Functional characterization of histone H3.3-G34 mutations in development & cancer

Sima Khazaei

Department of Human Genetics
McGill University, Montreal, Canada

October 2020

Supervisor: Nada Jabado, MD, PhD

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy Human Genetics

"Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world."

Louis Pasteur

Abstract

Recurrent mutations impacting the glycine 34 codon of *H3F3A* encoding noncanonical H3 variant H3.3 arise in specific cancers: G34R and rarely G34V mainly occur in high grade gliomas (HGGs) of the cerebral cortex, while G34W characterize 90% of a specific bone cancer, giant cell tumor of the bone (GCTB). There is limited information on H3.3G34-mutant's oncogenic mechanisms. Using an innovative initial mosaic step approach, we generated direct knock-in (DKI) mice carrying heterozygous germline H3.3 (*H3f3a*) G34R, V or W mutations. Our data in two different mouse genetic backgrounds consistently show drastically distinct phenotypes between G34R and W mutant mice, while H3.3G34V mice show a similar but milder phenotype to that observed in H3.3G34R DKI mice and no phenotype is observed in mice carrying a deletion of one *H3f3a* allele. Indeed, severe neurological issues are observed in G34R mutant mice and to a lesser degree in G34V, while minor neurological insults are observed in G34W mice, which in contrast have bladder and uretero-genital issues. All G34 mutant mice exhibit an obesity phenotype, which is most severe in G34W mice. This is consistent with patient data where G34R/V mutations in *H3F3A* seem exclusive to HGG of the brain while H3.3G34W are specific to mesenchymal tumors namely GCTB. GCTB are characterized by a neoplastic H3.3G34W-mutant stromal population and a bone-resorptive giant osteoclast population. The mechanism by which the H3.3G34W mutation alters the epigenome to give rise to GCTB remains unknown. Here, we generated isogenic cell lines from patient derived GCTB stromal cultures and defined the epigenetic and transcriptomic changes mediated by H3.3G34W. G34W induced H3.3K36me3 loss and H3.3K27me3 gain on mutant H3.3 histones. We show that H3.3G34W promotes H3K27me3 redistribution from intergenic to genic regions. This epigenetic dysregulation and associated transcriptional changes are linked to cell identity, with upregulation of extracellular matrix organization genes and downregulation of contractile actomyosin pathways in H3.3G34W cells. Single-cell transcriptomics of GCTB tumors reveals that G34W stromal cells resemble osteoblast progenitors with distinct populations comprising a neoplastic trajectory from an *SPP1+IBSP+* osteoblast-like population towards an *ACTA2+* contractile population. Correction of the H3.3G34W mutation by CRISPR-Cas9 in GCTB stromal cells results in increased progression to myogenic differentiation along this trajectory *in vitro*. We further show that *ACTA2+* contractile H3.3G34W stromal cells secrete ligands promoting extracellular matrix remodeling, likely

facilitating association and recruitment of osteoclasts. Importantly, orthotopic tibial injection of H3.3G34W, but not isogenic CRISPR-edited stromal cells, resulted in aggressive osteolytic tumors with recruitment of murine multinucleated osteoclasts, indicating a persistent requirement for H3.3G34W mutation in GCTB tumorigenesis.

Résumé

Des mutations récurrentes du codon glycine 34 (G34) du gène *H3F3A*, qui code pour le variant non canonique H3.3 de l'histone H3, sont retrouvés dans certains cancers specifiques. En effet, H3.3G34R et rarement G34V surviennent principalement dans les gliomes de haut grade (HGG) du cortex cérébral, tandis que G34W caractérise 90% d'un cancer des os spécifique, les tumeurs à cellules géantes de l'os (GCTB). La compréhension des mécanismes d'oncogenèse des cancers mutants H3.3G34 est limitée. C'est pourquoi, à l'aide d'une méthode innovante dite « initial mosaic step approach », nous avons généré des souris Knock-In (DKI) portant des mutations germinales hétérozygotes H3.3 (*H3f3a*) G34R, V ou W. Nos données obtenues dans deux fonds génétiques différents de souris montrent de manière systématique des phénotypes radicalement distincts entre les souris mutantes G34R et W. Les souris H3.3G34V montrent un phénotype similaire mais plus modéré par rapport aux souris H3.3G34R. En revanche, aucun phénotype n'est observé chez les souris portant une délétion d'un allèle *H3f3a*. En effet, de graves problèmes neurologiques sont observés chez les souris mutantes G34R et à un moindre degré chez les souris G34V, tandis que ces atteintes neurologiques sont mineures chez les souris G34W. Ces dernières présentent en revanche des problèmes de vessie et urétro-génital. Toutes les souris mutantes G34 présentent un phénotype d'obésité, qui est plus sévère chez les souris G34W. L'ensemble de ces résultats est cohérent avec les données de patients où les mutations G34R / V sur le gène *H3F3A* semblent exclusives aux HGG tandis que la mutation H3.3G34W est spécifique aux tumeurs mésenchymateuses que sont les GCTB. Les GCTB sont caractérisés par une population stromale néoplasique mutante H3.3G34W et une population d'ostéoclastes géants à résorption osseuse. Le mécanisme par lequel la mutation H3.3G34W modifie l'épigénome pour donner naissance aux GCTB reste inconnu. C'est pourquoi nous avons généré des lignées cellulaires isogéniques à partir de cultures stromales de GCTB dérivées de patients et défini les changements épigénétiques et transcriptionnels médiés par la mutation H3.3G34W. Ainsi, nous avons observé que G34W induit une perte de H3.3K36me3 et un gain de H3.3K27me3 sur des histones H3.3 mutantes. Nous montrons également que H3.3G34W favorise la redistribution de H3K27me3 des régions intergéniques vers des régions géniques. Cette dérégulation épigénétique et les changements transcriptionnels associés sont liés à l'identité cellulaire, avec une régulation positive des gènes d'organisation de la matrice extracellulaire et une régulation négative des voies

de l'actomyosine contractile dans les cellules H3.3G34W. L'analyse de transcriptomes en cellule unique des tumeurs GCTB révèle que les cellules stromales G34W ressemblent à des progéniteurs d'ostéoblastes avec des populations distinctes comprenant une trajectoire néoplasique d'une population de type ostéoblaste *SPP1*+ *IBSP*+ vers une population contractile *ACTA2*+. La correction de la mutation H3.3G34W par CRISPR-Cas9 dans les cellules stromales de GCTB entraîne une progression accrue vers la différenciation myogénique *in vitro*. Nous montrons en outre que les cellules stromales contractiles H3.3G34W *ACTA2*+ sécrètent des ligands favorisant le remodelage de la matrice extracellulaire, facilitant probablement l'association et le recrutement des ostéoclastes. Surtout, l'injection tibiale orthotopique de H3.3G34W, mais pas de cellules stromales isogéniques éditées par CRISPR, a engendré des tumeurs ostéolytiques agressives avec recrutement d'ostéoclastes murins multinucléés. L'ensemble démontre une dépendance persistante à la mutation H3.3G34W du processus de tumorigénèse des GCTB.

Table of Contents

List of Abbreviations

2HG	2-hydroxyglutarate
ACVR1	Activin A receptor 1
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
ALT	Alternative lengthening of telomeres
AML	Acute myeloid leukemia
ATRX	α- thalassaemia/ mental retardation syndrome X-linked
BAT	Brown adipose tissue
BOS	Base of support
CABIN1	Calcineurin binding protein 1
CAM	Chick chorioallantoic membrane
CBC	complete blood count
CENP-A	Centromere protein A
ChIP-seq	chromatin immunoprecipitation sequencing
DAXX	Death Associated Protein
DEGs	Differentially expressed genes
DIPG	Diffuse intrinsic pontine gliomas
DKI	Direct knock-in
DKI	Direct knock in
DNMTs	DNA methyltransferases
ECM	Extracellular matrix
ECM	Extracellular matrix
ERV	Endogenous retroviral element
ESC	Embryonic stem cell
FGFR	Fibroblast growth factor receptor
FGFR1	Fibroblast growth factor receptor 1
GCLJ	Giant cell lesions of the jaw
GCTB	Giant cell tumor of the bone
GCTB	Giant cell tumor of bone
GFAP	Glial fibrillary acidic protein
gRNA	Guide RNA
GSEA	Gene-set enrichment analysis
H&E	Haemotoxylin and Eosin
HAT	Histone acetyltransferase
HD	Huntington's diseases
HDAC	Histone deacetylase

HDACs	Histone deacetylases
HDM	Histone demethylase
HGGs	High grade gliomas
HIRA	Histone regulator A
HMT	Histone methyltransferase
HSC	Hematopoietic stem cell
hTERT	Human telomerase reverse transcriptase
IHC	Immunohistochemistry
Ki67	Proliferation marker protein Ki-67
KO	knockout
M-CSF	Macrophage colony stimulating factor
MBDs	Methyl-CpG binding domain proteins
MBP	Myelin basic protein
MEF	Mouse embryonic fibroblast
MS	Mass spectrometry
MSC	Mesenchymal stem cell
NASH	Non-alcoholic steatohepatitis
NSC	Neural stem cells
NSPC	Neural stem/progenitor
NueN	Neuronal nuclei
O/E	Overexpression
Olig1/2	Oligodendrocyte Transcription Factor 1/2
OPG	Osteoprotegerin
PCA	Principal component analysis
PD	Parkinson's disease
PDGFB	platelet-derived growth factor subunit B
PDGFRA	platelet-derived growth factor receptor
PFA-EP	Posterior fossa ependymomas group A
PRC2	Polycomb repressive complex 2
PTMs	Post-translational modifications
RANK	Receptor activator of nuclear factor κ B
RANKL	Receptor activator of nuclear factor kB ligand
RNA-seq	RNA sequencing
scRNA-seq	Single cell RNA sequencing
ssODNs	Single-stranded donor oligonucleotides
SVZ	Subventricular zone

TRAP	Tartrate-resistant acid phosphatase
TSS	Transcription start sites
UBN1	Ubinuclein1
WAT	White adipose tissue
WES	Whole-exome sequencing
WHS	Wolf-Hirschhorn syndrome
WVS	Weaver syndrome
XCI	X-chromosome inactivation

List of Figures

List of Tables

Acknowledgments

I would like to express my special thanks to my supervisor, **Dr Nada Jabado**, for her great support and supervision during the 6-years of my PhD program. She continuously encouraged me and explored my scientific ideas, even they seemed like impossible at first. She gave me the opportunity to do a variety of research, I learned many new subjects and techniques, which will form a solid basis to fulfill my dream of being a strong and independent researcher like her. Nada, thanks you so much for taking me as PhD student, it makes me proud.

I would also like to extend my appreciation to my supervisory committee members, **Dr. Livia Garzia**, **Dr. Claudia Kleinman**, **Dr. Loydie Majewska** and **Dr. Peter Siegel** for their invaluable guidelines during my PhD years.

My warm thanks to the labs project managers, **Dr Eef Harmsen** and **Dr Nikoleta Juretic** for their vison, friendship and empathy. You always encouraged me and showed me to look at the bright side of life. Thank you both for all your support and listening to me when I needed it.

I would like to thanks to my wonderful lab colleagues for being my big family during the last 6 years. For sure, without your help and support, I wouldn't have accomplished my PhD thesis. I learned a lot from you and enjoyed working with you. You created beautiful memories during my PhD years.

I would like to take this opportunity to thank you my parents and family. Thank you, my **mom** and **dad**, for helping me to shape my life, find my way and follow my dreams. Without your inspiration, help and support, I'd never could the person I am today. Thank you for everything! I would also like to thank my brother, **Mamad**, my sister, **Mina**, my sister in law, **Fereshteh** and my brother in law, **Reza**. Thank you for your incredible supports during this journey. Thanks for all your advices and sharing your experiences, your kindness cannot be counted!

I would like to specially thanks, my beloved, husband, **Fahim**, for all his support and understanding during my pursuit of PhD. Thanks for your endless love. Your positive attitude in

life helped me a lot in this journey. I always had you and your help by my side even though we were miles apart. Saying ‘Thank you’ is hardly enough!

Contribution to Original Knowledge

This study comprehensively defines and characterizes the role of histone H3.3G34 mutations in the contexts of developmental processes and cancer.

Chapter 3 We characterized for the first time the developmental effects of histone H3.3 G34 mutations using a direct knock-in (DKI) mouse model with germline H3.3 G34 mutations in the endogenous locus. We show that H3.3G34R and H3.3G34W mutations differentially affect mouse development, consistent with related cancer types where H3.3G34R occurs in pediatric HGGs and H3.3G34W occurs in GCTB.

Chapter 4 We comprehensively characterized the epigenome, transcriptome (including at the single-cell level) and secreted proteome of patient samples and tumor-derived cells CRISPR/Cas9-edited for H3.3G34W. Also, using orthotopic tibial injections, we are the first to show that the H3.3G34W mutation is necessary for tumor formation *in vivo.* Our findings suggest that H3.3G34W leads to GCTB by sustaining a transformed state in osteoblast-like progenitors which promotes neoplastic growth, pathological recruitment of giant osteoclasts, and bone destruction.

Format of the Thesis

This thesis is written in the traditional format and contains six chapters. Chapter I is the introduction, with a review of literature related to my study as well as rationales, hypothesis and objectives. Chapter II details the materials and methods used in this study. Chapter III describes the functional characterizations of histone H3.3 (*H3f3a*) G34 mutations on development in murine models. Chapter IV reports the effects of H3.3G34W mutation on tumor growth in GCTB and the differentiation of osteoblast-like progenitors. This chapter is part of a study that has been submitted to Cancer Discovery, with I am one of the first co-authors. Chapter V is a discussion of all the results, while Chapter VI describes overall conclusion and discusses some of the future directions.

Contribution of Authors

Chapter 3, I designed the study and all the experiments, generated DKI mice using CRISPR/Cas9, performed mouse genotyping, carried out functional studies (assessment of behavioral phenotypes, investigating gross morphology and histopathology of major organs, performing immunostaining (using multiple markers) of affected tissues and target organs), analyzed and interpreted the functional and RNA-seq data.

Pariya Azarafshar, Josiane França and Bethany Tong contributed to performing mouse genotyping, behavioral phenotypes and immunostaining of tissues. Tianna Sihota contributed to performing mouse genotyping and behavioral phenotypes. Nisha Kabir led bioinformatic analysis. Carol Chen, Augusto Faria Andrade, Aurèle Besse-Patin, and Michele Zeinieh contributed to interpretation of the results. Xiaohong Li performed body composition and metabolic chamber tests. Veronica Sandy, Caterina Russo contributed to mouse breeding and colony maintenance. Mitra Cowan contributed in animal model generation and maintenance. Maia Kokoeva, Heidi McBride, Loydie majewska, Indra Gupta, Livia Garzia, Claudia Kleiman contributed to the study conceptualization and data interpretation. Nada Jabado supervised all the aspects of the project.

The work presented in **Chapter 4** was included in a manuscript (Cancer Discovery, 2020) and authored by Sima Khazaei*, Nicolas De Jay*, Shriya Deshmukh*, Liam D. Hendrikse*, Wajih Jawhar, Carol C.L. Chen, Leonie G. Mikael, Damien Faury, Dylan M. Marchione, Joel Lanoix, Éric Bonneil, Takeaki Ishii, Siddhant U. Jain, Kateryna Rossokhata, Tianna Sihota, Robert Eveleigh, Véronique Lisi, Ashot S. Harutyunyan, Sungmi Jung, Jason Karamchandani, Brendan C. Dickson, Robert E. Turcotte, Jay Wunder, Pierre Thibault, Peter W. Lewis, Benjamin A. Garcia, Stephen C. Mack, Michael D. Taylor, Livia Garzia, Claudia L. Kleinman, Nada Jabado. Sima Khazaei, Nicolas De Jay, Shriya Deshmukh, Liam D. Hendrikse contributed equally.

I led the study conceptualization, generated isogenic GCTB G34W edited patient-derived cell lines using CRISPR/Cas9, performed *in vitro* assays (proliferation, clonogenicity, and cell differentiation analysis), carried out and analyzed patient-derived subcutaneous and orthotopic xenograft mouse models, contributed to interpretation of mass spectrometry of histone PTMs data,

performed overexpression of H3.3G34W in CRISPR-edited cell lines, contributed to analysis and interpretation of genomic data.

Shriya Deshmukh led the study conceptualization, contributed to creating the isogenic cell lines and functional studies, perfomed Chip-seq and contributed to analysis and interpretation of genomic data. Nicolas De Jay led the study conceptualization, performed genomics data (RNA-seq and Chip-seq) analysis and interpretation. Liam D. Hendrikse led the study conceptualization, performed genomics data (scRNA-seq) analysis and interpretation.

W.J., K.R., and T.S. performed functional studies, analyzed and interpreted data. C.C.L.C. and S.U.J. contributed to study conceptualization and data interpretation. L.G.M. facilitated patient sample acquisition, generated primary tumor-derived cell lines, and contributed to manuscript preparation. D.F. processed samples for single-cell transcriptomics and contributed to data interpretation. D.M.M. and B.A.G. led histone proteomics experiments and analysis. J.L., E.B., and P.T. led secreted proteomics experiments and analysis. T.I. performed PDOX experiments. R.E. and V.L. contributed to bioinformatics analyses of exome sequencing and single-cell transcriptomic data respectively. A.S.H. performed ChIP-seq experiments and contributed to data interpretation. S.J. and J.K. assessed and interpreted histopathology results. B.C.D., R.E.T., J.W. facilitated patient sample acquisition and contributed to data interpretation. P.W.L., S.C.M., M.D.T., and L.G. contributed to study conceptualization and data interpretation. C.L.K. and N.J. co-led and supervised all aspects of the project.

Chapter I: Introduction

1.1. Definition of epigenetics and epigenetic factors

Waddington was the first to use the term epigenetics in 1942 to refer to the study of heritable traits (phenotypes) regulated by factors that do not alter DNA sequences[1]. Epigenetic regulation occurs either via direct DNA methylation, or via post-translational modifications (PTMs) of histones[2]. Both processes play a central role in fine-tuning underlying transcription of genes and are essential for the precise control of developmental processes and the steering of cellular fate. Disruptions of epigenetic regulators can lead to developmental disorders and many cancers[3,4].

1.1.1. DNA methylation

DNA methylation involves the addition of a methyl group to the fifth carbon of cytosine of CpG dinucleotides by DNA methyltransferases (DNMTs). DNA methylation of CpG islands in promoter regions and constitutive heterochromatin is associated with silencing of gene expression. Also, DNA methylation is involved in epigenetic phenomena such as genomic imprinting and X-chromosome inactivation (XCI)[5]. Methyl-CpG binding domain proteins (MBDs) and methyl-CpG binding zinc-finger proteins of the Kaiso family recognize methyl-CpG marks, resulting in activation of histone deacetylases (HDACs) and subsequent depletion of the acetyl group from histone proteins, which ultimately leads to downregulation of gene transcription[6]. Cancers show changes in DNA methylation compared to normal cells, both with gain of methylation (hypermethylation) at promoter regions of tumor suppressor genes as well as global loss of methylation (hypomethylation) at regions normally silenced for genome stability[7].

1.1.2. Histones

Histones are the building block of nucleosomes, the basic unit of chromatin. A nucleosome is formed by an octamer consisting of heterodimers of the core histones H2A, H2B, H3 and H4 around which 147 base pairs of DNA are coiled[8]. Histone H1 is attached to DNA and links the nucleosomes to stabilize the higher order chromatin structure and helps to regulate gene expression[9]. Each of the core histones has a globular domain that mediates histone–histone interactions within the nucleosomes and an amino terminal (N-terminal) tail that is enriched in

basic amino acids 20–35 residues in length. The histone tails can be modified through post-translational modifications (PTMs) which are central in epigenetic regulation and affect transcriptional activation, silencing, chromatin assembly and DNA replication[10]. PTMs play an important role in remodeling of the compacted DNA/histone structure (unwrapping of DNA from the nucleosome core) to allow access of the protein's molecular machinery for genomic DNA to be read, copied or transcribed[11].

