

Characterizing the transcriptional impact of histone H3 mutations on cell identity in cancer and development

Nisha Kabir

Department of Human Genetics
Faculty of Medicine and Health Sciences
McGill University, Montreal, Canada
April 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree
of Master of Science Human Genetics

© Nisha Kabir, 2022

Abstract

Histone proteins have vital roles in the regulation of gene expression during development, where dysregulation of this processes results in diseases such as cancer and neurological disorders. Pediatric high-grade glioma (pHGG) is a highly aggressive cancer with a two-year survival rate of less than 30%. With the advent of whole genome sequencing, recurrent somatic mutations in histone H3 proteins and common co-occurring partner mutations have been identified to drive a subset of cancer. Histone H3 lysine 27 to methionine (H3K27M) cause a subset of pHGGs occurring in the brain midline in younger children, while mutations in glycine 34 to arginine or valine (G34R/V) cause another subset of pHGGs occurring in the cerebral cortex of adolescents and young adults. Histone H3 mutants also drive cancer outside of the brain, where mutations in glycine 34 to tryptophan (G34W) cause over 90% of giant cell tumors of bone (GCTB) in adults. The restricted pattern of histone mutations with anatomical location and age of incidence in brain tumors gave rise to the hypothesis that histone mutations only occurring in a specific developmental window and cellular context give rise to cancer. While the cellular context is known for these tumors, how the mutations subsequently impact cell identity along differentiation are not well characterized.

In this thesis, I study the transcriptomic effect of H3 and partner mutations on cellular identity using in silico integrative analysis of genome-wide transcriptomic data from model systems and normal references of the developing brain at single-cell resolution. I will use a variety computational methods to map genome-wide transcriptomic data to normal references to infer changes to cell identity in disease states. First, I study the role of partner mutations in a kinase, *ACVR1*, in the context of mutant H3 K27M. Gene expression profiling shows that mutations in *ACVR1* impact cellular identity in distinct ways depending on the histone context. Second, I study the effect of H3.3 K27M mutations using cell lines subjected to differentiation experiments and orthotopic xenograft models using an isogenic system. *In vivo* K27M-KO experiments show a difference in cell state towards oligodendrocyte and astrocyte lineages, while *in vitro* K27M-KO experiments show upregulation of extracellular matrix and cellular adhesion gene programs. Third, I study the transcriptomic effects of germline heterozygous H3.3 G34R/V/W mutations in murine models. Germline G34 mutants have surprising tissue

specificity; G34R mutants cause progressive neuronal loss coupled to increased immune infiltration in the brain, while G34W mutants cause an unhealthy obesity phenotype with decreased expression of PPARG, a master regulator of adipocyte differentiation. Together, this work provides insight into changes in cell identity and neurological abnormalities induced by H3 mutations at the level of transcription. This represents an important step in characterizing the cellular contexts of these disease states for the development of therapeutics.

Résumé

Les histones jouent un rôle essentiel dans la régulation de l'expression des gènes au cours du développement alors que la dérégulation de ces processus entraîne des maladies telles que le cancer et des troubles neurologiques. Le gliome pédiatrique de haut grade (pHGG) est un cancer très agressif dont le taux de survie après deux ans est inférieur à 30 %. Grâce à la technologie du séquençage du génome entier (WGS), des mutations somatiques récurrentes dans les histones H3 en combinaison avec des mutations partenaires ont été identifiées comme étant à l'origine d'un sous-ensemble de cancers. Les mutations de la lysine 27 en méthionine de l'histone H3 (H3K27M) sont à l'origine d'un sous-ensemble de pHGG survenant dans la ligne médiane du cerveau chez les jeunes enfants, tandis que les mutations de la glycine 34 en arginine ou en valine (G34R/V) sont à l'origine d'un autre sous-ensemble de pHGG survenant dans le cortex cérébral des adolescents et des jeunes adultes. D'autres mutants de l'histone H3 provoquent également des cancers en dehors du cerveau; les mutations de la glycine 34 en tryptophane (G34W) sont à l'origine de plus de 90% des tumeurs osseuses à cellules géantes (GCTB) chez les adultes. Le schéma restreint des mutations des histones en fonction de la localisation anatomique et de l'âge d'incidence des tumeurs cérébrales a donné lieu à l'hypothèse selon laquelle les mutations des histones ne se produisant que dans une fenêtre de développement et un contexte cellulaire spécifiques donnent lieu à un cancer. Bien que le contexte cellulaire soit connu pour ces tumeurs, comment les mutations impactent l'identité cellulaire durant la différenciation n'est pas bien caractérisé.

Dans cette thèse, j'étudie l'effet transcriptomique des mutations de H3 et de ses partenaires sur l'identité cellulaire en utilisant des analyses intégratives in silico de données transcriptomiques du génome entier provenant de systèmes de modèles et de références normales du cerveau en développement à une résolution de cellules individualisées. J'utiliserai diverses méthodes informatiques pour cartographier les données transcriptomiques à l'échelle du génome entier et les références normales afin de déduire les changements d'identité cellulaire dans un contexte de maladies. Tout d'abord, j'étudie le rôle des mutations partenaires dans une kinase, ACVR1, dans le contexte du mutant H3 K27M. Le profilage de l'expression des

gènes montre que les mutations dans ACVR1 ont un impact sur l'identité cellulaire de manière distincte selon l'état de l'histone. Ensuite, j'étudie l'effet des mutations H3.3 K27M en utilisant des lignées cellulaires soumises à des expériences de différenciation et des modèles de xénogreffes orthotopiques utilisant un système isogénique. Les expériences K27M-KO in vivo montrent une différence d'état cellulaire vers les lignées d'oligodendrocytes et d'astrocytes, tandis que les expériences K27M-KO in vitro montrent une régulation positive des programmes de gènes de la matrice extracellulaire et de l'adhésion cellulaire. Troisièmement, j'étudie les effets transcriptomiques des mutations germinales hétérozygotes H3.3 G34R/V/W dans des modèles de souris. Les mutants de lignée germinale G34 ont une spécificité tissulaire surprenante ; les mutants G34R provoquent une perte neuronale progressive couplée à une infiltration immunitaire accrue dans le cerveau, tandis que les mutants G34W provoquent un phénotype d'obésité malsaine avec une diminution de l'expression de PPARG, un régulateur principal de la différenciation des adipocytes. Ensemble, ces travaux permettent de mieux comprendre les changements d'identité cellulaire et les anomalies neurologiques induits par les mutations de H3 au niveau de la transcription. Cela représente une étape importante dans la caractérisation des contextes cellulaires de ces maladies pour le développement de thérapies.

Table of Contents

Abstract.....	2
Résumé	4
Table of Contents.....	6
List of Abbreviations	8
List of Figures	9
List of Tables	10
Acknowledgements.....	11
Preface	12
Format of the thesis.....	12
Contribution of the authors	12
Chapter 1: Introduction	13
1.1 Introduction to histone mutant pediatric high grade gliomas	13
Chapter 1.2 Histone H3 in Development	14
1.2.1 Introduction to Histone Modifications	14
1.2.2 H3K27 methylation	16
1.2.3 H3K36 methylation	17
1.2.4 DNA methylation.....	18
1.3 Dysregulation of H3 and partner mutations in cancer and human disease	19
1.3.1 H3K27M mutations in pHGGs	19
1.3.2 ACVR1 partner mutations in H3K27M pHGGs	20
1.3.3 H3.3 G34 mutations in cancer and neurodevelopmental syndromes	22
1.4 Hypothesis and Aims.....	25
Chapter 2: Materials and Methods.....	26
2.1. Analysis of bulk RNA seq data.....	26
2.2. Analysis of single-cell RNA seq data.....	27
Chapter 3: Results.....	30
3.1: Transcriptomic effect of ACVR1 in H3.1 and H3.3 K27M mutant tumors in isogenic cell lines.....	30
3.1.1. ACVR1 impacts global transcriptome in H3.3 K27M, to a lesser degree in H3.1 K27M	30
3.1.2. ACVR1 induces switch in glial cell state in H3.3 K27M, not in H3.1 K27M	36
Chapter 3.2: Transcriptomic effects of H3.3 K27M in experimental model systems	40
3.2.1. H3.3 K27M-KO show upregulation of extracellular matrix gene programs <i>in vitro</i>	40
3.3.2 H3.3 K27M-KO map oligodendrocytes and astrocytes <i>in vivo</i>	55

3.3. Transcriptomic effects of germline G34 mutations in a murine model.....	63
3.3.1. Early postnatal G34 mutants have minor transcriptomic changes in the brain	63
3.3.2. Juvenile G34R have transcriptionally normal astrocytes.....	70
3.3.3. Adult G34 mutants have disease-associated immune response in the brain	75
3.3.4. Adult G34R have cortical neuronal loss	80
3.3.5. Adult G34R mice share transcriptional overlap with Rett syndrome patients	83
3.3.6. G34W have an unhealthy obesity phenotype at 10W	86
Chapter 4: Discussion.....	90
4.1. The role of ACVR1 on cell state is dependent on histone context	90
4.2. Effect H3.3 K27M mutations in <i>in vitro</i> and <i>in vivo</i> model systems	93
4.3. Lineage specificity of germline <i>H3f3a</i> G34R/V/W mutations	95
Chapter 5: Conclusion and Future Directions	99
5.1. Conclusion.....	99
5.2. Future Directions	100
Chapter 6: References.....	103

List of Abbreviations

ACVR1; Activin A receptor type 1

BMP; Bone morphogenic protein

CRISPR; Clustered Regularly Interspaced Short Palindromic Repeats

Cas9; CRISPR Associated Protein 9

CpG; 5'-Cytosine-Phosphate-Guanine-3'

DIPG; Diffuse intrinsic pontine glioma

DEG; Differentially expressed gene

GSEA; Gene set enrichment analysis

GCTB; Giant cell tumor of bone

GO; Gene ontology

LOESS; Locally estimated scatterplot smoothing

MAPQ; Mapping quality

MSC; Mesenchymal stem cell

NPC; Neural progenitor cell

OPC; Oligodendrocyte precursor cell

PCA; Principal component analysis

PC1; Principal component 1

PC2; Principal component 2

pHGG; Pediatric high-grade glioma

PTM; Post-translational modification

RNA seq; RNA sequencing

scRNA seq; single-cell RNA sequencing

ssODN; Single strand donor nucleotides

tSNE; *t*-distributed stochastic neighbor embedding

UMAP; Uniform manifold approximation and projection

UMI; Unique molecular identifier

WGBS; whole genome bisulfite sequencing

List of Figures

Figure 1: ACVR1 mutation status is secondary to H3 K27M	31
Figure 2: ACVR1 has heterogeneous effects in driving transcriptomic changes in a cell-line specific manner	32
Figure 3: ACVR1 status results in large changes in H3.3, but not H3.1, K27M cell lines	33
Figure 4: Pathway analysis shows terms related to differentiation in H3.3 K27M cell lines.....	35
Figure 5: ACVR1 blocks astrocytic maturation in H3.3 K27M	37
Figure 6: Removal of ACVR1 does not induce astrocytic maturation in H3.1 K27M	39
Figure 7: Media has a greater effect than H3.3 K27M status on global gene expression	43
Figure 8: ECM programs upregulated by media	44
Figure 9: Developmental and cellular adhesion pathways induced by media for pontine H3.3 K27M and H3.3 K27M-KO cell line	46
Figure 10: Developmental pathways induced by media for thalamic H3.3 K27M and H3.3 K27M-KO cell lines.....	47
Figure 11: Enrichment of non-neuroectodermal gene programs in differentiation media	49
Figure 12: Non-neuroectodermal cell state induced by differentiation media.....	52
Figure 13: Non-neuroectodermal cell states mask astrocyte gene programs in differentiation media	54
Figure 14: Genotype drives most source of variation in xenografts derived from thalamic cell lines	58
Figure 15: Developmental terms are enriched in K27M-KO xenografts.....	58
Figure 16: K27M-KO cells express astrocyte and oligodendrocyte gene programs in xenograft models ..	60
Figure 17: Glial cell heterogeneity of K27M and K27M-KO cells in xenograft models	62
Figure 18: Germline knock-in of H3.3 G34R/V/W mice result in tissue specific defects.....	65
Figure 19: G34R/V/W mutation does not drive major source of variation in gene expression one week post birth.....	67
Figure 20: G34R/V/W mutation induces few transcriptomic changes at one week post birth.....	69
Figure 21: G34R/V/W mutations display changes related to signaling in the cortex at one week post birth	69
Figure 22: G34R and wild type cortex and cerebellum samples have similar composition of cell types ...	71
Figure 23: G34R and wildtype astrocytes are transcriptionally similar and arise from spatially distinct populations	73
Figure 24: G34R and wild type astrocytes form distinct clusters	74
Figure 25: G34 mutant status drives major changes in immune and astrocyte gene expression in the cortex and cerebellum at 10 weeks	76
Figure 26: Upregulation of immune signaling in G34 mutant cortex and cerebellum	77
Figure 27: G34R display disease associated microglial signature in the cortex at 10 weeks	79
Figure 28: G34R mutants display neuronal loss.....	81
Figure 29: G34 mutants do not display pan-neuronal loss in cerebellum at 10 weeks.....	82
Figure 30: Enrichment of G34R signatures in cortex of Rett syndrome patients	85
Figure 31: G34 mutants display increased immune signaling in white adipose tissue by 10 weeks.....	87
Figure 32: G34 mutants display transcriptomic changes consistent with an unhealthy fat phenotype	89

List of Tables

Table 1: RNA seq cohort for ACVR1 cell lines	30
Table 2: RNA seq cohort for H3.3 K27M serum differentiation experiments	40
Table 3: scRNA seq cohort for H3.3 K27M serum differentiation experiments	49
Table 4: RNA seq cohort for H3.3 K27M orthotopic xenograft models.....	56
Table 5: scRNA seq cohort for H3.3 K27M orthotopic xenograft models	60

Acknowledgements

I would like to express my thanks and gratitude first and foremost to my supervisor Dr. Claudia Kleinman, for her unwavering support, mentorship, and discussions. Claudia, thank you for taking a chance on me and giving me such a wonderful opportunity to grow as a scientist in your laboratory.

I would like to extend my appreciation to my supervisory committee members, Dr. Livia Garzia and Dr. Jacek Majewski, for their guidance throughout the years.

I would like to express my gratitude to all the people in the laboratories of Dr. Claudia Kleinman and Dr. Nada Jabado with whom I had the pleasure to work. Marie, Nicolas, Steven, Selin, Samantha, Yang, Maria, Zaza, and Simon from the Kleinman lab; Sima, Carol, Pariya, Shakour, and Brian from Jabado lab; thank you for the wonderful discussions and support in exploring science and learning together. Finally, I would like to extend my gratitude towards my friends and family for their support.

Preface

Format of the thesis

This thesis is written in the format of a traditional thesis following the guidelines of the Human Genetics Department at McGill University. The format of this thesis is divided into five chapters.

Chapter 1 introduces histone H3 and relevant co-occurring mutations in the context of pediatric high grade glioma (pHGG) and neurodevelopmental syndromes; provides background on known mechanisms and functions in these disease states; and provides the objectives. **Chapter 2** consists of materials and methods used in the body of the text. **Chapter 3.1** deciphers transcriptomic effects of a co-occurring kinase mutation in histone H3 K27M pHGGs and finds the functions differ depending if the K27M mutation is in canonical or variant H3. **Chapter 3.2** further explores the transcriptomic effects of K27M in variant H3 through the use of *in vitro* and *in vivo* experiments and finds divergence in gene programs upregulated by experimental model. **Chapter 3.3** shifts gears to characterize the transcriptomic effects of germline heterozygous G34R/V/W mutations in development using a murine model and finds a neurodegenerative phenotype characterized by inflammation specific to G34R mutants. **Chapter 4** discusses the findings and limitations of the studies, while **Chapter 5** provides a conclusion.

Contribution of the authors

The work described in this thesis was performed under the supervision of Dr. Claudia Kleinman in collaboration with the laboratory of Dr. Nada Jabado.

I lead all transcriptomic analysis in **Chapter 3.1** with the help of Steven Hebert for quality control and genomic alignment of raw data. Generation of CRISPR-Cas9 edited cell lines and experimental characterization was done by Abdulshakour Mohammadnia, Caterina Russo, Elias Jabbour, Michele Zeineih, and Jing Zhou. In **Chapter 3.2**, I lead the transcriptomic analysis while generation of CRISPR-Cas9 edited cell lines, differentiation experiments, patient-derived orthotopic xenograft models and experimental characterization was done by Brian Krug. In **Chapter 3.3**, I lead the transcriptomic analysis while generation of the murine mouse model and experimental characterization was done by Sima Khazaei. Carol Chen lead the epigenetic analysis while Pariya Azarafshar contributed to genotyping. BioID experiments were performed by the lab of Dr. Eric Campos (University of Toronto).

Chapter 1: Introduction

1.1 Introduction to histone mutant pediatric high grade gliomas

The central nervous system is one of the most common sites of cancer in children with an average incidence of 5.65 cases per 100,000 individuals¹. Cancers in the central nervous system are one of the leading causes of cancer-related death in children aged 14 years or less. Unlike other cancers which are classified into stages, tumors of the central nervous system are classified into grades which are essential for determining prognosis and treatment²⁻⁴. Grading of tumors is split into two large categories previously based on histology, low grade (I and II) or high grade (III, IV). High grade tumors are comprised of a heterogeneous set of specific cancers but are characterized by nuclear atypia, brisk mitotic activity (grade III), and necrotic areas (grade IV). Within high grade tumors, pediatric glioblastoma (GBM) and diffuse intrinsic pontine glioma (DIPG) have an especially poor prognosis with a two-year survival rate of less than 10%⁵. Despite advances in the field, standard treatment consists of maximal surgical resection followed by a combination of radio- and chemotherapy⁶, aggressive therapies that leave a long-lasting series of side effects which severely affect the quality of life of the few surviving patients. Therefore, a deeper molecular understanding of the disease is required for the development of effective therapeutics.

With the advent of genomic profiling, tumor classification moved away from a solely histology-based characterization and towards a molecular and histologic characterization. Molecular markers divide tumors into distinct biological groups which are prognostically and clinically significant^{5,7-16}. Notably, mutations in histone genes drive a subset of pediatric and adult cancers including pediatric high grade gliomas (pHGGs), giant cell tumor of bone (GCTB), and chondroblastoma^{7-9,14,17}. In rare cases, mutations in histone genes have been shown to promote tumorigenesis in undifferentiated sarcoma, head and neck squamous cell carcinoma, and acute myeloid leukemia¹⁸⁻²¹.

In pHGGs, recurrent somatic mutations at specific residues in histone genes segregate with anatomical location and age of incidence. Given this pattern of occurrence and the normal function of histones in development, it is thought that these mutations only cause brain tumors

if they occur in a specific developmental window and cellular context^{22–25}. Mutations at lysine 27 to methionine (K27M) occurring in *H3F3A* (encoding H3.3) and less commonly in *HIST1H3B* or *HIST1H3C* (encoding H3.1, H3.2, respectively) define a subset of high-grade glioma and DIPG occurring along the midline in young children; where H3.3 K27M mutations drive approximately 63% of DIPGs and approximately 60% of non-brainstem midline tumors^{5,9,14,26}. Mutations in glycine 34 to arginine (G34R) or less commonly valine (G34V) in *H3F3A* define approximately 16% of high-grade gliomas occurring in the cerebral cortex (primarily temporal and parietal lobes) in adolescents and young adults. Glycine 34 to tryptophan (G34W) in *H3F3A* define above 90% of GCTB occurring in adults¹⁷. Mechanistically, these mutations result in distinct epigenetic landscapes which promote tumorigenesis through global dysregulation of chromatin marks^{27–30}. pHGGs are thought to arise from a stem-like ‘cell of origin’ characterized by a primed epigenetic state which mirror cell types from the developing brain^{29,31–34}. The occurrence of histone and partner mutations in this particular cellular context is thought to promote tumorigenesis^{8,23,24,35}. However, how the mutations impact cellular differentiation toward terminal lineages is poorly understood. Moving forward, insight on how cell identity and differentiation is impacted by the mutations in a specific cellular context may provide insight onto potential molecular dependencies and thus aid in the development of therapeutics.

In this chapter, I will first introduce concepts related to normal functions of histone proteins and known roles associated with development in **Chapter 1.2**. Next, in **Chapter 1.3** I will provide background on how specific histone (H3K27M, H3G34R/V/W) and partner (*ACVR1*) mutations work to promote oncogenesis in the context of pHGGs, and GCTB. The role of histone mutations in neurodevelopmental disorders will be briefly addressed. Finally, **Chapter 1.4** outlines the hypothesis and aims addressed in the body of this thesis.

Chapter 1.2 Histone H3 in Development

1.2.1 Introduction to Histone Modifications

Gene expression is regulated in the cell through heritable factors which do not alter DNA sequence but influence chromatin structure^{36,37}. The basic repeating unit of chromatin is a nucleosome which consists of 146-147 base pairs (bp) of DNA wrapped around a histone

octamer^{38,39}. The histone octamer consists of two copies of H2A, H2B, H3 and H4 which form a globular core that DNA is wrapped around. Importantly, the histone amino-terminal (N-terminal) tails are free in space. Repeating nucleosomes along the genome are condensed into dense arrays through linker H1 proteins and further condensed into chromatin fiber^{40,41}. The degree of condensation determines the accessibility to transcriptional machinery in the cell. There are two large categories of chromatin; 'euchromatin' which is less condensed and transcriptionally active, and 'heterochromatin' which is highly condensed and generally transcriptionally silent. Heterochromatin is further subdivided into two categories; facultative heterochromatin which has the potential to be transcriptionally active, and obligative heterochromatin which is always transcriptionally silent. Chromatin state is a blueprint for cellular identity and is regulated over the course of development in a cell-cycle and tissue-specific manner⁴²⁻⁴⁴. Key players in regulating chromatin state are histone modifications^{43,44}.

Histone tails serve as the site of chemical post-translational modifications (PTMs) which have essential roles in remodeling chromatin to mediate DNA transcription, replication, recombination, and response to damage⁴⁵. Histone PTMs promote or repress gene expression depending on the specific mark deposited, histone residue, and cross-talk interactions with other histone marks. Examples of histone PTMs include but are not limited to acetylation, methylation, phosphorylation, and ubiquitination. The majority of histone PTMs are dynamically regulated and are deposited by 'writer' complexes, their function mediated through recognition by 'reader' complexes, and removed by 'eraser' complexes^{46,47}. Writer complexes can recognize a variety of histone residues but deposit a single type of chemical mark⁴⁴. Histone modifications enact their function through two known mechanisms. One mechanism is that the PTM modifies the chemical charge of the nucleosome such that interactions with DNA or nearby nucleosomes are disrupted. A second mechanism is through recognition of the PTM by reader complexes which function in promoting repair, transcription, or chromatin condensation.

Variants of histone proteins play a key role in chromatin remodeling during development and cellular differentiation⁴⁸. In mammals, variants of all histone proteins except H4 exist. Several properties of histone variants mediate their specific functionalities in the cell, including

differences in protein sequence, deposition in the genome, cell cycle- and tissue-specificity. In particular, variant H3 (H3.3) plays a key role in early embryogenesis including maintenance of genome integrity^{49,50}. H3.3 expression is required for the oocyte, and complete depletion leads to early embryonic lethality. Depletion of H3.3 leads to disruption of heterochromatin structures, mitotic defects, and disruption neural stem cell differentiation^{50,51}. With age, H3.3 accumulates in neurons where it constitutes over 90% of total H3 in post-mitotic neurons by adulthood^{52,53}. Over the course of life, H3.3 expression in neurons and glial cells contributes to regulation of cell-identity gene programs.

H3.3 differs from canonical histone H3 in a variety of ways which contribute towards its specialized functions. Canonical histone H3 is encoded by genes located in multiple gene clusters along the genome and consists of H3.1 and H3.2 proteins which differ by a single amino acid^{48,54}. Canonical histones are deposited in a cell cycle dependent manner during DNA replication. In contrast, H3.3 is encoded by two genes in mammals, *H3F3A* and *H3F3B*, and deposited in a replication and cell cycle-independent manner. H3.3 differs in sequence from H3.1 and H3.2 in four to five amino acids, respectively^{48,55}. It is these differences in sequence which contribute to the differential recognition of histone H3.3 by chaperone proteins which deposit H3.3 in specific locations in the genome. H3.3 deposited by the HIRA chaperone complex is associated with actively transcribed gene bodies, while deposition by the ATRX chaperone complex is associated with heterochromatin⁵⁶. The differing role of H3.3 in transcription is due to the specific patterns of deposited chromatin marks⁵⁰. Among different chromatin marks, lysine methylation does not change the chemical charge of the nucleosome, but functions through recognition by downstream effector proteins.⁴³. Due to its dysregulation in pHGGs, lysine methylation is discussed in more detail in the following subsections.

1.2.2 H3K27 methylation

Lysine 27 is located on the N-terminal tail of H3 and serves as a site of methylation⁵⁷. It is dynamically regulated, where lysine 27 is methylated by histone methyltransferases and demethylated by histone demethylases. Although H3K27me3 is canonically associated with gene silencing in heterochromatin, H3K27 methylation exists in different degrees of mono-, di- and trimethylated states (H3K27me1, H3K27me2, H3K27me3, respectively) which correspond

to functional roles in activating transcription (H3K27me1) or repressing transcription (H3K27me2/3)^{58,59}. H3K27 methylation is established and propagated in distinct genomic regions by the PRC2 complex which contains core subunits EED, SUZ12, AEBP2 and one of the mutually exclusive catalytic subunits EZH1 and EZH2. H3K27me1 is catalyzed by PRC2 in gene bodies and is associated with active transcription^{59,60}. The conversion of H3K27me1 to H3K27me2 is regulated through crosstalk with H3K36 methylation. In contrast to the activating role of H3K27me1, H3K27me2/3 are progressively catalyzed by PRC2 and associated with transcriptional repression. H3K27me2 forms broad domains in both inter- and intragenic regions and is the most abundant methylation state, accounting for over 70% of the total methylated H3K27 pool. It is associated with lowly transcribed genes and thought to prevent activation of non-cell type specific gene programs. H3K27me3 is found in broad domains in the genome and functions to repress transcription^{59,61}. It is propagated in the genome to form broad domains through a positive feedback loop where H3K27me3 is recognized as a substrate and functions to allosterically activate the EED domain of PRC2, causing increased deposition of H3K27me3 by PRC2⁶².

Dynamic regulation of H3K27 methylation in facultative chromatin is critical for normal development and cellular differentiation. In human embryonic stem cells, H3K27 methylation targets developmentally regulated genes including pluripotency markers and is critical for differentiation^{61,63,64}. In combination with the transcriptionally activating mark of H3K4 methylation, it marks promoters of developmentally regulated genes and serves to control lineage-specific gene programs⁶⁵. It is therefore perhaps unsurprising that dysregulation of H3K27 methylation results in disease.

1.2.3 H3K36 methylation

In contrast to the repressive transcriptional function of H3K27 methylation, methylation on lysine 36 (H3K36) is associated with transcriptional activation^{66,67}. H3K36me3 promotes transcription through regulating the spread of repressive H3K27me3 and functions in silencing transcription of cryptic intragenic regions. H3K36 methylation also exists in different degrees of mono-, di-, and trimethylated states (H3K36me1, H3K36me2, H3K36me3) which correspond to functional states. However, unlike H3K27me3 methylation which is solely deposited by PRC2,

H3K36 methylation is deposited by multiple enzymes containing a catalytic SET domain which separate into two major classes: mono- and dimethylases and trimethylases. H3K36 mono- and dimethylases include NSD family proteins (NSD1, NSD2, and NSD3) and ASH1L⁶⁷⁻⁷¹. H3K36me1 has no clear functional role, while H3K36me2 forms large intergenic domains in euchromatin⁷². In intergenic regions H3K36me2 serves to recruit DNA methylase Dnmt3a to promote transcriptional repression and represses the conversion of H3K27me2 from H3K27me3^{73,74}. While there are multiple mono- and dimethylases, SETD2 and SETD5 are the only known enzymes to catalyze H3K36me3^{75,76}. H3K27me3 is co-transcriptionally deposited in actively transcribed genes through interaction of SETD2 with RNA polymerase II^{67,77}. H3K36me3 serves to recruit DNA methylase Dnmt3b through a PWWP reader domain to promote transcriptional repression of cryptic intragenic regions⁷⁸. Due to its crosstalk with H3K36 methylation, DNA methylation is briefly discussed below.

1.2.4 DNA methylation

DNA methylation occurs on cytosine residues primarily in the context of a cytosine followed by a guanine nucleotide (CpG)⁷⁹. A small minority of cytosine methylation occurs in non-CpG sequences (mCH, where H denotes A, C, T) found in postnatal neurons⁸⁰. Patterns of DNA methylation are established during development in a dynamic process through DNA methylases and demethylases where they serve to regulate cell-identity programs in differentiated cells, maternal- or paternal-allele specific expression through genomic imprinting, silence repeat element expression, and function in X-chromosome inactivation⁷⁹⁻⁸¹. In the CpG context, methylation at genic promoters functions to repress gene expression through inhibition of transcription factor binding. In intragenic regions, CpG methylation is associated with repression of spurious transcription initiation. In the non-CpG context, mCH methylation characterizes specific neuronal subtypes and is associated with transcriptional repression in gene bodies^{82,83}.

DNA methylation in both contexts is catalyzed by the Dnmt family of DNA methyltransferases. *De novo* methylation patterns are established in the genome by Dnmt3a or Dnmt3b, while methylation is maintained at CpG sites by Dnmt1 and at mCH sites by Dnmt3a^{84,85}. Although both Dnmt3a and Dnmt3b are capable of establishing DNA methylation, there are differences in

tissue expression and recruitment. Dnmt3a is ubiquitously expressed and recruited by H3K36me₂, while Dnmt3b expressed in testis, thyroid and bone marrow and recruited by H3K36me₃^{73,78,86}. DNA is demethylated through the ten–eleven translocation (TET) family of enzymes⁷⁹. Dysregulation of DNA methylases or downstream effector proteins are implicated in neurological syndromes.

1.3 Dysregulation of H3 and partner mutations in cancer and human disease

1.3.1 H3K27M mutations in pHGGs

Mutation of lysine 27 to methionine in histone H3 (H3K27M) drives a subset of pHGGs including 80% of deadly diffuse intrinsic pontine glioma (DIPG)^{5,8–10,14,26}. A common partner mutation in H3K27M tumors is loss of TP53, which functions to suppress tumor formation through a variety of mechanisms including regulation of cellular division, DNA repair in response to damage or stress, and inducement of apoptosis in response to irreversible cellular damage^{9,26,87}. H3K27M mutations induce DNA hypomethylation and global reduction of H3K27me₂ and H3K27me₃ levels in a genome-wide manner^{25,88–90}. While H3K27M mutated nucleosomes constitute a minority of the total H3 pool, H3K27M mutations act in a dominant-negative manner to inhibit PRC2 activity through interaction with the catalytic EZH2 subunit^{25,88}. How PRC2 inhibition leads to global loss of H3K27me_{2/3} is still the subject of debate. Several mechanistic models have been proposed: recruitment and retention of PRC2 by H3K27M mutated nucleosomes resulting in inhibition of H3K27me_{2/3} deposition^{25,89,91}; retention of PRC2 at strong PRC2 targets⁹²; PRC2 interactions with H3K27M leading to partly irreversible inhibition⁹³; impaired genomic spreading of H3K27me₃ with unaltered PRC2 recruitment and propagation⁹⁰; and PRC2 inhibition leading to ‘step-down’ reduction of one degree of K27M methylation⁹⁴.

The vast epigenetic dysregulation induced by H3K27M mutations serves to stall cells in an epigenetic landscape reflective of a cell-of-origin²⁹. At the level of gene expression, tumor cells mirror a primitive stem-cell like state with high proliferative capacity with impaired ability to differentiate towards terminal states^{23,24,90,95}. Given the restricted pattern of occurrence in the midline and efforts to induce high-grade tumors in murine models, H3 K37M tumors are thought to arise in a specific cellular context in a restricted neurodevelopmental window. Early

experiments of inducing an H3.3 K27M transgene with p53 loss in postnatal neural progenitor cells (NPCs) *in vitro* using murine cells and *in vivo* using a murine model failed to produce tumor-like lesions^{24,25}. However, H3.3 K27M transgene delivery with p53 loss and PDGFRA activation induced glioma-like lesions *in vitro* using embryonic NPCs derived from human ES cells²³. The same genetic combination induced *in vivo* using murine embryonic NPCs in the forebrain or hindbrain faithfully recaptured the invasive high-grade glioma phenotype²⁴. A second *in vivo* approach using H3.3-K27M transgene delivery with tp53 loss in NPCs induced glioma-like lesions using orthotopic xenograft models⁹⁶. While the general consensus is that tumors arise from an embryonic cell-of-origin, the exact cell of origin remains elusive. In addition to NPCs, oligodendrocyte precursor cells (OPCs) have been proposed as a putative cell-of-origin from single-cell studies of primary tumor samples³⁴. However, a third *in vivo* approach using NPCs, but not OPCs, derived from human induced pluripotent stem cells injected with H3.3 K27M and TP53 loss gave rise to tumors in orthotopic xenograft models⁹⁶. This is suggestive of a cell-of-origin committed towards the oligodendrocyte lineage, but an earlier progenitor may be possible.

H3.3 K27M also impairs differentiation of cells which may contribute towards the proliferative capacity of the tumor. Delivery of H3.3 K27M transgene in NPCs resulted in a severe differentiation block towards the astrocyte lineage and to some extent the oligodendrocyte lineage *in vitro*²³. Removal of the H3.3 K27M mutation resulted in differentiation of cells towards the astrocyte lineage *in vitro*³¹. However, astrocytes and oligodendrocytes are distinct terminal cell types arising from different lineages derived from NPCs. Thus the precise cellular context of where along differentiation the tumors arose remains a subject of active investigation.

