

**Förderkennzeichen:** 01KU1201E

**Vorhabenbezeichnung:** Internationales Krebsgenomkonsortium (ICGC) PedBrainTumor

**Arbeitspaket:** AP8 – Transkriptomanalyse

**Ausführende Stelle:** Max-Planck-Institut für Molekulare Genetik, Berlin

**Zuwendungsempfänger:** Prof. Dr. Hans Lehrach

**Projektleiterin:** Dr. Marie-Laure Yaspo

## **SCHLUSSBERICHT FÜR DEN VORHABENSZEITRAUM 01.07.2012 bis 31.12.2014**

---

### **I. Kurze Darstellung**

#### **1. Aufgabenstellung**

In den vergangenen Jahren haben Studien an Krebsgenomen eine Vielzahl von Mutationen in vielfältigen Kombinationen aufgedeckt. Daraus resultieren zahlreiche neue Erkenntnisse zu Tumor-Pathogenesemechanismen mit potentieller Anwendung in der Diagnose und der Behandlung der betreffenden Tumorerkrankungen. Dank neuer Entwicklungen im Bereich von *Next Generation*-Sequenziermethoden ist es nun möglich, ganze Genome und Transkriptomte in kurzer Zeit zu sequenzieren und zu analysieren. Auf der Basis von Genom- und Transkriptomsequenzierung war es das Ziel des ICGC PedBrainTumor-Projekts, eine umfassende Analyse einer großen Zahl an pädiatrischen Hirntumoren durchzuführen, die die pädiatrischen Tumore mit der höchsten Mortalitätsrate darstellen.

Die genomische Sequenzanalyse von pädiatrischen Hirntumoren sollte dabei durch detaillierte Transkriptom- und Methylomanalyse ergänzt werden. Aufgabe des Arbeitspakets 8 – Transkriptomanalyse – war dabei die Erstellung und Auswertung von hoch aufgelösten Transkriptomdatensätzen. Die molekularen Daten sollten mit klinischen Parametern korreliert werden, um neue prognostische und prädiktive Marker sowie mögliche Ansatzpunkte für Therapien zu identifizieren.

#### **2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde**

In den vergangenen Jahren sind Hirntumore zur Hauptursache von krebisbedingten Todesfällen bei Kindern geworden, obwohl sie nur halb so häufig auftreten wie Leukämien. Darüber hinaus leiden betroffene Kinder unter schwerwiegenden Folgen nach neurochirurgischen Eingriffen an sensiblen anatomischen Hirnstrukturen, die während einer anfälligen Phase der Gehirnentwicklung stattfinden. Auch führen die notwendigen intensiven multi-modalen Chemotherapien langfristig zu neurologischen, kognitiven, endokrinologischen und psychologischen Nebenwirkungen.

Die häufigsten pädiatrischen Hirntumore sind Astrozytome (42%, meist Pilozytische Astrozytome), Medulloblastome (18%), Ependymome (10%), Supratentoriale Primitive Neuroektodermale Tumore (PNET, 7%) und Craniopharyngiome (4%). Die 5-Jahres-Überlebensrate für alle Hirntumore lag bei nur 64%, was ein vergleichsweise schlechtes Resultat bei den pädiatrischen Krebsfällen darstellt.

Im Gegensatz zu pädiatrischen Leukämien und anderen soliden Tumoren bei Kindern waren zum Projektstart noch kaum molekulare Marker bekannt, mithilfe derer eine individuelle Risikoabschätzung bei pädiatrischen Hirntumoren erfolgen könnte. Das könnte auch eine Erklärung für die geringe Verbesserung der Heilungsraten bei diesen Tumoren sein. Tiefere Einblicke in die molekularen Pathomechanismen und Korrelation mit klinischen Folgedaten sollten daher verwendet werden, um die Situation der Patienten verbessern zu können.

Im Arbeitspaket 8 – Transkriptomanalyse – des ICGC PedBrainTumor-Projekts sollte der Fokus vor allem auf den beiden häufigsten pädiatrischen Hirntumorarten liegen (Astrozytome und Medulloblastome), welche in den ICGC-Richtlinien beide auch explizit als besonders relevante pädiatrische Entitäten aufgeführt wurden. Darüber hinaus sollte aber auch eine erste Charakterisierung weiterer pädiatrischen Hirntumore möglich sein, je nach Verfügbarkeit von Probematerial aus betroffenen Patienten.

Für das Projekt konnte die am MPI-MG vorhandene Infrastruktur zur RNA-Sequenzierung und -Analyse verwendet werden. Das MPI-MG ist seit 1995 an großen Sequenzierprojekten beteiligt. Dr. Marie-Laure Yaspo hat bereits die Arbeiten an der Sequenzierung des humanen Chromosoms 21 koordiniert, ihre Gruppe war an zahlreichen anderen Sequenzierstudien beteiligt und verfügt über die notwendige Expertise. Im Rahmen von ICGC PedBrainTumor sollte diese Expertise zur Datengenerierung und -analyse verwendet werden und dabei auch eine Weiterentwicklung der verwendeten Analysetechniken stattfinden.

### **3. Planung und Ablauf des Vorhabens**

Um das Projektziel zu erreichen, hat sich ein Netzwerk von Experten im ICGC PedBrainTumor-Konsortium zusammengeschlossen. Dieses Konsortium verfügt über (1) eine exzellente Sammlung von Tumoren mit gut dokumentierten molekularen und klinischen Charakteristiken, (2) international anerkannte Experten für die Identifikation von genomischen Rearrangements und DNA-Mutationen mittels Paired-end-Sequenzierung sowie für Hochdurchsatz-Genom-, Transkriptom- und Methylom-Sequenzierung, und (3) eine international führende Bioinformatik mit Zugang zu großskaligen Ressourcen für Datenhaltung und -analyse. Das Netzwerk baute auf bereits gut etablierten Interaktionen zwischen Partnern mit ausgewiesener klinischer, molekulargenetischer und genomischer sowie bioinformatischer Expertise auf.