1.2. Histone variant H3.3

1.2.1. Histone H3.3 vs H3.1/2

Histones H3 and H4 are the most evolutionarily conserved histones. Multiple H3 variants with distinct PTM "signatures" play an important role in development and cell differentiation. Several H3 variants are expressed in human including H3.1, H3.2, H3.3, the centromere protein A (CENP-A) as well as two primate-specific H3.X and H3.Y, and the testis-specific histones H3t and H3.5[12] .The canonical histones H3.1 and H3.2 are synthesized during the S phase and account for the majority of the histone pool during replication[13]. H3.3, an ancient and conserved histone H3 variant, is a protein with 135 amino acids which has the typical structure similar to the four core histones: a histone tail and three α-helices connected by two loop domains[14]. In contrast to H3.1 and H3.2, H3.3 is expressed throughout the cell cycle. It differs from its H3.1 and H3.2 counterparts by four (Ser31, Ala87, Ile89, Gly90) and five (Ser31, Ala87, Ile89, Gly90, Ser96) amino acids, respectively[15,16]. H3.3 is expressed from two separate genes, *H3F3A* which is on chromosome 1 and *H3F3B* on chromosome 17. This is in contrast to the cluster of coding genes of H3.1 and H3.2, encoded by ten genes and three genes respectively[17]. *H3F3A* and *H3F3B* have poly-A tails and introns. Although the *H3F3A* and *H3F3B* genes express identical proteins, they have different regulatory sequences and produce different transcripts[18–21].

1.2.2. Chaperones of H3.3

An important role for H3.3 in nucleosomes is its exchange with H3.1 and H3.2, resulting in less compact chromatin to facilitate transcription of active genes. In addition, H3.3 is enriched in heterochromatin regions such as telomeres and centromeres[22,23]. H3.3 marks active genes[24], maintains epigenetic memory[25], and maintains heterochromatin and telomeric integrity[26]. Its

exchange with H3.1 or H3.2 is aided by specific chaperone proteins such as histone regulator A (HIRA), Ubinuclein1 (UBN1), and calcineurin binding protein 1 (CABIN1) that all mediate the distribution of H3.3 at gene transcription start sites (TSS), regulatory elements including promoters and enhancers as well as gene bodies of transcribed (active) genes[27]. The chaperone duo DAXX (Death Associated Protein)/ATRX (α- thalassaemia/ mental retardation syndrome X-linked) mediates the incorporation of H3.3 into heterochromatin, such as telomeres, pericentromeric regions, repetitive elements and imprinted genes[28–30].

1.2.3. The role of H3.3 in development

Both *H3F3B* and *H3F3A* genes are highly conserved in mammals[17] and are expressed ubiquitously during embryo development and throughout life[31]. In mouse adult tissues, H3.3 is expressed broadly with relatively high expression levels in reproductive organs (testes and ovaries) as well as in immune organs such as thymus and bone marrow[31,32].

In the brain, which is enriched in terminally differentiated postmitotic cells, H3.3 can build up to nearly 90% of the total H3 in post-mitotic neurons[33,34]. In both mouse and human postnatal brains, H3.3 replaces H3.1 and H3.2 over time to reach this near saturation levels by mid-age[23,33]. This increased H3.3 incorporation into actively transcribed DNA regions and its efficient turnover in neuronal and glial cells are crucial to maintain cell type–specific gene expression as well as synaptic connectivity and brain cognition[33] . Similar H3.3 accumulation patterns were observed in liver and kidney, although not in heart tissues [23].

In mice, single knockouts (KO) and double KO of *H3f3a* and *H3f3b* have been studied extensively. The phenotypes depend on the genetic background and techniques used to generate the mouse models, with some discrepancies between reported phenotypes. Hypomorphic retroviral gene traps of *H3f3a* resulted in mice with a reduced viability. The survivors had reduced growth rates due to their inability to compete with wild-type littermates for food. Furthermore, the males were sub-fertile/infertile and exhibited reduced copulatory activity[35].

Heterozygote *H3f3a*-KO (*H3f3a*$^{+/-}$) mice were found to be normal and fertile. Homozygote *H3f3a*-KO (*H3f3a*$^{-/-}$) females were of normal size, although males were slightly smaller compared to *H3f3a*$^{+/-}$ and WT males at 3-6 weeks of age[36].

When the *H3f3b* gene is disrupted, it results in a more severe phenotype than when *H3f3a* is disrupted. While *H3f3a*$^{-/-}$ mutants of both sexes were normal in behaviour and appearance[36], heterozygote *H3f3b*-KO (*H3f3b*$^{+/-}$) male mice were infertile due to spermiogenesis failure[36]. *H3f3b*$^{+/-}$ mutants were also growth deficient. Homozygote *H3f3b*-KO (*H3f3b*$^{-/-}$) mutants died at birth[36].

Mouse embryos (E18.5) with loss of a single H3.3- encoding gene (*H3f3a*$^{-/-}$ or *H3f3b*$^{-/-}$) contained the same amount of residual H3.3^{A} (encoded from *H3f3a*) and H3.3^{B} (encoded from *H3f3b*) in various somatic tissues such as kidney, liver and lung. The notable difference between the *H3f3a*$^{-/-}$ and *H3f3b*$^{-/-}$ mice was found in the brain, where *H3f3b*$^{-/-}$ animals had a relatively higher level of residual H3.3^{A} [36].

A complete KO of both *H3f3a* and *H3f3b* totally obliterated H3.3 protein levels and resulted in preimplantation embryonic lethality. RNA-seq analysis on H3.3 depleted embryos compared to WT embryos showed only small changes in the global transcriptome and only ∼5% of transcribed genes were affected. However, H3.3 loss did lead to dysfunction of heterochromatin structures at telomeres, centromeres, and pericentromeric heterochromatin sites. Furthermore, H3.3 loss in mouse embryonic stem cells (ESCs) cells and in embryonic fibroblasts (MEFs) led to mitotic defects (such as anaphase bridges and lagging chromosomes) and cell death[31].

1.3. Histone H3 post-translational modifications

Processes such as replication, transcription, regulation of gene expression, chromosome segregation, and DNA repair are modulated by N-terminal post translational modifications (PTM) of specific histones residues (the histone code)[10]. PTMs are regulated by well-studied enzymes and can be grouped into enzymes that deposit (writers: such as histone methyltransferases), remove (erasers: such as histone demethylases) and interpret (readers: proteins that contain specialized binding regions, such as bromodomains) these modifications. The context of histone PTMs are

well known. In histone H3.3 at least five lysine amino acids (K4, K9, K27, K36 and K79) can be methylated[37]. Methylation of the lysine residues H3K9 and H3K27 is associated with gene silencing, while methylation of H3K4, H3K36, and H3K79 is associated with transcriptional activation. In addition to methylation, at least 16 more kinds of PTMS, such as acetylation, phosphorylation and ubiquitination, have been detected on more than 30 amino acid residues of human histone H3 variants[38].

This chapter will focus on the methylation of H3K27 and H3K36, as the interest of this thesis is the function of H3.3G34 mutations in cancer and development. H3.3G34 mutations are thought to play a role as epigenetic modifiers via a reduction of H3K36me3[39] and H3K27me3 gain[40] in *cis* on the mutant histone, discussed in the following sections in more detail.

1.3.1. H3K27 and H3K36 post-translational modifications

Lysine residues on histones can be methylated and demethylated by histone methyltransferases (HMTs) and histone demethylases (HDMs), respectively, while the acetylation of lysine is mediated by histone acetyltransferases (HATs) and histone deacetylases (HDACs). H3K27 can either be mono-, di-, or trimethylated (H3K27me1, H3K27me2 and H3K27me3) or acetylated (H3K27ac). H3K27 methylation is established by the two mutually exclusive catalytic enzymes EZH1 or EZH2 as part of the polycomb repressive complex 2 (PRC2)[41–43]. Moreover, EZH2 has both reader and writer roles, which means that EZH2 is activated by reading its own catalytic product H3K27me3, thereby helping to further propagate the H3K27me3 mark[44]. The functional effects of H3K27me1 are less well studied, although some reports show that H3K27me1 might play a role in transcriptional activation [45–47]. H3K27me2 is widely distributed in the genome and plays a role in silencing non-cell-type-specific enhancers[43], while H3K27me3 at gene promoters acts to repress transcriptional activity.

H3K27 can be acetylated by the acetyltransferases CREBBP (CBP) and EP300 (p300)[46,48,49]. H3K27ac is a hallmark of active regulatory elements including promoter regions and enhancers[50]. H3K27ac neutralizes the positive lysine charge of histones and loses its interaction with the negatively charged phosphate groups of DNA, so enabling an open chromatin conformation. The relation between H3K27me3 and H3K27ac and their switch is well studied. A

loss/reduction of PRC2 activity results in diminished K27me3, increased acetylation of H3K27, and consequent activation of adjacent genes[51].

Similar to H3K27, H3K36 also plays an important role in the regulation of gene expression. H3K36 can also be mono-, di-, and trimethylated (H3K36me1, H3K36me2 and H3K36me3). H3K36me1 might function as a substrate for fast production of H3K36me2. H3K36me3 is found within the gene body of actively transcribed genes, while H3K36me2 spreads out in large intragenic regions, including those associated with active regulatory regions[52–54]. Also, H3K36me2 establishes repressive DNA methylation by recruiting DNMT3A to intergenic regions[54,55], while H3K36me3 recruits DNMT3B to actively transcribed gene bodies[56].

The methylation of H3K36 is highly regulated. Mammalian tissue contains several H3K36 methyltransferases, such as the family of NSD proteins (NSD1, NSD2, NSD3), ASH1L, SETMAR, and SETD2 (SET domain containing 2), that all perform mono- and di-methylation of H3K36 (H3K36me1/2). SETD2 and SETD5 (SET domain 5) are the only H3K36 tri-methyltransferases[55,57]. Moreover, in mammals, there are two families of HDMs for H3K36; KDM2A/2B which specifically demethylate K36me1/2, and KDM4A/4B/4C which demethylate H3K36me3. In addition, KDM4A/4B/4C demethylate H3K9me2/me3 as well. The reader proteins of H3K36me usually have either Tudor, PWWP, or chromodomain domains, and some epigenetic writers can also act as readers. Readers of H3K36me3 with a PWWP domain include DNMT3A and DNMT3B that also directly methylate DNA, MSH6 and LEDGF that are involved in DNA repair, and ZMYND11, a specific reader of H3.3K36me3[58]. ZMYND11 is a regulator of RNA polymerase II (Pol II) elongation and a putative tumor suppressor[59]. Intriguingly, the p.S31 residue that is specific to H3.3 is required by ZMYND11 for the recognition of H3K36me3[59,60].

It is important to note that there is mutually exclusive cross talk between the different H3 PTMS. For example, regions of the genome marked by the active marks H3K4me3 and H3K36me2/3 are generally depleted of the repressive mark H3K27me3[61]. H3K36me2/3 has an antagonistic effect on PRC2-mediated H3K27 methylation and mediates balance between activating and repressive chromatin states[62].

1.4. Deregulation of H3K27/H3K36 methylation in human diseases

1.4.1. Deregulation of H3K27 and H3K36 methylation in neurodevelopmental syndromes

Appropriate regulation of the different levels of H3K27 and H3K36 modification is very important in development as germline mutations in the writers, erasers and readers of H3K27 and H3K36 methylation have been implicated in several neurodevelopmental syndromes, which often phenocopy each other. I will briefly discuss these mutations, their clinical outcomes and the related animal models.

Loss-of-function mutations in *EZH2*, the catalytic methyltransferase component of PRC2, result in a reduction of H3K27me3 and a consequential activation of several genes that are normally suppressed. Germline mutations in *EZH2* cause Weaver syndrome (WVS), an autosomal dominant disorder. WVS is characterized by overgrowth and intellectual disability. In mice, EZH2-null mutations cause abnormal neurogenesis and cerebellar hypoplasia during embryonic development[63,64].

Germline heterozygous mutations in the *NSD1* methyltransferase gene result in *NSD1* haploinsufficiency, reducing both H3K36me1/me2 and associated H3K36me3 levels[65]. These germline mutations cause Sotos syndrome, an autosomal dominant disorder which is characterized by prenatal childhood overgrowth, craniofacial dysmorphic features, macrocephaly and learning difficulties[66]. In contrast, duplication of the genomic region of *NSD1* has an opposite phenotype compared to the Sotos syndrome, with a small birth size, microcephaly, brachydactyly, a delayed bone age and mild to moderate intellectual disability. This indicates the importance of proper *NSD1* expression during growth and brain development[67]. Mutations in the methyltransferase *NSD2* also reduce H3K36me1/2/3 levels and *NSD2* is the gene associated with Wolf-Hirschhorn syndrome (WHS). Key features of this syndrome include severe delayed growth and mental retardation, low muscle tone (hypotonia), seizures and different midline defects such as craniofacial anomalies and congenital heart defects[68]. Mice with *NSD2* gene deletions showed growth-retardation, craniofacial, congenital cardiovascular anomalies and ocular anomalies. In addition, these mice showed other phenotypes such as cerebellar hypoplasia as well as a shortened cerebral cortex. However, these mice did not show any learning abnormalities, in contrast to mice

with NSD1 mutations[69,70]. Therefore, although NSD1 and NSD2 are presumed to have similar functions and both are important for development, they clearly have a different role in the developmental process. Complete KO mouse models of *NSD1* and *NSD2* are both lethal in early stages of development: <E10.5 or postnatal day P10 for *Nsd1* and *Nsd2* KO mice respectively[69,71]. However, heterozygous KO of *Nsd1* or *Nsd2* in mice were viable and fertile although with variable growth rates[69,71].

The H3.3K36me3 reader ZMYND11 is implicated in a distinct developmental syndrome. It was found to be mutated in patients with autism spectrum disorders, or chromosome 10p15.3 microdeletion syndrome, with associated phenotypes such as intellectual disability, behavioral abnormalities, seizures, hypotonia etc.[72,73]. Subsequently, a *de novo* missense mutation in *ZMYND11* (p.Ser421Asn) was identified in a patient with hypotonia, seizures, dysmorphic facial features and developmental delays[72,74]. Recently, 16 additional cases with heterozygous mutations in the *ZMYND11* gene have been reported. The patients have a spectrum of phenotypes including intellectual disability, behavioral abnormalities, and some new features such as brachydactyly[75].

1.4.2. Cancer associated H3 mutations ("oncohistones")

A growing number of cancers, especially in children and young adults, are attributed to epigenetic dysfunction, resulting in oncogenesis, tumor progression and metastatic processes. Our group was one of the first to identify histone H3 mutations in pediatric brain cancers[76,77].

Whole-exome sequencing (WES) of high-grade gliomas (HGGs) in pediatric and young adult patients revealed recurrent somatic, gain-of-function mutations at specific residues in histone H3 genes leading to amino acid changes of lysine 27 (K27) to methionine (K27M) in H3.3 (encoded by *H3F3A*) or H3.1 (encoded by *HIST1H3B* or *HIST1H3C*). In addition, mutations causing amino acid changes of Glycine 34 (G34) to Arginine (H3.3G34R) and rarely valine (H3.3G34V) in H3.3 (encoded by *H3F3A*)[76–78] were found. Eighty percent of HGGs in the brain midline (pons, thalamus, cerebellum, spine), including diffuse intrinsic pontine gliomas (DIPG), carry the H3K27M mutation, while H3.3G34R/V occur in more than 30% of HGGs in the cerebral hemispheres[76–79] (Fig. 1.1).

In addition, H3.3G34 substitutions to Tryptophan (H3.3G34W) or Leucine (H3.3G34L, although sporadic) were found in 92% of giant cell tumors of the bone (GCTB)[80]. Also, H3.3 Lysine 36 to Methionine (K36M) (encoded by *H3F3B*) mutations were found in 95% of chondroblastomas [80]. Moreover, K36M mutations have been shown to occur in a subset of head and neck squamous cell cancers[81] (Fig. 1.1).

Furthermore, in two cases of undifferentiated sarcomas: H3.1K36M (encoded by *HIST1H3C*) and H3.1K36I (encoded by *HIST1H3B)* mutations were found[82]. Although rare, recent studies have identified K27-mutations in acute myeloid leukemia (AML)[83,84] and posterior-fossa ependymomas group A (PFA-EP), with mutations more often in *HIST1H3B* and *HIST1H3C* than *H3F3A*[85]. PFA-EP is a deadly brain tumor which affects infants and young children (Fig. 1.1).

In pediatric HGGs, other mutations overlap with histone K27M or G34R/V mutations, depending on the mutated histone variant, the age of the patient and the location of the tumor in the brain. Among these are loss of function mutations in *TP53* and/or other related cell-cycle checkpoints genes such as *PPM1D*- protein phosphatase 1D and *CHEK2*- checkpoint kinase 2[76,79,86–89]. Also, loss of function mutations in chromatin remodeller *ATRX* and rarely in *DAXX* are found[76]. Moreover, mutually exclusive gain-of-function mutations or genetic amplifications of growth factor receptors involved in brain development, such as *ACVR1*-activin A receptor, type I, *FGFR1*- fibroblast growth factor receptor 1 and *PDGFRA*- platelet-derived growth factor receptor A (receptor tyrosine kinase) co-occur with oncohistones[86–89]. Importantly, G34R/V mutations always co-occur with loss of function mutations in *ATRX* and *TP53*.

In addition, our lab identified loss-of-function mutations in *SETD2* in subsets of HGG in the cortex[90]. These mutations are mutually exclusive with H3.3G34R/V and similarly co-occur with ATRX and TP53 alterations[86,90,91]. Recent data indicates that G34 mutations inhibit SETD2 activity[39,40], which might play a role in the oncogenic effect of G34 mutations. Moreover, in a rare parietal young adult HGG tumor without H3.3G34R/V, SETD2 or IDH1/2 mutations, our group found mutations in *ZMYND11/BS69* co-occurring with *ATRX* and *TP53*[92]. Importantly, the ability of ZMYND11 to recognize H3.3K36me3 is hindered by H3.3G34R/V[59], suggesting that there may

be convergent mechanisms implicating SETD2 and ZMYND11 in the oncogenicity of G34 mutants.

Recently, several *de novo* germline missense mutations in *H3F3A* or *H3F3B* were reported, such as *H3F3B* G34V and *H3F3A* K36E, with core phenotypes of progressive neurologic dysfunction and congenital anomalies, but no malignancies (E. Bhoj, U. of Pennsylvania, manuscript in revision). In addition, in a girl with secondary microcephaly (normal brain size at birth, with post-natal development of microcephaly)[93], a heterozygous *de novo* missense variant in the *H3F3A* gene was found that affect a highly conserved leucine residue at position 62 p.(Leu62Arg). This mutation was associated with a severe developmental delay, intellectual disability, growth retardation and dysmorphic features[94].

1.5. Proposed mechanisms and consequences of K27M and K36M mutations

K27M mutations occur in both the canonical H3.1 and H3.2 (25%) and non-canonical H3.3 (75%). *In vitro* studies showed that H3K27M dominantly inhibits the enzymatic activity of EZH2 of the PRC2 complex due to strong binding of the enzyme to the H3K27M-containing nucleosomes[95]. As a result, H3 K27M reduces the global levels of the H3K27me3 mark. This *in trans* effect means that nucleosomes without the K27M mutation are also hypomethylated[39,96] (Fig. 1.2).. In DIPG tumors H3K27M-mutant histones constitute only a small fraction (3–17%) of the total H3 proteins[39]. This indicates that mutation in only one out of the 32 H3 encoding alleles is sufficient to drive DIPG tumorigenesis, indicating that the K27M mutation has a dominant-negative effect[97]. Moreover, by using isogenic human tumor cells *in vitro*, our lab showed that cells with the K27M mutation are unable to spread the repressive H3K27me3 and me2 marks to achieve further differentiation, and therefore seem to be stuck in a progenitor state of unlimited self-renewal capacity[98].

