Research Report

Group 6:

Abrar Faruque, David Tawse, Nasim Idris, Robert Soroka, Samantha Lenet

Introduction

The disease investigated for this study is medulloblastoma (MB). Medulloblastomas are cancers that arise from the cerebellum and can spread through the cerebrospinal fluid to other parts of the nervous system. Common symptoms include back pain, headaches, vision changes, and issues related to maintaining balance among others. Although medulloblastomas can occur at any age, it is most frequently observed in children 16 years old or younger. If caught early treatment may involve chemotherapy, radiation therapy, surgery or any combination of these approaches. Once treated patients have a 5 year survival rate of 80%.

Medulloblastoma can be divided into 4 subgroups: Wingless activated (WNT), Sonic hedge hog (SHH), Group 3, and Group 4. Each subgroup presents with a different phenotype and is caused by a different subset of genes. Groups 3 and 4 are the most common, yet the genetic correlation linking these subgroups is the least known. Despite the lack of knowledge, common patterns seen in Group 3 and 4 are the levels of the NMYC and MYC genes, with MYC amplification being a characteristic of Group 3 but not typically seen in Group 4 cases (Sarvode and Gajjar, 2023).

Target Genes

The subtypes of medulloblastoma are dynamic, making clinical outcomes highly variable. Numerous genes have been associated with the disorder as more tumours were sequenced, however the genes selected for analysis were; CTNNB1, TP53,

MYCN, KMT2D, KCNA1 and PRDM6. We chose these genes due to mutation in each one having a characterized role in the formation, propagation, or diagnosis of medulloblastoma The CTNNB1 gene encodes for a protein that is a subunit of the adherens junction (AJ) complex. It is responsible for regulating cell growth and adhesion between cells. Mutations in CTNNB1, specifically exon 3, are frequently associated with the WNT subtype of medulloblastoma. The TP53 gene encodes a critical tumour suppressor protein that regulates cell division, initiates DNA repair, and initiates apoptosis. It resides within the nucleus and binds directly to DNA in the event of damage to promote transcription of repair elements. While not associated with a specific subtype of medulloblastoma, mutations in TP53 are associated with worse outcomes across all subtypes. MYCN is a proto-oncogene that promotes normal embryonic development through the regulation of gene expression and cell growth. It primarily acts as a transcription factor for genes involved in the development of the nervous system, lungs, and heart. Given its prominent role in embryonic development, MYCN sees mutations across all subtypes of medulloblastoma. KMT2D encodes a histone methyltransferase that methylates the Lys-4 position of histone H3. This encoded protein is part of a larger protein complex called ASCOM. ASCOM is a transcriptional regulator of the beta-globin and estrogen receptor genes. Mutations to this gene can cause Kabuki syndrome, and are also associated with the SHH subtype of medulloblastoma. KCNA1 provides instructions for making the Kv1.1 potassium channel. These channels are critical for regulating electrical signals in nerve cells by controlling the flow of potassium ions into and out of the cell. This function is essential for proper nerve cell function. KCNA1 is used as a molecular marker of group 4

medulloblastoma. Finally, PRDM6 encodes a protein which acts as a transcriptional repressor and epigenetic modifier. It plays a crucial role in several developmental processes, including heart development and the regulation of vascular smooth muscle cells and proliferation and differentiation. It causes widespread repression of chromatin accessibility and complex alterations of gene expression patterns in patients with medulloblastoma.

Methodology:

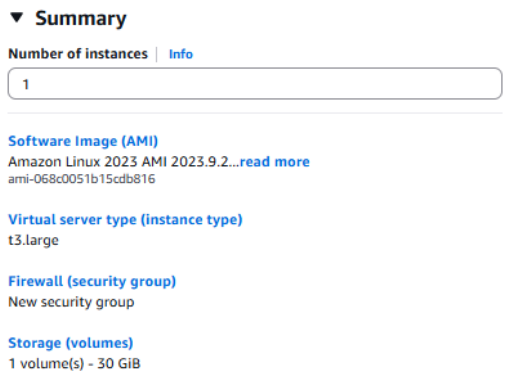
Genes were derived from commonly known or suggested markers for each sub-group. Once genes were linked to each subgroup, the NCBI database for genes was searched to get relevant information on genes, including the reference sequence.

A preliminary search was done in gnomAD to find variants observed for each gene as well as searching the SRA database for sequence data.

SRA database was searched using keywords such as by subgroup and medulloblastoma (example: (WNT AND medulloblastoma)). Desired sequences were desired if size was relatively low (no whole genome/exome), was Illumina short-read sequence data (no Nanopore, Pacbio, ATAT, CHiP, RNA-seq, etc.).

Once both the desired refseq and disease sequence were determined, analysis could begin.

Our pipeline used the Amazon Web Services Learner Lab (AWS), running an instance of EC2 with the following settings (see below).



The dependencies installed and code used in this pipeline can be viewed in the supplementary.

The general pipeline is as follows:

The following tools were used for the pipeline:

Downloading Seq data -> SRA tools, entrez-direct, fastq-dump

QC analysis of sequence -> fastqc, fastp

Indexing and Alignment ->samtools (bwa mem)

Filtering and post-alignment processing -> samtools (for indexing), gatk (Variant calling)

Variant annotation-> snpEFF

Once the annotated vcf file was obtained, key variants were identified (missense, nonsense, frameshift, splice-site variants) and referenced with the gnomAD database to determine if variants were observed in any populations with high frequency to rule out polymorphic changes. If not, the observed variant was considered disease-causing, with further literature searching to estimate the correlated prognosis of the variant. If the variant was not observed in population and literature supported the prognosis of the variant was high, we conclude that the variant was disease-causing if the gene was an

oncogene, or if only mutated reads were observed (via tview) for tumor suppressor genes.

