Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Franklin L Chien, MD Date

Medulloblastoma Circulating Tumor Cell Clusters as a Novel Tumor Biomarker and Mechanism for Hematogenous Spread

By

Franklin L Chien, MD
Master of Science

Computer Science

Manoj Bhasin, Ph.D.
Advisor

Matthew Reyna, Ph.D.
Committee Member

Swati Bhasin, Ph.D.
Committee Member

Zhaohui Qin, Ph.D.
Committee Member

Accepted:

Kimberly J. Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Medulloblastoma Circulating Tumor Cell Clusters as a Novel Tumor Biomarker and Mechanism for Hematogenous Spread

By

Franklin L Chien, MD
B.A., Swarthmore College, PA, 2012
M.D., Jefferson Medical College, PA, 2017

Advisor: Manoj Bhasin, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Computer Science
2023

Abstract

Medulloblastoma Circulating Tumor Cell Clusters as a Novel Tumor Biomarker and Mechanism for Hematogenous Spread
By Franklin L Chien, MD

Medulloblastoma is an embryonal tumor of the cerebellum and accounts for most pediatric malignant disease of the central nervous system (CNS). RNA sequencing and DNA methylation studies reveal four major subgroups of disease: SHH-activated, WNT-activated, Group 3 and Group 4 [22, 23]. While grouping carries prognostic significance, metastatic spread remains the single most important indicator of outcomes [25].

Previously, medulloblastoma spread was thought to occur only by direct tumor shedding. However, there is a growing recognition for a hematogenous route for metastasis [13].

Literature in several adult cancers recognize the role of circulating tumor cells (CTCs) as a mechanism of seeding disease at remote sites [2], with CTC clusters (CTCCs) carrying even greater potential for metastasis [6]. Traditional methods of CTC detection rely on immunolabeling. This requires a universal and tumor-specific surface marker, whereas no similar markers are known in medulloblastoma cells. A novel microfluidic chip (Cluster-Chip) was developed to capture CTCCs from unprocessed blood using a label-free approach, achieving 99% efficiency on cluster sizes of 4 or more cells [24]. Using the Cluster-Chip technology, we describe the presence of CTCCs in patients with medulloblastoma. CTCCs therefore may represent a novel biomarker for tumor and present an exciting new direction to study hematogenous disease spread in medulloblastoma.

We enrolled 44 pediatric patients in a longitudinal study. CTCCs are quantified at enrollment and additionally at routine 3-month intervals in medulloblastoma patients. From this study cohort, we report the presence of CTCCs in all patients with medulloblastoma and at all phases of therapy, while none are detected in patients without malignancy.

Identifying CTCCs in blood from patients with medulloblastoma is promising, but their role in metastasis pathophysiology is not currently understood. We use multiple change point detection to identify significant changes in CTCC quantity from sequential peripheral whole blood data, and next-generation high-throughput RNA sequencing to describe CTCC transcriptome compared to whole blood through gene set enrichment analysis and immune cell deconvolution.

Medulloblastoma Circulating Tumor Cell Clusters as a Novel Tumor Biomarker and Mechanism for Hematogenous Spread

By

Franklin L Chien, MD
B.A., Swarthmore College, PA, 2012
M.D., Jefferson Medical College, PA, 2017

Advisor: Manoj Bhasin, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Computer Science
2023

Acknowledgments

The author recognizes the generous support by the V Foundation for funding the 3-year pilot study titled "Circulating Tumor Cell Clusters as a Novel Biomarker for Medulloblastoma". Data presented in this thesis is a direct result from funding of the pilot study.

Additionally, the author wishes to acknowledge the contributions of the Emory Integrated Computational Core for their part in alignment of FASTQ read data and analysis, and Sarioglu lab for their development of the Cluster Chip from which circulating tumor cell clusters were captured.

Lastly, the author would like to thank Drs. Tobey MacDonald, Manoj Bhasin, and Swati Bhasin for their joint research mentorship as well as other members of the scholarly oversight committee who contributed to the project including Drs. Matthew Reyna, Jeffrey Switchensko, and Zechen Chong.

Contents

1 Introduction and Background **1**

1.1 Medulloblastoma 1

1.1.1 Medulloblastoma Genomics 2

1.1.2 Metastasis and Survival Outcomes 2

1.2 Circulating Tumor Cells and Cell Clusters 3

1.2.1 Circulating Tumor Cell Capture 4

2 Methods **5**

2.1 Patient Enrollment and Human Subjects Research 5

2.2 Change Point Detection 6

2.3 CTCC Enrichment and Bulk-RNA Seq 7

2.3.1 Enrichment 7

2.3.2 Sequencing 7

2.3.3 Differential Gene Expression Analysis 9

2.3.4 Batch Effect 9

2.3.5 Stemness 10

2.3.6 Gene Set Enrichment Analysis 10

2.3.7 Immune Deconvolution 11

2.4 Contributions from Collaborators 12

2.5 Funding 12

3 Results **13**

3.1 Patient Enrollment and Cell Capture 13

3.2 CTCC Quantification and Change Point Detection 15

3.3 Bulk RNA-Seq Quality Control 17

3.4 Bulk RNA Sequencing 20

3.5 Gene Set Enrichment Analysis 23

3.5.1 Immune Deconvolution 25

4 Discussion **27**

4.1 Circulating Tumor Cell Cluster Quantification 27

4.2 Transcriptomic Analysis of Captured CTCCs 28

4.3 CTCC Gene Set Enrichment Analysis 29

4.4 Immune deconvolution 30

4.5 Batch Effect 31

4.6 Future Studies 31

5 Conclusions **33**

Bibliography **35**

List of Figures

3.1 Number of peripheral blood collections per patient enrolled in pilot cohort. Each bar represents a patient with medulloblastoma. 14

3.2 Immunofluorescent staining of CTCC in blood and CSF. Clusters of large cells stain positively with synaptophysin (green) and negatively with CD45 (red) indicate tumor cells of neuronal origin. CD45+ cells from hematopoietic lineage associating with CTCC from blood are hypothesized to constitute a CTCC immune microenvironment. CD45+ cells are not seen in CSF, an immune privileged space 15

3.3 Overall distribution of CTCC concentration (CTCC/ml) enriched from peripheral blood and CSF samples from patients diagnosed with Medulloblastoma. Several extreme outlying data points are noted. Distributions are plotted both with and without these outliers 16

3.4 CTCC concentrations (CTCC/ml) trends over time in seven patients with medulloblastoma having the most number of blood draws, both with and without extreme outlying data. The addition of outlying data demonstrates a spike in CTCC concentration at the end of therapy seen in some patients. Excluding the spike, an overall shape of decreasing CTCC is noticed as patients are treated with cancer directed therapy. 17

3.5 CTCC concentrations (CTCC/ml) trends over time with detected change points shown in three patients, two with 10 or more blood collections (MB003, MB008), and one patient with the post therapy spike in CTCC. In patient MB0003 (3.5a), first change point occurs in the midst of chemotherapy and second coincides with therapy completion. In patient MB008 (3.5b), first change point is associated with initiation of therapy for relapse disease, the second at the end of radiation therapy and initiating chemotherapy, and the 3rd during chemotherapy. Patient MB022 (3.5c) show the off therapy spike in CTCC/ml, with change points detected at the spike. 18

3.6 FastQC results from ultra-low input bulk RNA-sequencing of CTCC enriched samples. Alignment, GC content, Phred quality scores, and read lengths are given. 19

3.7 Adaptor content of sequenced CTCC samples pre and post algorithmic adaptor removal 19

3.8 Comparison of gene expression distribution between CTCC sample and peripheral whole blood controls, showing right skew and relative scarcity in CTCC data. Log transformed DESeq2 normalized gene expression counts were to plot density distribution. 20

3.9 Principal component analysis using top 1000 genes by variance showing Medulloblastoma CTCC samples (red) and peripheral whole blood control (blue). Medulloblastoma CTCC and whole blood sequencing show distinct clustering by group, supporting that the Cluster Chip was able to capture and distinguish CTCCs from background whole blood. 21

3.10 Heat map of top 40 genes with highest variance in differential gene expression on bulk RNA sequencing comparing medulloblastoma CTCC (red columns) to peripheral whole blood control (blue columns). . . . 22

3.11 Volcano plot of 46 genes known to be markers of cancer stem cells within the adult CTCC literature, showing several genes in the upper right quadrant compared to a single gene differentially down regulated reaching significance. The result supports stemness as a potential feature for medulloblastoma CTCCs. 22

3.12 Gene set enrichment analysis by C2 canonical pathway analysis show several gene sets involved in collagen, extracellular matrix, basement membrane, and cell junction proteins. 23

3.13 Gene set enrichment analysis by Gene Ontology Biological Processes showing enrichment of several gene sets involved in neuronal processes. These processes including membrane depolarization, action potential, and neurotransmitter uptake. 24

3.14 Gene set enrichment analysis by oncologic signatures collection demonstrating enrichment with statistical significance several gene sets known to be involved in cell cycle regulation and proliferation, including NFκB, PTEN, MTOR, and KRAS. Additionally, a gene set upregulated in embryonic stem cells (ESC J1 UP LATE V1 UP) is notably enriched, supporting CTCC stemness. 24

3.15 Relative fraction of myeloid and lymphoid lineage immune cells deconvoluted through CIBERSORTx using LM22 signature matrix. Paired t-test show significant difference in relative fraction of CD4+ memory resting t-cells (figure 3.15b) with $AdjustedP = 0.0001$. 25

3.16 Violin plot of medulloblastoma CTCC (red) compared to peripheral blood control (blue) showing increased relative fractions of monocytes in whole blood samples compared to increased macrophages in CTCC samples. This was true of M0, M1, and M2 macrophages 26

List of Tables

3.1 Number of patients enrolled in CTCC study over 3-year study period between 2020-2023 by sex, diagnosis, and medulloblastoma subgroup. Average and age ranges at time of enrollment are given for groups with 8 or more patients. 14

Chapter 1

Introduction and Background

1.1 Medulloblastoma

Medulloblastoma is the most common malignant brain tumor of the CNS in childhood [29]. It is an embryonal small round blue cell tumor of the cerebellum originating from neuronal progenitor cells. The malignancy occurs in the posterior fossa, an intracranial anatomical space housing the cerebellum. Medulloblastoma is more common in males compared to females at a ratio of 1.6 to 1 [9], and is most common in non-Hispanic white patients.

General approaches to medulloblastoma treatment include maximally safe surgical resection, adjuvant chemotherapy, and radiation therapy [19]. Due to significant adverse cognitive effects of radiation, high intensity chemotherapy followed by autologous stem cell rescue is becoming an increasing popular method of treatment in the youngest and thus most radiosensitive patient population. With these interventions, five- and ten-year overall survival in the United States are 70.1% and 63.3%, respectively [9]. Profound and permanent neurologic deficits may result from both the disease and treatment sequelae. In this way, medulloblastoma represents one of the most significant causes of childhood cancer related morbidity and mortality.

