Characterization of the genetic alterations in RTEL1 deficient mouse medulloblastoma model

by

Nikho Angelo Araullo Hizon

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biochemistry and Medical Genetics

University of Manitoba

Winnipeg

Abstract

Background: Medulloblastoma (MB) is a common pediatric brain tumour that originates from the stem and progenitor cells in the cerebellum and could be caused by a variety of genetic factors. Several omics-based studies have shown that MB is a heterogenous tumour which can be divided into 4 molecular subgroups: WNT, SHH, Group 3, and Group 4. Development of these subgroups of MB differs and is poorly understood.

Regulator of telomere length-1 (RTEL1) is a DNA helicase-like protein which has been demonstrated to be required for the maintenance of telomere and genomic integrity. This function of RTEL1 could be essential for development as indicated by RTEL1 knockout study. More recently, by using a conditional knockout approach, our lab demonstrated that loss of RTEL1 function in cerebellar stem cells is able to induce MB formation which recapitulates pathological features of human MB, indicating that RTEL1 could represent a new genetic pathway required for protecting cerebellar stem cells from MB formation. However, how loss of RTEL1 function transforms these stem cells to form MB is unknown. Given the demonstrated role of RTEL1 in the maintenance of telomere and genomic integrity, we hypothesize that loss of RTEL1 function in cerebellar stem cells could induce some genetic alterations which act as driver mutations in this tumorigenesis. My thesis aims to use the combined genetic analysis and bioinformatic approaches to identify these genetic alterations in RTEL1-deficient mouse MB model, which could have potential to better understand the pathogenesis of MB as well as to establish specific therapeutic targets for treating this malignant pediatric brain tumour.

Research findings: In order to determine whether the RTEL1-deficient mouse MB model is valuable for identifying the genetic alterations that can help us understand the pathogenesis of MB, in the first part of my thesis I characterized which molecular subgroup of human MB this mouse

model could belong to. For this, I used both supervised (i.e. SVM method) and unsupervised (i.e. PCA and t-SNE method) bioinformatic approaches to compare the gene-expression profiles of RTEL1-deficient mouse MBs with ones of human MBs. Both analyses showed RTEL1-deficient mouse MB model could have increased expression of MycN and many other genetic signatures presented in human SHH MB, indicating that this mouse model could belong to SHH MB. Therefore, RTEL1-dificent mouse MB model can serve as a genetic tool for studying human SHH MB.

To search for the driver mutations in RTEL1-deficient mouse MB model, in the second part of my thesis, I used several genome-wide approaches, such as array-based CGH, whole exome and RNA sequencing to characterize the recurrent genetic alterations in this mouse model. These analyses revealed massive genetic alterations, including genomic rearrangement, amplification, point and insertion/deletion mutations, in this mouse model. Among them, MycN amplification was presented in more than 65% RTEL1-deficient mouse MBs. This genetic alteration was also found to be associated with increased MycN expression and enriched MycN targeted genes in these tumours. Even in RTEL1-deficient mouse MBs without MycN amplification, high MycN expression was also observed. Given the demonstrated oncogenic role of MycN amplification or overexpression in tumorigenesis, my findings strongly support that this genetic alteration could be an important driver mutation that mediates MB formation in this mouse model. Future studies with this mouse model to identify the downstream effectors of MycN could better understand the role of this oncogene in this tumour formation.

To determine the mechanism that could contribute to genomic rearrangement and MycN amplification in RTEL1-deficient mouse MB model, in the third part of my thesis I used array-based CGH to characterize the copy number profiles of these tumours. 13 out of 21 tumours were

found to display chromothripsis-like pattern, such as oscillating copy number changes with no more than three states on a particular chromosome. Since chromothripsis is a catastrophic genetic event that can lead to a massive genomic rearrangement including amplification, this finding indicates that chromothripsis could be one main mechanism contributing to the genomic instability and MycN amplification in RTEL1-deficient mouse MB model.

Significance: My thesis not only demonstrates that RTEL1-deficient mouse MB model could belong to SHH MB, but also identifies MycN amplification as a potential driver mutation in this mouse model. This progress could stimulate interesting future research on using this mouse model to further characterize the genetic alterations for a better understanding the pathogenesis of this group of MB. In addition, the finding of chromothripsis in this mouse model by this thesis work also strengthens the pathogenic role of this genetic event in MB formation.

Key words: Medulloblastoma; RTEL1; telomere maintenance; conditional knockout; cerebellar stem cells; MycN; genomic rearrangement/amplification; chromothripsis; comparative genomic hybridization (CGH); whole exome sequencing; RNA sequencing; bioinformatics

Acknowledgements

I would like to express my sincerest appreciation of Dr. Hao Ding for letting me join his lab, his continued unwavering guidance, not just throughout this project but also in navigating my career choices. Thank you for allowing me this invaluable opportunity of learning and showing me what I am capable of.

Thank you to Xiaoli Wu, who has been an excellent role model. I will forever be inspired by your research expertise and kindness.

To my committee members, you have my deepest gratitude for your guidance. I appreciate the time all of you have given me to ensure my success in this department and in my future. This thesis would not have been possible without your expert suggestions and advice.

I owe my loving family and friends, for their encouragement in this endeavor of mine. Without your unwavering love and understanding, this work would never have been finished.

I would also like to thank the Department of Biochemistry and Medical Genetics for welcoming me into the family and providing support. Special thanks to the student mentors, specifically Berardino Petrelli and Lucile Jeusset, whose guidance and advice has greatly improved the quality of my stay here, as I'm sure it has many others.

Table of contents

List of figures

List of tables

List of abbreviations

EGL	External granular layer
IGL	Internal granular layer
SVZ	subventricular zone
GNP	Granule neuron precursor
WNT	Wingless / integrated
SHH	Sonic hedgehog
DN	Desmoplastic / anaplastic
MBEN	Medulloblastoma with extensive nodularity
LCA	Large cell anaplastic
MB	Medulloblastoma
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor embedding
RTEL1	Regulator of telomere length 1
RTEL1-MB	Rtel1-deficient medulloblastoma mouse model
WES	Whole-exon sequencing
PCA	Principal component analysis

Chapter 1: Introduction

1.1 The molecular characteristics of human medulloblastoma

Medulloblastoma (MB) is a malignant tumour that occurs in the cerebellum and accounts for 20% of pediatric brain tumours[1–3]. Although 70-80% of patients survive their encounter with MB[4–6], MB treatment options are aggressive and can result in morbid long-term side effects. The current treatment options are surgical resection, craniospinal irradiation, and chemotherapy. As the tumours present in the cerebellum, which is a critical structure for both motor and nonmotor functions[7], it is crucial that normal tissue is left undisturbed during surgical resection of tumour tissue. However, even with careful consideration and a skilled clinical team, surgical resection is a highly invasive procedure that can result in persistent cognitive deficits, speech deficits, and ataxia in as high as 24% of patients[8]. Craniospinal irradiation and chemotherapy are also highly taxing procedures which here are aimed to affect the brain. Along with the well-known ototoxicity that comes with chemotherapeutics such as cisplatin, these patients suffer from intellectual decline, variable learning disabilities, and psychological problems[9–11]. Here, we must be extra cautious as MB patients can be as young as 3 years old or even younger[12–14], and younger patients have been shown to show more severe cognitive deficits as a result of radiation[15]. Thus, it remains important to improve current therapeutic options or develop novel therapeutics for this disease.

Difficulties in MB treatment can be attributed to its heterogeneous nature. To further improve patient experience, MB heterogeneity is being accounted for by the classification of MB via histological and genomic hallmarks. There are four histological variants found in human MB[16]: Classic, large cell/anaplastic (LCA), desmoplastic/nodular (DN), and MB with extensive nodularity (MBEN). Classic histology, the most common occurring one, is characterized by its

composition of small round cells which have a high nuclear-to-cytoplasmic ratio and are undifferentiated[17,18]. These tumour cells are usually arranged as a rosette (Figure 1-1a). DN MBs are defined by the presence of nodes in the cancer tissue[17,18] (Figure 1-1b). DN MBs can also be characterized by the presence of collagen fibers, which are not present in the other MB histological classifications. MBENs are a subset of DN MBs that have larger and more frequent nodes composed of more differentiated neural cells[17,18]. Of all the histological variants, DN MBs and MBEN have the best prognosis with 55% of patients having no recurrence after 8 years[5]. LCA MBs includes all MBs with larger cells and/or anaplasia (Figure 1-1c and d). MBs with large cells frequently display anaplastic nuclei (Figure 1-1c) and presence of anaplasia indicates similar prognosis to large cell MBs. As such, MBs displaying anaplasia, even without the large cell characteristic, are grouped together with the large cell MBs[17–19]. The LCA histological variant has the worst overall prognosis with only 14% of patients having no recurrent after 8 years[5].

Histological classification does not account for all the heterogeneity in human MBs. Several genome-wide genetic analyses have been performed since 2012 which revealed that human MBs can be divided into four subgroups based on multi-omics data: WNT, SHH, Group 3, and Group 4[14,20,21]. These four subgroups differ transcriptionally, genetically, demographically, and prognostically.

1.1.1 The WNT subgroup

The WNT subgroup consists of ~10% of sporadic MBs[20,22]. Majority of WNT MBs have a classic histology[18] with ~90% survival rate and are rarely metastatic[13,14,20] which makes this group of MB the low risk patient group[23]. These tumours are more commonly found in children

(3-16 years old) and adults (>16 years old) than in infants (<3 years old)[13,14,20]. The WNT signaling pathway is highly upregulated in this subgroup implying it as the driving mechanism. The most common genetic alteration in WNT MBs is an activating mutation in β-catenin (CTNNB1) which is the downstream signal transducer in the WNT signaling pathway that is responsible for binding to transcriptional activators thus promoting transcription. The β-catenin protein is found in the cytoplasm and activation of the signaling pathway leads to accumulation

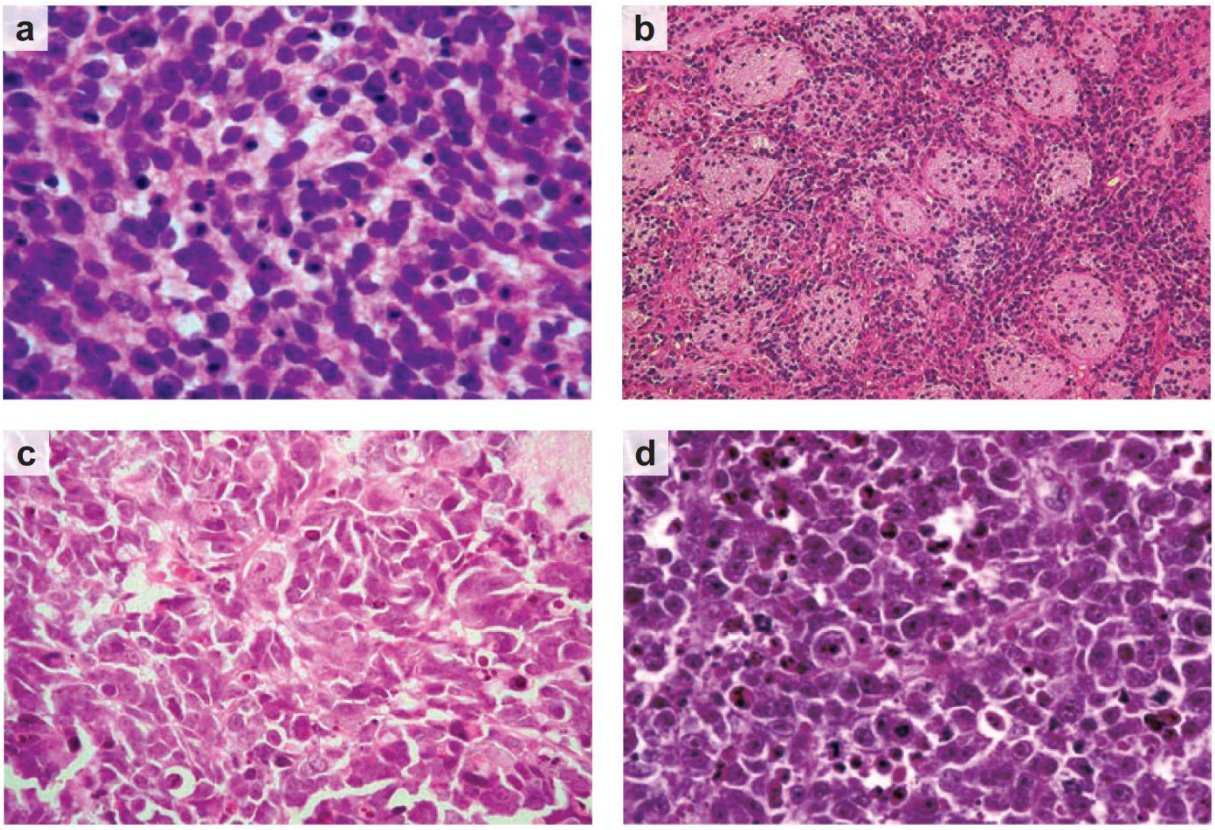


Figure 1-1. Histological classification of human MB. (a) Classic MB is composed of small round cells with a high nuclear-to-cytoplasmic ratio. (b) DN MBs contain nodes within the cancer tissue which contain collagen. (c) Anaplastic MBs, a subset of large cell MB, are composed of neuronal granular cells which lose normal morphology with altered nuclear-cytoplasmic ratio and high mitotic index. (d) Large cell MBs are have large round cells in which nuclei contain many vesicles that does not stain evenly under H&E staining. This image was adapted from Gilbertson and Ellison[17] with copyright permission from the Annual Review of Pathology.

of β-catenin and its eventual translocation to the nucleus where it can interact with transcriptional activators. Mutations in CTNNB1 act by stabilizing β-catenin facilitating its localization to the nucleus[14,22].

Another common mutation in WNT MBs occurs in the P53 gene. P53 is a well-known tumour suppressor gene (TSG) that is important for several cancer types[24–26]. Aside from these mutations, the overall genome of these MBs are mostly balanced. WNT MBs rarely have somatic copy number changes, with the exception of the deletion of chromosome 6 which rarely presents in other subgroups[14,21,22]. A possible gene targeted by this deletion is TULP4[27], which has been suggested as a TSG based on its role in neuronal cell apoptosis.

Predisposition to WNT MB is seen in people with Type II Turcot syndrome, which is characterized by an inherited inactivating mutation in adenomatous polyposis coli (APC), a key inhibitor of the WNT signaling pathway[22].

1.1.2 The SHH subgroup

The SHH subgroup contains ~30% of all MBs characterized by the classical histological features[14,20]. However, DN and LCA histological variants are also seen in this subgroup[14,20]. SHH MBs occurs in both infants (<3 years old) and adults (>16 years old), with favorable and variable prognosis respectively[14,20]. SHH subgroups are thought to result from disruptions in the SHH signaling pathway commonly caused by mutations in PATCHED1 (PTCH1), suppressor of fused (SUFU), or smoothened (SMO), and the amplification of GLI1 and GLI2[14,20,28,29]. Consistent with this, SHH MBs often have chromosomal deletion of 9q, where the PTCH1 gene is located[13,14,20]. SHH MB prognosis varies from patient to patient. Patients presenting with

metastasis or MYCN amplification are considered to be at high risk (50 – 75% survival)[23], while those with P53 mutations are at a very high-risk (<50% survival)[23,30].

Predisposition to SHH MBs is seen in people with germline mutations in PTCH1, called Gorlin syndrome[14,20]. Germline mutations of SUFU are also possible, and often lead to infantile MB[31,32].

1.1.3 The Group 3 subgroup

The Group 3 subgroup includes ~20% of all MBs and mainly display either a classic or LCA histology. This subgroup is known for having the poorest prognosis (<50% survival)[14,20,23]. Group 3 MBs predominantly present in infants (<3 years old) and children (3-16 years old)[14,20]. Unlike WNT and SHH MBs, Group 3 MBs are characterized by genomic instability. They commonly contain gains of chromosomes 1q, 7, and 17q, and deletions of 5q, 10q, 11, 16q, and 17p[13,14,20]. Gain of chromosome 17q and loss of chromosome 17p is often due to the formation of an isochromosome 17q. In addition, group 3 MBs have significantly higher c-MYC expression than the other subgroups except WNT group, possibly indicating c-MYC signaling pathway as the tumour driving mechanism[14,20,28]. c-MYC amplification appears more frequently in Group 3 MBs (10% - 17%) than in any other subgroup[14,22]. Other commonly upregulated pathways are Aberrant Notch and TGFβ signaling which could also play some important pathogenic roles in these MBs[22].

1.1.4 The Group 4 subgroup

Group 4 MB is the most common subgroup of MBs, accounting for ~40% of all MB occurences[13,14,20,28]. Most Group 4 MBs show classic histology, although there are cases with

LCA features[14,20]. This group of MBs frequently metastasize to the brain stem or spinal cord, resulting in a poor prognosis[20]. Group 4 MBs are rarely found in infants (<3 years old), most patients are children (3-16 years old) and adults (>16 years old). Similar to Group 3, Group 4 MBs often develop genomic instability[14,20], this includes recurrent amplification of MYCN and cyclin-dependent kinase 6 (CDK6) and generation of an isochromosome 17q[20]. Besides the genetic mutations that enhance MYCN and CDK6 activities, 30% of Group 4 MBs were also found to contain mutations involved in histone lysine modifications, such as ZMYM3, a histone deacetylase, and KMT2C, a histone lysine N-methyltransferase[22]. These two genetic mutations could play important pathogenic roles as they generally affect gene regulation of proto-oncogenes or TSGs[33,34]. In Group 4 MBs, neuronal differentiation and neuronal development genes, such as the glutamatergic receptors GRM1 and GRM8, are also highly expressed and may be promoting MB formation[14]. However, compared to the other subgroups, there is a lack of scientific evidence supporting the tumorigenic function of the upregulation of these genes. Thus, the driving mechanism behind Group 4 MBs is less characterized and requires further research.

Group 4 MB prognosis varies depending on metastasis status. As mentioned above, Group 4 MBs are frequently metastatic and patients that present with metastasis have poor prognosis and are considered to be in high-risk[23]. Rarely, patients with Group 4 MB have no metastasis. The patients with no metastasis and whole chromosome 11 loss or whole chromosome 17 gain have much better prognosis[23], with survival rates reaching >90%, they are considered to be at very low risk.

Group 3 and Group 4 MBs share certain genetic alterations. Both subgroups recurrently show amplification and overexpression of OTX2[14] as well as aberrant expression of growth factor independent 1 transcriptional repressor (GFI1), a zinc finger protein whose function is tied

to a complex that controls histone modifications leading to silencing of target genes, and growth factor independent 1B transcriptional repressor (GFI1B), another zinc finger protein involved in a complex that controls development and maturation of erythrocytes and megakaryocytes. Interestingly, GFI1 and GFI1B are mutually exclusive. Interestingly, the aberrant expression of GFI1 or GFI1B is not due to mutations directly occurring in these genes, but is instead due to structural variants that affect their expression through rearrangement of enhancers, a process that has been termed "enhancer hijacking"[22,35].

1.1.5 The subtypes of Group 3 and 4 human MBs

More recently, based on epigenetic, proteomic, and phosphoproteomic data, the Group 3 and 4 of human MBs have been further divided into α, β, and γ subtypes[36]. Group 3α is characterized by frequent losses of chromosome 8q and is composed of infant (<3 years old) patients with metastatic presence and favorable prognosis. Group 3β is enriched for OTX2 amplifications, are less metastatic, and frequently present with GFI1 and GFIB activation, also with a similar favorable prognosis as Group 3α. Group 3γ is characterized by frequent gains in chromosome 8q, which result in c-MYC amplification, and higher frequencies of isochromosome 17q formation. Group 3γ is associated with the poorest prognoses. Group 4α is characterized by recurrent gains of CDK6 and chromosome 7q and loss of chromosome 8p. All MYCN amplified Group 4 tumours fall under the subtype Group 4α. Group 4β MBs are characterized by SNCAIP duplication, almost ubiquitous i17q, and have the most stable genomes (small number of SCNAs). Group 4γ are characterized by CDK6 amplification, similar to 4α, loss of p arm of chromosome 8 and gain of the q arm of chromosome 7.

The molecular characteristics of different groups of human MBs are summarized in Table 1-1.

Table 1-1. Summary of the molecular characteristics of the human MB subgroups[13,14,36,37].

	SHH	**WNT**	**Group 3**	**Group 4**
Histology	DN Classic LCA MBEN	Classic LCA (rarely)	Classic LCA	Classic LCA
Point mutations	PTCH1 SMO SUFU TP53 TERT	CTNNB1 APC	No known recurrent mutations	ZMYM3 KMT2C
Expression signature	Hedgehog signaling MYCN overexpression	WNT signaling MYC overexpression	Photoreceptor GABAergic MYC overexpression OTX2 overexpression GFI1 or GFIB activation	Neuronal Glutamatergic MYCN overexpression MYC overexpression OTX2 overexpression GFI1 or GFI1B activation
Amplifications	GLI1 GLI2 MYCN YAP1	No known recurrent amplifications	MYC OTX2 1q 8q	MYCN CDK6 OTX2 SNCAIP 7q
Deletions	PTEN 9q 10q	Monosomy 6	DDX31 5q 10q 8q Isochromosome 17q	X 8p Isochromosome 17q

1.2 The cell origin of MBs

1.2.1 The development of the cerebellum

Cerebellar development in mice has been thoroughly studied and provides many fundamental information that helps to understand human cerebellar formation. In mice, cerebellar development begins after formation of the neural tube, during embryonic development, when anterior regions of the neural tube form a structure called rhombomere 1

(Figure 1-2). Rhombomere 1 formation is directed by the expression of a combination of transcription factors, including OTX2 and GBX2[38,39]. The development of Rhombomere 1 into cerebellum begins at the 4th subventricular zone (SVZ). Here, a group of cerebellar stem cells proliferate until embryonic day 10.25 (E10.25) when the stem cells start differentiating into the neurons which will later make up the cerebellar nuclei[40,41]. At E11.5, a subset of these stem cells migrates radially to form a cell layer that covers the dorsal surface of the neural tube, setting the foundation for cerebellar development. In this foundation, precursor cells are formed, i.e. Purkinje and interneuron precursors, which then migrate, with the guidance of radial glial cells, to form the infrastructure of the cerebellum (E11 – E14)[40].