1.3.2 ACVR1 partner mutations in H3K27M pHGGs

While H3K27M mutations are a founding event in tumorigenesis, evidence suggests the presence of obligate partner mutations to promote and maintain oncogenesis in high-grade tumors^{8,9,14,23,25,26,35}. Common partner mutations include loss of *ATRX* and *TP53*, and activation of kinases including *PDGFRA*, *PI3K*, *ACVR1*. The role of partner mutations, particularly kinase

mutations, in promoting oncogenesis is of particular therapeutic interest since histone mutations are not directly targetable but a wealth of kinase inhibitors exist.

Recurrent activating mutations in activin A receptor type I (*ACVR1*) encoding the kinase ALK2 are reported in 20-25% of DIPGs harboring a H3 K27M mutation^{8,13,16,97}. *ACVR1* mutations display preferential segregation with canonical H3.1 K27M DIPGs occurring in the pons but are also present in H3.3 K27M DIPGs occurring along the midline. In addition, mutations in specific domains of *ACVR1* show enrichment in specific histone contexts. Mutations in *ACVR1* affecting the inhibitory glycine-serine (GS) domain (arginine 206 to histidine; R206H) segregate with H3.3 K27M, while mutations affecting the ATP binding pocket of the kinase domain (glycine 328 to glutamic acid or valine, G328E/V, respectively) segregate with H3.1 K27M. From a structural basis mutations in both the GS domain and kinase domain are thought to inhibit interactions which stabilize the inactive state of the kinase and shift it towards a structurally active state⁹⁸. At the protein level, these mutations induce a weak constitutive gain of function^{13,16,98}.

ALK2 belongs to the bone morphogenic protein (BMP) growth signaling pathway in the transforming growth factor β (TGF β) superfamily⁹⁹. Inducement of BMP signaling occurs upon heteromeric complex formation of serine-threonine kinases belonging to type II and type I receptors¹⁰⁰. Ligand binding occurs on constitutively active type II receptors which is then recognized by class I receptors. Class I receptors are recruited to the complex and phosphorylated by type II receptors and then propagate downstream signaling mediated by receptor-regulated and pathway-specific SMAD proteins^{101,102}. SMAD proteins converge upon binding with co-mediator SMAD4, where they translocate to the nucleus to induce transcription of target genes including *ID1/2*.¹⁰³ Gain of function mutations in *ACVR1*, which encodes a type I receptor, causes constitutive activation even in the absence of a ligand in DIPG^{13,16}.

BMP signaling has a wide variety of functions including embryonic patterning, tissue morphogenesis and cellular differentiation^{104–106}. Given the important roles of BMP signaling in development, it is not surprising that dysregulation of BMP signaling pathways results in human disorders. Congenital mutations in *ACVR1* result in fibrodysplasia ossificans progressiva (FOP), a disorder characterized by ectopic bone formation in soft tissue usually occurring in children, but

with no predisposition for malignancies^{107–113}. A subset of the same mutations which occur in FOP have been reported in *ACVR1* mutant H3K27M pHGGs^{114,115}. The biological functions of *ACVR1* mutations in promoting oncogenesis is still a subject of active investigation. Furthermore it is possible that *ACVR1* mutations may have different functions in canonical H3.1 versus variant H3.3 K27M DIPGs due to distinct epigenetic landscapes and cellular contexts of H3.1 versus H3.3 K27M tumors¹¹⁶.

ACVR1 mutations in DIPG are found in a younger age group and associated with slightly longer overall survival^{13,16}. *ACVR1* mutations are an early clonal event in tumorigenesis which indicate the importance of *ACVR1* mutations in promoting tumorigenesis³⁵. To further support the idea that *ACVR1* mutations promote tumorigenesis, inducement *ACVR1* R206H or G328E/V mutations with H3.1 K27M and p53 loss in postnatal brainstem precursors induced glioma-like lesions in a murine model¹¹⁷. Using another murine model, *ACVR1* G328V mutations in progenitor cells were shown to impair differentiation of cells along the oligodendrocyte lineage but the mutation alone was insufficient to induce tumorigenesis¹¹⁸. While *ACVR1* mutations promote tumorigenesis, the role of *ACVR1* impacting the cellular differentiation in differing histone contexts (canonical versus variant) H3K27M DIPGs has yet to be elucidated.

1.3.3 H3.3 G34 mutations in cancer and neurodevelopmental syndromes

Mutations at glycine 34 to arginine or valine (G34R/V) in the tail of H3.3 (encoded by *H3F3A*) drive approximately 30% of pHGGs occurring in the cerebral cortex of adolescents and young adults^{9,10,14,26}. Common partner mutations include loss of TP53 and ATRX and activation of PDGFRA. Intriguingly, mutations at glycine 34 to tryptophan (G34W) in H3.3 do not present in pHGGs but drive over 90% of GCTB, a locally aggressive tumor affecting the epiphysis of long bones in adults¹⁷. The specificity in each mutation in driving a specific cancer is a subject of active investigation.

G34 mutations cause DNA hypomethylation and disrupt H3K36 methylation on the same histone tail^{10,19,25,119,120}. There are two proposed overlapping mechanisms by which G34 mutations act to influence H3K36 methylation: glycine is the smallest amino acid and substitutions to larger amino acids are thought to sterically hinder deposition of H3K36 methylation; and / or glycine substitutions lead to altered recognition by reader complexes for

H3K36 methylation^{25,121–123}. In support of the first mechanism, glycine 34 exists in a very narrow tunnel upon H3K36 tri-methylase SETD2 binding¹²³. Mutations to larger amino acids are thought to inhibit SETD2 interactions through steric hinderance in the narrow tunnel. Indeed, G34 mutations in cell lines inhibit SETD2-mediated deposition of H3K36me3³⁰. In contrast to promoting loss of H3K36 methylation, G34W mutations do not affect catalytic activity of EZH2 in the PRC2 complex and promote gain of repressive H3K27me3¹²⁴. Indeed, in G34W in cell lines promote abnormal PRC2-mediated deposition of repressive H3K27me3 at enhancers targeted by SETD2³⁰. G34W mutations additionally alters landscape of H3K27me3 by inducing a shift of H3K27me3 from intergenic to genic regions²⁸. In support of the second mechanism, G34W mutant cell lines display reduced binding of ZYMD11, a reader for H3K36 methylation¹²⁴. Thus G34 mutations impact local epigenetic landscapes through alteration of H3K36 and H3K27 methylation.

In addition to disruption of histone methylation, G34 mutations are associated with global DNA hypomethylation in gliomas¹²⁰. The mechanism by which this occurs is not well understood. One potential route may be through disruption of DNA methylase Dnmt3 family members which are known to be recruited by H3K36 methylation^{73,74,125}. Dnmt3a recognizes H3K36me2 and is recruited to deposit DNA methylation in non-coding regions. Dnmt3b recognizes H3K36me3 and is recruited to gene bodies⁷⁸. Local disruptions of H3K36 methylation by G34 mutations may impact binding of Dnmt3 family members, resulting in an altered DNA methylation landscape.

The epigenetic dysregulation induced by G34 mutations serves to transcriptionally alter the cellular context of G34 mutant tumors and converge on inhibition of cellular differentiation^{28,30,32}. Similar to H3K27M pHGGs, G34R/V pHGGs follow a spatial and temporal specificity leading to proposed origins in neurodevelopment and are proposed to arise in a prenatal interneuron cell-of-origin³². At the epigenetic level, G34R/V mutations induce gain H3K27me3 at promoters of mature neuronal genes and which is thought to transcriptionally inhibit differentiation of cells towards mature neurons. G34W mutations in GCTB are thought to follow a similar model of impaired differentiation of the mesenchymal lineage. In GCTB, G34W stromal cells mirror a proliferative osteoblast-like progenitor cell-of-origin which are derived

from mesenchymal stem cells (MSCs)^{28,30}. The stromal cells isolated from GCT tumors are refractory to differentiation towards a terminal myofibroblast state using *in vitro* differentiation experiments. This supports the model that G34 mutations alter the epigenome and transcriptome in a tissue and cellular context-specific manner to promote oncogenesis.

Outside of cancer, H3.3 G34 mutations are implicated in neurological disorders through their effect of disrupting H3K36 methylation. H3K36 methylation is required for proper normal development^{126,127}. Mutations in epigenetic modifiers of H3K36 methylation, among others, result in neurological syndromes characterized by neurological dysfunction and often intellectual disability, but no overall predisposition for cancer¹²⁸. Other features of individual syndromes include immune dysfunction and overgrowth. Genetically, there is a clear dosage dependency contributing to these syndromes, where loss of a single allele recapitulates the phenotypes for most known disorders affecting reader, writer, and eraser protein complexes. Germline loss of function in one copy of H3K36 writer *NSD1* results in Sotos Syndrome; mutations in *NSD2* are implicated in a subset of Wolf–Hirschhorn syndrome; mutations in *SETD2* result in Luscan-Lumish syndrome and other Sotos-like syndromes^{129–135}. Mutations in H3K36 readers such as DNA methylase *DNMT3A* results in Tatton-Brown Syndrome; mutations in *ZYMND11* result in 10p15.3 deletion syndrome; mutations in reader *MECP2* result in Rett Syndrome^{136–140}. The dosage dependencies often noted in these syndromes are evident in murine models; homozygous loss of H3K36 writers *NSD1*, *SETD2*, or *SETD5* results in embryonic lethality with clear developmental defects, while homozygous loss of writer *NSD2* results in severe developmental delay followed by postnatal lethality^{126,127,141,142}. However, heterozygous loss of *NSD1* or *NSD2* result in viable mice^{126,127}. This points to an underlying sensitivity for regulation of H3K36 methylation in the developing brain.

Recently, mutations in *H3F3A* and *H3F3B* were identified in a neurodegenerative disorder with neurological dysfunction but no predisposition for cancer-related malignancies^{143,144}. While there was one patient identified bearing a G34V mutation in *H3F3B*, the same G34 mutations in *H3F3A* observed in cancer were not reported. This is suggestive of key roles of H3.3 mutations in development which may potentially overlap with those observed in pHGGs. How G34

mutations function to dysregulate downstream changes in gene expression over the course of development and in the brain is poorly understood.

1.4 Hypothesis and Aims

We model that driver mutations in pediatric high grade gliomas (pHGGs) arise in a temporal and spatially restricted cellular context during neurodevelopment and maintain oncogenicity by stalling cells in a progenitor-like state. Characterizing the effect of histone and partner mutations on cellular differentiation may provide insight onto cell-specific dependencies required by the tumor and thus open therapeutic avenues. My aims are to use transcriptomic data to understand the effect of histone mutations on cell identity programs using several model systems by mapping to the developing brain.

In the body of this text, I will first demonstrate that mutations in *ACVR1* have different transcriptomic effects on cell state dependent on the specific histone context, using data generated by Abdulshakour Mohammadnia (Jabado lab). I will secondly show that K27M mutations stall differentiation towards astrocyte and oligodendrocyte states using *in vivo* patient-derived orthotopic xenograft models, but this effect is not primarily recapitulated using *in vitro* serum-differentiation experiments using data generated by Brian Krug (Jabado lab). I will thirdly demonstrate that germline heterozygous G34R / V / W mutations display surprising specificity in affected lineages; G34R mutants display progressive neuronal loss coupled to immune infiltration while G34W mutants display an unhealthy fat phenotype using data generated by Sima Khazaei (Jabado lab). The work here will improve our understanding of specific cell-type gene programs impacted by histone and partner mutations in cancer and development which have implications for the further development of therapeutics.

Chapter 2: Materials and Methods

2.1. Analysis of bulk RNA seq data

Read Processing. Raw reads were trimmed using trimmomatic¹⁴⁵ (v0.32) to remove adaptor sequences four bp from the start of each read and Illumina-specific sequences using palindromic mode. Reads were truncated where the average base quality was too low (phred score < 30) over a four nucleotide sliding window. After this procedure, reads which were less than 30 bp were discarded.

Read Alignment. Reads which passed preprocessing thresholds were aligned to reference genomes using STAR¹⁴⁶ (v2.3.0e) with default parameters. The reference genome build used were hg19 (GRCh37) or mm10 (GRCm38), respectively. Reads obtained from two species were simultaneously aligned to both genomes. Reads mapping to more than 10 locations in the genome (MAPQ < 1) were discarded. Multiple control metrics were obtained using FASTQC (v0.11.2), samtools¹⁴⁷ (v0.1.20), BEDtools¹⁴⁸ (v2.17.0), and custom scripts.

Gene expression analysis. Gene expression was estimated using featureCounts¹⁴⁹ (v1.4.4) by counting the number of primary alignments mapping to at most two locations (MAPQ ≥ 3) to exonic regions of the Ensembl gene genome. Reads from multiple isoforms were collapsed to the gene-level to obtain a gene by sample counts matrix. Reads were normalized using DESeq2¹⁵⁰ (v1.18.1) mean-of-ratios and variance-stabilizing transformation was done for visualization.

Principal component analysis. Principal component analysis was run using the top 10000 most variable genes calculated from DESeq2 mean-of-ratios normalized counts.

Hierarchical clustering. Hierarchical clustering was performed on top 10000 most variable genes calculated from DESeq2 mean-of-ratio normalized counts. Complete linkage was used for the agglomeration method and Euclidean distance as the distance metric. Variance-stabilized counts were plotted for visualization purposes.

Differential gene expression analysis. Differential gene expression analysis was run using DESeq2 (v1.18.1) comparing samples using normalized counts without batch correction or any further transformation of the data. The Benjamini-Hochberg method was used for multiple testing correction. Significance of genes was called using a threshold of sufficient coverage with

an average normalized read count of > 100, absolute value of log2FoldChange > 0.58, and adjusted p-value < 0.05, unless otherwise noted.

Pathway analysis. Pathway analysis was run using the R package gProfiler¹⁵¹ using the `gost` function with default parameters. The default parameter 'gSCS' was used for multiple testing correction. Input gene lists were genes significantly downregulated or upregulated, respectively, from differential expression analysis. We filtered on gene ontology GO categories 'BP', 'CC', and 'MF' for interpretability of results, unless otherwise noted.

Gene set enrichment analysis. Gene set enrichment analysis was run using the R package fgsea¹⁵² with parameters minimum gene set size 15, maximum gene set size 500, and 10000 permutations. Genes were ranked using the log2FoldChange divided by the standard error of the log2FoldChange, computed from the Wald test from DESeq2. Input gene lists were inputted from a selection of databases and publications. Hallmark and GO gene signatures were obtained from msigdb¹⁵³ (v7.4.1). Gene signatures defining cell type identity derived from single-cell RNA sequencing data were obtained for the developing murine forebrain and hindbrain, cerebellum, adult cortex, and disease-associated microglia^{31,33,154,155}. The top 100 genes were used ranked by log2FoldChange and adjusted p-value, where available. The Benjamini-Hocberg method was used for multiple testing correction. Significantly enriched or depleted pathways were defined as those with an adjusted p-value < 0.05.

2.2. Analysis of single-cell RNA seq data

Initial processing and quality control of sequencing data. Cell Ranger (10X Genomics, v3.1.0) was used with default parameters to demultiplex and align sequencing reads, to distinguish cells from background and obtain gene counts per cell. Alignment was performed using a joint hg19+mm10 genome reference build, coupled with Ensembl transcriptome build GRCh37 v.82 for hg19 and GRCm38 v.84 for mm10. Intronic counts were excluded. For xenograft samples, human cells were extracted if cells were either assigned as human by Cell Ranger or if the cell contained greater than 75% of total reads mapping to hg19 in order to obtain adequate numbers of cells per sample. Cells were filtered based on the following quality control metrics: mitochondrial content (indicative of cellular damage), number of genes and number of unique molecular identifiers (UMIs) using the R package Seurat¹⁵⁶ (v3.1.0). For xenograft samples,

thresholds were set on a per-sample basis where cells were excluded if they had greater than 50% of total reads mapping to mitochondrial read counts, had less than 500 genes or UMIs per cell, or were outside 2 standard deviations from the mean number of genes or UMIs, respectively. For mouse samples, thresholds were set on a per-sample basis where cells were excluded if they had greater than 20% of total reads mapping to mitochondrial read counts, had less than 1000 genes or UMIs per cell, or were outside 2 standard deviations from the mean number of genes or UMIs, respectively. Libraries were scaled to 10,000 UMIs per cell and natural log-normalized. Log normalized counts were used for computing correlations for gene expression and assessing expression of specific genes. Samples were combined by cell line or tissue, respectively, without any additional transformation of the data.

Dimensionality reduction. The top 2000 most variable genes were computed using 'vst' selection parameter from the FindVariableFeatures function the R package Seurat. Briefly, the method fits a line using loess regression between the observed $\log(\text{variance})$ and $\log(\text{mean})$. Then it standardizes the gene counts using the observed mean and expected variance. Gene variance is then calculated using standardized gene counts followed by regression of nUMIs and mitochondrial counts to eliminate technical artifacts. The top 2000 genes which have the greatest variance are then selected for downstream analysis. PCA is then run using the top 2000 most variable genes. The top 30 PCs are used as input for UMAP¹⁵⁷ and tSNE¹⁵⁸ dimensionality reduction. Cells were clustered using the shared nearest neighbor graph¹⁵⁹ (SNN) construction inputting the top 30 PCs, followed by unsupervised clustering using a resolution parameter of 0.8 and the Louvain algorithm¹⁶⁰ for modularity optimization.

Differential gene expression analysis. A Wilcoxon rank-sum test was run to identify differentially expressed genes using the FindMarkers function from the R package Seurat requiring that genes are expressed in at least 25% of cells in either population to test.

Nearest normal cell type assignment. For xenograft samples, cells were projected using a Spearman correlation between log normalized counts and reference expression profiles. The cell type in the reference with the highest correlation was assigned as the cell type identity. Proportions of cells were calculated as the number of cells belonging to a given cell type over the total number of cells for a given condition. For mouse samples, cells were projected to

nearest normal cell types using consensus prediction between a neural network ACTINN¹⁶¹ and single sample gene set enrichment analysis (ssGSEA)¹⁶².

Chapter 3: Results

3.1: Transcriptomic effect of *ACVR1* in H3.1 and H3.3 K27M mutant tumors in isogenic cell lines

3.1.1. *ACVR1* impacts global transcriptome in H3.3 K27M, to a lesser degree in H3.1 K27M

While histone H3 K27M mutations are necessary for tumor initiation, evidence suggests the synergistic role of co-partner mutations to maintain tumorigenesis. However, the mechanisms of co-partner mutations in the cell are poorly understood. *ACVR1* is a cell-surface kinase receptor which functions in BMP growth signaling and gain-of-function mutations in *ACVR1* are present in approximately 67% of H3.1 K27M pHGGs and 18% of H3.3 K27M pHGGs^{13,16}. Despite being an amenable drug target, little is known of the functional role of *ACVR1* in promoting tumorigenicity. In particular, H3 K27M pHGGs maintain stem-like properties through inhibition of cellular differentiation. To test whether *ACVR1* is directly involved in maintaining a differentiation blockade in H3 K27M pHGGs, CRISPR-Cas9 editing of *ACVR1* was performed in four cell lines and sent for RNA sequencing (RNA seq). This allows us to study the effects of *ACVR1* using an isogenic system and decouples the effect of the mutation from cell-of-origin or cell-line specific effects. Briefly, the guide RNA was designed to target exon 6 of *ACVR1*. Following CRISPR-Cas9 editing genotypes were verified as frameshift or complete KO using Illumina MiSeq. The *ACVR1* sample cohort is presented in Table 1.

Table 1: RNA seq cohort for *ACVR1* cell lines

Cell line	Histone Mutation	<i>ACVR1</i> status	Replicates
BT869	H3.3 K27m	R206H	6
		R206H-KO	7
DIPG007		R206H	5
		R206H-KO	6
DIPG36	H3.1 K27M	G328E	3
		G328E-KO	5
DIPG4		G328V	3
		G328V-KO	4

PCA for all samples was used to assess global transcriptomic changes. Visualizing samples across PC1 and PC2, which account for the most and second most sources of variation in the data showed that samples separate primarily by histone status (H3.3 or H3.1) and cell line, but not ACVR1 status (Fig 1). This indicates that ACVR1 status may drive changes secondary to histone status.

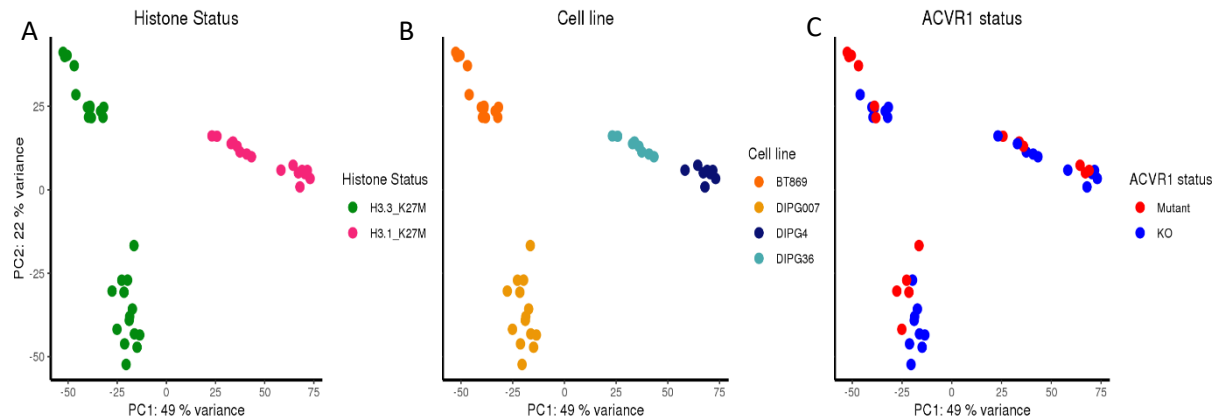


Figure 1: ACVR1 mutation status is secondary to H3 K27M

(A) PCA on top 1000 most variable genes highlighting histone H3 K27M status. **(B)** Same as A but highlighting cell line. **(C)** Same as A but highlighting ACVR1 mutation status.

Due to the overwhelming effect of H3.1 or H3.3 K27M histone status and cell line-specific differences, cell lines were analyzed separately to assess the effect of ACVR1 status on the global transcriptome. PCA on H3.3 K27M cell lines showed separation by ACVR1 status, although this was not significant along PC1 and PC2 (Fig 2 A, B). In the case of BT869, there were two KO samples along PC1 closer to the mutant samples along PC1 than other KO. This indicates that those samples may be more similar to the mutant than other KO samples. However, this difference could not be mapped to technical factors such as mapping percentage, exonic coverage, mitochondrial content, or type of ACVR1 perturbation. In DIPG007 the same trend appeared and likewise could not be mapped to technical factors. Together these results indicate ACVR1 status influences global transcriptomic profiles in H3.3 K27M cell lines but removing the mutation may have a heterogeneous effect.

In contrast, PCA on H3.1 K27M cell lines showed separation by *ACVR1* status along PC2 for one replicate (DIPG4, Fig 2C). For this replicate, *ACVR1* status is driving the second most source of variation in the samples, indicating a large effect on the transcriptome. However, for the second H3.1 K27M replicate DIPG36, samples did not segregate by *ACVR1* status which indicates little effect of genotype in driving major transcriptomic changes (Fig 2D). Together, these results indicate that *ACVR1* status effects the global transcriptome with substantial heterogeneity in H3.1 K27M cell lines. As a result, downstream analyses were kept separate per cell line.

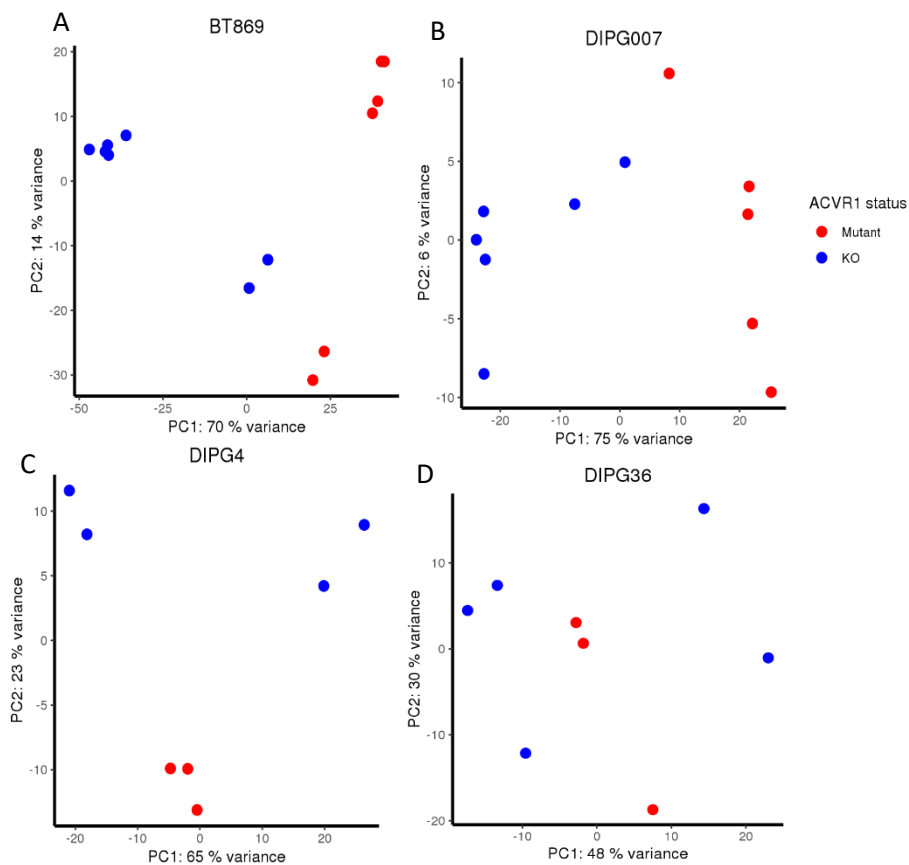


Figure 2: *ACVR1* has heterogeneous effects in driving transcriptomic changes in a cell-line specific manner

(A) PCA using top 1000 most variable genes for H3.3 cell line BT689; **(B)** H3.3 K27m cell line DIPG007; **(C)** H3.1 K27M cell line DIPG4; **(D)** H3.1 K27M cell line DIPG36

Differential gene analysis was used to assess for significantly upregulated or downregulated genes. In H3.3 cell lines, differential gene expression analysis showed many significantly dysregulated genes (BT869, 3427; DIPG007 1381) (Fig 3). In H3.1 cell lines, differential gene expression analysis showed notably few dysregulated genes (DIPG4, 229; DIPG36). This indicates that *ACVR1* mutation status plays a different role in the two histone (H3.1 K27M, H3.3 K27M) contexts. In H3.3 cell lines, *ACVR1* is majorly influencing the transcriptome. In contrast in H3.1 K27M cell lines, although there is an effect on global transcriptome by PCA for one replicate, this effect is not replicated in the second replicate. Furthermore, there are few differentially expressed genes in H3.1 cell lines indicating a weak effect of *ACVR1* status in this histone context.

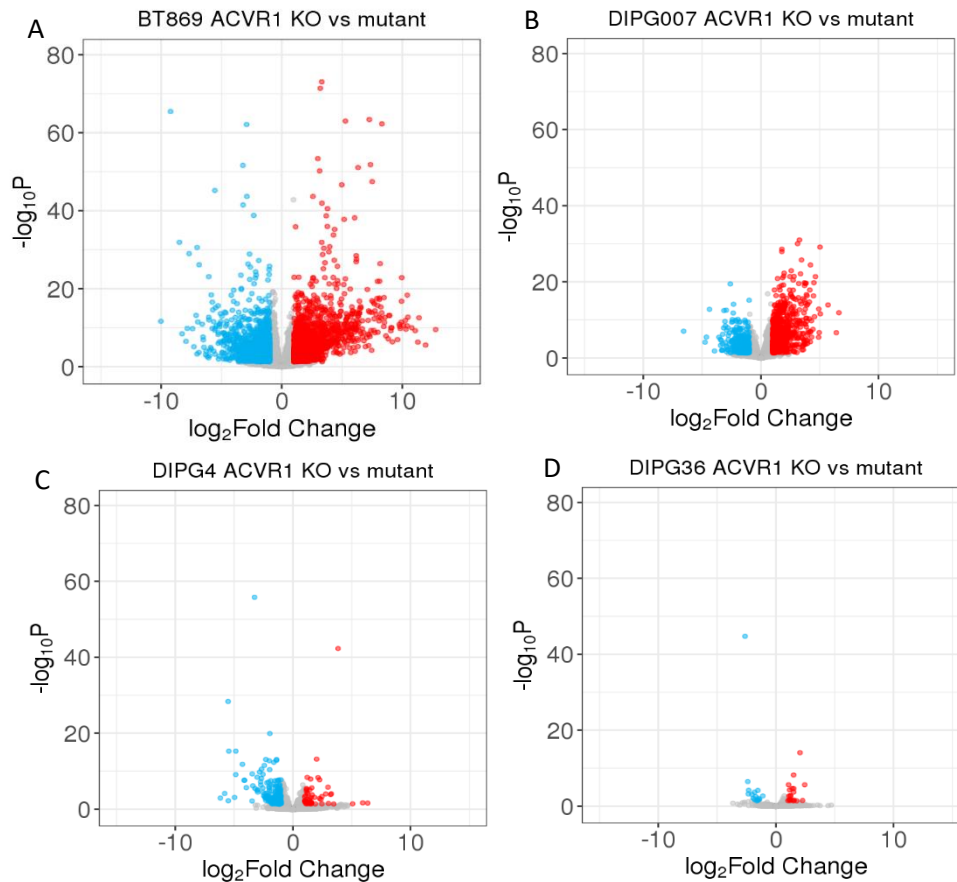


Figure 3: *ACVR1* status results in large changes in H3.3, but not H3.1, K27M cell lines

(A) Volcano plot of bulk RNA seq highlighting differentially upregulated genes (red) and downregulated genes (blue) in *ACVR1* KO using a threshold of mean normalized read count >

100, absolute log2FoldChange > 1 and adjusted p-value < 0.05 for H3.3 K27M cell line BT869; **(B)** H3.3 K27M cell line DIPG007; **(C)** H3.1 K27M cell line DIPG4; **(D)** H3.1 K27M cell line DIPG36.

To identify potential functional terms associated with *ACVR1* mutational status, pathway analysis on differentially expressed genes was run. For both H3.3 K27M cell lines, upregulated genes in *ACVR1*-R206 samples were associated with differentiation (neurogenesis, nervous system development, system development) (Fig 4 A, B). Upregulated genes in *ACVR1*-R206H-KO samples were associated with extracellular processes (cell adhesion, cell projection, cell migration) and developmental processes (cell differentiation, system development) (Figure 4 D, E). This indicates that in the context of H3.3, the effect of *ACVR1* is potentially involved in neuron development and cells may have characteristics of an early progenitor. This is compatible with the scenario where removal of *ACVR1* may switch the cell towards a more mature state and induce cell surface remodeling.

For H3.1 cell lines, pathway analysis did not result in shared upregulated gene programs since one replicate (DIPG36) had few differentially expressed genes. As such, results obtained from pathway analysis were not deemed reliable. For the other 3.1 K27M replicate (DIPG4), upregulated genes in *ACVR1*-G328V samples were associated with extracellular responses (cellular response to chemical stimulus, cell periphery, extracellular region) and developmental processes (developmental process, anatomical structure development) (Fig 4C) . Upregulated genes in *ACVR1*-G328V-KO were associated with homeostatic responses (cellular response to metal ion, cellular response to calcium ion) and extracellular pathways (cell periphery, extracellular matrix) (Fig 4F). These results indicate that while *ACVR1* may be implicated in some aspect related to development, there seems to be a substantial component involved in other biological pathways not related to cellular differentiation upon removal of the mutation. This points to the idea that function of *ACVR1* in these two histone contexts (H3.1 K27M, H3.3 K27M) may be distinct and have cellular functions in promoting oncogenesis.