In H3K36M tumors, the mutant dominantly inhibits the H3K36 methyltransferases NSD1/NSD2 and SETD2, resulting in a reduction of the abundant H3K36me2/3 mark and a genome wide gain of the repressive H3K27me3 mark. This in turn redistributes the PRC1 complex

from repressed genic loci and results in aberrant gene expression (Fig. 1.2). It leads to a subsequent blockade of cellular differentiation, as discussed in detail below[82].

In cells with either the K36M and K27M mutation, the redistribution of epigenetic marks involved in 'crosstalk' of H3K27 and H3K36 methylation, may participate in oncogenesis[54]. Furthermore, our lab has shown that such globally epigenome reprogramming, with loss of the H3K27me3 mark and gain of the H3K27ac mark in H3K27M containing gliomas, activates the expression of repetitive elements in the DNA such as endogenous retroviral elements (ERVs). Treating these tumors with DNA demethylation and histone deacetylase inhibitors generates exquisite vulnerabilities in H3K27M mutant tumors and primes cells to produce an innate immune response[99].

1.6. The role of G34 mutations in cancer

As mentioned before, histone H3.3G34 mutations occur in two different types of tumors (HGGs and GCTB). The following is a description of the role of each G34 mutation in the respective tumors, their cell of origin, as well as their known molecular mechanisms.

1.6.1. H3.3 G34 mutations in pediatric HGGs

G34R/V mutations are found in 36% of HGGs and are specific to the *H3F3A* gene. G34R/V mutations have only been found in cerebral hemispheric tumors, mainly in temporo-parietal regions, and predominantly in young adults[78,79]. G34R/V mutations always coexist with loss of function mutations in the chromatin remodeler gene *ATRX* and *TP53*[76,100]. In pediatric HGGs nearly all *ATRX* mutations occur either within the C-terminal helicase domain or lead to truncation of the protein upstream of this domain. Pediatric HGGs with loss of function *ATRX* mutations exhibit alternative lengthening of telomeres (ALT) phenotype, which depends on homologous recombination between telomeric DNA sequences [101].

Pediatric HGGs with the G34R/V mutations have a distinct gene expression profile compared to *H3F3A*/*IDH1* WT pediatric HGGs, or pediatric HGGs with IDH1 or K27M mutations. However, no significant differences were found between G34R and G34V mutations.

Bjerke et al. showed that G34 tumors were enriched for gene signatures of early embryonic brain development and especially of the ganglionic eminences, which are a source of neuroglial progenitors. This could indicate that G34 mutations occur in a specific cell-of-origin context, and at a specific time of tumor initiation[102]. Furthermore, G34-mutant tumors show global DNA hypomethylation patterns especially in non-promoter and subtelomeric regions, while a few loci are hypermethylated. For example, the *OLIG1* and *OLIG2* genes (oligodendrocyte transcription factor) are hypermethylated, resulting in low levels of *OLIG1/2* expression in these tumors[78,100]. Lastly, two recent studies on cerebral hemispheric of HGGs which are positive for the G34R/V mutations revealed a morphology with dual neuronal and glial compartments, which is consistent with an earlier used primitive neuroectodermal tumor (PNET)-like classification. The glial compartments were positive for GFAP (an astrocyte marker), and negative for the Olig2 marker as well as ATRX protein[103,104].

Moreover, there are rare individuals with *de novo* germline ZMYND11[74,105], or H3.3G34V and Lysine 36 (K36) to Glutamate (H3.3K36E) mutations (E. Bhoj, U. of Pennsylvania, manuscript in revision). Patients with these germline mutations show varying degrees of autism spectrum disorder, neurodevelopmental defects and often microcephaly. This could indicate that G34 mutations and specific K36 readers, such as ZMYND11, act via convergent mechanisms on the epigenome.

1.6.1.1. *In vitro* and *in vivo* efforts to model H3.3G34R/V mutations and identify the HGG cell of origin

To understand cancer development, it is crucial to identify the original cell that gives rise to tumors and identify specific cell types that can form a tumor. Alcantara Llaguno et al. found that a combined inactivation of *Nf1* (tumor suppressor) and *Trp53* leads to increased Ras mitogenic signaling and enhanced cell proliferation due to genomic instability in the subventricular zone (SVZ) of adult mice, which would be sufficient to induce HGG formation. In other words, specifically targeting Nestin expressing neural stem/progenitor cells with these oncogenic drivers is sufficient to induce tumor formation, suggesting that that neural stem/progenitor cells are

candidates as cancer initiating cells[106]. Friedmann-Morvinski and colleagues generated a novel mouse model for HGGs by using an oncogenic lentiviral system with two shRNAs: one targeting *Nf1* and the other one targeting *Trp53*. They showed that GBM can originate from a variety of cells in the brain, including terminally differentiated cortical astrocytes and neurons[107]. They proposed that oncogenic induced dedifferentiation and reprogramming of neurons or glia cells to neural progenitor stem cells can give rise to the heterogeneous populations observed in malignant gliomas. Intriguingly, as the tumor progresses, the differentiation marks decrease, and stem/progenitor markers become predominantly expressed. Therefore, differentiated cortical astrocytes and neurons can be also added to the list of candidate cells-of-origin for HGGs[107].

Recurrent K27M and G34R/V mutations in pHGGs and their ability to induce genome wide changes in global expression and DNA methylation profiles likely points to their nature as oncogenic drivers, and are considered the pediatric counterparts to IDH1 mutations identified in adult HGGs[76,78,108]. IDH1-mutant HGGs also promote epigenetic dysfunction through the 'oncometabolite' 2HG (2-hydroxyglutarate) which allosterically inhibits lysine demethylases targeting H3- K9, K27 and -K36 methylation[109–111]. Therefore, to have a better insight into the role of the H3 variants K27M and G34V/R in the development of HGGs, it is critical to identify their mechanism of action, their cellular specificity and the developmental timeline in which they promote tumor formation.

The generation of specific mouse models with the oncohistones mutations will help clarify the contributions of histone H3 mutants to gliomagenesis, especially in combination with genetic alterations that are considered oncogenic by themselves (e.g. loss of TP53 or PDGFRA gains or mutations). These models are also important for future studies, because they can provide invaluable screening tools for subgroup-specific therapeutic designs. These mice will therefore fill a critical gap absolutely required for pre-clinical validation of therapeutic targeting of these mutations in this group of deadly pediatric cancers.

The generation of mouse models with histone mutations is fraught with difficulties. Mouse embryos expressing the H3.3K27R mutation exhibited a reduced development rate before the blastocyst stage[112]. This was confirmed by us, where we found that direct knock-in (DKI)

H3.3K27M into mouse embryonic stem cells using zinc finger nuclease technology resulted in a severe embryonic phenotype such that embryos did not go beyond the four-cell stage[113].

Importantly, in the context of HGG, many studies have shown that the expression of oncohistones alone is not sufficient to form tumors in mice. Previous studies showed that expression of H3.3K27M combined with p53 loss (*Trp53*−/−) in mouse neural stem cells (NSCs) when injected into mice pons did not generate tumors[99,114]. In addition, in human embryonic stem cells-derived-NSCs, p53 loss and PDGFRA amplification in conjunction with expression of H3.3K27M did show induction of neoplastic transformation, however, this effect was not seen in cells with H3.3G34R/V mutations[115]. Also, expression of K27M did not affect the proliferation rates in other cell types such as undifferentiated HES cells or differentiated somatic cells[115]. In addition, injection of NSCs co-expressing platelet-derived growth factor subunit B (PDGFB) and H3K27M into mice developed tumor faster than NSCs co-expressing PDGFB and H3 WT[114].

Recently, our group has shown that mice with the H3.3K27M mutation and p53 loss develop tumors faster and are more aggressively with a gain of *Pdgfra*. However, no tumors formed with the H3.3G34R mutation in combination with *p53* loss [113]. Lastly, Larson et al. showed that induction of a conditional knock-in mice with H3.3K27M, p53 loss and activation Pdgfra at P0-P1 accelerated tumor formation in the brainstem[116].

Altogether, efforts to introduce oncohistones into specific lineages and at different time points emphasizes the dependence on specific cell types and developmental time windows for the tumor formation. It is consistent with observations in patients, where specific oncohistones cause tumors at specific age of the patients and in specific brain region (e.g. the K27M mutation is specifically found in young children in the pons/thalamus).

1.6.2. H3.3 G34 mutations in GCTB

Giant cell tumor of bone (GCTB) are locally aggressive osteolytic tumors. GCTB accounts for 5% of all primary bone tumors and 20% of all benign bone tumors[117]. It is usually found in

young adults between 20-40 years old (about 65% of the cases). GCTB typically occurs in long bones, meta epiphyseal and eccentrical (distal femur, proximal tibia, and distal radius). It can also arise in small bones (of hand, foot, spine, sacrum, and pelvis) but is uncommon in the immature skeleton. 2–5% of GCTB metastasize, usually to the lung[118].

GCTB are characterized (>90% cases) by *H3F3A* G34W mutations and less frequently by G34L mutations[80]. These mutations are restricted to the mononuclear stromal cell population and are not found in multinucleated giant cells of GCTB[80]. In contrast to pediatric HGGs, the degree of co-occurrence of ATRX/TP53 mutations with H3.3G34W/L mutations is unknown in GCTB cases, probably because to date only a few GCTB cases have been analyzed for genetic alterations other than *H3F3A* mutations[119,120].

1.6.2.1. Cell identities of GCTB

GCTB contain three different histological cell types: multinucleated, osteoclast-like giant cells (40-100 nuclei) round, macrophage-like monocytic cells and ovid to spindle-shaped, fibroblast-like undifferentiated mononuclear stromal cells[121,122]. Mononuclear stromal cells, the presumed neoplastic cell-of-origin in GCTB, are ill-defined in the literature and vaguely referred to as mesenchymal stem cells, fibroblasts, or pre-osteoblasts based on *in vitro* differentiation assays with few criteria to reliably distinguish them. Joyner and James[123,124] showed that GCTB derived stromal cells can form bone *in vivo* and mineralized nodules *in vitro* in a passage dependent manner, while Wulling et al. showed that GCTB derived stromal cells express MSC markers and can differentiate into osteoblasts, chondroblasts and adipocytes *in vitro*[125].

Mononuclear stromal cells are derived from bone marrow mesenchymal stem cells (MSCs), which originate from the mesodermal embryonic layer and express markers such as FGFR (fibroblast growth factor receptor) that may play a role in osteoblastic differentiation through increasing downstream *SPP1*-osteopontin, *BGLAP*-osteocalcin and *ALPL*-alkaline phosphatase activity. However, osteoclast-like giant cells (CD68+, CD14+, markers of macrophage/ monocyte lineage) are derived from hematopoietic stem cells (HSCs). They express

markers of bone resorption such as TRAP (tartrate-resistant acid phosphatase), *CTSK*-cathepsin K, *VTN*-vitronectin and calcitonin receptors[118].

Furthermore, stromal cells highly express RANKL (receptor activator for nuclear factor-kB ligand), a soluble factor which interacts with its receptor, RANK, expressed by osteoclast precursors and monocytes. This interaction leads to osteoclast differentiation facilitated by cytokines and macrophage colony stimulating factor (M-CSF). The stromal cells also express and secrete OPG (osteoprotegerin) which blocks osteoblast/osteoclast interactions as it is a competitive RANK antagonist. In summary, the balance between osteoblast and osteoclast is regulated by the RANK/RANKL/OPG pathway[126,127] and in GCTB, the formation of giant multinucleated osteoclasts which do not harbour G34W is required for pathogenesis and is thought to be mediated by G34W-dependent secretion of factors such as RANKL[128].

1.7. Summary and evidence for mechanism of G34 mutants

The downstream consequences of G34 mutations and their specific roles in tumorigenesis are not clear. In contrast to modifications in H3K27 or H3K36, the G34 amino acid residue does not carry an epigenetic marker, hence mutations in this amino acid do not seem to directly change the epigenetics of H3. It is however plausible that the effects of G34 mutations are driven through modifications of the methylation status of the nearby K36 amino acid residue due to steric hindrance[129,130]. This effect on H3K36 methylation could play a role in tumorigenesis and tumor progression in brain and or other tissues, because H3K36 methylation is highly controlled in mammals. A similar effect as G34 mutations is found in missense or truncating mutations in SETD2 (a H3K36 tri-methyltransferase), occurring in 15% of pediatric and young adult HGGs of the cerebral hemisphere[90]. Also, in a cohort of 140 GCTBs, we found 1 GCTB with a germline SETD2 mutation, but is WT for all histones. All findings suggest a convergence in mechanism between SETD2 loss and G34 mutations.

H3.3 G34R/V mutations are found to be mutually exclusive with SETD2 and K27M mutations[90]. Furthermore, G34 mutations act only on the same histone tail (*in cis*), in contrast to K27M and K36M mutations that are dominant negative and promote global loss of K27me3 or

K36me2/3 on additional histone tails as well[39,82]. This also suggests that H3.3 G34 could influence the binding of writer or reader proteins of K36. This possibility has been strengthened by crystal structures studies of SETD2/KDM2A that shows that bulky amino acids at G34 (even alanine) impede access to H3K36 methyltransferases/demethylases in a narrow tunnel in the SETD2 and KDM2A complexes and hence may impede the K36 methylation[129,130]. This is further established by the fact that ectopic expression of G34 mutants in two different cell models have shown loss of H3K36me2/3 and an increase of the H3K27me3 histone mark *in cis*. H3K27me3 is a secondary change resulting from changes in K36 methylation as Shi et al. (2018) showed, while G34 mutants abrogate SETD2 activity (and so reduce H3K36me2/3 levels), do not have any effect on EZH2 enzymatic activity (which controls H3K27me3)[40]. Moreover, H3K27me3 and H3K36me3 are known to be antagonistic marks. Recently it has been shown that EZH2 has a specific sensing pocket near its catalytic site that recognizes H3K36 modifications, explaining PRC2's preference for methylating H3K27 in the presence of unmethylated K36 nucleosome substrates[131].Therefore, the absence of H3K36me3 on G34-mutant histones may make these nucleosomes a better substrate for the PRC2 complex (Fig. 1.2).. Therefore, these observations could indicate that G34 mutations might affect tumorigenesis via modulation of the epigenetic modification of K36 (primary) and K27 (secondary). However, contrasting data from Voon et al. suggests that the major effect of G34 is by inhibiting the KDM4 family, which results in increased H3K36me3 and H3K9me3 levels at select loci[132]. Comparison of H3K36me3 distributions between a single G34V cell line to a WT pHGG cell line by ChIP-seq substantiates this finding by showing increased H3K36me3 levels and RNA Pol II occupancy on several key oncogenes including MYCN, which has potential to cause glioblastoma when expressed in the correct developmental context [102].

Most of the above observations therefore suggest a strong interaction between the H3.3 G34 mutations and K36 methylation, although the molecular basis of each G34 mutation and its role in specific tumor type specificity remains unclear and therefore needs further study.

1.8. Effects of histone K27 and K36 modification and mutations on mesenchymal stem cells (MSCs) biology

As my thesis comprises the study of the H3.3 G34W mutation in a mesenchymal tumor type, in the following section, I will highlight some of the studies discussing the deregulation of K27 and K36 methylation in MSC related lineages.

MSCs originate from the mesodermal layer in an embryo during its development and consists of multipotent cells (CD105+, CD90+, CD73+ and CD34-, CD31-, CD45-). MSCs were initially isolated from bone marrow (BM-MSC), but are now isolated from mesenchymal tissues such as skeletal muscle tissue, placenta tissue, adipose tissue, liver tissue etc. MSCs are able to differentiate into lineages, such as osteocytes, chondrocytes, adipocytes and myocytes both *in vivo* and *in vitro*[133,134]. This differentiation is guided by dynamic modifications of histones in the promoter regions of master transcription factors for osteogenesis such as RUNX2 and OSX,; for adipogenesis such as PPARG and CEBPA; and for chondrogenesis such as SOX9; All are well studied[135,136]. ChIP-seq analyses showed that the landscape of histone modifications in MSCs that are differentiating into osteoblasts are more similar to undifferentiated MSCs than the histone modifications controlling the adipogenic and chondrogenic differentiated MSCs[135,137].

Lu et al. showed that H3K36M and H3K36I (isoleucine) mutations impair the differentiation of murine MSCs into chondrocytes and so cause undifferentiated sarcomas, while H3.3K27M or H3.3G34W/L did not have such an effect on chondrocyte differentiation. H3.3K36M also blocks the differentiation of murine MSC to adipocytes and osteocytes[82]. Lu et al. also observed a redistribution of H3K27me3 from genic to intergenic regions which seems to facilitate the expression of genes blocking mesenchymal differentiation[82]. In addition, Wang et al. showed that conditional Knockout (KO) of *SETD2* in mouse BM-MSC preferentially causes adipogenesis rather than osteogenesis[138]. This in contrast to the KO of NSD2 that seems to hinder adipogenic differentiation of MSCs, likely by inhibiting PPARγ expression[139]. Furthermore, Zhuang et al. also showed that ectopic expression of H3K36M in both brown and white preadipocytes inhibits adipogenesis by inhibiting PPARγ target gene expression *in vitro,* and adipogenesis and muscle development in mice[139].

KDM6A/B (H3K27me3 demethylases) and EZH2 act differently in adipogenesis and osteogenesis. KDM6A and EZH2 both regulate H3K27me3 on the promoters of both osteogenic and adipogenic genes, including Runx2, Oc, Pparg and Cebpa. Furthermore, overexpression of EZH2 or knockdown (KD) of KDM6A promotes adipogenic differentiation of MSCs in vitro[140]. However, EZH2 directly inhibits adipogenesis by increasing H3K27me3 on the promoters of WNT genes in preadipocytes. Wnt/β-catenin signaling inhibits adipogenesis and EZH2 has enrichment binding sites on Wnt genes (Wnt1, Wnt6, Wnt10a and Wnt10b)[141]. In addition, KDM6A/6B have also been shown to mediate the chondrogenic differentiation of MSCs. Yapp et al. showed that an inhibitor of KDM6A/B (GSK-J4) inhibits SOX9 and COL2A1 expression (the master regulators of chondrogenesis), and impairs chondrogenic differentiation[142].

1.9. Rationale, Hypothesis, and Objectives

1.9.1. Rationale for studying H3.3 G34 mutations in this study

Human data show that different H3.3G34 mutations occur in distinct cancers (HGGs and GCTB). All previous efforts to study G34 mutations are based on exogenous overexpression systems of the mutated H3.3 histones which can vary in levels of expression or may amplify the effects of G34 mutations beyond that observed in tumors. Modeling of H3.3G34 mutations is therefore the best accomplished through expression of endogenous mutated *H3f3a* via its intrinsic regulatory mechanisms. Moreover, the effects of some G34 mutations have been characterized mainly in cell types distant from the elusive cell-of-origin or neoplastic cells of the tumors, largely because of the inaccessibility of G34R/V-mutant pHGGs and consequent paucity of G34-mutant pHGG cell line models. In addition, the impact of site-specific H3.3G34 mutations in untransformed primary cells or in mice has not been explored sufficiently. Lastly, cell-lineage specific effects of H3.3G34 mutations is expected because of the expression and functions of H3K36 methylation writers, erasers and readers is tissue-specific.

We therefore require an isogenic mouse model to better characterize the effects of G34 mutations on the epigenome, transcriptome, and developmental phenotypes. In addition, we focus on G34W GCTB as a model to study the tumorigenic effect of G34 mutations. This model will

provide insights into the mechanism of G34R pHGG as well since the molecular mechanisms of the two mutations likely share similarities, despite driving oncogenesis in different cancers.

