dies als „zweiter Treffer“, die in Beisein des „ersten Treffers“ die Entstehung von Tumoren hervorrufen kann (62,63). Für CNV gilt daher, dass sie in einigen seltenen Fällen den „ersten Treffer“ darstellen und folglich die Tumörprädisposition erhöhen (18), obwohl die Mehrheit der CNVs in nicht kodierenden Regionen des Genoms liegt (64). Man vermutet, dass etwa 3% aller Krebserkrankungen auf vererbte Veränderungen in sog. Krebsprädispositionsgenen (Cancer Predisposition Genes, CPG), zu denen man auch Tumorsuppressorgene zählt, zurückzuführen sind (65).

Die Rolle von somatischen Kopienzahlveränderungen wurde von Zack et al. (60) hervorgehoben, die SCNAs in 11 Krebsarten untersuchten. Dabei wurde besonderes Augenmerk auf jene Veränderungen gelegt, von denen schon bekannt war, dass sie tatsächlich zur Tumorevolution und Progression beitragen. Diese Studie ergab, dass SCNAs krebsartübergreifend gehäuft in den gleichen Regionen zu finden sind. Weiters entdeckten sie, dass Zentromer-nahe SCNAs kürzer sind, als ihre Telomer-nahen Pendanten und schlussfolgerten eine unterschiedliche Genese dieser Kopienzahlveränderungen. Sie vermuteten, dass der Ort der SCNAs während der Zellteilung ausschlaggebend für diese unterschiedliche Genese ist und folglich für deren Längendifferenz. Für interne

SCNAs gilt, dass ihre Häufigkeit umgekehrt proportional zur Länge der SCNA war während Telomer-gebundene SCNAs insgesamt häufiger zu finden waren. Prinzipiell schienen große Veränderungen (*whole genome duplications*) die Anzahl an SCNAs zu erhöhen (60). Dies stellt zumindest in Brustkrebs einen wichtigen Schritt in der Tumorevolution dar, da es Gene, die für das Überleben eines Tumors essentiell sind, sehr häufig zu amplifizieren scheint. Außerdem konnten solche Ereignisse in zwei Dritteln der untersuchten Mammakarzinome zumindest einmal als „whole genome duplication“ nachgewiesen werden. Solche Duplikationen treten gehäuft in sehr späten Stadien der Tumorerkrankungen auf und gehen mit massiven Genamplifikationen und Deletionen einher (14).

Die Bedeutung von SCNAs wird auch in einer Arbeit von Wei et al. (66) betont, die nicht nur die Häufigkeit von SCNAs in unterschiedlichen Krebsarten hervorhebt, sondern auch deren direkte biologische Auswirkung in einer Pankrebs-Studie beschreibt (66). Die Arbeitsgruppe um Wei entdeckte, dass mindestens eine CNV in einem von 49 CPG in über 50% der untersuchten Krebsarten - darunter in Prostata-, Pankreas-, Blasen-, Lungen-, Ovar-, Uterus-, Magenkarzinomen sowie Glioblastomen und Sarkomen - vorhanden waren und dabei mehrheitlich als CNL vorlagen. Die enorme Bedeutung der Rolle dieser CNVs in CPGs und besonders der CNL wurde durch Feststellung einer verminderten Genexpression der CPGs, bei Vorliegen einer CNL, in den oben genannten Krebsarten verdeutlicht (66).

Betrachtet man die Bedeutung von Kopienzahlveränderungen für die Tumorbiologie, macht es auch Sinn, selbige in Einzelzellen zu untersuchen. So konnte gezeigt werden, dass man das von Navin et al. (12) untersuchte Mammakarzinom und die korrespondierende Lebermetastase trotz der Monoklonalität des Tumors hauptsächlich aufgrund von SCNAs voneinander unterscheiden kann (12).

Die gefundenen Subpopulationen bei Casasent et al. wiesen größtenteils gemeinsame SCNAs aber auch populationsspezifische Kopienzahlveränderungen auf (47). Bei näherer Beleuchtung der SNVs innerhalb der Populationen, erkannten sie, dass diese, sogar in Tumoren, die bislang als homogen galten, eine noch nähere Einteilung in Subpopulationen erlaube. Diese Beobachtung ließ sie vermuten, dass die SCNAs größtenteils vor den SNVs entstanden sind.

Die Bedeutung der Kopienzahlveränderungen zeigt sich weiters in der veränderten Häufigkeit mit der sich Klone, die SCNAs in Genabschnitten, die mit Migration assoziiert sind, enthalten, in den invasiven Teilen der Karzinome finden (47).

Diese Beobachtung teilen auch Wong et al. Sie wiesen SCNAs als sehr frühe Veränderung der Tumor DNA nach, die in Salven auftreten und quasi als Fundament für eine kontinuierliche Tumorprogression gelten. Dies geschehe vor allem, weil für SCNAs nur wenige Zellteilungen notwendig sind, während SNVs hingegen im Laufe der Zeit mit der Tumorentwicklung und der damit einhergehenden großen Anzahl an Zellteilungen entstehen (46).

1.5 Ziel der Diplomarbeit

Die Entwicklung von NGS hat zu bedeutenden Studien im Bereich der intertumoralen Tumorheterogenität geführt. Vor wenigen Jahren wurde begonnen, damit auch die intratumorale Tumorheterogenität zu erforschen, von der man sich bahnbrechende Ergebnisse für die Antikrebstherapie erhofft und damit einen Schritt in Richtung personalisierter Medizin setzen will. Dies gelang nicht zuletzt durch SCS, welches unser Verständnis über die Evolution von Tumoren nachhaltig geprägt hat. Nach wie vor jedoch sind Einzelzellstudien durch die sehr aufwendige Zellisolierung aus soliden Tumoren, die Notwendigkeit einer WGA, und der komplexen bioinformatischen Auswertung mit erheblichem Aufwand und Kosten verbunden. Dieser Umstand wird durch die hohe Anzahl an zu untersuchenden Einzelzellen zusätzlich verstärkt und limitiert daher die Resultate von Einzelzellstudien und die Einführung dieser Methoden in die klinische Diagnostik. In dieser Diplomarbeit soll SC-mFAST-SeqS daher als Möglichkeit untersucht werden, um Kopienzahlveränderungen in Einzelzellen mittels eines einzelnen Primerpaars und konsekutiver bioinformatischer Auswertung feststellen zu können. Dies könnte eine erhebliche Senkung der Kosten für derartige Studien bedeuten und auch den Computer- und Laboraufwand drastisch reduzieren.

2 Material und Methoden

2.1 Etablierung von SC-mFAST-SeqS

Zur Etablierung von SC-mFAST-SeqS wurden Leukozyten aus peripherem Blut einer gesunden weiblichen Kontrollperson, sowie einer männlichen gesunden Kontrollperson untersucht. Diese sollten als Normalisierung für die im Anschluss untersuchten Zelllinien dienen.

Zur Kontrolle der Ergebnisse der Einzelzellen bei den verwendeten Zelllinien wurde DNA der gesamten jeweiligen Population (bulk DNA) mit SC-mFAST-SeqS sowie der höher-auflösenden SCNA-Seq Methode untersucht. Pro Zelllinie sollten mindestens 48 Zellen untersucht werden, die bei der Sequenzierung mindestens 100.000 *Reads* erreichten.

2.1.1 Zellkulturen

Als Zelllinien wurden folgende humane Zelllinien kultiviert und untersucht: die eines humanen Prostatakarzinoms (VCaP; ATCC® CRL-2876™), eines humanen Mammakarzinoms (SKBr3; ATCC® HTB-30™) und Zellen einer akuten Leukämie (HL-60; ATCC® CCL-240™). Zusätzlich wurde eine lymphoblastoide Zelllinie (zur Verfügung gestellt von Unité de Biologie Fonctionnelle et Adaptative, Paris, Frankreich, bei der eine Trisomie 21 bekannt ist, in die Untersuchungen inkludiert. Alle Zelllinien wurden unter der Anleitung und Aufsicht von Martina Auer MSc. und Christine Beichler MSc. nach ATCC Richtlinien kultiviert.

Analysierte Zellen waren vor ihrer Isolation nicht länger als 4 Wochen in Kultur.

2.2 FAST-SeqS

FAST-SeqS wurde das erste Mal von Kinde et al. (67) publiziert. Ihr Ziel war es, das Screening von fetalen Aneuploidien aus zirkulierender fetaler DNA im Blut der Mutter zu vereinfachen. Bisherige Untersuchungen dieser Art, benötigten einen großen Aufwand bei der Präparation der „*DNA-library*“ sowie bei der Analyse und waren daher mit hohen Kosten verbunden. Laut Autoren wäre eine Aufwands- und Kostenreduktion möglich, wenn es gelänge, eine definierte Menge an Fragmenten, die über das Genom verstreut liegen, mit einem einzelnen Primerpaar zu amplifizieren. Dies würde die Präparation der „*DNA-library*“ deutlich vereinfachen und die nachfolgende bioinformatische Analyse beschleunigen. Tatsächlich gelang

dieser Optimierungsschritt durch die Entdeckung von Primern, die an Regionen binden, die zur Gruppe der sogenannten „*long interspersed nucleotide element-1*“s (*LINE-1*) gehören und es einem einzelnen Primerpaar ermöglichen, geschätzte 22.000 *Loci* zu amplifizieren.. Diese *LINE-1* Retrotransposons haben sich auf das gesamte Genom verteilt, sind jedoch unterschiedlich genug, um sie auf einzelne Chromosomen und, wie nachfolgend beschrieben, sogar auf Chromosomenarmebene zurückführen zu können.

Kinde et al. bezogen FAST-SeqS auf Chromosomenebene (67). Dieser Ansatz wurde auch von Belic et al. (68) verwendet, um den Gehalt an Kopienzahlveränderungen bereits vor Analyse durch eine hochauflösende und teure Methode abzuschätzen und folglich nur ausgewählte Proben gezielt höher auflösenderen Methoden zuzuführen. Die Analyse der als modified FAST-SeqS (mFAST-SeqS) bezeichneten Methode erfolgte auf Chromosomenarmebene.

Während Kinde et al. die erste Amplifizierung nur einmal wiederholten, erhöhten Belic et al. diesen Vorgang auf 4 Wiederholungen. In der zweiten Amplifikation wurden 13 PCR-Zyklen durchgeführt, diese wurden bei dem mFAST-SeqS auf 15 erhöht, wobei für die zweite PCR unterschiedliche Indexprimer verwendet wurden. Die Aufreinigung erfolgte bei beiden Studien mit AMPure Beads, nach Herstellerprotokoll. Beide Studien verwendeten dieselbe Quantifizierungsmethode mittels Agilent Bioanalyzer (Agilent, Santa Clara, USA).