Im Rahmen dieses Netzwerks erhielt das MPI-MG RNA-Material aus Patientenproben, welches zuerst einer Qualitätskontrolle unterzogen wurde. Alle RNAs mit ausreichender Quantität und Integrität wurden zur Herstellung von RNA-Sequenzierbibliotheken für *Next Generation Sequencing* auf der Illumina-Plattform verwendet. Im Vorhabenszeitraum konnte so für insgesamt 372 Patientenproben die RNA sequenziert und die Sequenzierdaten dem Projektkonsortium zur Verfügung gestellt werden.

Nach der Sequenzierung wurden die Reads auf das humane Referenzgenom gemappt. Anschließend erfolgte die Bestimmung der Genexpressionswerte (RPKM-Werte), die Identifizierung von Fusionsgenen, alternativer Exon-Nutzung, differentiell exprimierter Gene inklusive Identifizierung von molekularen Subclustern, die Identifizierung von Genregulationsnetzwerken und Transkriptom-Assemblierung zur Identifizierung bisher unbekannter neuer Transkripte. Als Teil dieser Arbeit hat das MPI-MG neue Bioinformatikpipelines für die Transkriptomanalyse entwickelt, die Transkriptom- und Genomdaten integrieren. Für dieses Projekt wurden insbesondere Verfahren zur sensitiven Identifikation von *bona fide*-Fusionsgenen und alternativen Spliceformen, die spezifisch mit diesen Tumoren assoziiert sind, entwickelt. Alle fertigen Datensätze wurden den Projektpartnern zur Integration mit Daten aus anderen Sequenzierplattformen und mit klinischen Daten zur Verfügung gestellt. Für den Medulloblastom-Datensatz befindet sich die Auswertung noch in Arbeit, eine Veröffentlichung ist im Rahmen einer weiteren Konsortium-Publikation geplant.

#### **4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde**

- **Angabe bekannter Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden**

Es wurden keine bekannten Konstruktionen, Verfahren oder Schutzrechte bei der Durchführung des Vorhabens verwendet.

- **Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste**

Internationale medizinische, biomedizinische, biochemische und molekularbiologische Fachzeitschriften.

#### **5. Zusammenarbeit mit anderen Stellen**

Im Rahmen dieses Projekts mit keinem anderen externen Partner als den Konsortiumsmitgliedern des ICGC PedBrainTumor-Projekts.

## II. Eingehende Darstellung

### 1. Verwendung der Zuwendung und erzielte Ergebnisse im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

Die Projektziele sind die Erstellung von hoch aufgelösten, auf RNA-Sequenzierung (RNA-seq) basierten Expressionsprofilen von Tumoren und Kontrollgeweben sowie die zugehörige umfassende Transkriptomanalyse in Zusammenarbeit mit WP10 (Bioinformatik). Im Laufe des Projekts wurde für folgende 372 Patientenproben die RNA sequenziert und die Sequenzierdaten zur Verfügung gestellt:

- Medulloblastome (MB): 189
- Pilozytische Astrozytome (PA): 90
- Glioblastome (GBM): 50
- Ependymome: 25
- ETANTR: 7
- Primäre Glioblastome (pGBIV): 3
- Kontrollgewebe: 8

Die Herstellung der Sequenzierlibraries erfolgte mit dem TruSeq Strang-spezifischen Protokoll mit RiboZero Gold (Illumina), sequenziert wurden die Libraries auf der Illumina HiSeq 2000-Plattform mit 2x51 Zyklen. Das Demultiplexing der fastq-Dateien erfolgte mit der bcl2fastq Konversionssoftware (Illumina, Version 1.8.4). Die resultierenden Sequenzreads wurden mit der Software bwa (Version 0.7.7-r441) auf das menschliche Referenzgenom (1000Genomes-Version) Phase1 (human\_g1k\_v37.fasta) gemappt, wobei die Chromosomen 1 bis 22, X, Y und M verwendet wurden.

Die ursprüngliche Vorhabensbeschreibung umfasste die Sequenzierung von 300 Medulloblastomen und 300 Pilozytischen Astrozytomen. Im Laufe des Vorhabenszeitraums und nach Analyse der ersten resultierenden Datensätze wurde diese Ziele nach Zustimmung des Scientific Advisory Boards von den Projektpartnern und Projektträger dahingehend modifiziert, dass die Pilozytischen Astrozytome mit 90 Proben bereits weitestgehend charakterisiert sind, da diese größtenteils durch Fusionen oder Mutationen des *BRAF*-Gens hervorgerufen werden. Auch wurde beschlossen, weniger Proben zu sequenzieren und dafür die einzelnen Proben in viel höherer Sequenziertiefe zu analysieren (>300 Million Reads pro Probe). Zusätzlich wurden auch noch Glioblastome sequenziert, bei denen neuartige *MET*-Fusionen gefunden wurden (siehe unten).