Here, we evaluate the effects of H3.3G34-to-R/V/W histone mutations in *in vivo* and *in vitro* model systems upon complex biological processes, including tissue homeostasis and development, cellular differentiation and cancers.

1.9.2. Hypothesis

H3.3G34R and G34W mutations will each give rise to a distinct spectrum of phenotypes affecting neural and mesenchymal lineages respectively, because of their differential effects on H3K36 methylation and readers of H3K36 methylation.

1.9.3. Objective 1

Derive direct knock in (DKI) mouse models carrying histone H3.3G34R/V/W mutations to study the effects of these mutations on phenotypes of NSC and MSC related lineages during their development.

1.9.3.1. Experimental plan

- Generating mice carrying germline histone H3.3G34R/V/W mutations using the CRISPR/Cas9 system
- Evaluating behavioral phenotypes

  Rotarod performance, Open Field (OF) paradigm, CatWalk gait analysis
- Gross morphology and histopathology of major organs (full body necropsy)

  Weight, Length, Brain, Bone, Muscle
- Histopathology and immunostaining (using multiple markers) of affected tissues and target organs
- Molecular analysis of the importance of abnormal tissues

  RNA-seq

1.9.4. Objective 2

Study the effects of histone H3.3G34W mutation in the context of giant cell tumor of bone (GCTB) using patient tumors and isogenic cell lines and to study tumorigenicity and its effects on the epigenome and transcriptome.

1.9.4.1. Experimental questions

- Is G34W mutation the sole driver mutation in the development GCTB?
- Does local *in-cis* reduction of H3K36me3 on the mutant G34W histone tail contribute to epigenetic remodeling and tumorigenesis?
- What are global (*in-trans*) effects of the G34W mutation on the epigenome and transcriptome in of GCTB cells?
- What is the cell of origin of G34-mutant GCTB?
- How does G34W mutation in stromal cells enable recruitment of giant osteoclast cells?

1.9.4.2. Experimental plan

- Generating isogenic GCTB G34W edited patient-derived cell lines using CRISPR/Cas9
- *In vitro* assays

  Proliferation assay and clonogenicity assay
- *In vivo* injections

  Subcutaneous & orthotopic tibial injections; Characterizing of the PDOX tumors
- Mass spectrometry of histone PTMs
- Genomic distribution of histone marks (H3K36me2/3, H3.3K27me3,) attributable to the G34W mutation using Chip-seq
- Transcriptomic profiling using RNA-seq and scRNAseq

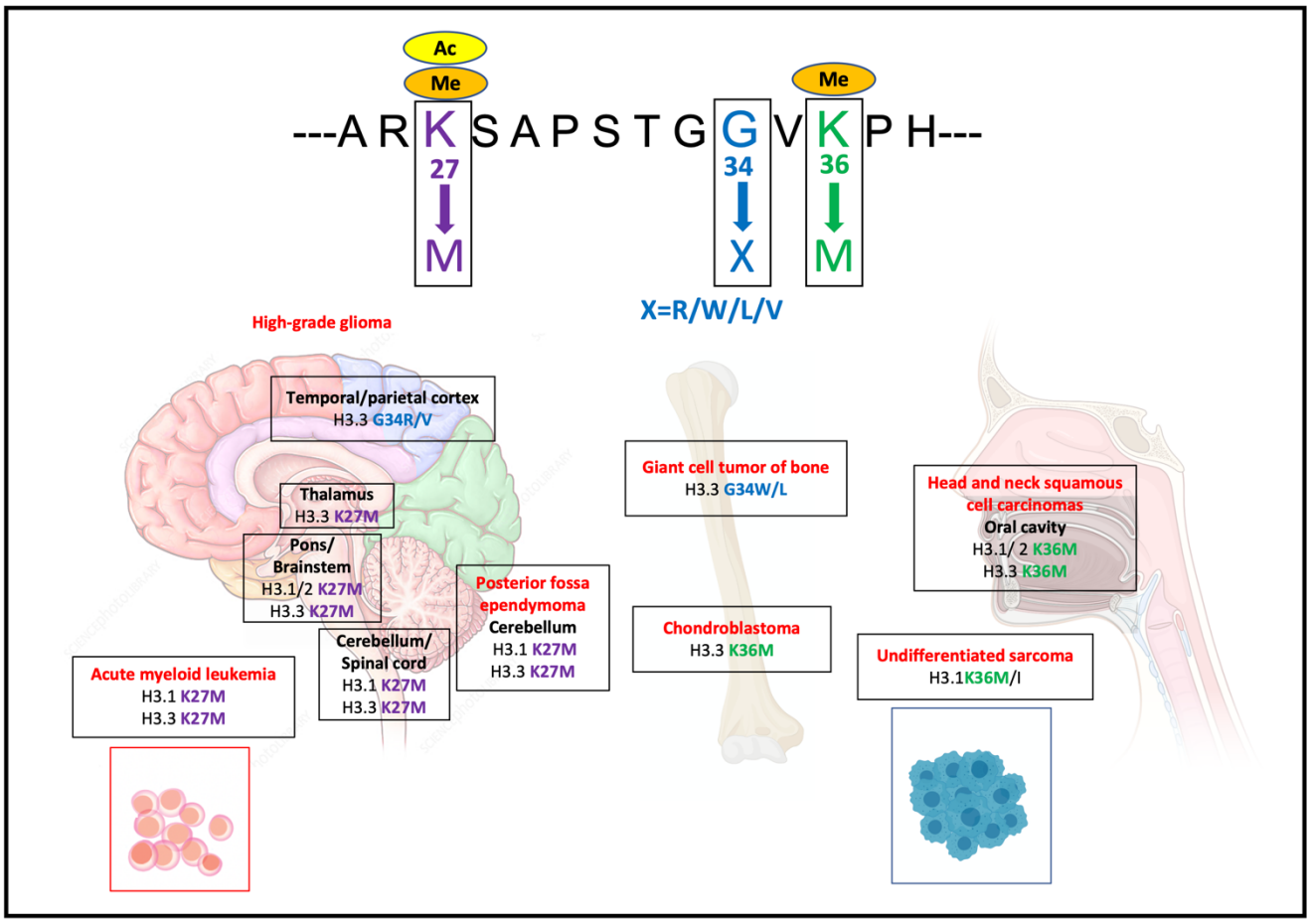


Figure 1. 1 H3 oncohistones and their association with specific cancer types.

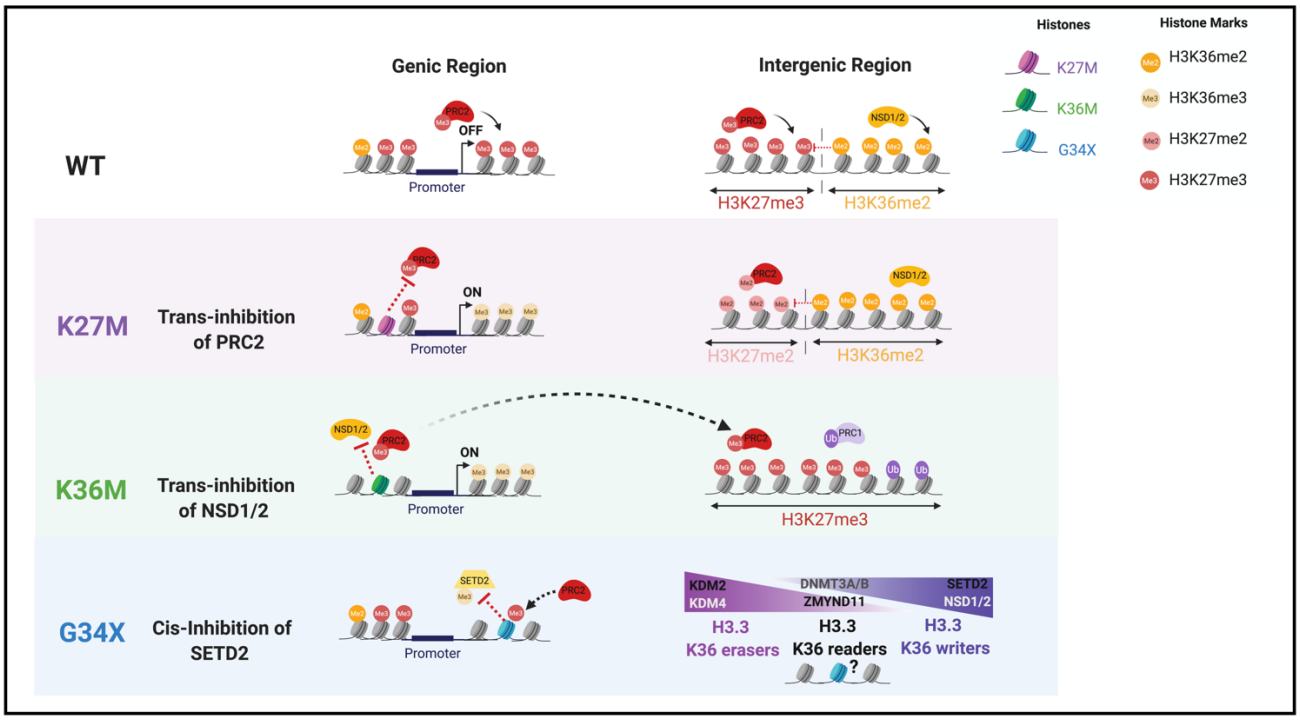


Figure 1. 2 H3 oncohistone-mediated remodeling of the chromatin landscape

Schematic representation of chromatin-remodeling mechanisms of H3K27M, H3K36M, and H3.3 G34R/V/W oncohistones at genic and intergenic regions. H3K27M & H3K36M exhibit dominant-negative effects on methyltransferases PRC2 and NSD1/2 respectively, while H3.3 G34X impede SETD2 locally at mutant histones; potentially impacting H3.3K36 readers.

Chapter II: Materials and Methods

M&M related to Chapter 3

2.1. Animals

All mice were housed, bred, and subjected to the listed procedures according to the McGill University Health Center Animal Care Committee and in compliance with the Canadian Council on Animal Care guidelines. All procedures and experiments were performed according to the guidelines of the Canadian Council of Animal Care and approved by the Animal Care Committee of the RI-MUHC. The mice were monitored weekly and euthanized at clinical endpoints when recommended by the veterinary and biological services staff members.

2.2. Generation of direct knock in (DKI)-G34 mutant mice using the CRISPR/Cas9-system

2.2.1. gRNA selection and *in vivo* validation

To generate DKI-G34 mutant mice, three gRNA sequences were designed, which are listed in Table 2.1 The gRNAs were purchased from IDT as Alt-R crRNAs along with Alt-R S.p. Cas9 Nuclease (catalog # 1081058). Each gRNA sequence was tested for cutting efficiency by microinjection (conc. 50 ng/ul gRNA:50ng/ul Cas9) into zygotes of B6C3F1 embryos. These embryos were cultured to the blastocyst stage (E3.5) and the DNA was collected for PCR, followed by T7 endonuclease I assay and sequencing as previously described[143]. The ssODN donor template included 3-point mutations near to the PAM site that when integrated at the Cas9 cleavage site prevented recognition of the target seed sequence of gRNA and cleavage by Cas9. All ssODN donor templates are listed in Table 2.1.

2.2.2. Microinjection and generation of germline mutant mice

Cas9: gRNA was simultaneously injected with ssODN (50 ng/ul:50 ng/ul :20 ng/ ul) into mouse zygotes from B6C3F1 females that were mated with male B6C3F1 mice. The embryos were transferred into pseudo-pregnant CD-1 females for gestation. The born pups were used as founder mice.

2.2.3. Genotyping germline mutant mice

Genomic DNA from 7-21 days old mice was extracted from tail snip using mouse direct PCR kit (Biomake.com). The DNA was PCR amplified with MyTaq™ HS Red Mix (Bioline), screened for potential founders using Sanger sequencing and confirmed through targeted deep sequencing using Illumina MiSeq. The sequencing reads were mapped to the wild-type locus for mutation Identification. Once established, PCR-Sanger-seq based genotyping was used. The sequences of genomic primers for PCR-Sanger-seq and MiSeq-PCR are listed in supplemental Table 2.1.

2.3. Whole mouse perfusion fixation/post-fixation

Adult mice were anesthetized and transcardially perfused with 20 ml cold PBS followed by cold 25-30 ml cold 4% paraformaldehyde (PFA). The brains were post-fixed with 4% PFA for 24-48 hr. Usually the brains were put directly in 30% sucrose/PBS for 12-36 hours. Other tissues were postfixed in 10% formalin for 48-72hr. Following fixation, the samples were dehydrated, embedded in paraffin wax and sectioned in slices with a thickness of 5µm.

2.4. Immunofluorescence staining of paraffin-embedded tissue sections

The sections were deparaffinized twice with xylene for 10min, followed by rehydration with decreasing ethanol concentrations (100% 2X each 10 min, 95% Ethanol 1X 5 min, 70% Ethanol 1X 5 mins, 50% Ethanol 1X 5 mins, 30% Ethanol 1X 5 mins and PBS wash 5min). Heat induced antigen retrieval was performed using a decloaking chamber (Biocare Medical) in antigen retrieval buffer (sodium citrate buffer 0.05% Tween 20, pH 6.0, citrate buffer 0.05% Tween 20, pH 6.0 or Tris-EDTA buffer 0.05% Tween 20, pH 9.0). After cooling, the sections were rinsed in PBS-0.03% Triton. The sections were blocked with 5-10% normal goat serum (NGS) in PBS- x Triton (1 hr, room temperature) followed by an overnight incubation at 4°C with diluted primary antibodies (for details, see Table 2.2). The following day, the sections were washed 6 times: 3 times with PBS-0.03% triton (each 5min), 3 times with PBS- x triton (each 5min) on the shaker and the sections were subsequently incubated with alexa-fluor conjugated secondary antibodies for 90 min at room temperature, followed by a final washing 2X in PBS-0.03% Triton 2min on the

shaker. The sections were mounted with ProLong™ Gold Antifade (ThermoFisher). The sections were incubated 1 hr at room temperature and stored overnight in 4°C and the results were visualized with a Zeiss LSM780 Laser Scanning Confocal Microscope and IR lasers + OPO (LSM780 IR Laser + OPO).

2.5. Immunohistochemistry (IHC)

Automated IHC on the tissue slides was performed with a Ventana Discovery Ultra. The slides were deparaffinized and rehydrated. Antigen retrieval was done using an EDTA buffer and slides were incubated with primary antibodies (for details, see Table 2.2). After washing, a secondary antibody was added (anti-Mouse or Rabbit/Mse HRP). The DAB kit chromogen was used to detect the signal.

2.6. H&E staining procedure (Autostainer, Leica)

Tissue slides were deparaffinized and rehydrated with graded ethanol to distilled water. The slides were stained with Hematoxylin, defined and incubated in a blue buffer solution, followed by staining with Eosin. The slides again were dehydrated through graded alcohols, cleared in Xylene and mounted.

2.7. Crystal Violet staining (Nissl Staining) procedure

The slides were deparaffinized and rehydrated through graded ethanol to distilled water. Slides were stained with 0.1% cresyl violet solution (3-10 min). Rinsed in distilled water. Slides were dehydrated through graded alcohols, cleared in Xylene and mounted.

2.8. Study of FAT and metabolism

2.8.1. Body weight measurements

The body weight and length (from nose to the base of tail) of the mice were measured weekly from week 1 to week 8 and then monthly until the end of the study period. At least 10 females and 10 males from all genotypes were measured for each time point.

2.8.2. Body composition, metabolic chamber and food intake

Body composition of adult male and female mice (body-, lean- and fat mass) was measured using quantitative nuclear magnetic resonance (EchoMRI system). Metabolic chambers (TSE systems) were used to assess the daily food and water intake of single housed mice. The mice were adapted to the chamber environment for 7 days prior to data collection. For each mouse, the same amount of food and water was available as ad libitum fed mice. The metabolic chamber data collection was conducted for 9 consecutive days.

2.8.3. FAT fixation and H&E staining

Mouse liver, brown and white adipose tissues were fixed in 10% neutral formalin in PBS for 24-48 hr. Following fixation, the samples were dehydrated, embedded in paraffin wax and sectioned in slices with a thickness of 5µm. Tissue slides were deparaffinized and rehydrated with graded ethanol to distilled water. The slides were stained with Hematoxylin, defined and incubated in a blue buffer solution, followed by staining with Eosin. The slides again were dehydrated through graded alcohols, cleared in Xylene and mounted.

2.8.4. Blood cell counting and blood chemistry

At the end of each experiment, and after fasting overnight (16h), the blood was collected from the heart. For blood biochemistry, serum was isolated at 2000g at 4 °C for 10 minutes and stored at -20°C until analysed. Serum cholesterol, triglycerides, high-density lipoprotein-cholesterol (HDL-C) and glucose were measured in the Comparative Medicine and Animal Resources Centre (CMARC), McGill University, Montreal, Canada. Fasting serum insulin and leptin concentrations were determined by ultra-sensitive mouse insulin (Crystal Chem 90080) and mouse leptin (Crystal Chem 90030) ELISA kits according to the manufacturer protocol.

2.9. Behavioural tests

2.9.1. Hindlimb clasping

For the Hindlimb clasping test, the mice were lifted by their tails and the position of their hind limbs was observed and scored for ten seconds. Unaffected mice consistently splayed their hindlimbs outward away from the abdomen, and this behaviour would be scored 0. A score of 1 was given to mice that pulled one hindlimb hind legs partially toward their abdomen for more than 5s. When this was observed for both hind limbs, the mouse received a score of 2. A score of 3 was assigned to mice that retracted both hind limbs completely for more than 50% of the observation time[144].

2.9.2. Rotarod test

Briefly, mice were trained to stay on a steady rod during an adaptation period of 30s. Thereafter, the rod started rotating with an acceleration of 4–40 rounds per minute for 300s. The trial ended when the mice fell off the rod or were able to keep standing for 300s. If a mouse clings on the rod and completes a full passive rotation, this is considered a failure in performance similar to falling. The tests were performed for four consecutive days and on each day three test trials were performed.

2.9.3. Open field test

Briefly, the animals were placed in an empty arena, which was novel for the mice, and allowed them to explore this space for five minutes. The movements of the mice were automatically tracked by video tracking to analyze their behavior. All the behaviors were analyzed using the Any-Maze software (Stoelting Co., IL, U.S.A.)

2.9.4. Gait analysis

We used CatWalk XT gait analysis system (Noldus, The Netherlands) as previously described[145]. Briefly, each mouse walked on an illuminated glass plate of the CatWalk walkway in an unforced manner. The position of each paw was measured and analysed using the Noldus CatWalk XT 8.1 software system and at least three complete runs for each mouse were recorded.

M&M related to Chapter 4

2.10. Data and software availability

Raw and processed sequencing data for RNA-Seq, scRNA-Seq and ChIP-Seq data was deposited into the Gene Expression Omnibus (GEO) under the accession code.

To review GSE149211, go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149211 and enter token yxuxmwcsxzextij

2.11. Establishing primary cell lines from GCTB stromal cells

Fresh GCT tumor specimens were collected in DMEM medium. The tissue was washed twice in PBS, minced with surgical blades and subsequently dissociated in Collagenase-Dispase (Sigma-Aldrich) at 37ºC in 5% CO2 for 30 min. The cell pellet was treated with NH_4Cl to remove red blood cells, washed twice in DMEM medium (4.5 g/L glucose, L-glutamine, phenol red; Wisent Bioproducts, Montreal) containing 10% FBS and penicillin/streptomycin, and plated in T-75 flasks in the same medium. When the cells reached 75% confluence, they were trypsinized and passed in culture using the same medium. DNA was extracted from various passages and assessed for *H3F3A* G34W mutation status and allele frequency by droplet digital PCR (Biorad). We thank Dr. David Allis for generously sharing a primary GCTB cell line (GCT-2611). All lines were tested monthly for mycoplasma contamination (MycoAlert Mycoplasma Detection kit by Lonza) STR fingerprinting was regularly performed to confirm the identity of the cells.