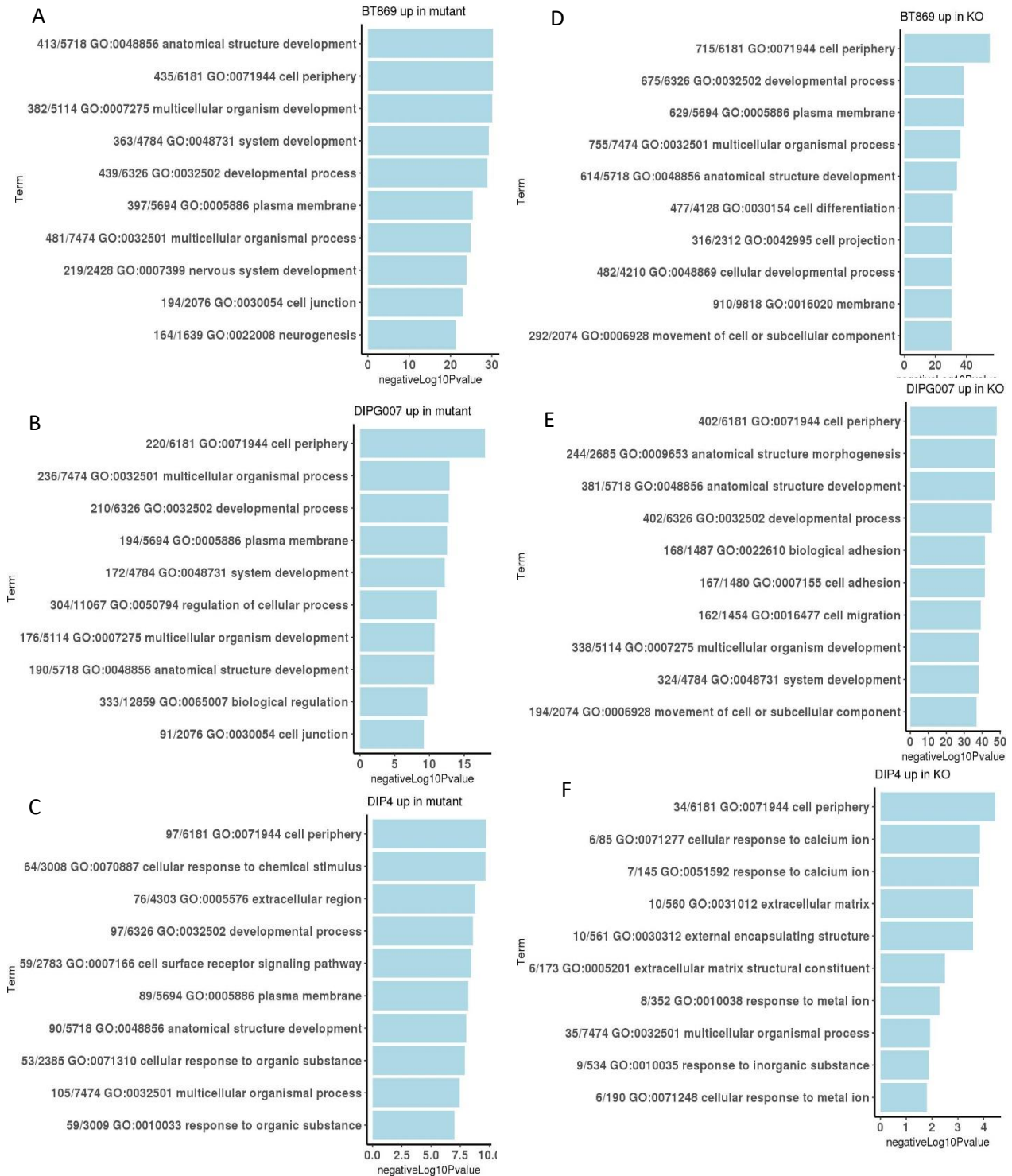


Figure 4: Pathway analysis shows terms related to differentiation in H3.3 K27M cell lines

(A) Barplot showing top 10 gene ontology pathways in H3.3 cell line BT869 *ACVR1* mutant; **(B)** in BT869 *ACVR1*-KO; **(C)** in H3.3 cell line DIPG007 *ACVR1* mutant; **(D)** in DIPG007 *ACVR1*-KO; **(E)**

in H3.1 cell line DIPG4 *ACVR1* mutant; **(F)** in DIPG4 *ACVR1*-KO. Y-axis labels show numbers of DEGs and size of pathway followed by ontology term.

3.1.2. *ACVR1* induces switch in glial cell state in H3.3 K27M, not in H3.1 K27M

Since *ACVR1* mutations upregulated genes associated with potentially distinct biological pathways in H3.1 and H3.3 K27M cell lines, we next asked if *ACVR1* had different impacts on cellular differentiation in these histone contexts. Since H3 K27M pHGGs are thought to arise during a specific cellular context and developmental window, assessment for changes in developmental cell types used a single-cell murine developmental atlas spanning embryonic and early postnatal time points in the forebrain and pons³¹. The reference contained a wide variety of cell types in the brain spanning neuronal, glial, and non-neuroectodermal lineages which allowed for detection of diverse gene programs. To derive cell-identity signatures, clusters were first manually assigned cell-type identity through expression of canonical marker genes followed by extraction of cluster markers to define robust gene signatures.

To assess whether specific brain cell types were relatively enriched in the *ACVR1* KO vs *ACVR1* mutants, gene set enrichment analysis (GSEA) was run using the cell-type identity gene signatures. Removal of *ACVR1* mutations in H3.3 K27M cell lines resulted in an enrichment of astrocyte and ependymal gene programs (Fig 5A). However, astrocytes were the only robustly enriched gene programs in both replicates. The same conclusion was independently found through robust expression of the astrocytic markers GFAP and CD44 through immunofluorescence staining. (Figure 5B). This implies that in H3.3 K27M cell lines, a function of *ACVR1* mutations on cell identity may be to hinder differentiation along the astrocytic lineage. To confirm whether an astrocytic enrichment was found in tumor samples, comparison 5 H3.3 K27M *ACVR1* WT tumors to 3 H3.3 K27M *ACVR1*-R206H tumors did not demonstrate the astrocytic increase. However, comparison of tumor samples has several confounding variable such as potentially differing cells-of-origin and different genetic backgrounds. Future work will be required to demonstrate the effect of a differentiation blockade towards astrocytes using genome-wide primary tumor data or *in vivo* models.

There were no consistently identified enriched gene programs in both H3.3 K37M *ACVR1* mutant replicates. One H3.3 K27M replicate, DIPG007, had few consistently enriched gene

programs in *ACVR1* mutants. The other H3.3 K27M replicate, BT869 showed a strong depletion of oligodendrocyte and oligodendrocyte precursor cell (OPC) gene programs and *ACVR1* mutants significantly expressed canonical gene markers *SOX10*, *PDGFRA* and *CNTN1* (Figure 5C). However, mature oligodendrocytes do not have the capacity to differentiate towards astrocytes, thus further analysis in untangling the exact mechanisms contributing to this state are required. The component of OPC enrichment is partially consistent with the proposed OPC-like cell-of-origin in H3.3 K27M tumors^{34,163}.

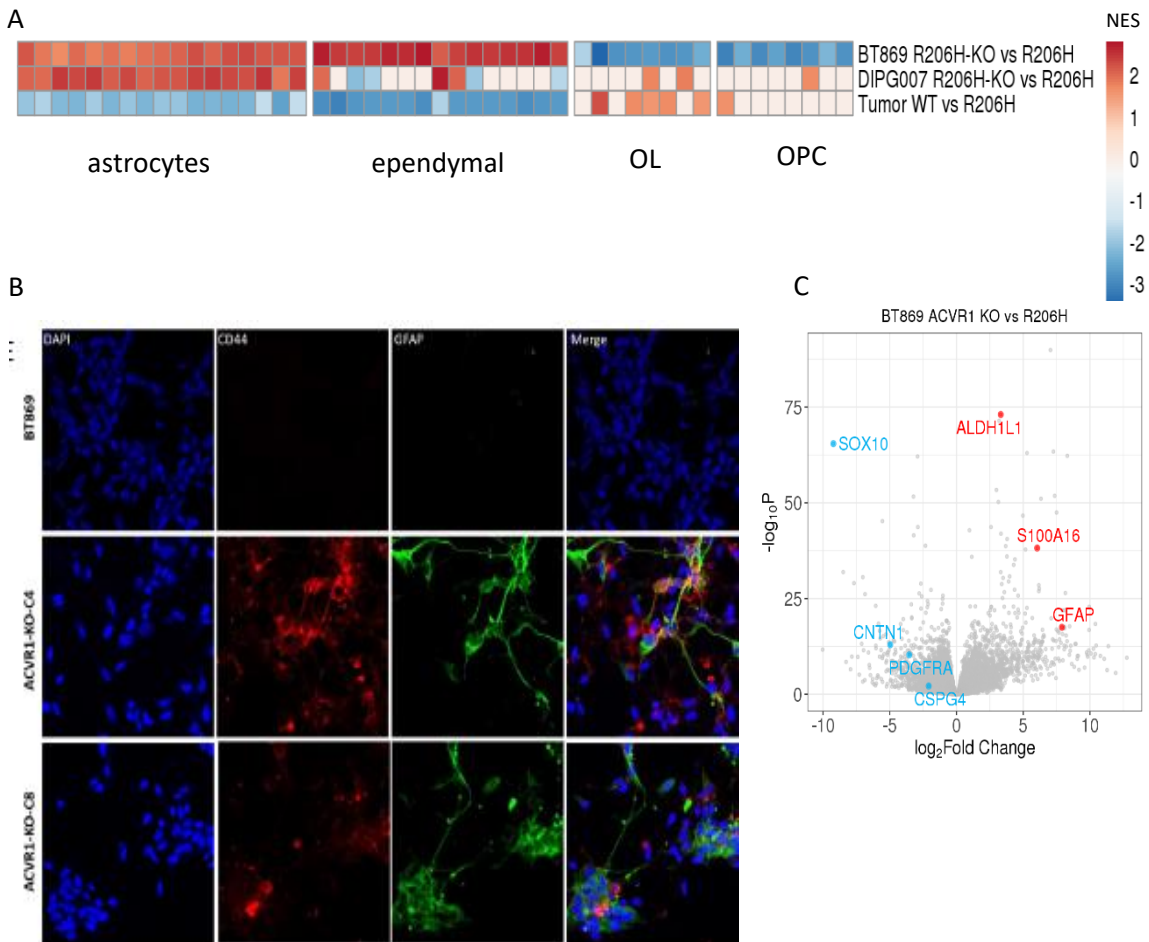


Figure 5: ACVR1 blocks astrocytic maturation in H3.3 K27M

(A) Gene set enrichment analysis of cell identity signatures (Jessa et al) in H3.3 K27M cell lines and tumor samples. **(B)** representative immunofluorescence images of BT869 for astrocyte markers GFAP and CD44. Figure generated by Abdulshakour Mohammadnia. **(C)** Oligodendrocyte, OPC, and astrocyte markers in BT869. Abbreviations: NES, normalized

enrichment score; OL, oligodendrocytes; OPC oligodendrocyte precursor cells. Scores which are not significant (adjusted p-value < 0.01) are set to 0.

In contrast to H3.3 K27M cell lines, removal *ACVR1-G328V* in the context of H3.1 K27M did not yield an astrocytic enrichment (Figure 6A). Surprisingly, astrocytic signatures and the canonical astrocyte marker GFAP were strongly downregulated in *ACVR1*-KO cells, (Fig 6B). Results are only shown for one replicate (DIPG4) due to the handful of differentially expressed genes in the second replicate (DIPG36) which did not yield robust results. However, removal of *ACVR1* showed decreased proliferation *in vitro* and upregulation of very few transcriptomic cell-identity gene programs, which consisted of a handful of non-neuroectodermal cell-identity signals (Figure 6A, C). This is consistent with the extracellular matrix programs upregulated through pathway analysis. Furthermore, *ACVR1-R206H* mutations in H3.1 K27M tumors promote a mesenchymal phenotype partly mediated through Stat3 signaling¹¹⁷. While our cell line does not bear the same *ACVR1* mutation (G328V instead of R206H), it is possible that the downstream effects mediated by *ACVR1* converge towards a common cellular response in H3.1 K27M tumors. Therefore, our results are consistent with a model where *ACVR1* in H3.1 K27M cell lines does not to promote changes in cell identity, but mesenchymal-related gene programs.

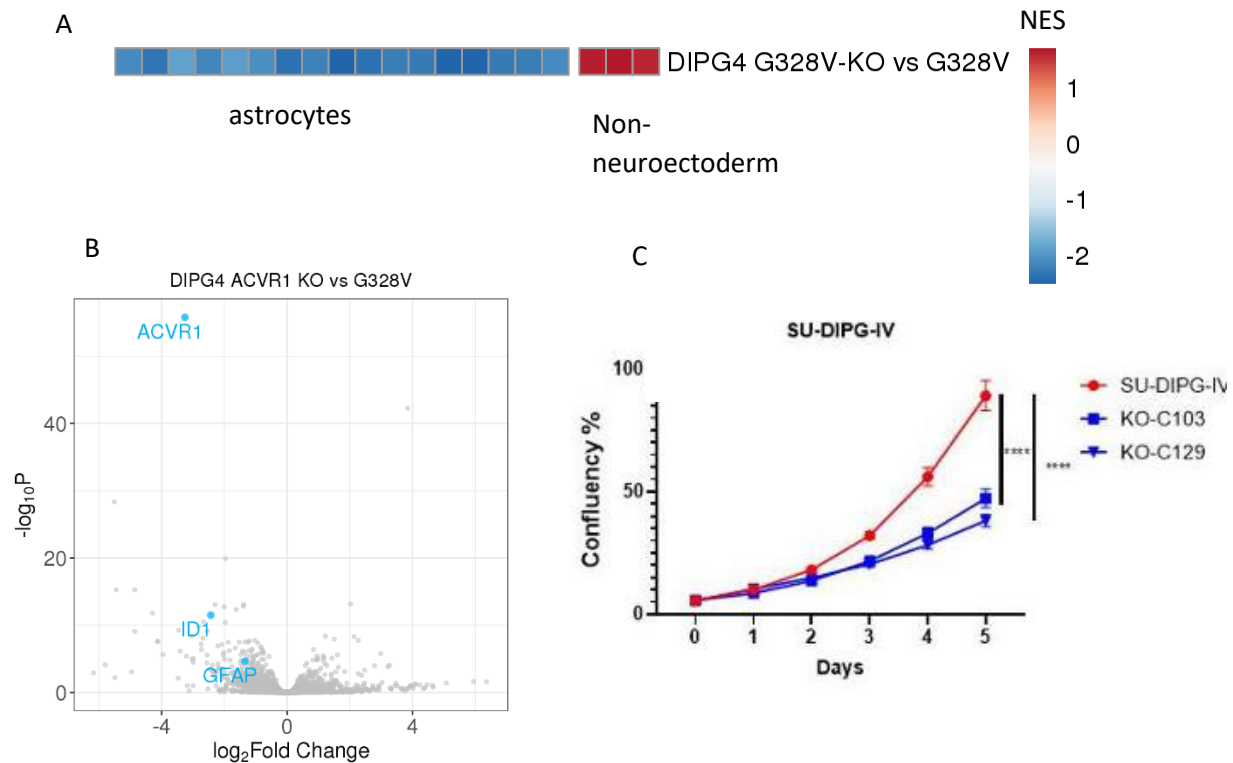


Figure 6: Removal of ACVR1 does not induce astrocytic maturation in H3.1 K27M

(A) Gene set enrichment analysis of astrocyte signatures (Jessa et al, 2019) in H3.1 K27M cell line DIPG4. Scores not significant are set to 0. **(B)** Decreased expression of astrocyte marker GFAP in DIPG4. **(C)** median ssGSEA scores for non-neuroectodermal cell identities in ACVR1-KO vs ACVR1 mutant **(D)** ACVR1-KO reduces proliferation *in vitro*. Figure generated by Abdulshakour Mohammadnia. NES: normalized enrichment score.

In summary, *ACVR1* has unique functions in the cell which is dependent on histone status (H3.1 vs H3.3 K27M). In H3.1 K27M cell lines, *ACVR1* does not cause major transcriptomic changes but pathway analysis indicates it may be implicated in promoting developmental processes and extracellular responses. *In vitro*, *ACVR1* promotes proliferation, which may tie into developmental processes results from pathway analysis. In the context of H3.3 K27M cell lines, *ACVR1* has a role in promoting a differentiation block towards the astrocytic lineage. Given that H3.3 K27M is thought to promote a differentiation block towards astrocytes as well, these results indicate a potential synergistic role of *ACVR1* in H3.3 K27M pHGGs in promoting

oncogenesis³¹. We next turned our attention towards the effects on K27M mutations in H3.3 K27M pHGGs to on cell identity.

Chapter 3.2: Transcriptomic effects of H3.3 K27M in experimental model systems

3.2.1. H3.3 K27M-KO show upregulation of extracellular matrix gene programs *in vitro*

H3.3 K27M mutated pediatric pHGGs are thought to arise during neurodevelopment in a specific developmental window and cellular context^{23,25}. K27M mutations are an initiating event in tumorigenesis and lock the tumor cells into a stem-like state with high proliferative capacity through epigenetic dysregulation^{23,90,95}. The proposed cell-of-origin is thought to be an early glial progenitor cell which is blocked in differentiation potential towards the astrocytic lineage³¹.

To assess whether the H3.3 K27M (K27M) mutation has a direct effect in blocking differentiation of progenitors, CRISPR-Cas9 editing of the K27M mutation was performed in three cell lines derived from thalamic tumors (BT245, HSJ019) or a pontine tumor (DIPG13) to generate H3.3 K27M-KO (KO) cell lines. Cell lines were maintained in different media conditions; neural stem-cell proliferation media (SCM), neural stem-cell proliferation media with DMSO (DMSO), or serum-differentiation media (DM) for two weeks prior to RNA sequencing.

We first investigated transcriptomic changes related to cell identity using previously published samples³¹ and additional unpublished samples for H3.3 K27M cell lines and H3.3 K27M-KO cell lines. The data cohort is presented in Table 2.

Table 2: RNA seq cohort for H3.3 K27M serum differentiation experiments

Cell line	Genotype	Media	Replicates
DIPG13	K27M	DMSO	3
		SCM	2
		DM	2
	KO	DMSO	2
		SCM	8
		DM	4
BT245	K27M	DMSO	3
		SCM	5

	KO	DM	2
		DMSO	2
		SCM	4
		DM	4
HSJ019	K27M	SCM	4
		DM	2
	KO	SCM	6
		DM	2

We first used PCA to assess global differences in gene expression across cell lines and found that the pontine cell line (DIPG13) separated from the thalamic cell lines (BT245, HSJ019) which indicates that tumor location is accounting for larger differences in gene expression profiles than tumor-specific differences (Figure 7A). Consequently, tumor cell-of-origin expression profiles affects transcriptional profiles and may lead to different downstream effects. PCA for each cell line showed that culture media contributed the greatest source of variation in expression profiles for the pontine cell line (DIPG13) and one thalamic cell line (BT245), and genotype contributed the second largest source of variation in the other thalamic cell line (HSJ019) (Figure 7B-G). As the effect of media was larger than that of genotype for the majority of cell lines, subsequent comparisons were made comparing the effect of media while fixing genotype.

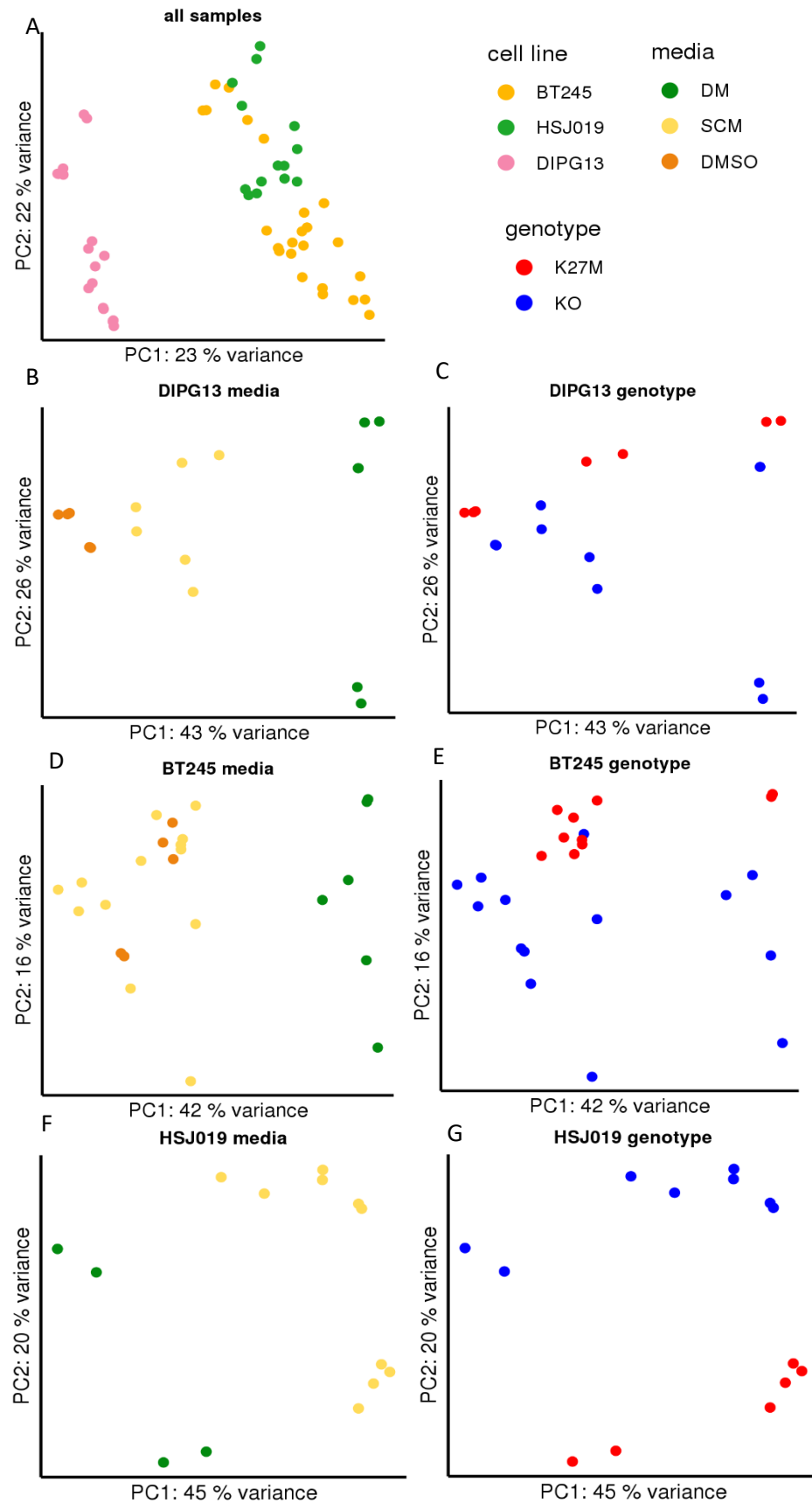
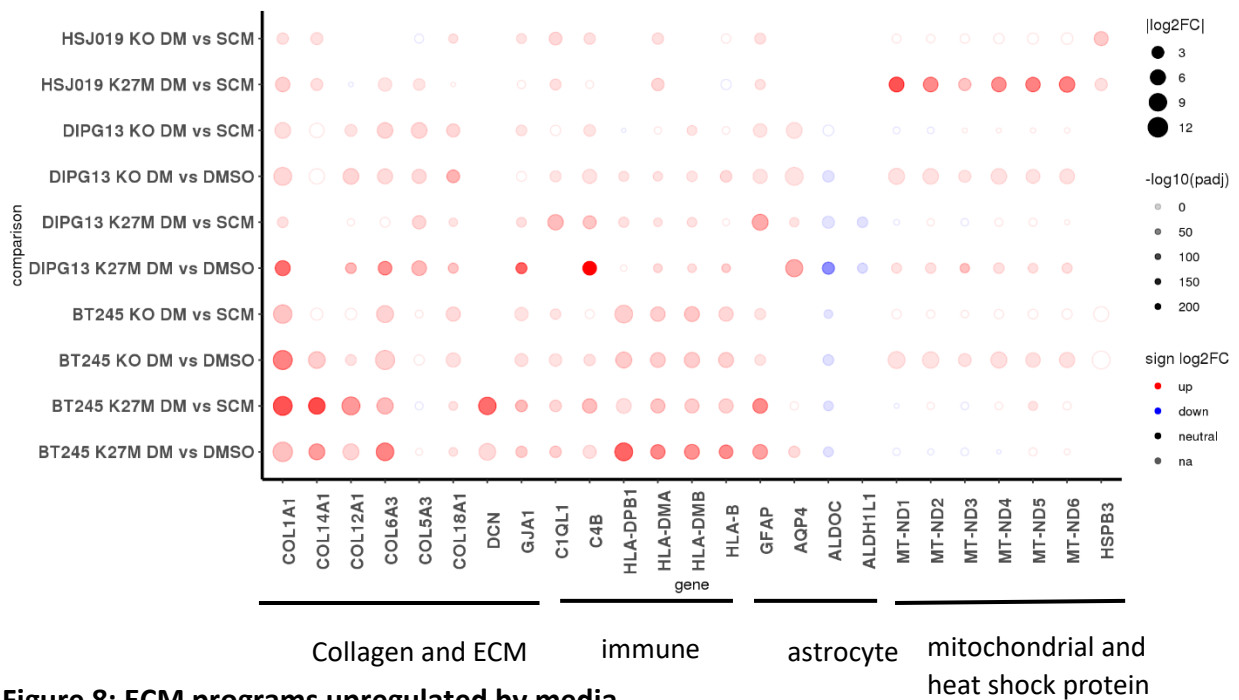


Figure 7: Media has a greater effect than H3.3 K27M status on global gene expression

(A) PCA using top 10000 most variable genes of all samples; **(B-C)** DIPG13; **(D-E)** BT245; **(F-G)** HSJ019. SCM: stem cell media; DM: differentiation media; DMSO: stem cell media with DMSO.

We used differential expression gene analysis comparing samples by media for each cell line to find genes significantly affected by the media. In the pontine cell line (DIPG13), differentiation media induced upregulation of potentially different pathways in each genotype (Fig 8). Top differentially expressed genes included upregulation of immune genes were noted in the K27M condition, while an upregulation of collagen genes was noted in the KO condition. For one of the thalamic cell lines, differentiation media was associated with an upregulation of collagen and extracellular genes in both genotypes and immune genes in the K27M condition (BT245). However, changes induced by differentiation media in the second replicate showed upregulation of cellular stress programs involving mitochondrial, ribosomal, and heat shock protein genes (HSJ019). Commonalities across two cell lines from different tumor locations (BT245, DIPG13) was the upregulation of collagen / extracellular and immune related genes. The collagen related genes may point to adaptive properties for cell culture gained induced by media. Interestingly, though an astrocytic enrichment is consistent with previous results³¹, the addition of canonical astrocyte marker genes showed both-down and up-regulation by differentiation media in this cohort.



While top differentially expressed genes pointed to a few signals, we performed pathway analysis to glean a wider characterization of biological pathways affected by differentiation media. We kept the same comparisons as those made in differential gene expression analysis; we compared the effect of differentiation media to stem cell media and DMSO separately within the same cell line. In the pontine cell line (DIPG13), comparing the effect of DM in both genotypes resulted in upregulation of genes associated with extracellular matrix related terms

(cell adhesion, extracellular matrix) and developmental terms (cell differentiation, developmental processes) (Fig 9A-D). For both of the thalamic cell lines, genes associated with developmental processes were upregulated in both genotypes in differentiation media (system development, cell differentiation) (Fig 10A-F). Additionally, upregulation of extracellular matrix and cellular adhesion associated pathways in for one replicate (BT245) in both genotypes. In all cell lines, the developmental signals gained in differentiation media may be due to the loss of gene expression promoting stem-like properties actively maintained in stem cell media. The gain of extracellular matrix and adhesion programs is consistent with the results from differential expression analysis, indicating that an overwhelming function of differentiation media may help cells survive in culture conditions.

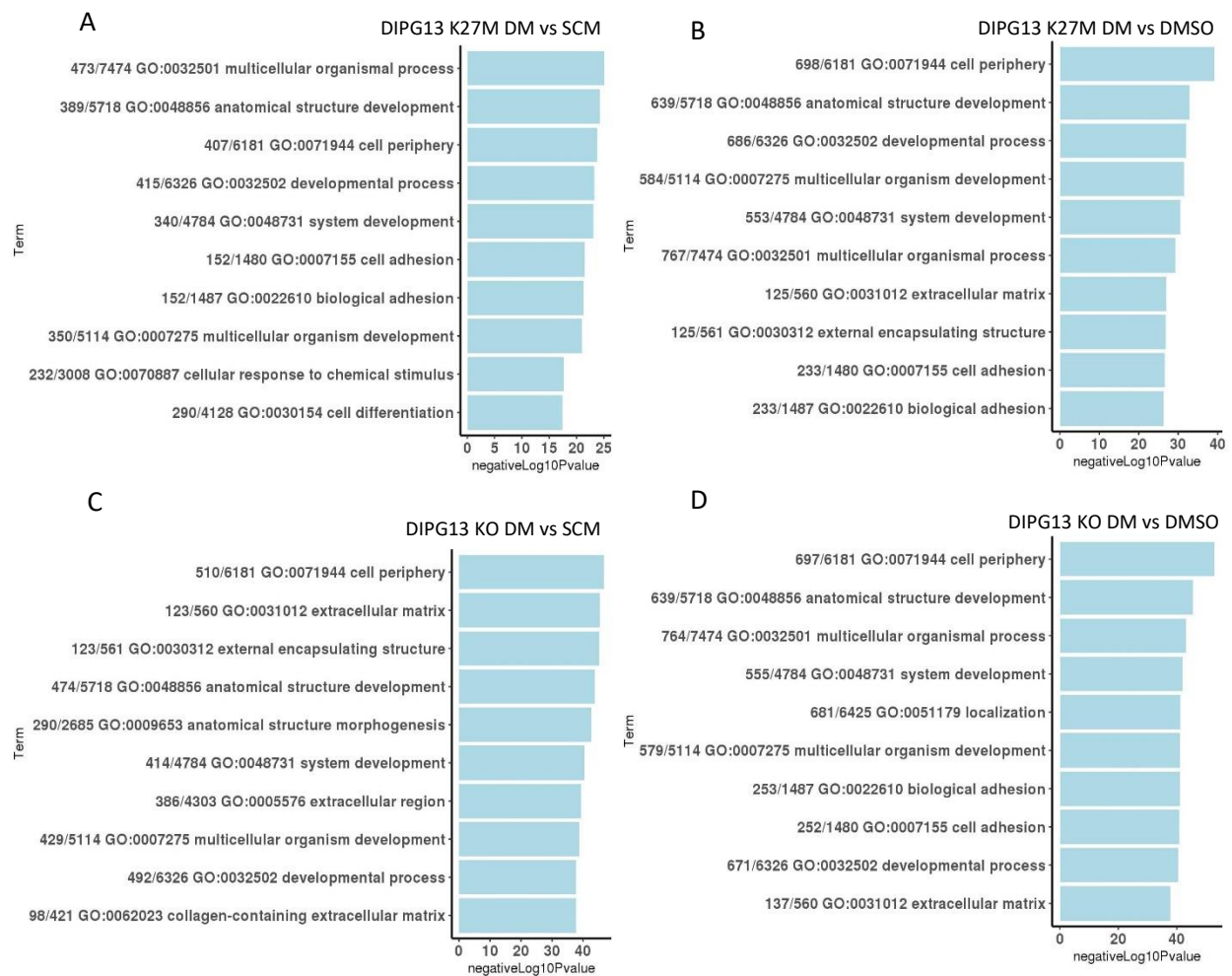


Figure 9: Developmental and cellular adhesion pathways induced by media for pontine H3.3 K27M and H3.3 K27M-KO cell line

(A) Pathway analysis using significantly upregulated genes (mean normalized read count > 100, adjusted p-value < 0.05, log2FoldChange > 1) in DIPG13 comparing K27M DM to SCM; **(B)** K27M DM to DMSO; **(C)** KO DM to SCM; **(D)** KO DM to DMSO DM: differentiation media; SCM: stem cell media; DMSO: stem cell media with DMSO

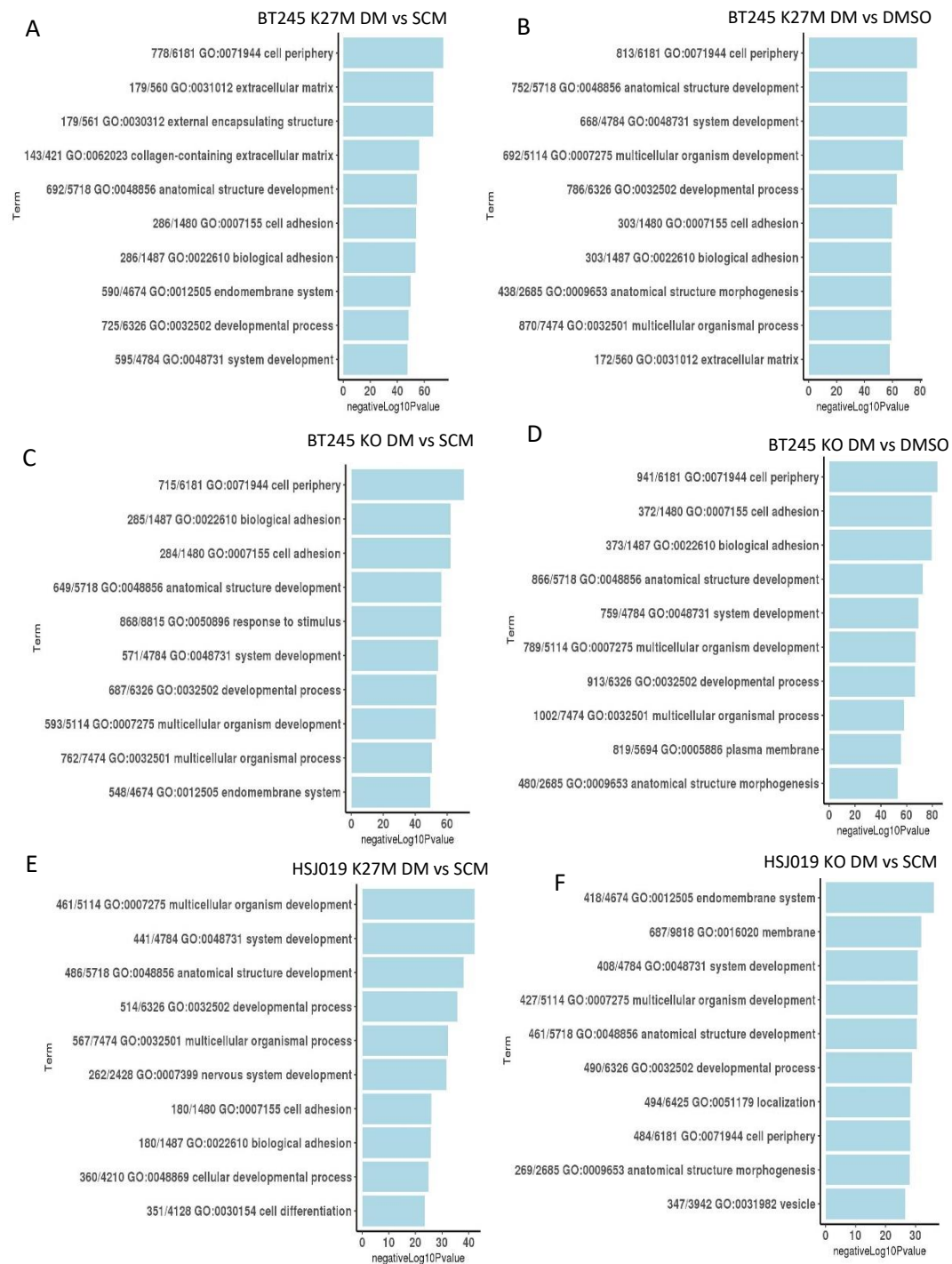


Figure 10: Developmental pathways induced by media for thalamic H3.3 K27M and H3.3 K27M-KO cell lines

(A-D) Pathway analysis using significantly upregulated genes (mean normalized read count > 100, adjusted p-value < 0.05, log2FoldChange > 1) in BT245; **(E-F)** HSJ019. DM: differentiation media; SCM: stem cell media; DMSO: stem cell media with DMSO

We next determined if the developmental signal gained in differentiation media mapped towards a particular cell state. From previously published results with fewer replicates, the K27M mutation stalled differentiation towards the astrocytic lineage for the pontine cell line³¹. Since H3.3 K27M tumors are thought to arise from an OPC-like cell during neurodevelopment, we used a single-cell reference of cell-type identity signatures spanning the developing murine forebrain and pons as described in section 3.2. Using gene set enrichment analysis with this reference, a commonality across all cell lines was an enrichment of non-neuroectodermal cell-identity signatures (pericytes, vascular smooth muscle, endothelial) coupled to a depletion of proliferating progenitor signals (Fig 11A). To help corroborate these results, genes which drive the enrichment, or leading edge genes, were plotted in the data. A strong enrichment of leading edge genes was observed for non-neuroectodermal signatures (Fig 11B) and strong depletion for proliferating progenitors (Fig 11E). Since stem cell media actively maintains self-renewal and stem-like properties in culture, the depletion of proliferating progenitor signatures is perhaps unsurprising. However, the enrichment towards non-neuroectodermal cell type signatures is unexpected since the cell-of-origin is an early glial progenitor, which should not have the capability towards differentiating towards a completely different cellular lineage. A more plausible rationale for this result, given the earlier findings of upregulation of collagen and extracellular matrix associated genes and pathways in differentiation media, may be that these upregulated genes are masking glial cell identity signals.