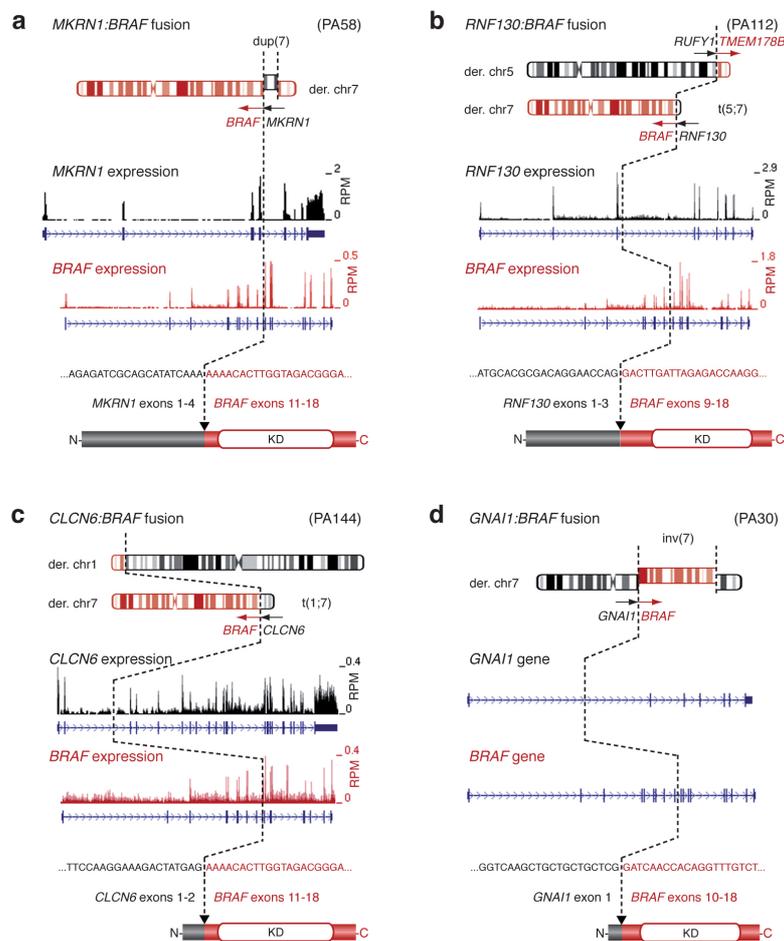
#### *Identifizierung von Fusionsgenen*

Einen wichtigen Teil der Transkriptomanalyse stellt die Suche nach Fusionsgenen dar. Wir haben im Rahmen von ICGC PedBrainTumor unsere Detektionspipeline weiterentwickelt (basierend auf den Programmen TopHat2-Fusion und SOAPfuse sowie eigener Postprozessierungs-Software) und für die Analyse von RNA-seq-Daten aus 79 Pilozytischen Astrozytomen, 49 Glioblastomen (GBM), 48 Ependymomen, 23 Atypische Teratoide/Rhabdoide Tumoren (ATRT), 164 Medulloblastomen (MB) sowie 3 Primitiven

Neuroektodermalen Tumoren des Zentralen Nervensystems (PNET) verwendet. Im Detail wurden daraus folgende Erkenntnisse gewonnen:

- 79 Pilozytische Astrozytome (PA):

Unsere Fusionsgen-Analyse identifizierte sowohl bekannte MAPK-Pathway-aktivierende Genomveränderungen wie Fusionsvarianten von KIAA1549-BRAF (70 Fälle) und FAM131B-BRAF (1 Fall) als auch vier neue BRAF-Fusionen (Abbildung 1). Die BRAF-Fusionen wurden sowohl durch den Vergleich mit Mate-Pair-Daten als auch mit RT-PCR-Sequenzierung bestätigt. Wie erwartet resultierten alle Varianten im Verlust der aminoterminalen regulatorischen Region von BRAF. Eine RNF130-BRAF-Fusion, entstanden durch reziprokale t(5;7)(q35;q34)-Translokation, wurde in zwei Patienten gefunden, andere Fälle von neuen BRAF-Fusionen zeigten die Fusionsgene CLCN6-BRAF, MKRN1-BRAF und GNAI1-BRAF.



**Abbildung 1. Neue BRAF-Fusionsgene in Pilozytischen Astrozytomen (PA).** (a) MKRN1-BRAF-Fusion in PA58. (b) RNF130-BRAF-Fusion in PA112. (c) CLCN6-BRAF-Fusion in PA144. (d) GNAI1-BRAF-Fusion in PA30. RPM, reads per million; KD, Kinase-Domäne.

Unsere Daten zu bekannten und neuen Fusionsgenen in Pilozytischen Astrozytomen bildeten einen wichtigen Teil der Veröffentlichung im Rahmen eines Konsortiumsartikels in Nature Genetics (Jones et al. 2013).

- 49 Glioblastome (GBM):

Aus der Fusionsgen-Analyse ergab sich eine Liste von 85 potentiell schädigenden Fusionsgenen. 51 besonders interessante Fusions-Kandidaten wurden mittels RT-PCR und

Sanger-Sequenzierung als echte Fusionsgene bestätigt. In Kooperation mit dem DKFZ Heidelberg wurde ein auf diesen GBM-Fusionsgenanalysen basierendes Manuskript erstellt, welches besonders das wiederholte Auftreten von *MET*-Fusionen, deren Beteiligung bei der Glioblastom-Entstehung sowie potentielle Therapiemöglichkeiten beschreibt und momentan für eine Veröffentlichung eingereicht ist (Bender *et al.* submitted).

- 48 Ependymome:

In den RNA-seq-Daten aus 48 Ependymomen wurden fünf potentiell schädliche Genfusionen identifiziert, darunter keine mehrfach auftretenden Fusionen. Die Ergebnisse der Fusionsanalysen wurden an den Projektpartner am DKFZ Heidelberg übermittelt und im Rahmen eines Konsortiumsartikels in *Cancer Cell* veröffentlicht (Pajtler *et al.* 2015).

- 23 Atypische Teratoide/Rhabdoide Tumore (ATRT):

In den RNA-seq-Daten aus 23 ATRTs wurden acht potentiell schädliche Genfusionen identifiziert, darunter keine mehrfach auftretenden Fusionen. Die Ergebnisse der Fusionsanalysen wurden an den Projektpartner am DKFZ Heidelberg übermittelt.

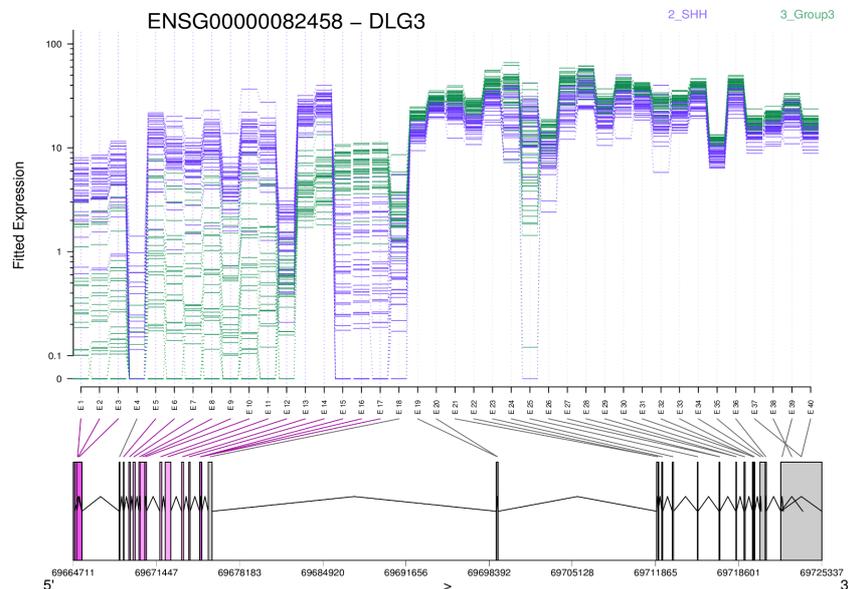
- 164 Medulloblastome (MB):