2.12. hTERT immortalization

The pBABE-hygro-hTERT plasmid (Addgene #1773) was a gift from Dr. Sidong Huang (McGill University). Phoenix Ampho cells were transfected with the plasmid and used to harvest the retrovirus. Primary GCTB cell lines were transduced with the hTERT retrovirus and the transduced cells were selected with 100 µg/ml of hygromycin for 4 days.

2.13. CRISPR/Cas9 gene editing

Human isogenic GCTB cell lines were generated using CRISPR/Cas9 gene editing, as previously described[1]. In short: a single-guide RNA (sgRNA, IDT) targeting the *H3F3A* G34W mutation (Table 2.1) was cloned into the plasmid pSpCas9(BB)-2A-Puro (PX459 V2.0) (Addgene #48139). A single-stranded donor oligonucleotide (ssODN, IDT) template was designed with the wild type *H3F3A* sequence (Table 2.1). The construct with the G34W-sgRNA, and the ssODN template, were transfected in the patient-derived GCTB cells using lipofectamine 3000 (Thermo Fisher Scientific) as per manufacturer's protocol. Transfected cells were placed under puromycin selection for 72h and then single-cell sorted into 96-well plates. Single-cell clones were expanded and screened for editing events at the target locus by Sanger sequencing and confirmed through targeted deep sequencing using Illumina MiSeq and G34W immunoblotting.

2.14. Proliferation assay

Cells were seeded at a density of 1,500 cells/well into a 96-well plate, and nuclei were stained with NucLight Rapid Red Cell Reagent (1:500 dilution, Essen Bioscience) as per manufacturer's instructions. Cells were imaged using the IncuCyte ZOOM System real-time instrumentation (Essen Bioscience) every 2 hours (10X magnification), and scanned images were analyzed after 120 hours with IncuCyte ZOOMTM 2015A software. Measurements were collected from triplicate wells and two independent experiments were performed. Statistical significance was assessed with Student's t-test.

2.15. Colony formation assay

Cells were seeded at a density of 250 cells/well of a 6-well plate. After 3 weeks, colonies were washed 3 times with PBS, fixed with 100% methanol for 20 min, and stained with 0.5% crystal violet for 30 min. Colonies were manually counted under an inverted microscope (colonies containing at least 50 cells were scored). This assay was independently performed 3 times.

2.16. Immunoblotting

Total histones were extracted from cells using the EpiQuick Total Histone Extraction Kit (Epigentek). Histone concentration was determined using Bradford assay reagent (Bio-Rad). 1-3 µg of total histones were prepared according to NuPAGE Bis-Tris Mini Gel Protocol (Thermo Fisher Scientific) and separated on NuPAGE Bis-Tris Gel. Wet gel electrotransfer to a PVDF membrane (GE Healthcare) was performed. Blots were blocked for 2 hours in Tris buffered saline (50 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.4) (0.1% TBS-T) containing 5% skim milk and incubated overnight at 4°C in primary antibody solutions (Table 2.2). Blots were washed 3 times in 0.1% TBS-T and incubated for 1 hour in 0.1% TBS-T containing 5% skim milk supplemented with ECL anti-rabbit IgG-conjugated HRP (1:5000 dilution; GE Healthcare). Membranes were washed 3 times in 0.1% TBS-T and the signal was resolved with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) and imaged on a ChemiDoc MP Imaging System (Bio-Rad).

2.17. Myogenic Differentiation and Immunofluorescence

Cells were seeded at a density of 6000 cells/chamber of 8-well chamber slides in standard medium. When cells reached 70% confluency, myogenic differentiation was induced by adding media containing 5ng/mL TGF-β1 (R&D systems 240-B-002/CF), 30µM L-Ascorbic acid 2-phosphate (Sigma A8960), 5ng/mL PDGF-AB (Peprotech 100-00AB). Fresh differentiation media was added twice a week for 3 weeks[146]. For immunofluorescence, cells were fixed with 4% paraformaldehyde and 15% sucrose in PBS solution for 20 minutes at 4°C. Permeabilization was performed with 0.1% Triton X-100 for 3 mins on ice, followed by incubation in blocking buffer of 5% goat serum for 1 h at room temperature. Cells were incubated overnight at 4°C with primary antibodies followed by Incubation with secondary anti-rabbit Alexa Fluor 488 or 594 antibodies (ThermoFisher Scientific) was performed at 1:1000 dilution in PBS under light protection for 90 minutes. ProLong Diamond Antifade Mountant with blue DNA DAPI stain (ThermoFisher Scientific) was applied. Fluorescent signal was captured the following day.

2.18. Histone post-translational modification quantification with nLC/MS

The workflow for histone extraction, LC/MS, and data analysis was processed as described before[147]. Briefly, cell pellets (~1×10^6 cells) were lysed on ice in nuclear isolation buffer supplemented with 0.3% NP-40 alternative. Isolated nuclei were incubated with 0.4 N H2SO4 for 3h at 4 °C with agitation. 100% trichloroacetic acid (w/v) was added to the acid extract to a final concentration of 20% and samples were incubated on ice overnight to precipitate histones. The resulting histone pellets were rinsed with ice cold acetone + 0.1% HCl and then with ice cold acetone before resuspension in water and protein estimation by Bradford assay.

Approximately 20µg of histone extract was then resuspended in 100 mM ammonium bicarbonate and derivatized with propionic anhydride. 1µg of trypsin was added and samples were incubated overnight at 37 °C. After tryptic digestion, a cocktail of isotopically labeled synthetic histone peptides was spiked in at a final concentration of 250 fmol/µg and propionic anhydride derivatization was performed for second time.

The resulting histone peptides were desalted using C18 Stage Tips, dried using a centrifugal evaporator, and reconstituted using 0.1% formic acid in preparation for nanoLC-MS analysis.nanoLC was performed using a Thermo ScientificTM Easy nLCTM 1000 equipped with a 75µm × 20 cm in-house packed column using Reprosil-Pur C18-AQ (3µm; Dr. MaischGmbH, Germany). Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in 80% acetonitrile. Peptides were resolved using a two-step linear gradient from 5 to 33% B over 45 min, then from 33 to 90% B over 10 min at a flow rate of 300nL/ min. The HPLC was coupled online to an Orbitrap Elite mass spectrometer operating in the positive mode using a Nanospray FlexTM Ion Source (Thermo Scientific) at 2.3 kV. Two full MS scans (m/z 300–1100) were acquired in the orbitrap mass analyzer with a resolution of 120,000 (at 200 m/z) every 8 DIA MS/MS events using isolation windows of 50 m/z each (e.g., 300–350, 350–400, ...,650–700).

MS/MS spectra were acquired in the ion trap operating in normal mode. Fragmentation was performed using collision-induced dissociation (CID) in the ion trap mass analyzer with a normalized collision energy of 35. AGC target and maximum injection time were 10e6 and 50ms

for the full MS scan, and 10e4 and 150ms for the MS/MS can, respectively. Raw files were analyzed using EpiProfile.

2.19. Secreted proteome -Golgi Apparatus purification and Mass Spectrometry

Cells were lysed using a 7mL Dounce stainless tissue grinder, and a 2M sucrose solution was added to homogenates to obtain final 1.7M sucrose concentration. Golgi membranes were isolated by isopycnic centrifugation using discontinuous sucrose gradients[148]and processed for MS analyses as described before[149].

2.19.1. Analysis of mass spectrometry data: Data processing

The abundances of proteins enriched in the Golgi Apparatus of isogenic lines were quantified as intensities using the PEAKS proteomics software program. The intensities were then normalized to the median intensities of the TMT ion reporters to adjust for the amount of material. We derived n=4 technical replicates for mutant lines and n=4 for edited lines from ImGCT-4072.

2.20. Animal models

All mice were housed, bred, and subjected to listed procedures according to the McGill University Health Center Animal Care Committee and in compliance with Canadian Council on Animal Care guidelines. The mice were monitored weekly and euthanized when xenograft tumor volume (subcutaneous and tibial) reached 2,000 mm^3 (maximum size) or immediately at a clinical endpoint when recommended by veterinary and biological services staff members.

2.21. Tagging cells with GFP-Luciferase

pSMAL-GFP-Luc lentiviral vector was a gift from Dr. Kolja Eppert (McGill University). 96 hours after isogenic GCTB cells were transduced with the virus, GFP-positive cells were sorted using fluorescence-activated cell sorting and expanded for use in xenograft experiments.

2.22. Mouse subcutaneous implantation

1.2-1.5 × 10^6 cells were prepared as a single-cell suspension in PBS-50% Matrigel (volume of 200 µL) and were implanted subcutaneously into the left flank of 8- to 12-week-old immunodeficient NSG mice (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ, Jackson Laboratory). Tumor formation was monitored weekly using bioluminescent imaging (see section below) and palpation starting at 4 weeks post-implantation.

2.23. Mouse orthotopic intratibial implantation

70% ethanol was used to clean the leg of 8- to 12-week-old immunodeficient NRG mice (NOD.Cg-Rag1t^{m1Mom} Il2rg^{tm1Wjl}/SzJ, Jackson Laboratory). A pre-drilled hole was created in the tibial plateau using rotating movements with a 25G needle. 3 µl of PBS-20% Matrigel containing 1.8-2 × 10^5 cells were implanted into the tibial diaphysis with a Hamilton syringe (27G needle). Tumor formation was monitored weekly using bioluminescent imaging (see section below) and tumor volume was measured with digital calipers.

2.24. Bioluminescence imaging

Bioluminescence imaging was performed as previously described[4] using an IVIS 100 system (Caliper). Briefly, anesthetized mice were placed in the imager 7 min. post-intraperitoneal injection of D-luciferin (150 mg/kg).

2.25. Immunohistochemistry (IHC)

Xenograft tumors were fixed in 10% buffered formalin for 48 hours. After fixation, the bone samples were decalcified using 10% EDTA for 3-4 weeks (solution changed weekly). This was followed by embedding in paraffin wax and sectioning at a thickness of 5µm. Automated IHC was performed with Ventana Discovery Ultra. Slides were deparaffinized and rehydrated. Antigen retrieval was done using EDTA buffer and slides were incubated with primary antibodies. After washing, the secondary antibody was added (anti-Mouse or Rabbit/Mse HRP). DAB kit chromogen was used to detect the signal.

2.26. Tartrate-resistant acid phosphatase (TRAP) staining

Slides were deparaffinized and rehydrated through graded ethanol to distilled water. Slides were placed in pre-warmed TRAP Staining Solution Mix (4% pararosaniline, 4% sodium Nitrite, 0.1 Acetate buffer pH11, Naphthol AS-TR Phosphate, Sodium Tartare, pH 5.0) and incubated at 37°C until control was developed, and then rinsed in distilled water. Slides were counterstained with 0.1% Fast Green for 1 min, air-dried and mounted. Osteoclasts stain red violet whereas background stains green using TRAP staining.

2.27. Next-generation sequencing

Nikolas De. Jay performed the RNA-seq and Chip-seq data analysis, and Liam Hendriks performed scRNA-seq data analysis. A short description of the techniques follows.

2.27.1. RNA-seq Library Preparation and Sequencing

Total RNA was extracted from cell pellets using Aurum Total RNA Mini Kit (Bio-Rad). Library preparation was performed with ribosomal RNA depletion according to instructions from the manufacturer (Epicentre) to achieve greater coverage of mRNA and other long non-coding transcripts. Paired-end sequencing was performed on Illumina HiSeq 4000 (RRID:SCR_016386) and Illumina NovaSeq 6000 S4 (RRID: SCR_016387) platforms.

2.27.2. Chromatin immunoprecipitation library preparation and sequencing

ChIP-seq was performed as previously described[150]. Briefly, cells were fixed with 1% formaldehyde (Sigma). Fixed cell preparations were washed, pelleted and stored at −80 °C. Sonication of lysed nuclei (in buffer containing 1% SDS) was performed on a BioRuptor UCD-300 for 60 cycles, 10s on, 20s off. Samples were checked for sonication efficiency using the criteria of 150–500 bp by gel electrophoresis. After sonication, chromatin was diluted to reduce SDS level to 0.1% and before ChIP reaction, 2% of sonicated Drosophila S2 cell chromatin was spiked-in the samples.

ChIP reaction for histone modifications was performed on a Diagenode SX-8G IP-Star Compact using Diagenode automated Ideal ChIP-seq Kit. Twenty-five microliter Protein A beads (Invitrogen) were washed and then incubated with antibodies that are listed in Table 2.2, and 2 million cells of sonicated cell lysate combined with protease inhibitors for 10 h, followed by 20 min wash cycle with provided wash buffers.

Reverse cross linking took place on a heat block at 65°C for 4 h. ChIP samples were then treated with 2 µl RNase Cocktail at 65°C for 30 min followed by 2 µl Proteinase K at 65°C for 30 min. Samples were then purified with QIAGEN MiniElute PCR purification kit as per manufacturers' protocol. In parallel, input samples (chromatin from about 50,000 cells) were reverse crosslinked and DNA was isolated following the same protocol.

Library preparation was carried out using Kapa HTP Illumina library preparation reagents, following manufacturer's instructions. Briefly, 25 µl of ChIP sample was incubated with 45 µl end repair mix at 20°C for 30 min followed by Ampure XP bead purification. A tailing: bead bound sample was incubated with 50 µl buffer enzyme mix at 30°C for 30 min, followed by PEG/NaCl purification. Adaptor ligation: bead-bound sample was incubated with 45 µl buffer enzyme mix and 5 µl of TruSeq DNA adapters (Illumina), for 20°C 15 min, followed by PEG/NaCl purification (twice). Library enrichment: 12 cycles of PCR amplification. Size selection was performed after PCR using a 0.6 × /0.8x ratio of Ampure XP beads (double size selection) set to collect 250–450 bp fragments. Single end sequencing was performed on the Illumina HiSeq 2500/4000 platforms.

2.27.3. Single-cell RNA-seq Library Preparation and Sequencing

Library preparation and sequencing was performed as previously described[151]. Briefly, fresh tumours collected after surgery were dissociated in Collagenase-Dispase (Sigma-Aldrich). Red blood cells were lysed by ammonium chloride treatment for 5 min on ice. Cell viability was assessed with Trypan Blue. For samples with low viability (<60%), dissociated cells were enriched for live cells using Dead Cell Removal kit (Miltenyi Biotech). 10,000 dissociated cells per sample were loaded on the 10X Genomics Chromium controller. The Chromium Single Cell 3′ (10X Genomics, Version 3) protocol was strictly followed to prepare libraries. The 10X libraries were

sequenced on Illumina HiSeq 4000 (RRID: SCR_016386) sequencing platform and Illumina NovaSeq 6000 S4 platforms (RRID: SCR_016387).

2.27.4. Quantification and statistical analysis

Description of statistical details for each experiment can be found in figure legends. Statistical significance was always adjusted for multiple testing and considered to be attained when $P<0.05$.

2.27.5. Analysis of ChIP-seq data

The ChIP-seq pipeline from C3G's GenPipes toolset (v3.1.0)[152] was used.

2.27.5.1. Genome-wide chromatin mark enrichment analysis

The abundance of each mark was quantified by counting the number of reads in genomic 10kb bins and normalized to RPKM with the "multiBamSummary bins" functionality of deeptools[153] and normalized to RPKM. Genome-wide changes in deposition patterns were assessed using principal component analysis (PCA) over RPKM-normalized read counts for all marks except H3.3, where we used DiffBind-normalized read counts consensus peaks. Differential enrichment of each mark at each bin was calculated as the $\log_2$ fold-change (LFC) between average abundance in each condition.

Bins were classified as promoter, intragenic or intergenic regions based on overlap with promoter (i.e. ±1.25kb of a transcription start site), intergenic (i.e. covered in the reference), and intergenic regions (i.e. complement of promoter and intergenic regions) using the UCSC table browser's "whole" annotation for Ensembl's ensGene reference (GRCh37; n=60234 genes). The significance of differences in proportion of bins overlapping these regions that also gain or lose a chromatin mark was determined using a chi-square test.

2.27.5.2. Chromatin mark enrichment analysis at genes

Enrichment of chromatin marks at promoter and gene bodies was obtained by counting reads over promoter and intragenic regions, respectively. Similar to differential gene expression,

differential enrichment analysis was computed using DESeq2[154] (v1.18.1), using unit-normalized library depth as size factors (i.e. total number of uniquely mapped reads).

2.27.6. Analysis of bulk RNA-seq data

Bulk RNA-seq was analyzed as previously described[99].

2.27.7. Analysis of single-cell RNA-seq (scRNA-seq) data

2.27.7.1. Quality control and normalization

Low-quality cells were excluded based on outlier mitochondrial content (indicative of cellular stress or damage) and number of genes expressed using the R package Seurat[155,156] (v3.0.0). The upper threshold on mitochondrial content varied from >15-20% of total genes detected. Since GCTB samples contain multinucleated osteoclasts, where a high number of UMI counts is a biological feature of this cell type, we applied a more lenient upper threshold of 3 times the inter-quartile range. Filtered cells were then normalized together using SCTransform (v0.2.0)[157], (with parameter variable.features.n=3000), and mitochondrial content percent was regressed.

2.27.7.2. Identification of cell types

To robustly identify cell types associated with each cell cluster, we first derived markers for each cluster through differential expression (DE) analysis using a Wilcoxon Rank Sum test configured to only test genes expressed in >25% of cells and with log-fold change $\geq$0.25. We then combined the three following approaches: (1) comparing cluster markers to known cell type markers from the literature, (2) classifying cell types from reference datasets based on correlation of gene expression profiles, and (3) pathway enrichment analysis of cluster-specific markers. We used SingleR package (v1.0.5) [158] with default parameters to assign cell types to each cluster based on their Spearman correlation to gene expression profiles of known cell types in the Human Primary Cell Atlas (HPCA) reference. Since the stromal and osteoclast cell types were not present in the reference, we performed pathway enrichment analysis on cluster-specific gene markers using G:Profiler with term size set to 1000 and deriving pathways from GO:BP only [159].

2.27.7.3. Identification of heterogeneity within stromal cells

To examine heterogeneity within stromal cell populations, stromal cells were re-clustered (i.e. in the absence of non-stromal cells), and highly variable genes recalculated. PCA revealed that PC1 was defined by *SPP1*-correlated genes, and PC2 by *ACTA2*-correlated genes. To determine if these genes, which represent major sources of variation within the data, defined specific subsets of cells, correlation analysis was performed using the correlateGenes function from R package Scran (v1.12.1)[160]. This revealed distinct gene sets differentially enriched in stromal cells, notably "*SPP1* module" and "*ACTA2* module".

Technical (i.e. sequencing batch) and biological variance (i.e. divergent tumor evolution and genetic backgrounds) can obscure similarities between neoplastic cells from different tumors. To validate that the gene modules identify distinct stromal cell subsets or states, we used the Harmony data integration method (v0.99.6)[161] to correct these sources of variation on the PCA embeddings.

2.27.7.5. Pseudotemporal ordering and trajectory analysis

To determine cell trajectory within each lineage, we employed two complementary trajectory inference methods, Slingshot (v1.3.1)[162] and velocyto.R (https://github.com/velocyto-team/velocyto.R, v0.6) [163]. To calibrate parameters for trajectory inference in stromal cells, we first analyzed a well-defined lineage, myeloid cells. To identify lineages, we applied unsupervised Slingshot analysis on a three-dimensional (3D) diffusion map embedding. A 3D diffusion map was constructed using the DiffusionMap function from the R package destiny (v2.14.0) [164], with only high-quality cells included. High quality cells were defined here as cells having < 5% mitochondrial gene content and genes with >10 counts in at least 100 cells. We then used Slingshot (with default parameters and without specifying a start cluster) to identify lineages in the diffusion map embedding. 3D visualizations were created using R package plotly (v4.9.1) [165]. Since Slingshot cannot confidently establish the start of the lineage, we employed velocyto.R as described[163].