Cerebellar stem cells from the SVZ also contribute to the formation of a germinal zone in the rhombic lip (RL). The progenitor cells generated at the RL form the external granular layer (EGL) of the cerebellum (Figure 1-2). The EGL is the source of granular neuron precursors (GNPs), which differentiate into the neurons and astroglial cells of the cerebellum[41,42]. Purkinje neurons derived from the cerebellar infrastructure provide the mitogenic signal, which is a SHH signalling protein, that induces GNP proliferation in the EGL. Math1 and MYCN are also required for the GNP proliferation and neural lineage cell identity[28,43–46]. Math1 expression at this stage of development is regulated by the antagonistic relationship between BMP and Notch1 receptor signaling. GNP proliferation in the EGL continues until the second postnatal week when a subset of GNPs begin migrating towards the inner cerebellum to form the inner granular layer (IGL)[47,48]. During this migration, the GNPs differentiate into granular neurons through downregulation of Math1[49,50] and upregulation of NeuroD1[51,52], Zic1,3[53,54], and p27 Kip1[55]. Part of the differentiation process is the development of axonal processes which is marked by the expression of Tag1[41,56].

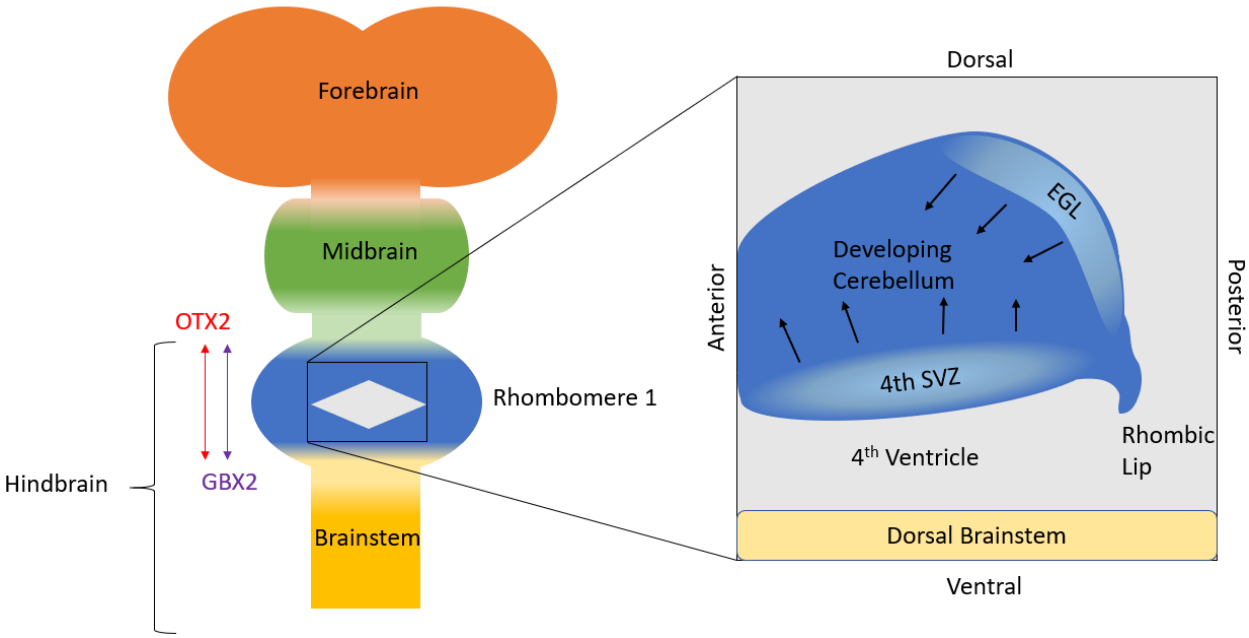


Figure 1-2. Dorsal view of the development of mouse cerebellum from the neural tube. Embryonal cerebellar development begins in rhombomere 1, marked by OTX2 and GBX2 expression, where two major germinal zones are created: external granular layer (EGL) and the 4th subventricular zone (SVZ). Stem/progenitor cells from these two zones are responsible for the development and differentiation of the cerebellum.

1.2.2 The cerebellar cells that could initiate MB formation

MBs arise from genetic dysfunctions in transient cerebellar stem cells and EGL progenitor cells that rely on diverse signals for proliferation and differentiation (Figure 1-2). Accordingly, the genetic aberrations required to induce tumorigenesis differ in these two groups of cells. Mouse models of MB have been key in understanding the different cell origins of the subgroups of MB and their varying driver genes.

Unlike other MBs, the cells of origin for WNT MBs are the stem or progenitor cells located in the dorsal brainstem[12,27,28], not in the cerebellum. The cells of origin for SHH MBs consists of both cerebellar stem cells and GNP cells in the EGL[14,20,28,57], which is consistent with the genetic profile of SHH MBs as SHH is the mitogenic factor for GNPs.

Group 3 MBs are thought to arise from Sox2+ astrocyte progenitors[58], although Prominin1+ lineage cerebellar stem cells or GNPs could also be involved[28]. GABAergic progenitors have also been suggested due to the similarity of gene expression signatures between Group 3 MBs and GABAergic neurons[28]. The cells of origin for Group 4 is unclear. Identification of these cell origins will aid in determination of driving events as it has for the WNT, SHH, and Group 3 subgroups.

1.3 The role of genomic instability in the development of MB

1.3.1 Genomic stability and its maintenance

The genome holds the complete set of genes required by the living organism. This includes tissue-specific genes required for cells to fulfill their roles, e.g. neuron-specific genes required for nerve impulse transmission. It is vital that the integrity of the genome is maintained, especially since there are several ways in which a genome can be damaged. It is estimated that tens of thousands of errors are introduced into the human genome daily, mainly due to the genomic toxicity of reactive oxygen species generated from cellular metabolism[59]. These mutations can be detrimental if not appropriately addressed. The phenotypic severity depends on the mutation type and location. For example, if a mutation were to occur in a silenced gene, no phenotypic effect is expected, however, if it occurs in a critical DNA repair gene, then the DNA repair pathway could be interrupted. These mutations can be as small as a point mutation, single nucleotide base exchange, or large enough to affect whole chromosomes, either a large genomic deletion, translocation or a massive rearrangement[60].

Human cells have developed several mechanisms to protect the genome from mutations. Due to the types of genomic mutations, these mechanisms, generally referred to as DNA-

Damage Response (DDR), have a variety of functions which basically share similar steps, i.e. detecting DNA damage, recruiting repair proteins, and repairing the errors[61]. For example, if a double-stranded break (DSB) occurs, it can be repaired by either non-homologous end-joining (NHEJ) or homologous recombination (HR)[62–65]. While NHEJ can directly ligate DNA breaks without utilizing a homologous template[64], HR requires a homologous template to guide repair but can more accurately repair DNA breaks[65]. If a point mutation occurs, it would be repaired via mismatch repair[66], base-excision repair[67], or nucleotide excision repair[68], depending on the type of mutation.

Another major mechanism important for protecting the genome from mutations is through telomere maintenance. Telomeres are nucleoprotein complexes that form at the ends of chromosomes in most eukaryotes. They harbor two key functions to maintain genomic integrity in cells. The first function is to solve the end-replication problem by recruiting telomerase. The end-replication problem is caused by the discontinuous nature of semi-conservative DNA replication that results in DNA shortening after each replication cycle[69]. Human and mouse somatic cells do not express telomerase and thus, telomeres progressively shorten with each cell division until they reach a critical length below which it is recognized as a DSB activating the DNA damage checkpoint to induce replicative senescence[69]. The second function of telomeres is to form a T-loop structure through the Shelterin protein complex, which is assembled on telomeric repeats, to protect the single-stranded chromosome ends[70,71]. T-loops are formed when the telomeric 3’ overhang invades the double-stranded region of the telomeric DNA to form a lariat structure (Figure 1-3a). Shelterin is composed of the double stranded DNA-binding proteins TRF1 and TRF2, and single stranded DNA-binding telomere-binding proteins, such as

POT1, along with a network of interacting proteins (RAP1, TIN2, TPP1) that are collectively able to block the DDR response to telomeric ends (Figure 1-3a)[71].

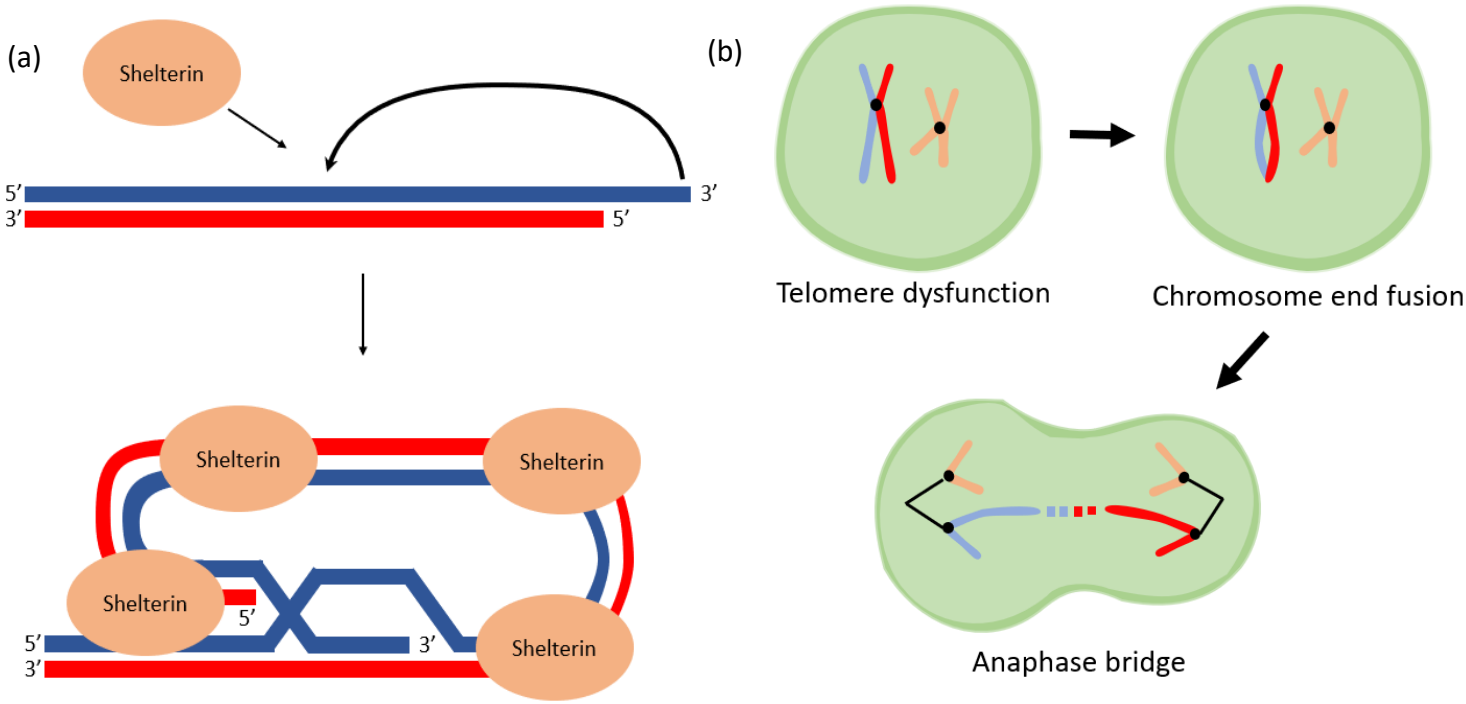


Figure 1-3. The role of telomere maintenance in protecting human cells from genomic instability. (a) The formation of T-loop of telomere which acts as a protective cap for human chromosome ends. The Shelterin complex can block DDR response to telomeric ends either by its ability to directly suppress the response to DSBs or partly through the formation of T-loops. (b) Critical telomere shortening can disrupt T-loops and expose telomere ends as DSBs. These DSBs will initiate DDR, leading to apoptosis. These DSBs on telomere ends can also induce chromosomal fusions, which further undergo break-fusion-bridge (BFB) cycles, resulting in stepwise- or chromothripsis-mediated genomic instability.

Loss of telomeric end-protection, due either to progressive telomere shortening, genetic mutations, or through the formation of a DSB in the subtelomere or telomeric repeats is a profoundly dangerous event as the chromosome end is now free to become a substrate for the DSB repair and signaling machineries. Unprotected telomeres can also engage in sister chromatid fusions or end-to-end fusions which will further initiate break-fusion-bridge (BFB) cycles, leading to genomic instability (Figure 1-3b)[60,72]. More recently, unprotected telomeres

have also been found to initiate chromothripsis through BFB cycles, resulting in a massive genomic rearrangement (see following discussion)[73].

1.3.2 Chromothripsis: one type of genomic instability frequently occurring in human MB

Genomic instability plays an important role in cancer development. It is generally accepted that genomic instability could allow normal human cells: (1) to acquire a selective growth advantage (driver mutations) for the initial clonal expansion; and (2) to accumulate genetic mutations in many other genes, such as oncogenes or tumour suppressors, that are responsible for tumour progression and invasion[74].

Classically, the accumulation of genetic mutations in cancer cells has been though to occur gradually with only a few mutations gained from each cell cycle[75,76]. However, recent studies demonstrated that some cancers could rapidly gain genetic mutations through chromothripsis[77]. In this process, a single catastrophic genomic event could shatter one or a few chromosomes into numerous pieces that upon DNA repairing, could be ligated in the wrong order to cause massive de novo genomic rearrangements. In addition, during DNA repair, some of these chromosome pieces could be lost, resulting in gene deletions, or form double-minute chromosomes, which are known to typically harbor oncogenes (Figure 1-4)[78]. Therefore, chromothripsis could have the potential to rapidly transform normal human cells into cancer cells by simultaneously targeting multiple tumorigenic genes[77].

Human MBs, specifically the group 3 and 4 MBs, are characterized by genomic instability. This genomic instability frequently exhibits complex genomic rearrangements which could have arisen through chromothripsis as revealed by several genome-wide analyses, including comparative genomic hybridization and whole genome sequencing[14,37]. In addition,

human MBs with chromothripsis have also been found to have poor prognosis. All of these indicate that chromothripsis could have an important tumorigenic function in human MB formation.

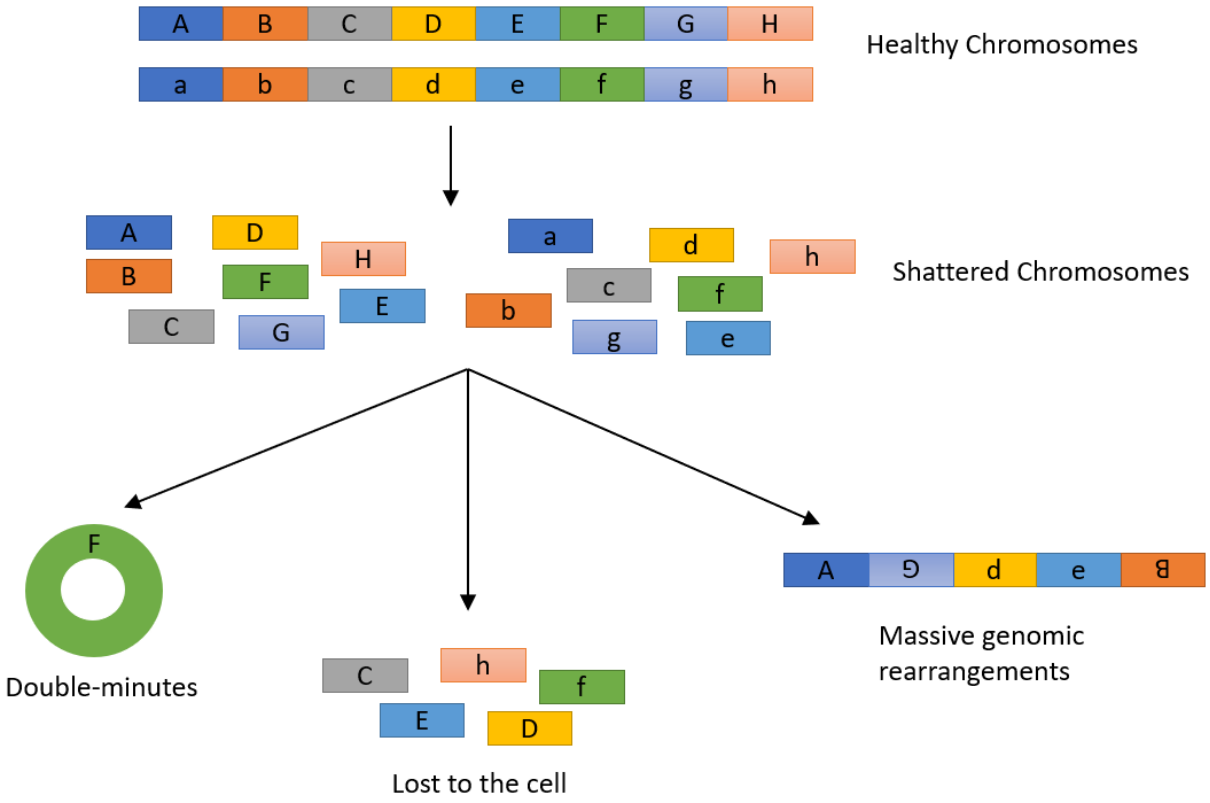


Figure 1-4. Complex chromosome rearrangements resulting from chromothripsis. In this process, a single catastrophic DNA event could shatter a single or a few chromosomes into numerous fragments which could re-ligate in a wrong order or with some missed chromosomal fragments. Some of these chromosomal fragments could also ligate to form double minutes that typically contain oncogenes.

1.3.3 The pathogenic mechanism of chromothripsis in MB formation

Although chromothripsis and its associated genetic defects have been characterized in human MB, how and when chromothripsis might occur are still not fully understood. In several mouse MB models, chromothripsis has been found to be caused by defective NHEJ and HR

activities[79]. The genomic rearrangements arising through this catastrophic event can lead to the amplification of Myc/MycN, which could be the major driving mutation for MB formation in these mouse models. This phenomenon was also observed in human MBs, indicating that chromothripsis is associated with Myc/MycN or other oncogene amplification that could be an important tumorigenic factor for MB.

Telomere dysfunction has also been implicated as an important driver for chromothripsis in human cancers[73,80]. This is mainly because the critical telomere shortening caused by defective telomere maintenance could lead to chromosome fusion, resulting in dicentric chromosome formation. In the subsequent cell cycle, this dicentric chromosome will be pulled to opposite spindle poles, forming a fragile anaphase bridge between two daughter cells that can lead to chromosome fragmentation which will be randomly assembled in each daughter cell to form genomic arrangements (Figure 1-4). This telomere dysfunction-mediated chromothripsis has been demonstrated in several human cancers, including bone sarcoma and leukemia[78]. It will be interesting to determine whether this genetic event could also be presented in MB, contributing to this tumour formation.

In summary, human MB could arise from neuronal stem or progenitor cells located in the cerebellum or brain stem due to the genetic alterations induced by genomic instability or other molecular mechanisms. In the following section, I will introduce RTEL1, a DNA helicase involved in telomere maintenance and DNA damage-repair whose dysfunction could transform cerebellar stem cells to form MB.

1.4 The role of Regulator of Telomere Length 1 (RTEL1) DNA helicase in protecting cerebellar stem cells from MB formation

1.4.1 The protein structure of RTEL1

RTEL1 belongs to the highly conserved helicase super family II[81,82]. It consists of 1209 amino acids which constitute 2 protein domains made up of 10 motifs: Fe-S motif, MLH1 motif, PIP motif, and motifs I, Ia, II, III, IV, V, and VI (Figure 1-5). Seven of these motifs (I-VI) are highly conserved in helicases[83]. The helicase activity of RTEL1 is believed to come from the Fe-S motif at the N-terminal[84] (Figure 1-5). The Fe-S motif is only found in four other DNA helicases, suggesting that these DNA helicases could form a subfamily[85]. The other two motifs in RTEL1 are Mut L homolog 1 (MLH1) binding motif and PCNA interacting protein (PIP) motif which are located at the N- and C-terminus, respectively (Figure 1-5). MLH1 motif is thought to interact with MLH1, a mismatch repair protein that interacts with PMS2 to repair DNA mismatches during DNA replication[86], whereas the PIP motif was found to mediate the interaction of RTEL1 with DNA replication machinery to promote genome-wide DNA replication[87].

Figure 1-5. The protein structure of RTEL1 DNA helicase. It consists of seven helicase domains marked by I, Ia, II, III, IV, V and VI. RTEL1 also contain several other motifs responsible for interacting with iron-sulfur clusters (Fe-S motif), with a mismatch repair protein (MLH1 motif) on the N-terminal and with PCNA (PIP motif) on the C-terminal.

1.4.2 The expression of RTEL1 during mouse development

In adult mice, RTEL1 is highly expressed in organs with frequent proliferation, such as the spleen, the testis, the thymus, the Peyer's patches, and the intestine. RTEL1 is also highly

expressed during mouse embryos development in areas of active proliferation. In addition, the expression of RTEL1 was found in lymphocytes upon stimulation, and in muscle and liver cells during regeneration, further indicating that RTEL1 expression is associated with cell proliferation[83].

To further investigate RTEL1-expressing cells during development, a knock-in mouse model (*Rtel1-CreERT2*) was generated. In this mouse model, a tamoxifen-inducible Cre (CreERT2) was specifically targeted into the Rtel1 genomic locus. Therefore, the expression of CreERT2 in *Rtel1-CreERT2* mice is tightly controlled by endogenous Rtel1 regulatory elements[88]. By breeding *Rtel1-CreERT2* mouse strain with *Rosa26-LacZ* reporter mice, the induced Cre activity upon tamoxifen administration can specifically turn on LacZ expression in RTEL1-expressing cells. These labeled RTEL1-expressing cells can be chased at late developmental stages to determine whether they are short-lived progenitor cells or function as stem cells that have the capacity to self-renew and differentiate into progeny cells[88]. Using this transgenic lineage assay, RTEL1 has been found to be expressed in the stem cells located in several adult mouse tissues[83,88]. In the intestine, RTEL1-expressing cells are in both +4 intestinal stem cells and crypt base columnar stem cells in intestinal crypts. Both cells can self-renew, proliferate, and generate all types of intestinal epithelial cells required for intestinal homeostasis. In the testis, RTEL1 is expressed in spermatogenic stem cells located along the base membrane of the seminiferous tubule. These stem cells can give rise to more committed chained spermatogonia cells and ultimately mature sperm inside the tubules. In the cerebellum, RTEL1-expressing cells are located in the 4th SVZ and act as cerebellar stem cells required for the formation of cerebellum.

1.4.3 The role of RTEL1 in the maintenance of telomeres and genomic stability

A. The requirement of RTEL1 in the maintenance of telomeres

RTEL1 is required for proper and full mouse development as shown by the embryonic lethality of its complete knockout[83]. *Rtel1*$^{-/-}$ mouse embryos died between E8.5 to E.10.5 with multiple developmental defects in the proliferative tissues in the nervous system, heart, vasculature, and extra-embryonic tissues. Using several chromosomal analysis tools, such as Q-FISH and SKY, RTEL1 deficient cells were found to contain shorter telomeres which lead to a global genomic instability[83].