Results:

genetic variation analysis

297 variants were identified in total across all Subgroups, PRDM6 contributed the most variant calls with 184 variants. (See supplementary for full .vcf files of corresponding genes)

Gene	**Mapped Variants in gnomAD**	**Mapped Variants of Unknown Significance absent in gnomAD**
CTNNB1	None found	10 intergenic variants 17 intronic variants
KCNA1	None found	32 intronic variants
KMT2D	None found	21 intergenic variants 21 upstream variants
MYCN	None found	Several variants in non-coding region (look at supplementary for full list)
PRDM6	None found	187 intronic variants
TP53	17-7674109-G-A -> intronic 17-7674797-T-C -> intronic 17-7676483-G-C -> intronic	Several variants in non-coding region (look at supplementary for full list)
Gene	**Pathogenic MD Variants absent in gnomAD**	**Benign MD Variants present/absent in gnomAD**
CTNNB1	c.379G>T|p.Glu127* -> nonsense	c.573T>C|p.Ser191Ser -> synonymous c.2340C>T|p.Asp780Asp -> synonymous
KCNA1	None found	None found
KMT2D	None found	None found
MYCN	c.1303G>A|p.Ala435Thr -> missense	None found
PRDM6	None found	None found
TP53	c.743G>A|p.Arg248Gln -> missense c.578A>G|p.His193Arg -> missense c.524G>A|p.Arg175His -> missense	17-7674109-G-A -> intronic 17-7674797-T-C -> intronic 17-7676483-G-C -> intronic

Discussion:

Review of variants found in genes for corresponding MB subgroupings

After variant annotation was done using snpEFF, all variant calls were contrasted with the gnomAD database. Variants that were not observed in the database were then filtered based on if they were protein coding, if they were protein coding, depending on whether they were missense or nonsense mutations. All nonsense and frameshifts were considered pathogenic as per ACMG guidelines. All synonymous mutations were considered benign.

WNT: A pathogenic nonsense variant was found in CNNB1, although the finding could be due to pure chance because of the small sample size, it is evidence to support it as a strong candidate for identifying this subgroup.

SHH: A number of missense variants were found in the MYCN and TP53 genes, although further analysis should be done to determine if they're disease-causing. Due to low frequency in the gnomAD database, we can determine these as disease-causing.

Group 3 and Group 4: A large number of variants were found in non-coding regions of respective genes. Since variants are not observed in the population database, further investigation is needed to identify phenotypic effect and have been left with VUS status

Limitations of Analysis

Most of the sequencing data we used for our analysis was from short-read illumina sequencers, the drawback of this sequencing method is the inability to identify large structural variants (large insertions and deletions), our pipeline also filtered for variants that could have mapped to potential fusions. BWA mem is best suited for reads of 70 bp - 100 bp in length.

Our sample size was incredibly small and if our findings are not representative of any conclusive findings seen in MB, it is possible that genes analyzed are representative of each subgroup; we just happened to use a sequence that represented the minority. VUS calls always have and will continue to be a challenge to determining the clinical significance of findings.

Final thoughts

Although our analysis is limited, it is a good proof of concept to identify potential biomarkers to identify subgrouping of medulloblastomas. If pathogenic variants can be found in genes in a reasonable proportion of randomly sequenced samples of a given subgrouping then it can be concluded that the gene is a good target for identifying subgrouping for MB.

Sources:

Medulloblastoma. (2025, August 27). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/22591-medulloblastoma

Medulloblastoma - Symptoms and causes. (n.d.). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/medulloblastoma/symptoms-causes/syc-20579268

Supriya Sarvode, Amar Gajjar. Review of the impact of molecular analysis on the therapy of medulloblastoma. Pediatric Hematology Oncology Journal.Volume 8. Issue 2,2023. Pages 121-128.ISSN 2468-1245.https://doi.org/10.1016/j.phoj.2023.05.001. (https://www.sciencedirect.com/science/article/pii/S2468124523000268)

Taylor, M. D., Northcott, P. A., Korshunov, A., Remke, M., Cho, Y., Clifford, S. C., Eberhart, C. G., Parsons, D. W., Rutkowski, S., Gajjar, A., Ellison, D. W., Lichter, P., Gilbertson, R. J., Pomeroy, S. L., Kool, M., & Pfister, S. M. (2011). Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathologica, 123(4), 465–472. https://doi.org/10.1007/s00401-011-0922-z

Sanghrajka, R. M. et al. KMT2D suppresses sonic hedgehog-driven medulloblastoma progression and metastasis. iScience 26, 107831 (2023).

Shrestha, S., Morcavallo, A., Gorrini, C. & Chesler, L. Biological role of MYCN in Medulloblastoma: Novel therapeutic opportunities and challenges ahead. Frontiers in Oncology 11, (2021).

Zhukova, N. et al. Subgroup-specific prognostic implications of tp53 mutation in Medulloblastoma. Journal of Clinical Oncology 31, 2927–2935 (2013).

Chiang, J. et al. A ctnnb1‐altered medulloblastoma shows the immunophenotypic, DNA methylation and transcriptomic profiles of shh‐activated, and not wnt‐activated, Medulloblastoma. Neuropathology and Applied Neurobiology 48, (2022).