However, there were some changes to glial cell identity in differentiation media in a cell-line specific manner. In the pontine cell line DIPG13, an enrichment of astrocyte and ependymal gene programs was enriched in differentiation media for one K27M comparison, but not as robustly in the KO (Fig 11A, C). This indicates that an effect of media may be to promote a baseline level of induction towards the astro-ependymal lineage in DIPG13, which is consistent with previous results³¹. Additionally, in a thalamic cell line (BT245), a depletion of oligodendrocyte and OPC gene programs was observed (Fig 11D). While cells are not thought to

originate from mature oligodendrocytes, the movement away from OPCs is consistent with the idea of a differentiation blockade stalling cells in an OPC-like glial progenitor state^{34,163}.

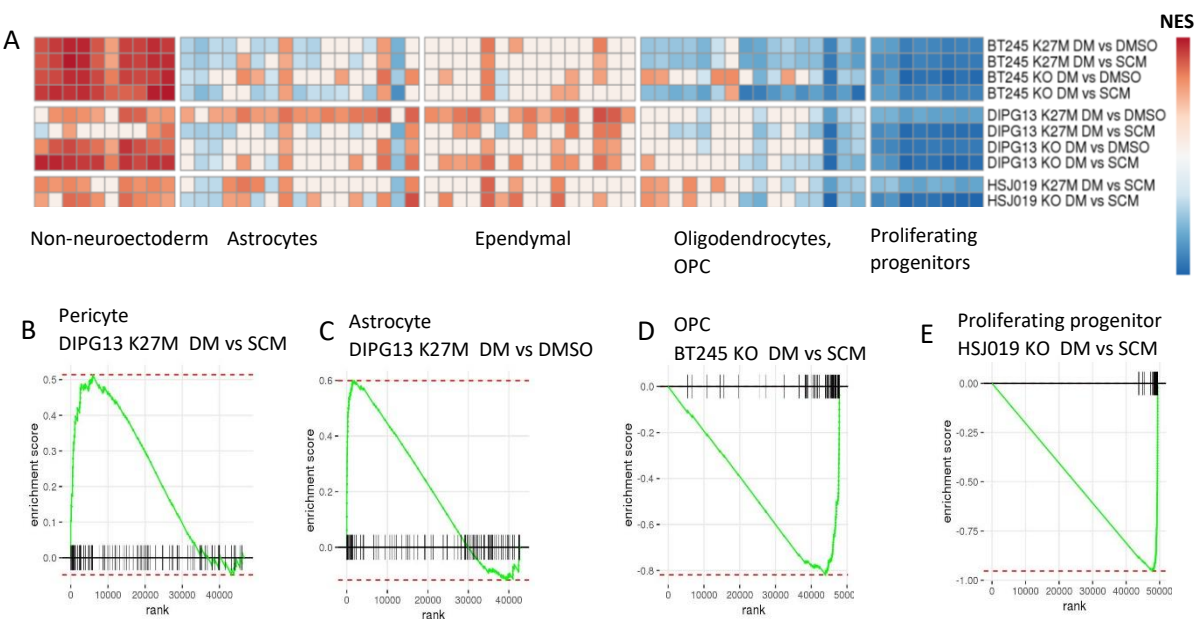


Figure 11: Enrichment of non-neuroectodermal gene programs in differentiation media

(A) Gene set enrichment analysis using cell-type identity signatures from a single cell developing murine reference (Jessa et al, 2019) comparing differentiation media to stem cell media and DMSO conditions. **(B)** representative leading edge plots for pericyte signature in DIPG13 K27M condition; **(C)** astrocyte signature in DIPG13 K27M condition; **(D)** OPC signature in BT245 KO condition; **(E)** proliferating progenitor signature in HSJ019 KO condition. Scores not significant at adjusted p-value < 0.05 set to 0. NES: normalized enrichment score; OPC: oligodendrocyte precursor cell.

Since a potential switch glial cell identities in differentiation media may be masked by overwhelming expression of collagen and extracellular matrix gene programs, we asked whether this signal was common to all cells or driven by a subset. Single-cell RNA sequencing data (scRNA seq) was generated for the same three cell lines in stem cell media (SCM) and differentiation media (DM) conditions. The data cohort is presented in Table 3.

Table 3: scRNA seq cohort for H3.3 K27M serum differentiation experiments

Cell line	Genotype	Media	Replicates
DIPG13	K27M	SCM	1
		DM	1
	KO	SCM	2
		DM	3
BT245	K27M	SCM	1
		DM	1
	KO	SCM	1
		DM	1
HSJ019	K27M	SCM	1
		DM	1
	KO	SCM	2
		DM	2

To assign cell-type identity, individual cells were projected to the same developing murine atlas used earlier through spearman correlation between expression profiles of the cell and reference cell types. The reference cell type with the highest correlation was assigned as the nearest-normal cell type in the data. To assess broad changes to cell type identity caused by media or genotype conditions, cells proportions were computed by summing the total number of cells for a given nearest-normal cell type over the total number of cells in the specific genotype and media condition (K27M cells in stem cell media, K27M cells in differentiation media, KO cells in stem cell media, KO cells in differentiation media) within cell lines. Of note, the majority of cells in stem cell media across cell lines mapped to RGCs which are a more primitive glial progenitor than the consensus OPC-like state in K27M cells (Fig 12A-C). However, mean expression of canonical RGC marker genes (VIM, NES, PAX6, HES1, SOX2) were not highly specific to cells labeled as RGCs across cell lines and therefore the match to RGCs may be due to stem-like factors actively maintained in stem cell media (Fig 12D-F). The predominant effect of differentiation media across cell lines was to promote cells along a differentiation trajectory towards a non-neuroectodermal cell state. In the thalamic cell lines BT245 and HSJ019, a

minority of cells in both K27M and KO conditions progressed in differentiation towards mature glial cell types of astrocytes and oligodendrocytes in differentiation media, an effect which was greater in the KO condition. This is compatible with the idea that differentiation media promotes a baseline level of cellular differentiation and that removal of the K27M mutation acts in synergy to promote differentiation.

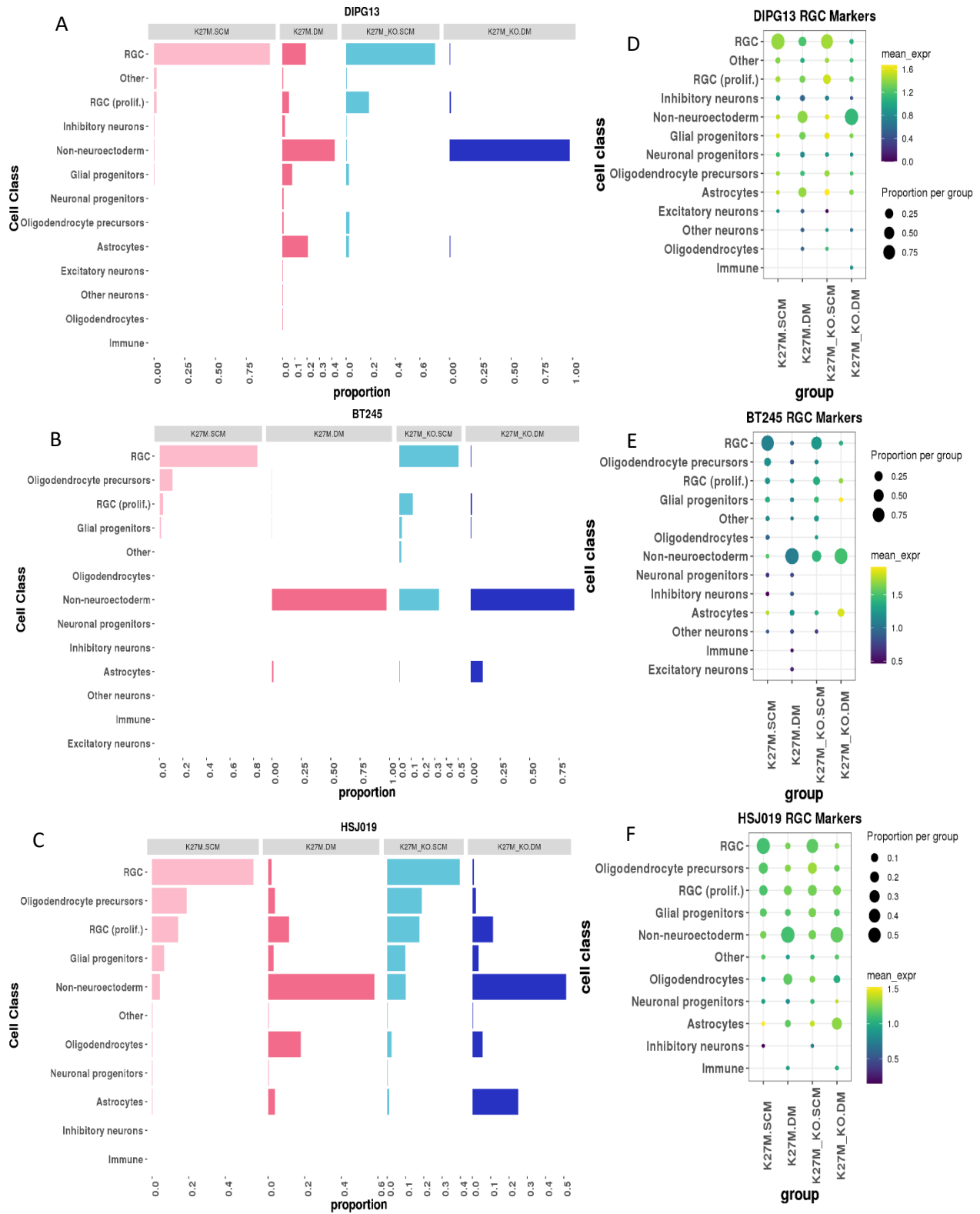


Figure 12: Non-neuroectodermal cell state induced by differentiation media

(A) Proportions of cell types assigned by Spearman correlations in DIPG13; **(B)** BT245; **(C)** HSJ019. **(D)** Mean expression of normalized counts for RGC marker genes VIM, NES, PAX6, SOX2 and HES1 in DIPG13; **(E)** BT245; **(F)** HSJ019.

However, the majority of cells in differentiation media mapped toward non-neuroectodermal cells, which is a completely different lineage than glial cells. Given the overwhelming expression of collagen and extracellular matrix promoting gene programs promoted by differentiation media in the same cell lines in bulk RNA seq, it is likely that the high expression of similar gene programs is driving a match towards non-neuroectodermal cell identity. To assess if other cell-identity related programs are masked by non-neuroectodermal gene programs in the same cells, non-neuroectodermal cells were extracted and projected to the same cell-type reference without non-neuroectodermal cell identities. The highest correlation was assigned as the 'second match' cell-type and cell proportions were computed per genotype-media condition. A large switch in cell proportions towards an astrocyte-like state was observed in differentiation media across cell lines, with the one exception of the KO condition of the pontine cell line (Fig 13). Additionally the thalamic cell line HSJ019 showed progression towards the oligodendrocyte lineage which may reflect tumor-specific differences. Therefore, a secondary effect of differentiation media may be to promote cells in progression along the astrocytic lineage, which consistent with prior results³¹. The primary effect of differentiation media, however, may be to induce non-neuroectodermal gene programs which help cells survive in culture.

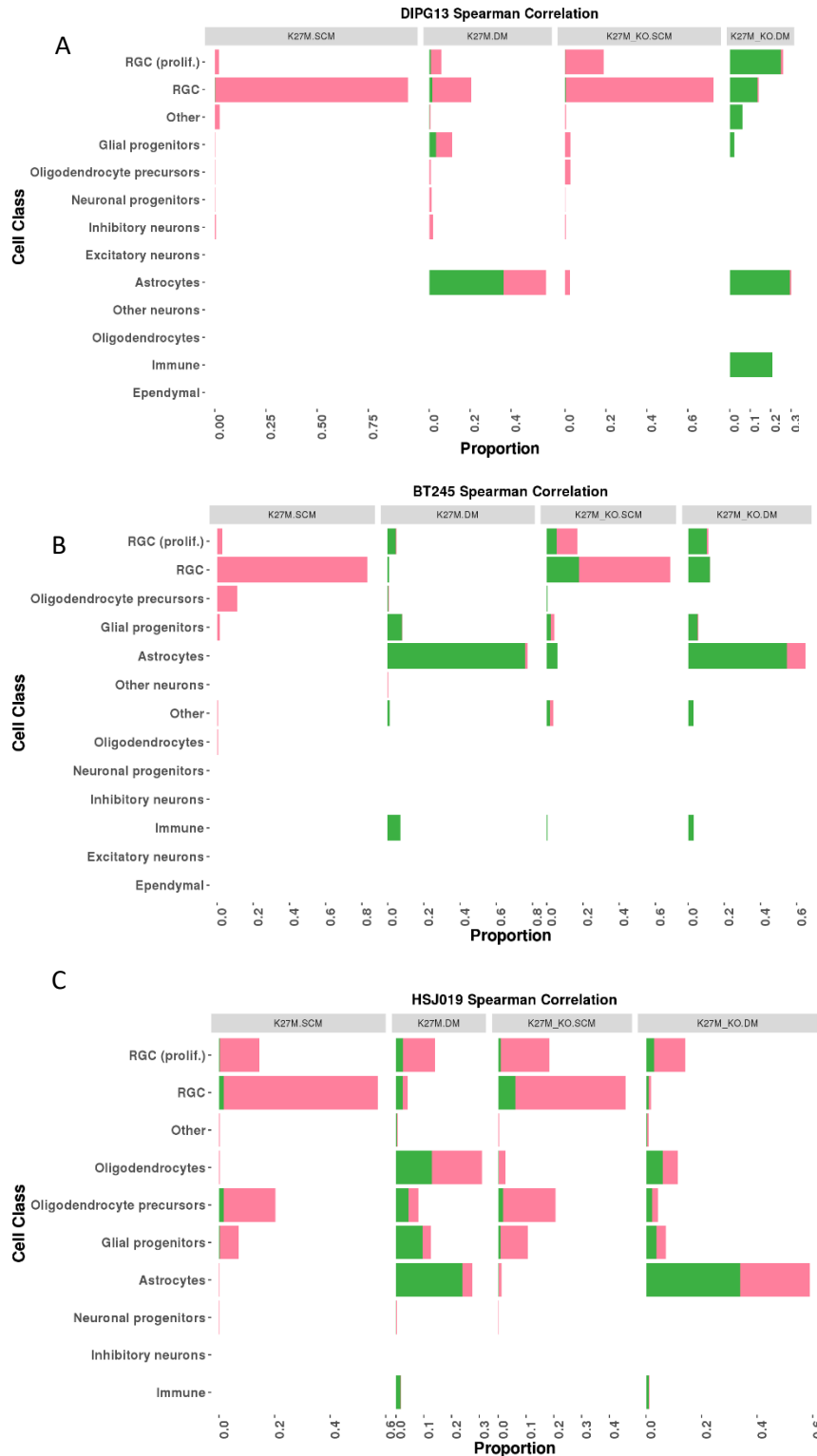


Figure 13: Non-neuroectodermal cell states mask astrocyte gene programs in differentiation media

(A) Cells originally mapping to non-neuroectodermal cell identity (green) were extracted, projected to neuroectodermal cell types, and cell proportions per condition re-computed in DIPG13; (B) BT245; (C) HSJ019. Cells originally mapping to neuroectodermal cell types are displayed in pink.

We concluded that the effect of media was greater than that of the genotype in serum-differentiation experiments, and that the primary effect of differentiation media was the upregulation of collagen and extracellular matrix associated gene programs which may help in adaption to culture conditions. In assessing whether differentiation media promoted progression along a differentiation trajectory, we observed an upregulation of predominantly non-neuroectodermal cell states which may mask mature glial cell identity signals in single cells. In a minority of cells, a progression in differentiation towards astrocytes was observed, consistent with previously published results³¹. We observed an increase in oligodendrocyte gene expression in one thalamic cell line, potentially indicating maturation along the oligodendrocyte lineage may be tumor-specific. To explore changes in cell state and uncouple the effects of differentiation media, we moved away from an *in vitro* model and used an *in vivo* model.

3.3.2 H3.3 K27M-KO map oligodendrocytes and astrocytes *in vivo*

In vitro cellular models present the limitation of inducing expression changes for adaptation to culture conditions. To move away from potential effects of culture of media induced *in vitro*, we next investigated differentiation state of tumor cells using orthotopic xenograft models. Briefly, brain tumor cells were transfected with lentiviruses expressing GFP. Cells were injected into orthologous regions in the mouse brain; into the caudate putamen for thalamic cell lines BT245 and HSJ019, and into the pons for pontine cell line DIPG13. Mice were monitored for clinical symptoms (weight loss, epilepsy) and euthanized upon clinical endpoint which was reached 100 days post-injection to upwards of one year. Dissected tumor and normal brain tissue was sorted for GFP+ human tumor cells before sending for sequencing. Samples were aligned to a joint human and mouse reference, human counts were extracted and used for further downstream analysis. Due to a batch effect, we analyzed batches separately. The data cohort is presented in Table 4.

Table 4: RNA seq cohort for H3.3 K27M orthotopic xenograft models

Cell line	Genotype	Replicates	Batch
DIPG13	K27M	3	1
		3	2
	KO	3	1
		4	2
BT245	K27M	3	1
		2	2
	KO	2	1
		3	2
HSJ019	K27M	4	1
		4	2
	KO	5	1
		4	2

To assess changes in global gene expression, we performed PCA of all samples. We found separation by cell lines in both batches, with DIPG13 separating from HSJ019 and BT245 (Fig 14A-B). The separation of DIPG13 from HSJ019 and BT245 may indicate differences due to anatomical location of the original tumors, which is consistent with differences found in serum differentiation experiments. PCA keeping cell lines separate revealed separation of samples by genotype for the thalamic cells, but not the pontine cell lines (Fig 14C-G). Thus we concluded that genotype drove the most greatest source of variation in gene expression profiles for thalamic cell lines, but not the pontine cell line. This is consistent with previous results in the pontine cell line from differentiation experiments.

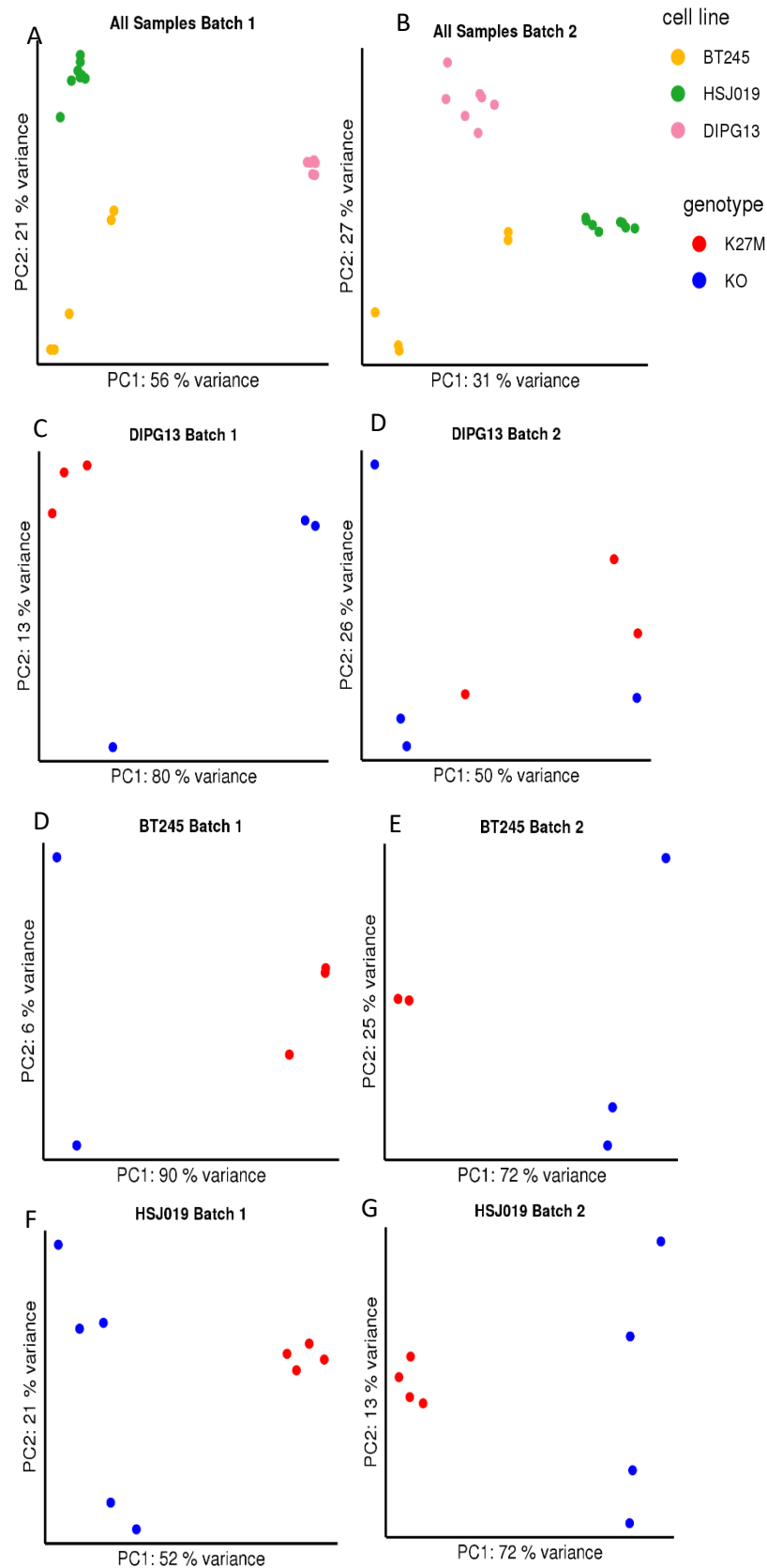


Figure 14: Genotype drives most source of variation in xenografts derived from thalamic cell lines

(A) PCA on top 10000 most variable genes across cell lines in batch 1; **(B)** batch 2; **(C)** pontine cell line DIPG13 in batch 1; **(D)** DIP13 in batch 2; **(E)** thalamic cell line BT245 in batch 1; **(F)** BT245 in batch 2; **(G)** thalamic cell line HSJ019 in batch 1; **(H)** HSJ019 in batch 2.

To identify changes in gene expression and associated biological pathways between genotypes, we performed differential gene expression analysis followed by pathway analysis (Fig 15). We found an upregulation of terms relating to development in xenografts derived from the thalamic cell lines. In one xenograft (BT245), terms relating to the extracellular matrix were additionally observed. Therefore, while removal of the K27M mutation may cause direct effects relating to development, there may also be an upregulation of mesenchymal gene programs.

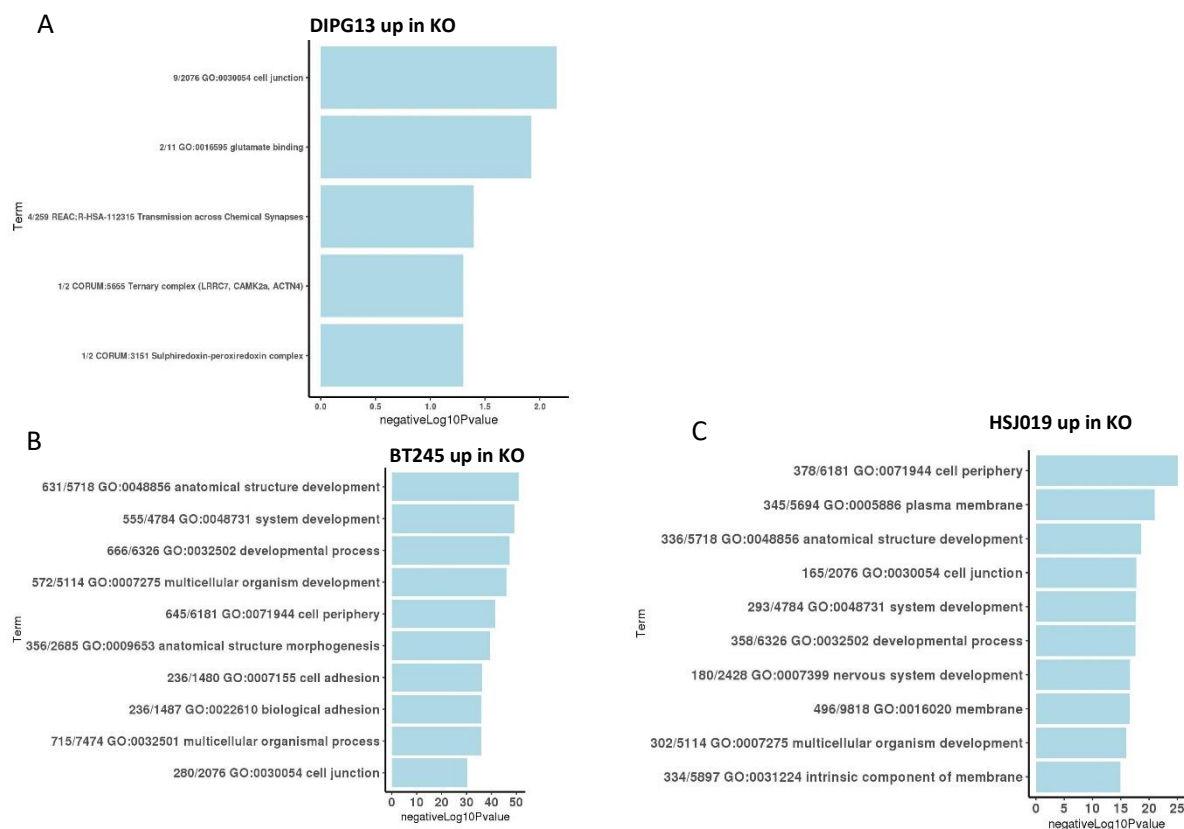


Figure 15: Developmental terms are enriched in K27M-KO xenografts

(A) pathway analysis using significantly upregulated genes (baseMean > 100, adjusted p-value (Benjamini-Hochberg) < 0.05, log2FoldChange > 1) in pontine cell line DIPG13 KO batch 1; **(B)**

thalamic cell line BT245 batch 1; **(C)** thalamic cell line HSI019 batch 1. The first two numbers of the label indicates number of differentially expressed genes found in gene list over total size of the list.

Given upregulated genes were associated with developmental signals in the K27M-KO condition, we asked if the K27M mutation directly impacted cellular differentiation towards a more terminal state. To investigate changes related to cell type identity, GSEA was run using a developmental murine atlas described in section 3.2. We observed an enrichment in oligodendrocyte and OPC signals across xenografts in the KO (Fig 16A). There were additionally some cell-line specific differences; an additional enrichment in astrocytic cell identity was observed for one thalamic cell line (BT245) and the pontine cell line (DIPG13), and an additional enrichment in non-neuroectodermal cell identity in one thalamic cell line (BT245).

Of note, genes driving the non-neuroectodermal cell identity signal observed in BT245 mapped to upregulation of collagen genes, and similar gene programs were upregulated from the serum-differentiation experiments for the same cell line (Fig 16B). Due to detection of the same signal in different experimental systems, it may be plausible that an effect of removing the H3.3 K27M mutation is in part to upregulate mesenchymal-related gene programs in addition to promoting glial cell maturation. However, this finding is not recaptured for the other cell lines, indicating that it may be a tumor-specific effect. Across cell lines and tumor locations, an upregulation of oligodendrocyte and OPC gene programs was observed. This is compatible with the idea that cell-of-origin is in an early glial progenitor in the OPC lineage¹⁶³. It is possible that removing the K27M mutation then directly functions to promote differentiation along the oligodendrocyte lineage to an extent. The increase in astrocytic gene programs in the pontine cell line is consistent with part of the serum differentiation experiments. Additionally, the upregulation observed in the thalamic cell line indicates that the switch towards astrocyte cell identity may be a common feature across tumor locations and potentially distinct cells of origin. The dual enrichment of astrocyte and oligodendrocyte cell-identity raises a question of whether these gene programs co-exist in the same cells, potentially suggestive of an abnormal differentiated state, or are expressed by distinct cells.

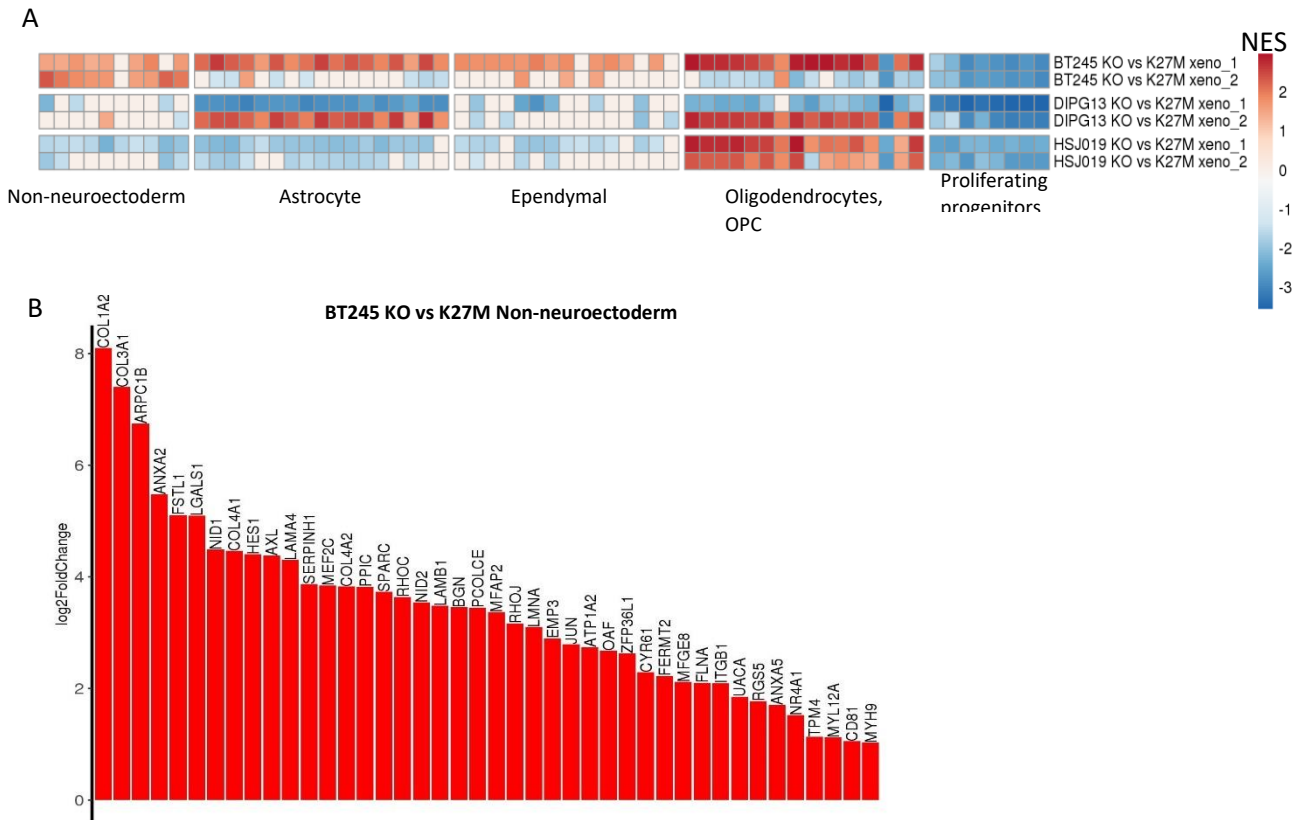


Figure 16: K27M-KO cells express astrocyte and oligodendrocyte gene programs in xenograft models

(A) Heatmap of normalized enrichment scores in xenograft comparisons. Scores which are not significant are set to 0. **(B)** Waterfall plot of log2FoldChange of leading edges driving non-neuroectodermal signal in BT245 xenograft batch 2. Gene significant with p-value threshold by differential expression analysis are in dark red. NES: normalized enrichment score; OPC: oligodendrocyte precursor cells.