1.1.1 Medulloblastoma Genomics

In the genomics era, molecular profiling of tumors have improved tumor categorization and prognostication. One notable example is the subdivision of medulloblastoma into four distinct groups based on transcriptomic landscape [22]. In 2006, discrete subgroups of medulloblastoma were first described based on Affymetrix oligonucleotide arrays [28]. In the following decade, these were later expanded upon and refined leading to a consensus statement defining 4 major molecular subgroups of Medulloblastoma: WNT-activated, SHH-activated, Group 3, and Group 4 [27]. Group 3 and Group 4 tumors are now recognized as two extremes of a continuum of tumors [31]. As molecular subgrouping continues to advance, 12 subtypes of the 4 major groups have been proposed [11].

Medulloblastoma molecular sub-grouping carries prognostic significance with most favorable survival in the WNT-activated group in contrast Group 3 tumors which predicts poorest survival outcomes [9].

1.1.2 Metastasis and Survival Outcomes

Despite significant recent advances to the molecular diagnosis and categorization of medulloblastoma, metastatic staging remains the most significant predictor of survival [25]. With the exception of extreme cases of severe tumor burden, the vast majority of medulloblastoma metastasis occurs in the leptomeningeal space. Traditionally, these have been referred to as "drop down" metastasis, suggestive of direct tumor shedding as a mechanism for spread.

Recent literature challenges the previously held assumption for direct tumor shedding as the sole means of metastasis. Murine xenograft studies have demonstrated the ability of medulloblastoma cells to travel through a hematogenous route and seed distant sites of disease in a sibling mouse with surgically anastomosed circulatory system [13].

1.2 Circulating Tumor Cells and Cell Clusters

Circulating tumor cells (CTCs) are cancer cells derived from solid tumors that can be detected away from the primary tumor site circulating in the bloodstream. CTCs were first discovered over a century ago, however, they were not able to be further characterized or studied in detail until more recently [5, 12]. Among the challenges to studying CTCs include their extreme rarity, accounting for 1 cancer cell among one billion normal circulating blood cells [3]. Most CTCs are fated to die in circulation due to sheer forces, oxidative stress, and peripheral immune surveillance. However, a small fraction survive and seeds metastasis at distant sites [14]. CTCs are thus a mechanism for hematogenous spread to distant site in several solid tumors. Their presence has not been previously described in medulloblastoma patients.

In addition to circulating as single clusters, recent data within the past decade have demonstrated that CTCs also cluster [3]. These clusters of two or more CTCs with strong cell-to-cell contact defines a circulating tumor cell cluster (CTCC) [12]. Clusters can consist of a single cell type (homotypic) or be composed of multiple distinct cell types (heterotypic) [1]. The mechanisms by which CTCCs originate is still under investigation, with two leading hypothesis including collective migration off the primary bulk tumor versus intravascular cell aggregation ("cell jamming") [5, 14].

Emerging evidence suggest that clustering confers certain survival advantages for the tumor cells including mesenchymal transformation, resistance to programmed cell death, and immune surveillance escape [5, 14, 12], though the specific mechanisms by which these advantages are achieved are not fully understood. While circulating, they have been observed to be quiescent with respect to cell cycle [14], allowing resistance to conventional chemotherapy. This ability to circulate and survive away from the primary tumor is necessary for metastasis.

Previous molecular studies of CTCC in non-medulloblastoma tumors have shown

that CTCC carries the tumor signature of the original bulk tumor. In addition, CTCCs also differentially express cell-cell adhesion proteins and markers of stemness, distinguishing them from singular CTCs [6]. Several markers of stemness have been described in a variety of cancers [4].

1.2.1 Circulating Tumor Cell Capture

Despite interest in CTCs and CTCCs as a novel disease biomarker, isolation of rare circulating cells from blood poses a significant challenge to their study. Several methods have been developed using a immune label based approach for cell detection and capture such as CellSearch [7]. Immune labeling approaches are difficult to apply to medulloblastoma, as their application requires the presence of a cell surface marker that is both sensitive and specific to medulloblastoma when no such candidate marker is known.

Recently, a novel microchip (Cluster Chip) was developed with the capacity to capture peripherally circulating CTCCs from unprocessed whole blood [24]. The Cluster Chip uses bifurcating triangular columns through which whole blood is passed in low-flow conditions, trapping tumor clusters. Using the label-free approach circumvents the requirement for an identifying surface marker. The low flow conditions are below the flow speed of human capillaries, minimizing mechanical injury to the cell clusters and artifactual dissociation of clusters. Capture efficiency was at 99% in a population of spiked CTCCs for clusters of 4 cells or greater. Lastly, the Cluster Chip captures and releases live cells, allowing for subsequent sequencing experiments of the cells.

Chapter 2

Methods

2.1 Patient Enrollment and Human Subjects Research

To study the role of CTCC in disease, 44 patients were enrolled into a pilot study including 24 with medulloblastoma, 10 with low grade glioma, and 10 with non-malignant hematological conditions. Patients were consented for study from Children's Healthcare of Atlanta's Aflac Blood and Cancer Disorder Center in Scottish Rite children's hospital. First samples of blood and cerebrospinal fluid (CSF) were obtained in medulloblastoma patients following surgical resection of the tumor and pathological diagnosis of disease, but prior to initiation of chemotherapy. Peripheral blood draws were repeated regularly at 3-month intervals throughout therapy during clinic visits when peripheral blood draws are otherwise medically indicated. Routine sequential spinal fluid evaluations were not obtained as repeat lumbar punctures are not medically indicated.

Controls patients with low grade glioma or non-malignant hematological conditions were similarly patients who sought care at Aflac Blood and Cancer Disorder Center. Patients with low grade gliomas were chosen as a control due to presence of

CNS mass. However, unlike Medulloblastoma, these lesions are not considered malignant. Non-malignant hematological controls include patients seeking care for iron deficiency anemia, hereditary spherocytosis, and sickle cell anemia but do not have any oncological disease. No patients received lab draws or procedures for research purposes except at times when they are otherwise medically indicated.

Human subjects research activities and protocol was reviewed and approved by Emory University Institutional Review Board (IRB) under the IRB approval ID MODCR003-IRB00116846 following submission for the expedited process. Original IRB documents were submitted Nov 2021. Re-application was performed the following year with re-approval in October 2022.

2.2 Change Point Detection

Multiple change point detection was used to detect changes in mean CTCC concentration from sequential lab draws using the R package "changepoint" developed by Killick and Eckley [16]. Multiple detection algorithms are implemented. The pruned exact linear time (PELT), an optimization of the segment neighborhood algorithm was selected as the method for multiple change point detection. Given no theoretical reason to limit to single change point, At Most One Changepoint (AMOC) was not used. A Modified Bayes Information Criterion (MBIC) penalty was applied. Multiple change point detection was performed on peripheral CTCC concentrations in patients with ten or more collections.

2.3 CTCC Enrichment and Bulk-RNA Seq

2.3.1 Enrichment

CTCCs were isolated and quantified from peripheral blood and CSF using a novel physics-based microfluidic device, the "Cluster Chip". Captured CTCCs undergo immunostaining with the cell surface markers CD45 and Synaptophysin. These two markers were selected for immunostaining after discussion with hematopathology and neuropathology to distinguish medulloblastoma cells from other cells in whole blood. CD45 was selected to stain cells of hematopoietic lineage, and synaptophysin to stain cells of neuronal lineage. CTC clusters enriched from whole blood are subsequently identified by pathology to be of tumor origin based on microscopic appearance and CD45-negative, synaptophysin-positive immunofluorescence staining pattern. Cells obtained in peripheral blood do not express synaptophysin under physiologic conditions.

2.3.2 Sequencing

To validate CTCC enriched samples as distinct from background whole blood and to describe the transcriptomic landscape of CTCC, next-generation bulk RNA sequencing was performed from CTCC enriched samples. Six patients in the study cohort were selected for sequencing, chosen to represent a range of medulloblastoma subgroups and relapse status. Of note, no patients were enrolled with WNT-activated disease at the time of sequencing. Sequenced samples were obtained from first lab draws prior to chemotherapy. Two to 4 replicates were performed for each biological sample for a total of 21 samples.

Following CTCC enrichment on the Cluster Chip, each sample typically contains fewer than 100 cells. To account for this, an ultra-low cell count protocol optimized for transcriptome analysis was used for sequencing. Following CTCC enrichment, RNA

was isolated using the Quick-RNA Microprep Kit by Zymo Research. Complementary DNA was generated using SMART-Seq v4 Ultra-Low Input RNA Kit by Takara bio. The final sequencing library is made by Nextera XT kit by Illumina. Samples were provided to the Emory Integrated Genomics Core (EIGC) who carried out ultra-low cell count sequencing. Once complementary DNA fragments are tagged and adaptors added, sequencing was performed at a read-depth of 30 million paired-end reads of 100 base pairs per sample on an Illumina NovaSeq6000 system. A read-depth of 30 million reads per sample was selected to detect large differences in gene expression and with consideration to sequencing cost. Significant differences in differential gene expression is anticipated between CTCC originating from cerebellar tumor compared to whole blood without malignancy.

Following sequencing, FASTQ files were then provided to the Emory Integrated Computation Core (EICC). Read fragments were first pre-processed by adaptor trimming using the Java based software Trimmomatic (v0.36). Read sequences are quality-checked using FastQC with read confidence reported in phred quality score. Trimmed reads are then aligned with STAR aligner v2.5.2 to the University of California Santa Cruz (UCSC) Hg38 human genome assembly, the most current major assembly of the human genome. Finally, gene quantification is performed using HTSeq-count.

The resulting count matrices obtained from ultra-low input bulk RNA sequencing of CTCC-enriched samples were then compared to whole blood control RNA expression. Whole blood control data derive from four healthy young adult donors aged 25-40, and were obtained from the Gene Expression Omnibus (GEO), a public genomics repository of the National Center for Biotechnology Information (NCBI). This comparison data series were selected to match major sequencing parameters used in CTCC-enriched sequencing in that it was also computed from STAR alignment to Hg38 assembly, with sequencing on the same NovaSeq6000 platform (GSE178388) [8]. Sequencing depth of whole blood controls was at 50 million paired end reads per

sample, compared to the 30 million in CTCC-enriched samples.

2.3.3 Differential Gene Expression Analysis

Differential expression analysis was performed with DESeq2, a component of the Bioconductor software packages. Sequencing counts typically have a skewed and over-dispersed nature, thus are incompatible with Gaussian models. Therefore, DESeq2 uses a negative binomial distribution for RNA-seq data to conduct statistical inferences [18]. Adjusted p-values were used rather than raw p-values to account for multiple hypothesis testing.