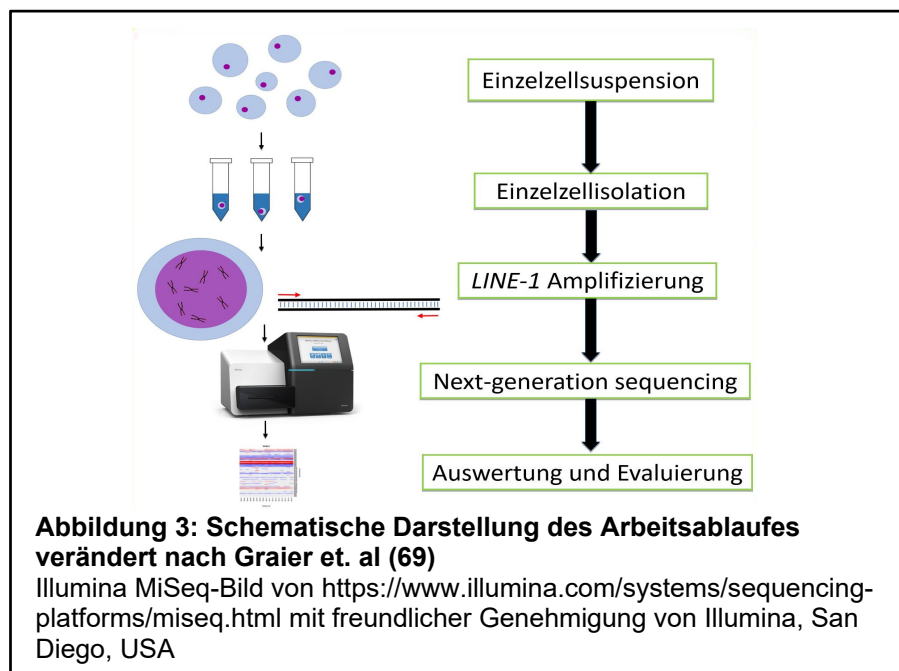
2.2.1 Optimierung von mFAST-SeqS für Einzelzellen

Um unser Vorgehen an die Untersuchung von Einzelzellen anzupassen, modifizierten wir das von Belic et al. (68) bereits adaptierte mFAST-SeqS Protokoll wie folgt:

- Da Belic et al. Plasmaproben untersuchten, mussten wir einen Proteinase-K-Verdau hinzufügen, um mFAST-SeqS für Einzelzellen zugänglich zu machen.
- Weiters wurden die PCR-Zyklen von FAST 1, von 5 auf 8 erhöht.
- Die Reaktionsbedingungen von FAST 2 wurden geändert. Dabei wurden nicht wie bei mFAST-SeqS 15 Zyklen verwendet, sondern die Anzahl der Zyklen für SC-mFAST-SeqS auf 18 erhöht.

Die Bedingungen der Aufreinigung und der verwendeten magnetischen Beads wurden nicht verändert. Wie Belic et al. verwendeten wir auch ein Quantifizierungsverfahren mittels Agilent Bioanalyzer (Agilent, Santa Clara, USA)

mit einem 7500 DNA Kit und ergänzten es um einen High Sensitivity Kit für Proben geringerer Konzentrationen oder mit Artefaktensignalen in der Analyse des 7500 DNA Kits. Bei der Sequenzierung wurden auch unsere SC-mFAST-SeqS *DNA-libraries* mit Plasma-Seq *DNA-libraries* gemischt, wobei zwischen zwei und 5 Plasma-Seq-*DNA-libraries* für die Sequenzierung mit SC-mFAST-SeqS Proben auf der Illumina MiSeq Plattform (Illumina, San Diego, USA) verwendet wurden. Die Präparation der *DNA-library* und die Sequenzierung wurden analog zu Belic et al. durchgeführt. Die bioinformatische Auswertung erfolgte analog der von Belic et al. beschriebenen mFAST-SeqS-Auswertung. Eine schematische Darstellung der Laborarbeit findet sich in **Abbildung 3** (verändert nach Graier et al. (69)). Sämtliche beschriebenen Schritte des Protokolls fanden, sofern nicht anders beziffert, bei Raumtemperatur statt.



2.2.2 Isolierung der Zellen

Die Isolierung der Zellen erfolgte einzeln in 1µL PBS mittels CellCelector (Aviso GmbH, Jena, Deutschland) in Eppendorf-PCR Tubes (Eppendorf Austria GmbH, Wien, Österreich) in dem 10µL NF-H₂O vorgelegt wurden.

2.2.3 Proteinase-K-Verdau

Mittels Proteinase-K-Verdau (PK-Verdau) wird die Extraktion der Nukleinsäuren ermöglicht. Dazu wurde isolierten Zellen 5µL einer 150µg/ml Proteinase-K-Verdünnung hinzugefügt und danach für 5 min bei 14.000g zentrifugiert.

Anschließend erfolgte der Verdau bei 37°C für 60 min, bei 65°C für 10 min, bei 95°C für 10 min und dann auf 4° abgekühlt (siehe **Tabelle 1**) und sofort weiterverarbeitet.

Tabelle 1: Thermocyclerprogramm des PK-Verdaus

PK-Verdau	
Temperatur	Dauer
37°C	60 min
65°C	10 min
95°C	10 min
4°C	∞

Dauer der des PK Verdaus: 100 min

2.2.4 Erste Amplifizierung

Zur Erleichterung des Laborablaufs, wurde die erste Amplifizierung als FAST 1 bezeichnet, die zweite Amplifizierung als FAST 2.

Den verdauten Zellen wurde direkt eine Mischung aus 5-fachem Phusion HF Buffer, 2U Phusion Hot Start II Polymerase, 10µM LINE-1-spezifischer Forward (R1-For) und Reverse (R1-RevA) Primer, die bereits in der Arbeit von Belic et al. beschrieben wurden, und 10mM dNTPs zugefügt (siehe **Tabelle 2**).

Tabelle 2: Reagenzien für FAST 1, für eine Reaktion

FAST 1	
Reagenz	Menge
<i>Probe nach PK-Verdau</i>	15µL
NF-H ₂ O	20,5µL
5× Phusion HF Buffer	10µL
10µM forward Primer (R1-For)	1,25µL
10µM reverse Primer (R1-RevA)	1,25µL
10mM dNTPs	1µL
2U Phusion Hot Start II Polymerase	1µL
Gesamtmenge	50µL

Anschließend wurden Proben folgendermaßen amplifiziert: Denaturierung bei 98 °C für 2 min, im Anschluss 8 Zyklen mit 98°C für je 10 sek, 57°C für 2 min, und 72°C für 2 min. Danach wurde das PCR-Produkt bei 4°C bis zur Weiterverarbeitung gekühlt (siehe **Tabelle 3**).

Tabelle 3: Thermocyclerprogramm für FAST 1

FAST 1		
Temperatur	Dauer	Zyklen
98°C	2 min	In Summe für 8 Zyklen
98°C	10 sek	
57°C	2 min	
72°C	2 min	
4°C	∞	

Dauer der 1. PCR: ca. 50 min

2.2.5 Erste Aufreinigung

Für die Aufreinigung wurden AMPure Beads (Beckman Coulter, Beverly, Massachusetts, USA) verwendet. Diese sollten 30 min vor Beginn aus dem 4°C Kühlschrank entnommen werden, um auf Raumtemperatur gebracht zu werden. Die PCR-Produkte werden dem Thermocycler entnommen und anschließend kurz abzentrifugiert, und in 1,5ml Tubes gebracht.

In der ersten Aufreinigung wurde jedes PCR-Produkt mit 70µL AMPure Beads vermengt und für 10 sek. mit einem Vortexmischer bei hoher Stufe gemixt. Anschließend erfolgte eine Inkubation für 10 min, danach wurden die Tubes für weitere 5 min auf einem Magneten platziert. Anschließend wurde der Überstand vorsichtig abgehoben, sodass die Beads nicht von der Wand des Tubes gelöst wurden und verworfen. Die Beads, die sich nun am Rande des Tubes befinden, werden mit 200µL 70% Ethanol gewaschen und für 1 min inkubiert. Der Alkohol wird abgehoben und verworfen. Dieser Schritt wird ein zweites Mal wiederholt. Nach der zweiten Waschung erfolgt die Trocknung der Beads unter Vermeidung direkter Lichteinwirkung für 5-10 min, bis sichtbare Alkoholrückstände verdunstet sind. Im Anschluss werden die Beads in 12µL 1x Tris-EDTA-Buffer resuspendiert, für 2 min inkubiert und danach für 2 min auf den Magneten gestellt. Im letzten Schritt der ersten Aufreinigung werden 10µL abgehoben und in ein neues PCR-

Tube zur weiteren Verarbeitung transferiert. Das Tube mit den verbleibenden Beads wird verworfen.

Dauer der ersten Aufreinigung: ca. 35 min

2.2.6 Zweite Amplifikation

Das aufgereinigte Produkt der ersten PCR wurde erneut einer PCR unterzogen, die als „FAST 2“ bezeichnet wird. Dabei werden die sog. Indexprimer, die die Trennung einzelner Proben beim Sequenzieren ermöglichen, verwendet. Die Menge des nukleasefreien Wassers wird auf 26µL erhöht. Als Forward-Primer werden FAST 2-spezifische Primer verwendet. Wesentlich sind die Reverse-Primer, wovon 46 zur Auswahl stehen. Diese ermöglichen die Wiedererkennung in der bioinformatischen Auswertung und erlauben damit das gleichzeitige Sequenzieren mehrerer SC-mFAST-SeqS Proben, wobei keine Proben mit gleichen Indexprimern sequenziert werden dürfen, da eine genaue Zuordnung sonst nicht mehr möglich ist. Ansonsten wurden keine weiteren Änderungen hinsichtlich der Reagenzien im Vergleich zu FAST 1 vorgenommen (siehe **Tabelle 4**).

Tabelle 4: Reagenzien für FAST 2, für eine Reaktion

FAST 2	
Reagenz	Menge
<i>Elution von FAST 1</i>	10µL
NF-H ₂ O	26µL
5× Phusion HF Buffer	10µL
10µM forward Primer (R2-For)	1,25µL
10µM reverse Primer (R2-RevA)- <u>Indexprimer</u>	1,25µL
10mM dNTPs	1µL
2U Phusion Hot Start II Polymerase	1µL
Gesamtmenge	50µL

Die Bedingungen der PCR-Zyklen wurden ebenfalls angepasst (siehe **Tabelle 5**) und dabei die Anzahl an Zyklen in der 2. PCR auf 18 erhöht. Im Anschluss wurde das PCR-Produkt bei 4°C bis zur Weiterverarbeitung gekühlt.

Dauer der 2. PCR: 40 min

Tabelle 5: Thermocyclerprogramm für FAST 2

FAST 2		
Temperatur	Dauer	Zyklen
98°C	2 min	In Summe für 18 Zyklen
98°C	10 sek	
65°C	15 sek	
72°C	15 sek	
4°C	∞	

Dauer der 2. PCR: ca. 30 min

2.2.7 Zweite Aufreinigung

Die zweite Aufreinigung erfolgt analog zur ersten mit 70µL AMPure Beads (Beckman Coulter, Beverly, Massachusetts, USA) (siehe 2.2.5).

Dauer der zweiten Aufreinigung: ca. 35 min

2.2.8 Lagerung

Die Lagerung der verarbeiteten SC-mFAST-SeqS Proben, sowie der isolierten Einzelzellen erfolgt bei -20°C.

2.2.9 Quantifizierung der Amplikon-Library

Die PCR-Produkte wurden auf einem Agilent 2100 Bioanalyzer (Agilent, Santa Clara, USA) auf einem Agilent DNA 7500 Kit (Agilent, Santa Clara, USA) laut Protokoll des Kits quantifiziert. Dabei wurden PCR-Produkte mit einer Konzentration von < 8nmol/L bzw. mit Artefaktsignalen auf einem Agilent DNA High-Sensitivity Kit (Agilent, Santa Clara, USA) wiederholt quantifiziert.

Dauer der Quantifizierung des Bioanalyzers bei voller Probenauslastung: ca. 1 Stunde

2.2.10 Library Vorbereitung und Sequenzierung

Nur Proben mit einer Konzentration von $> 3,5\text{nmol/L}$ wurden zur Sequenzierung zugelassen. Die L1-Amplicon-DNA-libraries wurden analog dem Protokoll von Belic et al. in äquimolaren Mengen zusammengeführt und sequenziert, wobei *single Reads* in der Länge von 150 Basenpaaren erzeugt wurden. Um die Komplexität der Basenzusammensetzung der Library-Fragmente auf der Flow Cell zu erhöhen, wurden SC-mFAST-SeqS immer gemeinsam mit Plasma-Seq Libraries sequenziert, wobei für jeden MiSeq-Lauf mindestens 2, maximal 5 Plasma-Seq verwendet wurden. Verhältnisse bei der Sequenzierung lagen daher zwischen 2 Plasma-Seq plus 40 SC-mFAST-SeqS Libraries und 5 Plasma-Seq plus mit 10 SC-mFAST-SeqS mFAST-SeqS. Die Sequenzierung erfolgte auf einem Illumina MiSeq (Illumina, San Diego, USA).

