

Schlussbericht MPIMG

BMBF Pilotprojekt CancerSys:

TREAT20 - Tumorforschung und -behandlung -

Ein Pilotprojekt mit 20 Patienten

Förderkennzeichen: 0315852b

Berichtszeitraum: 01. Oktober 2010 bis 30. September 2013

Zuwendungsempfänger: Max-Planck-Gesellschaft zur Förderung der
Wissenschaften e.V., vertreten durch das Max-
Planck-Institut für molekulare Genetik, Berlin,
Abteilung Analyse des Vertebratengenoms
Ihnestraße 63-73
14195 Berlin, Germany

Projektkoordinator: Prof. Dr. Hans Lehrach
Tel.: + 49 (0)30 - 84 13 1220
Fax: + 49 (0)30 - 84 13 1380
E-mail: lehrach@molgen.mpg.de

WP5: Deep sequencing of genomes and transcriptomes (Dr. Marie-Laure Yaspo, Dr. Tatjana Borodina, MPIMG)

I. Kurze Darstellung zu

1. Aufgabenstellung

Wir haben ursprünglich vorgeschlagen etwa 388 GB an Sequenzinformation für jeden der 20 Melanoma Patienten zu generieren. Für diese Sequenzierungen war geplant, eine ca. 30-fache Genomabdeckung mittels der Paired-End-Sequenzierungsstrategie zu erreichen. Für alle Patiententumore planten wir ganze Genom-Sequenzierungen (WGS) und Transkriptom-Sequenzierungen (RNAseq), für die Tumorzelllinien und Xenografts ganze Exom-Sequenzierungen und Transkriptom-Sequenzierungen. In einen zweiten Ansatz, sollte zirkulierenden Tumor-DNA und der epigenetischen Status von Tumoren und Tumor-abgeleiteten Material mittels Sequenzierung von angereicherte DNA durch Bindung an ein Methyl-C-spezifischen Antikörper (MeDIP-seq) untersucht werden.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Projekt wurde während des Beginns von NGS-Anwendungen für die Sequenzierung von Krebsgenomen begonnen. Unsere Abteilung gehörte zu den führenden Institutionen die NGS-Technologien anwendeten und war bereits mit den ersten SOLiD und Illumina (GAII) Sequenzer ausgestattet. Zu dieser Zeit waren wir maßgeblich an der Entwicklung, Erprobung und Verbesserung von NGS-Versuchsprotokolle, insbesondere für RNAseq beteiligt. Insbesondere entwickelten wir zuverlässige Protokolle für die Erhaltung der Strang-spezifischen Informationen in RNAseq Bibliotheken (Parkhomchuk et al., 2009, Sultan et al., 2012), um so wesentliche Informationen für die anschließenden Analysen zur Verfügung stellen zu können und etablierten Vergleiche zwischen Array- und NGS-basierten Analysen (Sultan et al., 2008). Damit waren alle notwendigen Techniken und Methoden, einschließlich der Datenverarbeitung und Qualitätskontrolle für die Durchführung des TREAT20 Projekts verfügbar. Das Projekt begann, als nur sehr begrenztes Wissen bezüglich der Genome von Melanomen zur Verfügung stand (Pleasant et al., 2010). Kein Melanom Transkriptom war zu dem Zeitpunkt sequenziert worden.

3. Planung und Ablauf des Vorhabens

Der ursprüngliche Plan war, innerhalb des Zeitrahmens von 3 Jahren 20 Melanom-Patienten zu sequenzieren und zu analysieren. Zu Beginn des Projektes im Jahr 2010 waren die NGS-Technologien noch relativ langsam im Vergleich zu heute, vor allem mit den ersten Geräten von Illumina (GAII), die nur im begrenzten Umfang Sequenzinformationen lieferten. Allerdings fanden enorme technologische Verbesserungen während des TREAT20 Projektes statt, so dass es möglich war, die geplanten 20 Patiententumore innerhalb des Zeitrahmens des Projekts zu sequenzieren. Das logistische Management für das Patientenmaterial für die Herstellung von qualitativ hochwertigen Nucleinsäuren für die NGS, das unserem Labor zur Verfügung gestellt wurde war etabliert.

4. wissenschaftlichem und technischem Stand, an den angeknüpft wurde, insbesondere Angabe bekannter Konstruktionen. Verfahren und Schutzrechte die für die Durchführung des Vorhabens benutzt wurden

Wir begannen das Projekt zu der Zeit, als das ICGC-Projekt (Internationale Krebsgenomkonsortium) initiiert wurden. Das Konzept der Verwendung der Krebsgenomdaten für die personalisierte Medizin und Behandlung fing gerade an zu wachsen. Nur begrenzte Informationen über Melanome andere als BRAF oder NRAS Mutationen standen zur Verfügung.

Außerdem wurden wichtige technologische Fortschritte gemacht, so dass wir den gesamten Prozess erheblich beschleunigen konnten. Wir begannen das Pilotprojekt mit den Illumina GAII-Maschinen, bei denen die Sequenzierung noch relativ langsam vonstatten ging, wodurch klinische Behandlungsmöglichkeiten oftmals verpasst wurden. Mittlerweile ist das Verfahren,

unter Verwendung der neuesten Illumina HiSeq2500 Maschinen (und neuer Sequenzier-Chemie) und mittels reduzierten Probenverarbeitungszeiten sowie verbesserter Analysepipeline, gut verträglich mit einer realistischen klinischen Interventionszeitlinie. Die benötigte Zeit von der Probenentnahme bis zur Abgabe der Medikamenten-Vorhersage beläuft sich derzeit nur noch auf etwa einen Monat. Darüber hinaus können wir nun erfolgreich auch kleine Mengen von Material aus Biopsien verarbeiten.

- Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste; (Pleasant et al., 2010). Die erste Veröffentlichung eines Melanom-Genoms. Obwohl nur aus einer Melanom-Zelllinie und nicht die komplexe Situation eines kutanen Melanoms reflektierend, waren diese Daten aufschlußreich als erster Bezugspunkt hinsichtlich der Zahl der zu erwartender Mutationen.

5. Zusammenarbeit mit anderen Stellen;

Keine

6. Beitrag des Ergebnisses zu den förderpolitischen Zielen. z. B. des Förderprogramms - (ggf. unter Angabe des Schwerpunkts) soweit dies möglich ist;

Die hier beschriebenen Resultate leisten einen wichtigen Beitrag zur Entwicklung und Validierung der Systemmedizin, eines Schlüsselthemas der BMBF Förderung im Bereich der biomedizinischen Forschung (<http://www.bmbf.de/de/6942.php>).

7. Präsentationsmöglichkeiten für mögliche Nutzer - z. B. Anwenderkonferenzen (Angaben, soweit die Art des Vorhabens dies zulässt);

Die Projektdaten wurden in verschiedenen Vorträgen (innerhalb von Tagungen und bei Seminaren) erwähnt. Eine Publikation ist in Vorbereitung.

8. Erklärung zur Einhaltung der Ausgaben- und Zeitplanung.

Im Laufe des Projektes konnten wir die für die Sequenzierung und Analyse der Patiententumoren nötige Zeitdauer wesentlich verbessern. Während der Projektdauer selbst wurde das Budget eingehalten, jedoch konnte der Zeitrahmen nicht ganz eingehalten werden, da die benötigten schnelleren Sequenzieretechnologien erst im Verlauf der Projektdauer verfügbar wurden.