To account for differences in read depth and gene length, normalization was applied using median of ratios method as implemented in DESeq2. Relative sparsity in gene expression is noted in the CTCC enriched samples. To account for the expected increase in variation due to sparsity, only differential up-regulation was considered as upregulation can be more robust to variance compared to down regulation.

2.3.4 Batch Effect

Peripheral whole blood control sequencing was selected to match several key experimental designs with the CTCC enriched sequencing. However, because the sequencing experiments were performed by different labs, institutions, and at different read depths, batch-effect is expected to contribute to differences in transcriptomic expression between the two sample groups.

To further evaluate technical differences in sequencing reads, DESeq2 normalized counts were plotted as two overlapping density curves. Due to substantial right skew in the sequencing alignment counts, normalized counts were globally log-transformed for ease of visualization.

Down sampling is a method of batch correction whereby data is randomly excluded to normalize gene expression raw count matrices. The CTCC sequencing was

performed at 30 million reads per sample - 40% fewer reads than the control which was sequenced at a read depth of 50 million reads. ComBat-seq is another batch-correction algorithm using a negative-binomial regression model to minimize variance in genomics data due to technical differences [33]. In discussion with Emory Integrated Computation Core, post sequencing batch-effect correction was not pursued as subsequent experiments are planned to minimize batch-correction a-priori. Down sampling was felt to be suboptimal due to resulting loss of data.

2.3.5 Stemness

Stemness and epithelial to mesenchymal transformation is known to play an important role in the biological function of CTCCs. To evaluate the expression of stemness markers in medulloblastoma CTCCs, a gene set of 46 markers of stemness was compiled from a review article of cancer stem cells from the circulating tumor cell literature [4]. Differential gene expression log-fold change and adjusted p-values for each of these 46 genes are plotted in a volcano plot.

2.3.6 Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) is a method whereby biological insight can be statistically inferred by comparing the pattern of differential gene expression against a-priori curated sets of genes that define a specific cellular behaviors. Gene set collections were obtained from the Molecular Signatures Database (MSigDB) jointly hosted by the Broad Institute and UC San Diego. Human collections C2, C5, and C6 corresponding to Canonical Pathway, Gene Ontology, and the Oncogenic Signatures gene sets were used as the basis of GSEA.

GSEA was scripted in R version 4.2.1 using the Bioconductor package “fgsea”. Resulting normalized enrichment scores (NES) and adjusted p-values are plotted in a bubble chart. Bubble chart are constructed with NES on the x-axis and bubble size

determined by the $-log(AdjustedP)$, transformed on negative log such that larger dots represent more significance.

2.3.7 Immune Deconvolution

CTCC captured from the peripheral blood are observed to be circulating with CD45+ cells of hematopoietic lineage, which are hypothesized to represent tumor associated immune cells. In order to identify immune cell populations associated with CTCC, CIBERSORTx was used to perform RNA deconvolution on CTCC samples. CIBERSORTx uses a support vector machine based machine-learning approach to deconvolute cell types from a mixture matrix [26].

A signature matrix defines RNA profiles in pure samples of different cell types and is used as the basis of mixture deconvolution. LM22 is a validated leukocyte gene signature matrix originally developed from Affymetrix microarray data [20], which contains 547 genes distinguishing 22 human cell types from hematopoietic lineage. This matrix was used as the basis of immune deconvolution.

To construct the CTCC mixture matrix, DESeq2 normalized RNA expression was averaged across replicates. Following deconvolution, relative fractions were used to construct heat maps of both lymphoid and myeloid lineage by patient sample. To illustrate differences in T, B, and Monocyte/macrophage cells, violin plots are constructed.

Differences in relative fraction between CTCC enriched samples and whole blood controls were evaluated statistically. Paired t-tests were performed on relative fractions for each cell type defined in the LM22 signature matrix. Resulting p-values are adjusted for multiple hypothesis using the False Discovery Rate (FDR) method.

2.4 Contributions from Collaborators

CTCC capture and quantification was performed by Sarioglu lab. Sequencing, alignment, and differential gene expression was performed by EIGC and EICC. The author of this manuscript as part of MacDonald lab was involved in the experimental design and interpretation of differential gene expression results, as well as selection of the data series to serve as whole blood control. Gene expression distribution comparison between CTCC enriched sequencing against whole blood sequencing was also performed by the author. Additionally, computation, scripting, and interpretation of change point detection, gene set enrichment analysis, and immune deconvolution was performed by the author.

2.5 Funding

Funding support for personnel, laboratory materials and reagents, and next-generation RNA sequencing was provided by the V Foundation for Cancer Research.

Chapter 3

Results

3.1 Patient Enrollment and Cell Capture

In the three-year period between 2020 and 2023, 44 pediatric patients were enrolled in the Medulloblastoma CTCC study (table 3.1). Of the 44 total patients, 24 have medulloblastoma, 10 have low grade glioma, and 10 have benign hematological conditions. Enrolled medulloblastoma patients include 16 with new diagnosis and 8 with relapse disease, and include all 4 molecular subgroups. Patient ages ranged from less than 1 year-of-age to 22 years-of-age at enrollment, with an average age of 11.

For the duration of the pilot study, 204 total samples were collected including 163 from blood and 41 from CSF. Serial blood draws were obtained at 3 month intervals with number of peripheral blood collections ranging from 1 to 13 (figure 3.1).

Medulloblastoma CTCCs, defined as large cells in clusters staining positive for synaptophysin and negative for CD45, are detected in every patient with medulloblastoma and absent in both low grade glioma controls and non-malignant controls. Representative images of captured clusters from peripheral blood and from CSF are given in figure 3.2 with green fluorescence indicating synaptophysin and red indicating CD45. CD45 positive cells were noted to associated with clusters from peripheral

Variable	n	Mean Age (Range)
Sex		
Male	24	10.0 (1, 22)
Female	20	13 (0, 22)
Medulloblastoma	24	9.8 (0, 22)
Relapse Status		
New Diagnosis	16	7.5 (0, 22)
Relapse	8	14.3 (6, 22)
Sub-type		
SHH	9	9.7 (0, 22)
WNT	1	
Group 3	3	
Group 4	6	
Non WNT/SHH	5	
Low Grade Glioma	10	12.6 (4, 21)
Non-Malignant Heme	10	13.9 (5, 19)
Total	44	11.34 (0, 22)

Table 3.1: Number of patients enrolled in CTCC study over 3-year study period between 2020-2023 by sex, diagnosis, and medulloblastoma subgroup. Average and age ranges at time of enrollment are given for groups with 8 or more patients.

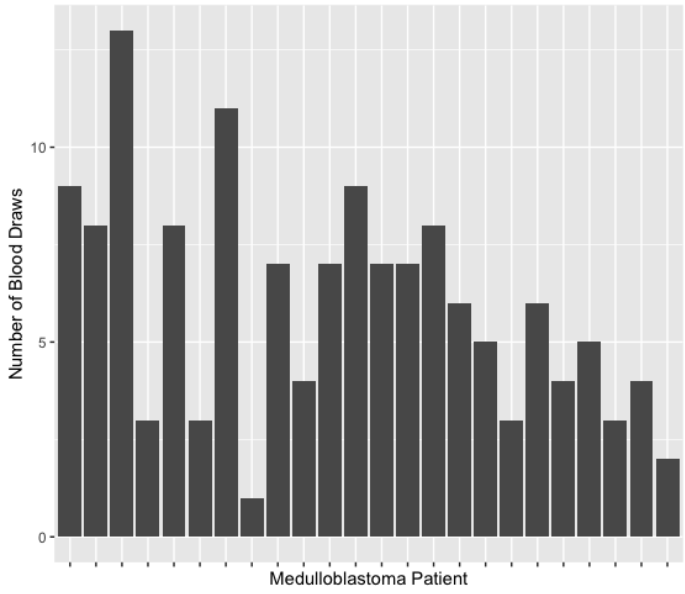
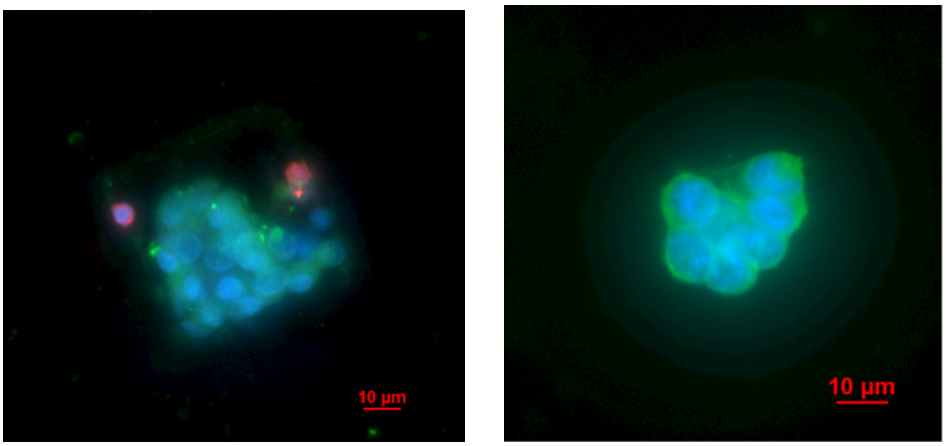


Figure 3.1: Number of peripheral blood collections per patient enrolled in pilot cohort. Each bar represents a patient with medulloblastoma.



(a) CTCC cluster in peripheral blood (b) CTCC cluster in CSF

Figure 3.2: Immunofluorescent staining of CTCC in blood and CSF. Clusters of large cells stain positively with synaptophysin (green) and negatively with CD45 (red) indicate tumor cells of neuronal origin. CD45+ cells from hematopoietic lineage associating with CTCC from blood are hypothesized to constitute a CTCC immune microenvironment. CD45+ cells are not seen in CSF, an immune privileged space

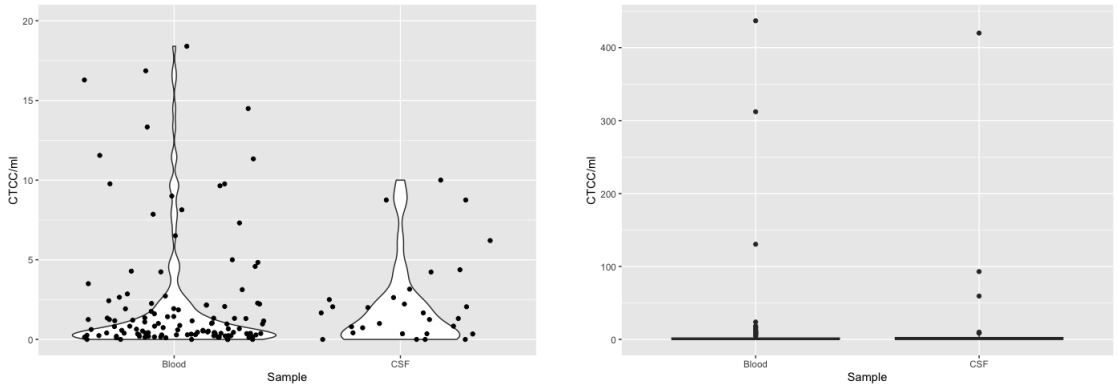
blood (figure 3.2a) but not from CSF (figure 3.2b), an immune privileged space. These associated CD45+ cells are predicted to comprise a peripheral CTCC immune microenvironment.