To test whether cells acquire expression of astrocytic or oligodendrocyte gene programs, we investigated gene expression changes of individual cells by scRNA seq for the same cell lines. The data cohort is presented in Table 5.

Table 5: scRNA seq cohort for H3.3 K27M orthotopic xenograft models

Cell line	Genotype	Replicates
DIPG13	K27M	2

	KO	2
BT245	K27M	1
	KO	1
HSJ019	K27M	2
	KO	3

Briefly, samples were aligned to genome builds hg19 and mm10, and hg19 counts extracted. Human cells were called through the 10X pipeline. Due to low numbers of human cells, cells which were labeled as multiplets through the 10X pipeline but contained greater than 75% of reads mapping to hg19 were included as human cells for downstream analysis. Cells were projected to nearest normal cell type through correlation analysis.

To identify changes in cell type identity associated with removal of the mutation, nearest normal cell types were counted for each genotype. Across xenografts, cells with the K27M mutation predominantly mapped to a variety of early glial progenitors including RGCs and OPCs, but few mature glial cells (Fig 17A-C). Removal of the mutation caused a heterogeneous shift along a differentiation trajectory, with a slight increase in the number of cells mapping to astrocytes and oligodendrocytes. The xenograft model which was enriched for extracellular matrix-related gene programs from bulk RNA sequencing displayed a near complete switch towards non-neuroectodermal cell identity. Therefore at the level of individual cells *in vivo*, xenograft cells may exist along a continuum of states. Removing the K27M mutation induces progression along a differentiation trajectory and is consistent with a progression along the glial lineage for some cells.

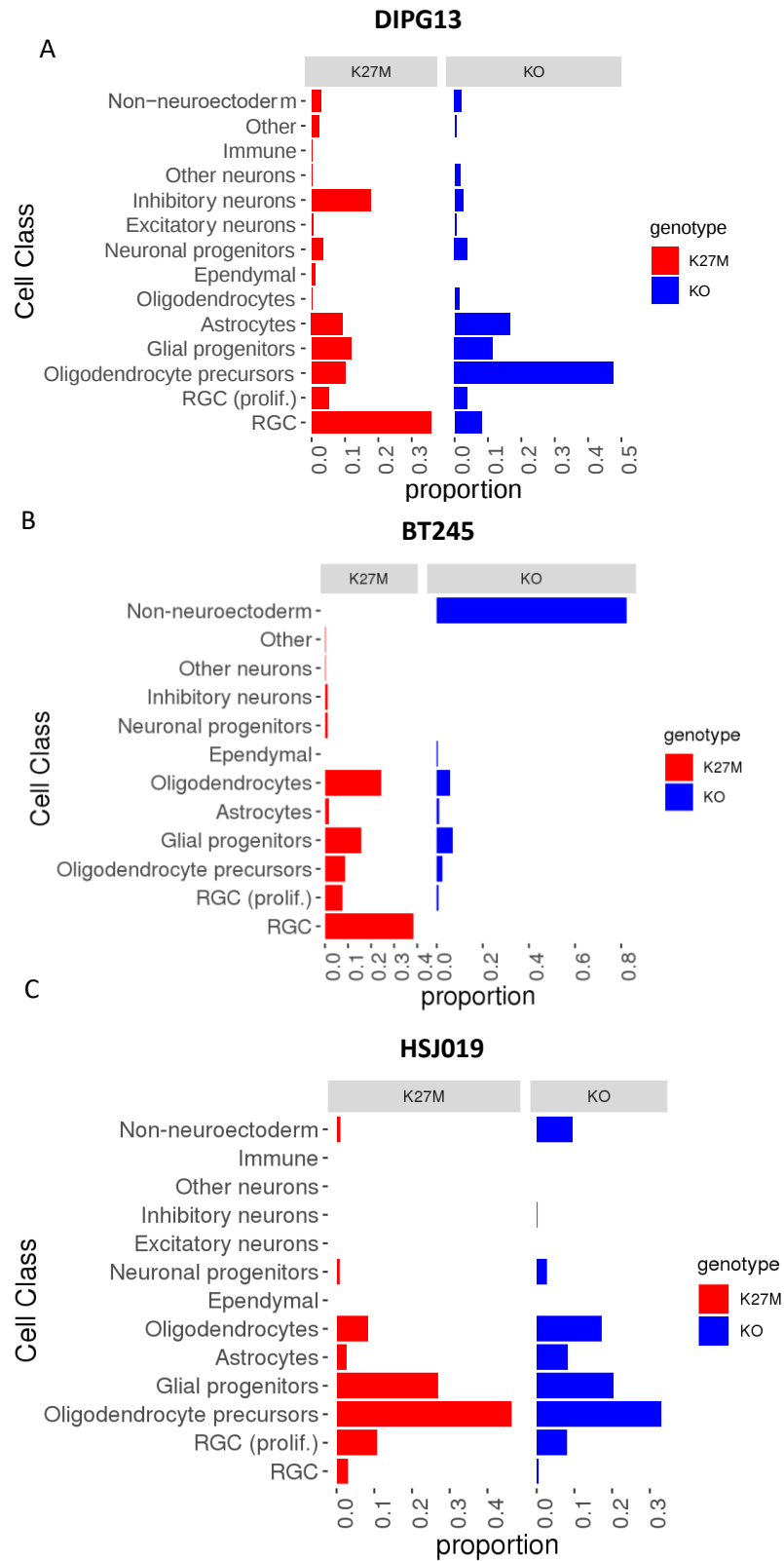


Figure 17: Glial cell heterogeneity of K27M and K27M-KO cells in xenograft models

(A) Cell proportions in K27M and KO conditions for xenograft cells derived from pontine cell line DIPG13; **(B)** thalamic cell line BT245; **(C)** thalamic cell line HSJ019.

In summary, our results are consistent with the idea that a direct effect of H3.3 K27M is to block differentiation along the glial lineage through both *in vitro* and *in vivo* models. We find that a major effect of *in vitro* serum-differentiation experiments is to upregulate collagen and extracellular matrix-related gene programs, suggesting cells are adapting to media using pathways normally associated with non-neuroectodermal cell types. Indeed, the same cells which map to non-neuroectodermal cell types also express astrocyte cell-identity related gene signatures. Therefore subjecting cells to differentiation media may serve to mask glial cell identity signals. Moving away from *in vitro* experiments to directly assess the effect of the K27M mutation on cell identity, we find upregulation of astrocyte and oligodendrocyte cell identity gene programs in xenografts in the KO condition through bulk RNA sequencing. At the level of single cells, however, a switch towards an exact cell identity is not clear. Altogether, these results support the idea of a direct differentiation blockade induced by H3.3 K27M along the glial lineage.

3.3. Transcriptomic effects of germline G34 mutations in a murine model

3.3.1. Early postnatal G34 mutants have minor transcriptomic changes in the brain

While effects of somatic histone mutations have been studied in pediatric high grade gliomas and are thought to arise as defects in neurodevelopment, little is known about the impact of histone mutations in development. Somatic mutations at glycine 34 in H3.3 cause different tumors with surprising tissue specificity. G34R/V mutations drive tumorigenesis in approximately 30% of pHGGs occurring in the cerebral cortex and are thought to arise in an interneuron lineage during brain development^{9,32}. In contrast, G34W mutations drive over 90% of GCTB and are thought to arise from an osteoblast-like progenitor^{17,28}. To explore the remarkable tissue specificity of G34 mutations and their roles in development, germline heterozygous *H3F3A* H3.3 G34R/V/W mice were generated using an innovative CRISPR direct-knock-in strategy (Fig 18A). Briefly, single-guide RNA (sgRNA) targeting exon 2 of *H3f3a* injected with a homology-directing repair (HDR) sequence containing the desired G34 nucleotide substitution sequences and silent blocking mutations at the PAM site. The sgRNA, HDR

sequence, and Cas9 protein were injected into embryos at the 2- or 4-cell stage. Embryos were transplanted into pseudo-pregnant hosts to generate mosaic founder mice. Founder mice were backcrossed with wild type (WT) mice to generate true germline heterozygous mice. To control for a potential dosage effect of *H3f3a*, germline hemizygous mice (INDEL) were generated with the same strategy.

The G34 mutant mice displayed a phenotypic spectrum. All mice were born small but gained weight and overshot their WT peers. G34R/V mice had predominantly neural defects affecting motor functioning including hindlimb clasping, abnormal gait, and paresis noticeable by adulthood; G34W mice had predominantly mesenchymal defects with enlarged bladder and the most marked obesity by adulthood (Fig 18B). Strikingly, male G34W mice died of urinary tract problems at 3 months of age. To investigate genome-wide gene expression changes contributing to these phenotypes, mice were sent for RNA-sequencing at one week following birth (postnatal day 7, P7) and adulthood (10 weeks, 10W). single-cell RNA-sequencing was performed at the juvenile stage (postnatal day 21, P21).

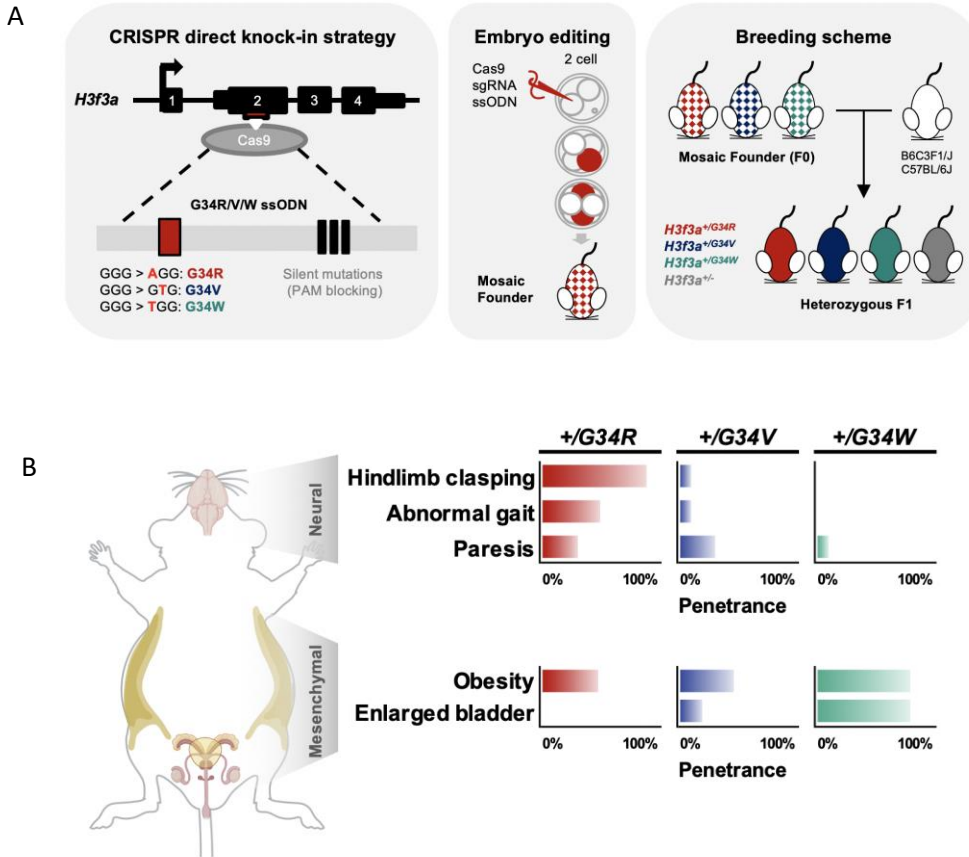


Figure 18: Germline knock-in of H3.3 G34R/V/W mice result in tissue specific defects

(A) CRISPR-Cas9 editing strategy to generate H3.3 G34R/V/W mice. Cells were injected with single-guide RNA (sgRNA) with ssODN repair template at the 2- to 4-cell stage. Mosaic founders were backcrossed with wild type mice to generate germline heterozygous H3.3 G34R/V/W mice. *Figure generated by Sima Khazaei and Carol Chen.* **(B)** G34R/V/W mice display distinct defects where phenotype was scored over at least 10 mice. *Figure generated by Sima Khazaei and Carol Chen.*

One week post birth, the mice do not exhibit any phenotypic symptoms. To investigate potential dysregulation in gene expression preceding onset of symptoms, bulk RNA seq data was generated at P7 for G34R, G34V, G34W, Indel, and WT mice in 4 regions of the brain (cortex, cerebellum, midbrain, pons) and bladder tissue, with 2 replicates each (N = 50). We first used PCA to assess if genotype contributed to a large source of variation in global gene expression profiles. Comparing tissues separately, we found that genotype did not contribute

greatly to changes in global gene expression in all sampled tissues (Fig 19A-E). To find if any genes were significantly up- or down-regulated in the mutants, we performed differential gene expression analysis. Across all tissues and comparisons, we found very few differentially expressed genes, indicating that histone status is not driving many transcriptomic changes one week post-birth (Fig 20A-E).

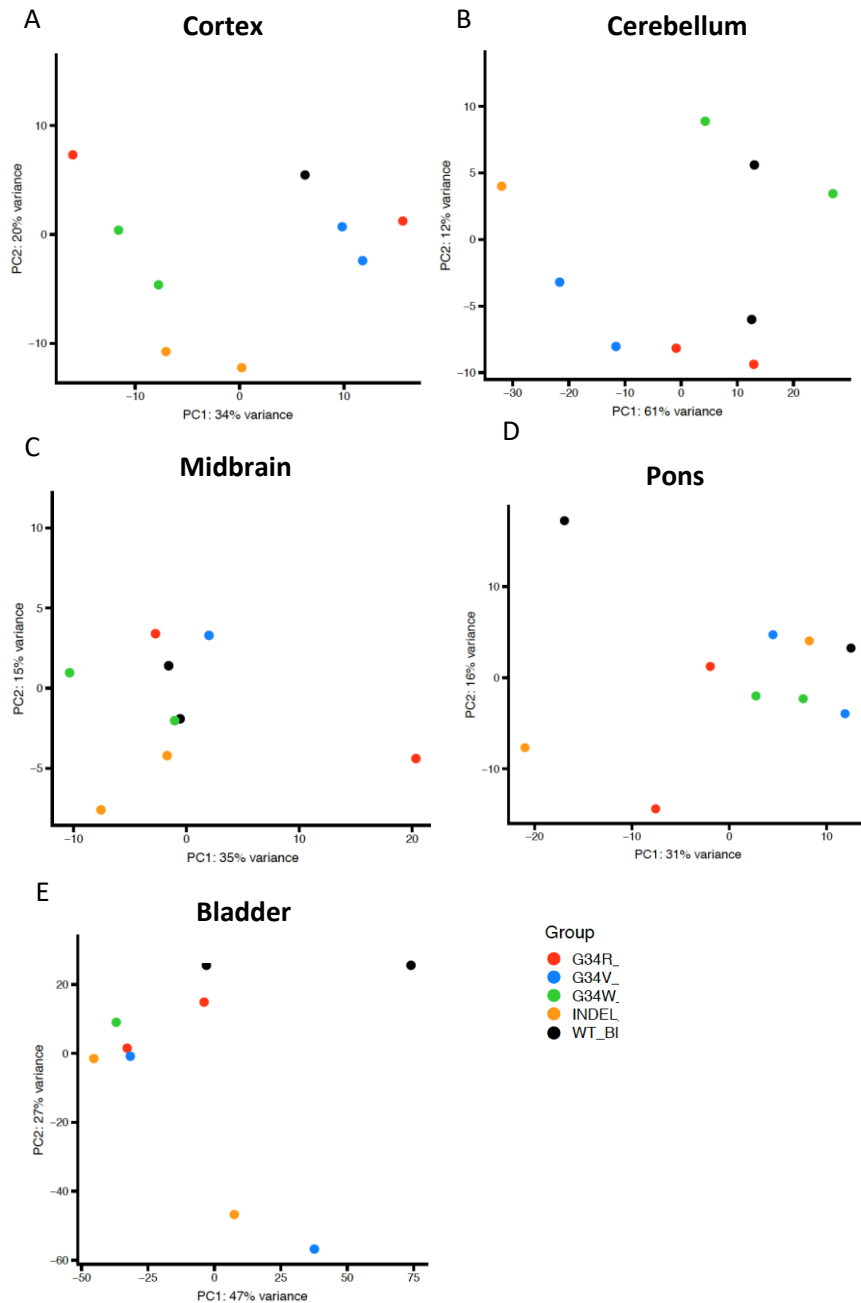


Figure 19: G34R/V/W mutation does not drive major source of variation in gene expression one week post birth

(A) PCA using top 10000 most variable genes in the cortex; **(B)** cerebellum; **(C)** midbrain; **(D)** pons; **(E)** bladder

To identify pathways associated with differentially expressed genes, pathway analysis was run. For all genotypes in the cortex, upregulated genes in the G34 mutants were associated with signaling pathways (Fig 21A-C). For cerebellar, midbrain, pons, and bladder tissues, few terms were found enriched. Altogether, concluded that histone mutation status was not significantly driving many changes in gene expression one week post-birth. These results were somewhat surprising given that H3.3 G34 mutants function to disrupt H3K36 methylation, and that loss of H3K36 methylation writer proteins NSD1, NSD2, and SETD2 result in embryonic lethality or severe developmental delay^{30,126,127,141}. However, given that H3.3 G34 mutants locally affect H3K36 methylation in H3.3-G34 mutant bearing nucleosomes and that H3.3 G34 mutant mice are phenotypically indistinguishable from wild type counterparts at this age, it is not entirely unsurprising that we failed to detect large changes in gene expression due to histone mutation status at this age.

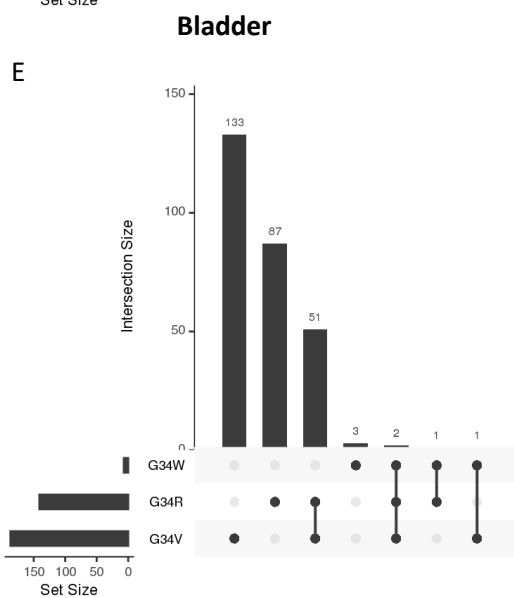
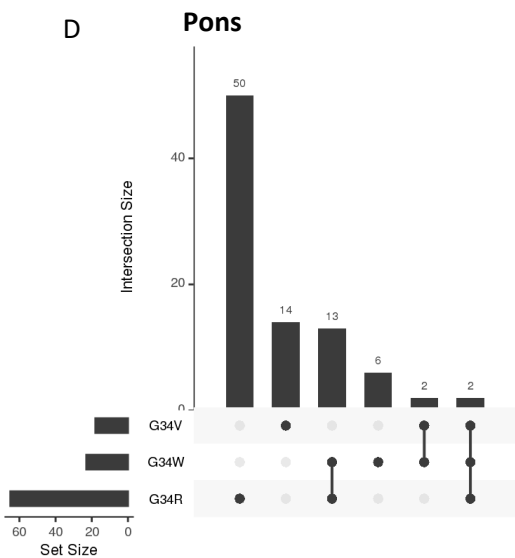
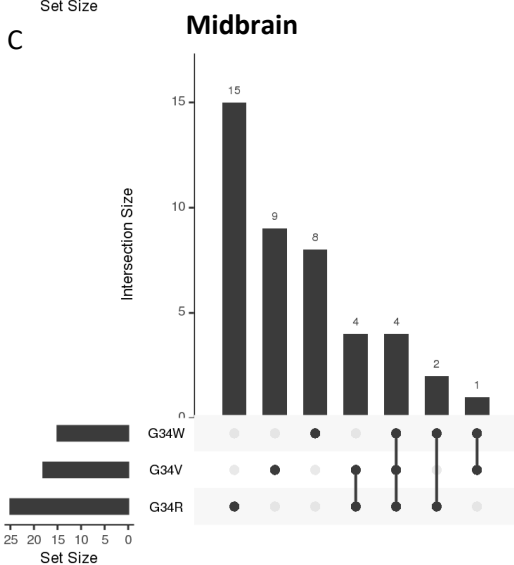
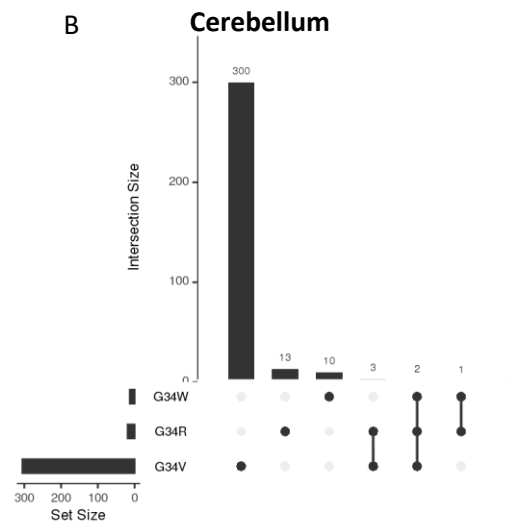
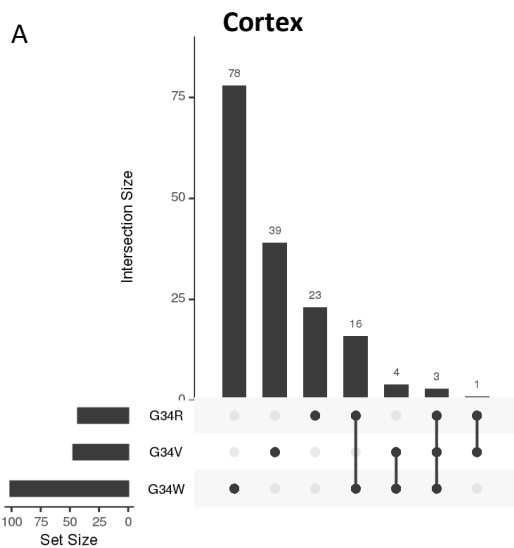


Figure 20: G34R/V/W mutation induces few transcriptomic changes at one week post birth

(A) Upset plot of differentially expressed genes per G34R/V/W mutants compared to wild type controls at a threshold of mean normalized expression > 100 counts, absolute log2FoldChange > 1, and adjusted p-value < 0.05 for cortex; **(B)** cerebellum; **(C)** midbrain; **(D)** pons; **(E)** bladder. Y-axis shows number of differentially expressed genes while x-axis shows overlaps of differentially expressed genes across genotypes.

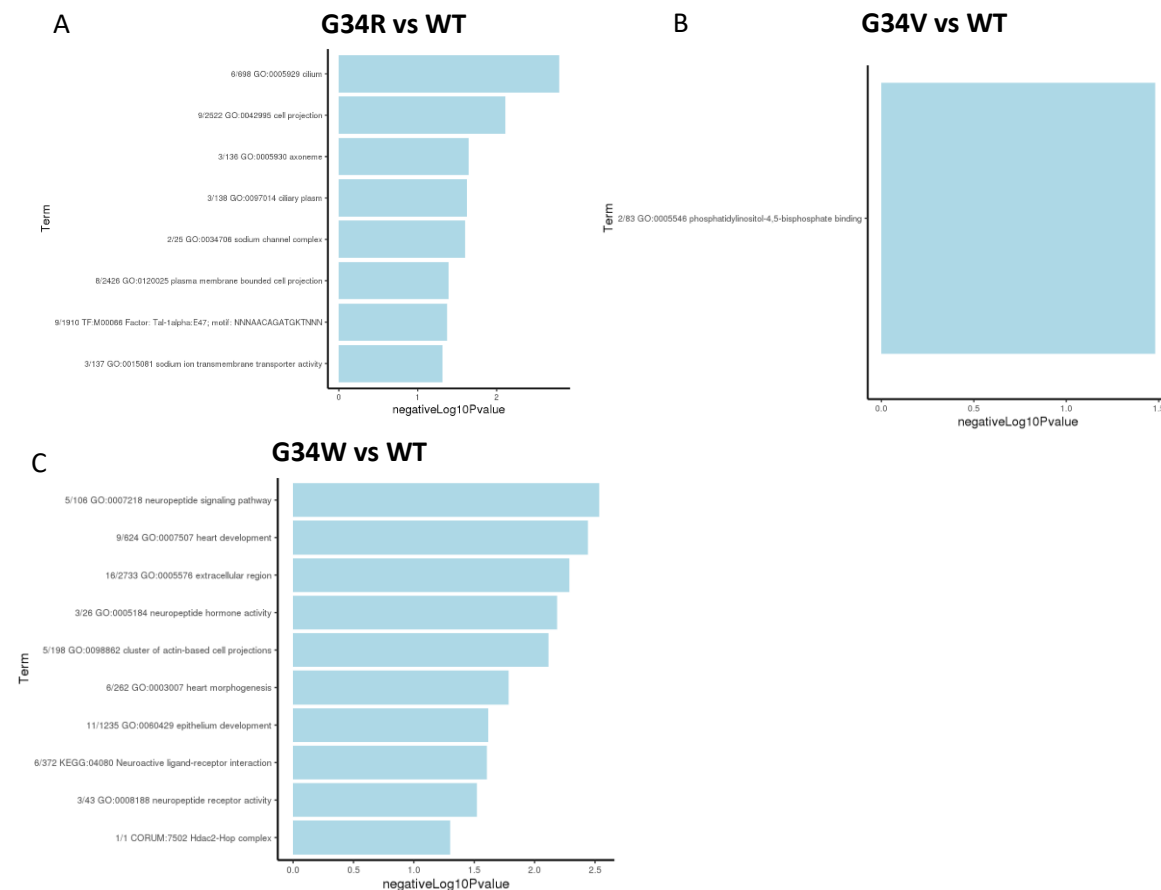


Figure 21: G34R/V/W mutations display changes related to signaling in the cortex at one week post birth

(A) Pathway analysis using differentially upregulated genes at a threshold of mean normalized expression > 100 counts, log2FoldChange > 0.58, and adjusted p-value < 0.05 in the cortex comparing G34R to wildtype; **(B)** G34V to wild type; **(C)** G34W to wildtype

3.3.2. Juvenile G34R have transcriptionally normal astrocytes

Although G34 mutant mice are phenotypically indistinguishable and do not display major differences in gene expression from wild type counterparts one week post-birth, the onset of neurological symptoms (hindlimb clasping) and an increase in astrocytes in the brain of G34R mice is observed by three weeks post-birth. Astrocytes are remarkably plastic and under disease conditions undergo morphological and molecular changes to become ‘reactive astrocytes’¹⁶⁴. To investigate whether astrocytes in the G34R brain were transcriptionally distinct from wild-type astrocytes and potentially reactive in response to a neurological disease-like condition, we generated single cell RNA sequencing data at postnatal day 27 (P27) for one G34R and one wild type (WT) male mouse in the cortex and cerebellum (N = 4). Tissues were dissected using sagittal cross-sections. Thus, while referred as cortical, cortex samples comprised of several anatomical regions; primarily the cerebral cortex tissue, and to a lesser extent hippocampal, thalamic, and hypothalamic tissue. Similarly, cerebellar samples comprised of primarily cerebellar tissue, but included to a lesser extent pontine and medullar tissue.

Post alignment and quality control, 6434, 3973, 7726 and 4857 cells were obtained for the G34R cortex, G34R cerebellum, WT cortex, and WT cerebellum, respectively. To identify cell types, cells were assigned nearest-normal cell identity through two complementary approaches using references cell types derived from single cell RNA sequencing. The cortical reference used was a developmental murine forebrain atlas described in section 3.2. Although the timepoints are not matched, we reasoned that cell identities in the reference should be similar to cell identities three weeks post birth. The cerebellar reference used was a developmental murine cerebellar atlas where cell types were characterized using a similar method to the forebrain atlas³³. Briefly, cell types were identified on a cluster-level through expression of established lineage marker genes and genes specifically expressed in each cluster were extracted to generate cell identity gene signatures. To assign cell identities, the first approach used a neural network (ACTINN) to automatically assigned cells to cell identities based on mean expression profiles of reference cell types¹⁶¹. The second approach used a variation of gene set enrichment analysis (single sample gene set enrichment analysis, ssGSEA) to assign an enrichment score per cell type based on a ranking of cell identity gene signatures in the expression profiles of cells¹⁶².

The highest enrichment score was extracted and assigned as the cell type. Cell type assignments between methods were consistent.

Predominant cell populations in the G34R and WT cortex and cerebellum were astrocytes, immune cells, and non-neuroectodermal cells. Very few neuronal cells were identified from either genotype. Cell type proportions were similar in the G34R vs WT, indicating that the loss of neuronal cells may be due to tissue sectioning or library preparation. Expression of canonical markers was used to validate cell projections. To evaluate transcriptomic differences between cell populations by genotype, G34R and WT samples were joined by tissue (cortex and cerebellum, respectively) without using any integration method or further correction of the data (Fig 22A-D). Cells were visualized in UMAP space. In both the cortex and cerebellum, cells did not separate by histone status, but clustered by cell type. This indicates that any effects of the G34R mutation at the level of cell populations was secondary to cell-identity signals.

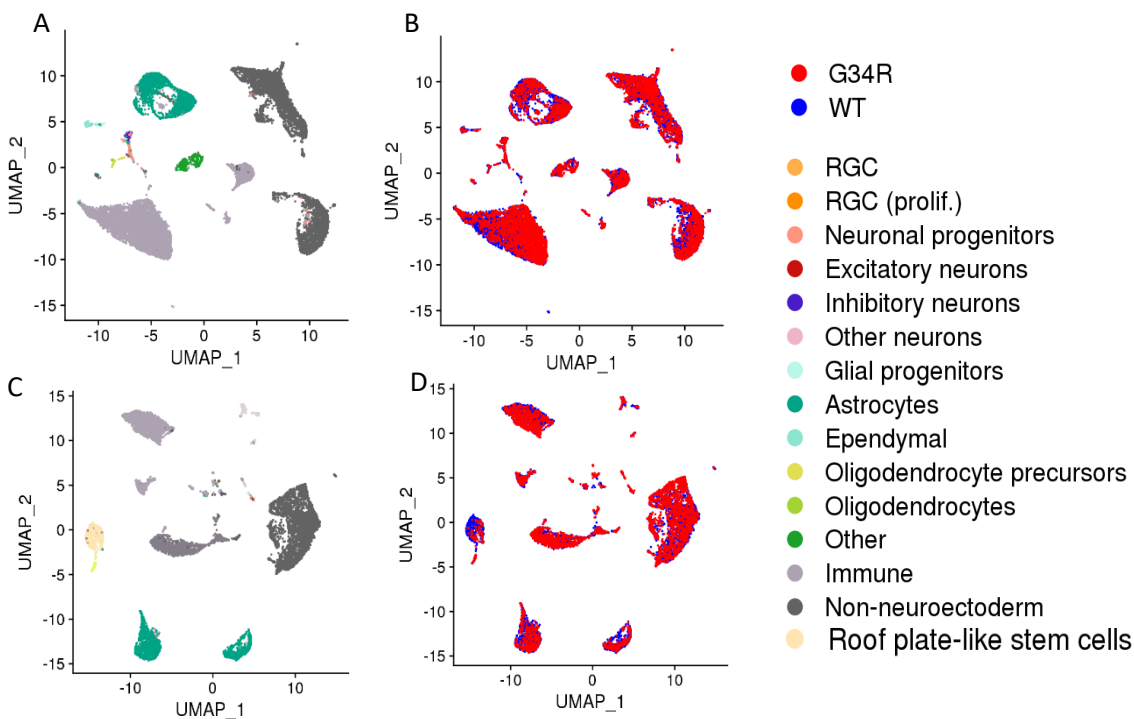


Figure 22: G34R and wild type cortex and cerebellum samples have similar composition of cell types

(A) Cell types in the G34R and wild type cortex; **(B)** cerebellum.

To assess for potential transcriptomic differences among astrocytic populations in the G34R or wild type brain, cells labeled as astrocytes through both projection methods were extracted and re-clustered without using any further transformation of the data. In the cortex, clusters of astrocytes were comprised of both G34R and wild type cells (Fig 23A). This indicates that G34R and wild type astrocytes did not have large differences in gene expression caused by the G34 mutation. Similarly, UMAP projections showed overlap between populations of G34R and wild type astrocytes indicating similarity in expression profiles (Fig 23B).

While the G34R mutation did not dramatically change astrocyte cell identity, the presence of discrete populations of astrocytes in UMAP space and defined clusters pointed to transcriptomic differences in populations of astrocytes. To further characterize differences in astrocyte populations, expression of marker genes from a single cell atlas which resolved spatially distinct astrocyte populations was plotted (Fig 23C)¹⁶⁵. Telencephalon astrocyte markers *Mfge8* and *Lhx2*, characterizing astrocyte populations from the olfactory bulb, cerebral cortex, striatum, amygdala, and hippocampus were specifically expressed in one population observed in the UMAP space. The second population expressed non-telencephalic astrocytes markers *Agt* and *Slca9*, found in astrocyte populations from the hypothalamus and thalamus. Therefore the two broad populations of astrocytes observed in UMAP space corresponded to spatially distinct populations. To additionally validate these results, we found the same marker genes in cluster markers, indicating that the difference in expression between populations is significant. We then asked if there were differences in G34R and wild type astrocytes within spatially distinct populations of astrocytes. However, we found very few differentially expressed genes between G34R and wild type astrocytes in both non-telencephalic and telencephalic populations. Therefore, we concluded that the G34R mutation did not affect the transcriptomic profile of cortical astrocytes at three weeks of age.