2.3 Bioinformatische Auswertung von SC-mFAST-SeqS

Die Auswertung von SC-mFAST-SeqS erfolgte anhand der von Belic et al. beschriebenen Verfahren (68). SC-mFAST-SeqS basiert auf der Amplifizierung von *LINE1*-Sequenzen, die über das gesamte Genom verteilt zu finden sind.

Die gewonnenen *Reads* werden mittels *Burrows-Wheeler-Aligner* an das hg19 Referenzgenom angeglichen. Ein selbst erstelltes Skript zählt nur jene *Reads*, die einer Zuordnungsqualität von 15 pro Chromosomenarm entsprechen.

Die Anzahl an *Reads* pro Chromosomenarm wurde mit der Gesamtanzahl an *Reads* pro Probe normalisiert. Dabei spiegelt der Quotient den prozentuellen Anteil der *Reads* eines Chromosomenarms mit den *Reads* der gesamten Probe wieder.

Eine Subtraktion des Quotienten der Proben sowie des Quotienten der Kontrollzellen und eine anschließende Division durch die Standardabweichung der normalisierten *Read*-Anzahl jedes Chromosomenarms der Kontrollzellen ermöglicht die Berechnung von s.g. „z-Werten“. Veränderungen wurden von uns ab einem z-Wert von ± 3 als „wahrscheinlich vorhanden“, ab ± 5 als „sehr wahrscheinlich vorhanden“ definiert.

Die kurzen Arme der akrozentrischen Chromosomen 13p, 14p, 15p, 21p, 22p und Y wurden ausgeschlossen, da ihnen keine *Reads* zugeordnet werden können.

Zur leichteren Darstellung der Veränderungen wurden mit den vorhandenen Daten mit Programm R, Version 3.2.3 eine *Heatmap* erstellt.

2.4 FACS-Analyse der HL60 Zellen

FACS wurde zu einem weltweit akzeptierten Standard in der Analyse und Sortierung von Zellpopulationen und ist besonders im Bereich der Hämatologie und Hämatookologie ein wichtiges Werkzeug. In dieser Arbeit wurde die Durchflusszytometrie und Zellsortierung mit einem FACS Aria (BD Biosciences, San Jose, USA) an der Abteilung Core Facility Flow Zytometrie, der Medizinischen Universität Graz, unter der Leitung von Mag. Dr.rer.nat. Heimo Strohmaier durchgeführt.

Die Sortierung der Zellen erfolgte in folgende 4 Gruppen nach vorheriger FACS Analyse durch ao.Univ.-Prof. Dr. med. univ. Albert Wölfler, Klinische Abteilung für Hämatologie, Medizinische Universität Graz:

1. CD45+ (unspezifisch)
2. CD45+/ CD11b-/ CD117+
3. CD45+/CD117-, CD11b+
4. nur CD45+ (CD117-/CD11b-)

Dazu wurden die Zellen von Frau Jenniger Ober, BSc., in 96-Well-Platten, in denen 10 µL NF-H₂O vorbereitet waren, einzeln oder in 10er Gruppen sortiert. Sortierte Zellen wurden maximal einmal eingefroren, ehe sie SC-mFAST-SeqS unterzogen wurden.

3 Ergebnisse

3.1 Etablierung von SC-mFAST-SeqS

Es wurden 48 Kontrollzellen einer Frau und 57 Kontrollzellen eines Mannes untersucht, deren Zellen alle nach der Sequenzierung eine *Read*-Gesamtanzahl von über 100.000 *Reads* pro Kontrolle aufwiesen. Im Durchschnitt waren es 194.120 *Reads* bei weiblichen Kontrollen (Maximum war SCf_38 mit 369.687 *Reads*) und 181.988 *Reads* bei männlichen Kontrollen (Maximum war EZmale44_S9 mit 474.757 *Reads*).

Mittelwerte und Standardabweichung pro Chromosomenarm wurden für weibliche und männliche Kontrollen getrennt berechnet und dienten der Normalisierung der im Anschluss untersuchten Krebszellen. Wie erwartet, zeigten bei einem z-score Cut-off von 5 alle Kontrollen balancierte Profile. Weder bei den weiblichen, noch bei den männlichen Kontrollzellen konnte mit dieser Methode eine chromosomenarmspezifische Veränderung gefunden werden, die mit einem z-score von über ± 5 mehr als einmal in der Gesamtheit der jeweils untersuchten Zellen auftrat.

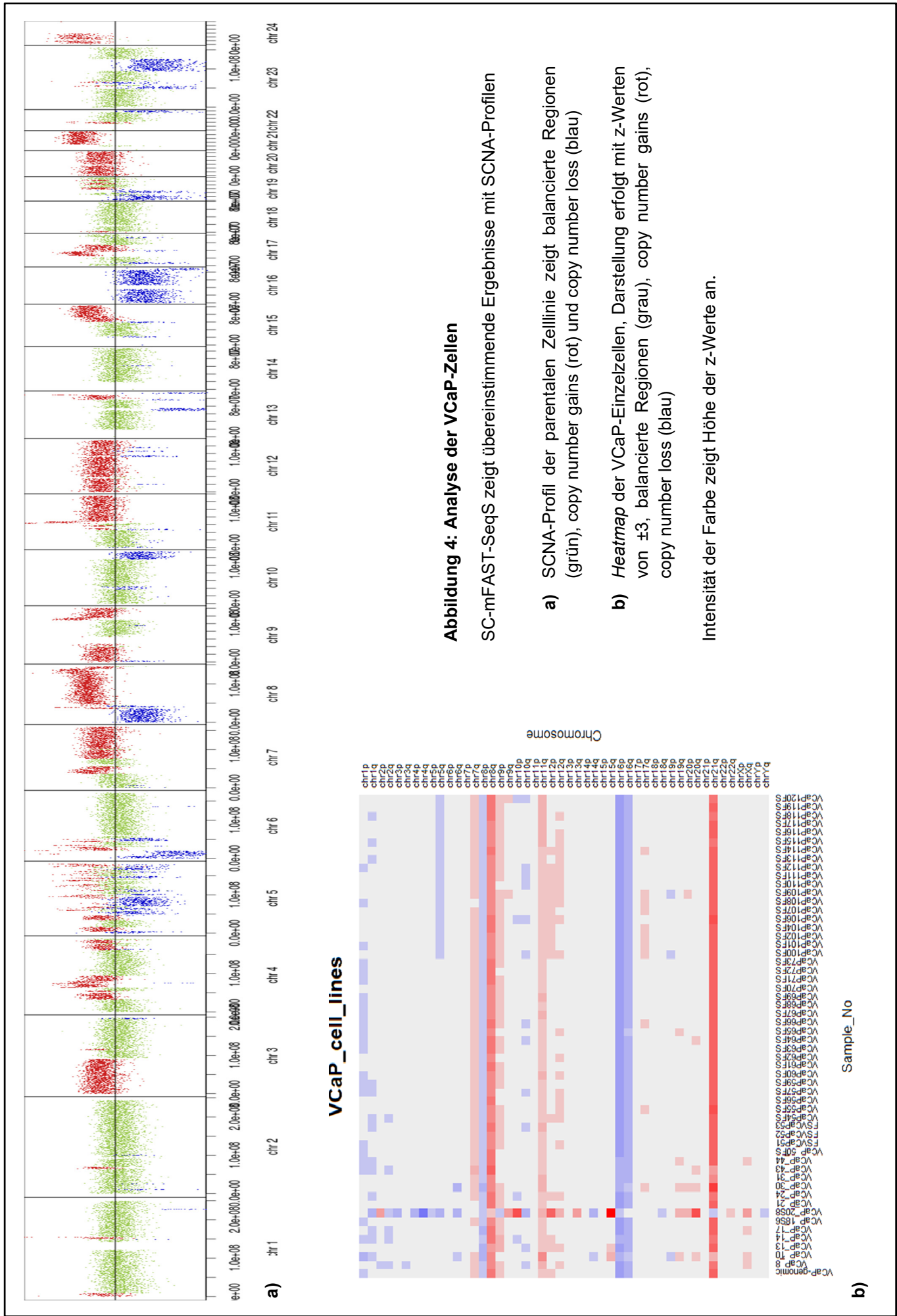
3.2 SC-mFAST-SeqS in Krebszelllinien

Um die diagnostische Sensitivität hinsichtlich der Detektion von Kopienzahlveränderungen zu testen, wurden Einzelzellen von Zelllinien mit bekannten Kopienzahlveränderungen untersucht.

. Hierfür wurden die Krebszelllinien VCaP, SKBr3 und HL-60., untersucht. Die Validierung der erhaltenen Kopienzahlprofile erfolgte mittels SCNA-Seq.

3.2.1 Untersuchung der humanen Prostatakarzinomzelllinie VCaP mit SC-mFAST-SeqS

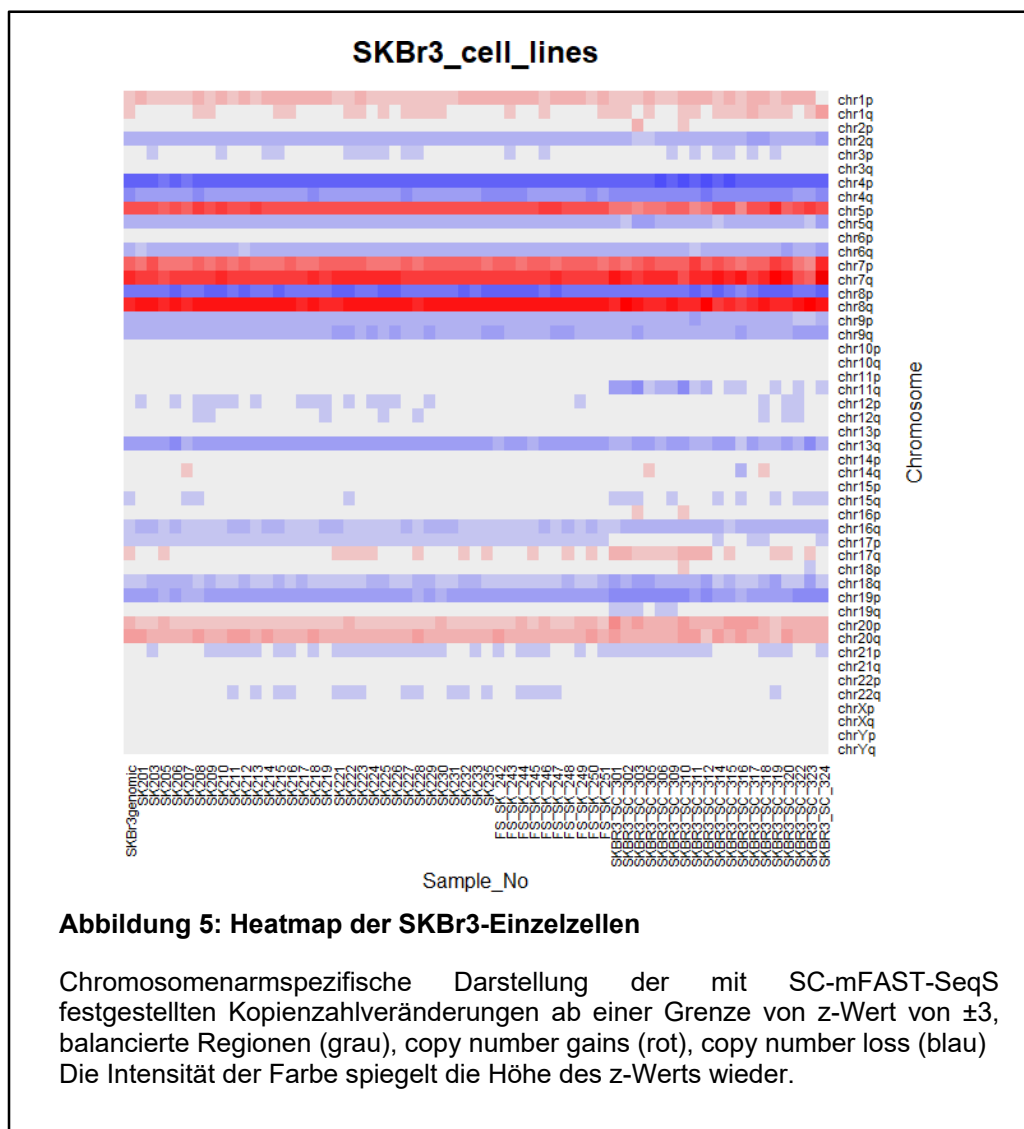
Wir konnten für 55 Zellen eine *Read*-Gesamtanzahl von über 100.000 mit einer durchschnittlichen Gesamtanzahl von 289.983 *Reads* pro untersuchter Probe (Maximum war VCaP67FS mit 733.215 *Reads*) erzielen (**Abbildung 4**). Die untersuchten Zellen wiesen hochkonkomidante Ergebnisse innerhalb der mit SC-mFAST-SeqS untersuchten Zellen auf. Zudem konnte die Mehrzahl der mittels dem SCNA-Seq detektierten Kopienzahlveränderungen auch mit SC-mFAST-SeqS nachgewiesen werden.