Im Projektverlauf war die Sequenzierung von zirkulierender Tumor DNA nicht möglich, da die Technologie dafür noch nicht ausgereift war. Die Methylierungsanalyse konnte ebenfalls nur unter hohem Kosten durchgeführt werden. In der Zwischenzeit ist die Verwendung von 450K-Arrays zu dem bevorzugten Standard über den MeDIP Ansatz geworden, den wir ursprünglich vorgeschlagen hatten. Allerdings waren diese Analysen nicht zwingend notwendig für die Modellierung und Medikamentenvorhersagen.

II. Eingehende Darstellung

1. des erzielten wissenschaftlich-technischen Ergebnisses, der erreichten Nebenergebnisse und die gesammelten wesentlichen Erfahrungen, bzw. der Arbeiten, die zu keiner Lösung geführt haben;

In der TREAT20-Pilotphase haben wir das Gesamtgenom von 20 metastatischen Melanomen verschiedene Typs (Kutan, Akral und Uveal) sowie den dazugehörigen Kontrollblutproben der entsprechenden Patienten sequenziert (whole genome sequencing (WGS) von P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20) und dabei eine durchschnittliche Sequenzabdeckung von 30-50 fach erreicht (100 bp paired-end Fragmente),

sodass eine Annotation und anschließende Analyse der somatischen Mutationen erfolgen konnte. Wir konnten auch das Material von P12 trotz der geringen Biopsiegröße erfolgreich prozessieren. Des Weiteren haben wir das mRNA Expressionsprofil aller Tumorproben mittels mRNAseq (50 bp paired-end) unter Zuhilfenahme eines Protokolls bestimmt, welches die Transkriptorientierung beibehält. Sämtliche Daten erfüllen die Qualitätskontrollkriterien die für dieses Projekt implementiert wurden (entsprechend den internationalen ICGC Richtlinien), zum Beispiel im Bezug auf die prozentuale Anzahl der zugeordneten Reads, die Komplexität der verwendeten Genom und Transkriptom Bibliotheken, der Verteilung der erhaltenen Reads sowie die Sequenzabdeckung auf Genom und Transkriptomebene. Von den 20 Melanom Patienten hatten wir auch zwei Rückfälle (P9 und P14), die ebenfalls sequenziert wurden.

Zusätzlich haben wir von den von Patienten P1, P2, P3, P6 und P9 abgeleiteten Xenograft Tumoren das Exom oder das Transkriptom bestimmt. Dies erlaubte vergleichende Analysen von Xenograft-Sequenz und originaler Tumorsequenz. Wir haben darüber hinaus auch die Transkriptome für sechs Zellenlinien sequenziert (P2, P3, P6, P9, P15, und P9). Außerdem haben wir zwei zusätzliche Melanozytenzelllinien sequenziert welche wir für die Analyse der differentiellen Expression im Tumorgewebe nutzen.

Im Bezug auf die verwendeten Technologieplattformen verwenden wir Illumina Hiseq Geräte für WGS sowie SOLiD Geräte für die Exomsequenzierung. Zu Beginn der Projektphase wurden Transkriptome mittels des Illumina GAll Geräts sequenziert. Wir haben die entsprechenden Bibliotheken erneut mit dem Illumina Hiseq Gerät sequenziert um eine optimale Abdeckung niedrig und mittelstark exprimierter Gene zu erreichen, die Identifizierung exprimierter Mutationen zu verbessern und eine erweiterte komplexe Analyse der sequenzierten Transkriptome, zum Beispiel im Bezug auf Genfusionsereignisse, zu ermöglichen. Um alle Informationen ausgehend von der Akquisition des Tumormaterials bis zu den generierten Sequenzdaten zu speichern haben wir ein LIMS System der Firma Genologics angeschafft. Dieses System wurde installiert und an die Illumina Sequenziermaschinen angeschlossen.

In der Anfangsphase des Projekts wurden wie eben erwähnt die Transkriptome mittels der Illumina GAll Plattform sequenziert. Wir haben nun begonnen, die vorhandenen Bibliotheken mittels der Illumina HiSeq Plattform erneut zu sequenzieren, um so eine optimale Abdeckung der Sequenzen gering oder mittelstark exprimierter Gene zu erzielen. Dies hat den klaren Vorteil, dass exprimierte Mutationen auch in dieser Klasse von Genen bestimmt werden können, was eine weitere komplexe Analyse des individuellen Transkriptoms erlauben wird. Abgesehen von den genannten Experimenten und Ergebnissen haben wir die Anschaffung zusätzlicher Melanozytenzelllinien in Auftrag gegeben; diese Zellen sollen als weitere Kontrollen dienen und die Ergebnisse der Expressionsanalysen konsolidieren.

Wie im vorherigen Zwischenbericht erwähnt untersuchen wir die Verwendbarkeit respektive den zusätzlichen Nutzen der Exom-Sequenzierung von Tumorproben mit geringem Anteil an Tumorzellen (geringer Tumor-Reinheit) in Ergänzung zur Gesamtgenomsequenzierung, sowie die Sequenzierung von Xenograph-abgeleitetem Material. Wir haben die Exom-Selektionsprotokolle von Life Technologies und Agilent (upgraded SOLiD platform) getestet und darauf basierend ein Test-Exom sequenziert wobei sich das Agilent System als überlegen herausgestellt hat und daher für die weiteren Arbeiten präferiert wurde.

Biomek FXP library Prozessierungsroboter (Beckman Coulter)

Der Biomek Roboter wurde im Dezember 2011 angeschafft und es fand eine von der Firma Beckman Coulter organisierte Schulung am MPIMG im Februar 2012 statt, um den Umgang mit dem Roboter zu erlernen (z.B. Programmierung von Protokollen, Kalibrierung usw.). In den darauf folgenden Monaten wurden die einzelnen Schritten von zwei Sequenzierungsprotokollen (DNA library Präparation, TruSeqRNA library Präparation) auf den Roboter programmiert, kalibriert und getestet. Einzelne Proben konnten erfolgreich prozessiert werden und die komplexen Arbeitsprozesse konnten dadurch erheblich vereinfacht werden sowie damit eine hohe Genauigkeit und Reproduzierbarkeit gewährleistet werden. Wir haben damit eine belastbare und nutzbare Struktur etabliert, die uns in Zukunft ein größer angelegtes Behandlungsverfahren für Krebspatienten zu ermöglichen.

Damit sind wir dem Ziel der Pilotphase, nämlich ein System für größere Patientenkohorten mit schnellen und qualitativ gleich bleibend hochwertigen Ergebnissen zu etablieren, einen ganzen Schritt näher gekommen.

Insgesamt sequenzierten und erhielten wir mehr als 10.000 Gb an Sequenzdaten, also mehr als ursprünglich geplant (388x20, d.h. 7.760 Gb) waren.

2. des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses;

„Deep sequencing“ von Genomen und Transkriptomen von mehreren Subtypen eines metastasierten Melanoms sind eine wertvolle Ressource für das Verständnis der molekularen Grundlagen dieser Krebsarten. Insbesondere haben die Daten gezeigt, dass die Analyse der Expressionsprofile wesentlich zum Verständnis der komplexen heterogenen molekularen Profilen von Melanomen beitragen.

Die Ergebnisse wurden verwendet, um Therapien für einige dieser Patienten vorherzusagen (siehe WP 8 Modellierung).