Aus der Fusionsgen-Analyse ergab sich eine Liste von 293 exprimierten Genfusionen in 86 verschiedenen Patienten. Wie erwartet traten Genfusionen vor allem in Medulloblastomen der Gruppen 3 und 4 auf, jedoch wurden auch eine Reihe von Fusionen in der SHH-Gruppe identifiziert und sogar einige wenige Fusionen in WNT-Medulloblastomen. Eine erste Veröffentlichung eines Teils der Medulloblastom-Daten fand bereits 2012 im Magazin *Nature* statt (Jones *et al.* 2012). Wiederholt auftretende Fusionen des *GFI1B*-Onkogens wurden im Rahmen einer weiteren Veröffentlichung in *Nature* beschrieben (Northcott *et al.* 2014). Nach Analyse des gesamten Datensatzes aus 164 Medulloblastomen wurden 23 Gene identifiziert, die in mehr als einem Patienten von Fusionen betroffen waren. Momentan analysieren wir die Fusionen im Detail, um ihre mögliche Funktion als Driver-Gene zu charakterisieren, indem die Daten mit den weiteren Datensätzen zu Single Nucleotide Variants (SNVs), Indels und Copy Number Variations (CNVs) sowie RNA-seq-Expressionsclustern und Expressionsmustern im Gehirn integriert werden. Die Veröffentlichung der Fusionsgen-Daten und anderer Ergebnisse zu allen analysierten Medulloblastomen ist im Rahmen einer umfassenden Transkriptomstudie von pädiatrischen Medulloblastomen für das Jahr 2015 geplant.

#### *Identifizierung alternativer Exon-Nutzung (AEN)*

Neben der Analyse von Fusionsgenen ist die Detektion alternativer Exon-Nutzung (AEN) ein weiterer Teil der Transkriptomanalyse. Im Rahmen von ICGC PedBrainTumor wurden Methoden zur Detektion, zur Visualisierung und zum Clustern von AEN entwickelt, die auf RNA-seq Daten basiert und alternative Junction-Nutzung (AIN) als zusätzlich Validierung integrieren. 164 Medulloblastome wurden auf subtypspezifische AEN untersucht. Insgesamt konnten 1362 Gene identifiziert werden, die AEN zwischen den Subtypen zeigen. Von diesen Genen sind 267 subtypspezifisch: 98 für WNT, 116 für SHH, 16 für Group 3 und 60 für Group 4. Die detektierten Fälle von AEN und AIN werden mittels Literaturvergleich ausgewertet, um mögliche funktionelle Analysen zu planen. Einige der so gefundenen alternativen Spleißformen, die mit bestimmten Medulloblastom-Subtypen assoziiert sind, wurden bereits vorher in bestimmten biologischen Kontexten beschrieben. Die interessantesten Beispiele

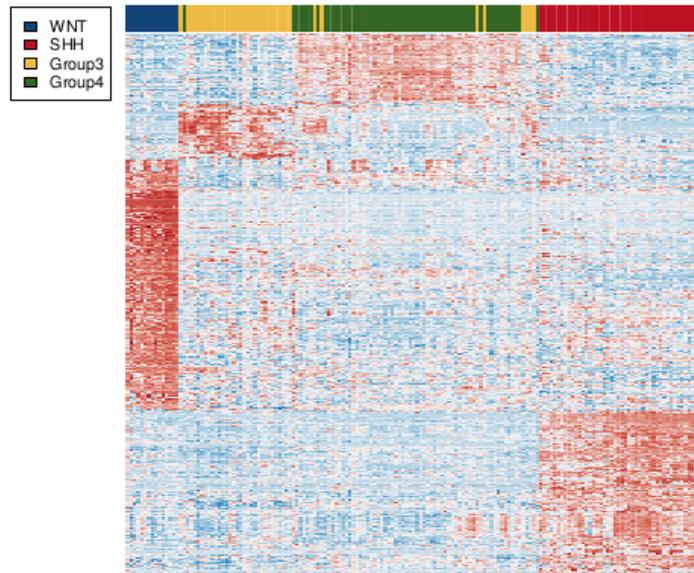
betreffen Gene, die an der Hirnentwicklung beteiligt sind, was Hinweise geben könnte sowohl auf den Ursprungszelltyp der Tumore als auch auf Ereignisse, die zu den transformierenden Prozessen in diesen Tumoren führen. Die in diesem Projekt gewonnenen Daten werden momentan eingehend analysiert.



**Abbildung 2. Alternative Exon-Nutzung im Gen DLG3 zwischen den Subtypen SHH und Group3.** Zu sehen ist die relative Expression der annotierten Exons von DLG3. Violett markierte Exons zeigen signifikante alternative Nutzung zwischen den beiden Subtypen.

### Identifizierung differentiell exprimierter Gene

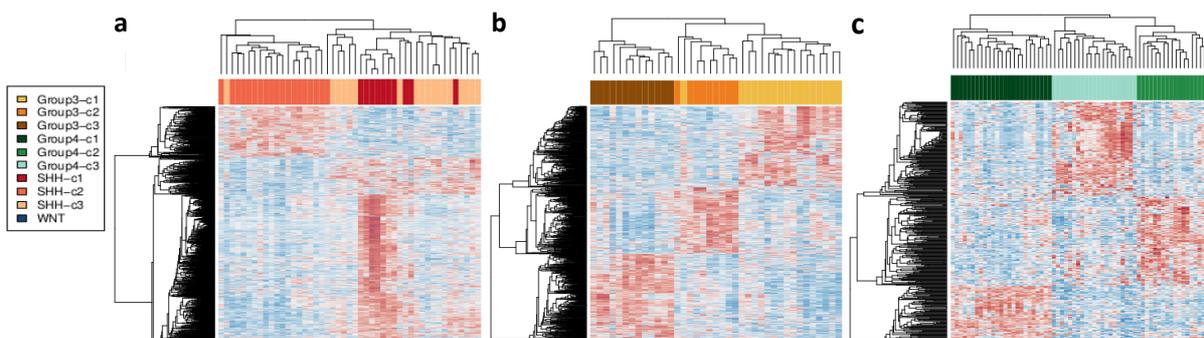
Parallel zur Detektion alternativer Exon-Nutzung wurden die (RiboZero) RNA-Sequenzierdaten der 164 Medulloblastome auf differentiell exprimierte Gene (DEG) untersucht (unter der Verwendung von edegR (Robinson *et al.* 2010) und eigener Postprozessierungs-Software). Hierbei standen die DGE zwischen den vier MB-Subtypen im Vordergrund, mit speziellem Fokus auf Group 3 und Group 4, welche am wenigsten verstanden sind. Neben den proteinkodierenden Genen stehen die langen nicht-kodierenden RNAs (lnc-RNA) im Fokus der differentiellen Genexpressionsanalyse. Die identifizierten deregulierten lnc-RNAs werden funktionell mittels *In-situ*-Hybridisierung (ISH) in Maus- und Makakengehirn untersucht, um Informationen zu Ort und Zeitpunkt ihrer Expression zu erhalten, was zur Aufklärung ihrer möglichen Funktion in genregulatorischen Netzwerken beitragen wird.