How RTEL1 is involved in the maintenance of telomeres is still not fully understood. Several studies indicate that RTEL1 could display multiple functions, such as (1) disassembling the T-loop of telomeres[89,90] (Figure 1-6a), (2) resolving G-quadruplex (G4) structures present in telomeres[82,90] (Figure 1-6b), and (3) functioning as a DNA replisome-associated helicase or interacting with Shelterin components, such as TRF2 and POT1, to facilitate telomeric replication[91] (Figure 1-6c).

B. The requirement of RTEL1 in DNA HR

RTEL1 was also independently identified as an antagonist of HR with functional similarity to the yeast anti-recombinase Srs2[92]. Biochemical analyses revealed that RTEL1 exhibits D-loop dissociation activity essential for suppressing DNA crossover during HR[92]. Dysfunction of RTEL1 in HR can lead to overactivity of HR when it is not required resulting in deleterious genomic rearrangements as a result of accumulation of toxic recombination intermediates[92].

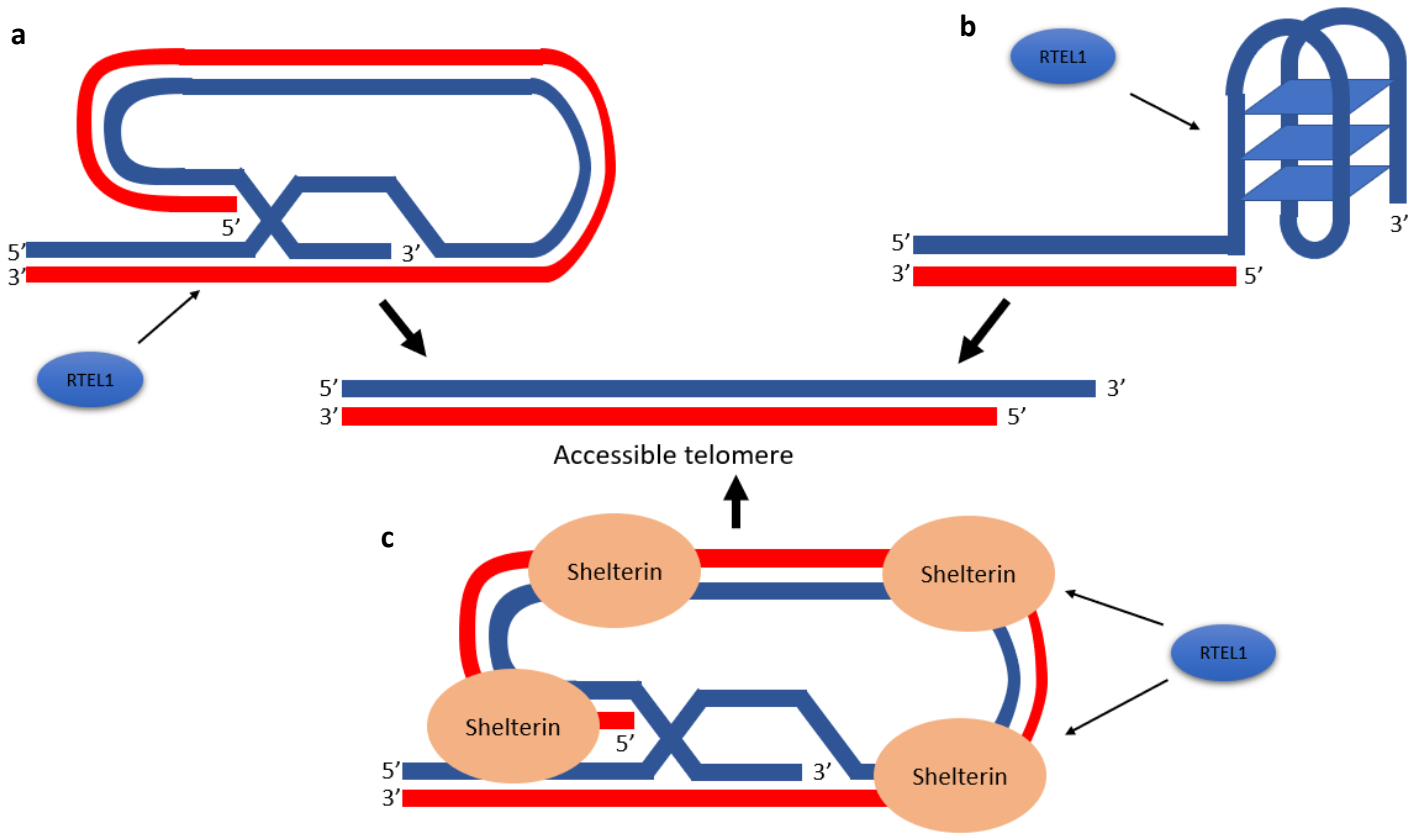


Figure 1-6. The functions of RTEL1 in telomere maintenance. RTEL1 could (a) disassemble T-loop structures to allow protein access to telomere and (b) disassemble G-quadruplex (G4) structures in telomeres, again to allow protein access to telomere, and (c) interact with Shelterin components and/or DNA replication complex to facilitate telomere replication.

C. The requirement of RTEL1 in DNA replication

RTEL1 interacts with PCNA, a core factor in DNA replication[90]. Disruption of RTEL1-PCNA interaction in mice leads to reduced replication fork extension rate, increased replication fork stalling and/or collapse, and telomere dysfunction[90], indicating that the role of RTEL1 in DNA replication is also crucial for the maintenance of genomic stability.

In summary, RTEL1 plays multiple functions, such as resolving T-loop and G4 structures of telomeres, suppressing HR and facilitating DNA replication, that all contributes to maintaining telomere length and genomic stability.

1.4.4 The tumour suppressive role of RTEL1 in MB formation

A. The formation of MB by conditionally knocking out RTEL1 in cerebellar stem cells

Given that RTEL1 is required for genomic maintenance and is expressed in cerebellar stem cells, the major cellular source for MBs, we asked whether RTEL1 is required to protect these stem cells from MB formation. To address this, we crossed the RTEL1 conditional knockout allele (*Rtel1*$^{F/F}$) with GFAP-Cre transgenic mice. GFAP-Cre mice exhibit a robust and specific level of Cre activity in mouse glial cells and neural stem cells including cerebellar stem cells[93]. To prevent any DNA damage-induced apoptosis that occurs in RTEL1-deficient cells, we also introduced a p53 null mutation into these mice, generating *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice. *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice developed MB between 6-12 months old (Figure 1-7a). Since the formation of MB was not found in *Rtel1*$^{F/F}$*/Math1-Cre/p53*$^{F/F}$ mice that deleted RTEL1 in cerebellar external granular layer (EGL) progenitor cells, this result indicates that inactivation of RTEL1 in cerebellar stem cells induce MB formation.

B. The pathological similarity between RTEL1-deficient mouse MB with human MBs

We further applied histological approaches to characterize the pathological and immunohistological characteristics of RTEL1-deficient mouse MBs. This work has been done in collaboration with Dr. Charles Eberhart at Johns Hopkins, who is a renowned pathologist in pediatric brain tumours. We found that RTEL1 deficient mouse MBs show a great similarity to human MBs, with both leptomeningeal spread and parenchymal invasion, including direct

infiltration into the brain stem (Figure 1-7b). In addition, in *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice, we also found that the MBs developed in these mice could metastasize to spinal cord.

(a) (b)

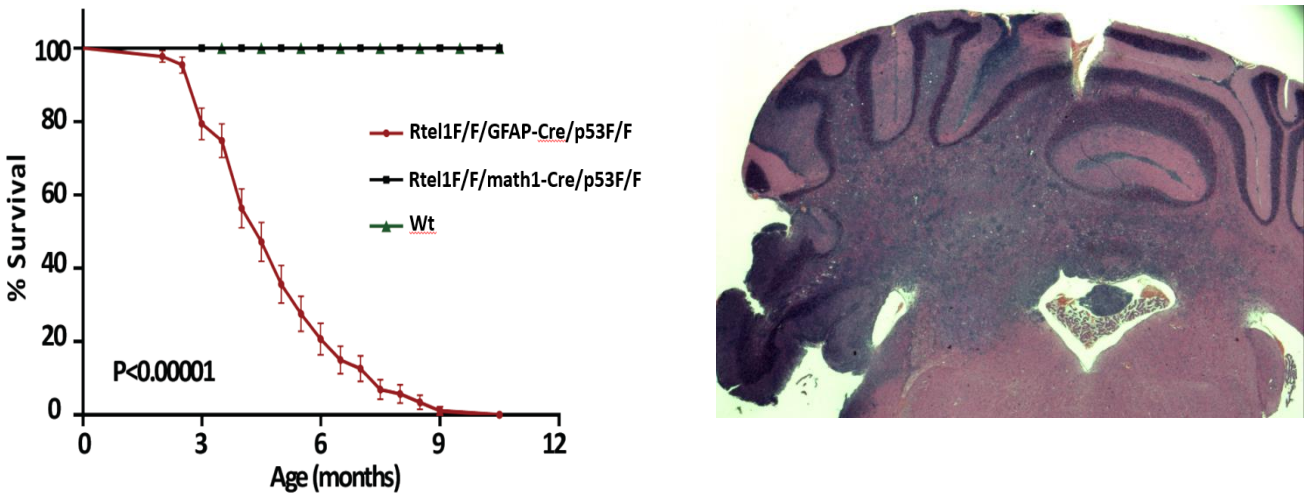


Figure 1-7. The formation of MB in *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice. (a) The Kaplan-Meier survival curve of *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice as compared to wild-type (Wt) and *Rtel1*$^{F/F}$*/Math1-Cre/p53*$^{F/F}$ mice. Only *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ develop MB within 12 months. (b) Hematoxylin and eosin staining show the widespread of MB tumour cells in the cerebellum of *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice.

Taken together, our data indicates the deletion of RTEL1 in cerebellar stem cells can induce a malignant form of MB, which recapitulates many pathological features of human MBs. Further characterizing the genetic alterations in this mouse model, as the main focus of this thesis, could help to better understand the pathogenic function of genomic instability in MB formation.

1.5 The contribution of mouse models in understanding the pathogenesis of MB

1.5.1 The application of mouse models for cancer research

Mouse models have been indispensable in understanding human biology, both physiological and pathological. In total, mice and humans share ~80% of protein coding genes[94]. The mouse model is also advantageous due to similar development with human but with a much shorter life span. In addition, many transgenic tools, such as the recently developed CRISPR/Cas9 based gene-editing approaches, allow generation of mouse models with specific mutations.

On top of this, the mouse genome has been extensively characterized. This data is stored and shared on public databases making downstream bioinformatic analyses easily accessible[95]. Research into the mouse genome has also led to the development of numerous proven genetic manipulation tools that are readily available, gene knockins/outs, mutation induction, and transgenesis just to mention a few[88,96].

In cancer research, transgenic mouse models have provided an *in vivo* model that recapitulates the intricate tumour environment that is difficult to mimic in an *in vitro* model. The tumour micro-environment is crucial for understanding the interactions between tumour and normal cells which are important for tumour survival, growth and metastasis[97]. Some of these interactions, for example, the interaction between tumour cells and immune cells which could aid in tumour growth, are lost in xenograft mouse models generated in immunocompromised mice[98].

Mouse models can also be used to identify potential driver genes by applying gene manipulation methods such as gene knock out and CRISPR/Cas9. As previously mentioned, due to the short life span of mice, the effect of a genetic mutation on a whole life cycle of mouse can be determined. Thus, mouse models of cancer can be used to identify the relevant pathways

involved in the different stages of cancer: initiation, progression, maintenance, and metastasis. Also, mouse models can be used for preclinical trials. In this case, mouse models of brain tumours will be extremely valuable for identifying the therapeutic approaches that can break blood-brain barrier and effectively treat these tumours.

1.5.2 The mouse models for MB

Up to date, several genetically modified mouse models have been created to investigate the pathogenesis of MBs (Table 1-2). The majority of these models were generated based on the genetic alterations in the SHH, WNT, and Myc signaling pathways identified in human MBs. These models have greatly helped not only to understand the pathogenesis of these human MBs, but also to substantially improve the targeted therapies, as evidenced by the inhibition of human MB growth by drugs blocking SHH signaling[99]. However, among these mouse models with MB, only a few models could represent other subgroups of human MB (Table 1-2).

For example, *GTML* mouse model which was generated by inducibly expressing MycN in mouse cerebellum, is the only one that can largely recapitulate the genetic features of human subgroup 4 MBs[100]. Since human subgroup 4 of MBs are characterized by the amplification of MycN, *GTML* model will be useful for studying the pathological mechanisms of this group of MBs.

To determine the role of DNA damage in the pathogenesis of MB, several mouse models have also been generated, which specifically inactivate key factors involved in NHEJ (i.e. DNA ligase 4) and HR pathways (i.e. Brca2 and XRCC2/4) in cerebellar neuronal progenitor cells[101] (Table 1-3). All of these mouse models were found to develop MB with great similarity with human SHH MBs[101]. More recently, these mouse models were also found to display

chromothripsis which frequently leads to complex genomic rearrangements harboring Myc/MycN amplification[79]. Therefore, these mouse models will be valuable to investigate the pathogenic role of genomic instability in MB formation.

As presented in this thesis, we demonstrated that inactivation of RTEL1 in cerebellar stem cells can induce MBs. The majority of these tumours not only contain MycN amplification, but also exhibit the genetic features that are commonly presented in human Group 4 MBs. Therefore, RTEL1-deficient MB mouse model is an additional mouse model to better understand the genomic instability as well as the cell origin in the pathogenesis of MB.

Table 1-2. Genetically engineered mouse models of MB based on identified defective signaling molecules in human MBs. Adapted from Wu et al[102]

Genotype	MB subgroup	Histology	MB incidence	Tumour latency (weeks)	Leptomeningeal metastasis	Reference
Ptc$^{+/-}$	SHH	Desmoplastic	14%	5–25	No	57
Ptc$^{+/-}$ *P53*$^{-/-}$	SHH	Desmoplastic	95%	4–12	No	29
Ptc$^{+/-}$ *Ink4c*$^{-/- \text{ or } +/-}$	SHH	Desmoplastic	30%	12–36	No	30
Ptc $^{+/-}$ *Kip1*$^{-/- \text{ or } +/-}$	SHH	Desmoplastic	60%–70%	16–18	No	55
Ptc$^{+/-}$ *Hic1*$^{-/-}$	SHH	Desmoplastic	~40%	by 25	No	103
Math1-Cre/Ptc$^{C/C}$	SHH	Desmoplastic	100%	8–12	No	104
GFAP-Cre/Ptc$^{C/C}$	SHH	Desmoplastic	100%	3–4	No	104
Sufu$^{+/-}$*/P53*$^{-/-}$	SHH	Desmoplastic	58%	by 28	No	105
Hemizygous ND2-SmoA	SHH	Desmoplastic	48%	25	No	106
Homozygous ND2-SmoA	SHH	Desmoplastic	94%	4–8	Yes	107
Glt1-tTA:TRE-MYCN/Luc (GTML)	Group 4	Classic or LCA	75%	by 29	Yes	100
Blbp-Cre/Ctnnb1$^{+/F(exon3)}$ */p53*$^{F/F}$	Wnt-subgroup		15%	~41	No	27
Sox2+ Myc-infection	Group 3	LCA	60%	By 9	No	108

Table 1-3. Genetically engineered mouse models of MB based on DNA damage repair factors involved in NHEJ and HR pathways. Adapted from Wu et al[102].

Genotype	MB subgroup	Histology	MB incidence	Tumour latency (weeks)	Leptomeningeal metastasis	Reference
Nestin-Cre/Brca2$^{F/F}$/p53$^{-/-}$	SHH	Desmoplastic	>90%	14–16	No	101
Nestin-Cre/Lig4$^{F/F}$/p53$^{-/-}$	SHH	Desmoplastic	>90%	14–16	No	101
Parp1$^{-/-}$/p53$^{-/-}$	SHH	Desmoplastic	49%	8–24	No	109
Nestin-Cre/Xrcc4$^{F/F}$/p53$^{-/-}$	SHH	Desmoplastic	87%	12–14	No	110
Nestin-Cre/Xrcc2$^{F/F}$/p53$^{-/-}$	SHH	Desmoplastic	>90%	14–16	No	101
GFAP-Cre/Rb$^{F/F}$/p53$^{F/F}$	SHH	Desmoplastic	>84%	12	No	111
Lig4$^{-/-}$/p53$^{-/-}$	SHH	Desmoplastic	100%	3–9	No	112

1.6 Rationale and Hypothesis

RTEL1 is a DNA helicase that is essential for the maintenance of genome stability and telomere integrity[83,89,90]. Conditional knockout of RTEL1 in the developing mouse cerebellum, specifically at cells located in the 4th ventricular zone, can lead to the formation of MBs. However, how the genomic alterations caused by RTEL1 deficiency leads to tumorigenesis is largely unknown.

The main aim of my thesis is to apply combined genomic and bioinformatic approaches to characterize the genetic features of RTEL1-deficient mouse MB model in order to better understand the pathogenic mechanism of MB. More specifically, in this study, I will: (1) use gene-expression analysis to classify which subgroups of human MB that RTEL1-deficient MB

mouse model could belong to; (2) apply several genome-wide approaches to identify the recurrent genetic alterations in RTEL1-deficient mouse MBs; (3) apply comparative genomic hybridization (CGH) approach to characterize genomic rearrangements in RTEL1-deficient mouse MBs to determine whether RTEL1 deficiency could induce chromothripsis as seen in human MBs.

Chapter 2: Materials and Methods

2.1 Mouse breeding

Mice were used in this thesis according to the Principles in the Current Guide to the Care and Use of Experimental Animals published by the Canadian Council of Animal Care (CCAC) and other guidelines published by the CCAC which were endorsed by the University of Manitoba.

To generate *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice, we first bred *Rtel1*$^{F/F}$ mice (established in our lab[89]) with *GFAP-Cre* mice (obtained from JAX lab) to produce *Rtel1*$^{F/+}$*/GFAP-Cre* mice. These mice were then bred with *p53*$^{F/F}$ mice (obtained from JAX lab) to form *Rtel1*$^{F/+}$*/GFAP-Cre/p53*$^{F/+}$ mice that were further intercrossed to generate *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice. The *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice were postnatally viable and indistinguishable from wild-type controls. However, starting two months old, some of the *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice started to develop MB with ataxia and unbalanced walk phenotypes. The majority of *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice formed MB around 6 months old.

2.2 Genotyping of *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice

To genotype *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice, PCR was performed using ear-punched DNA. Briefly, ear-punched tissues from two weeks old mice were incubated overnight at 55^{o}C with 100 μl lysis buffer containing 50 mM KCl, 100 mM Tris-HCl (pH8.3), 2 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.45% NP40, 0.45% Tween-20 and 100 μg/ml Proteinase K (Gibco, #2553001). 1 μl lysate was added with 16.5 μl dH_2O and then heated at 95^{o}C for 10 min to inactivate Proteinase K. This mixture was then mixed with 6.5 μl PCR reaction cocktail containing 1xPCR buffer, 25 mM MgCl2, 10 mM dNTPs, 100μg PCR primers (forward and reverse) and 1 unit Taq

polymerase for PCR which consisted with 35 cycles. The following primers were used to genotyping *Rtel1$^{F/F}$/GFAP-Cre/p53$^{F/F}$* mice:

To detect *Rtel1$^{F/F}$* allele: forward primer (5'-aggtaggctctgccattgtg-3'); reverse primer (5'-ggaggtggagtgaagcagag-3'). The PCR products for wild-type and floxed alleles are 174 bp and 415 bp, respectively.

To detect GFAP-Cre allele: forward primer (5'-actccttcataaagccct-3'); reverse primer (5'-atcactcgttgcatcgaccg-3'). The PCR product for this allele is 190 bp.

To detect GFAP-Cre allele: forward primer (5'-ggttaaacccagcttgacca-3'); reverse primer (5'-ggaggcagagacagttggag-3'). The PCR products for wild-type and knock-in alleles are 270 bp and 390 bp, respectively.

2.3 Collection of MB tissues from *Rtel1$^{F/F}$/GFAP-Cre/p53$^{F/F}$* mice

Once *Rtel1$^{F/F}$/GFAP-Cre/p53$^{F/F}$* mice showed ataxia and unbalanced walk, these mice were intraperitoneally injected with BrdU (5 mg total to label proliferating tumour cells). Two hours later, the mice with BrdU injection were sacrificed by cervical dislocation, and MBs formed in these mice were dissected. A portion of tumour tissue was collected for RNA and DNA preparation, while the remaining tumour tissue were immediately fixed with formalin for histological analysis.

2.4 Purification of genomic DNA from RTEL1-deficient mouse MBs

Qiagen genomic DNA preparation kit (Qiagen, #13343) was used to isolate genomic DNA from MBs collected from *Rtel1$^{F/F}$/GFAP-Cre/p53$^{F/F}$* mice according to the manufacturer's instruction. Briefly, around 100 mg tissue samples were homogenized with a homogenizer and

lysed with Proteinase K solution (50^{o}C for 2 hours). The samples were then loaded onto the Qiagen Genomic-tip which were washed three times with buffers that allowed to remove all contaminates. The genomic DNA bound to Qiagen Genomic-tip was then eluted with TE buffer. The amount and the quality of isolated genomic DNA were determined by spectrophotometer (NanoDrop, #ND-1000).

To prepare the control DNA for CGH, we also used the same method to isolate genomic DNA from the kidney of the same mouse which developed MB.

2.5 Purification of total RNA from RTEL1-deficient mouse MBs and tumour-free cerebellum

RNAeasy plus kit (Qiagen, #74134) was applied to purify total RNA from MB and normal cerebellum tissues collected from *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice according to the manufacturer's instruction. Briefly, in this procedure, around 10 mg tissues were homogenized with a homogenizer and the lysates were transferred to the genomic Eliminator spin column to remove genomic DNA. After removing genomic DNA, the samples were loaded onto RNAeasy spin column. The column was washed three times to remove the contaminates and the total RNA bound to the column was eluted with TE buffer. The amount and the quality of isolated total RNA were determined by spectrophotometer (NanoDrop, #ND-1000).

2.6 Microarray-based expression analysis of RNA from RTEL1-deficient mouse MBs

Total RNA prepared from 16 RTEL1-deficient mouse MBs and 3 normal cerebella of *Rtel1*$^{F/F}$*/GFAP-Cre/p53*$^{F/F}$ mice were hybridized on Mouse Gene 1.0 ST array chip by The Center for Applied Genomics (TCAG). The data was stored in binary CEL files and contained the

expression intensities of the transcript IDs used in Mouse Gene 1.0 ST array chip. The background correction, log transformation, and normalization of expression measurements were done using RMA normalization[113] in the "oligo" package[114] in R.