3. des während der Durchführung des Vorhabens dem Projektleiter bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen;

Mehrere Veröffentlichungen meist von Exom Sequenzierung von Melanomen Kohorten wurden von anderen Gruppen gegen Ende des Projektes (Berger et al. 2012, Hodis et al. 2012, Krauthammer et al. 2012, Nikolaev et al. 2012) veröffentlicht. Diese Daten deckten einige neue, in verschiedenen Patienten wiederholt auftretende Mutationen auf. Obwohl die Mutationen nur in einem relativ kleinen Teil der Patienten auftreten, waren diese Informationen als Vergleichsmaßstab nützlich, um zu vergleichen, ob wir einige dieser Mutationen auch sehen würden.

Allerdings integrieren diese keine RNAseq und fokussierten sich auf die Identifizierung von wiederkehrenden Varianten in großen Patientenkohorten im Gegensatz zu unserer Studie, die sich auf eine tiefe detaillierte Analyse der einzelnen Patienten, auch für die Identifizierung individuell relevanter Ereignisse, auf eine genaue Modellierung der Tumoren konzentriert.

4. der erfolgten oder geplanten Veröffentlichungen des Ergebnisses.

Die Ergebnisse der gesamten Serie von 20 Patienten, darunter Proteomik-Daten werden dem Journal of Clinical Investigation zur Veröffentlichung eingereicht werden.

Literatur:

Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-506

Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251-263

Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet* **44**, 1006-1014, doi:10.1038/ng.2359 (2012).

Nikolaev, S. I. *et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat Genet* **44**, 133-139, (2012).

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, Lehrach H, Soldatov A. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* Epub 2009 Jul 20.

Pleasant et al. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196.

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*. 2008 Aug 15;321(5891):956-60

Sultan M, Dökel S, Amstislavskiy V, Wuttig D, Sültmann H, Lehrach H, **Yaspo** ML. 2012. A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem Biophys Res Commun*. 2012 Jun 15;422(4):643-6.

WP6: Proteomeanalyse (Markus Ralser)

I. Kurze Darstellung zu

1. Aufgabenstellung

Ziel von TREAT20 ist die Einführung von genomweiten Sequenzierungsdaten und hierauf basierenden Computermodellen molekularer Netzwerke zur Behandlungsstratifizierung von Tumorpatienten. Dazu wurden in diesem Pilotprojekt 20 metastasierte Melanompatienten, welche auf die Standardtherapie nicht ansprachen, ausgewählt und auf Basis der aus den Analysen der molekularen Aberrationen des Tumors gewonnenen Daten, individuell im Rahmen einer klinischer Studie therapiert. Diese Strategie soll neue Möglichkeiten der personalisierten Behandlung eröffnen und damit beitragen die bisher meist noch global ausgerichteten Tumorthérapien, zu ersetzen.

Von den 20 Studienpatienten wurde das Genom, Methylohm und Transkriptom des Tumors mit Hochdurchsatzsequenzierung und eine zielgerichtete Proteomanalyse charakterisiert. Referenz ist das Genom des Normalgewebes (Blutzellen) des Patienten für genomische Untersuchungen, Melanozytenzelllinien für Transkriptom und Epigenome Untersuchungen. Auf der Basis dieser Daten werden prädiktive Modelle des einzelnen Tumors/Patienten erstellt, an Hand von Zell- und Xenograftmodellen der einzelnen Tumore validiert, und verwendet. Basierend auf den Ergebnissen der Analysen, wurde die Behandlung durch den Studienarzt individuell angepasst und durchgeführt. Der Einfluss dieser optimierten Therapie wird am Krankheitsverlauf (Therapieansprechen, Progressionsfreie Zeit) dokumentiert und bewertet.

2. Voraussetzungen, unter denen das FuE Verfahren durchgeführt wurde

Wichtige Teile der biologischen Netzwerke, die in den Tumorzellen ablaufen, sind post-transkriptionell reguliert. Sie involvieren Proteine, Proteinmodifikationen, Proteinkomplexe und Metabolite. Zum Zeitpunkt des Beginns des Projektes, waren nur sehr limitiert Informationen über diese Komponenten in Tumoren vorhanden. Insbesondere waren keinerlei Untersuchungen verfügbar, in denen Genom, Transkriptom und Proteomdaten der gleichen Proben analysiert worden waren.

3. Planung und Ablauf des Vorhabens

Auf Grund der technischen und finanziellen Limitationen war innerhalb von TREAT20 geplant, selektiv bestimmte Proteine und Proteinmodifikationszustände zu quantifizieren. Im Zuge des Verfahrens war es aber möglich, diesen Ansatz durch einen zweiten, datengetriebenen Analyseansatz, der systematischen *shotgun*-Analyse des Proteoms mit Hilfe einer HPLC gekoppelten hybrid Quadrupole/Orbitrap Massenspektrometers (Qexactive, Thermo Scientific) zu komplementieren. Die Resultate dieser Analysen sind seit kurzem verfügbar, und werden jetzt ebenfalls zur Verbesserung und Validierung der Computermodelle der Tumoren verwendet.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Die in TREAT20 vorgeschlagenen Proteomuntersuchungen basierten auf dem Stand der Technik, und den damals in den beteiligten Instituten verfügbaren Geräten (e.g. Lange et al., 2008). Basierend auf den Erfahrungen wurden gezielte Proteomanalysen auf Tumorproben durchgeführt und in verschiedenen Tumorproben quantifiziert. Eines der analysierten Proteine ist das glycolytische Enzym Pyruvatkinase (PKM) das als Regulator des Krebsstoffwechsel gilt. Die entwickelte massenspektrometrische Methode (Bluemlein & Ralser, 2011) wurde eingesetzt um zwei Splice-isoformen des Gens absolut zu quantifizieren.

Im Zuge des Projektes wurden des weiteren alternative getargeted Ansätze (SWATH) geprüft, die aber bisher, auf Grund der noch weiter notwendigen Entwicklungsarbeiten und der mangelnden Verfügbarkeit des dazu notwendigen Gerätes (Triple TOF5600, AB/Sciex) noch nicht endgültig eingesetzt werden konnten. Tests werden aber weiter fortgesetzt, und die Methodenentwicklung vorangetrieben (Vowinckel et al., 2014). Datengetriebene Proteomanalysen waren aber möglich mit Hilfe eines Orbitrap Exactive Massenspektrometers. Die dabei erhaltenen Daten können jetzt weiter zu Validierung der Modelle, sowie zum Entwurf weiterer SRS Assays verwendet werden.

5. Zusammenarbeit mit anderen Stellen

Keine

6. Beitrag des Ergebnisses zu den förderpolitischen Zielen zB des Förderprogram (ggf. unter Angabe des Schwerpunkts) soweit dies möglich ist;

Die hier durchgeführten Arbeiten tragen generell zur Entwicklung der Systemmedizin bei. Dies ist bereits seit einiger Zeit ein zentraler Aspekt der BMBF Förderung.

7. Präsentationsmöglichkeiten für mögliche Nutzer - z. B. Anwenderkonferenzen (Angaben, soweit die Art des Vorhabens dies zulässt);

Keine

8. Erklärung zur Einhaltung der Ausgaben- und Zeitplanung.

Da am Anfang des Projektes zu wenige Informationen zu Selektion von (hypothesengetriebenen) SRM Analysen verfügbar war, wurden in der Anfangsphase vor allem Technologieentwicklung vorangetrieben, um nicht wertvolles Tumormaterial suboptimal zu verwenden. Erst später war es möglich, auf der Basis eines datengetriebenen Proteomikansatzes und verbesserte Modelle die Voraussetzungen für eine systematische hypothesengetriebene Proteinanalyse zu schaffen.