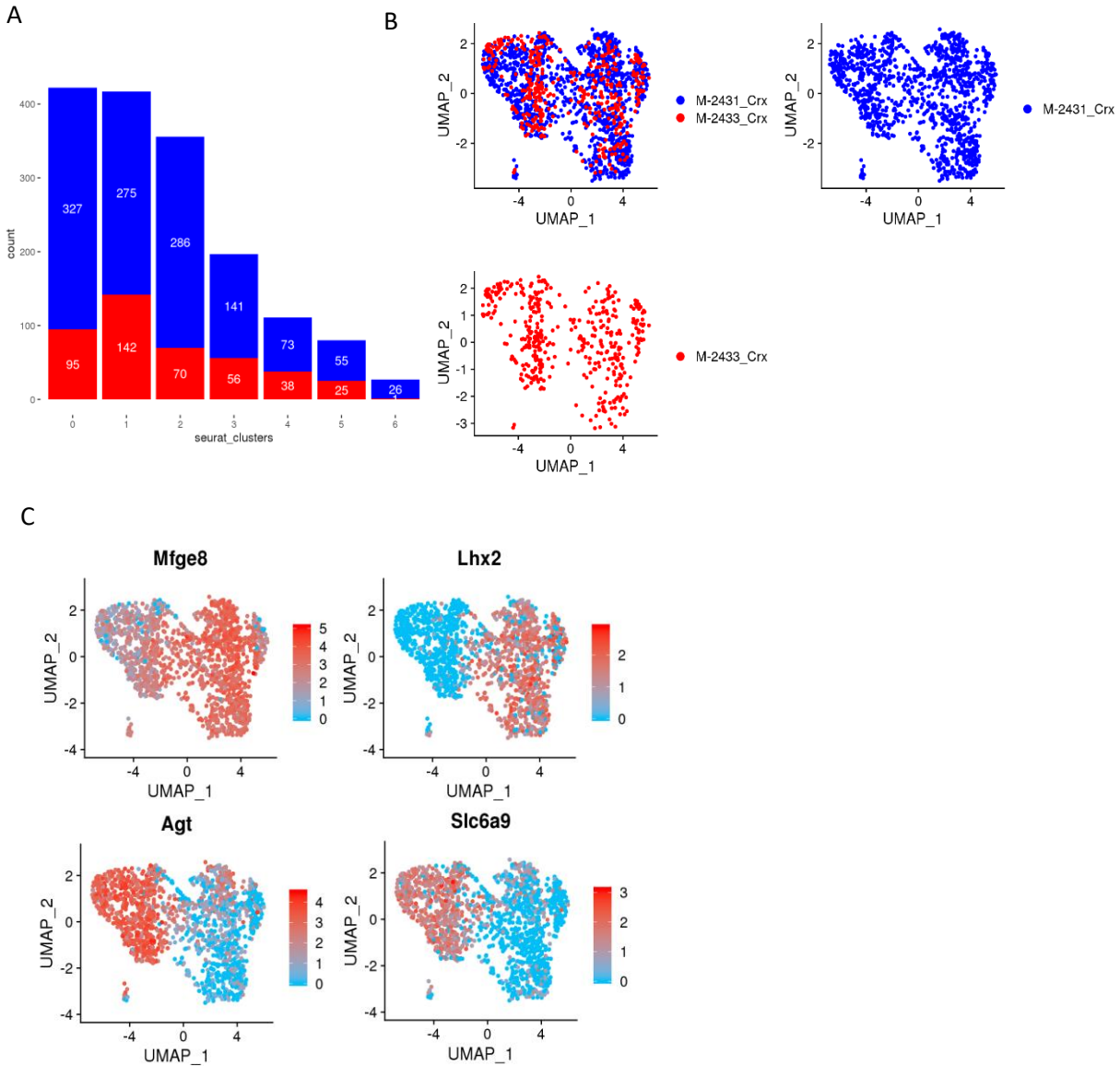


Figure 23: G34R and wildtype astrocytes are transcriptionally similar and arise from spatially distinct populations

(A) G34R (red) and wild type (blue) cells per cluster. Number of cells indicated on chart. **(B)** UMAP visualization of G34R and wild type astrocytes in joint space. **(C)** Gene expression of telencephalic astrocyte markers *Mfge8* and *Lhx2*, non-telencephalic astrocyte markers *Agt* and *Slc6a9*.

In the cerebellum, reprocessing G34R and wild type astrocytes revealed two distinct populations in the UMAP space. Of these, one population was distinct in UMAP space

segregating by genotype, indicating that genotype contributed to differences in gene expression (Fig 24A). This population corresponded to two clusters; cluster 1 was composed of mainly wild type cells (361 WT; 27 G34R), while cluster 2 was composed of mainly G34R cells (227 G34R; 9 WT). Thus, both UMAP and clustering analysis indicated transcriptomic differences in a subset of G34R and WT astrocytes (Fig 24B)

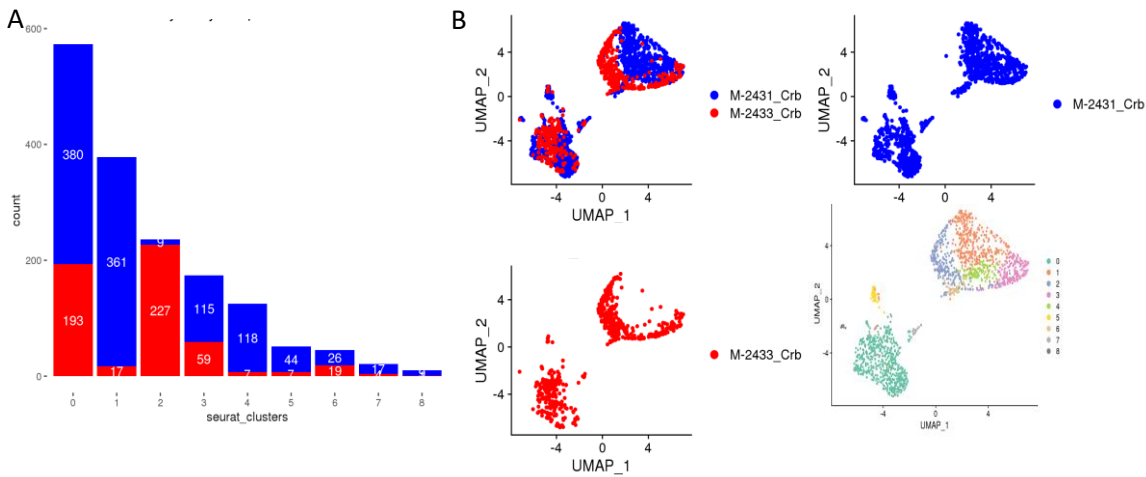


Figure 24: G34R and wild type astrocytes form distinct clusters

(A) G34R (red) and wild type (blue) cells per cluster. Number of cells indicated on chart. **(B)** UMAP visualization of G34R and wild type astrocytes in joint space.

To assess differences in gene expression between distinct G34R and WT astrocytes, differential gene expression analysis was run comparing wild type astrocytes in cluster 1 and G34R astrocytes in cluster 2. Of the few upregulated genes found in the G34R astrocytes, most were ribosomal and mitochondrial genes. While mitochondrial dysfunction is associated with reactive astrocytes in response to disease or injury, other pathways such as increased immune signaling also characterize reactive astrocytes^{166,167}. Therefore, it is unlikely that the transcriptionally distinct subset of astrocytes in the G34R cerebellum are truly reactive astrocytes. In summary, while G34R mutants gain astrocytes in the brain three weeks post birth, the astrocytes are transcriptionally similar to wild type astrocytes and not in a clear reactive state.

3.3.3. Adult G34 mutants have disease-associated immune response in the brain

By adult hood (10 weeks) G34R/V mutants display neurological symptoms including abnormal motor functions, with progressive microcephaly, hindlimb clasping, and paresis (Fig 18). To investigate changes in gene expression related to neurological symptoms, we generated bulk RNA seq data at 10 weeks for G34R, G34V, G34W and WT mice in the cortex and cerebellum, with 2 – 3 replicates each (N = 30). We used PCA to assess the effect of genotype on global gene expression profiles. In the cortex and cerebellum, G34 mutants were separate from wild type counterparts along the first or second principal components (Fig 25A-B). This indicates that G34 mutant status is responsible for major sources of variation in gene expression profiles by adult hood. To identify significant changes in gene expression in the G34 mutant brain, differential expression analysis was performed comparing each mutant to wild type counterparts (Fig 25C). Fitting with the phenotypic penetrance, we found larger numbers of significantly up- and down-regulated genes in G34R/V mutants (476, 297) than G34W mutants (144 genes) in the cortex. This dichotomy was not observed in the cerebellar samples (209 genes G34R, 408 genes G34V, and 151 genes G34W). Therefore G34R/V mutants may preferentially dysregulate processes in the cortex, as opposed to the cerebellum. Expression of complement cascade and other immune related genes were upregulated in the G34R cortex and cerebellum, and to a lesser extent in G34V/W (Fig 25C). Among top differentially expressed genes in the G34R cortex were complement cascade genes (C3, C4a, C4b, C1qb, C1qc), immune related genes (Clec7a, Ccl6, Cd52, Cd84, Mpeg1 Itgax), canonical microglial markers (Cd68, Ly86), and a canonical astrocytic marker (Gfap). To determine whether the increase in expression of microglial and astrocyte gene markers was due to an accumulation of cells or overexpression, we performed immunohistochemistry staining for an microglial marker (Iba1) and astrocytes (Gfap) (Fig 25D). We found an abnormal accumulation of microglia and astrocytes in the G34R cortex, which was not significant for G34V/W. Therefore the phenotype progressively appeared by age and was specific to G34R.

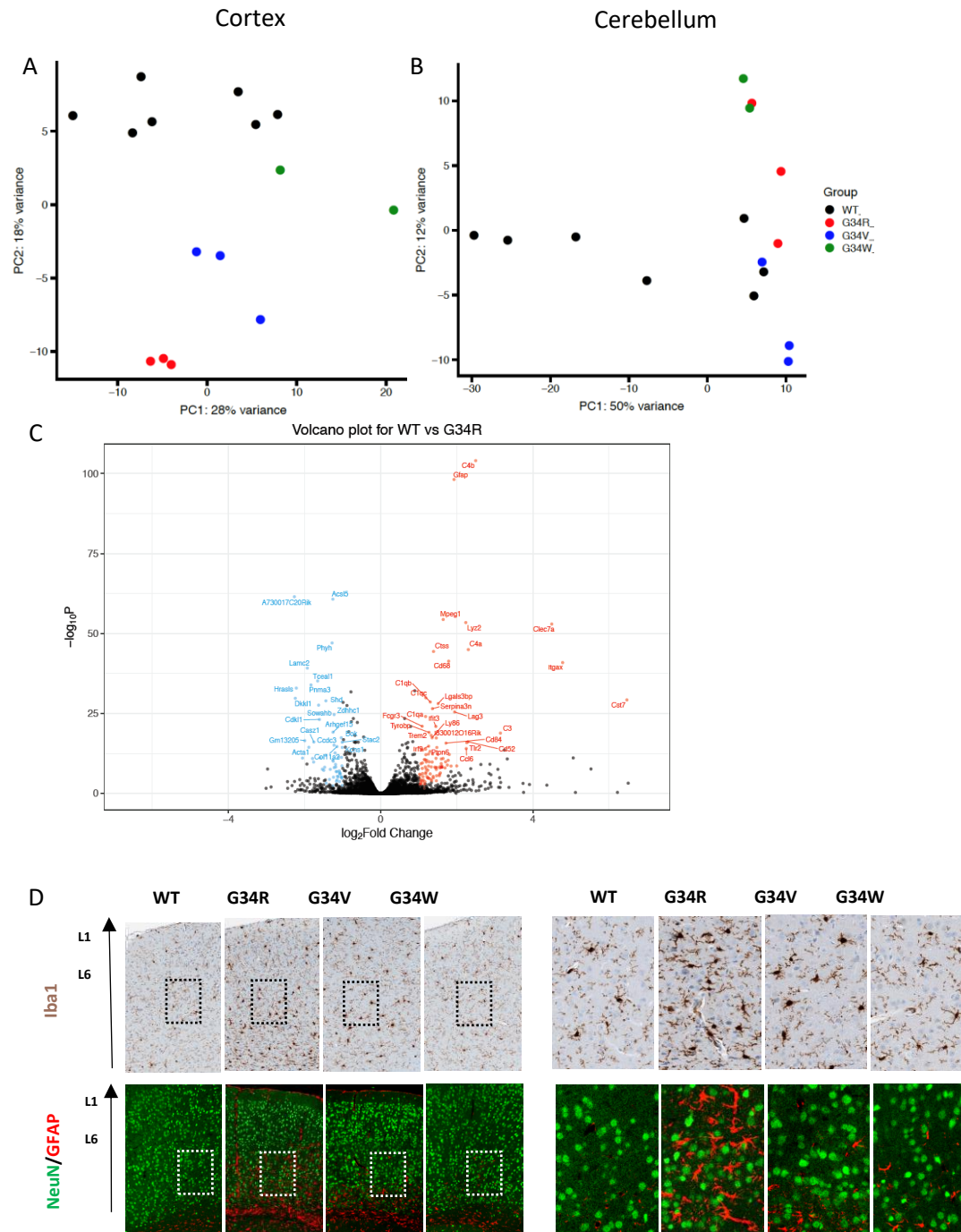


Figure 25: G34 mutant status drives major changes in immune and astrocyte gene expression in the cortex and cerebellum at 10 weeks

(A) PCA using top 10000 most variable genes in the G34 mutant and wildtype cortex; **(B)** cerebellum. **(C)** Volcano plot highlighting top 50 differentially expressed genes using thresholds mean normalized read count > 100, absolute log₂FoldChange > 1 and adjusted p-value < 0.05.

(D) Representative immunohistochemistry of cortical layers of 10 weeks old G34 mutant mice.
Figure generated by Sima Khazaei, Carol Chen, and Pariya Azarafshar.

To identify functional pathways associated with G34 mutants in an unbiased manner, pathway analysis on differentially expressed genes was run. In the G34R cortex and cerebellum, terms associated with immune responses (immune response, innate immune response, immune effector response) were found (Fig 26A, B). In the G34V cortex, terms related to immune response was found. In the G34W cortex, very few terms were found. In the G34W cerebellum, the majority of terms were related to transcription factor binding motifs. Taken together, this indicates that while all G34 mutants show dysregulated immune related genes through DEG analysis, pathway analysis corroborates dysregulated immune processes for the G34R.

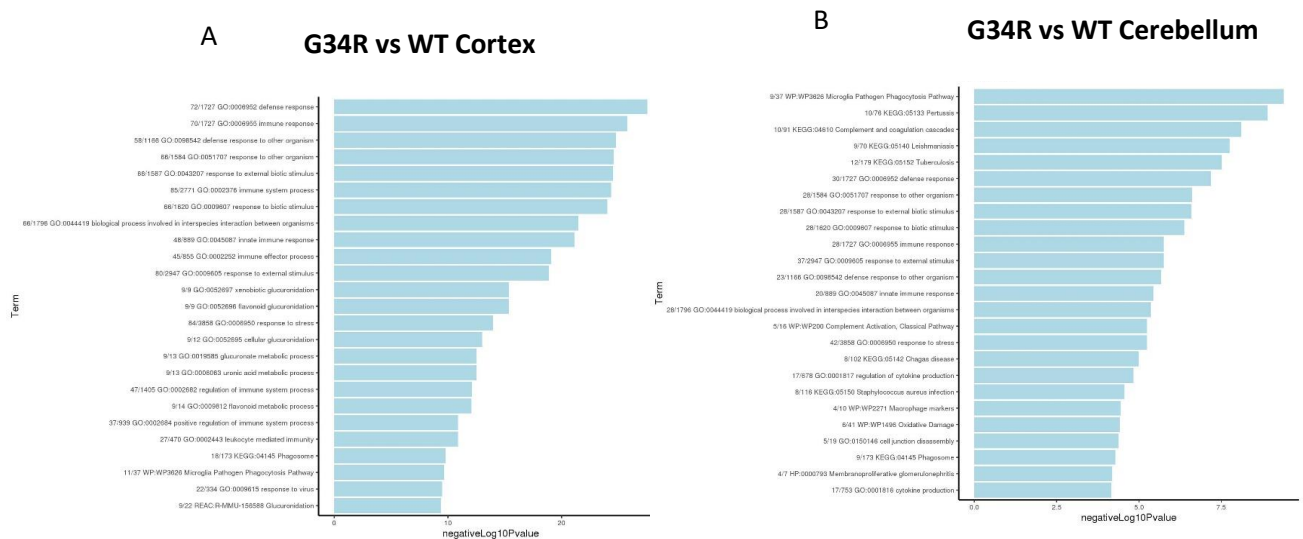


Figure 26: Upregulation of immune signaling in G34 mutant cortex and cerebellum

(A) Pathway analysis comparing G34R vs wild type samples in the cortex; **(B)** cerebellum. The first two numbers in the label indicates the number of differentially expressed genes found in the gene list over the size of the gene list.

To evaluate if the observed upregulation of immune genes was associated with a disease state, GSEA was run using a disease-associated microglia and homeostatic microglia signature obtained from an Alzheimer mouse model associated with neurodegenerative disease¹⁵⁵ (Fig 27A,B). In the G34R/V mutant cortex, significant enrichment was found for disease associated,

but not homeostatic, microglial signatures. Fitting with the non-neuronal phenotype observed, G34W mutants were not specifically enriched for disease-associated microglia. This indicates that while all G34 mutants are progressively expressing increased microglial genes in the cortex, the G34R mutant is specifically associated with an accumulation of disease-associated microglia in the cortex.

To evaluate if the observed increase of astrocytes in the G34R was associated with a disease state, GSEA was run using reactive astrocyte signatures using signatures profiled from single cell RNA sequencing data of a autoimmune encephalomyelitis mouse model¹⁶⁷. In the G34R, significant enrichment for the reactive astrocyte signature was observed in the cortex and cerebellum, indicating the potential presence of reactive astrocytes. However, since the reactive astrocyte signature significantly overlapped with immune signatures and bulk RNA seq provides an average expression profile (Fig 27C), the presence of reactive astrocytes cannot be confirmed by this analysis.

Therefore G34R mutants display progressively worsening neurological phenotypes which is characterized by abnormal accumulation of disease-associated microglia and astrocytes. Fitting with the phenotypic penetrance, G34W mutants share the increased immune response at the molecular level but are not specific for disease associated microglial infiltration.

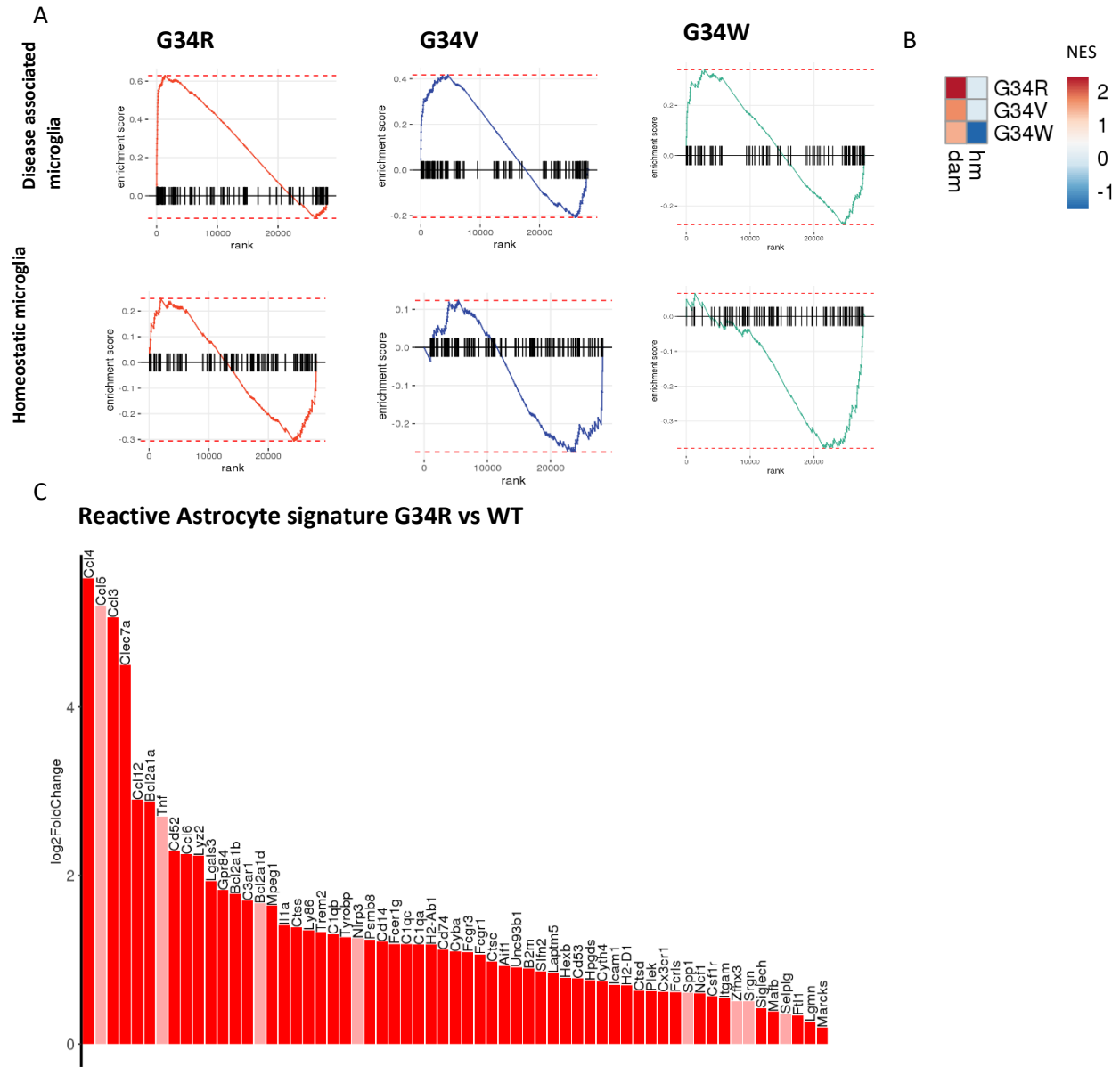


Figure 27: G34R display disease associated microglial signature in the cortex at 10 weeks

(A) Enrichment plots using disease associated and homeostatic microglial signatures. **(B)** Normalized enrichment scores for disease associated (DAM) and homeostatic (HM) microglia. Scores not significant (adjusted p-value < 0.05) are set to 0. NES: normalized enrichment score. **(C)** Log fold change of G34R compared to wild type for reactive astrocyte gene signature, genes meeting significance (adjusted p-value < 0.05) are highlighted in dark red.

3.3.4. Adult G34R have cortical neuronal loss

Due to the increase of astrocytes and microglia in the G34R cortex, we were curious if there were changes to other cell types in the G34 mutant brain. To assess for potential up- and down-regulation of cell types, we used gene set enrichment analysis selecting postnatal cortical cell-identity gene signatures from a developmental murine cortical reference³¹(Fig 28A). While not matched with age, this reference was chosen since signatures were derived from many distinct cell populations characterized through single cell RNA seq data. Thus, an enrichment in gene set is proxy for enrichment in cell type. Consistently, all G34 mutants had enrichment across microglial and macrophage signatures. We additionally found a depletion across oligodendrocyte and OPC signatures for all G34 mutants. Specific to G34R mutants, we found enrichment for astrocytic and radial glial cell signatures and depletion for all neuronal signatures. To further assess neuronal depletion in specific populations, GSEA was run on using cell type signatures from adult mouse cortical and hippocampus single cell RNA sequencing reference from the Allen Brain Atlas¹⁵⁴ (Fig 28B). We found depletion of multiple neuronal populations in the G34R cortex, with layer V / VI glutamatergic (excitatory) neurons showing the strongest depletion. We validated our results through immunohistochemistry staining for Oct6, a marker of cortical layer V neurons¹⁶⁸ (Fig 28C)

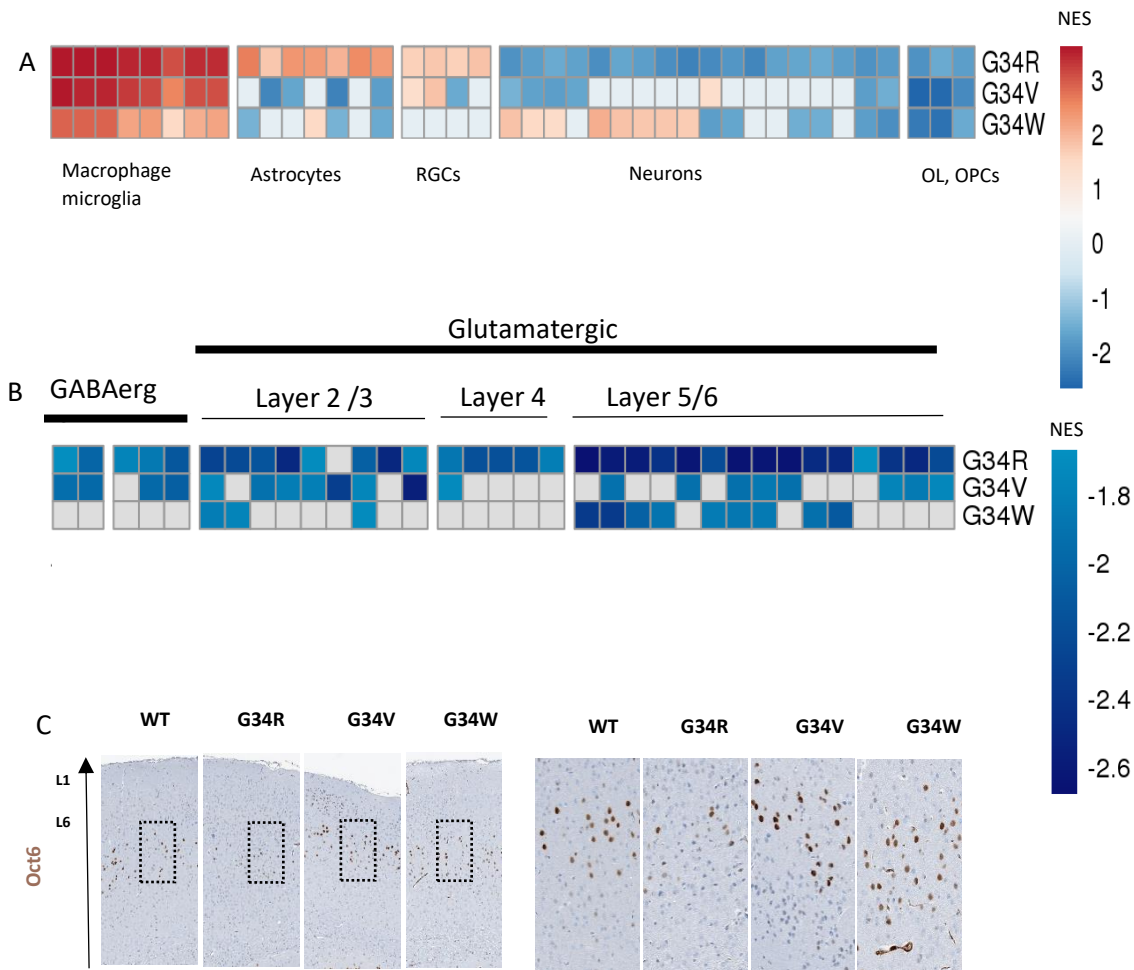


Figure 28: G34R mutants display neuronal loss

(A) Heatmap of normalized enrichment scores of postnatal cortical cell identity signatures from murine forebrain atlas; **(B)** normalized enrichment scores of cortical neuronal cell identity signatures from adult murine atlas. Scores which do not meet significance (adjusted p-value < 0.05) are set to 0. NES: normalized enrichment score; OL: oligodendrocyte; OPC: oligodendrocyte precursor cell; RGC: radial glial cell. **(C)** Representative immunohistochemistry staining of Oct6 in 10 week G34 mutant cortex. *Figure generated by Sima Khazaei, Carol Chen, and Pariya Azarafshar.*

Thus in the cortex, G34R mutants specifically display progressive abnormal accumulation of astrocytes and disease-associated microglia which is coupled to neuronal loss most pronounced in deep cortical layers. Loss of layer V-VI neurons is documented in neurodegenerative diseases

such as Huntington’s disease and Alzheimer’s, however, the mechanism linking G34R mutations to progressive neuronal loss specifically in layers V/VI is unclear through gene expression profiling alone^{169,170}.

We next assessed if G34R mutants displayed similar profiles in the cerebellum through GSEA analysis of cell-identity gene signatures from a murine cerebellar reference previously described in section 5.2. Similar to the cortex, we found an increase of astrocyte and microglial gene expression across all mutants, and depletion of oligodendrocyte expression specific to G34R/V mutants (Fig 29A). However, we did not find a complete depletion of neuronal signatures as in the cortex (Fig 29B)

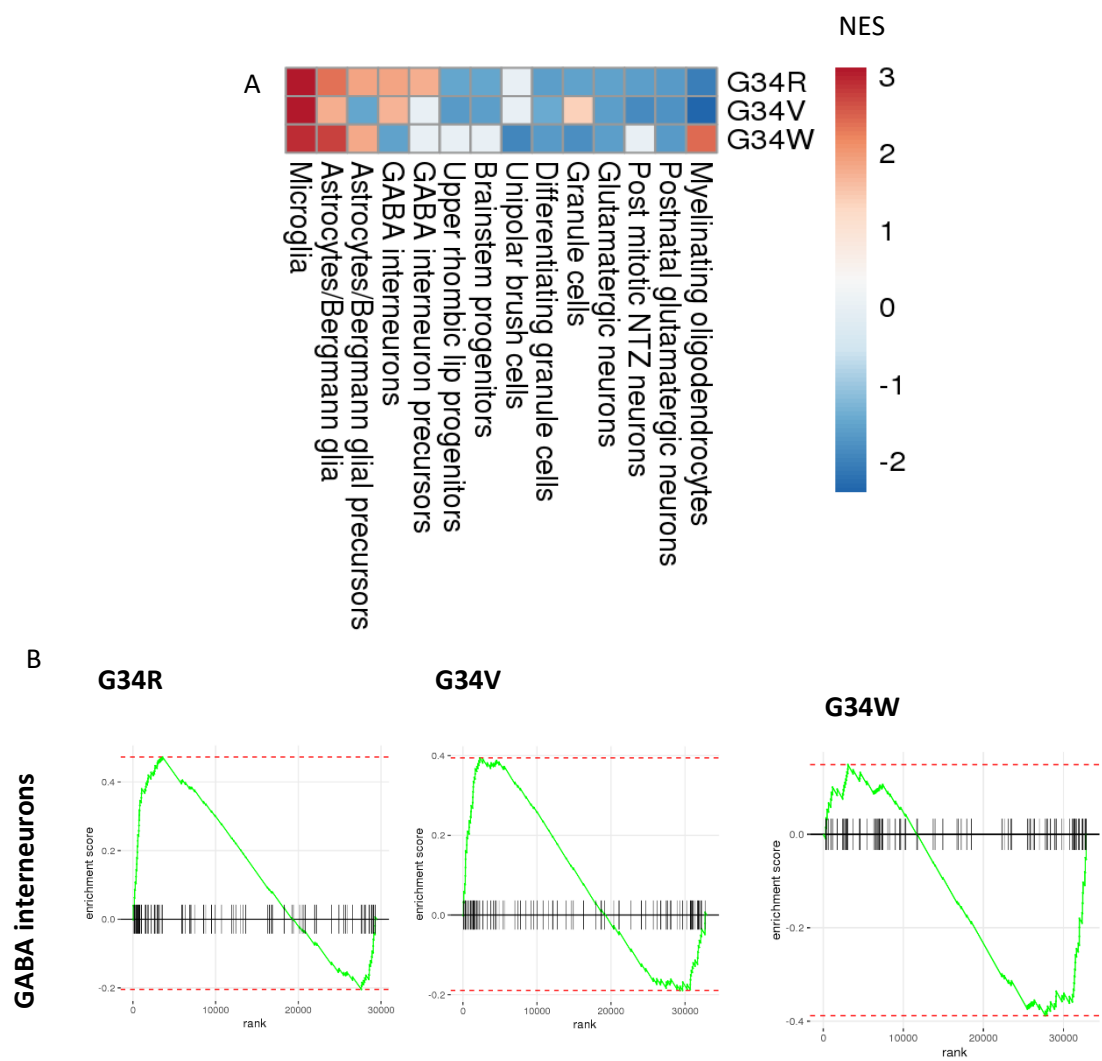


Figure 29: G34 mutants do not display pan-neuronal loss in cerebellum at 10 weeks

(A) Heatmap of normalized enrichment scores for cell type identity signatures in the G34 mutant cerebellum at 10 weeks. Scores not meeting significance (adjusted p-value < 0.05) are set to 0. NES: normalized enrichment score. **(B)** Enrichment plots for GABA interneurons in G34R and G34V cerebellum at 10 weeks

In sum, we observe a progressive neurodegenerative phenotype in G34R/V mutants which is most pronounced in G34R cortex. At the molecular level, G34R mutants specifically display accumulation of disease-associated microglia coupled to pan-neuronal loss most pronounced in layer V/VI neurons in the cortex. In the cerebellum, G34 mutants display increased microglial and astrocyte gene expression, but no pan-neuronal loss is observed.

3.3.5. Adult G34R mice share transcriptional overlap with Rett syndrome patients

To explore a mechanistic cause for the neurodegenerative phenotype observed in the G34R cortex, we performed BioID experiments which assess for protein-protein interactions in U2OS cells injected with the G34 mutant construct. We found decreased protein-protein contacts with several proteins bearing PWWP domains across all G34 mutants. Specifically to the G34R mutant, we found decreased interactions of DNA methylase Dnmt3a. In the cell, Dnmt3a is a writer complex which deposits methylation on cytosines at non-CpG islands (mCH) in post-natal neurons^{80,171}. MeCP2 is a reader complex for mCH methylation, mutations in which drive Rett syndrome; a neurodevelopmental disorder characterized by a period of normal development followed by developmental regression, impairment of motor functions, and impairment of cognitive abilities^{140,172,173}. Recent murine models demonstrated that loss of Dnmt3a or Mecp2 in inhibitory neurons converge to a similar neurological phenotype, and that MeCP2 is partially dependent on mCH methylation to perform its normal function^{174,175}. At the phenotypic level, G34R mutants share some overlap with Rett syndrome patients; seemingly normal phenotype post birth but progressively develop symptoms including impairment of motor functioning. Since we observed loss of Dnmt3a interactions in G34R, we were interested if G34R mice shared similarities in gene expression with Rett Syndrome patients at a molecular level.