**Abbildung 3. Subtypespezifische DEG in ICGC PedBrain Medulloblastom.** Hierarchisches Clustern subtypespezifisch exprimierter proteinkodierender Gene.

#### Identifizierung von molekularen Subclustern in SHH, Group 3 und Group 4

Um heterogene Eigenschaften der Medulloblastom Subtypen zu verstehen, wurde an einer molekularen Subklassifizierung innerhalb der Subtypen SHH, Group 3 und Group 4 gearbeitet, da diese Subtypen die schlechtesten Prognosen in der Klinik ausweisen. Hierzu wurden Techniken von nichtüberwachtem hierarchischem Clustering verwendet. Innerhalb der drei Subtypen SHH, Group 3 und Group 4 konnten jeweils drei molekulare Subcluster, die subclusterspezifische differenzielle Genexpression aufweisen, identifiziert werden (Abbildung 4).

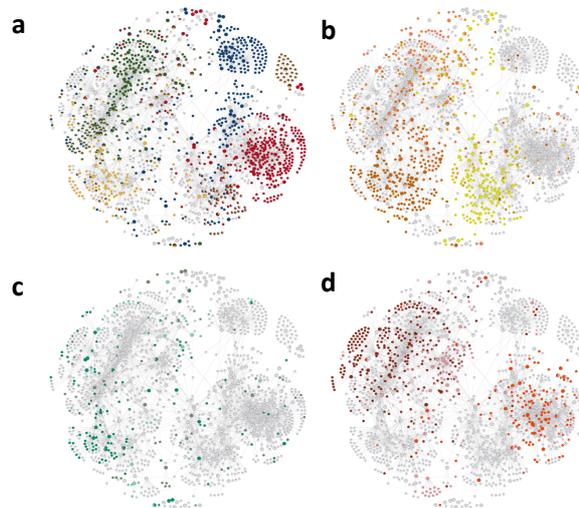


**Abbildung 4. Molekulare Subcluster in der ICGC PedBrain Medulloblastom Kohorte.** Differential exprimierte Gene zwischen den molekularen Subclustern in a) SHH, b) Group 3 und c) Group 4.

#### Identifizierung von Genregulationsnetzwerken

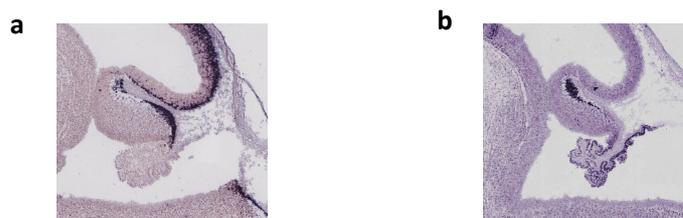
Die identifizierten differentiel exprimierten Gene deuten auf verschiedene aktivierte transkriptionelle Programme in den Subtypen und Subcluster. Zum besseren Verständnis der regulatorischen Mechanismen hinter diesen transkriptionellen Programmen wurde ein genregulatorisches Netzwerk für die DEG erstellt. Dazu wurden Target-Gene der differentiel exprimierten Transkriptionsfaktoren *in silico* bestimmt und das daraus resultierende

Genregulationsnetzwerk (Abbildung 5) in Subnetzwerke unterteilt (verwendete Software: Genomatix Genome Analyzer, Infomap (Rosvall M and Bergstrom CT 2010)).



**Abbildung 5. Genregulationsnetzwerke in ICGC PedBrain Medulloblastom Kohorte.** a) Genregulationsnetzwerke in den vier MB Subtypen. b) Genregulationsnetzwerke in den Group 3 Subclustern. c) Genregulationsnetzwerke in den Group 4 Subclustern. d) Genregulationsnetzwerke in den SHH Subclustern.

Diese Subnetzwerke reflektieren die aktivierten transkriptionellen Programme in der ICGC PedBrain Medulloblastoma Kohorte. Die identifizierten Subnetzwerke werden auf ihre Funktion und mittels *In-situ*-Hybridisierungsdatenbanken, die von der AG Yaspo im Rahmen des europäischen EUEXPRESS-Konsortiums erstellt wurden (Diez-Roux G *et al.* 2011, [www.eurexpress.org](http://www.eurexpress.org)) sowie dem Allen Brain Atlas (Allen Institute for Brain Science 2015) auf zellspezifische Expression hin untersucht. Diese zellspezifische Expression kann Aufschluss über die Ursprungszellen der Medulloblastom-Subtypen und -Subcluster geben (Abbildung 6).



**Abbildung 6. In-situ-Hybridisierung in Mausembryo.** Der Ausschnitt zeigt die ISH der Gene a) BARHL1 und b) LMX1A im sich entwickelnde Kleinhirn, Mittelhirn und Plexus choroideus in Mausembryo. Die ISHs stammen von der Datenbank Eurexpress.