3.2.2 Untersuchung der humanen Mammakarzinomzelllinie SKBr3 mit SC-mFAST-SeqS

Ähnliches gilt für die in **Abbildung 5** dargestellten Einzelzellen der SKBr3 Zelllinie. 52 Zellen erfüllten die von uns gesetzten Kriterien mit einem Mittelwert von 196.234 Reads (Maximum war FS_SK_249 mit 846.854 Reads).

Auch hier wiesen die untersuchten Zellen sehr ähnliche Ergebnisse auf und Veränderungen mit hohen z-Werten konnten durch SCNA-Seq bestätigt werden.



SCNA-Seq der SKBr3 Zellen ist in **Abbildung 6** dargestellt.

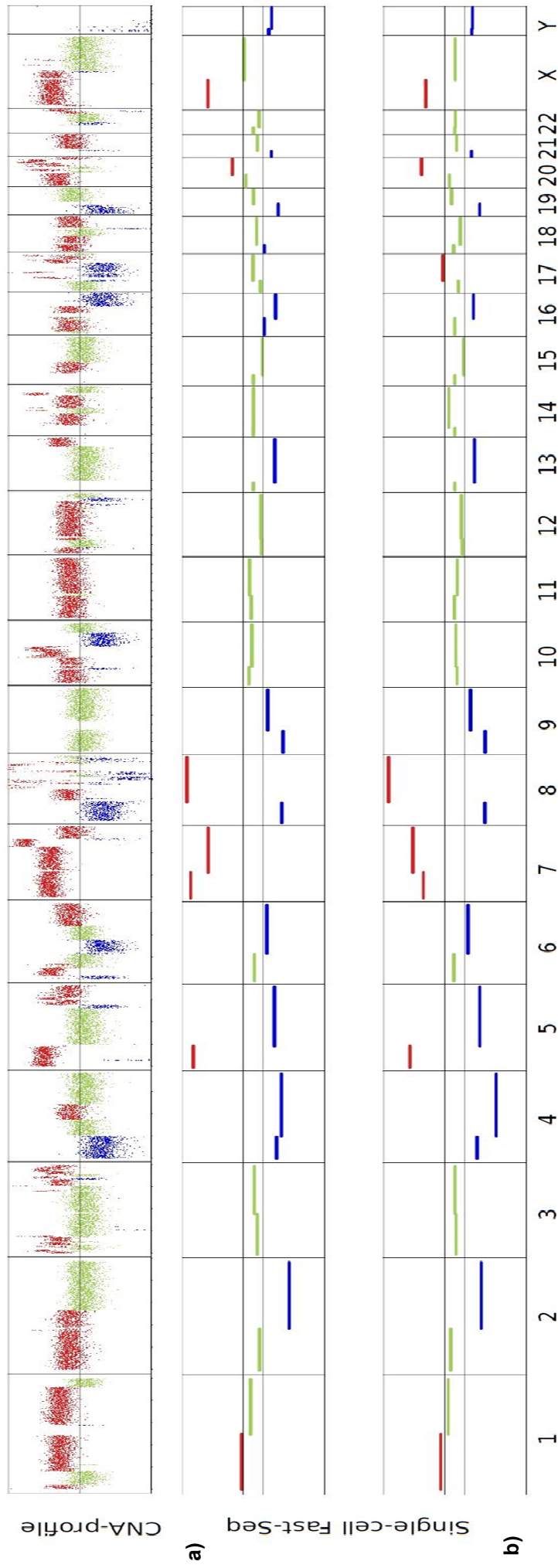


Abbildung 6: Direkter Vergleich des SCNA-SeqS und SC-mFAST-SeqS der SKBr3-Zellen

SC-mFAST-SeqS und SCNA-SeqS zeigen übereinstimmende Ergebnisse für große Kopienzahlveränderungen:

- a) SCNA-Profil der parenteralen SKBr3 Zelllinie zeigt balancierte Regionen (grün), copy number gains (rot) und copy number loss (blau)
- b) SC-mFAST-SeqS der parenteralen SKBr3-Zelllinie, balancierte Regionen (grau), copy number gains (rot) und copy number loss (blau)

3.2.3 Untersuchung der Leukämiezelllinie HL-60 mit SC-mFAST-SeqS

Da die Vereinzelung von Zellen in soliden Tumoren spezieller Methoden bedarf, wollten wir testen, ob sich SC-mFAST-SeqS eher für Krebszellen eignet, die keiner enzymatischen oder mechanischen Vorbehandlung bedürfen. Daher entschieden wir uns, die Zelllinie HL-60 zu analysieren, die von einer akuten Leukämie stammt. Nachdem SC-mFAST-SeqS in der Lage war Kontrollzellen, sowie VCaP- und SKBr3-Einzelzellen klar voneinander zu trennen (siehe **Abbildung 12 & 13**), stellten wir uns die Frage, ob dies auch für Subpopulationen in Zelllinien der Fall ist.

Abbildung 7 zeigt die Ergebnisse der FACS-Analyse bei Sortierung der Zellen.

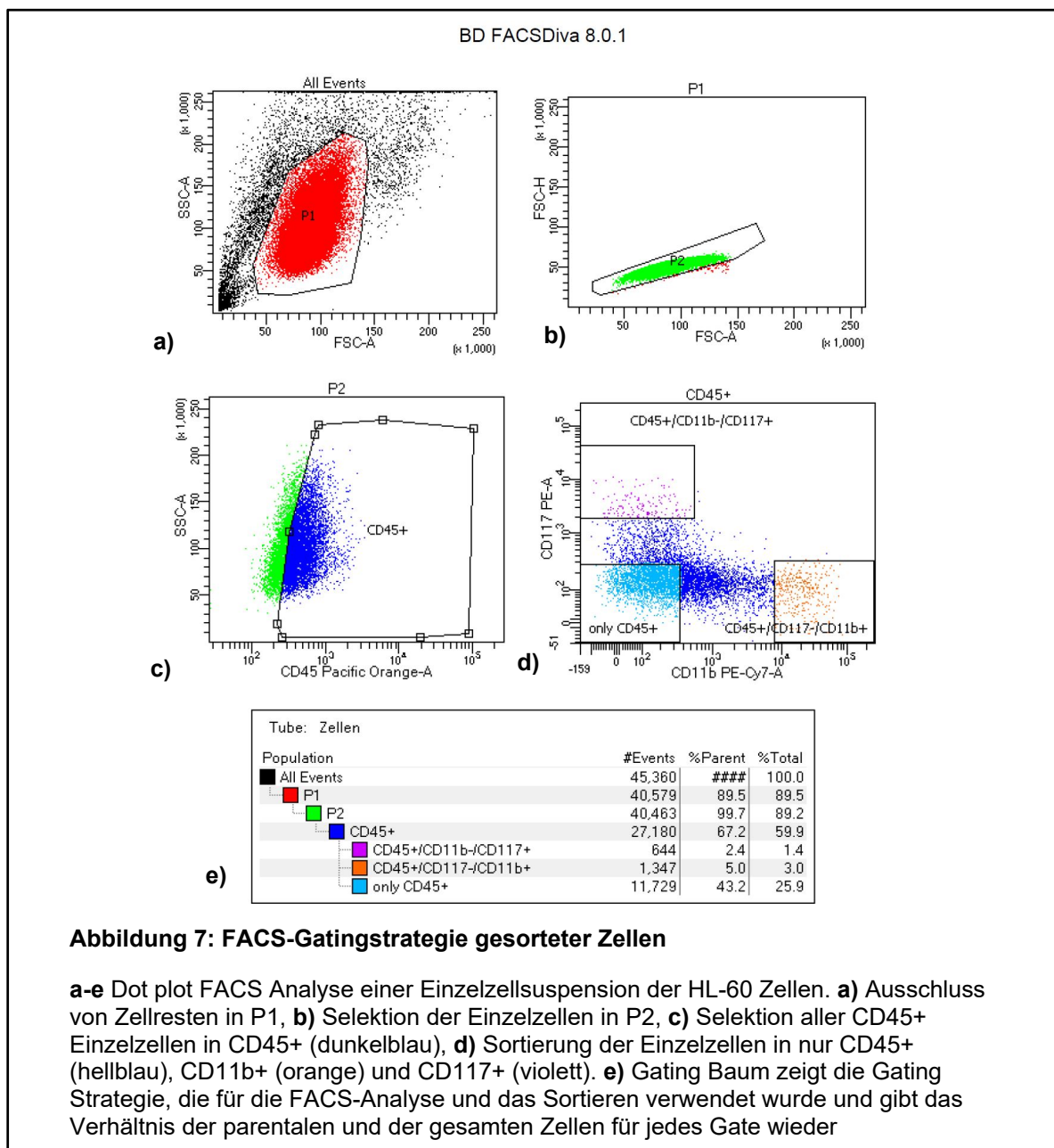


Abbildung 7: FACS-Gatingstrategie gesorteter Zellen

a-e Dot plot FACS Analyse einer Einzelzellsuspension der HL-60 Zellen. **a)** Ausschluss von Zellresten in P1, **b)** Selektion der Einzelzellen in P2, **c)** Selektion aller CD45+ Einzelzellen in CD45+ (dunkelblau), **d)** Sortierung der Einzelzellen in nur CD45+ (hellblau), CD11b+ (orange) und CD117+ (violett). **e)** Gating Baum zeigt die Gating Strategie, die für die FACS-Analyse und das Sortieren verwendet wurde und gibt das Verhältnis der parentalen und der gesamten Zellen für jedes Gate wieder

SC-mFAST-SeqS der HL-60 Zellen wurde in folgenden 5 Gruppen durchgenommen, diese entsprechen den 4 Gruppen, die bereits in Abschnitt 2 beschrieben wurden, die letzte Gruppe stellt HL-60 Einzelzellen dar, die nicht gesortet wurden, sondern direkt aus Kultur isoliert wurden.

1. HL-60 Kultur: Hier konnten 13 Proben unsere Kriterien erfüllen, mit einer durchschnittlichen *Read* Anzahl von 305.043 (Maximum war HL60_20 mit 368.571 *Reads*).
2. CD45+ (unspezifisch): Aus dieser Gruppe konnten 27 Zellen analysiert werden mit einem Mittelwert von 232.084 *Reads* pro Probe (Maximum war all45_HL60_1.4 mit 340.337 *Reads*).
3. CD45+/ CD11b-/ CD117+: 28 Proben erfüllten die Analysevoraussetzungen mit durchschnittlich 294.586 *Reads* pro Probe (Maximum war 117_HL60_10_16 mit 613.134 *Reads*)
4. CD45+/CD117-, CD11b+: 28 Proben konnten analysiert werden, die durchschnittliche *Read* Anzahl beträgt 200.770 *Reads* (Maximum war mit 391.971 *Reads* bei Probe 11b_HL60_1_13).
5. CD45+ (CD117-/CD11b-): Für diese gesorteten Zellen liegen keine Ergebnisse vor. Nach Beendigung des SC-mFAST-SeqS-Protokolls konnte für keine der bearbeiteten Zellen ein ausreichendes PCR-Produkt in der Quantifizierung gefunden werden.