The normalized data was then analyzed for differential expression using the "limma" package[115] in R according to the package author's instructions. In brief, limma first fits the data into a generalized linear model which is then used to calculate a moderated t-statistic using empirical Bayes method. The moderated t-statistic was calculated on transcript ID-level data by contrasting tumour expression to control expression. Obtained p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method[116]. Transcripts with adjusted p-values < 0.05 and a fold change larger than 2 ($logFC > 1$) were considered significant. The significantly differentially expressed transcripts were then summarized to gene level with each significant gene being represented by only one transcript for interpretation.

The genes found to be differentially expressed were used as input for gene set enrichment using GSEA[117], a BROAD institute software for enrichment analysis. LogFC values from the differential expression analysis were used as the weights for the genes in the gene set enrichment analysis.

2.7 Collection and merging of RTEL1-deficient MB and human MB microarray expression data

RNA microarray data obtained from both RTEL1-deficient mouse MBs (n = 16) and human MBs were used to determine which subgroup of MBs that RTEL1-deficient mouse MB could represent. The microarray data of human MB was taken from Ramaswamy et al[36] (n = 763) and includes subgroup information (70 WNT MBs, 223 SHH MBs, 144 Group 3 MBs, 326

Group 4 MBs). In this analysis, data collected from the same platform, i.e. RNA analyzed on the same microarray chips, were loaded together into the R environment and RMA normalized. Each platform was annotated to contain gene expression levels instead of probe expression. Gene-level expression was defined as the mean expression of all probes present in the gene. The metagene projection algorithm[118] was then used to determine metagenes, which are positive linear combinations of genes obtained from nonnegative matrix factorization, from the human gene-expression profiles that could describe the biological differences between the human MB subgroups. In this case, since there are 4 human MB subgroups, 4 metagenes were derived. Only genes with a fold change greater than 5 were included in metagene calculation. A threshold of 20 and a ceiling of 1000 applied to the gene expression values before metagene calculation.

After determining the metagenes, the metagene projection algorithm was then applied to the gene-expression profiles of MB mouse models to obtain the 4 metagenes of the mouse model so that cross-platform and cross-species comparison can be performed on these metagenes. All metagene values, both human and mouse models, were normalized together before downstream analyses.

Mouse models included in this study are GTML mouse model[100] (n = 32, Group 3), SMOM2 mouse model[119] (n = 21, SHH), WNT mouse model[27] (n = 3, WNT).

2.8 Unsupervised classification of RTEL1-deficient mouse MBs

t-SNE was applied onto the metagenes of both human and mouse models using the R package "Rtsne"[120]. In brief, the input metagenes were first centered to mean of 0 and scaled to standard deviation of 1. After this standardization a distance matrix is calculated using euclidean distances. Before implementing t-SNE, this package allows for a preliminary principal

component analysis (PCA) to begin feature reduction and help alleviate the curse of dimensionality. Finally, the t-SNE was run at 3 000 iterations, a perplexity of 50, and set to output data with 2 dimensions. The perplexity was set at 50 by testing which perplexity is able to clearly separate only the human MB expression data into their known subgroups.

The metagene projection algorithm also includes PCA capability which was applied to the normalized metagenes of both human and mouse models.

2.9 Supervised classification of RTEL1-deficient mouse MBs

The metagene projection algorithm[118] had a built in SVM functionality. SVM multi-class classification was done using class to class logic resulting in 6 classifiers being trained (k(k-1)/2 classifiers, where k is the number of classes) with the 4 metagene expression values as input. The probabilities are therefore the averages of the 6 different class to class tests (NOTE: Probability measures on SVM are tacked on after the SVM calculations. A radial SVM kernel was used with the following parameters: Cost = 1, gamma = 0.05. SVM does not normally calculate probabilities). Subgroup labels were assigned using the highest probability, without the need of meeting a threshold.

Being a multi-class problem, a receiver operating characteristic curve (ROC) could not be made for this SVM. To calculate AUC, Hand and Till suggested calculation of separate ROC curves for each classifier whose average is then the AUC for the multi-class classifier[121].

2.10 Sequencing of RNA from RTEL1-deficient mouse MBs

Total RNA isolated from 17 RTEL1-deficient MBs and 3 normal controls were sent to TCAG for RNA-sequencing. Quality control and read alignment were also done by TCAG using TopHat2[122] software, from which we received the pre-processed BAM files.

Read counts were extracted from the BAM files using Partek® Genomics Suite® software, version 7.0 Copyright ©; 2019 Partek Inc., St Louis, MO, USA. As TCAG already filtered mRNAs that had no read counts in any of the samples, no further filtering was done. The remaining read counts were normalized via the trimmed-means of M-values (TMM) method[123] in the "edgeR" package[124] in R. Differential expression analysis was done using the "limma" package[115]. For RNA-seq, the "limma" package uses a negative binomial generalized linear model since read counts are discrete data. P-values for each transcript were then calculated using empirical Bayes quasilikelihood F-tests. P-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method[116]. The transcripts were annotated to genes for ease of interpretation. Genes with adjusted p-values < 0.05 and a fold change larger than 2 ($\log FC > 1$) were considered significant.

For variant calling, we followed the proposed short variant discovery pipeline by GATK Best Practices[125]. In brief, the RNA-seq BAM files were realigned to the mm10 genome using STAR software[126]. After realignment, new quality control measures were taken by recalculating base quality values using the new alignment. Lastly, variant calling was done using the HaplotypeCaller method from the GATK software[125]. Mutations that did not have a PASS filter score were excluded from downstream analyses. For the recurrent homozygous analyses, silent mutations and any mutations occurring in noncoding regions, such as splice site or on noncoding RNAs, were excluded. Here, recurrent is defined as variants occurring in at least 8 tumour samples,

as determined by graphing number of tumour samples affected against variant frequency (Figure 2-1). The noncoding and silent mutations were however included in the 20/20+ algorithm implementation to account for background mutation rates.

To identify fusion transcripts, the BAM files were again realigned using STAR software[126]. This alignment is meant to better track the relative orientation of read mapping and allows for pipelining to STARFusion software[127], which was used to detect fusion transcripts.

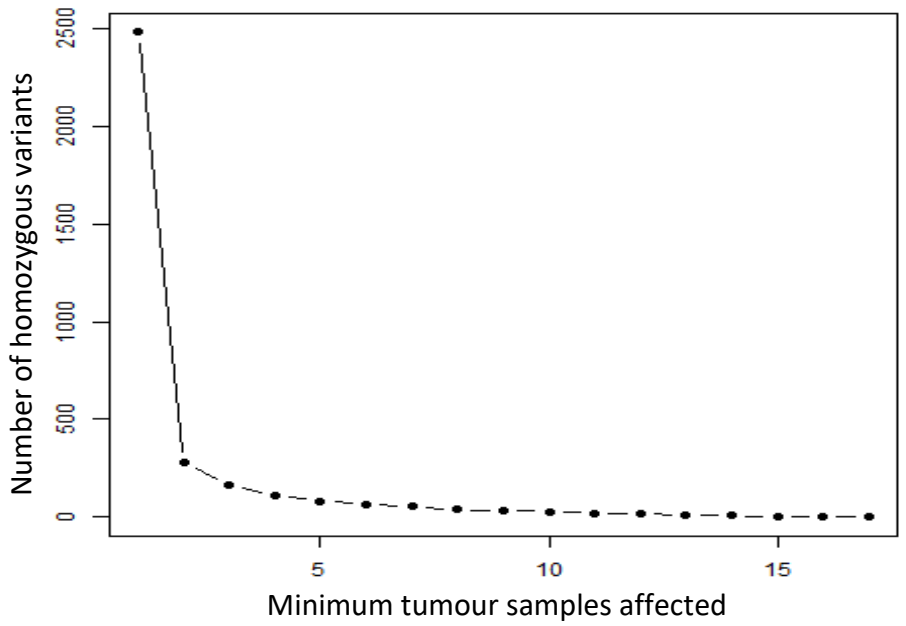


Figure 2-1. Cutoff determination for RNA-seq recurrent homozygous variant analysis. Graph of number of variants against the minimum number of tumour samples chosen as a cut off for recurrence analysis. Drastic change in number of homozygous variants can be seen at 5 samples affected from the graph, however looking at actual values another large change in recurrent variant number occurs from a minimum of 7 to 8 tumour samples.

2.11 Whole exome sequencing of RTEL1-deficient mouse MBs

Genomic DNA from 11 RTEL1-deficient mouse MBs and 3 controls (collected from the kidney of RTEL1-deficient mouse MB) was sent to TCAG for exome selection and whole exome sequencing. Quality control, read alignment, and variant calling were done by TCAG using

GATK Best Practices[128]. In brief, low quality reads, according to FastQC (Andrews S. 2010, FastQC: a quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), were removed from analysis. High-quality reads were aligned to the mm10 genome using BWA software[129]. Variants were called on both the tumour and control samples using Mutect2, a function in the GATK software[125]. Variants occurring in both tumour and control samples were excluded from the analysis. The resulting vcf files were annotated using ANNOVAR[130]. To search for tumorigenic mutations, we looked at recurrent homozygous mutations. In this thesis, recurrent is defined as variants occurring in at least 5 tumour samples as determined by graphing tumour sample number against variant frequency (Figure 2-2). For this analysis, silent mutations and mutations in noncoding regions were excluded from analysis.

To identify driver mutations, we also applied the 20/20+ algorithm[131]. Silent mutations and noncoding regions were included in the 20/20+ calculations as to account for background mutations.

2.12 CGH analysis of RTEL1-deficient mouse MBs

23 genomic DNA samples isolated from RTEL1-deficient mouse MBs were sent to TCAG for CGH analysis on Agilent oligonucleotide array chips. Copy number alterations (CNAs) were identified from the Log_2 ratios of the CGH analysis by the genomic segmentation algorithm on the Partek® Genomics Suite® software, version 7.0 Copyright ©; 2019 Partek Inc., St Louis, MO, USA. In brief, this algorithm begins with the mean log_2 ratio of a DNA segment from one end of a chromosome, which must contain a minimum of 6 probes, that is compared to the mean

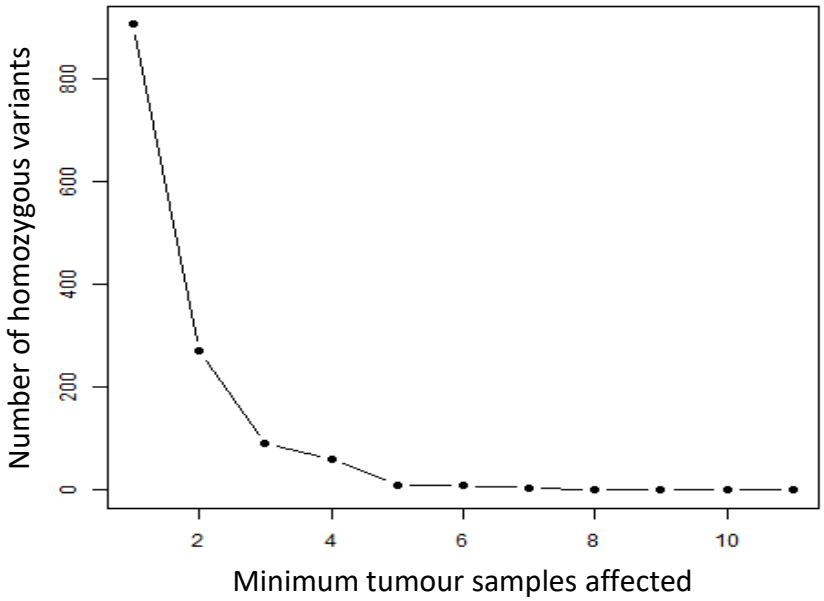


Figure 2-2. Cutoff determination for WES recurrent homozygous variant analysis. Graph of number of homozygous variants against the minimum number of tumour samples chosen as a cut off for recurrence analysis. Drastic change in number of homozygous variants can be seen at 5 samples affected from the graph.

$\log_2$ ratio of the adjacent segment. Significant differences are identified as differences in mean between the two segments larger than 0.3, to account for noise, with p-value < 0.001.

To identify recurrent genetic alterations in these CGH data, we applied the graph-based Maximal Clique algorithm[132,133]. For this analysis, an interval graph was created where each CNA was defined as a vertex and overlaps between CNAs were indicated by edges connecting the representing vertices. CNAs smaller than 5 kb were excluded. Identification of maximal cliques with these definitions not only allows identification of recurrent regions, it also allows detection of the innermost region of the CNA overlap which can further narrow down the likely driver mutations. Inner CNA regions that are either smaller than 5 kb, does not occur in at least 5 samples, or did not include any genes (putative or otherwise) were excluded from further analysis. Regions were annotated using the mm10 genome build.

To determine the genomic rearrangements that could resemble that of chromothripsis in RTEL1-deficient mouse MBs, we graphed $\log_2$ ratios from CGH analysis onto genomic coordinates. In this process, the $\log_2$ ratios of each probe, extracted from the CEL files of CGH array, were mapped to the middle of the probe and graphed as a dot plot along each chromosome. These images were manually checked for signs of chromothripsis using the first three criteria described by Korbel and Campbell[77]: (1) Clustering of DNA double stranded breakpoints, here defined as a single chromosome having at least 10 adjacent DNA breakpoints (2) Oscillation between the copy number states within the breakpoint cluster (3) Low number of copy number states, here defined as a maximum of 3 copy number states.

Chapter 3: Results

3.1 To use gene-expression analysis to classify RTEL1-deficient mouse MB model

3.1.1 Rationale

Specific knockout of RTEL1 in mouse cerebellar stem cells can induce a malignant form of MB, which recapitulates many pathological features of human MBs (Figure 1-7). However, whether this RTEL1-deficient MB mouse model can exhibit similar genetic alterations as seen in human MBs is still unknown. Recently, mainly based on gene-expression profiles, human MBs have been divided into four main subgroups with distinct genetic features (Table 1-1). This progress has greatly advanced our understanding of human MB formation. If RTEL1-deficient mouse MB model shares a similar genetic profile as one of these subgroups, it can serve as a genetic tool to further investigate the pathogenic mechanisms underlying this subgroup of human MB formation. Because of this, in this part of the study, I applied both unsupervised and supervised bioinformatic tools to determine which subgroup of human MB RTEL1-deficient MB mouse model simulates. This work should enable us to demonstrate whether RTEL1-deficient mouse MB model is a valuable tool for understanding the formation of human MBs.

3.1.2 Results

3.1.2.1 Classification of RTEL1-deficient mouse MB model based on t-stochastic neighbor embedding and principal component analysis

To achieve statistical significance for this part of the study, I included microarray-based RNA expression profiles from 16 RTEL1-deficient mouse MBs and 763 human MBs that cover four subgroups as described previously[36]. From these two datasets, gene-expression profiles were generated containing the expression level of thousands of genes. These profiles can be

considered high-dimensional (data with greater than 10 dimensions) data because each gene can be considered a dimension since it can have varying levels of expression. High-dimensional data is impossible to visualize thus cannot be used to infer similarity between the varying gene-expression profiles. Therefore, to classify RTEL1-deficient mouse MB model as one of the subgroups of human MB, I applied the t-stochastic neighbor embedding (t-SNE) technique, an unsupervised nonlinear dimensionality reduction method that allows to visualize relationships between high-dimensional data by reducing the dimensions. In brief, t-SNE uses the normalized microarray measured RNA expression data to calculate a similarity value between all possible pairs of gene-expression profiles. Highly similar pairs of profiles will be given higher similarity values, hereafter termed pairwise similarity probability. t-SNE then randomly graphs these same profiles as points on a separate, lower dimensional space, i.e. in 2- or 3-dimensions. New similarity values are calculated for the points in the lower dimensional graph that are then compared to the original pairwise similarity probability. t-SNE repeats the creation of a lower dimensional graph until the similarity values from the lower dimensional graph matches the pairwise similarity probability from the original high-dimensional data (gene-expression profiles). By doing so, we now have data that can be easily visualized in 2D or 3D where the distances between the sample points is a representation of the similarity between their gene-expression profiles. This graph can now be used to determine whether RTEL1-deficient mouse MB model simulates the gene-expression profiles of human MBs and which subgroup it most closely resembles by seeing where it is placed on the graph in relation to the human MB samples. To account for the inherent differences between mouse and human gene expression, the metagenes calculated from the metagene projection algorithm are used as input for t-SNE and PCA. It is important to note that t-SNE and PCA are not a clustering algorithm. Neither will

assign subgroup membership to any of the data points. It is simply used to visualize our data to identify possible data clusters. Thus, if RTEL1-deficient mouse MB model truly mimics one subgroup of human MB then we expect the data points of this model to merge with this group of human MB data points.

t-SNE has been successfully applied to separate a cohort of 1 256 human MBs into four subgroups based on DNA methylation profiles[22]. Here, we believe the t-SNE algorithm can also be applied to genome-wide expression profiles to separate human MBs into the four subgroups and test which subgroup RTEL1-deficient MB mouse model most closely recapitulates.

Using t-SNE method, the majority of RTEL1-deficient mouse MB were found to have similar gene-expression profile as human SHH MB (Figure 3-1). For t-SNE analysis, only one mouse model was compared to human at a time as because when more than 1 mouse model was compared to human, the similarities due to mouse expression overpower the subgroup differences in expression and all mouse models begin clustering together. The PCA is able to ignore these similarities as it does not rely on maintaining the pairwise similarity probability. In the PCA images, the gene-expression profiles of RTEL1-deficient mouse MBs were found to be largely overlapped with the ones in SMOM2 mouse MB model, which has been characterized as SHH, and close to the human SHH MBs (Figure 3-2). All of these indicate that RTEL1-deficient mouse MB gene-expression could mimic human SHH MB.

3.1.2.2 Classification of RTEL1-deficient mouse MB based on support vector machine method

The second approach for this part of the study is to use a supervised machine learning model, termed support vector machine (SVM), which uses data with known subgroup

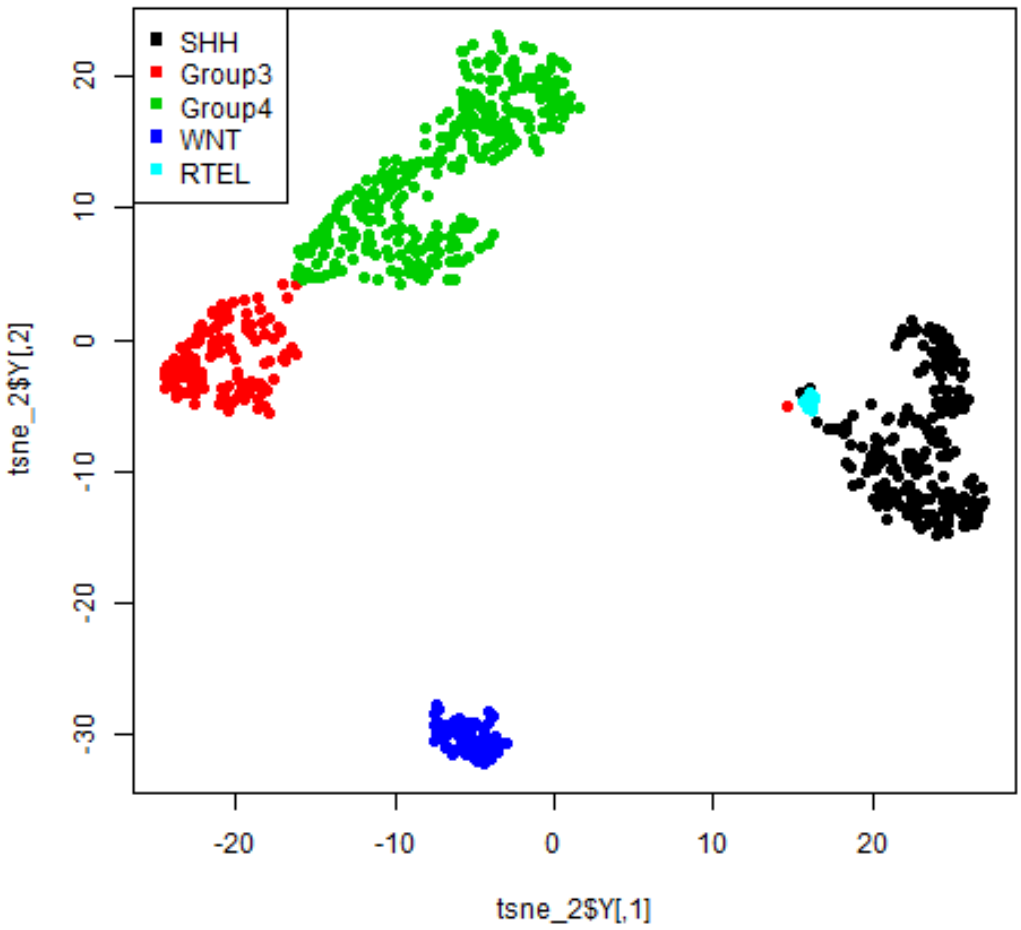


Figure 3-1. Classification of RTEL1 deficient mouse MB based on t-SNE method. Each dot represents the metagene-expression profile of a sample. In this analysis, microarray-based gene-expression profiles of 16 RTEL1-deficient mouse MBs and 763 human MBs were included. Metagenes based only on orthologous genes between mouse and human were used to compare expressions.

membership to determine differences in gene expression between the subgroups so that it may assign subgroup membership for unlabeled data. This is achieved by placing the metagene-expression profiles, consisting of the metagenes, in a high-dimensional space to determine a hyperplane that can separate the data points into their respective classes, in this case the four subgroups of human MB. SVM remembers the high-dimensional space and the hyperplane so that new data given to the SVM model are mapped to the high-dimensional and the hyperplane is

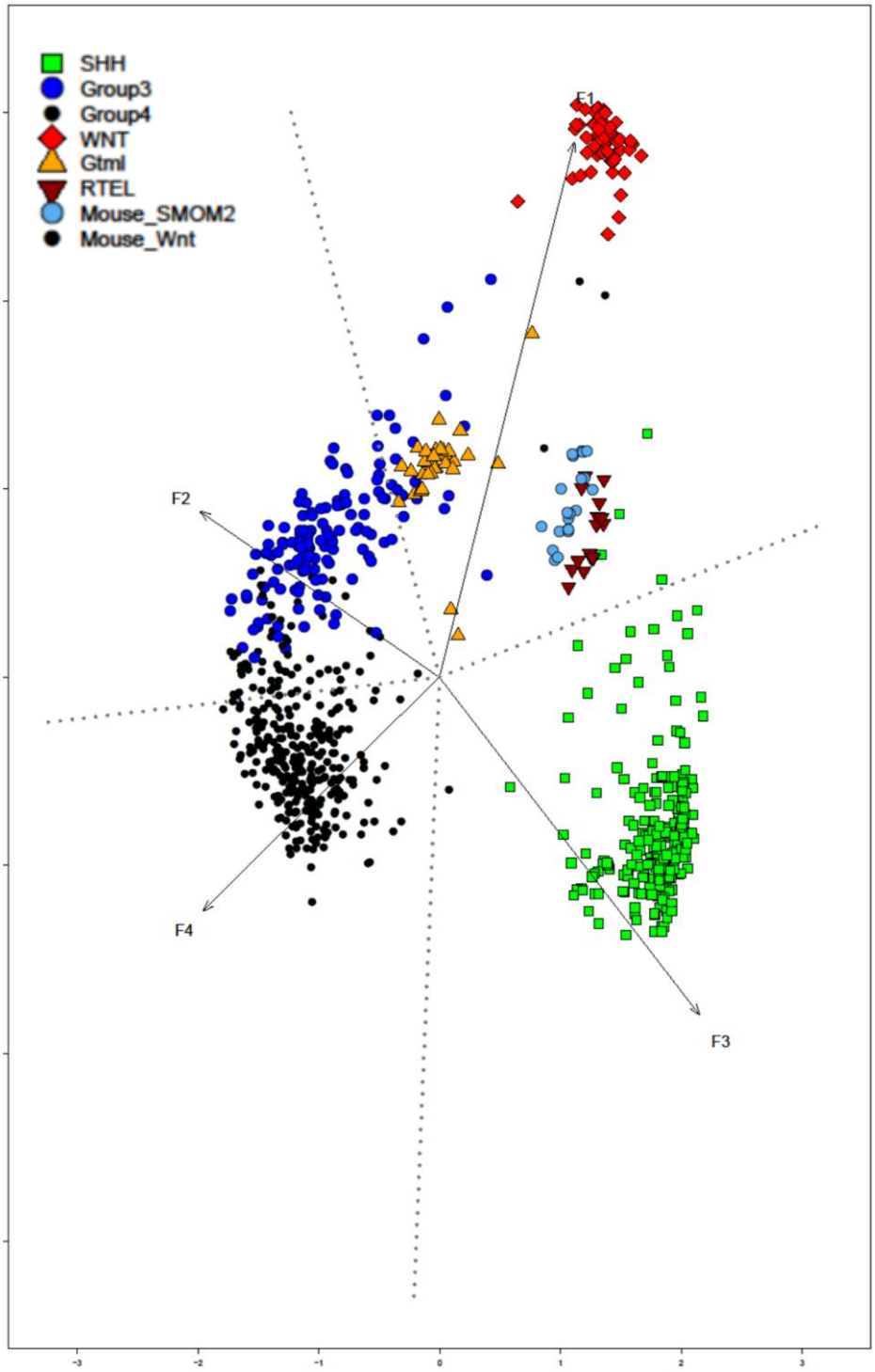


Figure 3-2. Classification of RTEL1-deficient mouse MB based on principal component analysis. Each dot represents the metagene-expression profile of a sample. In this analysis, microarray-based metagene-expression profiles of 16 RTEL1-deficient mouse MBs, 32 GTML mice (Group 3 mouse model), 21 mouse SMOM2 models (SHH mouse model), 3 mouse WNT models, and 763 human MBs were included. Metagenes based only on orthologous genes between mouse and human were used to compare expressions.

used as the reference to determine class membership. This technology has been previously applied to classify new human MB samples into the four subgroups[36,134,135].