2.27.7.6. Determining stromal cell identity

To better determine a putative cell-of-origin for stromal cells, we classified them using SingleR with custom references derived from two murine bone marrow scRNA-seq datasets Baryawno et al. 2019 [166] and Tikhonova et al. 2019 [167].

2.27.7.7. Cell to cell interaction analysis

Cell to cell interactions analysis between the various cell types in the GCTB tumor microenvironment was performed as described[168].

Table 2. 1 List of oligos

Oligo	**Sequence (5'-3')**
Mouse_*H3f3a* G34_gRNA	TAGAAATACCTGTAACGATG
Mouse_*H3f3a* G34_Sangerseq_Fwd primer Mouse_*H3f3a* G34_Sangerseq_Rev primer	CGATAGATGTAATCCGCGCC ATCCCACTGCTCGACGTT
Mouse_*H3f3a* G34_MiSeq_CS1primer Mouse_*H3f3a* G34_MiSeq_CS2primer	ACACTGACGACATGGTTCTACACGATAGATGTAATCCGCGCC TACGGTAGCAGAGACTTGGTCTATCCCACTGCTCGACGTT
ssODN donor template_G34R	CCGCAAATCCACCGGTGGTAAAGCACCCAGGAAACAACTGGC TACAAAAGCCGCTCGCAAGAGTGCGCCCTCTACTGGAAGGGT GAAGAAACCTCACCGCTATAGGTATTTCTAAAACGTCGAGCA GTGGGATAGTGTCTAAGCAGTATGTCCGTGTAATTTAACAGG AAGATAGTCATA
ssODN donor template_G34W	CCGCAAATCCACCGGTGGTAAAGCACCCAGGAAACAACTGGC TACAAAAGCCGCTCGCAAGAGTGCGCCCTCTACTGGATGGGT GAAGAAACCTCACCGCTATAGGTATTTCTAAAACGTCGAGCA GTGGGATAGTGTCTAAGCAGTATGTCCGTGTAATTTAACAGG AAGATAGTCATA
ssODN donor template_G34V	CCGCAAATCCACCGGTGGTAAAGCACCCAGGAAACAACTGGC TACAAAAGCCGCTCGCAAGAGTGCGCCCTCTACTGGAGTGGT GAAGAAACCTCACCGCTATAGGTATTTCTAAAACGTCGAGCA GTGGGATAGTGTCTAAGCAGTATGTCCGTGTAATTTAACAGG AAGATAGTCATA
H3F3A G34W sgRNA	TTCTTCACCCATCCAGTAG
H3F3A WT ssODN	ACCCAGGAAGCAACTGGCTACAAAAGCCGCTCGCAAGAGTGCG CCATCTACTGGAG GGGTGAA GAAACCTCATCGTTACAGGTATT AAAAAACAGGAAAAAAA

Table 2. 2 List of antibodies

Antibody	Company	Catalogue #	Antigen retrieval buffer	Blocking buffer	Dilution	Primary diluent buffer
Immunofluorescence Staining on the Brain Sections						
Calbindin	Millipore Sigma	C9848	Citrate buffer	5% NGS in PBS-0.3%T	1:200	PBS-0.03%T
GFAP	Cell signalling	12389S	Citrate buffer	5% NGS in PBS-0.3%T	1:200	PBS-0.03%T
NeuN	Millipore Sigma	MAB377	Citrate buffer	5% NGS in PBS-0.3%T	1:200	PBS-0.03%T
MBP	Abcam	ab62631	Citrate buffer	5% NGS in PBS-0.3%T	1:150	PBS-0.03%T
NF200	Millipore Sigma	N4142	Citrate buffer	5% NGS in PBS-0.3%T	1:100	PBS-0.03%T
Olig2	Danad-Farber	DF308	Sodium citrate buffer	5% NGS in PBS-0.1%T	1:2500	1% NGS in PBS-0.1%T
S100β	Abcam	ab52642	Tris/EDTA buffer	5% NGS in PBS-0.3%T	1:300	PBS-0.03%T
Vimentin	abcam	ab92547	Citrate buffer	3% NGS in PBS-0.03%T	1:200	3% NGS in PBS-0.03%T
Immunofluorescence Staining on the Bladder Sections						
α-SMA	Abcam	ab5694	Sodium citrate buffer	10% NGS in PBS-0.3%T	1:100	5% NGS in PBS-0.3%T
Col I	Abcam	ab21286	Sodium citrate buffer	10% NGS in PBS-0.3%T	1:250	1% NGS in PBS-0.03%T
Col III	Abcam	ab778	Sodium citrate buffer	10% NGS in PBS-0.3%T	1:100	10% NGS in PBS-0.3%T
Immunofluorescence Staining on the GCT Cells						
α-SMA	Abcam	ab5694		3% BSA in PBS	1:200	3% BSA in PBS solution
Calponin 1	Abcam	ab46794		3% BSA in PBS	1:200	1% BSA in PBS solution
Western Blot						
Total H3	Abcam	ab1791		5% skim-milk in TBS-0.1%T	1:1000	2 % BSA in PBS solution
G34R	REvMab	31-1120-00-S		5% skim-milk in TBS-0.1%T	1:500	2 % BSA in PBS solution
G34W	REvMab	31-1145-00		5% skim-milk in TBS-0.1%T	1:500	2 % BSA in PBS solution
Chip-seq						
H3K27me3	CST	9733			1:40	
H3K27ac	Diagenode	C15410196			1:80	
H3K36me3	Active Motif	61021			1:100	
H3.3	Millipore	09–838			1:66	
H3.3G34W	RevMAb	31-1145-00			1:100	
Immunohistochemical Staining on the Tissue Sections						
G34W	RevMAb	31-1145-00			1:100	
Ki67	Abcam	ab15580			1:300	
Col VI	Abcam	ab6588			1:50	
Biglycan	Abcam	ab49701			1:500	
α-SMA	Abcam	ab5694			1:200	
Osteopontin	Abcam	ab8448			1:500	
Vimentin	Cell signalling	5741			1:50	
Actin	Thermo Fisher	MA1-744			1:900	
Desmin	Abcam	ab15200			1:200	
CD34	Abcam	ab81289			1:200	
S100	Dako	Z0311			1:2000	
Sox10	Abcam				1:500	

Chapter III: Functional characterization of histone H3.3 (*H3f3a*) G34 mutations on the development in murine models

3.1. Generation of mice lines with germline *H3f3a* point mutations using the CRISPR/Cas9-system

To generate germline G34 mutated mice, we CRISPR-edit blastocyst cells of mice. For this we designed three single-guide RNAs (sgRNAs) targeting exon 2 of the mouse *H3f3a* gene (Fig. 3.1A, Table 2.1). We tested the sgRNAs' cutting efficiency by microinjecting the sgRNAs into the pronucleus of 1-cell stage BC3F (a hybrid of C3H, C57BL/6 mice) mouse strain embryos, followed by T7 endonuclease I analysis (Fig. 3.1B). The sgRNA with the highest cutting efficiency (Fig. 3.1.B) was co-injected together with Cas9 protein and a repair template (single-stranded donor oligonucleotide-ssODN) (Fig. 3.1A, Table 2.1) into the pronucleus of BC3F single cell zygotes that were then implanted into recipient mothers. Genotyping of tail tissue by targeted deep sequencing (Illumina MiSeq) revealed that founder animals carried various genotypes (WT, indels, or H3.3G34 point mutations) on different *H3f3a* alleles and were denoted mosaic founders (F0) (Fig. 3.1C).

To establish mouse lines with the different H3.3G34 mutations, we crossed all mosaic founder animals with histone H3.3 wild-type (WT) mice (two strains: C57BL/6J and B6C3F1/J), and selected 4 different true heterozygous founders (g1), namely: *H3f3a*+/- (het_KO), *H3f3a*+/G34R, *H3f3a*+/G34V and *H3f3a*+/G34W (Fig. 3.1D). To further confirm the presence of heterozygous point mutations in codon G34 in g1 founders, we used PCR to amplify the region flanking the site mutations in exon 2 of murine *H3f3a* genomic DNA followed by Sanger sequencing (Fig. 3.1E). In contrast to wild-type littermates, the heterozygous G34R mutant mice showed expression of the mutant H3.3G34R histone in all collected tissues, confirmed by immunoblotting with a highly specific antibody for H3.3G34R (Fig. 3.1F).

During crosses, we found that some female *H3f3a*+/G34 mutant mice had difficulty during parturition and exhibited dystocia (*H3f3a*+/G34R: 5/16, *H3f3a*+/G34W: 8/33, *H3f3a*+/G34V: 3/39), a rare phenotype in mice. Normal parturition in mice happens at 19 days post-coitum (d.p.c.), but mice experiencing dystocia either produce no births (≥20 d.p.c.) or have incomplete delivery with embryos stuck in the canal or utero[169]. *H3f3a*+/G34R mutant females also underwent abortions at different embryonic stages. In addition, *H3f3a*+/G34R male mice displayed a degree

of subfertility, in which mutant males were not able to copulate with females, as judged by the absence of copulatory plug or pregnancy.

When born alive, mice with germline heterozygous *H3f3a* mutations (*H3f3a*+/-, *H3f3a*+/G34R, *H3f3a*+/G34V and *H3f3a*+/G34W) seemed healthy, suggesting normal embryonic development. This is in contrast to the heterozygous *H3f3a*+/K27M mutation, which produced zygotes that were not viable past the four-cell stage[113].

3.2. *H3f3a*+/G34 mutations negatively affect growth during early postnatal development, but mutant mice acquire weight as adults

For both sexes, the mutant *H3f3a*+/G34R/V/W mice were born smaller compared to the WT and het_KO control mice. We measured the body lengths and weights of all *H3f3a* mutants and WT mice of both sexes in two different backgrounds (C57BL/6 and B6C3F1) and at different time points from birth (D0) up to mature adult ages (Fig. 3.2A-C). Their body weights and lengths were significantly smaller until 7-8 weeks (Fig. 3.2A-C). After that time, the *H3f3a*+/G34W mice showed significant weight gain compared to the WT, het_KO and G34R/V mutant mice. Around 3 months of age, the G34R/V mutant mice also gained weight to reach a similar weight as control mice (Fig. 3.2A). A similar effect was noticed with the body length. (Fig. 3.2B). These data suggest that G34 mutations negatively affect the growth of these mice during embryonic stage and early postnatal development.

The faster weight gain of G34W mice, followed by G34R/V mice compared with control mice, could indicate that G34 mutations play a role in metabolic regulation through the development of an obesity phenotype.

3.3. *H3f3a*+/G34 mutant mice have an obesity phenotype

To evaluate metabolic regulation in *H3f3a*+/G34 mutant mice, we investigated fat tissues (liver, epididymal white adipose tissue (epi-WAT) and brown adipose tissue (BAT)) of both sexes at a mature adult age (5 months) by histopathology (Fig. 3.3A-D). Staining with hematoxylin and

eosin (H&E) revealed a common obesity-related phenotype in G34 mutant mice. Among G34-mutant mice, the *H3f3a*+/G34W mice had the most severe obesity phenotype (Fig. 3.3B-D). *H3f3a*+/G34W mice displayed macrosteatosis (fatty liver with large lipid droplet phenotype), and *H3f3a*+/G34R had macrosteatosis but to a lesser extent. The *H3f3a*+/G34V mice showed microsteatosis. Also, *H3f3a*+/G34W mice had some immune cell infiltration in liver tissues which is not present in other mutants or control mice. Furthermore, *H3f3a*+/G34W mice presented with hepatocyte ballooning, a feature characteristic of non-alcoholic steatohepatitis (NASH), a more severe form of fatty liver disease[170,171](Fig. 3.3B-D).

All *H3f3a*+/G34 mutants contained larger adipocytes in epi-WAT compared with controls, with abundant crown-like structures. These structures are macrophages surrounding dying adipocytes, a clear sign of "low-grade" inflammation associated with obesity[172]. Moreover, in the interscapular BAT, there is a clear whitening of the tissue (associated with too much blood lipid) in all *H3f3a*+/G34 mutants, with the highest lipid accumulation in *H3f3a*+/G34W mice (Fig. 3.3B-D).

To further study the obesity phenotype in *H3f3a*+/G34 mutant mice, we evaluated the fat mass of young adult mice (2-3 months) by echo-MRI. We found that there is a significant increase in the fat mass of *H3f3a*+/G34 mutants compared to control mice, despite the fact that there is no significant body weight difference between the groups. (Fig. 3.3E). We also did H&E staining for liver, BAT and epi-WAT tissues of this cohort of mice. The BAT of G34W mice showed a whitening in color and an increased lipid accumulation (Fig. 3.3F). In addition, we noticed that lipids started to build up in their liver cells (data not shown). The epi-WAT in these *H3f3a*+/G34W mice showed additional changes in adipocyte morphology, including reduced numbers of adipocytes and adipocyte hypertrophy. Adipocyte hypertrophy is also seen in the *H3f3a*+/G34R/V mice (Fig. 3.3F). Together, these results suggest that the progressive acquisition of fat is the primary reason for the increased weight gain observed in G34-mutant mice postnatally, with the G34W mutation promoting the most severe phenotype.

A complete blood count (CBC) and blood chemistry test was performed in two-month-old *H3f3a*+/G34 mutants and control mice. *H3f3a*+/G34W mice had much higher levels of serum leptin, HDL and LDL than in control or *H3f3a*+/G34R/V mutant mice. Compared to control mice,

blood glucose levels were also higher in *H3f3a*+/G34W mice, despite the fact that their serum insulin levels were not different (Fig. 3.3G). In addition, there were no significant differences in the numbers of red or white blood cells across groups. However, platelet numbers were lower in *H3f3a*+/G34 mutants compared to control mice (data not shown). Because leptin is a signal from adipose tissue to the brain to regulate food intake[173], we evaluated whether the higher levels of leptin in *H3f3a*+/G34W mice impact their food intake. *H3f3a*+/G34W mice did not consume more food. *H3f3a*+/G34R mice seemed to have higher cumulative food intake than the others, although this did not reach significance (Fig. 3.3H). The combination of profound obesity, high leptin levels and normal food intake suggests that *H3f3a*+/G34W mice may be insensitive to satiety signals.

In summary, histology of fat tissues showed signs of unhealthy obesity such as low-grade inflammation, macrophage infiltration, and adipocyte hypertrophy, together with lipotoxicity from ectopic lipid deposition, fatty liver, and high blood glucose, especially for *H3f3a*+/G34W mice.

3.4. *H3f3a*+/G34W mice show common dysregulated pathways of the obesity phenotype

To further understand the effect of *H3f3a*+/G34 mutations on adipogenesis, we assessed gene expression differences between *H3f3a*+/G34 mutant and control mice in epi-WAT tissues at 2-3 months-age by RNA sequencing (RNA-seq). Principal component analysis (PCA) showed that epi-WAT samples of *H3f3a*+/G34W mice cluster robustly together and separately from control mice (Fig. 3.4A). Transcriptomic analysis revealed 3382, 1089, and 456 differentially expressed genes (DEGs) comparing *H3f3a*+/G34W, *H3f3a*+/G34R and *H3f3a*+/G34V mutants, respectively, to control mice (Fig. 3.4B). A selection of genes comparing *H3f3a*+/G34W to control mice, based on the highest fold change, is shown in the volcano plot (Fig. 3.4C).

Gene-set enrichment analysis (GSEA) of the DEGs between *H3f3a*+/G34W and control mice revealed several dysregulated pathways including down-regulated hallmark gene sets of oxidative phosphorylation, adipogenesis and fatty acid metabolism and up-regulated gene sets such as epithelial mesenchymal transition, inflammatory response, and complement (Fig. 3.4D).

These dysregulated pathways are similarly enriched in WAT and aorta samples from other models of obesity in mice[174].

Next, we focused on gene sets relevant to adipogenesis and fatty acid metabolism. Many genes were significantly downregulated in *H3f3a*+/G34W epi-WAT compared to the control samples, in contrast to *H3f3a*+/G34R and *H3f3a*+/G34V mice (Fig. 3.4E). For instance, an important gene implicated in adipogenesis that is significantly downregulated in *H3f3a*+/G34W epi-WAT is *Ennp2*-nucleotide pyrophosphatase/phosphodiesterase (Fig. 3.4E), an important regulator of BAT[175] function, energy expenditure, obesity and metabolic conditions related to adipose tissue expansion . *Ennp2* is normally highly expressed in adipose tissue such as epididymal fat and less in subcutaneous or brown fat[175].

In addition, one of the master regulators of WAT adipogenesis, peroxisome proliferator-activated receptor gamma gene (*Pparg*), is downregulated in *H3f3a*+/G34W compared to controls, but not in *H3f3a*+/G34R and *H3f3a*+/G34V epi-WAT samples (Fig. 3.4E-F). Using a database of mouse Pparg target genes[176], we found that several Pparg target genes such as *Apoe*-apolipoprotein, *Slc27a1*-long-chain fatty acid transport protein 1, *Adipoq*-Adiponectin and *Cat*-Catalase were also significantly downregulated in *H3f3a*+/G34W compared to control epi-WAT tissues (Fig. 3.4G). In addition, BAT marker genes such as *Ucp1* and *Cidea* are significantly downregulated in *H3f3a*+/G34W compared to control epi-WAT tissues (Fig. 3.4G).

Interestingly, the WNT1 inducible signaling pathway protein 2 gene (*Ccn5/Wisp2*) is highly upregulated in G34W WAT samples (Fig. 3.4F). Canonical WNT signalling is an important pathway in the commitment of mesenchymal stem cells (MSCs) to myocytes and osteocytes but inhibits MSC commitment to adipocytes. *Ccn5/Wisp2* is activated by the canonical WNT pathway and its expression is associated with WNT-regulated genes such as *Ccnd1* which is up-regulated *in H3f3a*+/G34W epi-WAT samples too *(*Fig. 3.4F). *Ccn5/Wisp2* is highly expressed in both human and mouse MSCs and preadipocytes but is very lowly expressed during adipocyte differentiation and not expressed in mature adipocytes. *Ccn5/Wisp2* expression can also be upregulated in adipose tissue under hypertrophic obesity/metabolic syndrome conditions[177]. This

high expression might be due to impaired adipogenesis rather than inappropriate regulation of *Ccn5/Wisp2* gene[177].

Altogether, the transcriptomic data suggests that G34W mutation may impede adipogenesis through activation of WNT signaling and downregulation of Pparg, the master regulator of adipogenesis. Low levels of Pparg consequently down-regulate its target genes, including BAT marker genes (Fig. 3.4F-H).

3.5. A high lethality rate in *H3f3a*+/G34W male mice can distinguish the effect of this mutation from the other G34 mutations

To investigate the effect of the G34 mutants on general physiology, we monitored the survival of a large cohort of *H3f3a*+/G34 mutants together with control mice (*H3f3a*+/+ and *H3f3a*+/-) on the two genetic backgrounds, C57Bl/6J and B6C3F1, in both sexes (Fig. 3.5A-D).

H3f3a+/G34W male mice showed a high level of lethality with only 11% surviving past 7 months of age. In contrast, *H3f3a*+/G34R and *H3f3a*+/G34V males lived substantially longer, with 67% and 87% surviving past 7 months of age respectively (Fig. 3.5A-B). *H3f3a*+/- male mice did not die during the 12-month study period. Survival curves among the female *H3f3a*+/G34 mice were closer to each other and had an overall lower mortality rate than the corresponding mutant male mice, with survival rates of 77%, 76%, and 93% in *H3f3a*+/G34W, *H3f3a*+/G34R and *H3f3a*+/G34V respectively (Fig. 3.5C-D).