### *De novo* Transkriptom-Assemblierung

Die Transkriptom-Analyse beinhaltet zudem das Assemblieren der RNA-seq-Reads, was eine Identifizierung unbekannter transkriptionell aktiver Genomregionen wie Gene, Transkripte oder Exons erlauben soll. Für das Assemblieren wurde eine Pipeline entwickelt, die eine Subtyp-spezifische Identifizierung dieser Genomregionen ermöglicht. Das Assemblieren in der Medulloblastom-Kohorte umfasste ca. 60 Milliarden RNA-Reads. Die ersten Schritte dieser Arbeit wurden bereits beendet, wodurch eine wertvolle Ressource an neuen

potentiellen transkriptionellen Einheiten entstanden ist. Eine weitreichende Auswertung der bisher unbekannt transkriptionell aktiven Genomregionen ist geplant.

Die oben beschriebenen Analysen bilden die Grundlage für eine Publikation über die molekulare phenotypische Charakterisierung von Medulloblastomen, die gerade ausgearbeitet wird.

#### *Andere Arbeiten*

Weiterhin haben wir basierend auf Ergebnissen der bisherigen Projektperioden an der Verfassung eines Manuskripts mitgearbeitet, welches Arbeiten zur Identifizierung von Medulloblastom-spezifischen Methylierungsmustern umfasst (Hovestadt et al. 2014).

#### *Referenzen*

Robinson MD *et al.* edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010** Jan 1;26(1):139-40. doi: 10.1093/bioinformatics/btp616

Rosvall M and Bergstrom CT. Multilevel Compression of Random Walks on Networks Reveals Hierarchical Organization in Large Integrated Systems. *Plos One* **2011** April doi:10.1371/journal.pone.0018209

Diez-Roux G *et al.* A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. *Plos Biology* **2011** Januar. doi:10.1371/journal.pbio.1000582

Allen Developing Mouse Brain Atlas [Internet]. Verfügbar unter: <http://developingmouse.brain-map.org>. Website: © 2015 Allen Institute for Brain Science.

## **2. Wichtigste Positionen des zahlenmäßigen Nachweises**

Im Rahmen des ICGC PedBrainTumor-Vorhabens wurden folgende Mittel vom Max-Planck-Institut für molekulare Genetik in Arbeitspaket 8 – Transkriptomanalyse – aufgewendet:

- Personalmittel: EUR 191.841,34
- Verbrauchsmittel für Sequenzierungen: EUR 313.292,79

Die Aufwendungen blieben damit im Rahmen des beantragten Finanzierungsplans. Für eine genauere Auflistung wird auf den detaillierten Verwendungsnachweis der Drittmittelabteilung des Max-Planck-Instituts für molekulare Genetik verwiesen.

## **3. Notwendigkeit und Angemessenheit der geleisteten Arbeit**

Die in Arbeitspaket 8 – Transkriptomanalyse – geleisteten Arbeiten waren notwendig und wichtig zum einen zur Bestätigung von Beobachtungen der anderen Arbeitspakete (Kreuzvalidierung von Ergebnissen der Genom- und Methylom-Sequenzierung) und zum

anderen aufgrund substantieller neuer Erkenntnisse, die die Transkriptomanalyse vor allem im Bereich der Subtypen-Klassifizierung und von Fusionsgenen und ihrer onkogenen Wirkung in verschiedenen pädiatrischen Hirntumorentitäten beigesteuert hat (siehe Abschnitt II.1.).

#### **4. Voraussichtlicher Nutzen, insbesondere Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans**

Das Ziel des ICGC PedBrainTumor-Projektes war es, Sequenzdaten von pädiatrischen Hirntumoren zu generieren und diese, wie auch die zugehörigen klinischen Daten, der internationalen Forschergemeinde zugänglich zu machen sowie bei renommierten internationalen wissenschaftlichen Zeitschriften zur Publikation einzureichen. Als Projektpartner haben wir dazu beigetragen, eine einzigartige Ressource an Daten über genetische Veränderungen bei kindlichen Gehirntumoren zu schaffen, die die Grundlage für weitere Forschungen national und international sein und so ein besseres Verständnis der Pathogenese pädiatrischer Hirntumore ermöglichen wird. Die Daten sind für Wissenschaftler über das ICGC-Datenportal zugänglich (<http://dcc.icgc.org/>) und zum anderen für Mediziner und interessierte Laien auch über das cBio-Portal (<http://www.cbioportal.org/>), welches eine Visualisierung und öffentlichen Zugang zu den Ergebnissen ermöglicht, wodurch eine Translation der Ergebnisse in die Klinik gefördert werden soll. Darüber hinaus werden sich weitere Beobachtungen und Hypothesen aus diesem Projekt ergeben, die durch andere Forschungsvorhaben bearbeitet werden können. Damit sind die wissenschaftlichen Hauptziele des Projekts erreicht worden.

Alle ICGC-Mitglieder haben sich zu Beginn des Projekts verpflichtet, keine Intellectual Property (IP)-Ansprüche aus primären Daten (einschließlich somatischer Mutationen) zu verfolgen bzw. IP-Schutzanmeldungen vorzunehmen, die den Zugang zu oder die Benutzung von Elementen der ICGC-Daten oder Schlussfolgerungen direkt aus diesen Daten verhindern oder blockieren würde. Dementsprechend wurden keine Patente angemeldet. Dennoch haben die geleisteten Forschungsarbeiten ein großes Potential zur Generierung von IP, sobald sie funktionell und prospektiv/klinisch weiter untersucht und bewertet worden sind. Beispiele sind die Entwicklung prognostischer/prädiktiver Marker, von Assays zur Durchführung der Diagnostik in einem klinischen Setting, neuer therapeutischer Angriffspunkte sowie präklinischer Testung neuer Therapieoptionen in vitro und in vivo. Sollten sich prädiktive Marker prospektiv validieren lassen, werden die Beteiligten auf klinische und industrielle Partner zugehen, die klinische Tests für zertifizierte Routine-Anwendungen herstellen können.

#### **5. Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Dem Zuwendungsempfänger wurde während des Vorhabens kein vergleichbarer Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen bekannt.

#### **6. Erfolgte oder geplante Veröffentlichungen des Ergebnisses**

Die Ergebnisse aus Arbeitspaket 8 wurden bisher in fünf Publikationen in hochrangigen Fachzeitschriften publiziert (*Nature*, *Nature Genetics*, *Cancer Cell*). Weitere Publikationen sind derzeit in Arbeit.