The metagene-expression profiles of 536 human MB samples were provided to the SVM for training. The remaining 227 human MB samples were used to test the efficacy of this SVM model. We measured the accuracy (ACC) and the area under the receiver operating characteristic curve (AUROC) of the SVM on the test set. Accuracy is measured as correctly classified test set samples / total test set samples. The receiver operating characteristic curve (ROC) is a graph of the true positive rate against the false positive rate at different significance threshold (from 0 to 1). The AUROC then measures how well the model differentiates true positives from false positives, where an AUROC = 1 means the model is able to accurately predict subgroup classification with no false-positives or false-negatives. Our trained SVM model successfully classified the test set samples with an AUROC of 0.9692 and an ACC of 0.9453 (Table 3-1).

The trained SVM model was then applied to RTEL1-deficient mouse MBs to determine which subgroup this mouse model could belong to. As shown in Figure 3-3 and Table 3-2, all RTEL1-deficient mouse MBs were classified as SHH MBs, with overlapping similarities with Mouse SMOM2 model of SHH MB. SVM analysis ably classified remaining mouse models into their respective MB subgroups indicating that this method is a reliable tool for classifying human or mouse MB.

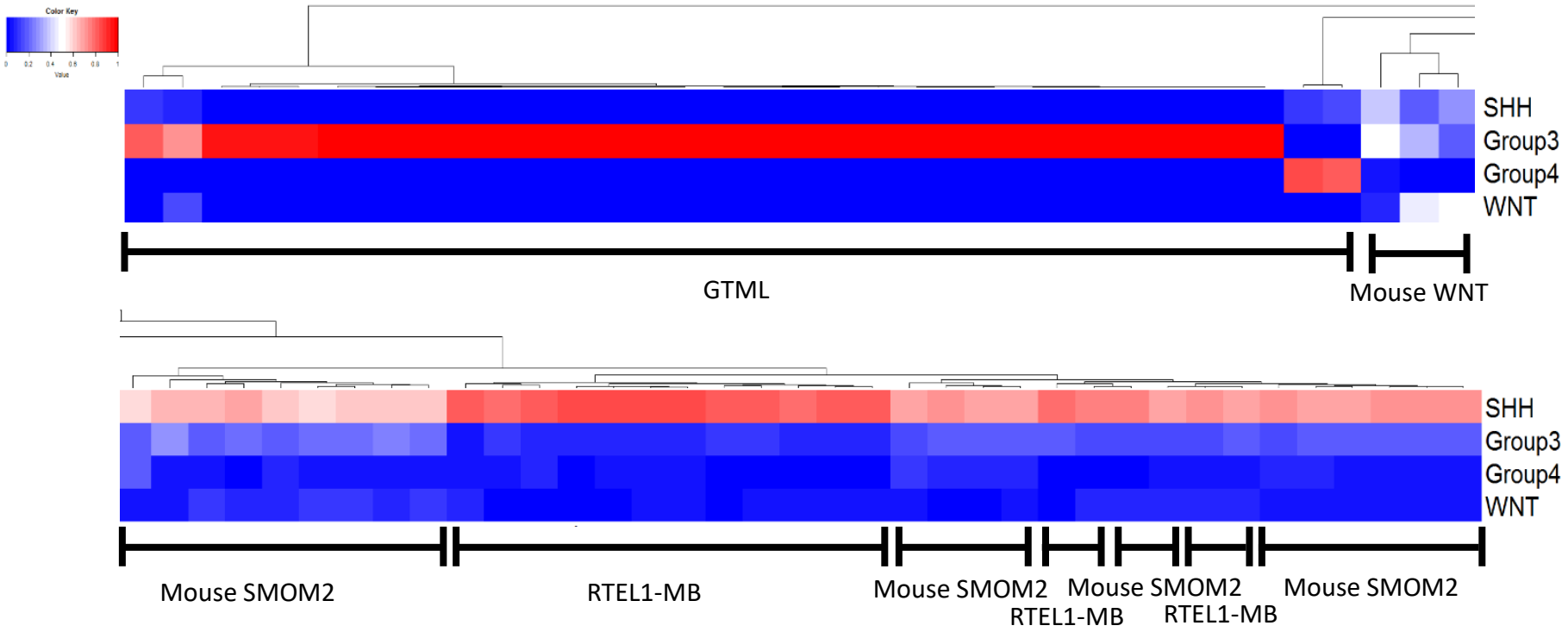


Figure 3-3. Heatmap of the classification of RTEL1-deficient mouse MBs based on support vector machine (SVM) method. RTEL1-deficient mouse MB (RTEL1-MB) are classified into SHH MBs along with the SMOM2 SHH MB mouse model. GTML was correctly classified into Group 3 as well as Mouse WNT mouse model being classified into the WNT MB subgroup.

Table 3-1. Performance measurements of support vector machine (SVM) on a test cohort of 227 human MBs.

Actual / Prediction	SHH	WNT	Group 3	Group 4
SHH	66	1	1	0
WNT	0	19	0	0
Group 3	0	1	42	4
Group 4	0	0	0	93
Accuracy	0.9692			
AUROC	0.967			

Table 3-2. RTEL1-deficient mouse MB subgroup probability assigned by SVM multi-class classification.

Samples	Group3	Group4	SHH	WNT
HD22	0.092	0.03	0.85	0.03
HD23	0.16	0.047	0.71	0.078
HD24	0.13	0.016	0.82	0.04
HD9	0.1	0.032	0.8	0.065
HD10	0.08	0.041	0.84	0.034
HD12	0.061	0.056	0.81	0.068
HD3	0.17	0.041	0.7	0.089
HD4	0.1	0.029	0.82	0.047
HD13	0.089	0.031	0.83	0.055
HD14	0.11	0.026	0.83	0.03

HD16	0.15	0.02	0.75	0.074
HD17	0.17	0.033	0.77	0.023
HD18	0.08	0.084	0.81	0.026
HD19	0.088	0.047	0.84	0.028
HD20	0.077	0.041	0.85	0.034
HD21	0.11	0.052	0.8	0.032

3.1.3 Summary

In this part of the study, I used both supervised and unsupervised bioinformatic approaches to demonstrate that the majority of RTEL1-deficient mouse MBs can recapitulate the genetic features seen in human SHH MBs. Since human MBs are characterized by genomic instability, which is consistent with the dysfunctional role of RTEL1-deficiency in the genomic maintenance, my finding indicates that RTEL1-deficient mouse MBs can be a valuable tool to study the pathological function of genomic instability in human MBs.

3.2 To characterize the recurrent genetic alterations in RTEL1-deficient mouse MBs

3.2.1 Rationale

Based on the essential role of RTEL1 in the maintenance of telomeres and genomic stability (Figure 1-6), we predict that RTEL1-deficiency in murine cerebellar stem cells will induce genetic alterations that can transform these stem cells to malignant cancer cells forming MBs that are similar to human SHH MBs. Therefore, characterization of genetic alterations in RTEL1-deficient mouse MBs has the potential to identify the genes that have specific tumorigenic functions in the formation of MB. This will greatly help us better understand the pathogenesis of MB and/or identify new therapeutic targets for treatment. Based on this rationale, in this part of the study, I applied several genome-wide approaches to characterize the genetic alterations in RTEL1-deficient mouse MBs.

The first approach is to use DNA array based CGH to characterize the recurrent copy number alterations (CNAs) in RTEL1-deficient mouse MBs. This is because the genes that could have important roles in tumour formation are more likely to be affected by the CNAs presented in most tumour samples. Therefore, identification of recurrent CNAs in RTEL1-deficient mouse MBs could be helpful to search for the driver genes in this mouse model. In order to increase the accuracy on detecting tumour-specific, recurrent CNAs in RTEL1-deficient mouse MBs, I used matching DNA samples prepared from the kidney of the same mouse with MB formation as the reference in this CGH study. Through a competitive fluorescence hybridization on a DNA array (Figure 3-4), we are able to more precisely measure the copy number changes in RTEL1-deficient mouse MBs. In addition, the CGH method applied in this study also allowed for a locus-by-locus measure of CNAs with a high resolution (~100 kilobases).

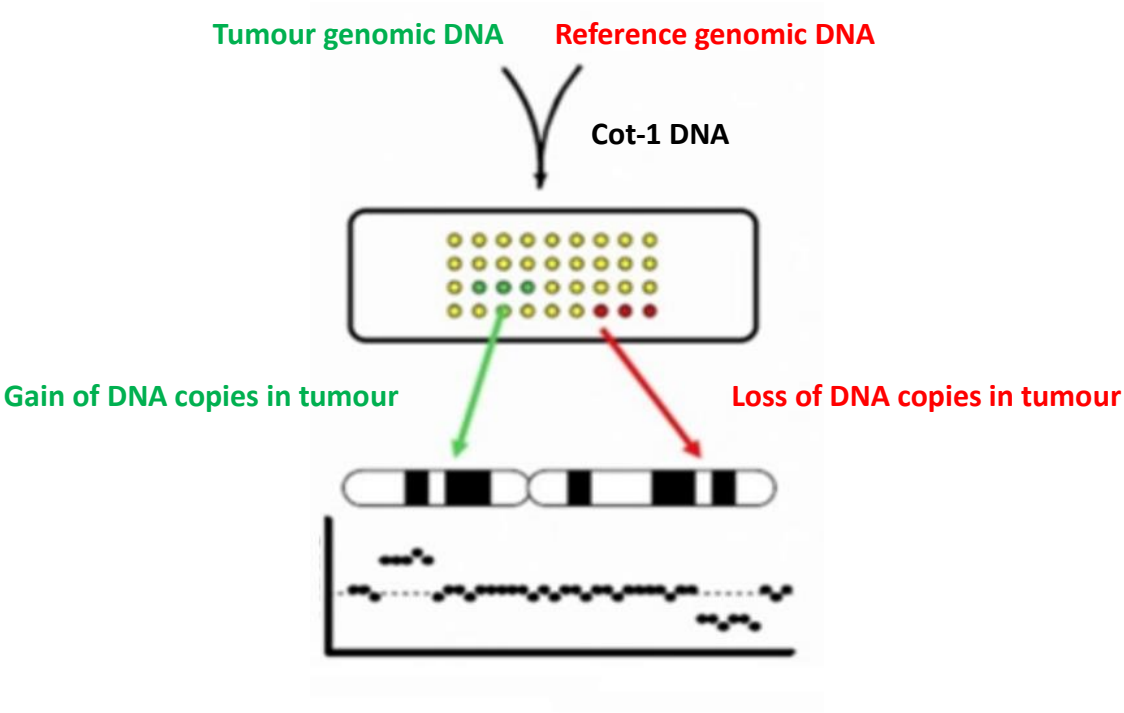


Figure 3-4. The principle of using array-based CGH to detect copy number alterations in tumour genomic DNA. Cot-1 DNA is used to block nonspecific hybridization.

The second approach used to characterize recurrent genetic alterations is next-generation sequencing: both whole exome sequencing (WES) and RNA sequencing (RNA-seq). These assays have been demonstrated as the most sensitive and reliable tools for identifying driver mutations in mouse genetic screenings, rare human genetic diseases and cancer[136–138]. In addition, the cost-effectiveness of these assays, together with available analysis software tools, allows an unbiased screen of driver mutations in multiple biological samples within a short time period. Based on this, I analyzed the WES and RNA-seq data collected from 10 and 17 RTEL1-deficient mouse MBs, respectively, to identify the recurrent mutations in these tumours

3.2.2 Results

3.2.2.1 Identification of recurrent genetic alterations in RTEL1-deficient mouse MBs by CGH analysis

To identify recurrent genetic alterations in RTEL1-deficient mouse MBs, I first used an array based CGH to detect CNAs (24 tumours analyzed). These CGH data were further analyzed by the maximal clique algorithm to identify the most recurrently affected genes presented in CNAs. The maximal clique algorithm is a graph-based algorithm that first maps CNAs onto an interval graph, as vertices, based on chromosomal locations (Figure 3-4a). Whenever CNAs have an overlapped region, the vertices on the graph which represent the overlapping CNAs are connected via an edge. Finally, this algorithm uses the interval graph to identify cliques that contain the largest possible number of interconnected vertices, i.e. maximal cliques. (Figure 3-5b). Therefore, using the maximal cliques algorithm, we can more precisely locate the start (inner start) and end (inner end) of recurrently affected regions (Figure 3-5a)[132,133].

Using this bioinformatic tool, several recurrent regions were found to be amplified in more than 60% of RTEL1-deficient mouse MBs (Table 3-3). Around 20 genes were found to have lower copy number in more than 90% of RTEL1-deficient mouse MBs (Table 3-4). The high frequency of CNAs in RTEL1-deficient mouse MBs is reminiscent of the genome instability found in human SHH, group 3, and group 4 MBs. Among these recurrent CNAs, we identified MycN amplification (Figure 3-6; Table 3-3) which is a well-established oncogene for MB formation[100]. Therefore, MycN amplification could be a driver gene mediating the transformation of murine cerebellar stem cells into MB in RTEL1-deficient mouse MBs.

(a) (b)

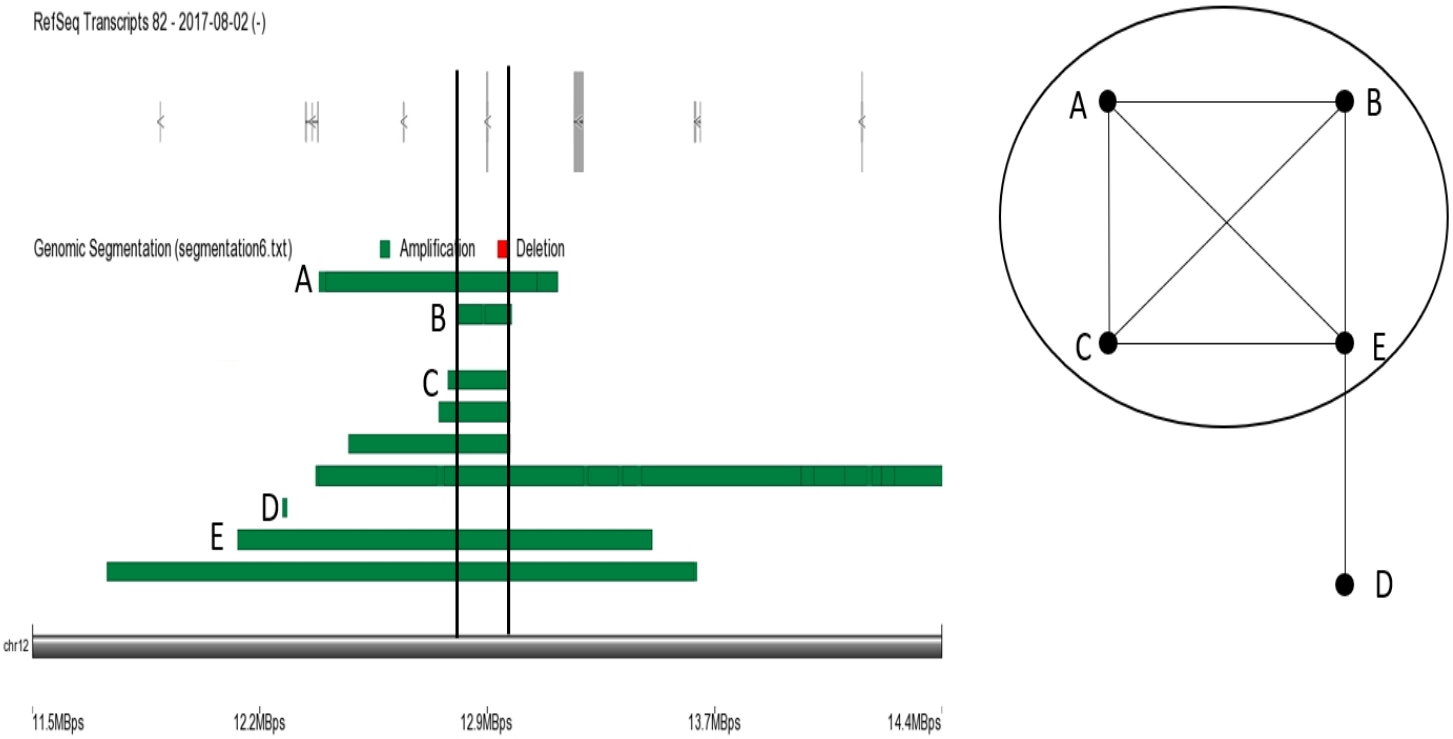


Figure 3-5. The principle of using graph-based maximal clique algorithm to identify affected genes among CNAs. (a) CNAs with gain of copy number (marked as green lines) are firstly identified through genomic segmentation analysis. The RefSeq annotation track shows the location and orientation of genes presented in these CNAs. Within the solid vertical black lines is the genomic region most commonly affected by CNAs identified by the maximal clique. (b) Interval graph of the CNAs presented in (a). Each vertex represents a CNA and edges are drawn between CNAs that overlap a common region. Within the circle is an example of a maximal clique.

Table 3-3. Recurrent CNAs with gain of copy number in RTEL1-deficient mouse MBs identified by maximal clique analysis.

Chromosome	Inner Start	Inner End	Genes overlapped	Gene Function	Samples affected (%)
6	55937788	55946570	Ccdc129	ITPR interacting domain containing 1	70%
6	56638966	56644300	Gm8239	Predicted lincRNA gene	70%
12	12852714	12928294	Gm48187 4930519A11Rik Rpl36-ps3 Gm40271	Predicted genes	65%
12	12928294	12991120	Gm48201	Predicted gene	65%

			Gm35208 Mycn	Myc-family member transcription factor	
6	56229541	56236151	Pde1c	3’5’-cyclic nucleotide phosphodiesterase specific for cAMP and GMP	65%
6	56355596	56388066	Pde1c		65%
6	55483197	55508566	Adcyap1r1	Membrane associated receptor protein for PACAP-27 and PACAP-38. Ligand binding activates adenylyl cyclase	14
6	55508566	55627004	Gm44352	Predicted gene	14
6	55646999	55937788	Neurod6 Gm3279 Ccdc129	E-box transcription factor that may be involved in nervous system development and differentiation Predicted gene	14
6	56012652	56224323	Gm44413 Ppp1r17 Pde1c	Predicted gene Protein phosphatase inhibitor primarily found in cerebellar Purkinje cells	14
6	56262463	56271582	Pde1c		14
6	56388066	56624424	Gm22493 Gm24709 Gm8239 Pde1c	Predicted gene Predicted gene Predicted gene	14
6	56624424	56636945	Gm8239	Predicted gene	14
6	63316540	63324971	Gm44071 Grid2	Predicted gene Delta receptor primarily found in cerebellar Purkinje cells with roles in synaptogenesis, synaptic plasticity, and motor coordination	14
6	119774108	119785615	Erc1	Regulatory protein in nerve active zones that regulates neurotransmitter release	14
6	148888870	148895064	Caprin2	Regulatory protein that promotes WNT signalling by phosphorylation of LRP6	14

Table 3-4. Recurrent CNAs with loss of copy number in RTEL1-deficient mouse MBs identified by maximal clique analysis.