II. Eingehende Darstellung

1. des erzielten wissenschaftlich-technischen Ergebnisses, der erreichten Nebenergebnisse und die gesammelten wesentlichen Erfahrungen, bzw. der Arbeiten, die zu keiner Lösung geführt haben;

Die targeted Proteomanalyse wurde auf weitere Proteine, welche durch bioinformatische Analysen identifiziert worden sind, ausgedehnt. Dadurch konnten zwei Proteine identifiziert werden, welche in Tumoren unterschiedlich reguliert sind als in Kontrollgewebe und darüber hinaus mit einem Regulator des Krebsstoffwechsels korrelieren.

Neue Methodenentwicklung: SWATH Proteomics

Für die markierungs-freie Analyse von Proteinproben wurde eine von Ruedi Aebersold (ETH Zürich) entwickelte Methode (SWATH-MS) etabliert und weiterentwickelt. Ähnlich wie beim SRM wird eine Daten-unabhängige Quantifizierung durchgeführt; allerdings ist diese nicht beschränkt auf die vorherige Definition des zu messenden Proteins, da der gesamte Massenbereich den das Massenspektrometer abdecken kann fragmentiert und analysiert wird. Spektren zur Proteinquantifizierung werden anschließend am Computer rekonstruiert. So durchgeführte Analysen sind von deutlich besserer quantitativer Aussage im Vergleich zu klassischer ‚Discovery Proteomics‘

und auf der anderen Seite nicht eingeschränkt auf eine vorherige Proteinauswahl (Abbildung 1). Dies ist vorteilhaft, da sich die Auswahl weniger Kandidatenproteinen durch die genetische Komplexität der TREAT20 Proben als schwierig erwiesen hat (andere WPs). Für diese Analyse wurden des weiteren verschiedenste Probenpräparationsmethoden getestet und entwickelt:

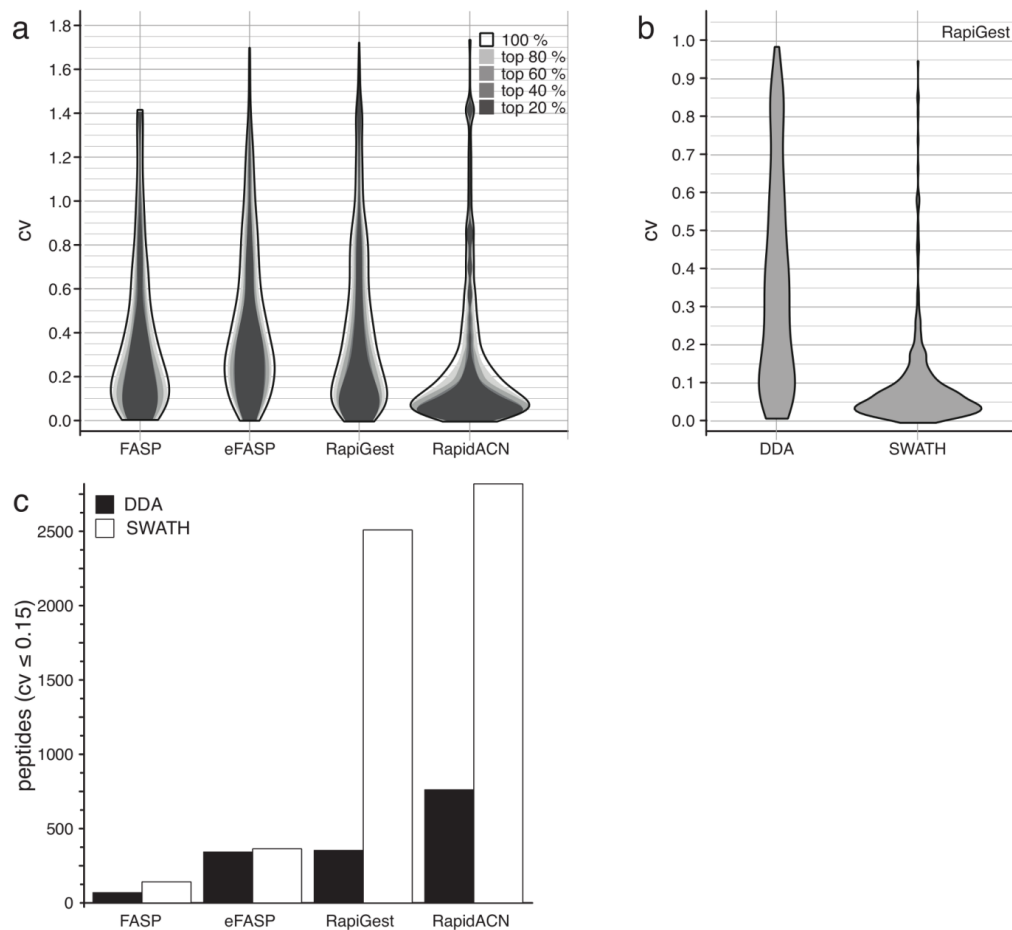


Abbildung 1: SWATH-MS Proteomics und SWATH Probenpräparationsmethoden (Vowinckel et al, 2013). a) Verschiedene Proteinpräparationen desselben Proteinextrakts analysiert mit SWATH-MS. In allen Fällen wird eine sehr niedrige durchschnittliche Varianz (dargestellt als CV Wert) erreicht; Die RpaidACN Methode entwickelt im Rahmen von TREAT20 (Bluemlein et al, 2011) erreicht die besten Werte. b) SWATH-MS im Vergleich zu klassischer Proteomik; CV Vergleich derselben Peptide gemessen mit SWATH-MS und klassischer datenabhängiger Proteomik (DDA). SWATH erreicht stabilere Quantifikationswerte. c) Anzahl präzise quantifizierter Peptide mittels SWATH-MS und klassischer DDA. In Präparationen mit hohem Peptid-Anteil werden bis zu 5x mal mehr Peptide präzise quantifiziert.

Die Analyseresultate aus SWATH wurden systematisch mit den Resultaten einer ebenfalls datengetriebenen Proteomanalyse mit einer seit kurzer Zeit am MPIMG verfügbaren Nanoflow-Umkehrphasen-Flüssigkeitschromatographie (RPLC) an einem Q-Plus-Exactive Orbitrap-Massenspektrometer (Thermo Scientific) verglichen.

Die Resultate, zum Beispiel bei der Proteomanalyse von Xenograftproben vor und während einer Behandlung mit entsprechenden Medikamenten, zeigten bereits die enorme Bedeutung dieser Information für die Modellvalidierung und Parameteroptimierung.

2. des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses;

Ein großer Teil der Regulationsprozesse in der Zelle laufen ‚post-transkriptionell‘, auf der Ebene von Proteinen, Proteinmodifikationszuständen, Metaboliten etc. ab. Die Verfügbarkeit systematisch und/oder punktuell generierter Proteomdaten eines Tumors stellen daher einen zentralen Beitrag zur Voraussage der Reaktion eines bestimmten Tumors auf ein bestimmtes Medikament dar.

Die Proteomkomponente des TREAT20 Projektes spielt daher eine wichtige Rolle in der Verbesserung und Validierung der Computermodelle des Tumors, kann aber auch wichtige Informationen zur Identifikation neuer Biomarkern liefern, die als ‚Companion Diagnostics‘ zu Stratifikation von Tumorpatienten verwendet werden könnten.

3. des während der Durchführung des Vorhabens dem Projektleiter bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen;

Während des TREAT20 Projektes ist das Wissen über die Biologie von Tumoren generell weiter gewachsen. So wurden in mehreren Projekten auch Proteomdaten generiert (<http://cancergenome.nih.gov/abouttcga/overview/howitworks/proteomecharacterization>).