Bisher erfolgte Publikationen von Daten aus ICGC PedBrainTumor Arbeitspaket 8 – Transkriptomanalyse:

Pajtler, K. W., H. Witt, M. Sill, D. T. Jones, V. Hovestadt, F. Kratochwil, K. Wani, R. Tatevossian, C. Punehiwa, P. Johann, J. Reimand, H. J. Warnatz, M. Ryzhova, S. Mack, V. Ramaswamy, D. Capper, L. Schweizer, L. Sieber, A. Wittmann, Z. Huang, P. van Sluis, R. Volckmann, J. Koster, R. Versteeg, D. Fults, H. Toledano, S. Avigad, L. M. Hoffman, A. M. Donson, N. Foreman, E. Hower, K. Zitterbart, M. Gilbert, T. S. Armstrong, N. Gupta, J. C. Allen, M. A. Karajannis, D. Zagzag, M. Hasselblatt, A. E. Kulozik, O. Witt, V. P. Collins, K. von Hoff, S. Rutkowski, T. Pietsch, G. Bader, M. L. Yaspo, A. von Deimling, P. Lichter, M. D. Taylor, R. Gilbertson, D. W. Ellison, K. Aldape, A. Korshunov, M. Kool and S. M. Pfister (2015). "Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups." *Cancer Cell* **27**(5): 728-743.

Northcott, P. A., C. Lee, T. Zichner, A. M. Stutz, S. Erkek, D. Kawauchi, D. J. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. Jones, M. Kool, M. Remke, F. M. Cavalli, S. Zuyderduyn, G. D. Bader, S. Vandenberg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jager, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H. J. Warnatz, T. Risch, M. L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turanyi, P. Hauser, E. Sanden, A. Darabi, P. Siesjo, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya and S. M. Pfister (2014). "Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma." *Nature* **511**(7510): 428-434.

Hovestadt, V., D. T. Jones, S. Picelli, W. Wang, M. Kool, P. A. Northcott, M. Sultan, K. Stachurski, M. Ryzhova, H. J. Warnatz, M. Ralser, S. Brun, J. Bunt, N. Jager, K. Kleinheinz, S. Erkek, U. D. Weber, C. C. Bartholomae, C. von Kalle, C. Lawerenz, J. Eils, J. Koster, R. Versteeg, T. Milde, O. Witt, S. Schmidt, S. Wolf, T. Pietsch, S. Rutkowski, W. Scheurlen, M. D. Taylor, B. Brors, J. Felsberg, G. Reifenberger, A. Borkhardt, H. Lehrach, R. J. Wechsler-Reya, R. Eils, M. L. Yaspo, P. Landgraf, A. Korshunov, M. Zapatka, B. Radlwimmer, S. M. Pfister and P. Lichter (2014). "Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing." *Nature* **510**(7506): 537-541.

Jones, D. T., B. Hutter, N. Jager, A. Korshunov, M. Kool, H. J. Warnatz, T. Zichner, S. R. Lambert, M. Ryzhova, D. A. Quang, A. M. Fontebasso, A. M. Stutz, S. Hutter, M. Zuckermann, D. Sturm, J. Gronych, B. Lasitschka, S. Schmidt, H. Seker-Cin, H. Witt, M. Sultan, M. Ralser, P. A. Northcott, V. Hovestadt, S. Bender, E. Pfaff, S. Stark, D. Faury, J. Schwartzentruber, J. Majewski, U. D. Weber, M. Zapatka, B. Raeder, M. Schlesner, C. L. Worth, C. C. Bartholomae, C. von Kalle, C. D. Imbusch, S. Radomski, C. Lawerenz, P. van Sluis, J. Koster, R. Volckmann, R.

Versteeg, H. Lehrach, C. Monoranu, B. Winkler, A. Unterberg, C. Herold-Mende, T. Milde, A. E. Kulozik, M. Ebinger, M. U. Schuhmann, Y. J. Cho, S. L. Pomeroy, A. von Deimling, O. Witt, M. D. Taylor, S. Wolf, M. A. Karajannis, C. G. Eberhart, W. Scheurlen, M. Hasselblatt, K. L. Ligon, M. W. Kieran, J. O. Korbel, M. L. Yaspo, B. Brors, J. Felsberg, G. Reifenberger, V. P. Collins, N. Jabado, R. Eils, P. Lichter, S. M. Pfister and P. International Cancer Genome Consortium PedBrain Tumor (2013). "Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma." Nat Genet **45**(8): 927-932.

Jones, D. T., N. Jager, M. Kool, T. Zichner, B. Hutter, M. Sultan, Y. J. Cho, T. J. Pugh, V. Hovestadt, A. M. Stutz, T. Rausch, H. J. Warnatz, M. Ryzhova, S. Bender, D. Sturm, S. Pleier, H. Cin, E. Pfaff, L. Sieber, A. Wittmann, M. Remke, H. Witt, S. Hutter, T. Tzaridis, J. Weischenfeldt, B. Raeder, M. Avci, V. Amstislavskiy, M. Zapatka, U. D. Weber, Q. Wang, B. Lasitschka, C. C. Bartholomae, M. Schmidt, C. von Kalle, V. Ast, C. Lawerenz, J. Eils, R. Kabbe, V. Benes, P. van Sluis, J. Koster, R. Volckmann, D. Shih, M. J. Betts, R. B. Russell, S. Coco, G. P. Tonini, U. Schuller, V. Hans, N. Graf, Y. J. Kim, C. Monoranu, W. Roggendorf, A. Unterberg, C. Herold-Mende, T. Milde, A. E. Kulozik, A. von Deimling, O. Witt, E. Maass, J. Rossler, M. Ebinger, M. U. Schuhmann, M. C. Fruhwald, M. Hasselblatt, N. Jabado, S. Rutkowski, A. O. von Bueren, D. Williamson, S. C. Clifford, M. G. McCabe, V. P. Collins, S. Wolf, S. Wiemann, H. Lehrach, B. Brors, W. Scheurlen, J. Felsberg, G. Reifenberger, P. A. Northcott, M. D. Taylor, M. Meyerson, S. L. Pomeroy, M. L. Yaspo, J. O. Korbel, A. Korshunov, R. Eils, S. M. Pfister and P. Lichter (2012). "Dissecting the genomic complexity underlying medulloblastoma." Nature **488**(7409): 100-105.