Thus far, three studies of RNA sequencing data of Rett patient autopsy samples have been published^{176–178}. Where available, raw data was reprocessed to ensure consistency across data

sets (Lin and Aldinger) or supplemental data was used as reported (Gogliotti) comparing autopsy samples of Rett Syndrome patients in the cortex and cerebellum to donors.

To assess for enrichment of G34 mutant signatures in the Lin data set, we performed gene set enrichment analysis using top upregulated genes in the G34 mutant as compared to WT (Fig 30A). We found a depletion of G34R and G34V derived signatures in Rett patients as compared to donors, and specifically found complement genes were causing this downregulation. The depletion of complement proteins is consistent with previous findings for this data set. From this analysis alone, H3.3 G34 germline mutant mice do not share similarities in gene expression with Rett syndrome patients when compared to controls. However, the controls used in this data set, while age and sex-matched, were from patients with cause of death due to multiple injuries including head injuries. Thus, it may be that absolute levels of complement and immune processes are highly expressed in controls due to trauma.

To assess for enrichment of G34 mutant signatures in the Aldinger data set, the same procedure was used as for the Lin data set. In contrast to the previous depletion, however, we found an enrichment of uniquely the G34R cortex signature in the cortex of Rett syndrome patients (Fig 30B). The genes driving this signal were complement (C1QA, C1QB, C1QC, C4A, C4B), immune (LYZ, CD68, MYO1F) and astrocytic genes (GFAP). This indicates that G34R, but not G34V or G34W, mice share similarities at the transcriptomic level with Rett Syndrome patients in this data set.

For the Gogliotti data set, raw data was not available and so we used top differentially expressed genes from the cortex and cerebellum of Rett syndrome patients to assess for enrichment in the G34 mutant brain. In the G34 mutant cortex, genes upregulated in Rett syndrome patients were enriched in the G34R, but not G34V/W (Fig 30C). These results are consistent with the Aldinger data set where we find enrichment of upregulated genes in Rett syndrome patients specifically in the G34R, but not G34V/W cortex.

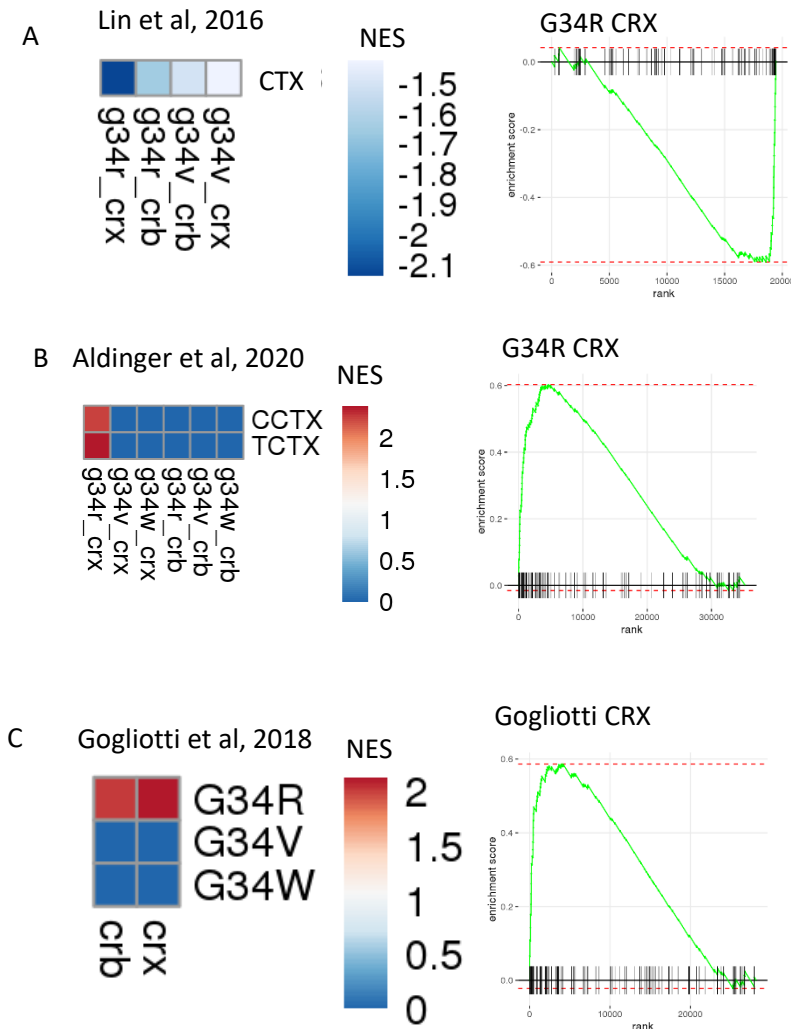


Figure 30: Enrichment of G34R signatures in cortex of Rett syndrome patients

(A) Heatmap of normalized enrichment scores for Lin dataset (Lin et al, 2016) using gene set enrichment analysis with top 100 differentially expressed genes in G34R compared to wild type cortex thresholding by average mean expression > 100 reads, log2FoldChange > 1, and adjusted p-value < 0.05. Scores which do not meet significance are set to 0. **(B)** Heatmap of the same for Aldinger dataset (Aldinger et al, 2020). **(C)** Heatmap of enrichment scores of signatures from Gogliotti et al in the G34R 10 week cortex. NES: normalized enrichment score; CTX: cortex; CCTX: cingulate cortex; TCTX temporal cortex; crb: cerebellum samples from Rett syndrome patients of Gogliotti data set; crx: cortex samples from Rett syndrome patients of Gogliotti dataset.

In summary, we find overlap in two out of three published data sets. To some extent, H3.3 G34R mutations cause symptoms and molecular changes partly consistent with a known neurodevelopmental disorder, Rett syndrome. The changes induced by germline H3.3 G34R mutations are not only critical for normal neurodevelopment, but also may bear some relevance to human disease. The lack of concordance in one data set may be due to the potentially non-homeostatic state of the controls. While the G34R point mutations partly phenocopy and have overlap in gene expression similarities with Rett syndrome patients, it is perhaps unsurprising there is not a complete overlap since one effect of many caused by G34R is altered Dnmt3a recruitment.

3.3.6. G34W have an unhealthy obesity phenotype at 10W

While G34R/V mutant mice display progressively worsening neurological symptoms, all G34W predominantly display an obesity phenotype. Of note, by this age all G34 mutants are obese, but the onset of obesity for G34W occurs at an earlier age. Since G34W mutations affect mesenchymal lineages, we were interested in gene expression changes relating to obesity in mesenchymal tissue. To this end, we generated RNA sequencing data of epidermal white adipose tissue (WAT) at 8-10 weeks for G34 mutant mice with INDEL and wild type mice as controls. We used PCA to find main sources of gene expression in the data were due to genotype; consistent with phenotypic observations the G34W were the most distinct in global gene expression profiles from other samples (Fig 31A). G34W mutants additionally displayed the largest amount of differentially expressed genes when compared against controls (3352 genes), as opposed to G34R/V mutants (1651, 828 genes). Upregulated genes in all G34 mutants in white adipose tissue corresponded to immune signalling (inflammatory response, cytokine production, defense response), while G34V additionally upregulated extracellular matrix degradation processes (collagen degradation, reduced muscle collagen VI) (Fig 31B-D). The combination of extracellular matrix degradation and immune infiltration is indicative of a disease state in aortic tissue of mice fed a high fat diet¹⁷⁹.

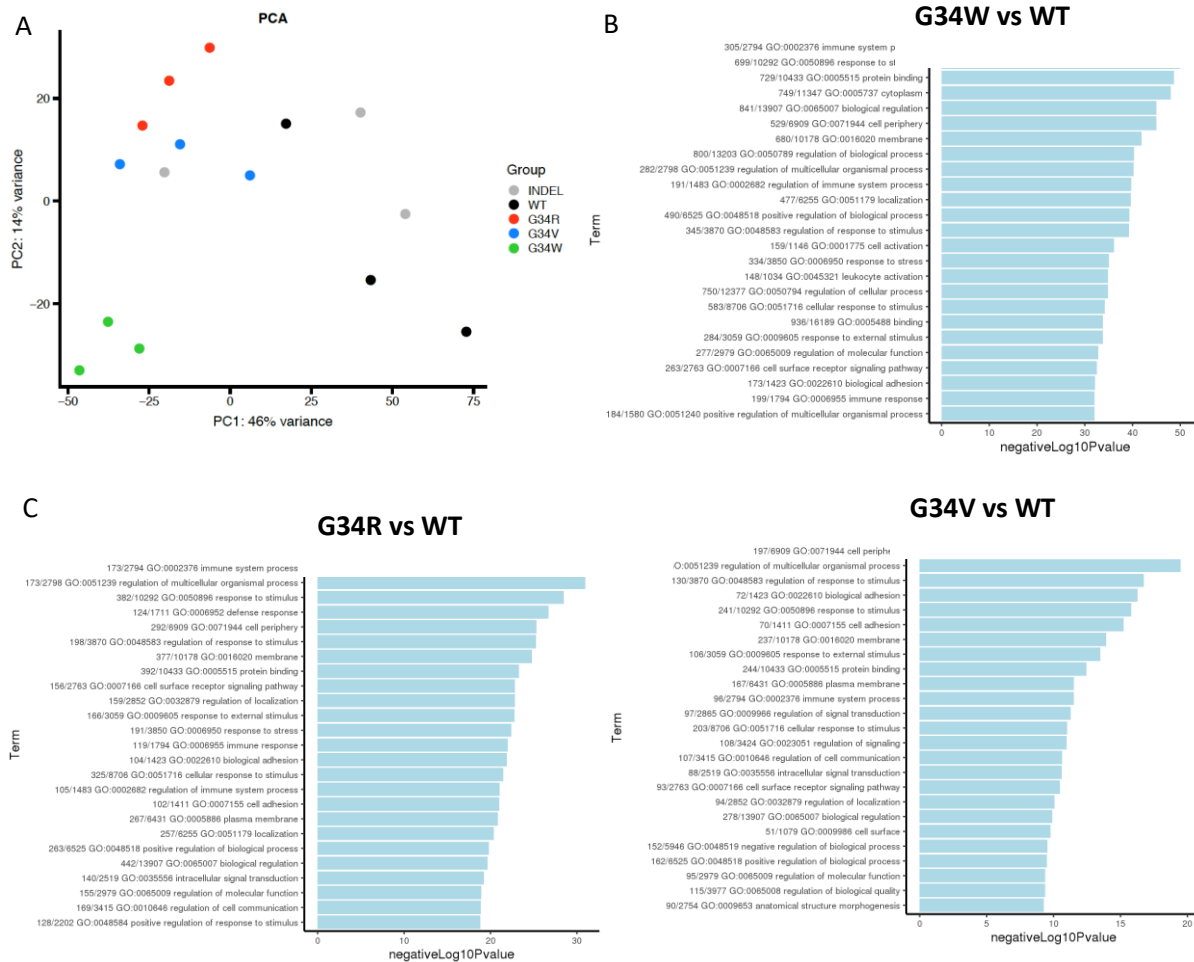


Figure 31: G34 mutants display increased immune signaling in white adipose tissue by 10 weeks

(A) PCA using top 1000 most variable genes in adipose tissue at 10 weeks for G34 mutants and wild type counterpoints. **(B)** pathway analysis for G34W versus wild type; **(C)** G34R versus wild type; **(D)** G34V versus wild type. The first two numbers in the label indicates the number of differentially expressed genes found in the gene list over the size of the gene list.

White adipose tissue is a highly metabolically adaptive tissue in response to normal changes in the body and in disease states such as obesity¹⁸⁰. To further asses changes in gene expression profiles of G34 mutants relating to metabolic or other biological pathways, we used gene set enrichment analysis using a manually curated data set of ‘hallmark pathways’ from Molecular Signature Database. Upregulated pathways common to all G34 mutants were Hallmark epithelial to mesenchymal transition, inflammatory response, interferon gamma response,

complement and myogenesis, while downregulated pathways included Hallmark oxidative phosphorylation, fatty acid metabolism, and adipogenesis (Fig 32A).

The molecular changes identified are consistent with an unhealthy fat phenotype in all G34 mutants. As seen by the strength of association in the heatmaps, G34W mutants most prominently display the unhealthy fat phenotype. To identify specific changes in G34W, we assessed the significance of genes driving the enrichment scores. Unique to G34W, we found downregulation of PPARG, a master regulator of adipocyte differentiation^{181,182} (Fig 32B). This indicates that in G34W, the decreased adipocytes may be due to impaired development of adipocyte progenitors.

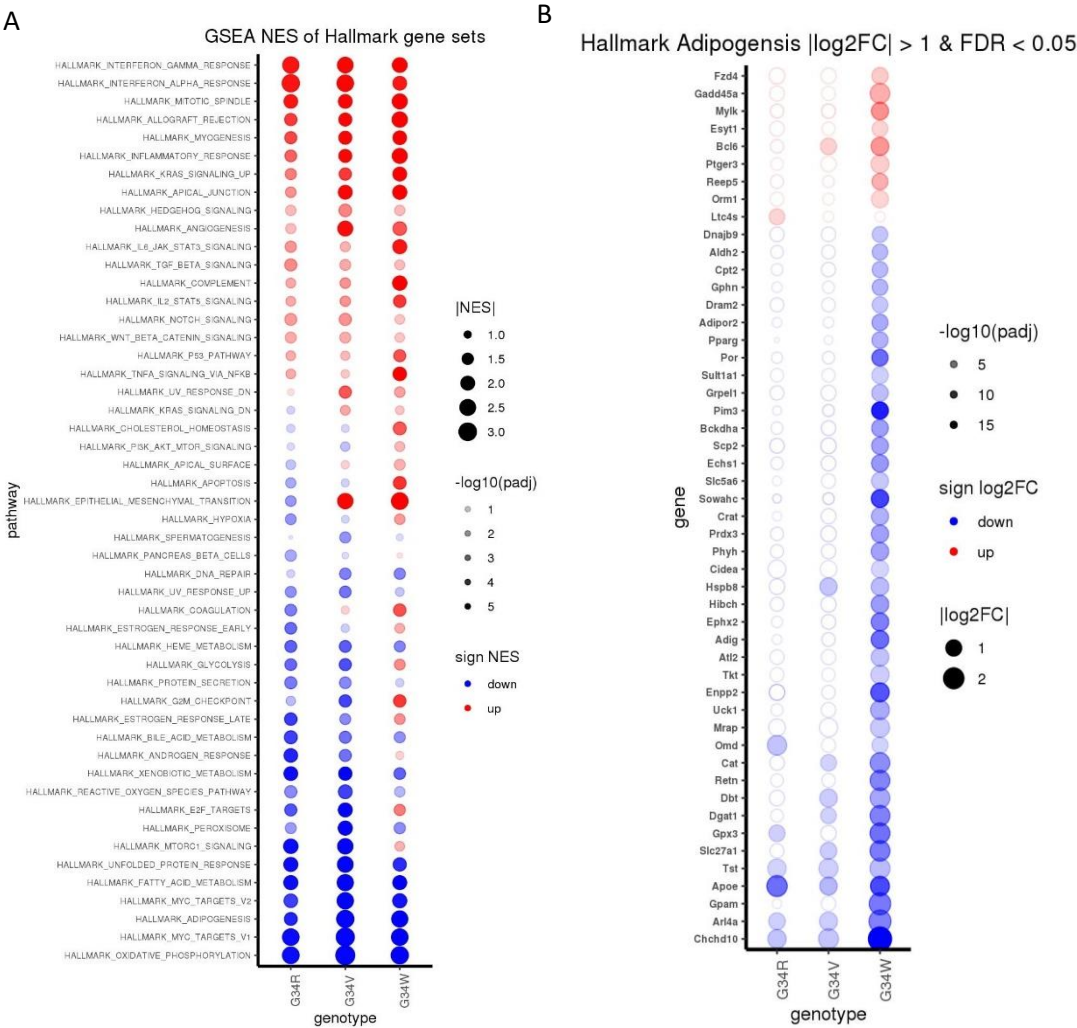


Figure 32: G34 mutants display transcriptomic changes consistent with an unhealthy fat phenotype

(A) Bubble plot of normalized enrichment scores from hallmark pathways. Size of bubble corresponds to absolute value of normalized enrichment score, color of bubble corresponds to enrichment (red) or depletion (blue), and transparency corresponds to the negative log₁₀ value of the adjusted p-value. **(B)** Bubble plot of leading edge genes from Hallmark Adipogenesis pathway in G34W. Size of bubble corresponds to absolute value of log₂FoldChange comparing G34 mutant to wild type. NES: normalized enrichment score.

In summary, G34 mutant mice have progressive transcriptomic dysregulation in a tissue-specific manner. Mice are transcriptionally normal at one week post birth before onset of symptoms, with an increase of astrocytes in the brain of G34 mutants also appearing transcriptionally similar to WT by the juvenile stage. However, by adulthood G34 mutants display tissue-specific alterations in gene expression. G34R mutants display a unique neuronal loss coupled to increased disease-associated immune activation in the brain, while G34W mutants display an unhealthy obesity phenotype marked by downregulation of master transcription factor PPAR γ regulating fat differentiation.

Chapter 4: Discussion

Histone H3 proteins have critical roles in regulating normal development, cellular differentiation, and cell identity. Dysregulation of chromatin marks on histone H3 proteins through mutations in histones proteins or through mutations in downstream effector proteins result in a variety of cancers and neurodevelopmental disorders. In the cancer setting, histone mutations are necessary to drive tumorigenesis but require obligate co-occurring partner mutations to form pediatric high-grade gliomas. While histone mutations are not therapeutically targetable due to their essential roles in the cell, co-partner mutations in proteins such as kinases are therapeutically actionable. Here, I presented my work on the analysis of transcriptomic data to better understand the impact these mutations have on the cellular context in pediatric high grade gliomas and development using a variety of model systems. I first described the impact of a partner mutation in a kinase, *ACVR1*, on cell identity is different depending on the driving histone H3 K27M mutation (H3.1 vs H3.3). Next, I reported results on the effect of K27M in H3.3 pHGGs on cell identity and find that an effect of removing H3.3 K27M is to promote differentiation towards astrocyte and oligodendrocyte cell-identity using *in vivo* patient-derived orthotopic xenografts, but signals pertaining to cell identity become dominated by mesenchymal gene programs induced by differentiation media using *in vitro* differentiation experiments. Finally, I analyzed the effects of germline heterozygous H3.3 G34R/V/W mutations in a murine model and identify progressive neuronal loss in G34R mutants and unhealthy obesity most marked in G34W mutants. Together, these results shed light onto cell identity programs impacted by histone and co-partner mutations in a cell-specific context.

4.1. The role of *ACVR1* on cell state is dependent on histone context

Although H3.3 K27M mutations cause a subset of pHGGs, the K27M mutation alone is insufficient to induce high grade-glioma like lesions^{23–25}. It is only in a restricted developmental window and cellular context that K27M mutations, in conjunction with partner mutations, induce neoplastic transformation. Common partner mutations include gain-of-function mutations in kinases *PDGFRA*, *EGFR*, *FGFR3*, and *ACVR1*. The pattern of kinase mutations in

pHGGs also segregates with histone mutation and anatomical location in pHGGs, suggesting specific functions of these mutations in promoting oncogenesis³⁵.

Mutations in *ACVR1*, encoding the kinase ALK2, are commonly found in H3K27M mutated tumors along the midline, occurring in both canonical H3 (H3.1) and variant H3 (H3.3) K27M mutated pHGGs^{8,13,16,97}. The gain of a mutation in *ACVR1* is an early clonal event for the tumor, underscoring its importance to tumor progression³⁵. Mutations in both the glycine-serine domain or kinase domain of *ACVR1* structurally shift the kinase conformation from an inactive to a constitutively active state⁹⁸. However, the function of these mutations in the two histone contexts remains poorly understood. Here, we use isogenic system through CRISPR-Cas9 mediated knock out to assess the transcriptomic effect of *ACVR1* mutations in the context of H3 K27M mutations. We use cell lines bearing *ACVR1*-R206H mutations in the context of H3.3 K27M tumors, and *ACVR1*-G328E/V mutations in the context of H3.1 tumors since this pattern of co-occurrence is preferentially found in patient tumors. We find the effect of *ACVR1* mutations on the transcriptome and cell state is dependent on canonical (H3.1) vs non-canonical (H3.3) histone H3 K27M status. *ACVR1* mutations co-segregating with H3.1 K27M do not greatly alter the transcriptomic landscape but rather promote proliferation and affect mesenchymal pathways such as extracellular matrix genes, consistent with previous reports¹¹⁷. *ACVR1* mutations co-segregating with H3.3 K27M alter cell state by switching cells from an OPC-like state towards the astrocyte lineage. Of note, the function of H3.3 K27M is also thought to block cells from further differentiating in the astrocytic lineage, thus it is possible *ACVR1* and K27M mutations work cooperatively to block differentiation³¹. Interestingly, germline *ACVR1* mutations alone are not sufficient to induce tumorigenesis, but induce a congenital disorder (fibrodysplasia ossificans progressiva) in which patients present with ectopic bone formation, but no predisposition for malignancies^{114,118}. Therefore our findings support the idea that *ACVR1* mutations cooperate with H3K27M mutations in a specific cellular context to promote oncogenesis.

Here we find that *ACVR1* mutations in pHGGs promote oncogenesis through distinct functions dependent on specific histone mutant (H3.1 versus H3.3 K27M) profiles. Since H3.1 and H3.3 K27M tumors are spatially and temporally distinct, we think the distinct effects of *ACVR1* are

intimately tied to different molecular profiles of the different cell-of-origins (Jessa et al, 2021, in revision). *ACVR1* is a member of the BMP family, which regulates cell fate over the course of neurodevelopment^{99,183}. BMP signaling forms a spatial- and temporal gradient of morphogens in conjunction with SSH signaling during neural development to specify distinct cell identities of progenitor cells in a dose-dependent manner. In H3.3 K27M tumors, increased BMP signaling due to constitutive activating mutations in *ACVR1* may serve to sensitize progenitors and alter cell fate decisions in the cell-of-origin. However, in H3.1 K27M tumors, *ACVR1* mutations may serve to promote mesenchymal lineages in progenitors potentially less sensitive to dosage of BMP morphogens.

While our work provides insight onto distinct cellular functions mediated by *ACVR1* dependent on the specific histone mutation in pHGGs, there are several limitations of this study. Firstly, we used few replicates per histone context and strengthened our conclusions using immunofluorescence staining *in vitro*. Using *in vitro* cell lines does not accurately recapitulate environmental factors in the tumor micro-environment, such as signalling molecules. While *ACVR1* mutations are ligand-independent gain-of-function mutations, it is possible that cross-talk among signaling pathways influence the cellular impact of *ACVR1* mutations in tumors. Furthermore, BMP signaling and *ACVR1* mutations have dual roles in either promoting oncogenesis or acting as tumor suppressors dependent on the cancer type, specific cellular context, and extracellular environment¹⁸⁴. This suggests the presence of several distinct mechanisms in promoting oncogenic adaptations. Using additional model systems which incorporate elements of the tumor micro-environment, such as patient-derived orthotopic xenograft models, will help assess whether the *in vitro* functions of *ACVR1* are recapitulated *in vivo*, or if there are potentially distinct responses *in vivo*.

In sum, *ACVR1* mutations have distinct cellular functions in promoting oncogenesis in DIPG dependent on the histone (H3.1 or H3.3 K27M) context. We report that *ACVR1* mutations in H3.3 K27M DIPGs stall cells from acquiring astrocyte-like properties, while *ACVR1* mutations in H3.1 K27M DIPGs may be implicated in upregulation of mesenchymal gene programs. This is of interest since distinct downstream cellular responses induced by *ACVR1* can be targeted through different therapeutic approaches.

4.2. Effect H3.3 K27M mutations in *in vitro* and *in vivo* model systems

H3.3 K27M mutations drive a subset of pHGGs occurring along the midline which arise in a restricted neurodevelopmental window in cells resembling OPCs^{34,163,185}. The effect of the mutation on the cell of origin is thought to block differentiation towards an astrocytic lineage³¹. However, since these tumors arise over a broad anatomical region and bear distinct epigenetic profiles, it is possible there are multiple distinct cells of origin which block differentiation towards potentially distinct terminal cell states^{31,116}. Here, we leverage both *in vitro* serum-differentiation experiments and *in vivo* orthotopic xenograft models to assess the direct effect of H3.3 K27M on cell identity. We find that the dominant effect of H3.3 K27M on gene expression is greatly impacted by choice of experimental model system. We report that removal of H3.3 K27M *in vitro* and subjecting cells to differentiation media cause an upregulation of mesenchymal-related gene programs in serum-differentiation experiments, which may partly mask astrocyte gene programs. In contrast, we report that removal of H3.3 K27M induces a predominant upregulation of both oligodendrocyte and astrocyte gene programs in orthotopic xenograft models.

Our work supports the idea that H3.3 K27M mutations alter the epigenome reversibly, and that removal of the mutation induces epigenomic changes permissive to changes in cell identities¹⁸⁵. Indeed, at the transcriptomic level we find tumor cells progress in differentiation towards mature glial cells upon removal of H3.3 K27M in orthotopic xenograft models. However, the exact glial cell identity is heterogeneous and a distinct terminal cell identity does not cleanly correspond to one original tumor location. We find that H3.3 K27M-KO cells express either primarily oligodendrocyte or a mixture of both oligodendrocyte and astrocyte cell identities in different xenograft models. These findings are consistent with the idea that H3.3 K27M directly causes a differentiation blockade and that the effects are transcriptionally reversible³¹. The changes observed in gene expression resulting in progression towards both astrocytes and oligodendrocytes in different tumors indicates that there may be potentially distinct cells of origin which are stalled different locations along a differentiation trajectory.

Our results demonstrate the importance leveraging multiple model systems to reach robust conclusions. *In vitro* models are an attractive choice due to the ability to manipulate a single

variable in isolation to study its effects. However, we find that culture conditions induce overwhelming upregulation of mesenchymal-related gene programs, and as such the choice of media must be carefully optimized. In serum-differentiation experiments, the overwhelming majority of upregulated gene programs are related to a mesenchymal-related signal caused by differentiation media, which is not fully recaptured by orthotopic xenograft models. This is suggestive that the mesenchymal-related gene programs may help cells acquire capabilities to survive in cell culture. We find these mesenchymal-related gene programs are co-expressed with astrocyte gene programs, suggestive of a potential masking effect of differentiation media on cell identity. Together, our results support a model where differentiation media induces a baseline level of differentiation towards an astrocyte state. However, due to the potential artifactual signal obtained in serum-differentiation experiments, we cannot conclusively identify whether removal of the H3.3 K27M mutation has a synergistic effect in promoting differentiation towards the astrocyte lineage. Moving to *in vivo* model systems to decouple the effect of differentiation media from H3.3 K27M mutations provides key insights onto the direct effect of H3.3 K27M on differentiation potential. Our results suggest a direct effect of H3.3 K27M in stalling cells from mature astrocyte and oligodendrocyte cell states, which is consistent with a progression observed for a minority of cells in serum-differentiation experiments. The expression of astrocyte and oligodendrocyte cell signatures warrants further investigation of whether cells co-express these identities, which may potentially suggest increased epigenetic plasticity of cells upon removal of the mutation. We additionally find upregulation of non-neuroectodermal cell types driven by mesenchymal gene programs in one xenograft model, which is consistent with serum-differentiation experiments. This raises the possibility that an effect of removing the mutation may induce mesenchymal properties, but it is unclear whether this is a tumor-specific effect.

Altogether, we find that *in vitro* differentiation results in a baseline level of differentiation towards an astrocyte-like state but is primarily masked by mesenchymal gene programs which may help cells adapt to culture conditions. We demonstrate a direct effect of H3.3 K27M is to stall cells from differentiating towards the astrocyte lineage is recaptured *in vivo*. We additionally find a potential progression towards the oligodendrocyte lineage. Both of these

findings support the idea that the effect of H3.3 K27M stalling cells into a progenitor-like state is reversible. Therefore, targeting processes which promote differentiation may induce tumor cells to change cellular state into ultimately less aggressive states.

4.3. Lineage specificity of germline *H3f3a* G34R/V/W mutations

Somatic histone mutations occurring in a restricted developmental window and cellular context are thought to give rise to cancer. However, little is known about the functions of histone mutations in development, and further understanding could yield insight onto the cellular context-specificities which are observed in cancer. Intriguingly, somatic mutations in *H3F3A* encoding H3.3 at glycine 34 to arginine or valine (H3.3 G34R/V) drive approximately 30% in pediatric high grade gliomas, while somatic mutations at the same residue mutated to tryptophan (G34W) drive over 90% of adult giant cell of bone tumors^{9,17}. G34 mutations are thought to promote oncogenesis through local alternations to K36 and K27 methylation landscapes in the mutant H3.3 nucleosomes by impeding interactions with specific chromatin modifiers³⁰. However, whether different amino acid substitutions have a specific effect in different tissues is unclear.

Here we report that germline *H3f3a* G34R/V/W mutations in a murine model preferentially affect different tissues. Although arginine, valine, and tryptophan have different biochemical properties, we find that G34R/V mutations affect neural lineages while G34W mutations affect mesenchymal lineages with surprising specificity at the level of gene expression. All germline G34 mutant mice are phenotypically indistinguishable from wild type counterparts and have similar transcriptional profiles in brain and bladder tissue at postnatal day 7 prior to onset of symptoms. By the juvenile stage at one month, at the cellular level G34R mice display an accumulation of astrocytes throughout the brain in comparison to wild type counterparts. However, the gene expression profiles of astrocytes from the G34R cortex and cerebellum are not robustly distinct from astrocytes of wild type mice. Therefore while we observe an astrocytic increase, astrocytes are not in a transcriptionally 'reactive' state in response to disease at one month of age. Progressively by adulthood at ten weeks, G34 mutants have distinct phenotypic profiles; G34R/V mutants display predominantly neurological symptoms

(paresis, altered gait) and G34W most markedly have an obesity phenotype. We find that at the cellular level, G34R mutants uniquely display abnormal accumulation of microglia in the brain, and that gene expression profiles match a 'disease-associated' microglial gene signature. Furthermore, G34R mutants strikingly display pan-neuronal loss in the cortex which is most pronounced in cortical layers V and VI. In white adipose tissue, all G34 mutants display transcriptomic changes related to unhealthy obesity marked by decreased adipogenesis, fatty acid metabolism, and increased immune signaling. Unique to G34W, we find downregulation of a master transcription factor of adipocyte differentiation, PPARG.

To assess for potential molecular underpinnings of our neuronal phenotype, we next turned our attention to epigenetic changes occurring in the G34 mutant brain. We found G34R uniquely causes depletion of H3K36me2/3 in H3.3 G34R nucleosomes which alters recruitment of several PWWP-domain bearing proteins that recognize H3K36me2, including Dnmt3a^{186,187} (Khazaei et al, submitted). Dnmt3a is a DNA methylase that acts to methylate cytosine at non-CpG islands (mCH) in postnatal neurons^{80,85}. In the postnatal brain, high levels of mCH are found in layer V/VI neurons, therefore this may explain the susceptibility to loss of these neurons found by RNA sequencing⁸³. The sole reader for non-CpG methylation is MeCP2, mutations in which cause Rett syndrome^{140,172}. Mutations in Dnmt3a or MeCP2 in inhibitory neurons result in phenotypic and molecular changes similar those observed in Rett Syndrome¹⁷⁵. Given the dysregulation of Dnmt3a and depletion of neurons in G34R mutants, we asked whether G34 mutants shared molecular overlap with Rett syndrome patients. We found a match between specifically our G34R, but not G34V/W, signatures with autopsy samples of Rett syndrome patients in two out of three published cohorts^{177,178,188}. Our G34R mutant signature contains primarily immune related gene programs such as complement genes and astrocyte-identity related genes. Therefore, assessment for certain immune processes may be relevant in other neurodevelopmental syndromes. The partial molecular overlap found may be due to other alterations in the G34R mutant brain which include, but are not limited to, dysregulation of Dnmt3a and potential downstream dysregulation of MeCP2. Assessing potential dysregulation of MeCP2 in the G34R mutant cortex will provide insights towards shared mechanisms between

Rett Syndrome patients and G34R mutants. Notably, mutations in Dnmt3 are also associated with neurodevelopmental syndromes including Tatton-Brown Syndrome, and Heyn-Sproul-Jackson Syndrome^{136,189}. Whether G34R mutants share similarities with other neurodevelopmental syndromes is of interest since we observe an abnormal immune infiltration in the brain which has therapeutic implications. Dissecting the roles of other affected PWWP domain bearing proteins which function to modify chromatin in the G34R cortex may also provide insight towards changes in the G34R transcriptome and potential overlap with neurodevelopmental syndromes.

Recently, 37 de novo germline mutations in *H3f3a* and *H3f3b* were reported in 46 patients presenting with a previously unidentified neurodegenerative disorder, but no predisposition for malignancies¹⁴³. In the reported cohort, there were no patients with *H3f3a* G34R/V mutations, but one patient with a *H3f3b* G34V mutation. The effects of other *H3f3a* and *H3f3b* mutations are consistent with the worsening neurological phenotype observed in the G34R/V mutant cortex. Assessment of molecular overlap between the G34R murine model and patients with mutations in *H3f3a* or *H3f3b* may provide insight towards more general mechanisms causing neuronal loss. However, a complete loss of either *H3f3a* or *H3f3b* have different developmental defects in murine models, and only somatic mutations in *H3f3a*, but not *H3f3b*, have been reported in cancer^{9,17,190}. Therefore there are potentially distinct cellular functions of *H3f3a* and *H3f3b* which may contribute uniquely to neurodevelopmental syndromes.