Abbildung 8 zeigt SCNA-Seq sowie SC-mFAST-SeqS Ergebnisse aller HL-60 Zellen, **Abbildung 9** die *Principal Component Analysis* aller HL-60 Zellen und **Abbildung 10** das hierarchische Clustering aller analysierten HL-60 Zellen.

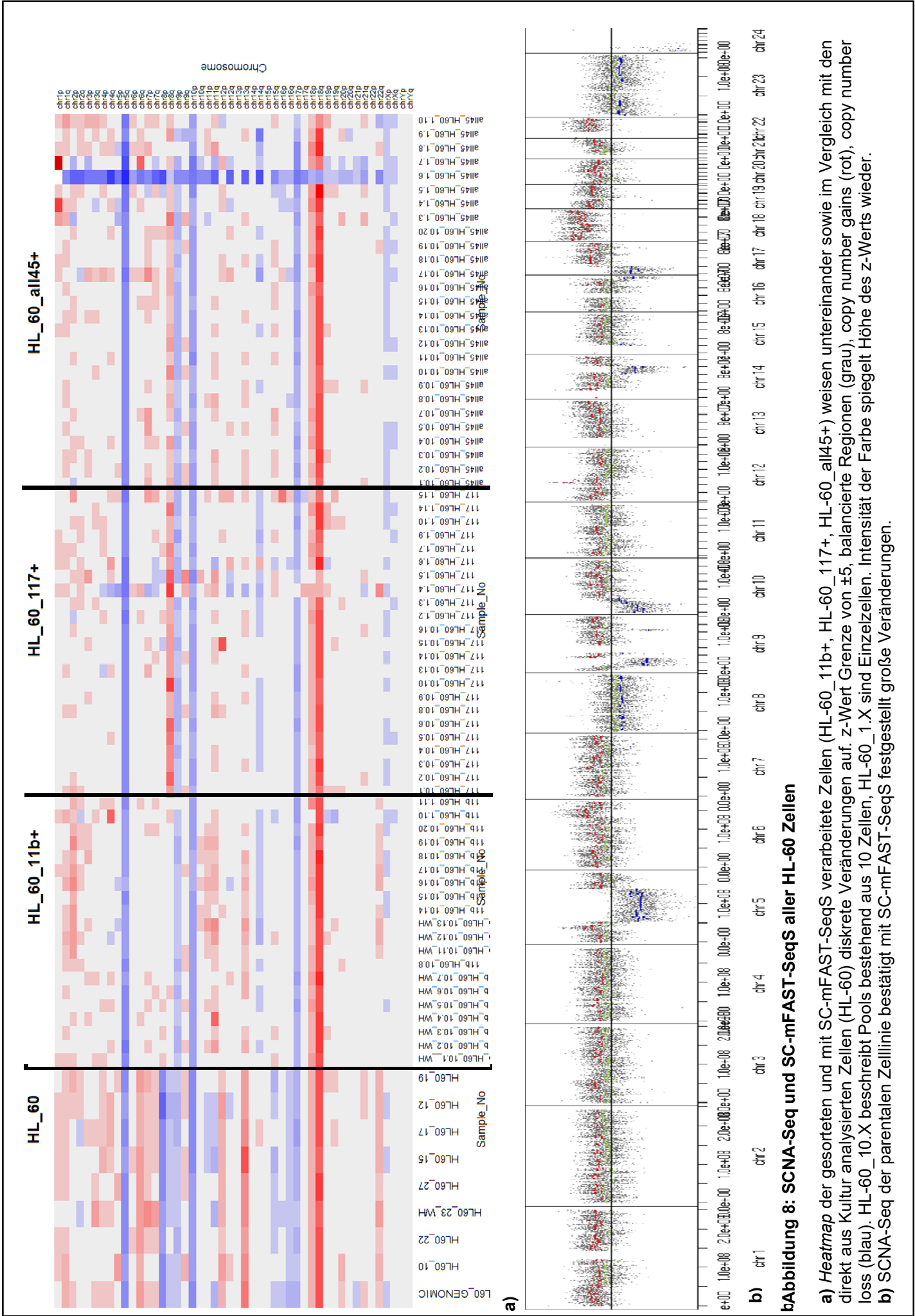


Abbildung 8: SCNA-Seq und SC-mFAST-SeqS aller HL-60 Zellen

a) Heatmap der gesorteten und mit SC-mFAST-SeqS verarbeitete Zellen (HL-60_11b+, HL-60_117+, HL-60_all45+) weisen untereinander sowie im Vergleich mit den direkt aus Kultur analysierten Zellen (HL-60) diskrete Veränderungen auf. z-Wert Grenze von ± 5 , balancierte Regionen (grau), copy number gains (rot), copy number loss (blau). HL-60_10.X beschreibt Pools bestehend aus 10 Zellen, HL-60_1.X sind Einzelzellen. Intensität der Farbe spiegelt Höhe des z-Werts wieder.

b) SCNA-Seq der parentalen Zelllinie bestätigt mit SC-mFAST-SeqS festgestellte große Veränderungen.

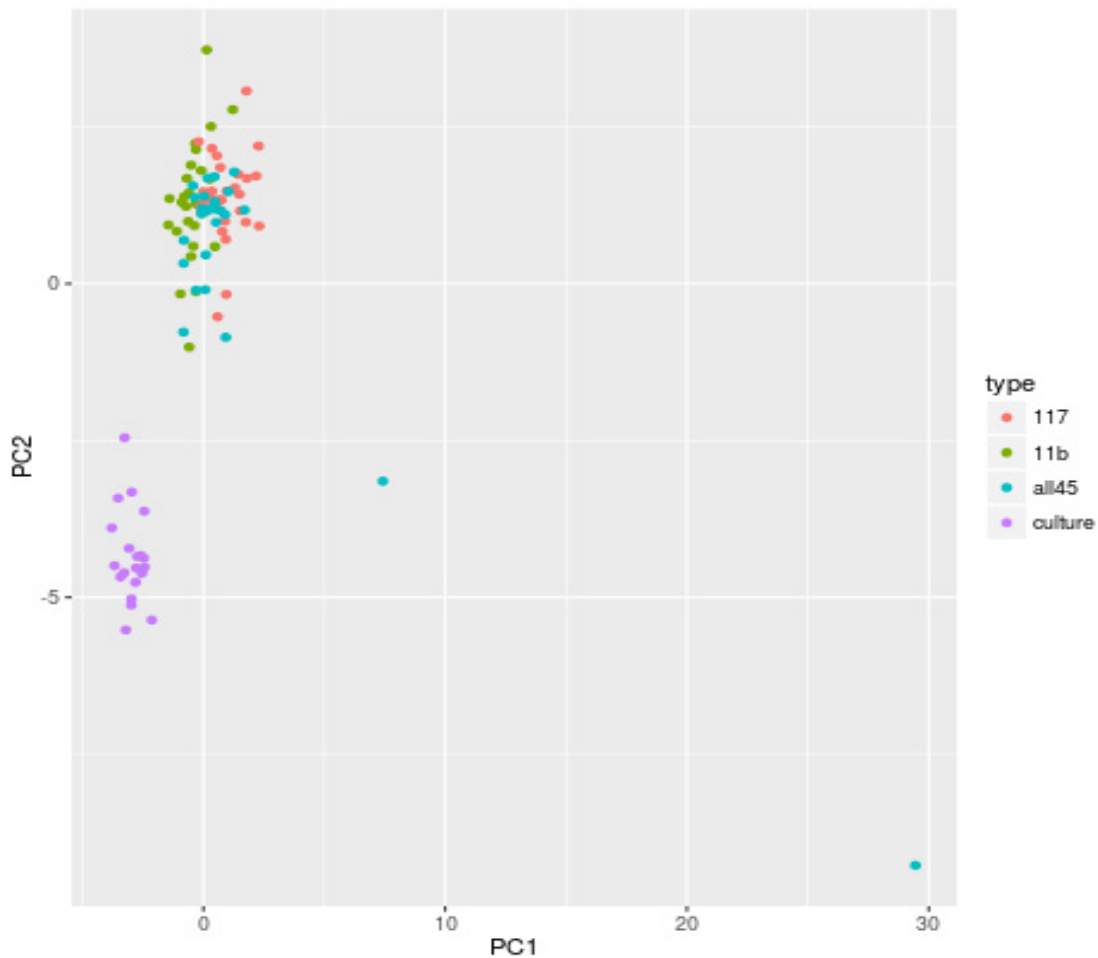


Abbildung 9: Principal component analysis aller HL-60 Zellen

Die PCA zeigt, dass gesortete Zellen einen Cluster bilden, in dem sich CD117+ Zellen von CD11b+ Zellen ansatzweise unterscheiden lassen. In diesem Cluster formieren sich auch unspezifisch positive CD45+ Zellen.

Zellen aus Kultur hingegen formen einen eigenen Cluster, der deutlich entfernt von den anderen sortierten Zellen liegt.

Alle untersuchten HL-60 Zellen, außer jene die „nur CD45+“ positiv waren, wiesen ähnliche Veränderungen innerhalb ihrer Gruppen, aber auch im Vergleich mit anderen auf (**Abbildung 8**). Es waren leichte Unterschiede zwischen den CD11b+ und CD117+ Zellen zu finden, während im allgemeinen CD45+ Cluster beide Unterschiede vertreten waren.

Die ungesorteten HL-60 Zellen zeigten für manche Chromosomenarme gröbere Veränderungen, im Vergleich mit sämtlichen gesorteten Zellen. PCA, sowie hierarchisches Clustering zeigen, dass CD11b+ und CD117+ Zellen trotz ihrer großen Ähnlichkeit aufgrund von wenigen Veränderungen dennoch differenziert werden können und sich auch die allgemeinen CD45+ Zellen auf diese beiden Cluster aufteilen. Ungesortete Zellen bilden jedoch gänzlich eigene Cluster und mischen sich nicht mit vorher erwähnten Clustern (siehe **Abbildung 10**).

3.3 Weitere Anwendungen von SC-mFAST-SeqS

Um die Anwendungsmöglichkeiten von SC-mFAST-SeqS auch außerhalb von Tumorerkrankungen, beispielsweise für ein Aneuploidiescreening von Polkörpern, zu testen wurde eine lymphoblastoide Zelllinie mit einer Trisomie 21 untersucht. 60 Zellen der Trisomie 21-Zelllinie erfüllten unsere 100.000 *Read* Marke mit einer durchschnittlichen *Read*-Anzahl von 280.245 *Reads* pro Probe (Maximum war bei TS_47 mit 659.906 *Reads*) vgl. **Abbildung 11**.