To investigate the cause of the lethality in mutant mice, all were observed twice per week and classified according to the gross structural findings on necropsy or visible symptoms identified before death (Fig 3.5B-D). The mice that were either found dead, or visibly incapacitated, were evaluated according to an end-point scoring system and euthanized. A high percentage (77.2%) of *H3f3a*+/G34W male mice had an enlarged bladder, detectable by palpation and by their inability to urinate, and was later confirmed during necropsy. In *H3f3a*+/G34R male mice the reasons of death included fecal impaction, and signs of pain and distress, but only 2.5% of these mice had an

enlarged bladder. Similarly, few *H3f3a*+/G34V male mice presented with an enlarged bladder at the time of death (Fig 3.5B).

Enlarged bladder and fecal impaction were also the main causes of death in female *H3f3a*+/G34 mice. *H3f3a*+/G34R and *H3f3a*+/G34W females had a higher number of deaths due to fecal impaction and enlarged bladder, respectively (Fig. 3.5B-D).

3.6. Mortality of male G34W mice is associated with urinary tract system impairment

As shown, *H3f3a*+/G34W male mice experienced a high lethality, and 50% died at 3-month-old. This high mortality does not occur in *H3f3a*+/G34R/V mutated male animals (Fig. 3.5A). However, *H3f3a*+/G34W female mice do not show this high mortality rate (only 13% died at 3-month-old) (Fig. 3.5C).

We observed that *H3f3a*+/G34W male mice had difficulty urinating and had a palpable distended bladder (Fig. 3.6A). All mice were observed weekly to detect abnormal behavior from 3-weeks of age onwards. We found that *H3f3a*+/G34W male mice started to have a palpable bladder (score 1) and seemed to have difficulty urinating (score >1) at two time points: at 2-3-months (48%) or 3-4-months (43%) (Fig. 3.6B-C). We euthanized the animals when both bladder size and urine flow reached a score of 2 (defined end point).

Necropsy of *H3f3a*+/G34W male mice revealed an enlarged bladder (72 out of 83 mice) that was often accompanied by inflammation of the urinary tract and sometimes hydronephrosis (Fig. 3.6A-D and Table 3.1). Histologic analysis of the bladder and kidney of terminally ill *H3f3a*+/G34W male mice displayed a thinning and deformation of the bladder wall, as well as kidney damage. In some cases, the kidneys displayed hallmarks of hydronephrosis, such as dilated collecting ducts (Fig. 3.6D and Table 3.1).

In *H3f3a*+/G34W male mice, histone H3.3G34W protein, as assessed by IHC, showed a homogenous pattern of expression throughout bladder wall's layers (Fig. 3.6E). We then evaluated

bladder histology in 3-week-old mice, prior to them presenting with symptoms (i.e. palpable bladder, urine retention). H&E and Masson's Trichrome staining showed some changes in the urinary bladder wall compared to control mice, likely due to stretching of the bladder muscle. (Fig. 3.6F). However, the muscular layer did not show hypertrophy, and additional evaluation of fibrosis-specific markers such as vimentin and the proliferation marker Ki67 did not show any change compared to controls (Fig. 3.6G).

To evaluate potential renal dysfunction in G34 mutant mice, we performed blood chemistry tests with 2-3-month-old *H3f3a* +/G34W male mice with enlarged bladders (>1 cm distension and inability to fully empty bladder post-palpation). These mice were often recommended to be euthanized within 24 to 48 h (Fig. 3.6B-C). *H3f3a*+/G34W male mice showed a significant increase in blood urea nitrogen (BUN), creatinine, and potassium (hyperkalemia) compared to control mice (Table 3.2), indicative of abnormal kidney function. Hyperkalemia can cause cardiac arrhythmias, which may explain the sudden death of more than 70% of male *H3f3a* +/G34W mice. In addition, urine from terminally- ill *H3f3a* +/G34W male mice collected directly from the bladder contained high levels of leukocytes (3/4 mice), blood (3/4 mice), and mild to high levels of proteins and glucose (Table 3.3), further suggesting that profound kidney dysfunction may be the proximal cause of death in male *H3f3a* +/G34W mice.

3.7. Striated muscle defects and urethral obstruction lead to an enlarged bladder and renal dysfunction in *H3f3a* +/G34W mice

To further investigate the reason for high mortality in *H3f3a*+/G34W mice and their presentation with an enlarged bladder, we performed RNA-seq analysis on fresh frozen urinary bladder tissue (including the bladder neck) of all *H3f3a*+/G34 mutants and control mice at postnatal day, P7. Urinary bladder tissue consists of urothelial cells, smooth muscle cells, fibroblasts, endothelial cells and other cell types, as shown in the Human Protein Atlas (http://www.proteinatlas.org/ENSG00000105668-UPK1A/tissue/urinary+bladder). We found only a very small number of significant DEGs between H3f3a+/G34 mutants and control samples at P7 age. To identify potential pathways that might be involved in the enlarged bladder, we performed GSEA analysis of DEGs between *H3f3a*+/G34W and control mice (Fig. 3.7A). Some

upregulated genes in *H3f3a* +/G34W mice converged on a subgroup of genes regulated by MYC, such as MYC target V1, V2 and E2F target genes (Fig. 3.7A). Importantly, we also find gene sets related to epithelial–mesenchymal transition (EMT) and myogenesis enriched among downregulated genes in *H3f3a*+/G34W (Fig. 3.7A-B), including genes associated with muscle fiber and myofibril assembly such as *Actc1*-alpha cardiac muscle actin, *Mypn*- myopalladin and *Acta1*-alpha skeletal muscle actin. We found that these genes are downregulated in *H3f3a*+/G34W male mice compared to other *H3f3a*+/G34 mutants and control samples, although this difference did not reach statistical significance, probably due to low sample size (Fig. 3.7C). It also could be that the effects in P7 bladder tissues were too small to see relevant transcriptomic changes, and tissue heterogeneity may further dilute the observed effect, because striated muscle may be under-represented in the bulk sample.

Interestingly, a mouse model with the *Acta1* H40Y mutation showed many similarities with the *H3f3a*+/G34W mice, including that the male mice died unexpectedly and prematurely (5-13 weeks of age), an effect that was not observed in females[178]. Tinklenberg et al. showed that urinary outlet obstruction is the reason for lethality in *Acta1* H40Y male mice[179]. *Acta1* H40Y males presented with bladder distension, difficulty urinating, and as secondary phenotypes, a degree of hemorrhage and necrosis in urogenital tracts and hydronephrosis. All these phenotypes are found to be due to urethral obstruction, where the urethral striated skeletal muscle showed marked nemaline myopathy in which many fibres have actin-positive inclusions[179]. Chan et al. further confirmed that *Acta1* H40Y causes stiffness of the actin filaments and therefore does not properly support actin-myosin binding[180]. Also, it is interesting to point that Acta1 H40Y males performed well on the rotarod- and open field tests.

We further analyzed the expression of *Acta1*-alpha skeletal muscle actin and *Acta2*-alpha smooth muscle actin genes in micro-dissected bladder tissues of mice at adult age. We found that the expression of smooth muscle actin was normal within the bladder wall and neck of *H3f3a*+/G34W mice, but expression of alpha skeletal muscle actin was not uniform and was lower in urethral striated muscle (Fig. 3.7D-E).

Altogether, our data strongly suggest that the enlarged bladder and renal dysfunction in *H3f3a*+/G34W male mice may be due to defects in striated muscle function through G34W-mediated downregulation of genes key to myogenesis.

3.8. *H3f3a*+/G34R mice show abnormalities in motor coordination

To evaluate neurological and neurobehavioral phenotypes of G34 mutated mice during maturation, we performed a panel of well-defined behavioral tests in G34 mutated mice and compared performance with age-matched wild-type and *H3f3a*+/- mice.

One of the hallmarks of neurological disorders is hindlimb-clasping[144,181]. Healthy WT mice, when picked up by their tail, extend all their four limbs in anticipation of contact. However, in some neurological pathologies (e.g. producing lesions in the cerebellum, basal ganglia, neocortex, or spinal cord), a flexion response characterized by limb-clasping, typically hindlimb-clasping and sometimes both hindlimb- and forelimb-clasping, is observed[181]. We used a scoring system to evaluate the progression of hindlimb-clasping in *H3f3a*+/G34 mutants (Fig. 3.8A-C; For a description, see materials & methods, chapter 2.9). As young adults (P30-90), controls and *H3f3a*+/G34V/W mutants mice have a score of 0 or ~1. In contrast, all *H3f3a*+/G34R mice (n=31) developed hindlimb-clasping behaviors (score 1-3) at this age (Fig. 3.8B), a phenotype that increased in severity with age (Fig. 3.8C). Notably, some showed clasping of both hind- and fore-limbs (bat-like posture, in addition to score 3) (Fig. 3.8A). This clasping behavior persists until death. Moreover, a tremor phenotype was observed in middle aged (10-12-months old) *H3f3a*+/G34R mice, further indicating a neurological abnormality.

To evaluate motor skill learning and balance, we used an accelerating rotarod test, measuring the time that mice stay on a rotating rod. At age 4-7 weeks, *H3f3a*+/G34R mice performed significantly worse compared to control mice. There were no significant differences in performances of other *H3f3a*+/G34 mutants, except at testing day 2 (Fig. 3.8D). In older mice (8 weeks onward), *H3f3a*+/G34R and *H3f3a*+/G34V, but not H3f3a+/G34W mice performed significantly worse compared to control mice (Fig. 3.8D), suggesting that *H3f3a*+/G34R/V mutations perturb functions related to balance.

To assess general behaviour and mobility of mutant mice, we performed an open field test. *H3f3a*+/G34R and *H3f3a*+/G34V mice showed a significant decrease in travelled distance and walking speed and reared less against the wall compared to age-matched (8 weeks) controls (Fig. 3.8E). These results are in line with their decreased motor coordination noted on the rotarod test. Mutant *H3f3a*+/G34W mice did not show these phenotypes. However, all *H3f3a*+/G34 mutants had defects in free rearing where the mice had difficulty standing on their hindlimbs and needed forelimb support on the wall (data not shown). This could indicate weakness of the hind legs and lack of balance during rearing. In addition, rearing against wall activity is a good measure of exploratory behaviour[182], and rearing deficits suggest a decrease in exploration.

To investigate in depth the motor coordination defects and possible gait abnormalities in *H3f3a*+/G34R mice, we used the Catwalk system to record the digital paw footprints of a walking mouse. *H3f3a*+/G34R mice displayed multiple placing of the same paw on or near the same site and exhibited a shortening of step length (Fig. 3.8F). We also observed a reduction of hind and front paw stride length ("the length between the successive placements of the same paw")[183] (Fig. 3.8F-G). The frequency of diagonal contacts in freely walking *H3f3a*+/G34R mice was also significantly lower than in control animals and led to an increase in trigonal supports (Fig. 3.8G). We also showed that *H3f3a*+/G34R mice walk with a wider distance between the front paws, an increased forepaws base of support (BOS, "distance between the forepaws or hindpaws")[184] (Fig. 3.8G).

In summary, *H3f3a*+/G34R mice acquired a profound tremor and demonstrated clasping of hind limbs, poor locomotion and balance. Moreover, *H3f3a*+/G34R mice were significantly slower and showed an abnormal walking pattern compared to control animals. This severe ataxia-like phenotype (tremor, hindlimb clasping, and abnormal gaiting) worsened progressively with age (Fig. 3.8H-I). These abnormal behaviours were present to a lesser degree or absent in *H3f3a*+/G34V mice and were entirely absent in *H3f3a*+/G34W mutant mice.

3.9. *H3f3a*+/G34R mice have defects in cerebellar foliation patterning and show a progressive shrinking of the brain

From P7 onwards, the brains of all *H3f3a*+/G34 mice were smaller than of WT and het-KO littermates (Fig. 3.9A). This corresponds with reduced body weight in the *H3f3a*+/G34 mice (Fig. 3.2A-C). However, from P30 onwards, the brains of *H3f3a*+/G34R mice, and to a lesser extent *H3f3a*+/G34V, progressively decreased in size relative to those of *H3f3a*+/G34W and control mice (Fig. 3.9A-B).

To investigate brains in *H3f3a*+/G34 mice, we evaluated brain structure by histology. The cerebellum of mammals and birds has a conserved foliation pattern, with a well-organized layered structure consisting of anteroposterior folds and lobules/folia[185]. Lobules are anatomically divided into medial vermis, paravermis and hemispheres. In mice, a sagittal section of the vermis consists of lobules which are further divided into four domains: anterior (lobules I-V), central (lobules IV-VIII), posterior (lobules VIII-IV) and nodular (lobules IV-X)[185]. The development of embryonic to complex mature cerebellum with all lobules formed is mostly complete by P21[186].

A Cresyl Violet (Nissl) stain at different developmental stages revealed abnormal cerebellar structure in *H3f3a*+/G34R animals at P90-180, with a less developed or simplified foliation pattern (Fig. 3.9C). These structural abnormalities were also evident at earlier stages: at P7, the shapes of lobule I-III, IV/V and Lobule VI were different. Furthermore, the fissures between lobules IV and V, and in lobule IX, were missing or poorly developed (Fig. 3.9D). After the cerebellum is fully developed, at P21, the *H3f3a*+/G34R mice have clearly affected cerebellum lobules. This is less pronounced in *H3f3a*+/G34V, and not observed in *H3f3a*+/G34W animals (Fig. 3.9E).

We conclude therefore that *H3f3a*+/G34R mice had a dramatic effect of cerebellar foliation, while there were not any gross structural abnormalities in other parts of the brain.

3.10. *H3f3a+/*G34R mice have partial to severe Purkinje cells loss in cerebellar lobules

The hindlimb clasping reflex has been described in several mice with cerebellar atrophy, and is mostly characterized by cell ectopias in the cerebellar cortex, combined with Purkinje and granule cell losses[187].

We therefore evaluated the potential loss of Purkinje cells (PC) which might underlie the cerebellar ataxia phenotype in *H3f3a*+/G34R mice that uniquely display the hindlimb clasping reflex starting at P21-35. We performed immunofluorescence staining of PC with anti-calbindin-D28k and indeed found a significant loss of PCs in some lobules of *H3f3a*+/G34R mice compared to control and *H3f3a*+/G34V mice at 12-months of age (Fig. 3.10A-B). This PC loss was not observed in P7 *H3f3a*+/G34R mice (Fig. 3.10C-D), suggesting that PC loss, and associated abnormalities of migration, order and axonal branching exacerbate as mice age (Fig. 3.10E-F).

Altogether, we found in *H3f3a*+/G34R mice a partial to severe PC loss in cerebellar lobules coupled with the hindlimb clasping reflex, with severe progression as mice age.

3.11. GFAP+ cells are increased in the brain of *H3f3a+/*G34R animals

Single-cell RNA sequencing (scRNA-seq) data of pediatric HGGs with G34R/V mutations showed that these tumors have a dual neuronal and astrocytic cell identity, and completely lack oligodendroglial signatures (Chen et al. submitted Cell 2020). We therefore decided to investigate the effect of G34 mutations on neuronal and glial populations by performing fluorescent immunohistochemistry using antibodies for NeuN, a marker of neuronal cells, and glial fibrillary acidic protein (GFAP), a marker for mature astrocytes. The number of neuronal cells (NeuN+) and their distribution in the cortex and cerebellum of H3f3a+/G34 mutants at all ages did not appear different from their littermate controls (Fig. 11A-E).

By contrast, we found increased GFAP positive (GFAP+) cells in *H3f3a*+/G34R mice throughout the brain at P60-90 (Fig. 3.11A-D). Analysis of the number of GFAP+ cells at P60-90

in specific regions of the cerebral cortex showed a significant increase of GFAP+ astrocytes in *H3f3a*+/G34R, but no significant increase in *H3f3a*+/G34V/W brains. In aging mice (P300-420), we found a significant increase of GFAP+ astrocytes also in *H3f3a*+/G34R/V (Fig. 3.11D).

Similarly, there was no significant GFAP intensity difference in cerebellar lobules at P7 in all groups, but at P21 a significant increase was notable in *H3f3a*+/G34R, but not in *H3f3a*+/G34V/W mice. The intensity further increased in *H3f3a*+/G34R/V mouse brains at P300-420 (Fig. 3.11E). We also observed an increase in vimentin and S100B immunoreactivity cells in *H3f3a*+/G34R mouse brains (data not shown). Vimentin is an astrocytic/glial precursor marker appearing before GFAP expression, while S100 expression in astrocytes defines a late mature developmental stage after GFAP-expressing cells have lost their NSC potential[188,189].

Altogether, in *H3f3a*+/G34R and to a lesser extent in *H3f3a*+/G34V mouse brains, there are increasing numbers of GFAP+ astrocytes with age. This might be due to the proliferation of astrocytes or more differentiation of neural progenitor cells towards astrocytes.

3.12. *H3f3a*+/G34R mice show reduced Olig2+ cells and an impaired myelination during aging

In *H3f3a*+/G34R mouse brains, we found that the number of Olig2+ cells in areas of the cerebral cortex with increased GFAP+ astrocytes was significantly lower compared to those of the littermate controls (Fig. 3.12A-B). Moreover, we found by staining myelin basic protein (MBP) that *H3f3a*+/G34R brains at P420 showed significantly reduced myelination in cerebral hemispheres, from densely myelinated axons near the corpus callosum to sparsely myelinated axons in outer layer of the cortex (Fig. 3.12C). We also found a reduced percentage of myelin in the cortex in *H3f3a*+/G34R brains at P60-90. However, the brains of *H3f3a*+/G34V/W mice had similar myelination levels as controls (Fig. 3.12D-E).

Altogether, the histological data demonstrate abnormalities in the development of cerebellum, severe reduction or loss of Purkinje cells in cerebellar lobules, a global proliferation of GFAP+ astrocytes, and a reduction in myelinating oligodendrocytes. Each of these histological

abnormalities, individually or in combination, can result in the observed neurobehavioral abnormalities.

3.13. Phenotype–tissue associations with G34 histone mutations

To further investigate affected tissues in an unbiased manner, we submitted a cohort of decapitated perfused adult mice from both sexes and all genotypes to a laboratory animal examination center (IDEXX BioAnalytics, Shari Hamiliton, DVM, Columbia, USA) for dissection/necropsy studies The necropsy/dissection focused on examining the requested tissues (Table 3.4) and to report any additional abnormalities that they have found. Table 3.1 shows a detailed diagnosis. Briefly, all G34V/W animals were obese, with hepatic inflammatory (associated with lipidosis) and enlarged brown fat. However, the G34R group demonstrated fewer animals with obesity phenotype. Furthermore, a dilated urinary bladder and hydronephrosis was noted in several animals (Table 3.1). Additionally, more than 50% of animals with *H3f3a*+/G34R genotype had lesions in their gastrointestinal tract, extending from the stomach to the rectal/anal junction (Table 3.1), consistent with the fecal impaction phenotype (Fig. 3.5B-D).

A heatmap plot in Fig. 3.13A summarizes the most significant phenotypes of the *H3f3a*+/G34 mutants, for both sexes (Fig. 3.13A). *H3f3a*+/G34R mice exhibit severe neurological and neuromuscular alterations, while *H3f3a*+/G34W mice have bladder and uretero-genital issues. Interestingly, the obesity phenotype is a common phenotype between all three G34 mutant genotypes and most pronounced in *H3f3a*+/G34W mice.

In summary, we observe strikingly distinct phenotypes between *H3f3a*+/G34R and *H3f3a*+/G34W compared to WT control or mice carrying a deletion of one *H3f3a* allele. *H3f3a*+/G34V mice show a similar but milder phenotype as *H3f3a*+/G34R/W (Fig. 3.13A). Together, these different phenotypes reinforce the codon-specific impact of H3.3G34 mutations on different tissues, likely mediated through their differential effects on the epigenome and tissue-specific reader proteins.