Germline G34 mutants display a phenotypic spectrum. While H3.3 G34W mutants display the most marked obesity phenotype, all G34 mutants are obese and display associated inflammatory responses characteristic of unhealthy obesity¹⁹¹. At the transcriptome level, we find downregulation of PPAR γ , a master transcription factor of adipocyte differentiation uniquely downregulated in white adipose tissue in G34W, further investigation at an earlier time point may yield insight onto further G34W-specific changes. Whether the obesity stems from a neurodevelopmental cause, as opposed to a peripheral cause, remains an open question. Assessing similarities of the H3.3 G34W transcriptome in white adipose tissue with

recently published single-cell atlases of mice fed a high-fat diet will determine if the transcriptome of G34W mutants reflect a transcriptome consistent with peripheral causes of obesity¹⁹². Obesity can also stem from epigenetic causes¹⁹³. H3K36M mutations, which alter H3K36 methylation, promote decreased adipogenesis and PPARG gene expression¹⁹⁴. H3.3 G34W impairs SETD2-, but not NSD2-mediated deposition of H3K36me3 *in vitro* which in turn promotes increased H3K27me2/3 and abnormal PRC2 activity³⁰. Whether this mechanism is recapitulated in germline H3.3 G34W white adipose tissue or whether PRC2 targets cause decreased expression of adipogenesis programs will provide insight toward potential epigenetic dysregulation. Of relevance to human disease, Prader-Willi syndrome is a neurodevelopmental syndrome characterized by morbid obesity due to alterations in genomic imprinting¹⁹⁵. Defects in genomic imprinting are also known to cause Beckwith–Wiedemann syndrome, an overgrowth syndrome with predisposition for malignancies¹⁹⁶. Assessing whether G34W mutants display similar defects in DNA methylation may additionally provide clues onto mechanisms causing obesity.

In sum, we report that germline G34 mutants preferentially affect neuroectodermal and mesenchymal lineages, providing insight onto progressive defect caused by histone mutations. We find that G34R mutants display pan-neuronal loss coupled to a gain in disease-associated microglial activity, while G34W mutants display a striking obesity phenotype characterized by downregulation of PPARG. Targeting abnormal immune processes may be of therapeutic interest in associated neurodevelopmental disorders.

Chapter 5: Conclusion and Future Directions

5.1. Conclusion

The epigenome serves as a blueprint for the cell to regulate gene expression through post-translational chemical modifications on histone proteins. Mutations in histones or downstream effector proteins result in epigenetic dysregulation driving a subset of cancers and neurodevelopmental syndromes. In pediatric high-grade gliomas, histone mutations rarely occur in isolation but require co-occurring mutations to promote oncogenesis. However, the changes to the cellular context induced by these mutations in cancer and neurodevelopment are not well characterized which present an impediment to the development of effective therapeutics. This thesis sheds light onto how histone and co-partner mutations affect cellular identity in cancer and development through analyzing transcriptomic data from model systems.

We first describe how the effect on cell identity of a partner kinase mutation, *ACVR1*, varies depending on the exact histone H3 K27M context. In canonical H3.1 K27M pHGGs *ACVR1* functions in promoting proliferation; while in canonical H3.3 K27M pHGGs *ACVR1* functions to block cellular maturity toward the astrocytic lineage. This implicates distinct targets for therapeutics since while histone mutations are not directly targetable due to their fundamental role in the cell, partner mutations in kinases are amenable drug targets. As we find *ACVR1* mutations directly impact cellular identity in H3.3 K27M mutations, we next focus on the direct effects of H3.3 K27M using different model systems. We find our choice of model system greatly impacts the transcriptome and that a dominating effect of differentiation media is the upregulation of mesenchymal gene programs. *In vivo* patient-derived orthotopic xenograft models reinforce the role of H3.3 K27M mutations blocking differentiation towards astrocytes and potentially oligodendrocytes while *in vitro* differentiation experiments confer a more mesenchymal phenotype which masks glial cell identity signals.

Outside of cancer, we explore the effects on cell identity of germline heterozygous H3.3 G34R/V/W mutations using a murine model. We find germline H3.3 G34 mutants display phenotypes with surprising tissue specificity in affected lineages similar to somatic H3.3 G34 mutations in cancers. G34R mutants preferentially affect neuroectodermal lineages and display

a progressive neurodegenerative phenotype marked by disease associated immune signaling; while H3.3 G34W mutants preferentially affect mesenchymal lineages and markedly display an unhealthy obesity phenotype characterized by downregulation of a master transcription factor of adipocyte differentiation, PPARG. Altogether, our findings shed light onto functions of histone and co-partner mutations in tumorigenesis and neurodevelopment in specific cellular contexts providing a step towards the development of therapeutics.

5.2. Future Directions

The specific molecular and cellular context in which histone mutation arise cause a subset of cancers and neurodevelopmental syndromes. Histone mutations drive a subset of pediatric high-grade gliomas which display remarkable specificity in the combination of histone and partner mutation, tumor location, and age of incidence. Here we show that mutations in a kinase, *ACVR1*, act to promote oncogenesis through distinct cellular functions dependent on the specific histone H3 mutation (canonical H3.1 or variant H3.3 K27M). However, mutations in *ACVR1* arise in distinct domains (glycine-serine or kinase) which display preferential segregation with H3.1 or H3.3 K27M mutations in pediatric high-grade gliomas. Although mutations in both glycine-serine and kinase domains shift the kinase to a structurally active state, the preferential segregation of domain mutations in *ACVR1* with histone status may point towards an underlying biological significance. While we studied mutations in the kinase domain of *ACVR1* in H3.1 K27M cell lines and mutations in the glycine-serine domain of *ACVR1* in H3.3 K27M cell lines, analyzing the opposite combination of mutations in *ACVR1* through RNA sequencing will determine whether the specific functions of *ACVR1* in the two histone contexts is robust across mutations in both domains. We find an effect of *ACVR1* is to stall cellular differentiation of progenitor cells towards astrocytes in H3.3 K27M pediatric high-grade gliomas, which is the same effect of H3.3 K27M mutations on differentiation potential. Therefore knocking out H3.3 K27M in *ACVR1* mutant cells may assess the potential synergistic effects of *ACVR1* and H3.3 K27M on cellular differentiation.

In pediatric high-grade gliomas, histone mutations are necessary but not sufficient for full oncogenesis. We find an effect of H3.3 K27M mutations in stalling differentiation of glial progenitor cells towards astrocytes *in vivo* using orthotopic xenograft models. While *in vivo*

models more closely recapture the tumor microenvironment, single cells display transcriptomic heterogeneity in cell identity in both the K27M and K27M-KO conditions. To assess whether these changes arise from altered epigenetic landscapes which do not resemble normal cells, analyzing open or closed chromatin states through chromosome conformation capture (Hi-C) experiments may provide insight towards the degree of reversibility of H3.3 K27M mutations.

Outside of cancer, dysregulation of chromatin marks stemming from mutations in histone proteins or proteins which deposit chromatin marks drive a subset of neurodevelopmental syndromes. We report that germline heterozygous mutations at glycine 34 of *H3f3a*, encoding H3.3, preferentially affect distinct developmental lineages. In the brain, *H3f3a* G34R mutants display a pan-neuronal loss most strikingly in deep cortical layers V and VI. Assessment of chromatin landscapes through chromatin immunoprecipitation assays (ChIP-seq) for marks known to be disrupted by G34 mutations in single cells followed by integration of genomic analysis (scRNA seq, WGBS) may provide a mechanistic basis for the cell-type specific dependencies and re-distribution of chromatin marks specific to each mutation. While *H3f3a* G34R mutants preferentially impact the brain, all G34 mutants display an obesity phenotype with variable penetrance. By adulthood, *H3f3a* G34W mutants display a marked obesity phenotype with complete penetrance. RNA sequencing analysis of white adipose tissue at an earlier age before the onset of obesity may provide insight onto G34W-specific changes. Assessing for enrichment of certain cell populations using published single cell atlases of white adipose tissue may provide a mechanism at the cellular level. Since histone mutations affect the epigenome, investigating potential epigenetic causes at the level of DNA methylation through whole genome bisulfite sequencing or chromatin landscapes through Chip-seq of H3K36, H3K27 methylation marks, H3.3 G34-interacting proteins, or BioID for protein-protein interactions in adipose and related mesenchymal tissue may provide insight onto specific chromatin modifications causing obesity.

While germline H3.3 G34R/V/W mutations in murine models do not form tumors, the parallel between corresponding somatic mutations in cancer is still of interest. In particular, since H3.3 G34R/V mutations arise in interneuron cells of origin in pHGGs, investigating changes to

protein-protein interactions of H3.3 G34 mutant nucleosomes in interneuron progenitors through BioID or DNA methylation through whole genome bisulfite sequencing may provide mechanisms for dependencies specific to cell-of-origin. Assessing immune infiltration in primary tumor samples may also provide therapeutic avenues.

Chapter 6: References

1. Ostrom, Q. T. *et al.* CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015. *Neuro-Oncology* vol. 20 iv1–iv86 Preprint at <https://doi.org/10.1093/neuonc/noy131> (2018).
2. Louis, D. N. *et al.* The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathologica* **114**, 97 (2007).
3. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803–820 (2016).
4. Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro-Oncology* **17**, iv1 (2015).
5. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520-537.e5 (2017).
6. Stupp, R. *et al.* Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine* **352**, 987–996 (2005).
7. Khuong-Quang, D. A. *et al.* K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathologica* **124**, 439–447 (2012).
8. Fontebasso, A. M. *et al.* Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. *Nature Genetics* **46**, 462–466 (2014).
9. Schwartzenuber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
10. Sturm, D. *et al.* Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
11. Buczkowicz, P. *et al.* Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nature Genetics* **46**, 451–456 (2014).
12. Castel, D. *et al.* Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathologica* **130**, 815–827 (2015).
13. Taylor, K. R. *et al.* Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. *Nature Genetics* **46**, 457–461 (2014).
14. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature Genetics* **44**, 251–253 (2012).
15. Faury, D. *et al.* Molecular profiling identifies prognostic subgroups of pediatric glioblastoma and shows increased YB-1 expression in tumors. *Journal of Clinical Oncology* **25**, 1196–1208 (2007).

16. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nature Genetics* **46**, 444–450 (2014).
17. Behjati, S. *et al.* Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nature Genetics* **45**, 1479–1482 (2013).
18. Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat Genet* **49**, 180–185 (2017).
19. Lu, C. *et al.* Cancer: Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science (1979)* **352**, 844–849 (2016).
20. Lehnertz, B. *et al.* H3K27M/I mutations promote context-dependent transformation in acute myeloid leukemia with RUNX1 alterations. *Blood* **130**, 2204–2214 (2017).
21. Boileau, M. *et al.* Mutant H3 histones drive human pre-leukemic hematopoietic stem cell expansion and promote leukemic aggressiveness. *Nature Communications* **2019 10:1** **10**, 1–12 (2019).
22. Sturm, D. *et al.* Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nature Reviews Cancer* **2014 14:2** **14**, 92–107 (2014).
23. Funato, K., Major, T., Lewis, P. W., Allis, C. D. & Tabar, V. Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. *Science (1979)* **346**, 1529–1533 (2014).
24. Pathania, M. *et al.* H3.3K27M cooperates with Trp53 loss and PDGFRA gain in mouse embryonic neural progenitor cells to induce invasive high-grade gliomas. *Cancer Cell* **32**, 684 (2017).
25. Lewis, P. W. *et al.* Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science* **340**, 857 (2013).
26. Khuong-Quang, D. A. *et al.* K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathologica* **124**, 439–447 (2012).
27. Harutyunyan, A. S. *et al.* H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me₂/me₃ and is essential for glioma tumorigenesis. *Nature Communications* **2019 10:1** **10**, 1–13 (2019).
28. Khazaei, S. *et al.* H3.3 g34w promotes growth and impedes differentiation of osteoblast-like mesenchymal progenitors in giant cell tumor of bone. *Cancer Discovery* **10**, 1968–1987 (2020).
29. Krug, B. *et al.* Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell* **35**, 782–797.e8 (2019).
30. Jain, S. U. *et al.* Histone H3.3 G34 mutations promote aberrant PRC2 activity and drive tumor progression. *Proc Natl Acad Sci U S A* **117**, 27354–27364 (2020).
31. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nature Genetics* **51**, 1702–1713 (2019).

32. Chen, C. C. L. *et al.* Histone H3.3G34-Mutant Interneuron Progenitors Co-opt PDGFRA for Gliomagenesis. *Cell* **183**, 1617–1633.e22 (2020).
33. Vladoiu, M. C. *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67–73 (2019).
34. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. <http://science.sciencemag.org/>.
35. Nikbakht, H. *et al.* Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. *Nature Communications* **7**, (2016).
36. Kornberg, R. D. Structure of chromatin. *Annu Rev Biochem* **46**, 931–954 (1977).
37. Waddington, C. The epigenotype. *Endeavour* **1**, 18–20 (1942).
38. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* **319**, 1097–1113 (2002).
39. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
40. Peterson, C. L. & Laniel, M. A. Histones and histone modifications. *Curr Biol* **14**, (2004).
41. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert Rev Proteomics* **2**, 719–729 (2005).
42. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
43. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
44. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* **21**, 381–395 (2011).
45. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218 (2012).
46. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074–1080 (2001).
47. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell* **25**, 15–30 (2007).
48. Buschbeck, M. & Hake, S. B. Variants of core histones and their roles in cell fate decisions, development and cancer. *Nature Reviews Molecular Cell Biology* **18**, 299–314 (2017).
49. Wen, D. *et al.* Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. *Proc Natl Acad Sci U S A* **111**, 7325–7330 (2014).
50. Jang, C. W., Shibata, Y., Starmer, J., Yee, D. & Magnuson, T. Histone H3.3 maintains genome integrity during mammalian development. *Genes & Development* **29**, 1377 (2015).

51. Xia, W. & Jiao, J. Histone variant H3.3 orchestrates neural stem cell differentiation in the developing brain. *Cell Death & Differentiation* 2017 24:9 **24**, 1548–1563 (2017).
52. Maze, I. *et al.* Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* **87**, 77 (2015).
53. Lowe, B. R., Maxham, L. A., Hamey, J. J., Wilkins, M. R. & Partridge, J. F. Histone H3 mutations: An updated view of their role in chromatin deregulation and cancer. *Cancers* vol. 11 1–24 Preprint at <https://doi.org/10.3390/cancers11050660> (2019).
54. Hake, S. B. & Allis, C. D. Histone H3 variants and their potential role in indexing mammalian genomes: The “H3 barcode hypothesis.” *Proc Natl Acad Sci U S A* **103**, 6428–6435 (2006).
55. Zink, L. M. & Hake, S. B. Histone variants: Nuclear function and disease. *Current Opinion in Genetics and Development* vol. 37 82–89 Preprint at <https://doi.org/10.1016/j.gde.2015.12.002> (2016).
56. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678 (2010).
57. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
58. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011 469:7330 **469**, 343–349 (2011).
59. Ferrari, K. J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell* **53**, 49–62 (2014).
60. Wang, L. *et al.* A role for monomethylation of histone H3-K27 in gene activity in Drosophila. *Genetics* **208**, 1023–1036 (2018).
61. Lee, T. I. *et al.* Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* **125**, 301–313 (2006).
62. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 2009 461:7265 **461**, 762–767 (2009).
63. Lavarone, E., Barbieri, C. M. & Pasini, D. Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nature Communications* 2019 10:1 **10**, 1–16 (2019).
64. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 2006 441:7091 **441**, 349–353 (2006).
65. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007 448:7153 **448**, 553–560 (2007).
66. Bannister, A. J. *et al.* Spatial Distribution of Di- and Tri-methyl Lysine 36 of Histone H3 at Active Genes *. *Journal of Biological Chemistry* **280**, 17732–17736 (2005).

67. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology* 2012 13:2 **13**, 115–126 (2012).
68. Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol Cell* **44**, 609 (2011).
69. Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* **284**, 34283–34295 (2009).
70. An, S., Yeo, K. J., Jeon, Y. H. & Song, J. J. Crystal Structure of the Human Histone Methyltransferase ASH1L Catalytic Domain and Its Implications for the Regulatory Mechanism. *The Journal of Biological Chemistry* **286**, 8369 (2011).
71. Miyazaki, H. *et al.* Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. *PLoS Genetics* **9**, 1003897 (2013).
72. Rao, B., Shibata, Y., Strahl, B. D. & Lieb, J. D. Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide. *Molecular and Cellular Biology* **25**, 9447 (2005).
73. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 2019 573:7773 **573**, 281–286 (2019).
74. Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Molecular Cell* **70**, 371-379.e5 (2018).
75. Strahl, B. D. *et al.* Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol* **22**, 1298–1306 (2002).
76. Sessa, A. *et al.* SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron* **104**, 271–289 (2019).
77. Kizer, K. O. *et al.* A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Molecular and Cellular Biology* **25**, 3305–3316 (2005).
78. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
79. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* 2013 38:1 **38**, 23–38 (2012).
80. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, (2013).
81. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 2017 543:7643 **543**, 72–77 (2017).

82. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
83. Liu, H. *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **2021** 598:7879 **598**, 120–128 (2021).
84. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* **17**, 215 (2014).
85. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu Rev Genomics Hum Genet* **16**, 55–77 (2015).
86. Xie, S. *et al.* Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* **236**, 87–95 (1999).
87. Aubrey, B. J., Strasser, A. & Kelly, G. L. Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harbor Perspectives in Medicine* **6**, (2016).
88. Bender, S. *et al.* Reduced H3K27me3 and DNA Hypomethylation Are Major Drivers of Gene Expression in K27M Mutant Pediatric High-Grade Gliomas. *Cancer Cell* **24**, 660–672 (2013).
89. Chan, K. M. *et al.* The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes Dev* **27**, 985–990 (2013).
90. Harutyunyan, A. S. *et al.* H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nature Communications* **2019** 10:1 **10**, 1–13 (2019).
91. Justin, N. *et al.* Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* **7**, 11316 (2016).
92. Mohammad, F. *et al.* EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat Med* **23**, 483–492 (2017).
93. Stafford, J. M. *et al.* Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. *Science Advances* **4**, 5935–5966 (2018).
94. Harutyunyan, A. S. *et al.* H3K27M in Gliomas Causes a One-Step Decrease in H3K27 Methylation and Reduced Spreading within the Constraints of H3K36 Methylation. *Cell Reports* **33**, (2020).
95. Weinberg, D. N., Allis, C. D. & Lu, C. Oncogenic Mechanisms of Histone H3 Mutations. *Cold Spring Harbor Perspectives in Medicine* **7**, (2017).
96. Haag, D. *et al.* H3.3-K27M drives neural stem cell-specific gliomagenesis in a human iPSC-derived model. *Cancer Cell* **39**, 407–422.e13 (2021).
97. Buczkowicz, P., Bartels, U., Bouffet, E., Becher, O. & Hawkins, C. Histopathological spectrum of paediatric diffuse intrinsic pontine glioma: diagnostic and therapeutic implications. *Acta Neuropathologica* **128**, 573–581 (2014).

98. Chaikuad, A. *et al.* Structure of the bone morphogenetic protein receptor ALK2 and implications for fibrodysplasia ossificans progressiva. *J Biol Chem* **287**, 36990–36998 (2012).
99. Attisano, L. *et al.* Identification of human activin and TGF beta type I receptors that form heteromeric kinase complexes with type II receptors. *Cell* **75**, 671–680 (1993).
100. Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF-beta receptor. *Nature* **370**, 341–347 (1994).
101. Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* 1997 390:6659 **390**, 465–471 (1997).
102. Wrana, J. L. TGF-beta receptors and signalling mechanisms. *Miner Electrolyte Metab* **24**, 120–130 (1998).
103. Schmierer, B. & Hill, C. S. TGF β –SMAD signal transduction: molecular specificity and functional flexibility. *Nature Reviews Molecular Cell Biology* 2007 8:12 **8**, 970–982 (2007).
104. Hemmati-Brivanlou, A. & Thomsen, G. H. Ventral mesodermal patterning in *Xenopus* embryos: expression patterns and activities of BMP-2 and BMP-4. *Dev Genet* **17**, 78–89 (1995).
105. Kobayashi, T., Lyons, K. M., McMahon, A. P. & Kronenberg, H. M. BMP signaling stimulates cellular differentiation at multiple steps during cartilage development. *Proc Natl Acad Sci U S A* **102**, 18023 (2005).
106. Stewart, A., Guan, H. & Yang, K. BMP-3 promotes mesenchymal stem cell proliferation through the TGF-beta/activin signaling pathway. *J Cell Physiol* **223**, 658–666 (2010).
107. EM, S. *et al.* A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet* **38**, 525–527 (2006).
108. Kaplan, F. S. *et al.* Classic and Atypical FOP Phenotypes are Caused by Mutations in the BMP Type I Receptor ACVR1. *Hum Mutat* **30**, 379 (2009).
109. Bocciardi, R., Bordo, D., di Duca, M., di Rocco, M. & Ravazzolo, R. Mutational analysis of the ACVR1 gene in Italian patients affected with fibrodysplasia ossificans progressiva: confirmations and advancements. *European Journal of Human Genetics* **17**, 311 (2009).
110. Gregson, C. L. *et al.* A novel ACVR1 mutation in the glycine/serine-rich domain found in the most benign case of a fibrodysplasia ossificans progressiva variant reported to date. *Bone* **48**, 654 (2011).
111. Kaplan, F. S. *et al.* Multi-System Involvement in a Severe Variant of Fibrodysplasia Ossificans Progressiva (ACVR1 c.772G>A; R258G): A Report of Two Patients. *Am J Med Genet A* **167**, 2265 (2015).
112. Petrie, K. A. *et al.* Novel Mutations in ACVR1 Result in Atypical Features in Two Fibrodysplasia Ossificans Progressiva Patients. *PLoS ONE* **4**, 5005 (2009).

113. Whyte, M. P. *et al.* Fibrodysplasia ossificans progressiva: middle-age onset of heterotopic ossification from a unique missense mutation (c.974G>C, p.G325A) in ACVR1. *J Bone Miner Res* **27**, 729–737 (2012).
114. Taylor, K. R., Vinci, M., Bullock, A. N. & Jones, C. ACVR1 mutations in DIPG: Lessons learned from FOP. *Cancer Research* vol. 74 4565–4570 Preprint at <https://doi.org/10.1158/0008-5472.CAN-14-1298> (2014).
115. Han, H. J., Jain, P. & Resnick, A. C. Shared ACVR1 mutations in FOP and DIPG: Opportunities and challenges in extending biological and clinical implications across rare diseases. *Bone* **109**, 91 (2018).
116. Nagaraja, S. *et al.* Histone Variant and Cell Context Determine H3K27M Reprogramming of the Enhancer Landscape and Oncogenic State. *Molecular Cell* **76**, 965–980.e12 (2019).
117. Hoeman, C. M. *et al.* ACVR1 R206H cooperates with H3.1K27M in promoting diffuse intrinsic pontine glioma pathogenesis. *Nature Communications* **10**, (2019).
118. Fortin, J. *et al.* Mutant ACVR1 Arrests Glial Cell Differentiation to Drive Tumorigenesis in Pediatric Gliomas. *Cancer Cell* **37**, 308–323.e12 (2020).
119. Lutsik, P. *et al.* Globally altered epigenetic landscape and delayed osteogenic differentiation in H3.3-G34W-mutant giant cell tumor of bone. *Nature Communications* 2020 11:1 **11**, 1–16 (2020).
120. Sangatsuda, Y. *et al.* Base-resolution methylomes of gliomas bearing histone H3.3 mutations reveal a G34 mutant-specific signature shared with bone tumors. *Scientific Reports* 2020 10:1 **10**, 1–13 (2020).
121. Wen, H. *et al.* ZMYND11 links histone H3.3 K36 trimethylation to transcription elongation and tumor suppression. *Nature* **508**, 263 (2014).
122. Cheng, Z. *et al.* A molecular threading mechanism underlies Jumonji lysine demethylase KDM2A regulation of methylated H3K36. *Genes & Development* **28**, 1758 (2014).
123. Zhang, Y. *et al.* Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Scientific Reports* 2017 7:1 **7**, 1–9 (2017).
124. Shi, L., Shi, J., Shi, X., Li, W. & Wen, H. Histone H3.3 G34 Mutations Alter Histone H3K36 and H3K27 Methylation In Cis. *Journal of Molecular Biology* **430**, 1562–1565 (2018).
125. Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X. & Chen, E. S. Structural and functional specificity of H3K36 methylation. *Epigenetics & Chromatin* 2022 15:1 **15**, 1–20 (2022).
126. Rayasam, G. V. *et al.* NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J* **22**, 3153–3163 (2003).
127. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. *Nature* 2009 460:7252 **460**, 287–291 (2009).
128. Bjornsson, H. T. The Mendelian disorders of the epigenetic machinery. *Genome Research* **25**, 1473–1481 (2015).

129. Sotos, J. F., Dodge, P. R., Muirhead, D., Crawford, J. D. & Talbot, N. B. Cerebral Gigantism in Childhood. <http://dx.doi.org/10.1056/NEJM196407162710301> **271**, 109–116 (2010).
130. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature Genetics* **2002 30:4** **30**, 365–366 (2002).
131. Kurotaki, N. *et al.* Fifty Microdeletions Among 112 Cases of Sotos Syndrome: Low Copy Repeats Possibly Mediate the Common Deletion. *Human Mutation* **22**, 378–387 (2003).
132. Waggoner, D. J. *et al.* NSD1 analysis for Sotos syndrome: Insights and perspectives from the clinical laboratory. *Genetics in Medicine* **2005 7:8** **7**, 524–533 (2005).
133. Boczek, N. J. *et al.* Developmental delay and failure to thrive associated with a loss-of-function variant in WHSC1 (NSD2). *Am J Med Genet A* **176**, 2798–2802 (2018).
134. Marzin, P. *et al.* SETD2 related overgrowth syndrome: Presentation of four new patients and review of the literature. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* **181**, 509–518 (2019).
135. van Rij, M. C. *et al.* Two novel cases expanding the phenotype of SETD2-related overgrowth syndrome. *Am J Med Genet A* **176**, 1212–1215 (2018).
136. Tatton-Brown, K. *et al.* Mutations in the DNA methyltransferase gene, DNMT3A, cause an overgrowth syndrome with intellectual disability. *Nat Genet* **46**, 385 (2014).
137. Descipio, C. *et al.* Subtelomeric deletion of chromosome 10p15.3: clinical findings and molecular cytogenetic characterization. *Am J Med Genet A* **158A**, 2152–2161 (2012).
138. Cobben, J. M. *et al.* A de novo mutation in ZMYND11, a candidate gene for 10p15.3 deletion syndrome, is associated with syndromic intellectual disability. *Eur J Med Genet* **57**, 636–638 (2014).
139. Oates, S. *et al.* ZMYND11 variants are a novel cause of centrotemporal and generalised epilepsies with neurodevelopmental disorder. *Clin Genet* **100**, 412–429 (2021).
140. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185–188 (1999).
141. Hu, M. *et al.* Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc Natl Acad Sci U S A* **107**, 2956–2961 (2010).
142. Osipovich, A. B., Gangula, R., Vianna, P. G. & Magnuson, M. A. Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. *Development* **143**, 4595–4607 (2016).
143. Bryant, L. *et al.* Histone H3.3 beyond cancer: Germline mutations in Histone 3 Family 3A and 3B cause a previously unidentified neurodegenerative disorder in 46 patients. *Science Advances* **6**, 58 (2020).
144. Maver, A., Čuturilo, G., Ruml, S. J. & Peterlin, B. Clinical Next Generation Sequencing Reveals an H3F3A Gene as a New Potential Gene Candidate for Microcephaly Associated with Severe

- Developmental Delay, Intellectual Disability and Growth Retardation. *Balkan J Med Genet* **22**, 65–68 (2019).
145. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114 (2014).
 146. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15 (2013).
 147. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 148. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 (2010).
 149. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
 150. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
 151. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191–W198 (2019).
 152. Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 060012 (2016) doi:10.1101/060012.
 153. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).
 154. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021).
 155. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017).
 156. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
 157. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **2018** 37:1 **37**, 38–44 (2018).
 158. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).
 159. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B* **2013** 86:11 **86**, 1–14 (2013).
 160. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).
 161. Ma, F. & Pellegrini, M. ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics* **36**, 533–538 (2020).

162. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108 (2009).
163. Monje, M. *et al.* Hedgehog-responsive candidate cell of origin for diffuse intrinsic pontine glioma. *Proc Natl Acad Sci U S A* **108**, 4453–4458 (2011).
164. Escartin, C. *et al.* Reactive astrocyte nomenclature, definitions, and future directions. *Nature Neuroscience* **24**, 312–325 (2021).
165. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
166. Motori, E. *et al.* Article Inflammation-Induced Alteration of Astrocyte Mitochondrial Dynamics Requires Autophagy for Mitochondrial Network Maintenance. (2013) doi:10.1016/j.cmet.2013.11.005.
167. Wheeler, M. A. *et al.* MAFG-driven astrocytes promote CNS inflammation. *Nature* **578**, 593–599 (2020).
168. Ubhi, K. & Price, J. Expression of POU-domain transcription factor, Oct-6, in schizophrenia, bipolar disorder and major depression. *BMC Psychiatry* **5**, (2005).
169. Hedreen, J. C., Peyser, C. E., Folstein, S. E. & Ross, C. A. Neuronal loss in layers V and VI of cerebral cortex in Huntington's disease. *Neuroscience Letters* **133**, 257–261 (1991).
170. Romito-DiGiacomo, R. R., Menegay, H., Cicero, S. A. & Herrup, K. Effects of Alzheimer's Disease on Different Cortical Layers: The Role of Intrinsic Differences in A β Susceptibility. *The Journal of Neuroscience* **27**, 8496 (2007).
171. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315 (2009).
172. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience* **17**, 215–222 (2013).
173. Kriaucionis, S. & Bird, A. DNA methylation and Rett syndrome. *Hum Mol Genet* **12 Spec No 2**, (2003).
174. Tillotson, R. *et al.* Neuronal non-CG methylation is an essential target for MeCP2 function. *Molecular Cell* **81**, 1260–1275.e12 (2021).
175. Lavery, L. A. *et al.* Losing dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting rett syndrome. *Elife* **9**, (2020).
176. Lin, C. Y. *et al.* Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
177. Gogliotti, R. G. *et al.* Total RNA-sequencing of Rett Syndrome Autopsy Samples Identifies the M4 Muscarinic Receptor as a Novel Therapeutic Target. *Journal of Pharmacology and Experimental Therapeutics* **365**, 291–300 (2018).

178. Aldinger, K. A. *et al.* Transcriptome data of temporal and cingulate cortex in the Rett syndrome brain. *Scientific Data* 2020 7:1 **7**, 1–8 (2020).
179. Kan, H. *et al.* Single-cell transcriptome analysis reveals cellular heterogeneity in the ascending aortas of normal and high-fat diet-fed mice. *Experimental & Molecular Medicine* **53**, 1379 (2021).
180. Rosen, E. D. & Spiegelman, B. M. What We Talk About When We Talk About Fat. *Cell* **156**, 20 (2014).
181. Chawla, A., Schwarz, E. J., Dimaculangan, D. D. & Lazar, M. A. Peroxisome proliferator-activated receptor (PPAR) gamma: adipose-predominant expression and induction early in adipocyte differentiation. *Endocrinology* **135**, 798–800 (1994).
182. Siersbæk, R., Nielsen, R. & Mandrup, S. PPARgamma in adipocyte differentiation and metabolism--novel insights from genome-wide studies. *FEBS Lett* **584**, 3242–3249 (2010).
183. Bond, A. M., Bhalala, O. G. & Kessler, J. A. The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Developmental Neurobiology* **72**, 1068–1084 (2012).
184. Valer, J. A., Sánchez-de-Diego, C., Pimenta-Lopes, C., Rosa, J. L. & Ventura, F. ACVR1 Function in Health and Disease. *Cells* **8**, (2019).
185. Krug, B., Harutyunyan, A. S., Deshmukh, S. & Jabado, N. Polycomb repressive complex 2 in the driver's seat of childhood and young adult brain tumours. *Trends in Cell Biology* vol. 31 814–828 Preprint at <https://doi.org/10.1016/j.tcb.2021.05.006> (2021).
186. Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat Struct Biol* **9**, 217 (2002).
187. Rona, G. B., Eleutherio, E. C. A. & Pinheiro, A. S. PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophysical Reviews* **8**, 63 (2016).
188. Lin, P. *et al.* Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q Genes in Rett syndrome. *BMC Genomics* **17**, 1–11 (2016).
189. Heyn, P. *et al.* Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat Genet* **51**, 96–105 (2019).
190. Tang, M. C. W. *et al.* Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. *PLoS Genet* **11**, 1–23 (2015).
191. de Heredia, F. P., Gómez-Martínez, S. & Marcos, A. Obesity, inflammation and the immune system. *Proc Nutr Soc* **71**, 332–338 (2012).
192. Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature* 2022 603:7903 **603**, 926–933 (2022).
193. Herrera, B. M., Keildson, S. & Lindgren, C. M. Genetics and epigenetics of obesity. *Maturitas* **69**, 41 (2011).

194. Zhuang, L. *et al.* Depletion of Nsd2-mediated histone H3K36 methylation impairs adipose tissue development and function. *Nature Communications* **9**, (2018).
195. G. Butler, M. Prader-Willi Syndrome: Obesity due to Genomic Imprinting. *Current Genomics* **12**, 204 (2011).
196. Soejima, H. & Higashimoto, K. Epigenetic and genetic alterations of the imprinting disorder Beckwith–Wiedemann syndrome and related disorders. *Journal of Human Genetics* **2013** 58:7 **58**, 402–409 (2013).