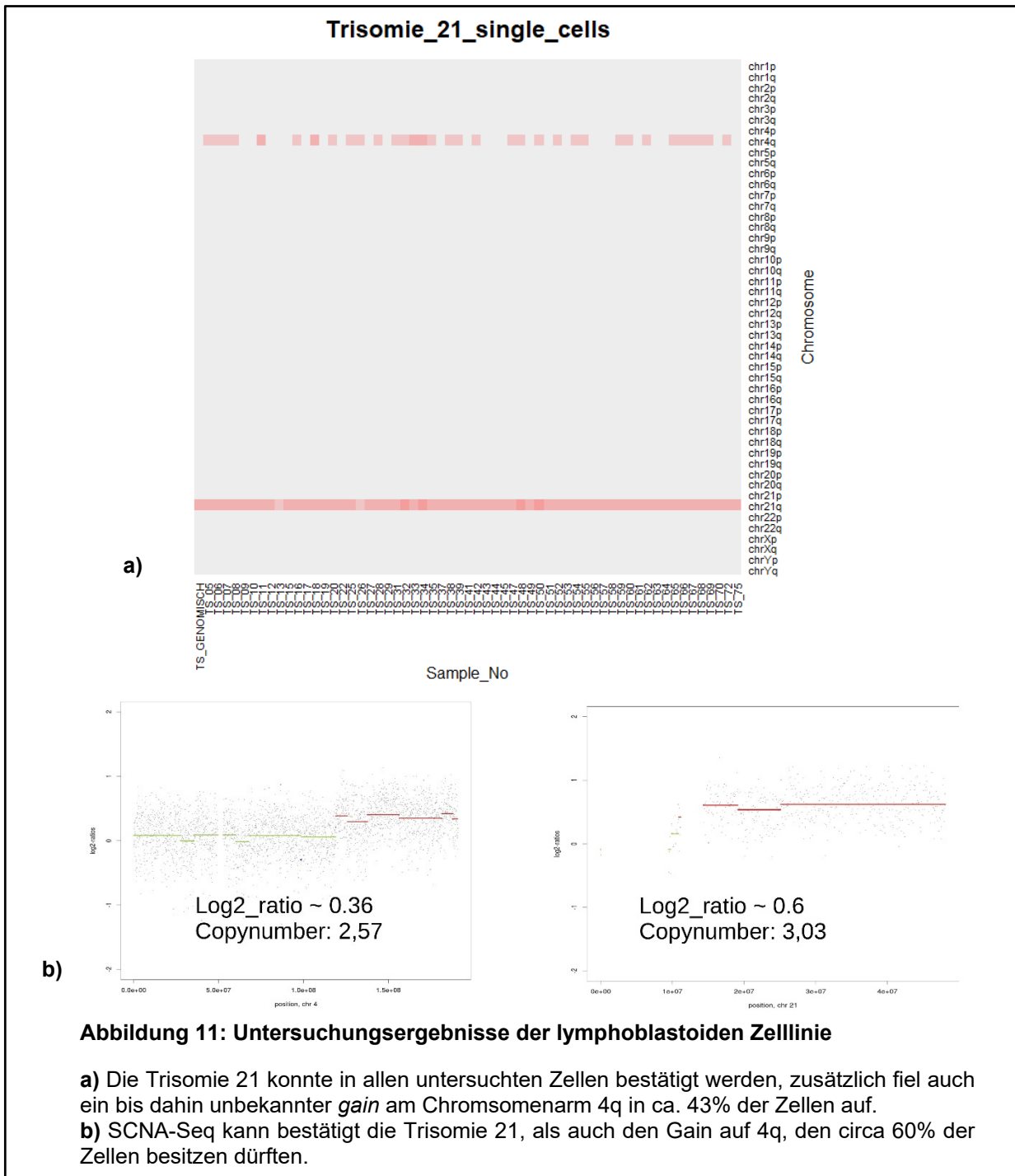


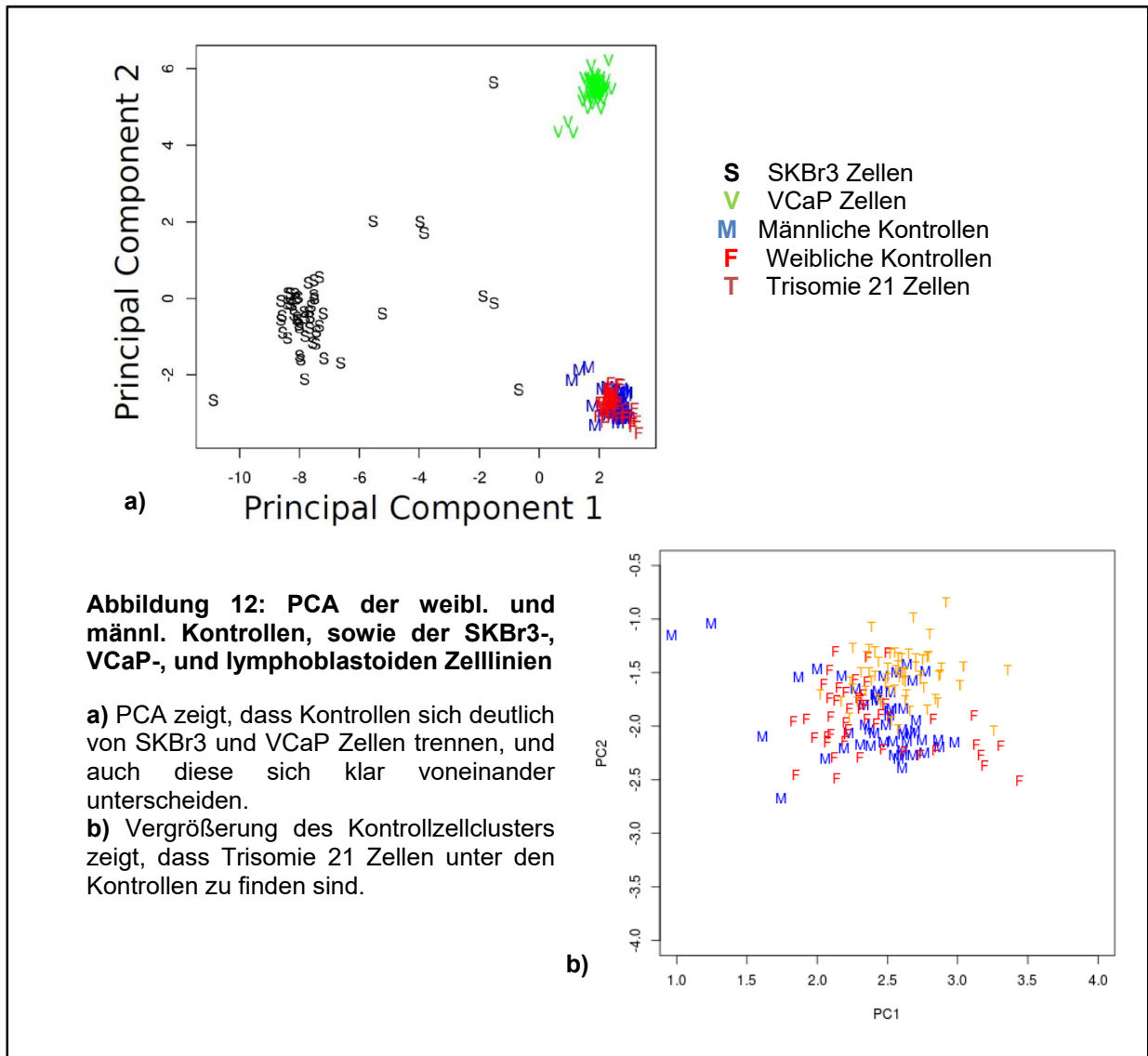
Abbildung 11: Untersuchungsergebnisse der lymphoblastoiden Zelllinie

a) Die Trisomie 21 konnte in allen untersuchten Zellen bestätigt werden, zusätzlich fiel auch ein bis dahin unbekannter *gain* am Chromosomenarm 4q in ca. 43% der Zellen auf.

b) SCNA-Seq kann bestätigt die Trisomie 21, als auch den Gain auf 4q, den circa 60% der Zellen besitzen dürften.

Alle mit SC-mFAST-SeqS untersuchten Zellen wiesen die zu erwartende Trisomie 21 auf, zusätzlich wurde eine Subpopulation entdeckt, die eine Veränderung auf Chromosomenarm 4q besitzt und laut SCNA-Seq etwa 60% der gesamten Population ausmachen dürfte. Diese Veränderung konnte in 43% der 60 untersuchten Zellen mittels SC-mFAST-SeqS bestätigt werden.

Da die lymphoblastoide Zelllinie an nur 2 Chromosomen Veränderungen aufweist, haben wir erneut eine PCA (**Abbildung 12**) und ein hierarchisches Clustering (**Abbildung 13**) durchgeführt, um die Grenzen der Auflösung dieser Methode zu verifizieren.



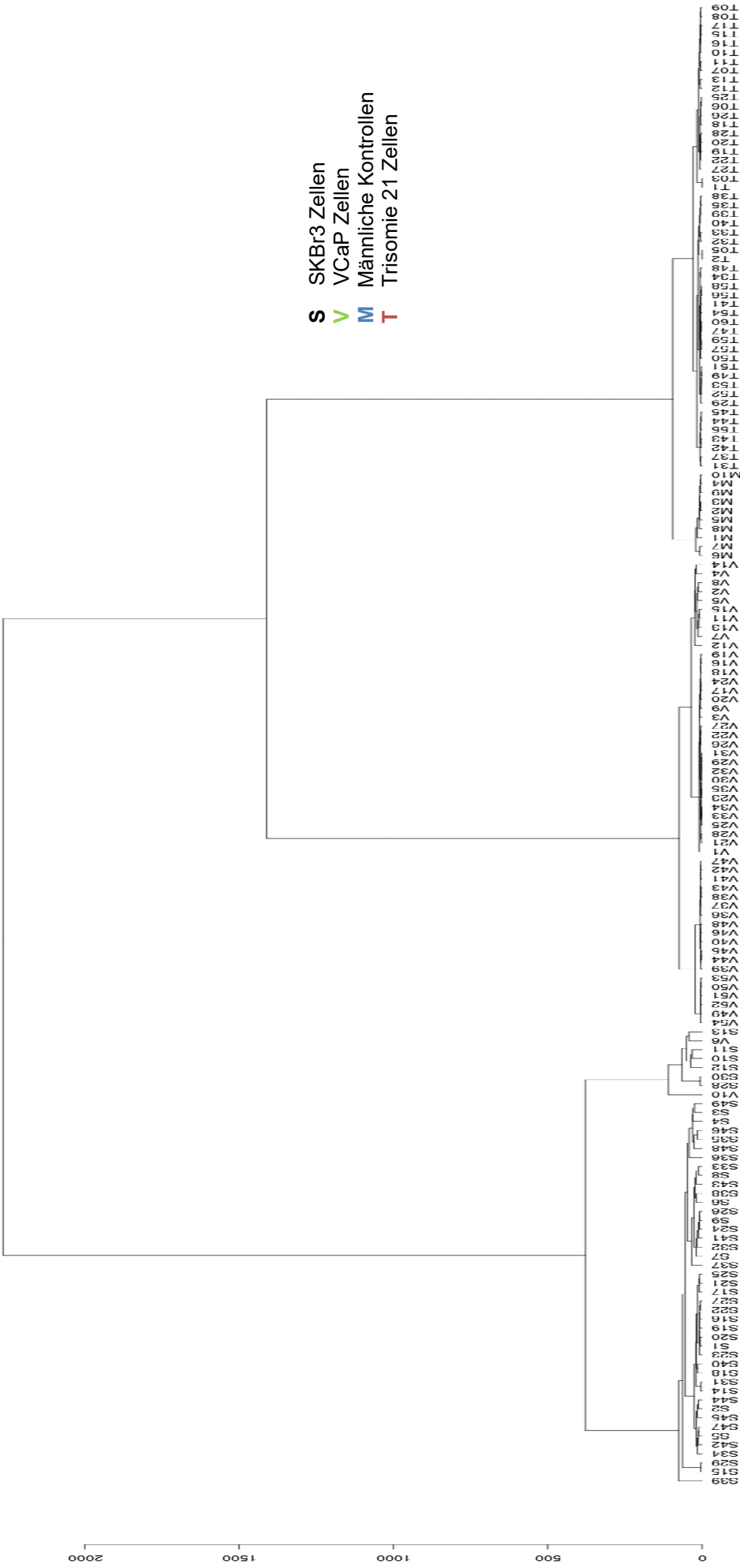


Abbildung 13: Hierarchisches Clustering der männl. Kontrollzellen, sowie der SKBr3-, VCaP-, lymphoblastoiden Zelllinie

SKBr3 Zellen (schwarzer Balken) und VCaP Zellen (grüner Balken) bilden eigene Cluster, wobei 2 VCaP- und 6 SKBr3-Zellen einen eigenen Cluster nahe den SKBr3-Zellen formen (gelber Balken). Männliche Kontrollen (blauer Balken) sowie lymphoblastoide Zellen (oranjer Balken) clustern sehr nahe, dennoch getrennt voneinander. Unter den männlichen Kontrollen ist keine einzige Trisomie-21 Zelle zu finden

4 Diskussion

Die Untersuchung von Einzelzellen hat ihren Platz in der Medizin bereits vor vielen Jahren im Bereich der Präfertilisations- bzw. Präimplantationsdiagnostik gefunden, sowie in der Diagnostik genetischer Erkrankungen, sowie von Krebserkrankungen, für die FACS-Untersuchungen klinische Routine geworden sind. Besonders im Hinblick auf Krebserkrankungen konnte man durch Einzelzellanalysen mittels NGS bedeutende Einblicke in die intratumorale Heterogenität gewinnen, die unser Verständnis der Evolution von Tumoren stark prägten und damit einen großen Schritt in Richtung personalisierter Medizin setzten. Dennoch limitieren der erhebliche Aufwand und die hohen Kosten, die für Einzelzellstudien nötig sind, die Resultate derartiger Untersuchungen bei Krebserkrankungen und eine Einführung in die klinische Diagnostik. Daher befasste sich diese Diplomarbeit mit der Fragestellung, ob sich SC-mFAST-SeqS als schnelle und kostengünstige Methode zur Feststellung von Kopienzahlveränderungen in Einzelzellen eignet.