Systematische Untersuchungen der gleichen Tumore durch eine Vielzahl verschiedener Untersuchungsverfahren, ähnlich der Untersuchungen in TREAT20, sind aber immer noch selten, und werden generell nicht als Bestandteil einer Modellierung des Tumors verwendet.

4. der erfolgten oder geplanten Veröffentlichungen des Ergebnisses.

Die Ergebnisse der gesamten Serie von 20 Patienten, darunter Proteomik-Daten werden dem Journal of Clinical Investigation zur Veröffentlichung eingereicht werden.

Literatur:

Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol.* 2008;4:222.

Bluemlein K, Ralser M. Monitoring protein expression in whole-cell extracts by targeted label- and standard-free LC-MS/MS. *Nat Protoc.* 2011 Jun;6(6):859-69.

Bluemlein K, Grüning NM, Feichtinger RG, Lehrach H, Kofler B, Ralser M. No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis. *Oncotarget.* 2011 May;2(5):393-400.

Bluemlein K, Glückmann M, Grüning NM, Feichtinger R, Krüger A, Wamelink M, Lehrach H, Tate S, Neureiter D, Kofler B, Ralser M. Pyruvate kinase is a dosage-dependent regulator of cellular amino acid homeostasis. *Oncotarget.* 2012 Nov;3(11):1356-69.

Vowinckel J, Capuano F, Campbell K, Deery MJ, Lilley KS, Ralser M. The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Research* 2013, **2**:272

WP7: Primary data analysis and modelling (Dr. Marie-Laure Yaspo, Dr. Ralf Herwig, Dr. Christoph Wierling, MPI-MG)

I. Kurze Darstellung zu

1. Aufgabenstellung

DNA- und RNA-Sequenz: Primary data analysis/Molekulare Analysen

In der TREAT20-Pilotphase haben wir NGS-Methoden zur Sequenzierung und Analyse der Genome/Exome und Transkriptome von 20 metastatischen Melanomen verschiedenen Typs (Kutan, Akral und Uveal) verwendet, um eine Basis für die Etablierung mechanistischer Modelle zu schaffen, mithilfe derer kombinatorische Effekte von molekularen Veränderungen in jedem einzelnen Patienten vorhergesagt werden. Das Ziel von WP6 war es eine NGS-Datenanalyse Pipeline zu etablieren, die in der Lage ist, eine umfassende Auflistung der krebisbedingten molekularen Veränderungen auf Genom- und Transkriptomebene zu erzeugen und diese Informationen zu integrieren. Dabei konnten wir eine Reihe neuartiger Melanom-assoziiertes biologischer Mechanismen und klinisch relevanter Varianten identifizieren. Diese Informationen waren essentiell, um das Computermodell optimal an diesen Tumortyp anzupassen und um die Voraussagen der Reaktion der Tumore auf bestimmte Medikamente zu verbessern. Damit wurden auch molekulare Daten zu den selteneren Melanomtypen erhalten, die auf Grund ihrer Unterrepräsentation trotz ihrer klinischen Bedeutung weniger genau untersucht sind.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde;

Die TREAT20-Pilotphase wurde vor mehr als drei Jahren eingeleitet, als die vorhandenen Analysewerkzeuge für die Genom- und Transkriptionsanalyse von Tumoren den Anforderungen nur teilweise gewachsen waren. Ein wichtiger Bestandteil unserer Arbeit war daher die Anpassung der Technologien und Datenanalysewerkzeuge an die enormen Anforderungen in der klinischen Anwendung. Neben anderen Herausforderungen, mussten wir die Programme auf die Frage der Analyse verdünnter Tumorproben, die eine höhere Sequenzierungstiefe, sondern auch zusätzliche zugeordnete Analyseschritte anzupassen. Innovative Schritte mussten auch für die Integration der Genom und Transkriptom-Daten überlegt und angepasst werden.

3. Planung und Ablauf des Vorhabens;

In der Anfangsphase, vor allem im ersten Projektjahr fokussierten wir uns auf die Feinabstimmung von Single Nucleotide Varianten Erkennungs- und Analyseprogramme. Wir benutzen unsere „inhouse“ Pipeline, die eine effiziente Identifizierung von somatischen Varianten (single nucleotide variants und small Insertions/deletions) sowie eine Bestimmung der chromosomalen copy number Variationen (ploidy status) ermöglicht.

Die zweite Herausforderung war es, die verschiedenen Ebenen der Informationen in einer angemessenen Weise zu integrieren, ein Ziel, das in der zweiten Phase des Projekts erreicht wurde.

4. wissenschaftlichem und technischem Stand, an den angeknüpft wurde, insbesondere

- Angabe bekannter Konstruktionen. Verfahren und Schutzrechte die für die Durchführung des Vorhabens benutzt wurden;
- Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste;

Die Errichtung der Analysepipeline stützte sich sowohl auf Open-Source-Programme als auch auf neu entwickelte maßgeschneiderte Algorithmen. Zu Beginn des Projektes testeten und bewerteten wir verschiedene „SNV caller“, einschließlich VARSCAN (Koboldt et al. 2012, SNVmix (Goya et al. 2010), und „inhouse“ entwickelte Programme, basierend auf Bayes-Statistiken. Die Herausforderung war die Spezifität-/Sensitivität-Schwellenwerte, die in der Tat probenabhängig sind, zu feinjustieren. Ein großer Teil der Arbeit wurde dieser Tätigkeit

gewidmet, da keine direkt verwendbare Lösung zur Verfügung stand.

5. Zusammenarbeit mit anderen Stellen;

Keine

6. Beitrag des Ergebnisses zu den förderpolitischen Zielen z.B. des Förderprogramms - (ggf. unter Angabe des Schwerpunkts) soweit dies möglich ist

Die hier beschriebenen Resultate leisten einen wichtigen Beitrag zur Entwicklung und Validierung der Systemmedizin, eines Schlüsselthemas der BMBF Förderung im Bereich der biomedizinischen Forschung (<http://www.bmbf.de/de/6942.php>).

7. Präsentationsmöglichkeiten für mögliche Nutzer - z. B. Anwenderkonferenzen (Angaben, soweit die Art des Vorhabens dies zulässt);

Keine

8. Erklärung zur Einhaltung der Ausgaben- und Zeitplanung.

Die entwickelte Analysepipeline läuft nun mit höchster Geschwindigkeit. Zum Projektende konnten wir die Analyseergebnisse mit Multi-Level-Informationen auf Genom- und Transkriptomebene der Tumore in einem zeitlichen Rahmen, der mit den klinischen Zeitrahmen vergleichbar ist generieren. Die benötigte Zeit von der Probenentnahme bis zur Abgabe der Medikamenten-Vorhersage beläuft sich derzeit nur noch auf etwa einen Monat, soll aber weiterhin verkürzt werden.

II. Eingehende Darstellung

1. des erzielten wissenschaftlich-technischen Ergebnisses, der erreichten Nebenergebnisse und die gesammelten wesentlichen Erfahrungen, bzw. der Arbeiten, die zu keiner Lösung geführt haben;

In der TREAT20-Pilotphase haben wir NGS-Methoden zur Sequenzierung und Analyse der Genome/Exome und Transkriptome von 20 metastatischen Melanomen verschiedener Typs (Kutan, Akral und Uveal) verwendet, um eine Basis für die Etablierung mechanistischer Modelle zu schaffen, mithilfe derer kombinatorische Effekte von molekularen Veränderungen in jedem einzelnen Patienten vorhergesagt werden. Dabei konnten wir eine Reihe neuartiger Melanom-assoziierten biologischer Mechanismen und klinisch relevanter Varianten identifizieren. Diese Informationen waren essentiell, um das Computermodell optimal an diesen Tumortyp anzupassen und die Voraussagen der Reaktion der Tumore auf bestimmte Medikamente zu verbessern. Damit wurden auch molekulare Daten zu den selteneren Melanomtypen erhalten, die auf Grund ihrer Unterrepräsentation trotz ihrer klinischen Bedeutung weniger genau untersucht sind.