3.2 CTCC Quantification and Change Point Detection

Global concentrations of CTCC in CTCC/ml following enrichment are plotted in violin plot in figure 3.3. The spread of the data shows significant skew, with few observations correlating with extremely high concentrations of CTCC. In blood, these extreme outlying data correspond clinically with a spike in CTCC concentration following completion of chemotherapy. These spikes are observed to return to non-outlying values without additional therapy. While noted as a recurring pattern, the spikes in CTCC concentration were not universally observed within our patient co-

hort.

Within CSF observations, outlying high CTCC concentrations coincides with lab draws preceding disease progression and death.

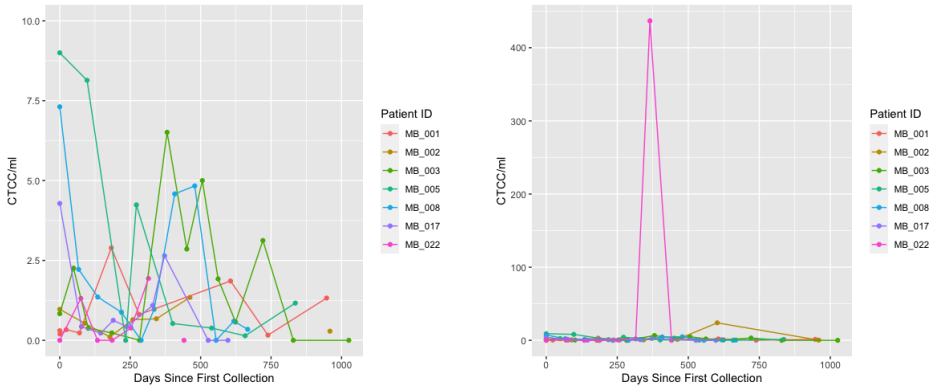


(a) Excluding extreme outlying data (b) With extreme outlying data

Figure 3.3: Overall distribution of CTCC concentration (CTCC/ml) enriched from peripheral blood and CSF samples from patients diagnosed with Medulloblastoma. Several extreme outlying data points are noted. Distributions are plotted both with and without these outliers

Peripheral blood CTCC quantity is hypothesized to indicate the body's overall tumor burden. To assess global trends in CTCC/ml with therapy and with off therapy follow-up, seven patients with the greatest number of serial whole blood collections were plotted on line plot (figure 3.4). The overall shape of the data is consistent with decreasing CTCC quantities with cancer directed care. Overall, CTCCs continues to be detected following therapy even in the absence of relapse or progression as assessed by traditional means. Figure 3.4b show a representative patient demonstrating spike in CTCC/ml following completion of therapy, with resolution in CTCC in the absence of further chemotherapy.

Multiple change point detection is predicted to correlate with significant changes to the biological system such as changes in chemotherapy, tumor progression, or relapse. Applied to patients with 10 or more collections, change points in mean are statistically detected. In one patient (figure 3.5b), the first change point corresponded



(a) Excluding outlying data (b) Including outlying data

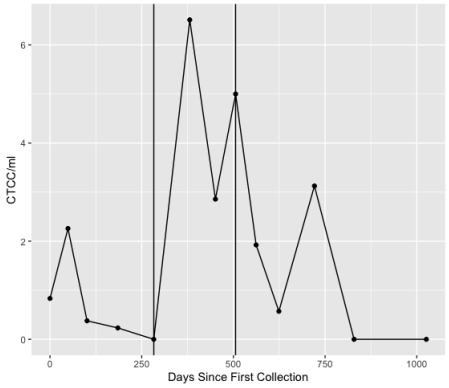
Figure 3.4: CTCC concentrations (CTCC/ml) trends over time in seven patients with medulloblastoma having the most number of blood draws, both with and without extreme outlying data. The addition of outlying data demonstrates a spike in CTCC concentration at the end of therapy seen in some patients. Excluding the spike, an overall shape of decreasing CTCC is noticed as patients are treated with cancer directed therapy.

with initiation of tumor directed therapy. Figure 3.5c shows one patient with a large spike in CTCC following therapy, triggering detection of two change points surrounding this spike.

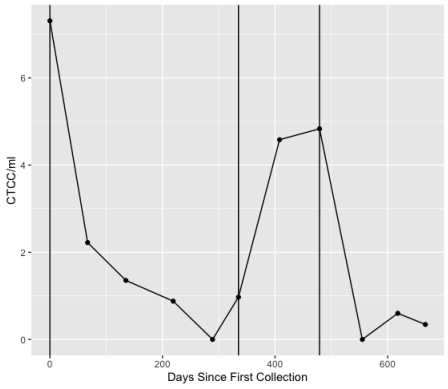
3.3 Bulk RNA-Seq Quality Control

Bulk RNA-Seq quality control was conducted by EICC and is given in the figure 3.6. Guanine and Cytosine are two of the four base pairs comprising human DNA and RNA with percentage content commonly used as a quality control metric. GC content of bulk RNA-sequencing are consistent with average GC-content of human genomes of 41% [17]. Phred score show high confidence in base pair calls (figure 3.6c), and read lengths were as expected in this short-read sequencing experiment (figure 3.6d). Relatively high adapter content was detected pre-trimming, in contrast to post-adapter trimming with Trimmomatic (figure 3.7).

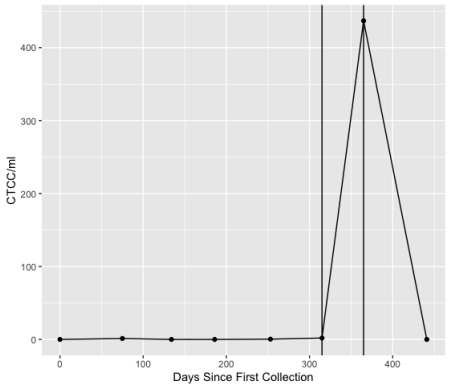
To assess differences in gene expression distribution between CTCC samples versus



(a) Patient MB003

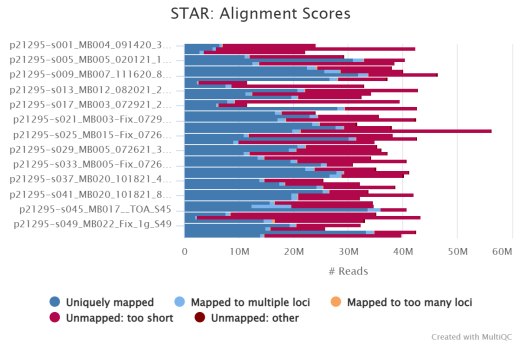


(b) Patient MB008. First change point correspond with initiation of tumor directed therapy

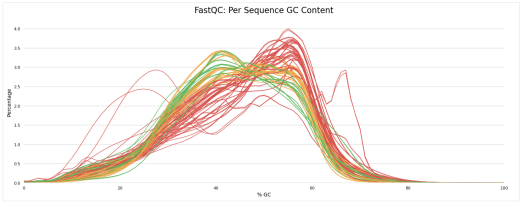


(c) Patient MB022. Change point coincide with the large spike in CTCC following completion of chemotherapy

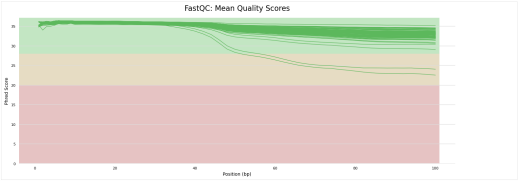
Figure 3.5: CTCC concentrations (CTCC/ml) trends over time with detected change points shown in three patients, two with 10 or more blood collections (MB003, MB008), and one patient with the post therapy spike in CTCC. In patient MB0003 (3.5a), first change point occurs in the midst of chemotherapy and second coincides with therapy completion. In patient MB008 (3.5b), first change point is associated with initiation of therapy for relapse disease, the second at the end of radiation therapy and initiating chemotherapy, and the 3rd during chemotherapy. Patient MB022 (3.5c) show the off therapy spike in CTCC/ml, with change points detected at the spike.



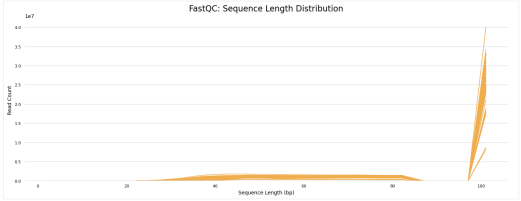
(a) Proportion of STAR Alignment uniquely mapped to reference genome



(b) RNA sequencing QC Content

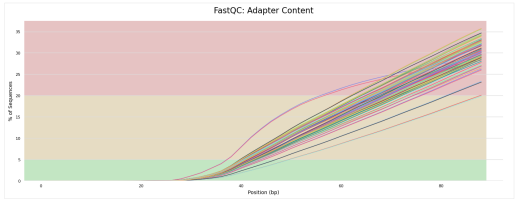


(c) Phred quality score by read position

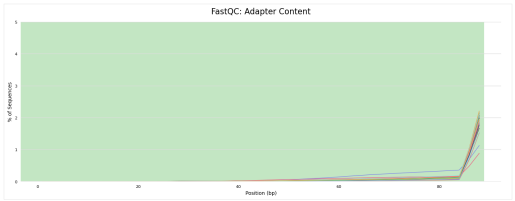


(d) Sequencing read-length

Figure 3.6: FastQC results from ultra-low input bulk RNA-sequencing of CTCC enriched samples. Alignment, GC content, Phred quality scores, and read lengths are given.



(a) Adaptor content pre trimming



(b) Adaptor content post trimming

Figure 3.7: Adaptor content of sequenced CTCC samples pre and post algorithmic adaptor removal

controls, density distribution of gene expression profiles using DESeq2 normalized gene counts were plotted (figure 3.8). Relative data scarcity is noted in the CTCC sample with greater number of genes with 0 reads compared to control, whereas more gene expression is detected in the control samples at log normalized counts within in the range less than 5 but greater than 0.