Die Etablierung von SC-mFAST-SeqS gelang durch Analyse von männlichen und weiblichen Kontrollzellen. Keine der untersuchten Zellen, wies eine Veränderung auf, die mehr als einmal mit einem z-Wert von über ± 5 vorkam. Derartige Veränderungen sind auf die Ebene von Chromosomenarmen beschränkt, die p-Arme der akrozentrischen Chromosome ausgeschlossen. Daher ist es nicht möglich mittels SC-mFAST-SeqS Punktmutationen oder kleine strukturelle Veränderungen zu finden.

Alle untersuchten Zellen wurden mit einer Gesamtleseanzahl von 100.000 *Reads* pro Probe analysiert. Diese Grenze wurde gesetzt, nachdem bei *Reads* unter 100.000 pro Probe gehäuft falsch positive bzw. falsch negative Ergebnisse gefunden wurden.

Dies lässt sich auch auf die Ergebnisse der untersuchten humanen Krebszelllinien SKBr3, VCaP und einer lymphoblastoiden Zelllinie umlegen. Dabei war es durch hierarchisches Clustering möglich, die Krebszelllinien deutlich von den Kontrollen und der lymphoblastoiden Zelllinie zu trennen, wobei auch letztere trotz ihrer nur geringen Veränderungen von Kontrollzellen klar getrennt werden konnten.

Ähnliche Ergebnisse erbrachte auch die Analyse mittels PCA, jedoch war in diesem Fall die Unterscheidung zwischen Kontrollzellen und Tumorzellen nicht mehr so eindeutig wie bei der SC-mFAST-SeqS.

Diese Ergebnisse lässt die Frage, ob SC-mFAST-Seq für die Einzelzelldiagnostik im pränatalen Bereich geeignet ist, zunächst noch nicht klar beantworten. Allerdings legt der wiederholte Nachweis der Trisomie 21 in allen untersuchten Zellen der lymphoblastoiden Zelllinie den Schluss nahe, dass Aneuploidien mittels SC-mFAST-Seq sehr effizient detektiert werden können und folglich eine Anwendung in der Präfertilisations- bzw. Präimplantationsdiagnostik durchaus möglich ist.

Weitere Einzelzellen der lymphoblastoiden Zelllinie sollten untersucht werden, um festzustellen, ob SC-mFAST-SeqS auch in der Lage ist, dasselbe von SCNA-Seq gefundene Verhältnis in den Subpopulationen darzustellen. Da die Untersuchung von Einzelzellen mittels FACS in der Hämatologie und Hämatookologie längst in die Routinediagnostik eingegliedert wurde, stellt auch die Untersuchung leukämischer Einzelzellen prinzipiell eine denkbare Einsatzmöglichkeit für unser SC-mFAST-SeqS dar.

CD117+ und CD11b+ Zellen wiesen diskrete Veränderungen auf, die dennoch groß genug waren um eine Differenzierung der beiden zu erlauben. Auch die Aufteilung der unspezifisch CD45+ Zellen auf die o.g. Subpopulationen stellt eine erfreuliche Entdeckung dar, da auch diese beiden Subpopulationen CD45+ sind. Als durchaus positiv ist auch der fehlende Unterschied zwischen den Einzelzellen und den untersuchten Zehnergruppen zu werten, die im hierarchischen Clustering keine eigene Gruppe bilden, sondern sich in ihre jeweilige Subpopulationen vermengten.

Ob SC-mFAST-SeqS nach FACS-Analysen, wie sie vor allem in der Hämatologie und Hämatookologie angewandt werden, eingesetzt werden kann, ist durch die deutlich entfernt clusternden, nicht gesorteten HL-60 Zellen jedoch fraglich. Aufgrund der Tatsache, dass keine der „nur CD45+“ Zellen analysiert werden konnte und da nur 13 ungesortete Proben analysiert werden konnten, kann nicht gänzlich ausgeschlossen werden, dass die ungesorteten HL-60 Zellen ausschließlich aus dieser Subpopulation stammen. Daher würde es durchaus Sinn machen, noch mehr ungesortete HL-60 Zellen zu untersuchen bzw. die „nur CD45+“ Zellen erneut zu sortieren. Sollten die „nur CD45+“ Zellen nicht innerhalb der ungesorteten HL-60 Zellen clustern, bzw. sich bei einer größeren Menge an untersuchten ungesorteten HL-60 Zellen keine Vermengung der Zellen in Bereiche der gesorteten finden, wie man es aufgrund der FACS-Verhältnisse für die

einzelnen Subpopulationen erwarten würde, spricht dies für einen möglichen Einfluss des FACS auf die SC-mFAST-SeqS Methode. Sollte sich dieses Szenario bewahrheiten, wäre es in einem nächsten Schritt sinnvoll, gesunde Kontrollzellen zu sortieren und dies tatsächlich zu verifizieren. Denn dies würde eine gravierende Limitation der Methode im Bereich der Hämatologie und Hämatonkologie bedeuten.

Zusammenfassend lässt sich daher sagen, dass SC-mFAST-SeqS im Vergleich mit anderen SCS Methoden gewisse Vorteile mit sich bringt. Zwar sind andere SCS Methoden deutlich höher auflösend, doch erlauben sie gleichzeitig nur eine eher geringe Anzahl an Einzelzellen zu untersuchen, was die Ergebnisse limitiert. Im Gegensatz dazu stellt SC-mFAST-SeqS eine deutlich kostengünstigere und weniger arbeitsintensive Alternative dar, die ihre Stärke durch die Anzahl der untersuchten Zellen erst entfaltet.

Daher ist es denkbar, in Studien Einzelzellen zuerst mit SC-mFAST-SeqS zu untersuchen und damit vorab den Kopienzahlgehalt von Einzelzellen abzuschätzen und eventuell bereits Subpopulationen zu detektieren.

D.h. unsere Anwendung könnte in Zellkulturuntersuchungen bzw. auch in Studien, die mit Patientenmaterial arbeiten, Verwendung finden. SC-mFAST-SeqS erlaubt ein schnelles, aufwand- und kostengünstiges Screening, das dazu dienen könnte, Patientenproben vorab bereits gezielt auszuwählen und dadurch Kosten für hochauflösende teurere und zeitaufwendige Verfahren einzusparen.

Eine weitere potentielle Anwendungsmöglichkeit von SC-mFAST-SeqS findet sich daher in der Untersuchung von Pleuraergüssen, die bisher mittels Zytogenetik nach wie vor auf maligne Zellen untersucht werden. Aber auch fernab der Tumorbilogie könnten Einzelzelluntersuchungen mittels SC-mFAST-SeqS möglicherweise verwendet werden, insbesondere im Bereich der Präfertilisations- bzw. Präimplantationsdiagnostik. Allerdings bleibt eine tatsächliche Anwendung in diesen Gebieten fraglich, da SC-mFAST-SeqS Kopienzahlveränderungen auf Chromosomenarmebene widerspiegelt und folglich eine Aussage über kleine strukturelle Veränderungen und Punktmutationen nicht zulässt. Dennoch legen die in dieser Arbeit erzielten Ergebnisse nahe, dass SC-mFAST-SeqS für derartige Untersuchungen großes Potential hat, konnte es doch die Trisomie in allen untersuchten Zellen verifizieren und sogar eine Subpopulation entdecken, die uns bis dato nicht bekannt war.

Zusammenfassend lässt sich sagen, dass SC-mFAST-SeqS eine sehr elegante und kostengünstige Möglichkeit ist, um Aneuploidien und große Kopienzahlveränderungen zu detektieren. Im Gegensatz zu anderen SCS Methoden erlaubt sie, eine Vielzahl an Einzelzellen getrennt zu untersuchen, was in Bereichen wie der Tumorbilogie aber auch der Präfertilisations- bzw. Präimplantationsdiagnostik Anwendung finden kann.

Literaturverzeichnis

1. Krebserkrankungen, Statistik Austria [Internet]. [cited 2018 Mar 16]. Available from: https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/krebserkrankungen/index.html
2. Yee N, S. N. Update in Systemic and Targeted Therapies in Gastrointestinal Oncology. *Biomed* 2018, Vol 6, Page 34 [Internet]. 2018 Mar 16 [cited 2018 Mar 17];6(1):34. Available from: <http://www-1mdpi-1com-1pubmed.han.medunigraz.at/2227-9059/6/1/34>
3. Yuh B, Wilson T, Bochner B, Chan K, Palou J, Stenzl A, et al. Systematic review and cumulative analysis of oncologic and functional outcomes after robot-assisted radical cystectomy. *Eur Urol* [Internet]. 2015;67(3):402–22. Available from: <http://dx.doi.org/10.1016/j.eururo.2014.12.008>
4. Battaglin F, Schirripa M, Buggin F, Pietrantonio F, Morano F, Boscolo G, et al. The PANDA study: A randomized phase II study of first-line FOLFOX plus panitumumab versus 5FU plus panitumumab in RAS and BRAF wild-type elderly metastatic colorectal cancer patients. *BMC Cancer*. 2018;18(1):1–11.
5. Blair BG, Bardelli A, Park BH. Somatic alterations as the basis for resistance to targeted therapies. *J Pathol*. 2014;232(2):244–54.
6. Statistik Austria, online, Prognose der Krebsinzidenz und Krebsmortalität [Internet]. [cited 2018 Mar 17]. Available from: https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/krebserkrankungen/prognose_krebsinzidenz_krebsmortalitaet/index.html
7. Koo KC, Park SU, Kim KH, Rha KH, Hong SJ, Yang SC, et al. Prognostic impacts of metastatic site and pain on progression to castrate resistance and mortality in patients with metastatic prostate cancer. *Yonsei Med J*. 2015;56(5):1206–12.
8. Tosello G, Torloni MR, Mota BS, Neeman T, Riera R. Breast surgery for metastatic breast cancer. *Cochrane Database Syst Rev* [Internet]. 2018 Mar 15 [cited 2018 Mar 17];3:CD011276. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29542106>
9. Schmidt F, Efferth T. Tumor heterogeneity, single-cell sequencing, and drug resistance. *Pharmaceuticals*. 2016;9(2).
10. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: A looking glass for cancer? *Nat Rev Cancer* [Internet]. 2012;12(5):323–34. Available from: <http://dx.doi.org/10.1038/nrc3261>
11. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast

- cancer heterogeneity. *Nat Rev Clin Oncol* [Internet]. 2015;12(7):381–94. Available from: <http://dx.doi.org/10.1038/nrclinonc.2015.73>
12. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature* [Internet]. 2011;472(7341):90–5. Available from: <http://dx.doi.org/10.1038/nature09807>
 13. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* [Internet]. 2015;27(1):15–26. Available from: <http://dx.doi.org/10.1016/j.ccell.2014.12.001>
 14. Wang E, Zou J, Zaman N, Beitel LK, Trifiro M, Paliouras M. Cancer systems biology in the genome sequencing era: Part 2, evolutionary dynamics of tumor clonal networks and drug resistance. *Semin Cancer Biol* [Internet]. 2013;23(4):286–92. Available from: <http://dx.doi.org/10.1016/j.semcancer.2013.06.001>
 15. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet*. 2006;38(4):468–73.
 16. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* [Internet]. 2013;152(4):714–26. Available from: <http://dx.doi.org/10.1016/j.cell.2013.01.019>
 17. Nadeu F, Clot G, Delgado J, Martín-García D, Baumann T, Salaverria I, et al. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia* [Internet]. 2017;32(3):645–53. Available from: <http://www.nature.com/doi/10.1038/leu.2017.291>
 18. Haynes B, Sarma A, Nangia-Makker P, Shekhar MP. Breast cancer complexity: implications of intratumoral heterogeneity in clinical management. *Cancer Metastasis Rev*. 2017;36(3):547–55.
 19. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res*. 2005;65(20):9155–8.
 20. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet*. 2003;33(1):49–54.
 21. Stoecklein NH, Klein CA. Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. *Int J cancer* [Internet]. 2010 Feb 1 [cited 2018 Mar 21];126(3):589–98. Available from: <http://doi.wiley.com/10.1002/ijc.24916>
 22. Klein CA. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* [Internet]. 2009 Apr 1 [cited 2018 Mar 21];9(4):302–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19308069>

23. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* [Internet]. 2010 Oct 28 [cited 2018 Mar 26];467(7319):1114–7. Available from: <http://www.nature.com/doi/10.1038/nature09515>
24. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* [Internet]. 2010 Apr 15 [cited 2018 Mar 26];464(7291):999–1005. Available from: <http://www.nature.com/doi/10.1038/nature08989>
25. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* [Internet]. 2009 Oct 8 [cited 2018 Mar 26];461(7265):809–13. Available from: <http://www.nature.com/doi/10.1038/nature08489>
26. Roche-Lestienne C, Lai JL, Darre S, Facon T, Preudhomme C. A mutation conferring resistance to imatinib at the time of diagnosis of chronic myelogenous leukemia. *N Engl J Med*. 2003;348(22):2265–6.
27. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* [Internet]. 2012 Jan 11 [cited 2018 Mar 26];481(7382):506–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22237025>
28. Johnson BE, Mazar T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science* (80-) [Internet]. 2014 Jan 10 [cited 2018 Mar 26];343(6167):189–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24336570>
29. Nowell PC. The clonal evolution of tumor cell populations. *Science* [Internet]. 1976 Oct 1 [cited 2018 Mar 21];194(4260):23–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/959840>
30. Tellez-Gabriel M, Ory B, Lamoureux F, Heymann MF, Heymann D. Tumour heterogeneity: The key advantages of single-cell analysis. *Int J Mol Sci*. 2016;17(12).
31. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* [Internet]. 2013 Sep 19 [cited 2018 Mar 26];501(7467):328–37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24048065>
32. Bissell MJ, Hall HG, Parry G. How does the extracellular matrix direct gene expression? *J Theor Biol* [Internet]. 1982 Nov 7 [cited 2018 Mar 26];99(1):31–68. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6892044>
33. Sun X, Mao Y, Wang J, Zu L, Hao M, Cheng G, et al. IL-6 secreted by cancer-associated fibroblasts induces tamoxifen resistance in luminal breast cancer.

- Oncogene [Internet]. 2014 Jun 9 [cited 2018 Mar 26]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24909173>
34. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* [Internet]. 2013 Sep 19 [cited 2018 Mar 26];501(7467):346–54. Available from: <http://www.nature.com/articles/nature12626>
 35. Pareja F, Reis-Filho JS. Triple-negative breast cancers - a panoply of cancer types. *Nat Rev Clin Oncol* [Internet]. 2018 Mar 19 [cited 2018 Mar 28]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29555966>
 36. McFarland CD, Korolev KS, Kryukov G V, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A* [Internet]. 2013 Feb 19 [cited 2018 Mar 28];110(8):2910–5. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.1213968110>
 37. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med* [Internet]. 2015 Aug 1 [cited 2018 Jan 29];21(8):846–53. Available from: <http://www.nature.com/articles/nm.3915>
 38. Haffner MC, Mosbrugger T, Esopi DM, Fedor H, Heaphy CM, Walker DA, et al. Tracking the clonal origin of lethal prostate cancer. *J Clin Invest* [Internet]. 2013 Nov 1 [cited 2018 Mar 26];123(11):4918–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24135135>
 39. Keats JJ, Chesi M, Egan JB, Garbitt VM, Palmer SE, Braggio E, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood* [Internet]. 2012 Aug 2 [cited 2018 Mar 26];120(5):1067–76. Available from: <http://www.bloodjournal.org/cgi/doi/10.1182/blood-2012-01-405985>
 40. Ellsworth DL, Blackburn HL, Shriver CD, Rabizadeh S, Soon-Shiong P, Ellsworth RE. Single-cell sequencing and tumorigenesis: improved understanding of tumor evolution and metastasis. *Clin Transl Med* [Internet]. 2017;6(1):15. Available from: <http://clintransmed.springeropen.com/articles/10.1186/s40169-017-0145-6>
 41. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, et al. Genome-wide copy number analysis of single cells. *Nat Protoc* [Internet]. 2012;7(6):1024–41. Available from: <http://dx.doi.org/10.1038/nprot.2012.039>
 42. Navin NE. Cancer genomics: one cell at a time. *Genome Biol* [Internet]. 2014 Aug 30 [cited 2018 Jan 29];15(8):452. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0452-9>
 43. Smeets DFCM. Historical prospective of human cytogenetics: From microscope to microarray. *Clin Biochem*. 2004;37(6 SPEC. ISS.):439–46.
 44. Fox EJ, Loeb LA. One cell at a time. *Nature* [Internet]. 2014 Aug 30 [cited 2018 Jan

- 29];512(7513):143–4. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/25079325>
45. Kuipers J, Jahn K, Beerenwinkel N. Advances in understanding tumour evolution through single-cell sequencing. *Biochim Biophys Acta - Rev Cancer* [Internet]. 2017;1867(2):127–38. Available from:
<http://dx.doi.org/10.1016/j.bbcan.2017.02.001>
46. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* [Internet]. 2014;512(7513):155–60. Available from: <http://dx.doi.org/10.1038/nature13600>
47. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell* [Internet]. 2018;172(1–2):205–217.e12. Available from:
<https://doi.org/10.1016/j.cell.2017.12.007>
48. Heselmeyer-Haddad K, Berroa Garcia LY, Bradley A, Ortiz-Melendez C, Lee W-J, Christensen R, et al. Single-Cell Genetic Analysis of Ductal Carcinoma in Situ and Invasive Breast Cancer Reveals Enormous Tumor Heterogeneity yet Conserved Genomic Imbalances and Gain of MYC during Progression. *Am J Pathol* [Internet]. 2012 Nov [cited 2018 Mar 26];181(5):1807–22. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/S0002944012005779>
49. Hernandez L, Wilkerson PM, Lambros MB, Campion-Flora A, Rodrigues DN, Gauthier A, et al. Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. *J Pathol* [Internet]. 2012 May [cited 2018 Mar 26];227(1):42–52. Available from: <http://doi.wiley.com/10.1002/path.3990>
50. Sakr RA, Weigelt B, Chandarlapaty S, Andrade VP, Guerini-Rocco E, Giri D, et al. PI3K pathway activation in high-grade ductal carcinoma in situ--implications for progression to invasive breast carcinoma. *Clin Cancer Res* [Internet]. 2014 May 1 [cited 2018 Mar 26];20(9):2326–37. Available from:
<http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-13-2267>
51. Nowakowska B. Clinical interpretation of copy number variants in the human genome. *J Appl Genet*. 2017;58(4):449–57.
52. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet* [Internet]. 2015;16(3):172–83. Available from:
<http://dx.doi.org/10.1038/nrg3871>
53. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* [Internet]. 2010;464(7289):713–20.

- Available from: <http://dx.doi.org/10.1038/nature08979>
54. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, et al. Challenges and standards in integrating surveys of structural variation. *Nat Genet* [Internet]. 2007 Jul [cited 2018 Jan 30];39(7 Suppl):S7-15. Available from: <http://www.nature.com/doi/10.1038/ng2093>
 55. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, McEntagart ME, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am J Hum Genet* [Internet]. 2014 Aug 7 [cited 2018 Jan 30];95(2):173–82. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0002929714003127>
 56. Cantsilieris S, White SJ. Correlating multiallelic copy number polymorphisms with disease susceptibility. *Hum Mutat* [Internet]. 2013 Jan [cited 2018 Jan 31];34(1):1–13. Available from: <http://doi.wiley.com/10.1002/humu.22172>
 57. Iacocca MA, Hegele RA. Role of DNA copy number variation in dyslipidemias. *Curr Opin Lipidol* [Internet]. 2018 Jan 4 [cited 2018 Jan 29];1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29303791>
 58. Takumi T, Tamada K. CNV biology in neurodevelopmental disorders. *Curr Opin Neurobiol* [Internet]. 2018;48:183–92. Available from: <http://dx.doi.org/10.1016/j.conb.2017.12.004>
 59. Sapkota Y, Narasimhan A, Kumaran M, Sehrawat BS, Damaraju S. A Genome-Wide Association Study to Identify Potential Germline Copy Number Variants for Sporadic Breast Cancer Susceptibility. *Cytogenet Genome Res*. 2016;149(3):156–64.
 60. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* [Internet]. 2013;45(10):1134–40. Available from: <http://dx.doi.org/10.1038/ng.2760>
 61. Li W, Lee A, Gregersen PK. Copy-number-variation and copy-number-alteration region detection by cumulative plots. *BMC Bioinformatics*. 2009;10(SUPPL. 1):1–11.
 62. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc Natl Acad Sci* [Internet]. 1971;68(4):820–3. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.68.4.820>
 63. Camacho N, Van Loo P, Edwards S, Kay JD, Matthews L, Haase K, et al. Appraising the relevance of DNA copy number loss and gain in prostate cancer using whole genome DNA sequence data. *PLoS Genet*. 2017;13(9):1–28.
 64. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* [Internet]. 2010 Mar 9 [cited 2018 Feb 23];12:e8.

Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20211047>

65. Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014;505(7483):302–8.
66. Wei R, Zhao M, Zheng CH, Zhao M, Xia J. Concordance between somatic copy number loss and down-regulated expression: A pan-cancer study of cancer predisposition genes. *Sci Rep* [Internet]. 2016;6(July):1–9. Available from: <http://dx.doi.org/10.1038/srep37358>
67. Kinde I, Papadopoulos N, Kinzler KW, Vogelstein B. FAST-SeqS: A simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLoS One*. 2012;7(7).
68. Belic J, Koch M, Ulz P, Auer M, Gerhalter T, Mohan S, et al. Rapid identification of plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. *Clin Chem*. 2015;61(6):838–49.
69. Graier T, Auer M, Heitzer E, Ulz P, Perakis S, Speicher MR, et al. Rapid single-cell copy-number aberration analysis using high-throughput sequencing. *Eur J Cancer* [Internet]. 2017 Feb 1 [cited 2018 Mar 7];72:S24. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0959804917301594>