## Berichtsblatt

1. ISBN oder ISSN geplant	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht	
3. Titel  Internationales Krebsgenomkonsortium (ICGC) PedBrainTumor: Arbeitspaket AP8 – Transkriptomanalyse		
4. Autor(en) [Name(n), Vorname(n)] Yaspo, Marie-Laure Lehrach, Hans Warnatz, Hans-Jörg Risch, Thomas	5. Abschlussdatum des Vorhabens 31.12.2014	
	6. Veröffentlichungsdatum geplant	
	7. Form der Publikation geplant	
8. Durchführende Institution(en) (Name, Adresse)  Max-Planck-Institut für Molekulare Genetik Inhnestrasse 63-73 14195 Berlin	9. Ber. Nr. Durchführende Institution –	
	10. Förderkennzeichen 01KU1201E	
	11. Seitenzahl 13	
12. Fördernde Institution (Name, Adresse)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 9	
	14. Tabellen 0	
	15. Abbildungen 6	
16. Zusätzliche Angaben		
17. Vorgelegt bei (Titel, Ort, Datum)		
18. Kurzfassung <p>In den vergangenen Jahren haben Studien an Krebsgenomen eine Vielzahl von Mutationen in vielfältigen Kombinationen aufgedeckt. Daraus resultieren zahlreiche neue Erkenntnisse zu Tumor-Pathogenesemechanismen mit potentieller Anwendung in der Diagnose und der Behandlung der betreffenden Tumorerkrankungen. Dank neuer Entwicklungen im Bereich von Next Generation-Sequenziermethoden ist es nun möglich, ganze Genome und Transkriptom in kurzer Zeit zu sequenzieren und zu analysieren. Auf der Basis von Genom- und Transkriptomsequenzierung war es das Ziel des ICGC PedBrainTumor-Projekts, eine umfassende Analyse einer großen Zahl an pädiatrischen Hirntumoren durchzuführen, die die pädiatrischen Tumore mit der höchsten Mortalitätsrate darstellen. Die genomische Sequenzanalyse von pädiatrischen Hirntumoren sollte dabei durch detaillierte Transkriptom- und Methylomanalyse ergänzt werden. Aufgabe des Arbeitspakets 8 – Transkriptomanalyse – war dabei die Erstellung und Auswertung von hoch aufgelösten Transkriptomdatensätzen.</p> <p>Im Vorhabenszeitraum konnte für insgesamt 372 Patientenproben die RNA sequenziert werden. Anschließend erfolgte die Bestimmung der Genexpressionswerte (RPKM-Werte), die Identifizierung von Fusionsgenen, alternative Exon-Nutzung, differentiell exprimierter Gene inklusive Identifizierung von molekularen Subclustern, die Identifizierung von Genregulationsnetzwerken und Transkriptom-Assemblierung zur Identifizierung bisher unbekannter neuer Transkripte. Alle fertigen Datensätze wurden den Projektpartnern zur Integration mit Daten aus anderen Sequenzierplattformen und mit klinischen Daten zur Verfügung gestellt. Die molekularen Daten wurden mit klinischen Parametern korreliert, um neue prognostische und prädiktive Marker sowie mögliche Ansatzpunkte für Therapien zu identifizieren. Im Rahmen von PedBrainTumor wurden unter anderem neue krebsauslösende Fusionsgene in Pilozytischen Astrozytomen, Medulloblastomen und Glioblastomen gefunden, die neuartige Ansatzpunkte für personalisierte Therapiemöglichkeiten bei den betroffenen Patienten darstellen.</p>		
19. Schlagwörter Krebs, Hirntumor, Transkriptomsequenzierung, Genexpression, Fusionsgene, personalisierte Therapie		
20. Verlag	21. Preis	

## Document Control Sheet

1. ISBN or ISSN planned	2. type of document (e.g. report, publication) Final report
3. title  International Cancer Genome Consortium (ICGC) PedBrainTumor: Work package WP8 – Transcriptome analysis	
4. author(s) (family name, first name(s)) Yaspo, Marie-Laure Lehrach, Hans Warnatz, Hans-Jörg Risch, Thomas	5. end of project 31.12.2014
	6. publication date planned
	7. form of publication planned
8. performing organization(s) (name, address)  Max-Planck-Institut für Molekulare Genetik Ihnestrasse 63-73 14195 Berlin	9. originator's report no. –
	10. reference no. 01KU1201E
	11. no. of pages 13
12. sponsoring agency (name, address)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. no. of references 9
	14. no. of tables 0
	15. no. of figures 6
16. supplementary notes	
17. presented at (title, place, date)	
18. abstract  During the last years, the study of cancer genomes has revealed numerous mutations in diverse combinations, resulting in new insights into tumor pathogenesis mechanisms with potential applications in diagnosis and treatment of the respective tumors. Due to new developments in the field of next generation sequencing methods, it is now possible to sequence and analyse whole genomes and transcriptomes in a short time. Based on whole genome and transcriptome sequencing, it was the aim of the ICGC PedBrainTumor project to conduct a comprehensive analyses of a large number of pediatric brain tumors, which constitute the pediatric tumors with the highest mortality rate. The genomic sequence analysis of pediatric brain tumors should be complemented by detailed transcriptome and methylome analyses. The task of work package 8 – Transcriptome analysis – was the generation and analysis of high-resolution transcriptome data sets for pediatric brain tumors.  During the project period, we performed RNA sequencing and analysis from 372 patient samples. We determined gene expression profiles (RPKM values), identified fusion genes, determined alternative exon usage and differentially expressed genes including the identification of molecular subclusters, analysed gene regulatory networks and performed de novo transcriptome assembly for the identification of so far unknown transcripts. All finished data sets were transferred to the projects partners for integration with other sequencing data types and with clinical data. The molecular data was correlated with clinical parameters to identify novel prognostic and predictive markers as well as new targets for therapies. In the framework of the PedBrainTumor project, we identified – among others – novel oncogenic fusion genes in pilocytic astrocytomas, medulloblastomas and glioblastomas constituting starting points for personalised therapy options for tumors with similar molecular profiles.	
19. keywords cancer, brain tumor, transcriptome sequencing, gene expression, fusion genes, personalised therapy	
20. publisher	21. price