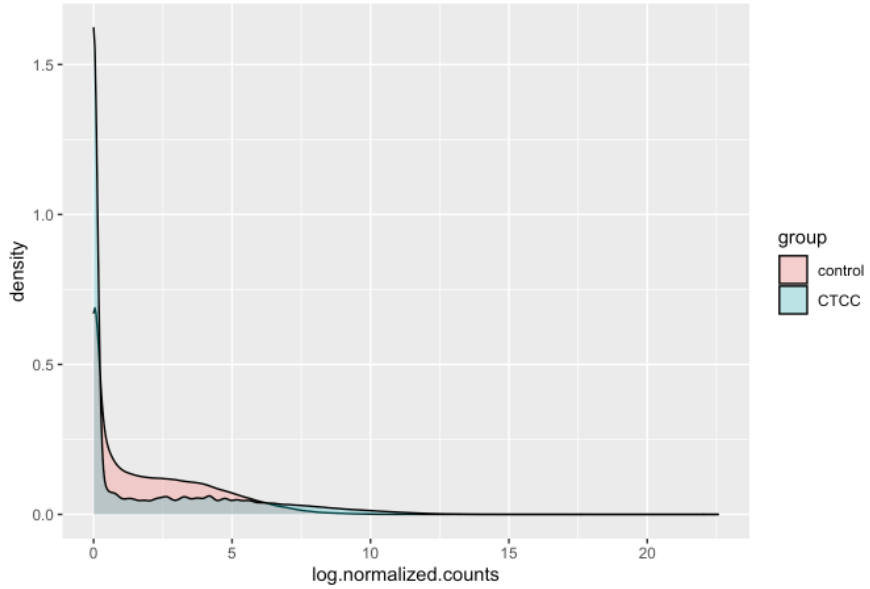


Figure 3.8: Comparison of gene expression distribution between CTCC sample and peripheral whole blood controls, showing right skew and relative scarcity in CTCC data. Log transformed DESeq2 normalized gene expression counts were to plot density distribution.

3.4 Bulk RNA Sequencing

Principal component analysis using top 1000 genes by variance demonstrate clear separation of the peripheral blood control samples from medulloblastoma sequencing samples, with 31% of variances explained by the first principal axis and 13% explained by the second (figure 3.9).

Heat map of gene expression in the top 40 most varied genes also reveal distinct gene expression profiles between medulloblastoma samples compared to the eight

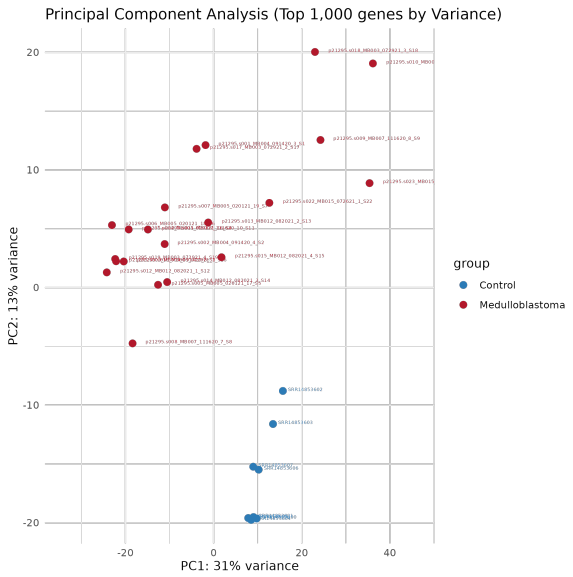


Figure 3.9: Principal component analysis using top 1000 genes by variance showing Medulloblastoma CTCC samples (red) and peripheral whole blood control (blue). Medulloblastoma CTCC and whole blood sequencing show distinct clustering by group, supporting that the Cluster Chip was able to capture and distinguish CTCCs from background whole blood.

control samples (figure 3.10). Among the highly differentially expressed genes, SOCS7 (suppressor of cytokine signaling 7) was upregulated in all medulloblastoma enriched samples sequenced compared to whole blood control. SOCS7 is predicted to play a role in brain development, and has also been implicated as an oncogene in the carcinogenesis of bladder cancer [21]. SOCS7 is not previously reported in association with medulloblastoma, and therefore represents a possible novel biomarker.

Volcano plot constructed using only the 46 genes known to be markers of cancer stem cells from adult CTCC literature show 13 genes differentially upregulated with statistical significance compared to whole blood. This is as opposed to only one gene differentially down regulated (figure 3.11). The remaining failed to meet statistical significance in differential expression analysis.

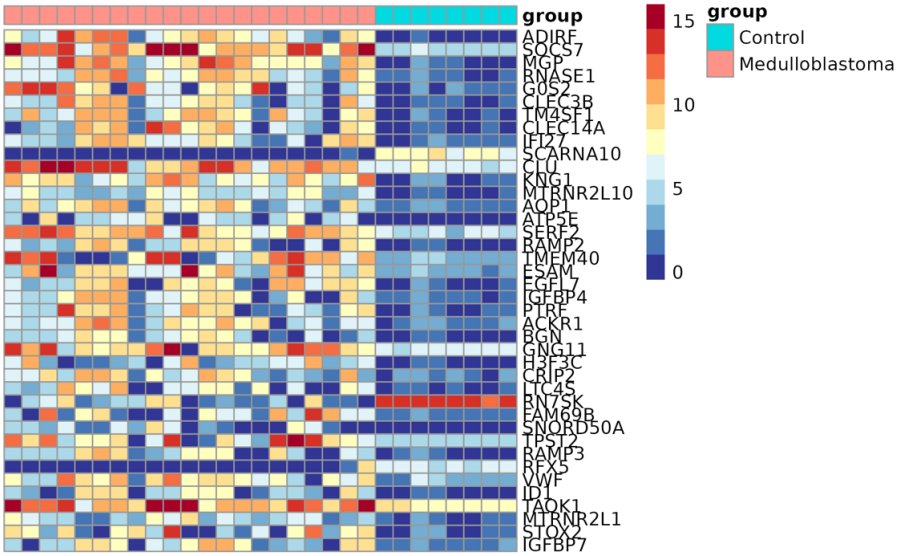


Figure 3.10: Heat map of top 40 genes with highest variance in differential gene expression on bulk RNA sequencing comparing medulloblastoma CTCC (red columns) to peripheral whole blood control (blue columns).

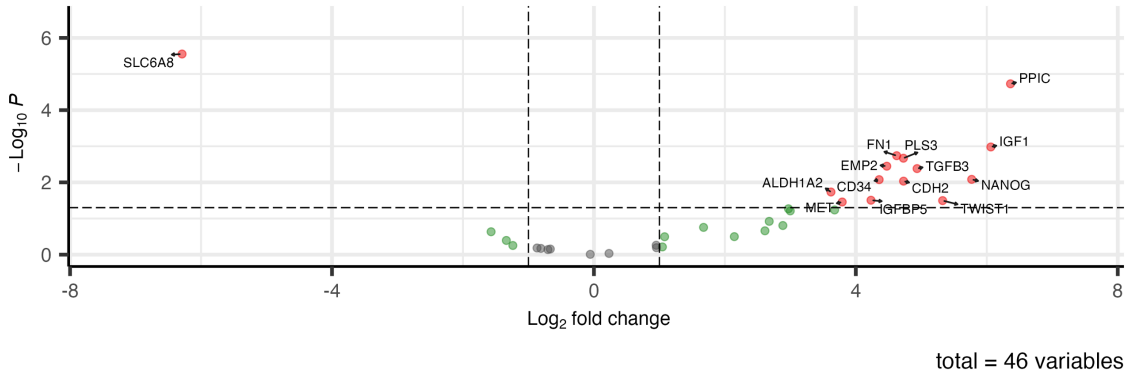


Figure 3.11: Volcano plot of 46 genes known to be markers of cancer stem cells within the adult CTCC literature, showing several genes in the upper right quadrant compared to a single gene differentially down regulated reaching significance. The result supports stemness as a potential feature for medulloblastoma CTCCs.

3.5 Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed using multiple gene set collections obtained from the Molecular Signatures Database. Normalized enrichment scores of the top 20 gene sets from the C2 canonical pathways collection is shown in figure 3.12. All top 20 gene sets achieve statistical significance at $AdjustedP < 0.05$ with collagens and collagen chain trimerization having the highest normalized enrichment scores.

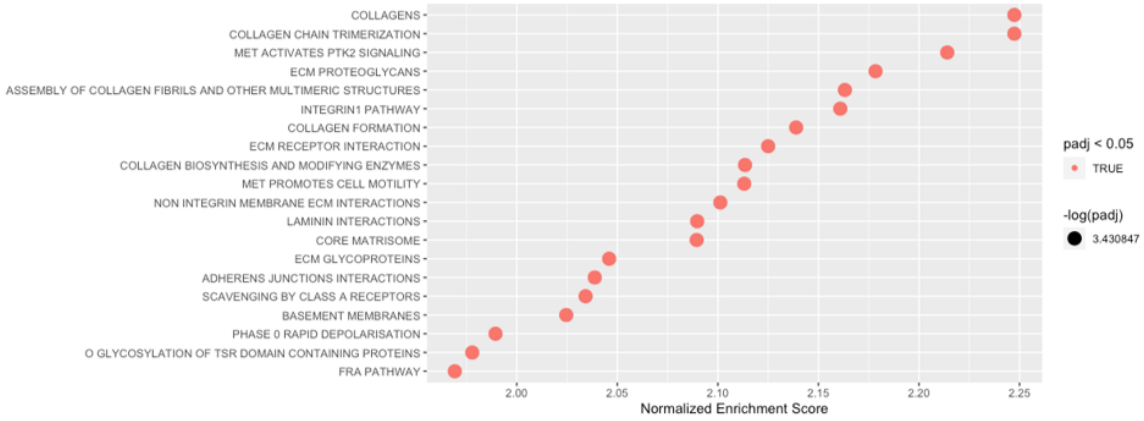


Figure 3.12: Gene set enrichment analysis by C2 canonical pathway analysis show several gene sets involved in collagen, extracellular matrix, basement membrane, and cell junction proteins.

GSEA by gene ontology biological processes shown in 3.13 demonstrate the highest normalized enrichment score in a gene set representing membrane depolarization. The second highest normalized enrichment was in basement membrane organization. Other gene sets annotated with biological processes involving action potentials are also represented. Each of the top 20 again reaches statistical significance at $AdjustedP < 0.05$.

Lastly, GSEA by oncological signature is shown in 3.14 demonstrating upregulation of oncologic pathways known to be over expressed in medulloblastoma tumor. NFκB related gene sets are the two most enriched pathways by this gene set. The mTOR and p53 pathways are noted in the top 20 of this collection, both well-known

Figure 3.13: Gene set enrichment analysis by Gene Ontology Biological Processes showing enrichment of several gene sets involved in neuronal processes. These processes including membrane depolarization, action potential, and neurotransmitter uptake.

pathways involved in cell-growth and cancers.

Of note, a gene set upregulated in embryoid bodies embryonic stem cells is also significantly enriched [15]. This is in agreement with both medulloblastoma cells' embryonal origin, and with stemness as a feature of CTCCs.