Ein wichtiger Bestandteil unserer Arbeit war daher die Anpassung der Technologien und Datenanalysewerkzeuge an die enormen Anforderungen in der klinischen Anwendung. Eine wichtige Komponente hierbei war die Entwicklung von OncoScan, einer integrierten und vielseitigen NGS-Datenanalysepipeline, die auch bei Tumoren mit relativ niedrigem Reinheitsgrad die multiplen Genom- und Transkriptionsänderungen in den Tumoren identifizieren und integrieren kann. Diese Analysepipeline ist jetzt voll funktionsfähig und liefert aufgrund der optimierten Filter sehr gute Resultate. Voraussetzung dafür war die Integration vieler neuer Funktionen und funktionelle Annotation, die Verbesserung der Filterschritte und schließlich die Bereitstellung von Katalogen von molekularen Daten, die bei der Modellierung berücksichtigt werden müssen.

Ausgehend von diesen Sequenzdaten haben wir die Variationen in der Anzahl der Chromosomenkopien (CNVs), somatische Einzelnukleotidaustausche (SNVs), kleine

Insertionen/Deletionen (InDels) sowie die Keimbahnvarianten der sequenzierten Gene in den Patienten bestimmt.

Die Patientenserie umfasste zwanzig Fälle (6 ALM, 6 UM, 7 CM und 1 Schleimhaut, mit einer Mehrzahl an sonnengeschützten Melanomen, inklusive Metastasen aus verschiedenen Körperregionen, einschließlich Haut, Lymphknoten, Leber, Lunge und Brust. Wir haben auch zwei Rückfälle analysiert, und zwar für die Patienten P9 (UM) und P14 (ALM). Weiterhin wurden die Mutationen der 20 Patienten identifiziert. Die Gesamtzahl der Mutationen variiert beträchtlich zwischen den Patienten.

Wir konnten zeigen, dass wichtige genetische somatische Veränderungen auf vielen Ebenen stattfinden können, darunter Austausch einzelner Basen, kleine Insertionen und Deletionen und Spleißvarianten, aber auch unerwartete Genfusionen aufgrund chromosomaler Deletionen und Translokationen, sowie fokale genomische Deletionen und Amplifikationen.

Viele der gefundenen Änderungen waren zuvor unbekannt und wären durch ein Mutations-Screening mit gängigen Arrays nicht gefunden werden können, selbst wenn diese nichtkanonische Mutationen abdeckten.

Durch Genomsequenzierung ist es zudem möglich, die Veränderungen auf der Ebene der Chromosomen zu untersuchen. Dabei zeigen die Genome der Patienten oft starke Abweichungen vom diploiden Chromosom mit komplexen Änderungen und Loss of Heterozygosity (LOH).

Die Analyse der Tumor-Expressionsprofile lieferte auf mehreren Ebenen weitere wichtige Informationen. Zum einen konnten wir den allelischen Expressionsstatus aller Mutationen bestimmen, die für die Modellierung relevant sind. Auch konnten wir Genexpressionssignaturen identifizieren, die spezifisch für bestimmte Melanom-Subtypen oder auch individuelle Tumorveränderungen sind. Die aus der Sequenzierung der Tumoren in der ersten Phase erhaltenen molekularen Ergebnisse haben eine große Bandbreite an neuem Wissen und auch unerwarteten Ergebnissen geliefert, wodurch unser Verständnis der Biologie des Melanoms und der genetischen und epigenetischen Heterogenität der Melanome vertieft wurde. Auch haben die Daten die Modellierung und die klinischen Anwendungen unterstützt.

Die molekularen Daten wurden von uns zusammengestellt und bilden die Basis eines Manuskripts über die Analyse des molekularen Profils, der biologischen Signalwege, Signaltransduktionsanalyse und Modellierung der 20 untersuchten Melanome, dessen baldige Einreichung im Journal of Clinical Investigation geplant ist.

2. des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses;

Zusammenfassend haben wir einen sehr wertvollen Datensatz zusammengetragen, der eine Vielzahl neuer Krebsmarker sowie für die Tumormodellierung nützliche Ansatzpunkte liefert. Auf Grund der Komplexität und Menge der produzierten Information dauert die Datenanalyse in enger Zusammenarbeit mit der Modellierung an.

3. des während der Durchführung des Vorhabens dem Projektleiter bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen;

Siehe WP5 Sequenzierung. Pleasance et al. 2010, Berger et al. 2012, Hodis et al. 2012, Krauthammer et al. 2012, Nikolaev et al. 2012.

4. der erfolgten oder geplanten Veröffentlichungen des Ergebnisses.

Die molekularen Daten wurden von uns zusammengestellt und bilden die Basis eines Manuskripts über die Analyse des molekularen Profils, der biologischen Signalwege, Signaltransduktionsanalyse

und Modellierung der 20 untersuchten Melanome, das in nächster Zukunft bei dem Journal of Clinical Investigation eingereicht werden soll.

Literatur (verwendete Algorithmen für die Analysepipeline)

Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman D, Murphy KP, Aparicio S, Shah SP. 2010. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics*. 2010 Mar 15;26(6):730-6.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012 Mar;22(3):568-76.

WP8: Establishment and refinement of predictive models (Dr. Christoph Wierling, Prof. Dr. Hans Lehrach, Prof. Dr. Bernhard Herrmann, MPI-MG, Prof. Dr. Reinhold Schäfer, Institute of Pathology, Charité)

Weiterentwicklung des Modellierungssystems und des Krebsmodells, Modellierung der Tumore und Tumormodelle

I. Kurze Darstellung zu Aufgabenstellung und Ablaufplanung des Teilprojekts 8

1. Aufgabenstellung

Das Teilprojekt 8 (Modellierung) umfasste im wesentlichen 3 Aufgaben:

Aufgabe 1: Aufsetzen der Modellierungsplattform und Adaption des Modells

Aufsetzen der Modellierungsplattform sowie Adaption und Verbesserung des existierenden Modells hinsichtlich der relevanten Krebsentitäten, z.B. durch Einpflegen zusätzlicher melanomspezifischer Signaltransduktionskaskaden. Eine weitere Aufgabe bestand in der Optimierung der Datenintegration, Simulation und Datenauswertung.

Aufgabe 2: Identifizierung der Effekte patientenspezifischer Mutationen und deren Integration in das Modell

Basierend auf den identifizierten Mutationsprofilen sollte das existierende Modell mit weiteren Pathways relevanter Signalwege und genregulatorischer Effekte erweitert werden. Des Weiteren sollten hierbei auch Ergebnisse der Transkriptom- und Proteomanalysen berücksichtigt werden.