Chromosome	Inner Start	Inner End	Genes overlapped	Gene Function	Samples affected (%)
13	111420604	111451347	AC118475.1 Gpbp1	GC-rich promoter-specific trans-activating transcription factor	91%
13	113585953	113603881	Snx18	Nexin family protein involved in endocytosis and intracellular trafficking	21
13	114000319	114011815	Arl15	GTP binding protein	21
13	117668618	117677206	Hcn1	Voltage-gated potassium channel with roles in pacemaking in heart and neuronal tissue	21
13	118574953	118593029	1700003P14Rik Gm47330	Predicted gene Predicted gene	21

13	103856671	103863921	Erbin	ERBB2 binding protein shown to affect RAS-RAF signaling pathway	21
13	103960364	103979927	Gm47849 Gm47851	Predicted gene Predicted gene	21
13	96762614	96778338	Gm48580 Ankrd31	Predicted gene Structural protein containing many Ankyrin repeats	21
13	97872071	97887621	Gm34388	Predicted gene	21
13	100139774	100152718	Naip2	Component of NLRC4 inflammasome responsible for innate immune response on variety of bacteria	21
13	86019471	86070755	Gm22574 Cox7c	Predicted gene Oxidase protein in the electron transport chain	21
13	86700747	86744696	Gm25700	Predicted gene	21
13	87197387	87254688	Gm47642	Predicted gene	21
13	87306105	87316094	Gm18265	Predicted gene	21
13	89112052	89133497	Edil3	Integrin ligand with a role in angiogenesis	21
13	90716814	90795666	Gm44338 Gm47521	Predicted gene Predicted gene	21
13	96128364	96138297	Sv2c Gm29543	Mediator of neurotransmitter release in neural and endocrine cells Predicted gene	21

3.2.2.2 Characterization of MycN amplification in RTEL1-deficient mouse MBs

MycN belongs to the Myc family of genes that code for transcription factors which have long been known to function as oncogenes[139,140]. It has been found to be frequently amplified in human SHH and group 4 MBs (Table 1-1)[14,37]. In addition, transgenic overexpression of MycN in cerebellar neuronal progenitor cells can induce MB formation which largely recapitulates the genetic features of human group 4 and SHH MBs[100]. All of these strongly indicate that amplification of MycN is a driver mutation for the formation of SHH MBs.

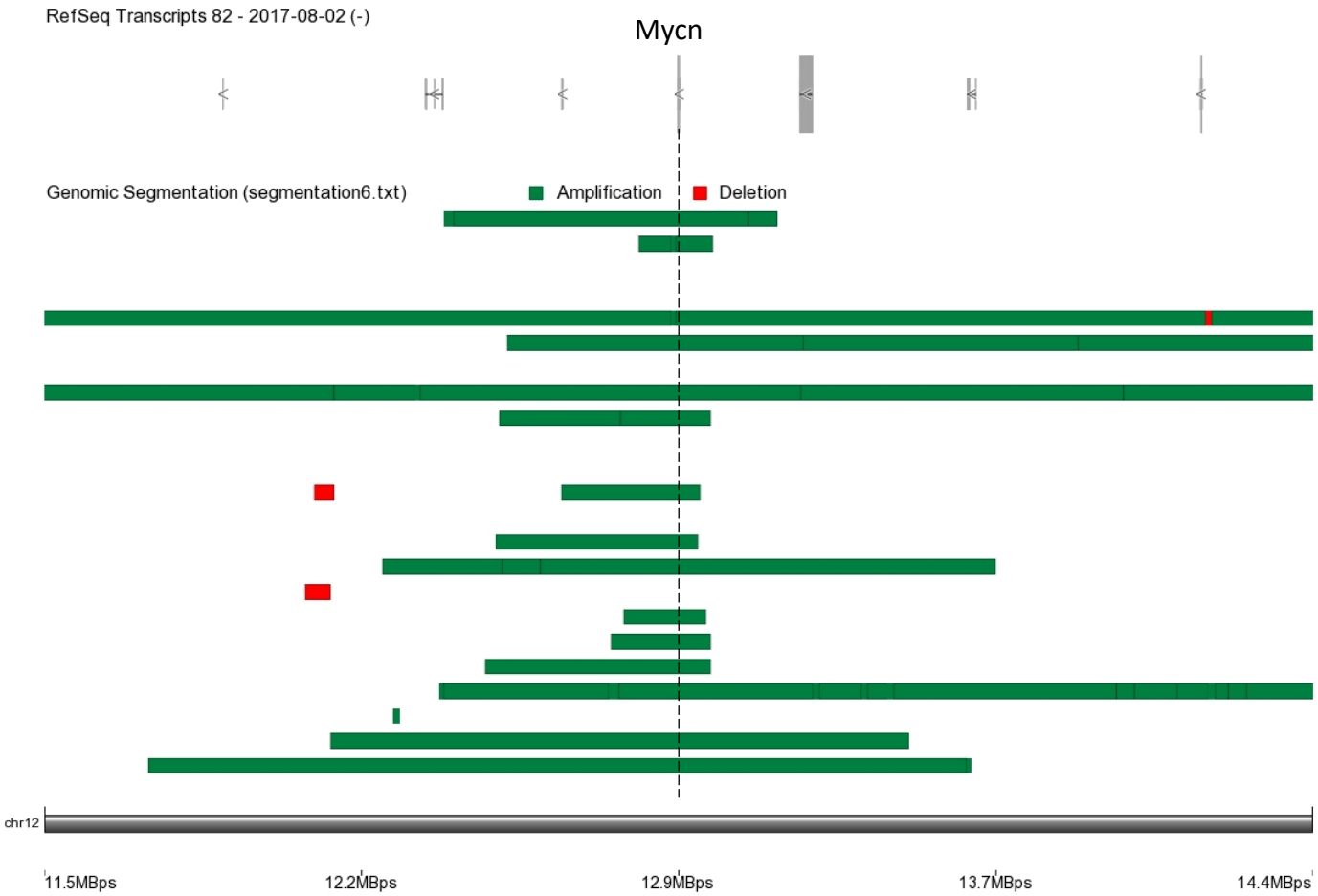


Figure 3-6. Amplification of MycN in RTEL1-deficient mouse MBs as detected by array based CGH. 15 out of 23 tumours showed CNAs with gain of copy numbers (marked as green lines) which contains MycN.

To validate the amplification of MycN detected by CGH analysis in RTEL1-deficient mouse MBs, Taqman copy number analysis was performed. In this assay, specific mouse MycN probes were applied to detect the copy number of this gene in DNA samples extracted from RTEL1-deficient mouse MB by using real-time PCR. Consistent with my CGH data, around 60% of RTEL1-deficient mouse MB also showed the amplification of MycN in this analysis (Figure 3-7). In addition, several tumours were also found to contain more than 60 copies of MycN gene.

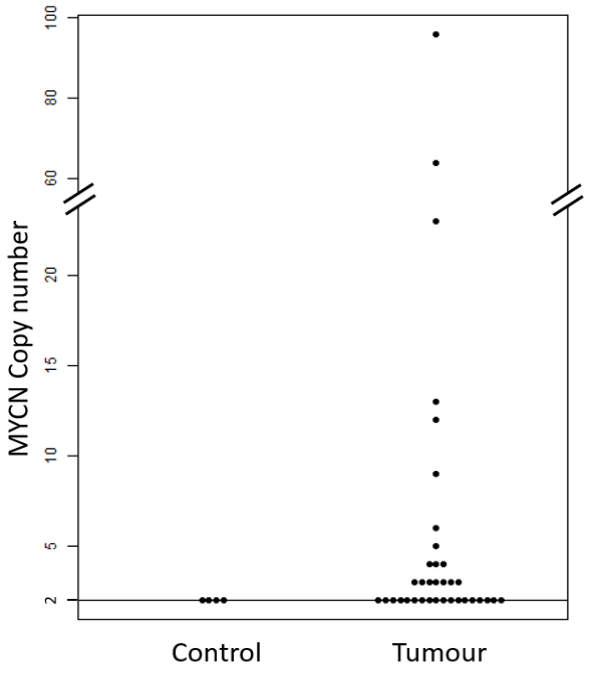
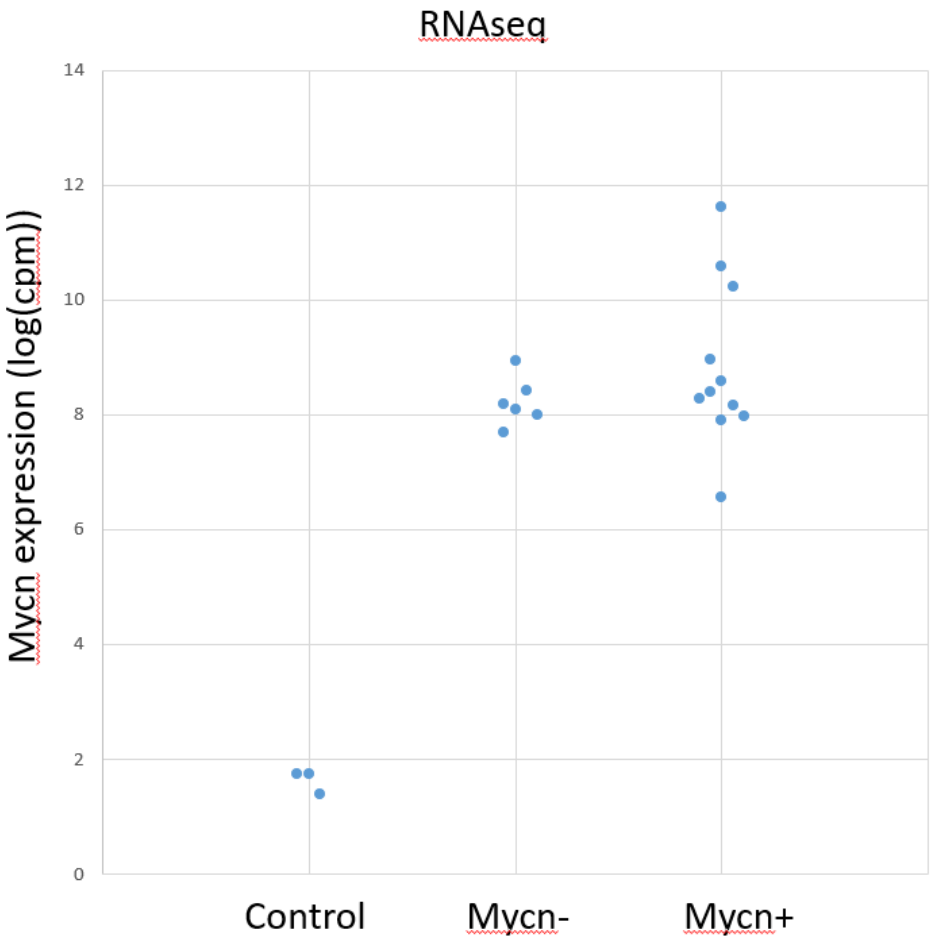


Figure 3-7. Taqman copy number analysis of MycN in RTEL1-deficient mouse MBs. In this analysis, around 60% tumours showed increased MycN copy number as compared to the control.

To correlate the genomic amplification of MycN with the expression of this gene in RTEL1-deficient mouse MBs, we also examined the expression levels of MycN by using both microarray and RNA-sequencing (Figure 3-8). MycN expression was greatly upregulated in RTEL1-deficient mouse MB models with MycN amplification (MycN+) compared to control ($\log_2$ fold change = 7.62, FDR = 2.78e-5). Interestingly, RTEL1-deficient mouse MB models without MycN amplification (MycN-) also showed Mycn overexpression with a similar magnitude to those with MycN amplification, suggesting that MycN activation could also be induced by other mechanisms in these tumours.



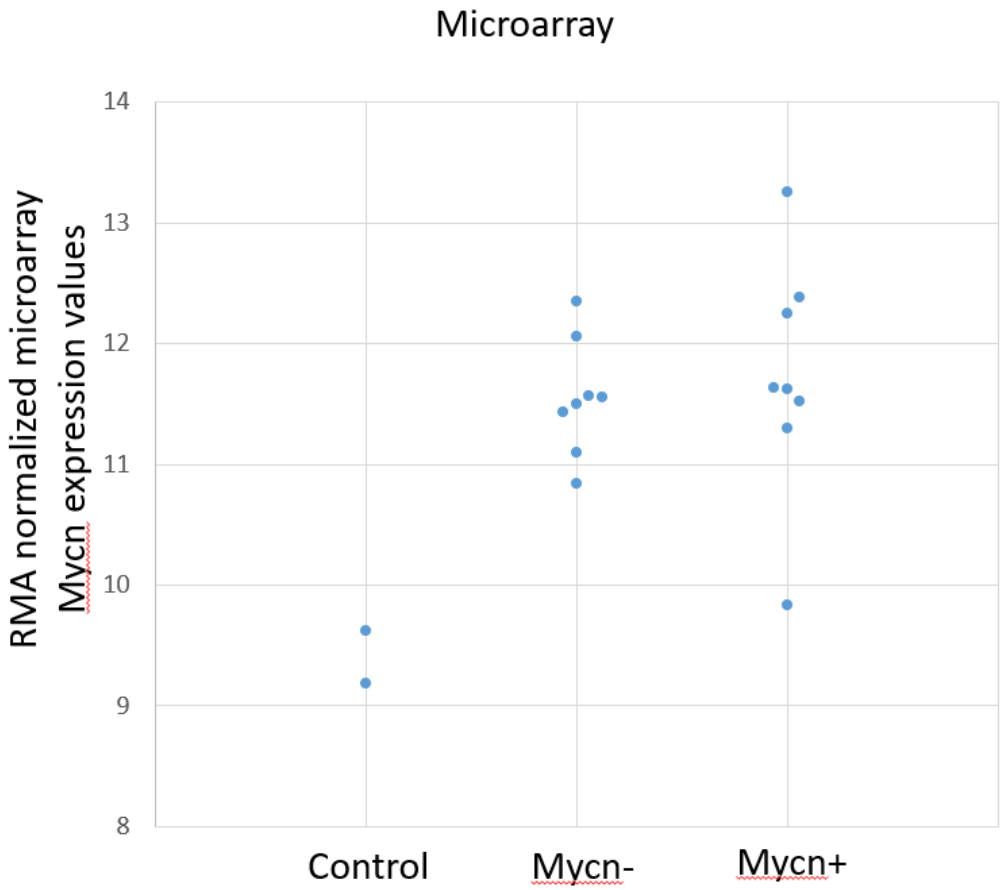


Figure 3-8. Upregulation of MycN expression in RTEL1-deficient mouse MBs. Both RNAseq and microarray-based expression analysis showed MycN overexpression in these tumours with and without MycN amplification.

Since the main function of MycN is to regulate gene expression by binding to the E-box sequence, a gene set enrichment analysis, using the Broad Institute software called GSEA[117,141], was performed to determine whether the genes with the E-box consensus sequence are enriched in RTEL1-deficient mouse MB. As shown in Figure 3-9, the hallmark signature gene set with E-box was found to be enriched in these tumours.

Taken together, both copy number and gene-expression analyses strongly indicate that the activation of MycN is an important genetic event to drive the formation of MB in RTEL1-deficient mouse MB model.

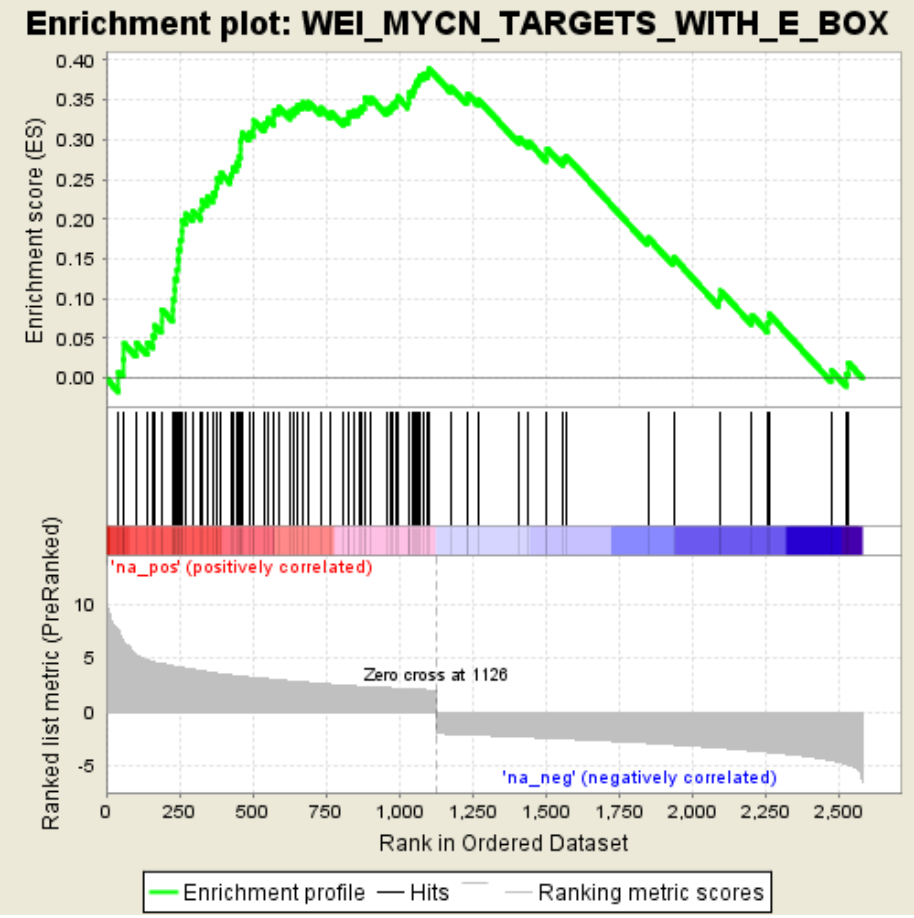


Figure 3-9. Gene set enrichment analysis of differentially expressed genes (DEG) between RTEL1-deficient mouse MBs and wildtype. MycN target genes, i.e. E-box genes, were found to be enriched in these tumours.

3.2.2.3 Identification of recurrent genetic alterations in RTEL1-deficient mouse MBs by next generation sequencing

To further search for potential driver mutations in RTEL1-deficient mouse MBs, I also used next generation sequencing, i.e. WES and RNA-seq, both allowing to identify point mutations, insertions/deletions (indels) and genetic fusions, in the coding sequences in these tumours. Only homozygous mutations that result in an amino acid sequence change were included in this analysis because these are the mutations most likely to have tumorigenic roles and RTEL1-deficient mouse MB genomes are highly unstable and carry many passenger mutations. 11 RTEL1-deficient mouse MBs were analyzed by WES (Table 3-5), and 17 were analyzed by RNAseq (Table 3-6).

An interesting mutation discovered is a homozygous stopgain mutation in the Ces1b gene identified by WES in over 90% of RTEL1-deficient mouse MBs. The Ces1b gene codes for carboxylesterase 1, a protein used as a marker for monocytes and thought to have a role in surveilling or killing cancerous cells[142]. Therefore, loss of Ces1b function in RTEL1-deficient mouse MBs could protect these tumour cells from monocyte-mediated cytotoxicity.

To better predict driver mutations from these sequencing data, I applied the 20/20+ method[131] which was established on the assumption that if 20% of mutations that occur in the particular gene are all loss of or gain of function, then these genetic mutations most likely play a role in tumour development.

Table 3-7 lists putative oncogenes of RTEL1-MB identified by 20/20+ from RNAseq data. Interestingly, the 20/20+ algorithm was unable to identify any putative TSGs. This could be the result of low sample size resulting in weaker statistical power, but the fact that no

Table 3-5. Recurrent homozygous genetic mutations in RTEL1-deficient mouse MBs identified by WES. Homozygous variants recurring in at least 5 tumours are considered recurrent. Silent mutations were not included in this analysis, only point mutations and small insertions and deletions that resulted in an amino acid sequence change were included. mm10 genome build was used for reference.

Locus ID (chromosome : start : end : reference allele : alternate allele)	Gene Symbol	Gene description	Mutation type	Samples affected (11 analyzed)
chr8:93068190:93068190:-:TTAG	Ces1b	carboxylesterase 1B	stopgain SNV	10
chr11:71181355:71181355:A:G	Nlrp1b	NLR family, pyrin domain containing 1B	nonsynonymous SNV	7
chr11:71181362:71181362:C:T	Nlrp1b	NLR family, pyrin domain containing 1B	nonsynonymous SNV	7
chr11:72215694:72215694:-:GGCGAGACAAGCAAGCCGCTC	4930563E22Rik	RIKEN cDNA 4930563E22 gene	nonframeshift insertion	7
chr11:71181353:71181353:T:C	Nlrp1b	NLR family, pyrin domain containing 1B	nonsynonymous SNV	6
chr15:79043075:79043075:A:G	Galr3	galanin receptor 3	nonsynonymous SNV	6
chr8:40681704:40681704:-:AC	Adam24	a disintegrin and metallopeptidase domain 24 (testase 1)	frameshift insertion	6
chr9:38143981:38143981:C:T	Olfr890	olfactory receptor 890	nonsynonymous SNV	6
chr11:71181365:71181365:T:-	Nlrp1b	NLR family, pyrin domain containing 1B	frameshift deletion	5
chr11:71181367:71181367:-:C	Nlrp1b	NLR family, pyrin domain containing 1B	frameshift insertion	5

Table 3-6. Recurrent homozygous genetic mutations identified by RNAseq. Homozygous variants occurring in at least 8 tumours are considered recurrent. Silent mutations and mutations on noncoding RNA were not included in this analysis, only point mutations and small insertions and deletions that resulted in an amino acid sequence change were included. mm10 genome build was used for reference.