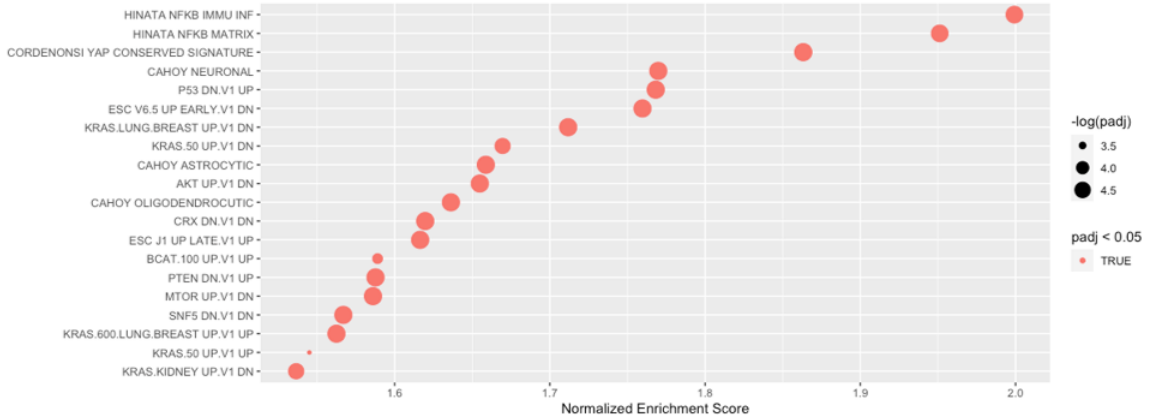
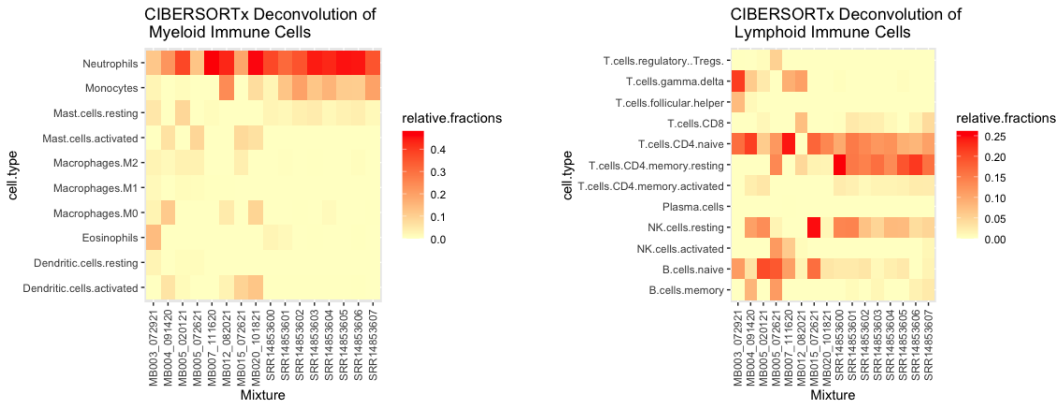


Figure 3.14: Gene set enrichment analysis by oncologic signatures collection demonstrating enrichment with statistical significance several gene sets known to be involved in cell cycle regulation and proliferation, including NFκB, PTEN, MTOR, and KRAS. Additionally, a gene set upregulated in embryonic stem cells (ESC J1 UP LATE V1 UP) is notably enriched, supporting CTCC stemness.

3.5.1 Immune Deconvolution

Immune deconvolution with CIBERSORTx was used to ascertain differences in relative fractions of immune cells between CTCC enriched samples compared to peripheral whole blood (figure 3.15). Among myeloid lineage cells, neutrophils had the highest relative fraction in both CTCC enriched samples as well as peripheral control. In the lymphoid lineage cells, both CTCC and control samples show high relative fractions of naive CD4+ T cells. High relative fraction of CD4+ memory resting T-cells is observed in whole blood control, and is notably different than that of CTCC samples.

Two-sample t-test of relative fractions from CTCC enriched samples compared to controls show statistically significant differences in CD4+ memory resting T-cells at $AdjustedP = 0.0001$, adjusted for multiple hypothesis testing. No other cell types achieved statistical significance at $p < 0.05$.



(a) Myeloid lineage cells

(b) Lymphoid lineage cells

Figure 3.15: Relative fraction of myeloid and lymphoid lineage immune cells deconvoluted through CIBERSORTx using LM22 signature matrix. Paired t-test show significant difference in relative fraction of CD4+ memory resting t-cells (figure 3.15b) with $AdjustedP = 0.0001$.

Figure 3.16 shows differences in relative fraction of immune cells from the mononuclear phagocyte system including monocytes and macrophages. Macrophages have a higher relative fraction in whole blood control samples whereas CTCC samples show

higher relative fractions of macrophages. This difference is biologically intuitive, but did not reach statistical significance under this experimental design.

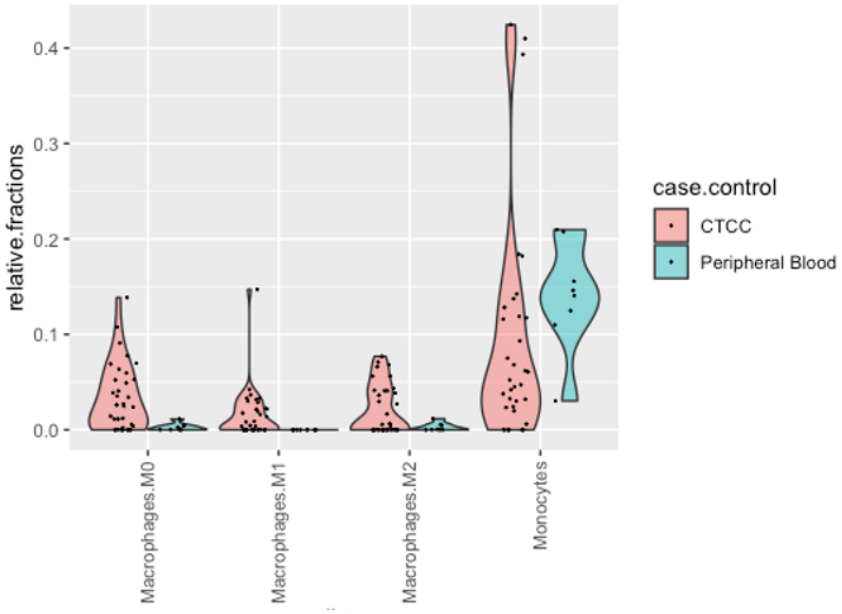


Figure 3.16: Violin plot of medulloblastoma CTCC (red) compared to peripheral blood control (blue) showing increased relative fractions of monocytes in whole blood samples compared to increased macrophages in CTCC samples. This was true of M0, M1, and M2 macrophages

Chapter 4

Discussion

4.1 Circulating Tumor Cell Cluster Quantification

Circulating tumor cell clusters (CTCCs) are not previously known to exist in pediatric patients with medulloblastoma. In our study cohort of 44 pediatric patients treated at Children's Healthcare of Atlanta, CTCCs were captured at some phase of therapy in all patients with medulloblastoma whereas none were observed from patients with either low grade glioma or non-malignant hematological conditions. Their universal presence in patients with medulloblastoma suggests that detection of CTCCs alone do not predict metastasis or survival outcomes.

CTCC concentrations are hypothesized to reflect overall tumor burden and thus correlate with disease outcomes. Supporting this is the recognition that CTCC concentrations vary with chemotherapy, often decreasing with treatment but notably rarely eliminated entirely following treatment. Treated as time-series data, CTCC concentrations were analyzed using change point detection analysis to identify inflexion points. While change points are statistically detected, to date patients with medulloblastoma have not clearly separated into distinct groups based on CTCCs concentrations, suggesting that the amount of CTCC in peripheral blood may not

necessarily distinguish population of tumors destined to metastasize.

A limitation of change point analysis is the relatively few data points available for each patient. While no theoretical lower limit is identified, it is felt that fewer than 10 observations is pushing the boundaries of statistical confidence. The study design and blood draw sampling schedule were finalized prior to the application of change detection methods, and so was not designed to optimize change point detection. Future studies of CTCCs are planned using monthly blood draws rather than an every 3 months schedule. This will provide increased data granularity and allow for more robust time-series analysis.

4.2 Transcriptomic Analysis of Captured CTCCs

CTCCs are identified as originating from medulloblastoma bulk tumor as opposed to from background whole blood on the basis of positive immunofluorescence staining for synaptophysin, a marker of neuronal lineage, and negative staining for CD45, a marker of hematopoietic lineage. Bulk-RNA sequencing of CTCC enriched samples corroborate this conclusion, as they demonstrate a transcriptomic landscape distinct to that of background whole blood.

First, fastQC results were reassuring in that GC content match the known GC% of 41% in human genomes [17]. Further, phred scores indicated high confidence in base calls. Our sequencing experiment show decreasing read quality as base pair position advances, a well-known phenomena in Illumina's sequencing by synthesis. Despite this, the majority of reads show phred scores in excess of 30 indicating fewer than 1 in 1000 incorrect base calls throughout all base pair positions.

Isolation of extremely rare circulating tumor cells from peripheral blood present a challenge in biomedical engineering addressed by the Cluster Chip. In both principal component analysis and gene expression heat map, CTCC samples separate from the

control whole blood in two distinct groupings. These results support that CTCC was successfully captured, and that sequencing was of a different population of cells compared to background whole blood.

Stemness and epithelial-mesenchymal transformations are known to be important properties of CTCCs from studies of other adult type tumors. We therefore hypothesized that stemness will also be important to CTCCs. Supporting this is the relative abundance of CTCC stem cell marker genes upregulated with statistical significance on volcano plot. Stemness as a feature of CTCCs is further corroborated by elevation of the normalized enrichment score in an embryonic stem cell gene set.

4.3 CTCC Gene Set Enrichment Analysis

Current CTCC literature from adult type tumors recognizes the importance of cell to cell adhesion and extracellular matrix remodeling. Indeed, the act of clustering and circulation requires cell adhesion and separation from the original extracellular environment. Clustered cells are resistant to anoikis, the programmed cell death process typically invoked upon separation with the extracellular matrix [6]. GSEA results are consistent with activation of these biological processes. In particular, 18 of the 20 gene sets with highest normalized enrichment scores from the C2 canonical pathway human gene collection are pathways involved in collagens, integrins, cell-cell adhesion, and basement membranes.

GSEA results by gene ontology biological processes reveal distinct neuronal functions. The pathway with the highest normalized enrichment score by a sizable margin is in membrane depolarization during action potentials, supporting CTCC cells being of neuronal origin. Several pathways indicating neuronal function are represented, including gene sets involved in action potentials, membrane depolarization, and neurotransmitter uptake. Under physiologic conditions, no cells from whole blood conduct

or relay action potentials. The result is not to suggest that CTCCs are literally firing action potentials, but rather is consistent with the fact that medulloblastomas arise from embryonal neuronal cells and so is expected to express several genes in common with mature neurons.