Aufgabe 3: Simulation der Mutationseffekte und Therapieoptimierung

Basierend auf den patientenspezifischen individualisierten Modellen sollten Simulationen zur Validierung der Mutationseffekte und zur Optimierung von Wirkstoffkombinationen durchgeführt werden. Die aus diesen Untersuchungen abgeleiteten Hypothesen sollten für die funktionale Analyse der patientenspezifischen Daten sowie für Modellvalidierung genutzt werden. Dies sollte Vorschläge für geeignete Therapieansätze liefern.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Eine wesentliche Komponente dieses Teilprojekts war die Integration der existierenden und im Verlaufe dieses Projekts anfallenden Daten, sowie deren Verwaltung, Analyse, Integration und Verknüpfung mit bereits verfügbaren Daten. Das Teilprojekt Modellierung stellt dabei den entscheidenden Schritt von der Datenerfassung zur Modellierung der biologischen Prozesse dar. Insbesondere stellte hierbei die mathematische Beschreibung krebsrelevanter Signaltransduktionswege und die Integration patientenspezifischer Expressions- und Mutationsinformationen eine zentrale Aufgabe dieses Teilprojekts dar. Derartige systemweite Ansätze sind erforderlich für das Verständnis der molekularen Komplexität von Krankheiten, wie Krebs.

Das Teilprojekt verfolgt dabei einen interdisziplinären Ansatz und steht an der Schnittstelle von Biologie, Medizin, Mathematik und Informatik. Leistungsfähige bioinformatische Werkzeuge standen bereits zu Beginn des Projekts zur Verfügung. Im speziellen kam hier die Modellierungs- und Simulationsplattform PyBioS¹ zum Einsatz, die bereits im Vorfeld am Max-Planck-Institut für molekulare Genetik entwickelt wurde. PyBioS ist ein Modellierungssystem, das es ermöglicht mathematische Modelle molekularbiologischer und zellulärer Prozesse, wie Signal- oder Stoffwechselwege sowie genregulatorischer Prozesse, zu erstellen und diese anschließend für *in silico* Analysen dieser Modelle zu verwenden.

1 <http://pybios.molgen.mpg.de>

3. Planung und Ablauf des Vorhabens

Zu Beginn des Projekts wurde der für die Modellierung im Rahmen dieses Projekts genutzte PyBioS Server aufgesetzt. Parallel hierzu wurde an der Optimierung des PyBioS Systems gearbeitet, das für die Erreichung der Projektziele erforderlich war. Zudem wurde parallel an der Modellverfeinerung und Optimierung, sowie der in silico Prüfung geeigneter Medikamente oder Medikamentenkombinationen für eine Behandlungsempfehlung gearbeitet. Hierbei handelt es sich um einen iterativen Prozess der während der gesamten Projektlaufzeit erfolgte. Die Details hierzu gehen aus der eingehenden Darstellung der Ergebnisse des Teilprojekts 8 hervor (siehe unten).

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Zu Beginn des Projekts stand das für die Modellierung und Simulation biologischer Systeme entwickelte Modellierungssystem PyBioS in der Version 1 zur Verfügung. Zudem existierte schon eine erste Generation des Computermodells, die 572 Gene umfasste sowie die jeweiligen mRNAs, Proteine und modifizierte Proteine (z.B. Phosphoproteine) als auch Komplexe aus diesen Proteinen. Zusätzlich waren mutierte Formen von 47 verschiedenen Onko- und Tumorsuppressor-Genen sowie 41 zielgerichtete Medikamente (wobei nur die Hauptziele der Medikamente abgebildet wurden) im Modell integriert. Insgesamt umfasste das Modell 2.839 Komponenten, welche durch 4.439 Reaktionen miteinander verbunden waren. Die in den Tumoren der Patienten identifizierten Mutationen mussten während des Prozesses der Erzeugung von patientenspezifischen Modellen manuell integriert werden. Auch die Bindungsaffinitäten der verschiedenen Medikamente zu ihren Hauptzielen mussten für die Drugoptimierung manuell eingestellt werden.

5. Zusammenarbeit mit anderen Stellen

Die Arbeiten zur Datenintegration und Modellierung wurden in intensiver Zusammenarbeit und im Austausch von Know-How und Projektergebnissen mit den Projektpartnern innerhalb des TREAT20 Verbundes durchgeführt.

6. Beitrag des Ergebnisses zu den förderpolitischen Zielen. z. B. des Förderprogramms - (ggf. unter Angabe des Schwerpunkts) soweit dies möglich ist;

Die hier beschriebenen Resultate leisten einen wichtigen Beitrag zur Entwicklung und Validierung der Systemmedizin, eines Schlüsselthemas der BMBF Förderung im Bereich der biomedizinischen Forschung (<http://www.bmbf.de/de/6942.php>).

7. Präsentationsmöglichkeiten für mögliche Nutzer - z. B. Anwenderkonferenzen (Angaben, soweit die Art des Vorhabens dies zulässt);

Die Projektdaten wurden in verschiedenen Vorträgen (innerhalb von Tagungen und bei Seminaren) erwähnt. Eine Publikation ist in Vorbereitung.

8. Erklärung zur Einhaltung der Ausgaben- und Zeitplanung.

Die Modellierung der einzelnen Patienten wurde entsprechend der Verfügbarkeit der jeweiligen Expressionsdaten durchgeführt und entspricht dem ursprünglich geplanten Umfang. Das PyBioS System wurde entsprechend der Erfordernisse erweitert, die für die Durchführung der geplanten Simulation erforderlich waren. Die entstandenen Kosten entsprechen der ursprünglichen Kostenplanung.

Die Modellierung der einzelnen Patienten wurde entsprechend der Verfügbarkeit der jeweiligen Expressionsdaten durchgeführt und entspricht dem ursprünglich geplanten Umfang. Das PyBioS System wurde entsprechend der Erfordernisse erweitert, die für die Durchführung der geplanten Simulation erforderlich waren. Die Modellentwicklung und Erweiterung sowie die Simulationen für

Medikamentvorhersage waren in wesentlichen Teilen abhängig von den Ergebnissen der Verfügbarkeit der Patientenproben und der entsprechenden nachgestellten Analysen des Genoms und Transkriptoms. Ausgehend von den hierbei identifizierten Mutationen konnte das Modell daher erst im Anschluss durch die entsprechenden funktionalen Effekte der Mutationen und weiterer hiermit assoziierter Signalwegskomponenten und Pathways erweitert werden. Hierdurch kam es zu Verzögerungen im Projektlauf, die eine einjährige kostenneutrale Verlängerung erforderlich machte. Die entstandenen Kosten entsprechen der ursprünglichen und im Rahmen der kostenneutralen Verlängerung bewilligten Kostenplanung.

II. Eingehende Darstellung der Projektergebnisse des Teilprojekts 8

1. Darstellung der erzielten wissenschaftlich-technischen Ergebnisse

Im Laufe des Pilotprojektes haben wir mehrere Signalwege um zusätzliche aktivierende Liganden, wie AREG, EREG, IGF2, IL13 und IL14 erweitert. Des Weiteren sind zusätzliche Signalwege, die insbesondere im Melanom relevant sind, in das vorhandene Krebsmodell integriert worden. Insbesondere haben wir Mutationen, die laut der COSMIC-Datenbank (Forbes et al. 2006) häufig in Melanomen gefunden wurden, sowie bisher noch unbekannte Mutationen und Genfusionen, die wir durch die Sequenzierung während des Projekts identifiziert haben im Modell implementiert. Dabei wurden entsprechende umliegende Signalnetzwerke integriert, erweitert bzw. modifiziert. Darüber hinaus haben wir weitere Inhibitor-Komponenten in das Modell integriert, die es uns ermöglichten, die Wirkungen von Medikamenten nun unter Berücksichtigung von mehr als 95 verschiedenen Wirkstoffzielen (d.h. auch sogenannten off-target-Effekten) zu simulieren. Insgesamt deckt das aktuelle Modell damit mehr als 620 Gene ab, die zu 3.397 Komponenten (mRNAs, Proteine, modifizierte Proteine und Komplexe) führen und durch 5.456 Reaktionen miteinander verbunden sind.