Locus ID (chromosome : start : end : reference allele : alternate allele)	Gene symbol	Gene description	Mutation type	Samples affected (17 analyzed)
chr11:99432951:C:T	Krt20	keratin 20 [Source:MGI Symbol;Acc:MGI:1914059]	Missense variant	15
chr4:129578336:T:C	Fam167b	family with sequence similarity 167, member B [Source:MGI Symbol;Acc:MGI:2668032]	Missense variant	15
chr11:99418396:T:C	Krt12	keratin 12 [Source:MGI Symbol;Acc:MGI:96687]	Missense variant	14
chr11:99438027:C:T	Krt20	keratin 20 [Source:MGI Symbol;Acc:MGI:1914059]	Missense variant	14
chr14:37098261:G:A	Cdhr1	cadherin-related family member 1 [Source:MGI Symbol;Acc:MGI:2157782]	5’ UTR premature start codon gain variant	14
chr14:5516321:T:G	Gm3488	predicted gene, 3488 [Source:MGI Symbol;Acc:MGI:3781665]	Missense variant	14
chr1:172057547:C:T	Nhlh1	nescient helix loop helix 1 [Source:MGI Symbol;Acc:MGI:98481]	5’ UTR premature start codon gain variant	13
chr3:4212088:A:G	Gm8775	predicted gene 8775 [Source:MGI Symbol;Acc:MGI:3646219]	Missense variant	13
chr3:62602623:C:A	Gpr149	G protein-coupled receptor 149 [Source:MGI Symbol;Acc:MGI:2443628]	Missense variant	13
chr5:92363176:T:TG	Cxcl11	chemokine (C-X-C motif) ligand 11 [Source:MGI Symbol;Acc:MGI:1860203]	Frameshift variant	13
chr8:39624140:T:C	Msr1	macrophage scavenger receptor 1 [Source:MGI Symbol;Acc:MGI:98257]	Missense variant	13
chr1:97025614:T:C	Gm6430	predicted gene 6430 [Source:MGI Symbol;Acc:MGI:3648857]	Missense variant	12

chr11:51651118:C:CG	D930048N14Rik	RIKEN cDNA D930048N14 gene [Source:MGI Symbol;Acc:MGI:2144709]	Frameshift variant	12
chr7:81692865:G:C	Homer2	homer scaffolding protein 2 [Source:MGI Symbol;Acc:MGI:1347354]	5' UTR premature start codon gain variant	12
chr14:14347024:G:T	Il3ra	interleukin 3 receptor, alpha chain [Source:MGI Symbol;Acc:MGI:96553]	5' UTR premature start codon gain variant	11
chr14:37079596:G:A	Cdhr1	cadherin-related family member 1 [Source:MGI Symbol;Acc:MGI:2157782]	Missense variant	11
chr14:51891584:G:GC	Mettl17	methyltransferase like 17 [Source:MGI Symbol;Acc:MGI:1098577]	Frameshift variant	11
chr14:51891587:A:G	Mettl17	methyltransferase like 17 [Source:MGI Symbol;Acc:MGI:1098577]	Missense variant	11
chr17:29262060:C:CT	Ppil1	peptidylprolyl isomerase (cyclophilin)-like 1 [Source:MGI Symbol;Acc:MGI:1916066]	Frameshift variant	11
chr17:9422390:C:T	Gm17728	predicted gene, 17728 [Source:MGI Symbol;Acc:MGI:4937362]	Missense variant	11
chr1:136216201:A:G	5730559C18Rik	NA	Missense variant	10
chr10:31249518:A:G	Gm5422	predicted pseudogene 5422 [Source:MGI Symbol;Acc:MGI:3643411]	Missense variant	10
chr10:93847307:T:A	Usp44	ubiquitin specific peptidase 44 [Source:MGI Symbol;Acc:MGI:3045318]	Missense variant	10
chr11:83504715:CTATGGTGGA GACAGAAGTGGGGGAGGT:C	Taf15	TATA-box binding protein associated factor 15 [Source:MGI Symbol;Acc:MGI:1917689]	Disruptive inframe deletion	10
chr13:67206991:C:T	Zfp455	zinc finger protein 455 [Source:MGI Symbol;Acc:MGI:3040708]	Missense variant	10
chr14:5976328:A:G	Gm3453	predicted gene 3453 [Source:MGI Symbol;Acc:MGI:3781629]	Missense variant	10
chr1:170963333:C:A	Fcgr2b	Fc receptor, IgG, low affinity IIb [Source:MGI Symbol;Acc:MGI:95499]	Missense variant	9
chr10:10346526:T:C	Adgb	androglobin [Source:MGI Symbol;Acc:MGI:3605549]	Missense variant	9
chr13:18031149:C:A	Vdac3-ps1	voltage-dependent anion channel 3, pseudogene 1 [Source:MGI Symbol;Acc:MGI:1270159]	Missense variant	9

chr13:23574606:G:C	Hist1h2ad	histone cluster 1, H2ad [Source:MGI Symbol;Acc:MGI:2448289]	Missense variant	9
chr14:36939929:T:C	Ccser2	coiled-coil serine rich 2 [Source:MGI Symbol;Acc:MGI:101859]	Missense variant	9
chr18:51865526:A:G	Gm4950	predicted pseudogene 4950 [Source:MGI Symbol;Acc:MGI:3649015]	Missense variant	9
chr5:138364248:C:T	Gm10874	predicted gene 10874 [Source:MGI Symbol;Acc:MGI:3704261]	5’ UTR premature start codon gain variant	9
chr9:59145760:C:T	Gm7589	predicted gene 7589 [Source:MGI Symbol;Acc:MGI:3644073]	Missense variant	9
chr1:107607004:A:C	Serpinb8	serine (or cysteine) peptidase inhibitor, clade B, member 8 [Source:MGI Symbol;Acc:MGI:894657]	Missense variant	8
chr1:97024836:T:G	Gm6430	predicted gene 6430 [Source:MGI Symbol;Acc:MGI:3648857]	Missense variant	8
chr17:46411488:ACCACCT:A	Zfp318	zinc finger protein 318 [Source:MGI Symbol;Acc:MGI:1889348]	Disruptive inframe deletion	8

significant TSGs were identified could also mean that in the context of RTEL1-MBs, point mutations only affect oncogenes. Perhaps, TSGs are disrupted instead by larger genetic alterations such as CNAs.

3.2.3 Summary

In this part of the study, I used CGH, WES, and RNAseq data to identify Mycn activation as a recurrent genetic alteration in RTEL1-deficient mouse MB model. This finding was further supported by expression analysis using microarray and Taqman copy number analysis. MycN is a well-established oncogene and could be the driver gene that mediates cerebellar stem cell transformation to MB in RTEL1-deficient mouse MB model.

In this study, I also identify genes that frequently carry homozygous, non-silent point mutations or small insertions and deletions in RTEL1-deficient mouse MB model. It will be interesting to see how these mutations affect the function of these genes and whether they have a role in tumour initiation, maintenance, or progression.

Table 3-7. Top 15 putative oncogenes identified in RTEL1-deficient mouse MBs by the 20/20+algorithm. The main criteria of the 20/20+ algorithm for driver gene detection is that 20% of all mutations occurring in a gene must be non-silent. This table shows the proportion of the varying mutations types occurring in each gene. The oncogene score is the fraction of trees that labelled the gene as an oncogene in a three-class (oncogene, tumour suppressor gene, or passenger gene) random forest analysis, based on 24 predictive features for tumorigenic genes[131]. Mutations were identified by mapping to the mm10 genome build.

Gene	Gene Length	Silent (%)	Nonsense (%)	Splice site (%)	Missense (%)	Frameshift Indel (%)	Inframe Indel (%)	Oncogene score	Oncogene p-value	Oncogene q-value
ABI1	1527	0	0	0	100	0	0	0.939	0	0
PRDX6	675	0	0	0	100	0	0	0.939	0	0
POLR2B	3525	0	0	0	100	0	0	0.932	0	0
HOXB5	810	0	0	0	100	0	0	0.93	0	0
CALCOCO2	1413	0	0	0	100	0	0	0.928	0	0
BEND7	1407	0	0	0	100	0	0	0.923	0	0
PSMD14	933	0	0	0	100	0	0	0.922	0	0
ATP2B3	3663	0	0	0	100	0	0	0.922	0	0
EXOC1	2685	25	0	0	75	0	0	0.921	0	0
EPS15	2691	0	0	0	100	0	0	0.92	0	0
BIRC2	1857	0	0	0	100	0	0	0.918	0	0
HSPA4L	2520	0	0	0	100	0	0	0.918	0	0
PEX19	900	0	0	0	100	0	0	0.917	0	0
CSNK2A2	1053	0	0	0	100	0	0	0.91	0	0
SIRT7	1203	0	0	0	100	0	0	0.909	0	0

3.3 To characterize chromothripsis in RTEL1-deficient mouse MBs

3.3.1 Rationale

As demonstrated above, RTEL1-deficient mouse MBs frequently contain the amplification of MycN oncogene. However, how this genetic alteration occurs in this mouse model is unknown. Since telomere dysfunction-induced chromothripsis could be a main cause for genomic rearrangement and amplification (Figure 1-3), plus RTEL1 is required for the maintenance of telomeres (Figure 1-6) we then asked whether MycN amplification could be caused by chromothripsis.

Chromothripsis is a single catastrophic genetic event which can lead to massive genomic rearrangements, including the formation of double-minute chromosomes which typically harbor oncogenes (Figure 1-4)[78]. Chromothripsis can be inferred by the detectable genomic signature it leaves behind using different operational definitions based on CGH copy-number profiling data, from which copy-number state information as well as the number of DNA breakpoints can be obtained[77]. If a genomic rearrangement is indeed caused by chromothripsis, three genetic signatures should be observed in the CGH copy-number profiling data: (1) Clustering of DNA breakpoints on a single chromosome, here we define this as 10 DNA breakpoints adjacent to each other, (2) A low number of copy number states within the cluster of breakpoints, defined here as a maximum of 3 copy number states, and (3) An oscillation between the copy number states within the cluster of breakpoints. Based on this definition, in this part of the study, I used CGH copy-number profiling data obtained from RTEL1-deficient mouse MBs to determine whether these tumours could display chromothripsis-like pattern.

3.3.2 Results

To identify chromothripsis-like pattern in RTEL1-deficient mouse MBs, CGH based copy-number profiling data were obtained from 23 RTEL1-deficient mouse MBs. 11 of these tumours were found to follow all three of the criteria of chromothripsis, i.e. at least 10 DNA breakpoints associated with oscillating copy-number alterations on a particular chromosome (Figure 3-10). Two tumours were also found to display chromothripsis-like pattern on three chromosomes, indicating either consecutive chromothripsis events or one large chromothripsis event affecting several chromosomes.

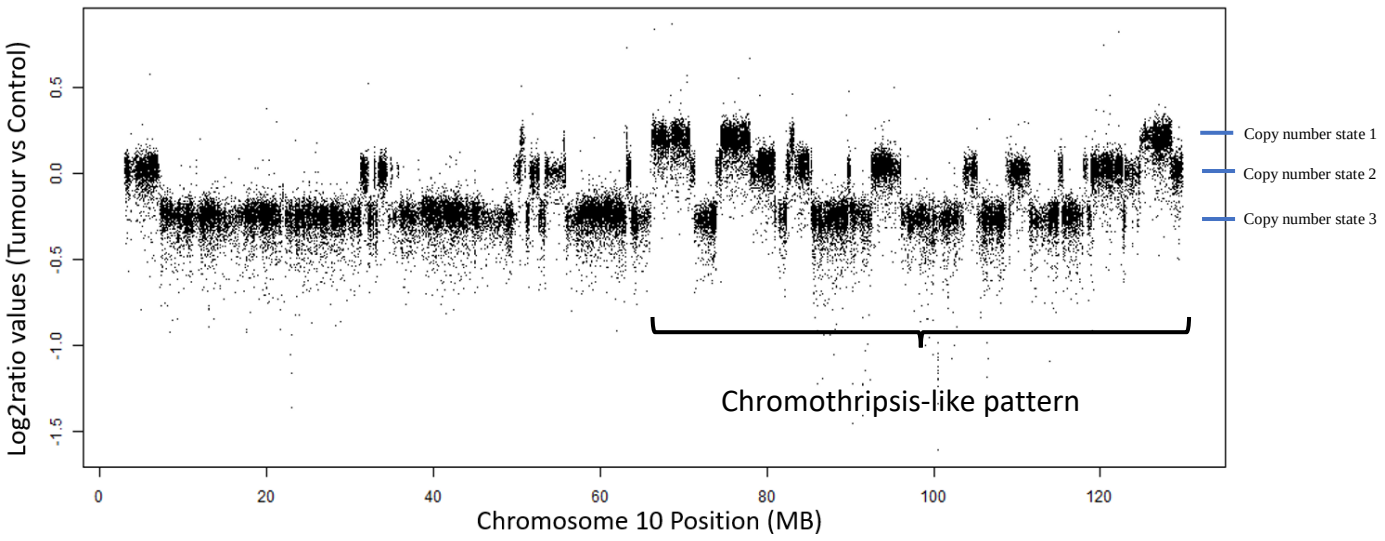


Figure 3-10. A chromothripsis-like pattern as detected by log2ratio profile of CGH from one RTEL1-deficient mouse MB. This profile perfectly matches with the signatures which indicates chromothripsis: (1) the presence of more than 10 DNA breakpoints that are associated with oscillating copy-number alterations on a particular chromosome (marked by abrupt changes in copy number states along the chromosome); (2) with the copy number changes oscillating between copy-number states; and (3) a low number of copy number states, in this case 3 visibly identifiable states.

3.3.3 Summary

In this study, I used log2ratio profiles obtained from CGH array to show signature genomic marks that could indicate a chromothripsis event occurring in RTEL1-deficient mouse

MBs. Since chromothripsis can result from telomere dysfunction and RTEL1 is responsible for telomere maintenance, chromothripsis could be the main mechanism in which RTEL1-deficient mouse MBs gain genetic alterations. However, to truly infer chromothripsis, whole genome sequencing data is required to identify the relative order and orientation of the DNA rearrangements, which is considered as the golden standard for chromothripsis as suggested by Korbel and Campbell[77]

Chapter 4: Discussion

4.1 Study background

Human MB is a common brain tumour that occurs mainly during childhood. It is also a highly heterogeneous tumour which can be classified into four broad molecular groups with distinct genetic features[14]. This tumour could also arise from multiple cell lineages, most of them being stem/progenitor cells, located in the cerebellum or dorsal brainstem[143], which further makes this tumour more heterogenic. Like many other human cancers, the pathogenesis of MB is still largely unknown.

During our characterization of the role of RTEL1 DNA helicase in development, we found that RTEL1 is expressed in cerebellar stem cells located at the 4th ventricular zone. Using a conditional knockout approach, we further showed that inactivation of RTEL1 in these stem cells can lead to the formation of MB. Since cerebellar stem cells have been implicated as the cell of origin of MB[143], we expect that this mouse model could be valuable for understanding how these stem cells are transformed to form MB. Based on the demonstrated role of RTEL1 in the maintenance of telomere and genomic stability, we hypothesize that loss of RTEL1 function in cerebellar stem cells could induce some genetic alterations with tumorigenic function. Therefore, identification of these alterations could not only help to increase our knowledge on MB formation but may also lead to establish new therapeutic targets for this tumour.

My thesis aims to characterize the genetic features of RTEL1-deficient mouse MB model in order to identify which subgroup of human MB this mouse model can represent and what are the potential driver mutations involved in this tumorigenesis. Since this mouse model also contains massive genomic rearrangements, my thesis also focused on understanding its mechanisms. My findings, as discussed following, indicate that RTEL1-deficient mouse MB

model could largely recapitulate the genetic features presented in human SHH MB, and MycN amplification may act as an important oncogenic event to drive this group of MB formation. In addition, my findings also implicate that RTEL1 could be required for protecting cerebellar stem cells from some catastrophic genetic events, such as chromothripsis, thus, preventing the formation of MB. This research progress will not only demonstrate that RTEL1-deificient mouse MB model can be a genetic tool to study human SHH MB, but also help to understand the mechanism of genomic instability presented in this group of MB.

4.2 Research findings from this study

4.2.1 RTEL1-deficient mouse MB model could represent human SHH MB

Mouse models have been demonstrated as important genetic tools for *in vivo* analyzing gene's function in development and in disease. This is largely because mice share high similarity with humans not only in genome but also in anatomy and physiology. In addition, mice have a short life span (less than 2 years) and embryonic development (around 19 days). These, together with several powerful transgenic tools allowing to specifically manipulate mouse genome function, make mouse model a most attractive and reliable genetic tool for modeling human disease or understanding a gene's *in vivo* function[96]. However, mice are not small humans. Mice have much fewer cell divisions and high metabolic rate compared to humans. Mice also contain some genetic features, such as relative long telomere, which are different from humans[144,145]. Therefore, before using a mouse model for studying a human disease or cancer, it is important to characterize it to make sure this model can largely recapitulate both genetic and pathological features of human disease or cancer.

Although RTEL1-deficient mouse MB model shared similar pathological changes with human MB, whether this model can largely recapitulate the genetic alterations as demonstrated in a given group of human MB was unknown. Therefore, the first part of my thesis was to determine which group of human MB that RTEL1-deficient mouse MB model can belong to. This was done by comparing the gene-expression files of this mouse model with a cohort of human MB (a total of 763 patients in this cohort). As demonstrated in several previous studies, microarray-based RNA expression could allow to divide human MB into four distinct molecular groups[13,14,20,146]. However, all these analyses were performed without normal control. To use this human dataset to compare with the microarray data from our mouse model, from these two datasets, I first created gene-expression profiles, which are high-dimensional data (more than 10 dimensions) because they contain the measurements of thousands of genes and each gene is considered a dimension. These high-dimensional data can be used to compare the similarity by t-SNE, an unsupervised nonlinear dimensionality reduction method which allows to calculate pairwise similarity probability that can be further graphed in lower dimensional space, i.e. in 2- or 3-dimensions. Using this bioinformatic approach, I can successfully separate a cohort of 763 human MB into four groups (Figure 3-1). This is consistent with a classification study on this cohort[36], demonstrating that my established t-SNE can be a reliable tool to group tumour samples based on normalized microarray signals. To determine whether t-SNE and PCA can be applied to group mouse MB samples with human MB samples, additional mouse models were included as a control. PCA was able to correctly subgroup these mouse models (Figure 3-2) but t-SNE tends to group all the mouse models together (data not shown), indicating that PCA and t-SNE can also be used for grouping samples across species, however t-SNE must be used on each mouse model

separately. Based on these, we concluded that t-SNE and PCA can be used to determine which group of human MB that RTEL1-deficient mouse MB can belong to.

Using t-SNE and PCA method, we found that RTEL1-deficient mouse MB can belong to the human SHH subgroup (Figure 3-1 and Figure 3-2). To further support this finding, I also applied a supervised learning machine model (SVM) to group RTEL1-deficient mouse MB model with human MB. This model was trained on group-specific metagenes identified from human MB and used to classify mouse MBs. With these metagenes, I was able to demonstrate that RTEL1-deficient mouse MBs belong to the SHH subgroup, similar to the SMOM2 mouse model (Figure 3-3).

Both unsupervised and supervised bioinformatic approaches, should allow us to conclude that this mouse model could largely represent human SHH MB.

4.2.2 MycN amplification could be a driver mutation in RTEL1-deficient mouse MB model

One interesting finding from my thesis was the identification of MycN amplification in more than 65% RTEL1-deficient mouse MBs. This was done by an array-based CGH approach that allowed to detect copy-number alterations (CNAs) in a genome-wide manner. Using this method, I found all RTEL1-deficient mouse MBs had massive copy number changes or loss of chromosomes, supporting a critical role of RTEL1 in the genomic maintenance. To more precisely identify the genes presented in CNAs, I used a maximal clique algorithm which can greatly help me to identify the genes by narrowing down the overlapped region among different CNAs in the same chromosomal region (Figure 3-5). Using this method, I found numerous recurrent CNAs with either gain of copy number or loss of copy number in RTEL1-deficient

mouse MBs (Table 3-3 and Table 3-4). Among them, amplification of MycN oncogene appears to be an important driver mutation in our mouse model.

MycN is a well-established oncogene which has been found to be frequently amplified in neuroblastoma[147]. In human MB, it has been found to be amplified in Group 4 MB (around 6.3% of all Group 4 MBs) and SHH MB (around 8.2% of all SHH MBs)[37]. Other groups of MBs seldomly contain MycN amplification. As demonstrated in this study, RTEL1-deficient mouse MB model was also found to belong to SHH MB. The presence of MycN amplification in this mouse model strongly supports that this genetic alteration is unique for this group of MB. Even in RTEL1-deficient mouse MBs which didn't show MycN amplification, these tumours still displayed high MycN expression (Figure 3-8), which further support an important role of MycN in MB formation.

MycN is a nuclear protein which contains MYCBoxes (numbered I, II, IIIa, IIIb and IV) and a basic helix-loop-helix-leucine zipper (bHLH-LZ) domain. It forms heterodimers with another bHLH-LZ protein (MAX) to specifically interact with E-box to regulate the expression of genes involved in cell proliferation and survival. This function of MycN is important for cellular transformation as demonstrated by both *in vitro* and *in vivo* studies[148]. The tumorigenic function of MycN in MB formation has also been supported by either xenograft mouse models or transgenic mouse models which contained overexpression of MycN in cerebellar stem/progenitor cells. In these mouse models, high MycN activity can efficiently induce MB formation[100]. In consistent with these findings, RTEL1-deficient mouse model also showed high MycN activity as reflected by enriched MycN targeted genes (Figure 3-9). This finding adds additional genetic evidence to support that MycN is an important oncogenic event in the development of SHH and group 4 MB, and MycN could be therapeutic target for this group of tumours.

In this thesis study, I also intended to use other genome-wide approaches to search for the mutations that could initiate MB formation in our mouse model. These approaches are whole exome sequencing and RNA sequencing (RNAseq), both of which are cost-effective, and together with available analysis software tools, allow for an unbiased screen of driver mutations in multiple biological samples proving these to be the most sensitive and reliable tools for detecting nucleotide mutations, such as point mutations, insertion/deletion (Indel) and fusions, in mouse genetic screenings, rare human genetic diseases and cancer[136–138]. Although using these assays, numerous recurrent genetic mutations were identified in RTEL1-deficient mouse MBs (Table 3-5 and Table 3-6), none of them appear to be a drive mutation based on the criteria that include: (1) a genetic mutation with demonstrated tumorigenic function; (2) a genetic mutation presented in human MB with a high frequency.

4.2.3 RTEL1-deficient mouse MB model could display chromothripsis

My CGH analysis shows RTEL1-deficient mouse MBs contain massive genomic rearrangements. In addition, my copy-number analysis of CGH data also reveals a chromothripsis-like pattern in many of these tumours (Figure 3-10). Chromothripsis is a single catastrophic event which can shatter a single chromosome or a few chromosomes into pieces that, upon DNA damage-repair, could lose or re-ligate in a wrong order to cause massive genomic rearrangement including genomic amplification[78] (Figure 1-4). Since this event has the potential to target multiple genes simultaneously, it has been considered as a rapid way to transform normal cells into cancer/tumour cells[78].

Several recent studies indicate that chromothripsis could present in human MB, responsible for genomic amplification of oncogenes, such as MycN[37,79]. Consistent with this, I

also found that RTEL1-deficient mouse MB model could also display chromothripsis. In addition, I also found that the tumours with chromothripsis are frequently associated with MycN amplification, suggesting that this genetic event could also drive MycN amplification in this mouse model. However, to confirm this, we would need to perform whole genome sequencing on these tumours to delineate the relative order or orientation of rearrangement.

Although chromothripsis has been identified in human MB, how it is induced in these tumours is still largely unknown. My finding from this study suggests that telomere dysfunction could be an important driving force for this event in this brain tumour. This is also supported by the knockout mouse studies showing critical telomere shortening could induce breakage-fusion-bridge cycle, leading to chromothripsis. Therefore, to fully establish a role of chromothripsis in MB formation, it will be interesting to determine whether human MB with chromothripsis could be associated with telomere dysfunction or genetic mutations involved in telomere maintenance.

4.3 Conclusion

In this thesis study, I applied the combined genetic and bioinformatic approaches to demonstrate the RTEL1-deficient mouse MB model could largely recapitulate the genetic alterations of human SHH MB. This mouse model could serve as a valuable genetic tool for investigating the pathogenesis of this group of MB. In addition, my genome-wide analysis also identified MycN amplification and enrichment of MycN targeted genes in this mouse model, indicating that this oncogenic amplification could be an important driving mutation for SHH MB. Finally, by analyzing microarray-based copy number profiling, I also uncovered a chromothripsis-like pattern in RTEL1-deficient mouse MBs, suggesting that this catastrophic

event could contribute to massive genomic rearrangement and genomic amplification as seen in human SHH MB.