Gene set analysis by oncological signatures reveal several pathways that are known to be upregulated in cancer and medulloblastoma including NFκB, YAP, p53, PTEN, MTOR, and KRAS. Additionally, gene sets upregulated by astrocytes and oligodendrocytes, two neuroglial cell types found in the CNS, are also enriched [10], further supporting CTCC similarity with neuronal cell types.

A limitation of this GSEA analysis is that the transcriptomic data from which GSEA is performed may not be representative of the entire breadth of medulloblastoma disease. At the time of sequencing, no patients were yet enrolled with WNT-activated tumor, the rarest of the 4 major subgroups. Further, this GSEA is based on differential gene expression comparing CTCC to background whole blood, and so widely differing transcriptomic expression is expected. Future RNA sequencing experiments comparing CTCC with bulk cerebellar tissue and subsequent downstream GSEA will inform cellular processes specific to the circulating tumor.

4.4 Immune deconvolution

To determine cell composition from bulk RNA sequencing data, deconvolution was performed with CIBERSORTx [26]. While not meeting statistical significance, deconvolution results suggests differences noted monocytes and macrophage cell types. Whereas monocytes physiologically circulate in whole blood, macrophages are typically found in target tissue. Our data suggests that increased relative fractions of macrophages are observed in CTCC enriched samples. Of note, tumor associated macrophages (TAMs) are known to play key roles in the bulk tumor immune mi-

croenvironment [32]. It is thus feasible that the macrophage signature detected in deconvolution arise from TAMs circulating in association with CTCCs. Increased number of samples in each comparison group may confer added statistical power to detect differences in relative fraction of mononuclear immune cells.

4.5 Batch Effect

The likely contribution of batch effect requires interpretation of the results with caution. Batch effect differences are expected to exaggerate the distinction between whole blood sequencing and CTCC samples on PCA and heat maps. DESeq2 normalization with median of ratios is expected to abate some variation due to technical differences, but does not alone overcome batch effect. We propose that while some difference in transcriptome is secondary to batch effect, the transcriptomic distinction of CTCC versus background whole blood cannot be fully explained by batch effect alone. Batch effect will not selectively introduce tumor specific and neuronal signatures detected by GSEA, nor would batch effect preferentially show upregulation in collagen and integrin proteins consistent with CTCC literature from other adult tumors. Post sequencing batch correction methods such as down sampling and ComBat-seq were considered. Ultimately, a-priori minimization of batch effect by experimental design in subsequent experiments was favored.

4.6 Future Studies

Differential gene expression analysis has yet to be performed comparing CTCC against the bulk tumor from which it originated. As part of Aflac Cancer and Blood Disorder Center's Precision Medicine Program, bulk tumor sequencing was performed with GEM ExTra [30], which sequences RNA reads at 100 million per sample. CTCC sequencing was not compared to GEM ExTra results as batch effect is expected to

be even more significant given the wide disparity in platform and read depth.

To further interrogate CTCC transcriptome against that of the bulk cerebellar tumor, further RNA sequencing experiments are planned. Five patients have been identified with high CTCC content and with available patient-matched samples of peripheral blood and bulk tumor tissue. Parallel ultra-low cell number bulk RNA sequencing will be performed on each of the three tissue types and for each of the five patients at four to five technical replicates each. Cell reduction on bulk tumor and tissue samples will be performed so as to match the cell and RNA content of CTCCs prior to sequencing. This experiment is designed to detect the unique transcriptomic signature of CTCC while minimizing batch effect a-priori.

Additionally, we hypothesize immune cells to substantially contribute to the microenvironments of both bulk cerebellar tumor as well as CTCC. Further, differences in immune cell profile and microenvironments in both circulating tumors as well as in bulk tumor may predict clinical outcome. To assess the tumor immune microenvironment, 17 medulloblastoma tumor tissues samples are identified for spatial sequencing. Spatial sequencing will allow for the identification of immune cells within tumor, while concurrently identifying their spatial orientation relative to cancer cells. To date, spatial sequencing has not been described in medulloblastoma.

Chapter 5

Conclusions

The presence of circulating tumor cells or cell clusters in blood from patients with medulloblastoma is not previously known. Similarly, their role in disease pathophysiology has yet to be elucidated. From a cohort of pediatric patients with medulloblastoma, low grade glioma, and non-malignant hematological conditions, we analyzed sequential peripheral blood CTCC concentrations with multiple change point detection and performed bulk RNA sequencing of CTCC enriched samples. Transcriptomic results are consistent with CTCC arising from neuronal lineage, while suggesting the importance of stemness and basement membrane reorganization consistent with adult CTCC literature from other tumors. Immune cell deconvolution analysis shows clusters enriched samples have different immune cell composition compared to background whole blood in a T-cell sub-type.

CTCCs in medulloblastoma patients hold promise as novel biomarkers and may represent a previously unknown route for hematogenous spread. As a biomarker, CTCC may function as a tool for prognostication or evaluation of drug response without invasive imaging and neurosurgery. If CTCCs contribute to disease spread, then CTCCs may become a pharmaceutically targetable site of disease decreasing risk of metastasis. To realize these two goals, future studies are needed to describe

their behavior, relationship to primary bulk tumor, and their role in disease spread.

Bibliography

[1] N. Aceto. Bring along your friends: Homotypic and heterotypic circulating tumor cell clustering to accelerate metastasis. *Biomed J*, 43(1):18–23, Feb 2020.

[2] N. Aceto, A. Bardia, D. T. Miyamoto, M. C. Donaldson, B. S. Wittner, J. A. Spencer, M. Yu, A. Pely, A. Engstrom, H. Zhu, B. W. Brannigan, R. Kapur, S. L. Stott, T. Shioda, S. Ramaswamy, D. T. Ting, C. P. Lin, M. Toner, D. A. Haber, and S. Maheswaran. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*, 158(5):1110–1122, Aug 2014.

[3] N. Aceto, M. Toner, S. Maheswaran, and D. A. Haber. En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. *Trends Cancer*, 1(1):44–52, Sep 2015.

[4] C. Agnoletto, F. à, L. Minotti, F. Baldassari, F. Crudele, W. J. J. Cook, G. Di Leva, A. P. d'Adamo, P. Gasparini, and S. Volinia. Heterogeneity in Circulating Tumor Cells: The Relevance of the Stem-Cell Subset. *Cancers (Basel)*, 11(4), Apr 2019.

[5] S. Amintas, A. Bedel, F. Moreau-Gaudry, J. Boutin, L. Buscail, J. P. Merlio, V. Vendrely, S. Dabernat, and E. Buscail. Circulating Tumor Cell Clusters: United We Stand Divided We Fall. *Int J Mol Sci*, 21(7), Apr 2020.

[6] S. Amintas, A. Bedel, F. Moreau-Gaudry, J. Boutin, L. Buscail, J. P. Merlio,

V. Vendrely, S. Dabernat, and E. Buscail. Circulating Tumor Cell Clusters: United We Stand Divided We Fall. *Int J Mol Sci*, 21(7), Apr 2020.

[7] P. C. Bailey and S. S. Martin. Insights on CTC Biology and Clinical Impact Emerging from Advances in Capture Technology. *Cells*, 8(6), Jun 2019.

[8] N. D. Beckmann, P. H. Comella, E. Cheng, L. Lepow, A. G. Beckmann, S. R. Tyler, K. Mouskas, N. W. Simons, G. E. Hoffman, N. J. Francoeur, D. M. Del Valle, G. Kang, A. Do, E. Moya, L. Wilkins, J. Le Berichel, C. Chang, R. Marvin, S. Calorossi, A. Lansky, L. Walker, N. Yi, A. Yu, J. Chung, M. Hartnett, M. Eaton, S. Hatem, H. Jamal, A. Akyatan, A. Tabachnikova, L. E. Liharska, L. Cotter, B. Fennessy, A. Vaid, G. Barturen, H. Shah, Y. C. Wang, S. H. Sridhar, J. Soto, S. Bose, K. Madrid, E. Ellis, E. Merzier, K. Vlachos, N. Fishman, M. Tin, M. Smith, H. Xie, M. Patel, K. Nie, K. Argueta, J. Harris, N. Karekar, C. Batchelor, J. Lacunza, M. Yishak, K. Tuballes, I. Scott, A. Kumar, S. Jaladanki, C. Agashe, R. Thompson, E. Clark, B. Losic, L. Peters, P. Roussos, J. Zhu, W. Wang, A. Kasarskis, B. S. Glicksberg, G. Nadkarni, D. Bogunovic, C. Elaiho, S. Gangadharan, G. Ofori-Amanfo, K. Alesso-Carra, K. Onel, K. M. Wilson, C. Argmann, S. Bunyavanich, M. E. n Riquelme, T. U. Marron, A. Rahman, S. Kim-Schulze, S. Gnjatic, B. D. Gelb, M. Merad, R. Sebra, E. E. Schadt, A. W. Charney, C. Agashe, P. Agrawal, E. Alibo, K. Alvarez, A. Amabile, S. Ascolillo, R. Bailey, P. Begani, C. C. Bozkus, P. Bravo, S. A. Brown, M. Buckup, L. Burka, L. Cambron, G. Carrara, S. Chang, S. T. Chen, J. Chien, M. Chowdhury, D. Cosgrove, F. Cossarini, A. Dave, T. Dawson, B. Dayal, M. Dhainaut, R. Dornfeld, K. Dul, N. Eber, F. Fabris, J. Faith, D. Falci, S. Feng, M. Fernandes, D. Geanon, J. Grabowska, G. Gyimesi, M. Hamdani, D. Handler, M. Herbinet, E. Herrera, A. Hochman, J. Hook, L. Horta, E. Humblin, J. S. Johnson, S. Karim, G. Kelly, J. Kim, D. Lebovitch, B. Lee, G. Lee, G. H. Lee,

J. Lee, J. Leech, M. B. Leventhal, K. Lindblad, A. Livanos, R. Machado, Z. Mahmood, K. Mar, G. Martin, S. Maskey, P. Matthews, K. Meckel, S. Mehandru, C. Mercedes, D. Meyer, G. Mollaoglu, S. Morris, M. Nisenholtz, M. Ounadjela, V. Patel, C. Pruitt, S. Rathi, J. Redes, I. Reyes-Torres, A. Rodrigues, A. Rodriguez, V. Roudko, E. Ruiz, P. Scalzo, A. S. Schanoski, P. Silva, H. Stefanos, M. Straw, C. Teague, B. Upadhyaya, V. Van Der Heide, N. Vaninov, D. Wacker, H. Walsh, C. M. Wilk, J. Wilson, L. Xue, N. A. Yeboah, S. Young, N. Zaks, and R. Zha. Downregulation of exhausted cytotoxic T cells in gene expression networks of multisystem inflammatory syndrome in children. *Nat Commun*, 12 (1):4854, Aug 2021.