Entwicklung von Datenbanken

Im Laufe des Projekts wurde eine Mutations/Fusions-Datenbank entwickelt, die derzeit 364 unterschiedliche Mutationen in 51 verschiedenen Onkogenen und Tumorsuppressorgenen sowie 72 verschiedene Fusionsereignissen in 7 verschiedenen Genen umfasst. Zudem wurde eine Medikament-Datenbank entwickelt, die derzeit 84 molekulare zielgerichtete Medikamente (sowohl Krebsmedikamente als auch für andere Krankheiten zugelassene Medikamente) umfasst. Die Mutations/Fusions-Datenbank beinhaltet die molekularen Konsequenzen der verschiedenen Mutationen/Fusionen in den verschiedenen Genen und erlaubt nun eine automatische Generierung von patientenspezifischen Tumormodellen. Die Medikament-Datenbank umfasst die jeweiligen Hauptziele und off-/side-targets-Profile der verschiedenen Medikamente sowie die dazugehörigen Bindungsparameter von mehr als 80 unterschiedlichen molekular zielgerichteten Medikamenten und ermöglicht so eine automatische Simulation der verschiedenen Medikamente während der Therapieoptimierung.

Erweiterung und Optimierung der Modellierungsplattform PyBioS

Während der Projektlaufzeit wurde zudem das Modellierungs- und Simulationssystem PyBioS (Wierling et al. 2007, Klipp et al. 2009) überarbeitet, optimiert und automatisiert, um Simulation individueller Mutations- und Medikamentwirkungen zu ermöglichen und verschiedene, individuelle Kombinationen von Mutationen und Medikamenten zu testen. Insbesondere wurde das Datenbankkonzept des PyBioS Systems vollkommen neu überarbeitet, um eine einfache, einheitliche und benutzerfreundliche Erstellung von Modellen zu ermöglichen. Hierfür wurde einerseits ein Postgres-Datenbankkonzept entwickelt und implementiert, das die Speicherung auch sehr großer Modelle und Modelldaten erlaubt, wie sie im Rahmen dieses Projektes erforderlich waren. Um eine einheitliche und korrekte Annotation von Genen, mRNA, Protein und Metabolitobjekten zu ermöglichen, wurden entsprechende Annotationsinformationen aus den Datenbanken Ensembl und ChEBI in die neue PyBioS Referenzdatenbank PyDB importiert. Desweiteren wurde die Benutzerschnittstelle optimiert um eine schnelle und fehlerfreie Modellannotation und -erweiterung zu ermöglichen. Insbesondere wurde hierfür das Konzept der „Templates“ eingeführt. Bei einem Template handelt es sich hierbei um eine Vorschrift zur systematischen Eingabe sehr ähnlicher zellulärer Abläufe, die jedoch auf unterschiedliche Objektentitäten angewendet werden können. So wurde z.B. ein Template für die Expression eines

Gens entwickelt. Dieses Template erlaubt dann, ausgehend von einem bestimmten Gen, alle für die Genexpression erforderlichen Reaktionen, wie eine Transkription oder Translation, automatisch zu erstellen. Das Template berücksichtigt dabei auch die korrekte Annotation der dem Gen entsprechenden mRNA- und Protein-Instanzen unter Berücksichtigung alternativer Splice-Formen. Entsprechende Templates wurden auch für andere zelluläre Prozesse entwickelt und implementiert, so z.B. für die Komplexbildung, dem Protein- oder Komplexabbau, oder der Phosphorylierung von Proteinen unter Berücksichtigung der Phosphorylierungspositionen. Diese Templates erlauben eine schnellere Erstellung von Modellen, wie es insbesondere für die Generierung großer Modelle unabdingbar ist. Neben entsprechenden Objektklassen für reguläre zelluläre Prozesse wurden auch Objektklassen für mutierte Gene oder Proteine, sowie Medikamente (Drugs) entworfen, die für die Simulation der patientenspezifischen individuellen Mutationsprofile und der in silico Untersuchung von Medikamentwirkungen erforderlich waren. Für alle Reaktionen der Templates wurden zudem geeignete kinetische Gesetze entwickelt und implementiert, die quantitative Simulation der Modelle, mit Hilfe gewöhnlicher Differentialgleichungssysteme, erlaubt. Das PyBioS Modellierungssystem liegt derzeit in der Version 3 vor.

Klinische Anwendung in der Pilotphase TREAT20

Während der Pilotphase haben wir einzelne Modelle für 21 verschiedene Proben von 20 metastasierten malignen Melanom-Patienten als auch 3 verschiedene Melanozyten-Zelllinien (als Kontrolle dienend) auf Basis der durch Sequenzierung gewonnenen Genom- und Transkriptomdaten mit Hilfe des Modellierungssystems ModCell, welches das PyBioS System für die Modellentwicklung nutzt, erstellt. Da in einigen Patienten neue Mutationen identifiziert wurden, deren funktionelle Konsequenz unbekannt waren, wurden für einige der Patienten mehrere Modelle erstellt, in denen für die unbekannt Mutationen hypothetische gain-of-function- bzw. loss-of-function-Konsequenzen angenommen wurden. Mit Hilfe des virtuellen Patientenmodells wurden Signalwegsanalysen durchgeführt, um Einsichten in die molekularen Strukturen der verschiedenen Tumoren zu erhalten.

Die Kombination aus NGS und Modellierung macht deutlich, dass sich auf molekularer Ebene deutlich mehr Melanom-Subtypen identifizieren lassen.

Anschließend wurden diese patientenspezifischen Modelle (es wurden nur die Patientenmodelle berücksichtigt, die keine hypothetischen Mutationsfunktionen beinhalteten) verwendet, um die Wirkung der in der Medikament-Datenbank enthaltenen Wirkstoffe mit Hilfe eines Monte-Carlo-Ansatzes (Wierling et al., 2012) zu simulieren. Alle Modellierungsergebnisse wurden den behandelnden Ärzten zur Verfügung gestellt, um diese bei ihrer Entscheidung zur Wahl einer möglichen Behandlung zu unterstützen.

2. Voraussichtlichen Nutzens, Verwertbarkeit der Ergebnisse

Die im Rahmen des Projekts entwickelten Grundlagen, Erkenntnisse und entwickelten Modelle krebsassoziierter zellulärer Signaltransduktionswege können in Folgeprojekten, die sich ebenfalls mit der Modellierung krebsassoziierter Signaltransduktionswege beschäftigen, genutzt werden. Das im Rahmen dieses Projekts entwickelte Modell krebsassoziierter Signalnetzwerke ist bereits Gegenstand eines geplanten Folgeprojekts.

3. Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Keine bekannt.

4. Veröffentlichungen der Ergebnisse

Kühn A & Lehrach H (2012) The 'Virtual Patient' system: modeling cancer using deep sequencing technologies for personalized cancer treatment. *Journal für Verbraucherschutz und Lebensmittelsicherheit* **7**: 55–62

Wierling C, Kühn A, Hache H, Daskalaki A, Maschke-Dutz E, Psycheva S, Li J, Herwig R & Lehrach H (2012) Prediction in the face of uncertainty: a Monte Carlo-based approach for systems biology of cancer treatment. *Mutat. Res.* **746**: 163–170