Chapter 5: Future directions

5.1 To identify the downstream effectors of MycN that mediate MB formation

One important finding of this study is to demonstrate the activation of MycN pathway in RTEL1-deficient mouse MB model (Figure 3-9). This, together with many human studies showing MycN is activated in human MB, specifically group 4 and SHH MB, strongly indicate that MycN pathway could be important for MB formation. However, how this pathway contributes to this tumorigenesis is largely unknown. Given the high similarity between RTEL1-deficient mouse MB with human SHH MB, we expect that this mouse model could be a valuable genetic tool to determine the downstream targets of MycN pathway that mediate MB formation.

MycN is a transcriptional factor. In order to identify the downstream targets of MycN, we will compare the gene-expression profiles collected from RTEL1-deficient mouse MB, other mouse MB models with MycN activation, human MB with demonstrated high MycN activity, and human neuroblastoma with MycN amplification. If a gene is indeed the target of MycN, it should be misregulated in all these tumour samples. Once these targets are identified, we will further apply CRISPR/Cas9 technology or other transgenic overexpression approaches to manipulate their expression in mice to determine whether they indeed can induce MB formation. This will allow us to discover the downstream targets of MycN involved in MB formation, which should have potential not only for understanding the pathogenesis of MycN in this brain tumour but could also lead to identify new therapeutic targets for treating this tumour.

5.2 To further characterize chromothripsis in RTEL1-deficient mouse MB model by using whole genome sequencing

My CGH analysis showed chromothripsis-like pattern in RTEL1-deficient mouse MB model (Figure 3-10), indicating that this could be a main mechanism for inducing genomic alterations and genomic amplification in this model. Since chromothripsis is presented in human MB, further characterization of chromothripsis in our model could have potential to better understand this genetic event in MB formation. For this, we plan to use whole genome sequencing technology to analyze the order and orientation of rearrangement in this model. This will not only allow to determine whether chromothripsis is induced by critical telomere shortening but can also demonstrate whether this genetic event is a main cause for MycN amplification in this mouse model. This study can provide molecular insight into chromothripsis in the formation of MB.

Chapter 6: References

1. Lanzkowsky, P. *Manual of Pediatric Hematology and Oncology*. *Manual of Pediatric Hematology and Oncology* (2011). doi:10.1016/C2009-0-30529-2
2. Sümer-Turanligil, N. C., Çetin, E. Ö. & Uyanikgil, Y. A contemporary review of molecular candidates for the development and treatment of childhood medulloblastoma. *Child's Nervous System* (2013). doi:10.1007/s00381-012-2014-3
3. Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro. Oncol.* (2017). doi:10.1093/neuonc/nox158
4. Gajjar, A. *et al.* Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol.* (2006). doi:10.1016/S1470-2045(06)70867-1
5. Rutkowski, S. *et al.* Survival and prognostic factors of early childhood medulloblastoma: An international meta-analysis. *J. Clin. Oncol.* (2010). doi:10.1200/JCO.2010.30.2299
6. Fattet, S. *et al.* Beta-catenin status in paediatric medulloblastomas: Correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics. *J. Pathol.* **218**, 86–94 (2009).
7. Stoodley, C. J. & Schmahmann, J. D. Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *Neuroimage* (2009). doi:10.1016/j.neuroimage.2008.08.039
8. Thompson, E. M. *et al.* Prognostic value of medulloblastoma extent of resection after accounting for molecular subgroup: a retrospective integrated clinical and molecular

analysis. *Lancet Oncol.* (2016). doi:10.1016/S1470-2045(15)00581-1

9. Mabbott, D. J., Penkman, L., Witol, A., Strother, D. & Bouffet, E. Core Neurocognitive Functions in Children Treated for Posterior Fossa Tumors. *Neuropsychology* (2008). doi:10.1037/0894-4105.22.2.159

10. Mabbott, D. J. *et al.* Serial evaluation of academic and behavioral outcome after treatment with cranial radiation in childhood. *J. Clin. Oncol.* (2005). doi:10.1200/JCO.2005.01.158

11. Spiegler, B. J., Bouffet, E., Greenberg, M. L., Rutka, J. T. & Mabbott, D. J. Change in neurocognitive functioning after treatment with cranial radiation in childhood. *J. Clin. Oncol.* (2004). doi:10.1200/JCO.2004.05.186

12. Northcott, P. A. *et al.* Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol.* **123**, 615–626 (2012).

13. Kool, M. *et al.* Molecular subgroups of medulloblastoma: An international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol.* **123**, 473–484 (2012).

14. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: The current consensus. *Acta Neuropathol.* **123**, 465–472 (2012).

15. Blomstrand, M. *et al.* Estimated clinical benefit of protecting neurogenesis in the developing brain during radiation therapy for pediatric medulloblastoma. *Neuro. Oncol.* (2012). doi:10.1093/neuonc/nos120

16. Perry, A. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).

17. Gilbertson, R. J. & Ellison, D. W. The Origins of Medulloblastoma Subtypes. *Annu. Rev. Pathol. Mech. Dis.* (2008). doi:10.1146/annurev.pathmechdis.3.121806.151518

18. Gajjar, A. J. & Robinson, G. W. Medulloblastoma—translating discoveries from the bench to the bedside. *Nat. Rev. Clin. Oncol.* **11**, 714–722 (2014).

19. Brown, H. G. *et al.* 'Large cell/anaplastic' medulloblastomas: A Pediatric Oncology Group study. *J. Neuropathol. Exp. Neurol.* (2000). doi:10.1093/jnen/59.10.857

20. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nat. Rev. Cancer* **12**, 818–834 (2012).

21. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29**, 1408–1414 (2011).

22. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).

23. Ramaswamy, V. *et al.* Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol.* (2016). doi:10.1007/s00401-016-1569-6

24. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* (2012). doi:10.1016/j.cell.2012.06.024

25. Ahmed, A. A. *et al.* Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* (2010). doi:10.1002/path.2696

26. Leiserson, M. D. M., Blokh, D., Sharan, R. & Raphael, B. J. Simultaneous Identification of Multiple Driver Pathways in Cancer. *PLoS Comput. Biol.* (2013). doi:10.1371/journal.pcbi.1003054

27. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).

28. Wang, J. & Wechsler-Reya, R. J. The role of stem cells and progenitors in the genesis of medulloblastoma. *Exp. Neurol.* **260**, 69–73 (2014).

29. Wetmore, C., Eberhart, D. E. & Curran, T. Loss of p53 but not ARF accelerates medulloblastoma in mice heterozygous for patched. *Cancer Res.* (2001).

30. Uziel, T. *et al.* The tumor suppressors Ink4c and p53 collaborate independently with Patched to suppress medulloblastoma formation. *Genes Dev.* (2005). doi:10.1101/gad.1368605

31. Taylor, M. D. *et al.* Mutations in SUFU predispose to medulloblastoma. *Nat. Genet.* (2002). doi:10.1038/ng916

32. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* (2014). doi:10.1016/j.ccr.2014.02.004

33. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48 (2012).

34. Roussel, M. F. & Stripay, J. L. Epigenetic Drivers in Pediatric Medulloblastoma. *Cerebellum* **17**, 28–36 (2018).

35. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

36. Cavalli, F. M. G. *et al.* Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754.e6 (2017).

37. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **487**, 49–56 (2012).

38. Li, J. Y. H. New regulatory interactions and cellular responses in the isthmic organizer region revealed by altering Gbx2 expression. *Development* (2005). doi:10.1242/dev.01727

39. Li, J. Y. & Joyner, A. L. Otx2 and Gbx2 are required for refinement and not induction of mid-hindbrain gene expression. *Development* (2001).

40. Morales, D. & Hatten, M. E. Molecular Markers of Neuronal Progenitors in the Embryonic Cerebellar Anlage. *J. Neurosci.* (2006). doi:10.1523/jneurosci.3493-06.2006

41. Roussel, M. F. & Hatten, M. E. *Cerebellum: Development and Medulloblastoma*. *Current Topics in Developmental Biology* **94**, (Elsevier Inc., 2011).

42. Okano-Uchida, T., Himi, T., Komiya, Y. & Ishizaki, Y. Cerebellar granule cell precursors can differentiate into astroglial cells. *Proc. Natl. Acad. Sci.* (2004). doi:10.1073/pnas.0307972100

43. Zhao, H., Ayrault, O., Zindy, F., Kim, J. H. & Roussel, M. F. Post-transcriptional down-regulation of Atoh1/Math1 by bone morphogenic proteins suppresses medulloblastoma development. *Genes Dev.* (2008). doi:10.1101/gad.1636408

44. Grimmer, M. R. & Weiss, W. A. BMPs oppose Math1 in cerebellar development and in medulloblastoma. *Genes and Development* (2008). doi:10.1101/gad.1657808

45. Oliver, T. G. *et al.* Transcriptional profiling of the Sonic hedgehog response: A critical role for N-myc in proliferation of neuronal precursors. *Proc. Natl. Acad. Sci.* (2003). doi:10.1073/pnas.0832317100

46. Ben-Arie, N. *et al.* Math1 is essential for genesis of cerebellar granule neurons. *Nature* (1997). doi:10.1038/36579

47. Fujita, S., Shimada, M. & Nakamura, T. H3-Thymidine autoradiographic studies on the cell proliferation and differentiation in the external and the internal granular layers of the mouse cerebellum. *J. Comp. Neurol.* (1966). doi:10.1002/cne.901280206

48. Rakic, P. Neuron-glia relationship during granule cell migration in developing cerebellar cortex. A Golgi and electonmicroscopic study in Macacus rhesus. *J. Comp. Neurol.* (1971). doi:10.1002/cne.901410303

49. Ben-Arie, N. *et al.* Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis. *Hum. Mol. Genet.* (1996). doi:10.1093/hmg/5.9.1207

50. Helms, A. W., Abney, A. L., Ben-Arie, N., Zoghbi, H. Y. & Johnson, J. E. Autoregulation and multiple enhancers control Math1 expression in the developing nervous system. *Development* (2000).

51. Lee, J. K. *et al.* Expression of neuroD/BETA2 in mitotic and postmitotic neuronal cells during the development of nervous system. *Dev. Dyn.* (2000). doi:10.1002/(SICI)1097-0177(200004)217:4<361::AID-DVDY3>3.0.CO;2-8

52. Pan, N., Jahan, I., Lee, J. E. & Fritzsch, B. Defects in the cerebella of conditional Neurod1 null mice correlate with effective Tg(Atoh1-cre) recombination and granule cell requirements for Neurod1 for differentiation. *Cell Tissue Res.* (2009). doi:10.1007/s00441-009-0826-6

53. Aruga, J. *et al.* The mouse Zic gene family: Homologues of the Drosophila pair-rule gene odd-paired. *J. Biol. Chem.* (1996). doi:10.1074/jbc.271.2.1043

54. Aruga, J. The role of Zic genes in neural development. *Molecular and Cellular Neuroscience* (2004). doi:10.1016/j.mcn.2004.01.004

55. Ayrault, O., Zindy, F., Rehg, J., Sherr, C. J. & Roussel, M. F. Two Tumor Suppressors, p27 Kip1 and Patched-1, Collaborate to Prevent Medulloblastoma . *Mol. Cancer Res.* (2009). doi:10.1158/1541-7786.mcr-08-0369

56. Furley, A. J. *et al.* The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. *Cell* (1990). doi:10.1016/0092-8674(90)90223-2

57. Goodrich, L. V., Milenković, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science (80-.).* (1997). doi:10.1126/science.277.5329.1109

58. Tao, R. *et al.* MYC drives Group 3 medulloblastoma through transformation of Sox2+ astrocyte progenitor cells. *Cancer Res.* (2019). doi:10.1158/0008-5472.can-18-1787

59. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. in *Cold Spring Harbor Symposia on Quantitative Biology* (2000).

60. Aguilera, A. & Gómez-González, B. Genome instability: A mechanistic view of its causes and consequences. *Nature Reviews Genetics* (2008). doi:10.1038/nrg2268

61. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* (2009). doi:10.1038/nature08467

62. Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* (2013). doi:10.3978/j.issn.2218-676X.2013.04.02

63. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* (2013). doi:10.1101/cshperspect.a012740

64. Lieber, M. R. The mechanism of human nonhomologous DNA End joining. *Journal of Biological Chemistry* (2008). doi:10.1074/jbc.R700039200

65. San Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annu. Rev. Biochem.* (2008). doi:10.1146/annurev.biochem.77.061306.125255

66. Jiricny, J. The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology* (2006). doi:10.1038/nrm1907

67. David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage.

Nature (2007). doi:10.1038/nature05978

68. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* (2001). doi:10.1038/35077232

69. Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W. & Harley, C. B. Telomere end-replication problem and cell aging. *J. Mol. Biol.* (1992). doi:10.1016/0022-2836(92)90096-3

70. Greider, C. W. Telomeres do D-loop-T-loop. *Cell* (1999). doi:10.1016/S0092-8674(00)80750-3

71. Bandaria, J. N., Qin, P., Berk, V., Chu, S. & Yildiz, A. Shelterin protects chromosome ends by compacting telomeric chromatin. *Cell* (2016). doi:10.1016/j.cell.2016.01.036

72. Artandl, S. E. *et al.* Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* (2000). doi:10.1038/35020592

73. León-Ortiz, A. M. *et al.* A Distinct Class of Genome Rearrangements Driven by Heterologous Recombination. *Mol. Cell* **69**, 292-305.e6 (2018).

74. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology* (2010). doi:10.1038/nrm2858

75. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* (1971).

76. Stratton, M. R., Campbell, P. J. & Andrew F, P. The cancer genome. *Nature* **458**, 719–724 (2009).

77. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* (2013). doi:10.1016/j.cell.2013.02.023

78. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic

event during cancer development. *Cell* (2011). doi:10.1016/j.cell.2010.11.055

79. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* (2012). doi:10.1016/j.cell.2011.12.013

80. Cleal, K., Norris, K. & Baird, D. Telomere length dynamics and the evolution of cancer genome architecture. *International Journal of Molecular Sciences* (2018). doi:10.3390/ijms19020482

81. Uringa, E. J., Youds, J. L., Lisaingo, K., Lansdorp, P. M. & Boulton, S. J. RTEL1: An essential helicase for telomere maintenance and the regulation of homologous recombination. *Nucleic Acids Res.* (2011). doi:10.1093/nar/gkq1045

82. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Research* (2016). doi:10.1093/nar/gkw079

83. Ding, H. *et al.* Regulation of murine telomere length by Rtel: An essential gene encoding a helicase-like protein. *Cell* (2004). doi:10.1016/j.cell.2004.05.026

84. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* (1989). doi:10.1093/nar/17.12.4713

85. Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. R. & White, M. F. The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains. **1**, 801–808 (2006).

86. Brieger, A. *et al.* Characterization of the nuclear import of human MutLα. *Mol. Carcinog.* (2005). doi:10.1002/mc.20081

87. Warbrick, E. The puzzle of PCNA's many partners. *BioEssays* (2002). doi:10.1002/1521-1878(200011)22:11<997::aid-bies6>3.3.co;2-r

88. Seshadri, N., Sandhu, S., Wu, X., Liu, W. & Ding, H. Generation of an Rtel1-CreERT2 knock-in mouse model for lineage tracing RTEL1+ stem cells during development. *Transgenic Res.* (2018). doi:10.1007/s11248-018-0093-y

89. Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H. & Boulton, S. J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* **149**, 795–806 (2012).

90. Vannier, J.-B. *et al.* RTEL1 Is a Replisome-Associated Helicase That Promotes Telomere and Genome-Wide Replication. *Science (80-.).* **342**, 239–242 (2013).

91. Deng, Z. *et al.* Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal–Hreidarsson syndrome. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1300600110

92. Barber, L. J. *et al.* RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination. *Cell* (2008). doi:10.1016/j.cell.2008.08.016

93. Zhuo, L. *et al.* hGFAP-cre transgenic mice for manipulation of glial and neuronal function in vivo. *Genesis* (2001). doi:10.1002/gene.10008

94. Guénet, J. L. The mouse genome The mouse genome. *Genome Res.* 1729–1740 (2005). doi:10.1101/gr.3728305

95. Smith, C. L., Blake, J. A., Kadin, J. A., Richardson, J. E. & Bult, C. J. Mouse Genome Database (MGD)-2018: Knowledgebase for the laboratory mouse. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1006

96. Bouabe, H. & Okkenhaug, K. Gene targeting in mice: A review. *Methods in Molecular Biology* (2013). doi:10.1007/978-1-62703-601-6_23

97. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for

effective therapy. *Nat. Med.* (2018). doi:10.1038/s41591-018-0014-x

98. Becher, O. J. & Holland, E. C. Genetically engineered models have advantages over xenografts for preclinical studies. *Cancer Research* (2006). doi:10.1158/0008-5472.CAN-05-3827

99. Kieran, M. W. Targeted treatment for sonic hedgehog-dependent medulloblastoma. *Neuro-Oncology* (2014). doi:10.1093/neuonc/nou109

100. F.J., S. *et al.* Pleiotropic role for MYCN in medulloblastoma. *Genes Dev.* **24**, 1059–1072 (2010).

101. Frappart, P.-O. *et al.* Recurrent genomic alterations characterize medulloblastoma arising from DNA double-strand break repair deficiency. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0806882106

102. Wu, X., Northcott, P. A., Croul, S. & Taylor, M. D. Mouse models of medulloblastoma. *Chin. J. Cancer* **30**, 442–449 (2011).

103. Briggs, K. J. *et al.* Cooperation between the Hic1 and Ptch1 tumor suppressors in medulloblastoma. *Genes Dev.* (2008). doi:10.1101/gad.1640908

104. Yang, Z. J. *et al.* Medulloblastoma Can Be Initiated by Deletion of Patched in Lineage-Restricted Progenitors or Stem Cells. *Cancer Cell* (2008). doi:10.1016/j.ccr.2008.07.003

105. Lee, Y. *et al.* Loss of suppressor-of-fused function promotes tumorigenesis. *Oncogene* (2007). doi:10.1038/sj.onc.1210467

106. Hallahan, A. R. *et al.* The SmoA1 mouse model reveals that notch signaling is critical for the growth and survival of Sonic Hedgehog-induced medulloblastomas. *Cancer Res.* (2004). doi:10.1158/0008-5472.CAN-04-1813

107. Hatton, B. A. *et al.* The Smo/Smo model: Hedgehog-induced medulloblastoma with 90%

incidence and leptomeningeal spread. *Cancer Res.* (2008). doi:10.1158/0008-5472.CAN-07-5092

108. Kawauchi, D. *et al.* A Mouse Model of the Most Aggressive Subgroup of Human Medulloblastoma. *Cancer Cell* **21**, 168–180 (2012).

109. Wang, Z. Q. *et al.* Mice lacking ADPRT and poly(ADP-ribosyl)ation develop normally but are susceptible to skin disease. *Genes Dev.* (1995). doi:10.1101/gad.9.5.509

110. Yan, C. T. *et al.* XRCC4 suppresses medulloblastomas with recurrent translocations in p53-deficient mice. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.0601938103

111. Marino, S., Vooijs, M., Van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev.* (2000).

112. Lee, Y. & McKinnon, P. J. DNA ligase IV suppresses medulloblastoma formation. *Cancer Res.* (2002).

113. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

114. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq431

115. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv007

116. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

117. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* (2005).

doi:10.1073/pnas.0506580102

118. Tamayo, P. *et al.* Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0701068104

119. Schüller, U. *et al.* Acquisition of Granule Neuron Precursor Identity Is a Critical Determinant of Progenitor Cell Competence to Form Shh-Induced Medulloblastoma. *Cancer Cell* (2008). doi:10.1016/j.ccr.2008.07.005

120. van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* (2014).

121. Hand, D. J. & Till, R. J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Mach. Learn.* (2001). doi:10.1023/A:1010920819831

122. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

123. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

124. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp616

125. Van der Auwera, G. A. *et al.* GATK Best Practices. *Curr. Protoc. Bioinformatics* (2002). doi:10.1002/0471250953

126. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013). doi:10.1093/bioinformatics/bts635

127. Haas, B. J. *et al.* STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv* (2017). doi:10.1101/120295

128. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* (2011). doi:10.1038/ng.806

129. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btp698

130. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010). doi:10.1093/nar/gkq603

131. Tokheim, C., Papadopoulis, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the Evaluation of Cancer Driver Genes. *bioRxiv* **113**, 060426 (2016).

132. Chi, C., Ajwad, R., Kuang, Q. & Hu, P. A novel graph-based algorithm to infer recurrent copy number variations in cancer. *Cancer Inform.* (2016). doi:10.4137/CIN.S39368

133. Gentleman, R. & Vandal, A. C. Computational algorithms for censored-data problems using intersection graphs. *J. Comput. Graph. Stat.* (2001). doi:10.1198/106186001317114901

134. Schwalbe, E. C. *et al.* Minimal methylation classifier (MIMIC): A novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. *Sci. Rep.* (2017). doi:10.1038/s41598-017-13644-1

135. Schwalbe, E. C. *et al.* DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathol.* (2013). doi:10.1007/s00401-012-1077-2

136. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and

Mutations. *Cell* (2018). doi:10.1016/j.cell.2018.02.060

137. Coudray, A., Battenhouse, A. M., Bucher, P. & Iyer, V. R. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. *PeerJ* (2018). doi:10.7717/peerj.5362

138. Nakagawa, H. & Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Science* (2018). doi:10.1111/cas.13505

139. RYAN, K. M. & BIRNIE, G. D. Myc oncogenes: the enigmatic family . *Biochem. J.* (2015). doi:10.1042/bj3140713

140. DePinho, R. *et al.* Myc family of cellular oncogenes. *J. Cell. Biochem.* **33**, 257–266 (1987).

141. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* (2003). doi:10.1038/ng1180

142. Markey, G. M. Carboxylesterase 1 (Ces1): From monocyte marker to major player. *Journal of Clinical Pathology* (2011). doi:10.1136/jcp.2010.084657

143. Huang, G. H. *et al.* Medulloblastoma stem cells: Promising targets in medulloblastoma therapy. *Cancer Sci.* **107**, 583–589 (2016).

144. Calado, R. T. & Dumitriu, B. Telomere dynamics in mice and humans. *Semin. Hematol.* (2013). doi:10.1053/j.seminhematol.2013.03.030

145. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci.* (2014). doi:10.1073/pnas.1413624111

146. KIJIMA, N. & KANEMURA, Y. Molecular Classification of Medulloblastoma. *Neurol. Med. Chir. (Tokyo).* **56**, 687–697 (2016).

147. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb. Perspect.*

Med. **3**, (2013).

148. Rickman, D. S., Schulte, J. H. & Eilers, M. The Expanding World of N-MYC–Driven Tumors. *Cancer Discov.* **8**, 150 LP – 163 (2018).