[9] S.M. Blaney, P.C. Adamson, and L.J. Helman. *Pizzo and Poplack's Pediatric Oncology*. M - Medicine Series. Wolters Kluwer Health, 2020. ISBN 9781975124793. URL `https://books.google.com/books?id=3TV4xwEACAAJ`.

[10] John D Cahoy, Ben Emery, Amit Kaushal, Lynette C Foo, Jennifer L Zamanian, Karen S Christopherson, Yi Xing, Jane L Lubischer, Paul A Krieg, Sergey A Krupenko, Wesley J Thompson, and Ben A Barres. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.*, 28(1):264–278, January 2008.

[11] F. M. G. Cavalli, M. Remke, L. Rampasek, J. Peacock, D. J. H. Shih, B. Luu, L. Garzia, J. Torchia, C. Nor, A. S. Morrissy, S. Agnihotri, Y. Y. Thompson, C. M. Kuzan-Fischer, H. Farooq, K. Isaev, C. Daniels, B. K. Cho, S. K. Kim, K. C. Wang, J. Y. Lee, W. A. Grajkowska, M. Perek-Polnik, A. Vasiljevic, C. Faure-Conter, A. Jouvet, C. Giannini, A. A. Nageswara Rao, K. K. W. Li, H. K. Ng, C. G. Eberhart, I. F. Pollack, R. L. Hamilton, G. Y. Gillespie, J. M. Olson, S. Leary, W. A. Weiss, B. Lach, L. B. Chambless, R. C. Thompson, M. K. Cooper, R. Vibhakar, P. Hauser, M. C. van Veelen, J. M. Kros, P. J. French,

Y. S. Ra, T. Kumabe, E. pez Aguilar, K. Zitterbart, J. Sterba, G. Finocchiaro, M. Massimino, E. G. Van Meir, S. Osuka, T. Shofuda, A. Klekner, M. Zollo, J. R. Leonard, J. B. Rubin, N. Jabado, S. Albrecht, J. Mora, T. E. Van Meter, S. Jung, A. S. Moore, A. R. Hallahan, J. A. Chan, D. P. C. Tirapelli, C. G. Carlotti, M. Fouladi, J. Pimentel, C. C. Faria, A. G. Saad, L. Massimi, L. M. Liau, H. Wheeler, H. Nakamura, S. K. Elbabaa, M. a Diazconti, F. n, S. Robinson, M. Zapotocky, A. Lassaletta, A. Huang, C. E. Hawkins, U. Tabori, E. Bouffet, U. Bartels, P. B. Dirks, J. T. Rutka, G. D. Bader, J. Reimand, A. Goldenberg, V. Ramaswamy, and M. D. Taylor. Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell*, 31(6):737–754, Jun 2017.

[12] A. Fabisiewicz and E. Grzybowska. CTC clusters in cancer progression and metastasis. *Med Oncol*, 34(1):12, Jan 2017.

[13] L. Garzia, N. Kijima, A. S. Morrissy, P. De Antonellis, A. Guerreiro-Stucklin, B. L. Holgado, X. Wu, X. Wang, M. Parsons, K. Zayne, A. Manno, C. Kuzan-Fischer, C. Nor, L. K. Donovan, J. Liu, L. Qin, A. Garancher, K. W. Liu, S. Mansouri, B. Luu, Y. Y. Thompson, V. Ramaswamy, J. Peacock, H. Farooq, P. Skowron, D. J. H. Shih, A. Li, S. Ensan, C. S. Robbins, M. Cybulsky, S. Mitra, Y. Ma, R. Moore, A. Mungall, Y. J. Cho, W. A. Weiss, J. A. Chan, C. E. Hawkins, M. Massimino, N. Jabado, M. Zapotocky, D. Sumerauer, E. Bouffet, P. Dirks, U. Tabori, P. H. B. Sorensen, P. K. Brastianos, K. Aldape, S. J. M. Jones, M. A. Marra, J. R. Woodgett, R. J. Wechsler-Reya, D. W. Fults, and M. D. Taylor. A Hematogenous Route for Medulloblastoma Leptomeningeal Metastases. *Cell*, 172(5):1050–1062, Feb 2018.

[14] M. Giuliano, A. Shaikh, H. C. Lo, G. Arpino, S. De Placido, X. H. Zhang, M. Cristofanilli, R. Schiff, and M. V. Trivedi. Perspective on Circulating Tumor

Cell Clusters: Why It Takes a Village to Metastasize. *Cancer Res*, 78(4):845–852, 02 2018.

[15] K. Hailesellasse Sene, C. J. Porter, G. Palidwor, C. Perez-Iratxeta, E. M. Muro, P. A. Campbell, M. A. Rudnicki, and M. A. Andrade-Navarro. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics*, 8:85, Mar 2007.

[16] Rebecca Killick and Idris Eckley. changepoint: An r package for changepoint analysis. *Journal of statistical software*, 58(3):1–19, 2014.

[17] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell,

M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. cker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, and J. Szustakowki. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, Feb 2001.

[18] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and

dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12):550, 2014.

[19] N. E. Millard and K. C. De Braganca. Medulloblastoma. *J Child Neurol*, 31(12): 1341–1353, 10 2016.

[20] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 12(5):453–457, May 2015.

[21] S. Noguchi, N. Yamada, M. Kumazaki, Y. Yasui, J. Iwasaki, S. Naito, and Y. Akao. induction through STAT3 nuclear translocation in bladder cancer cells. *Cell Death Dis*, 4(2):e482, Feb 2013.

[22] P. A. Northcott, A. M. Dubuc, S. Pfister, and M. D. Taylor. Molecular subgroups of medulloblastoma. *Expert Rev Neurother*, 12(7):871–884, Jul 2012.

[23] I. F. Pollack, S. Agnihotri, and A. Broniscer. Childhood brain tumors: current management, biological insights, and future directions. *J Neurosurg Pediatr*, 23 (3):261–273, Mar 2019.

[24] A. F. Sarioglu, N. Aceto, N. Kojic, M. C. Donaldson, M. Zeinali, B. Hamza, A. Engstrom, H. Zhu, T. K. Sundaresan, D. T. Miyamoto, X. Luo, A. Bardia, B. S. Wittner, S. Ramaswamy, T. Shioda, D. T. Ting, S. L. Stott, R. Kapur, S. Maheswaran, D. A. Haber, and M. Toner. A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat Methods*, 12(7):685–691, Jul 2015.

[25] D. J. Shih, P. A. Northcott, M. Remke, A. Korshunov, V. Ramaswamy, M. Kool, B. Luu, Y. Yao, X. Wang, A. M. Dubuc, L. Garzia, J. Peacock, S. C. Mack, X. Wu, A. Rolider, A. S. Morrissy, F. M. Cavalli, D. T. Jones, K. Zitterbart, C. C. Faria, U. ller, L. Kren, T. Kumabe, T. Tominaga, Y. Shin Ra, M. Garami,

P. Hauser, J. A. Chan, S. Robinson, L. r, A. Klekner, A. G. Saad, L. M. Liau, S. Albrecht, A. Fontebasso, G. Cinalli, P. De Antonellis, M. Zollo, M. K. Cooper, R. C. Thompson, S. Bailey, J. C. Lindsey, C. Di Rocco, L. Massimi, E. M. Michiels, S. W. Scherer, J. J. Phillips, N. Gupta, X. Fan, K. M. Muraszko, R. Vibhakar, C. G. Eberhart, M. Fouladi, B. Lach, S. Jung, R. J. Wechsler-Reya, M. vre Montange, A. Jouvet, N. Jabado, I. F. Pollack, W. A. Weiss, J. Y. Lee, B. K. Cho, S. K. Kim, K. C. Wang, J. R. Leonard, J. B. Rubin, C. de Torres, C. Lavarino, J. Mora, Y. J. Cho, U. Tabori, J. M. Olson, A. Gajjar, R. J. Packer, S. Rutkowski, S. L. Pomeroy, P. J. French, N. K. Kloosterhof, J. M. Kros, E. G. Van Meir, S. C. Clifford, F. Bourdeaut, O. Delattre, F. F. Doz, C. E. Hawkins, D. Malkin, W. A. Grajkowska, M. Perek-Polnik, E. Bouffet, J. T. Rutka, S. M. Pfister, and M. D. Taylor. Cytogenetic prognostication within medulloblastoma subgroups. *J Clin Oncol*, 32(9):886–896, Mar 2014.

[26] C. B. Steen, C. L. Liu, A. A. Alizadeh, and A. M. Newman. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol*, 2117:135–157, 2020.

[27] M. D. Taylor, P. A. Northcott, A. Korshunov, M. Remke, Y. J. Cho, S. C. Clifford, C. G. Eberhart, D. W. Parsons, S. Rutkowski, A. Gajjar, D. W. Ellison, P. Lichter, R. J. Gilbertson, S. L. Pomeroy, M. Kool, and S. M. Pfister. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol*, 123(4): 465–472, Apr 2012.

[28] M. C. Thompson, C. Fuller, T. L. Hogg, J. Dalton, D. Finkelstein, C. C. Lau, M. Chintagumpala, A. Adesina, D. M. Ashley, S. J. Kellie, M. D. Taylor, T. Curran, A. Gajjar, and R. J. Gilbertson. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol*, 24(12): 1924–1931, Apr 2006.

[29] Y. T. Udaka and R. J. Packer. Pediatric Brain Tumors. *Neurol Clin*, 36(3): 533–556, Aug 2018.

[30] T. White, S. Szelinger, J. LoBello, A. King, J. Aldrich, N. Garinger, M. Halbert, R. F. Richholt, S. D. Mastrian, C. Babb, A. A. Ozols, L. J. Goodman, G. D. Basu, and T. Royce. . *Oncotarget*, 12(8):726–739, Apr 2021.

[31] D. Williamson, E. C. Schwalbe, D. Hicks, K. A. Aldinger, J. C. Lindsey, S. Crosier, S. Richardson, J. Goddard, R. M. Hill, J. Castle, Y. Grabovska, J. Hacking, B. Pizer, S. B. Wharton, T. S. Jacques, A. Joshi, S. Bailey, and S. C. Clifford. Medulloblastoma group 3 and 4 tumors comprise a clinically and biologically significant expression continuum reflecting human cerebellar development. *Cell Rep*, 40(5):111162, Aug 2022.

[32] J. Zhang, X. Yuan, Y. Wang, J. Liu, Z. Li, S. Li, Y. Liu, X. Gong, Y. Sun, W. Wu, L. Sun, S. Du, and T. Wang. Tumor-Associated Macrophages Correlate With Prognosis in Medulloblastoma. *Front Oncol*, 12:893132, 2022.

[33] Y. Zhang, G. Parmigiani, and W. E. Johnson. : batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*, 2(3):lqaa078, Sep 2020.