Table 3. 1 IDEX metadata

IDEXX BioAnalytics Phenotyping Tissue List	*H3f3a* +/G34R
Stomach (glandular & non-glandular)	3/7 mice had the below lesions: Focal mild glandular dilatation Moderate to marked epithelial hyperplasia and marked hyperkeratosis in non-glandular stomach Suppurative gastritis, mild necrosis and ulceration in non-glandular stomach
Small intestine (Duodenum, Jejunum, Ileum)	3/7 mice had the below lesions: Mild intestinal dilatation Multifocally, lumen containing sloughed epithelial cells and mucous; mild epithelial hyperplasia of the crypts Focally, moderate apoptosis of surface epithelial cells and sloughing
Cecum	1/7 mice had lesion
Colon	3/7 mice had the below lesions: Mild to moderate colonic dilatation Abundant luminal mucous and sloughed cells in anterior colon; moderate crypt hyperplasia in the distal colon Abundant luminal mucous and sloughed cells in anterior colon; rare apoptotic surface epithelial cells
Rectal/Anal junction	4/7 mice had the below lesions: Ulcerative proctitis and transmural suppurative necrosis Circumferential rectal ulceration with hemorrhage annd edema; diffuse necrosis at the anus Moderate submucosal edema at the anus Mild diffuse hyperplasia of epidermal squamous cells at the rectal/anal junction, circumferential rectal ulceration
Liver	7/7 mice had the below lesions: Marked hepatic lipidosis Mild hepatic lipidosis Focal necrosis
Tissue masses/lesions	2/7 mice were markedly obese
Brown fat	1/2 mice had mixed inflammatory cells
Spleen	1/6 mice had enlarged spleen
Preputail gland	–
Urethra, penile	2/4 males had penile containing proteinaceous concretion
Kidenys	2/7 mice had mild hydronephrosis
Urinary bladder	1/7 mice had luminal dialation

IDEXX BioAnalytics Phenotyping Tissue List	*H3f3a* +/G34V
Stomach (glandular & non-glandular)	–
Small intestine (Duodenum, Jejunum, Ileum)	1/4 mice had mild multifocal epithelial hyperplasia of the crypts and mild to moderate mixed inflammatory cells
Cecum	–
Colon	–
Rectal/Anal junction	–
Liver	4/4 mice had the below lesions: Inflammatory reaction Markded steatohepatitis Mild hepatic lipidosis
Tissue masses/lesions	4/4 mice were markedly obese
Brown fat	3/4 mice had mixed inflammatory cells
Spleen	
Preputail gland	–
Urethra, penile	
Kidenys	3/4 mice had mild to moderate hydronephrosis
Urinary bladder	3/4 mice had luminal dialation, and mild inflammatory infiltrates within the wall

IDEXX BioAnalytics Phenotyping Tissue List	H3f3a +/G34W
Stomach (glandular & non-glandular)	–
Small intestine (Duodenum, Jejunum, Ileum)	1/4 mice had lesion Luminal sloughed cells
Cecum	1/4 mice had lesion
Colon	–
Rectal/Anal junction	–
Liver	4/4 mice had the below lesions: Marked steatohepatitis Hepatic lipidosis Focal necorsis and inflammatory reaction
Tissue masses/lesions	4/4 were markedly obese
Brown fat	2/4 mice had mixed inflammatory cells
Spleen	1/4 mice had enlarged spleen
Preputail gland	1/2 males had preputial glands containing thick exudate
Urethra, penile	1/2 males had penile containing proteinaceous concretion
Kidenys	4/4 mice had mild to moderate hydronephrosis
Urinary bladder	4/4 mice had luminal dialation, and mild inflammatory infiltrates within the wall

Table 3. 2 Chemistry blood test

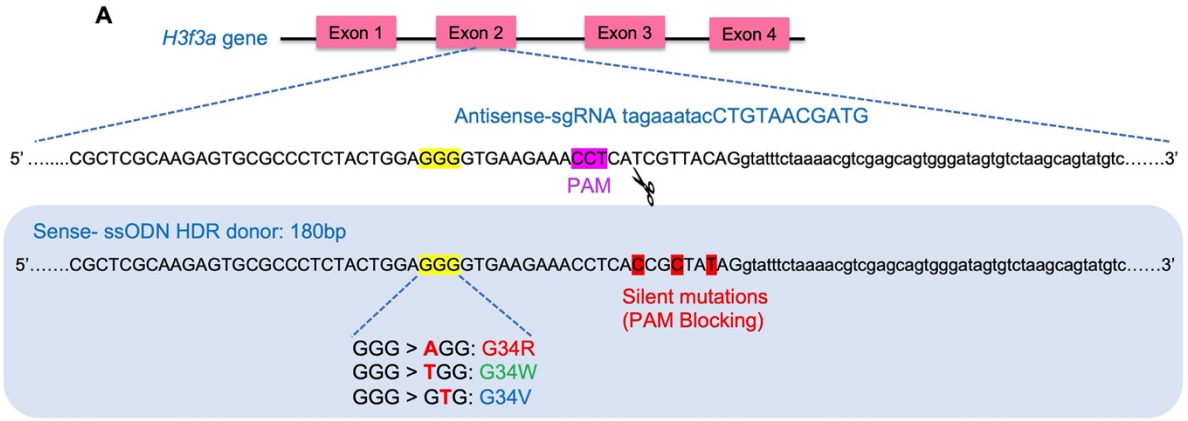
Genotype	Enlarged bladder	Urinating	Na (mmol/L)	K (mmol/L)	Cl (mmol/L)	iCa (mmol/L)	BUN (mg/dL)	Creatinine (mg/dL)
***H3f3a* +/+**	N/A	N/A	146	8	112	1.53	8.4	73
***H3f3a* +/+**	N/A	N/A	144	5.7	107	1.35	8.1	128
***H3f3a* +/+**	N/A	N/A	150	5	109	1.38	8.7	108
***H3f3a* +/G34W**	++	no urniating	141	>9.0	118	1.22	>50	172
***H3f3a* +/G34W**	++	no urniating	130	>9.0	112	1.29	>50	216
***H3f3a* +/G34W**	++	not measured	136	>9.0	115	0.65	17.9	<18
***H3f3a* +/G34W**	++	not measured	134	>9.0	111	1.17	>50	161
***H3f3a* +/G34W**	hard bladder	no urniating	136	6.4	107	1.08	>50	295
***H3f3a* +/G34W**	++	no urniating	134	4.8	106	1.08	>50	156

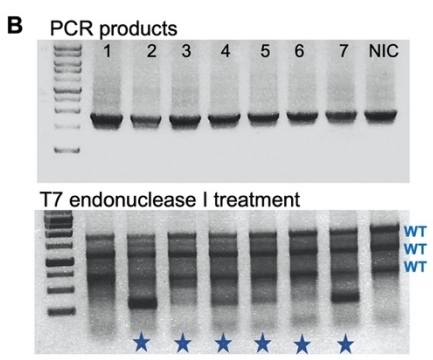
Table 3. 3 Chemical urine test

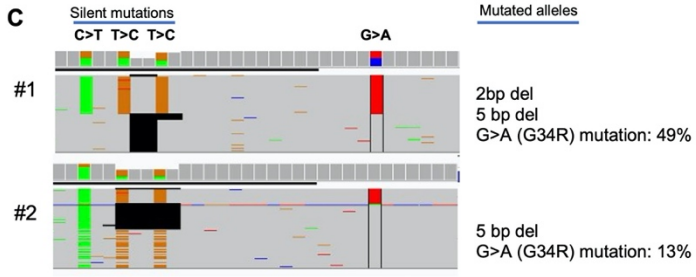
Genotype	pH	Leukocytes	Nitrates	Proteins	Glucose	Ketones	Urobilinogen	Bilirubin	Blood	Hemoglobin
***H3f3a* +/+**	5	neg	neg	neg	neg	neg	normal	neg	neg	N/A
***H3f3a* +/G34W**	6	2+	neg	2+	1+	neg	normal	neg	yes	4+
***H3f3a* +/G34W**	6	3+	pos	1+	2+	1+	normal	1+	yes	4+<
***H3f3a* +/G34W**	6	3+	pos	3+	1+	1+	1+	1+	yes	4+<
***H3f3a* +/G34W**	6	neg	neg	1+	neg	1+	normal	neg	neg	N/A

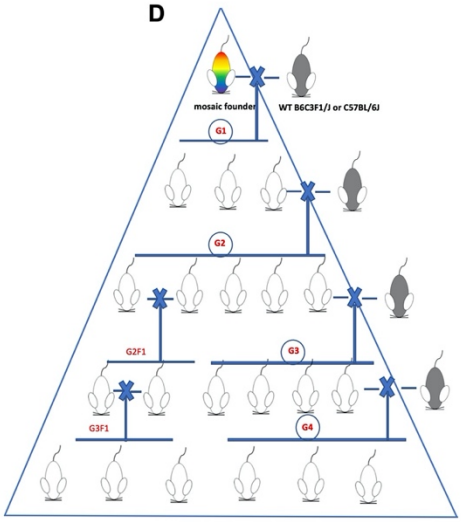
Table 3. 4 Phenotyping tissues list

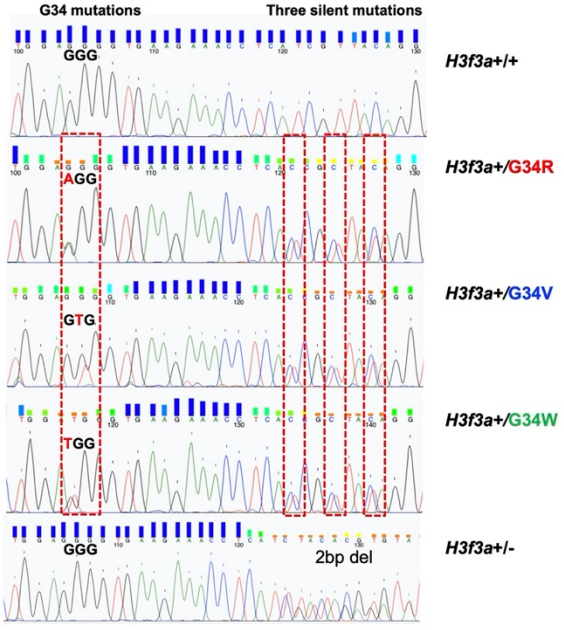
Cardiovascular	**Immune**	**Reproductive (male)**
Heart	Spleen	Bubourethral gland
Digestive	**Miscellaneous**	Epididymis
Duodenum	Tissue masses/lesions	Preputial gland
Cecum	**Musculoskeletal**	Prostate
Rectal/Anal junction	Diaphragm	Urethra, penile
Colon, distal	Femur/stifle	Testes
Colon, proximal	Muscle, quadriceps	**Respiratory**
Fat, brown	**Nervous**	Lung
Gall bladder	Spinal cord	**Urinary**
Ileum	**Reproductive (female)**	Kidneys
Jejunum	Ovary	Urinary bladder
Liver	Uterus	
Salivary gland, parotid	Vagina/Cervix	
Stomach (glandular & nonglandular)		


A
H3f3a gene
Exon 1
Exon 2
Exon 3
Exon 4
Antisense-sgRNA tagaaatacCTGTAACGATG
5' .....CGCTCGCAAGAGTGCGCCCTCTACTGGAGGGGTGAAGAAACCTCATCGTTACAGgtatttctaaaacgtcgagcagtgggatagtgtctaagcagtatgtc.....3'
PAM
Sense- ssODN HDR donor: 180bp
5'.....CGCTCGCAAGAGTGCGCCCTCTACTGGAGGGGTGAAGAAACCTCACCGCTATAGgtatttctaaaacgtcgagcagtgggatagtgtctaagcagtatgtc.....3'
Silent mutations
(PAM Blocking)
GGG > AGG: G34R
GGG > TGG: G34W
GGG > GTG: G34V


B
PCR products
1 2 3 4 5 6 7 NIC
T7 endonuclease I treatment
WT
WT
WT


C
Silent mutations
C>T T>C T>C
G>A
Mutated alleles
#1
2bp del
5 bp del
G>A (G34R) mutation: 49%
#2
5 bp del
G>A (G34R) mutation: 13%


D
mosaic founder
WT B6C3F1/J or C57BL/6J
G1
G2
G2F1
G3
G3F1
G4


E
G34 mutations
Three silent mutations
GGG
H3f3a+/+
AGG
H3f3a+/G34R
GTG
H3f3a+/G34V
TGG
H3f3a+/G34W
GGG
2bp del
H3f3a+/-

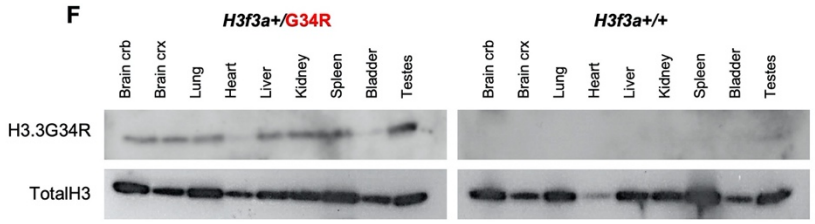

F
H3f3a+/G34R
H3f3a+/+
Brain crb
Brain crx
Lung
Heart
Liver
Kidney
Spleen
Bladder
Testes
H3.3G34R
TotalH3

Figure 3. 1 Generation of DKI mice lines with germline *H3f3a* point mutations using the CRISPR/Cas9-system

A) Schematic representation of the location of antisense-sgRNA and sense-ssODN and the mouse *H3f3a* locus. The scissor indicates the cleavage site. Purple letters show the PAM sequence. The ssODN homology-directed repair (HDR) donor sequence is shown inside the blue box at the corresponding position of the WT *H3f3a* locus. The yellow-, and red- highlighted bases in the sequence indicate the nucleotide substitutions site of the G34 codon and the three silent mutations for blocking PAM, respectively.

B) *In vivo* validation of the *H3f3a*-targeted sgRNA efficiency in 8 blastocysts using the T7 endonuclease I assay. Lanes 1–7: PCR products amplified from DNA of each Cas9 protein and sgRNA injected blastocyst. Lane NIC (no injection control) shows the PCR product amplified from DNA of a single un-injected control blastocyst. The asterisks indicate the additional 2-3 bands. The cutting efficiency of used sgRNA was 85%.

C) MiSeq sequences of the *H3f3a* edited alleles in two mosaic *H3f3a*+/G34R founders indicate the different genotypes (WT, different indels and G34R point mutation) in the unique alleles.

D) The breeding schema represents the strategy for establishing the lines of heterozygous *H3f3a*+/G34 mutant mice from the mosaic founders. G indicates generation number. F shows the number of brother x sister inbred intercrosses.

E) Representative electropherograms of Sanger sequencing of the different genotypes: *H3f3a*+/+, *H3f3a*+/G34R, *H3f3a*+/G34V, *H3f3a*+/G34W, and *H3f3a*+/-. The *H3f3a* locus around the gRNA target site was PCR amplified and sequenced. The results of the Sanger sequence analysis of the *H3f3a* locus of mice with a wild-type (GGG), *H3f3a*+/G34R (G>A: AGG), *H3f3a*+/G34V (G>T: GTG), *H3f3a*+/G34W (G>T: TGG), and *H3f3a*+/- (two nucleotides deletions). All genotypes contain the three silent mutations of PAM blocking.

F) Immunoblots of different tissues using a specific H3.3 G34R antibody. The total histones were extracted from different tissues of adult mice with two different genotypes (*H3f3a*+/G34R and *H3f3a*+/+).

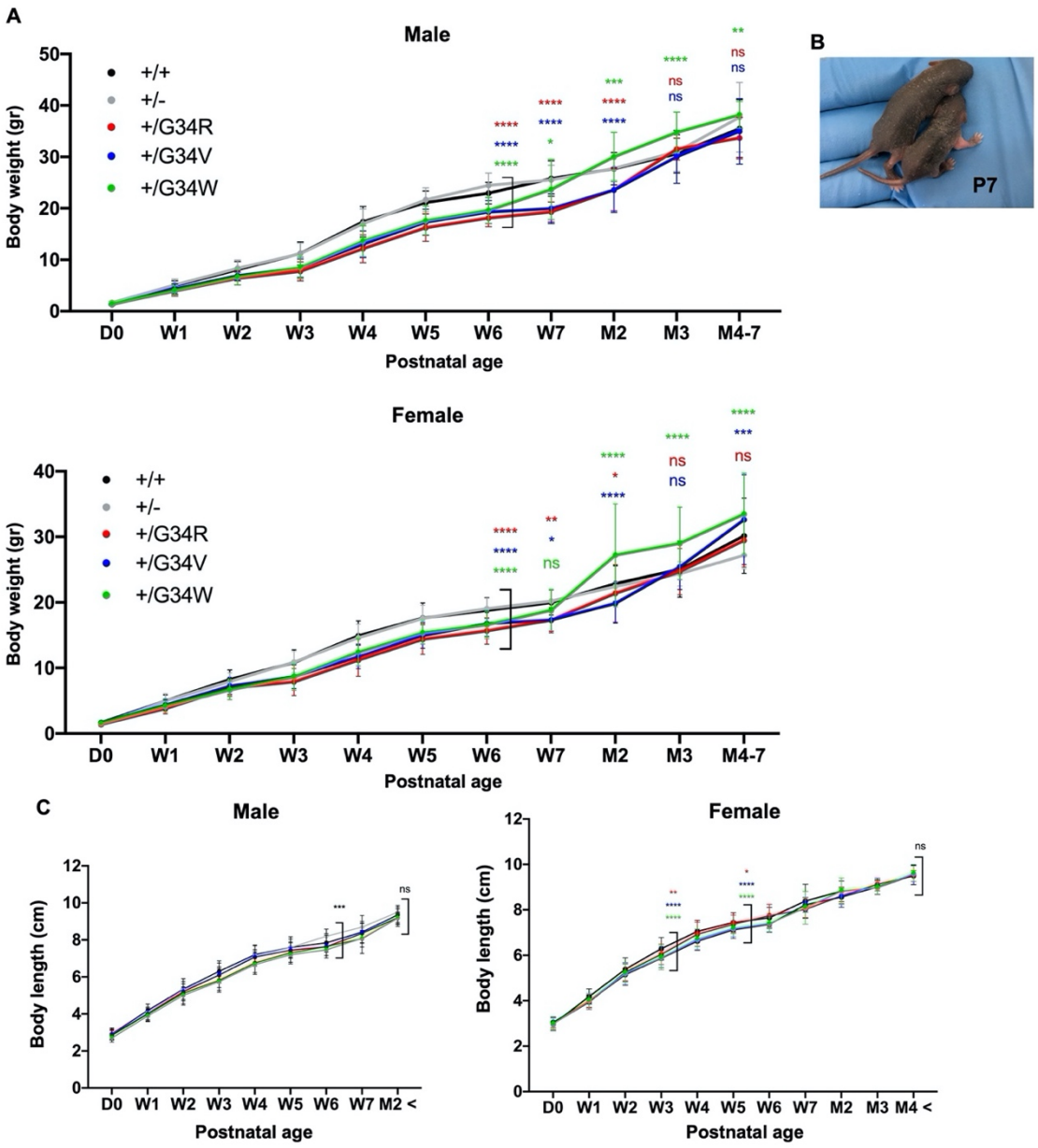


Figure 3. 2 *H3f3a*+/G34 mutations negatively affect growth during early postnatal development, but mutant mice acquire weight as adults

A-C) Mean body weights (A) and lengths (C) of all mutated G34 mice (generation: G1-G5, separated by male and female) on either a B6C3F1/Crl or a C57BL/6J DKI-background. All mice were fed a normal diet. The data is presented as mean ± SD. The statistical significance assessed using 2-WAY ANOVA compared to control mice. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, **** $P < 0.0001$.

B) A representative picture of WT and *H3f3a*+/G34R mice at P7 age.

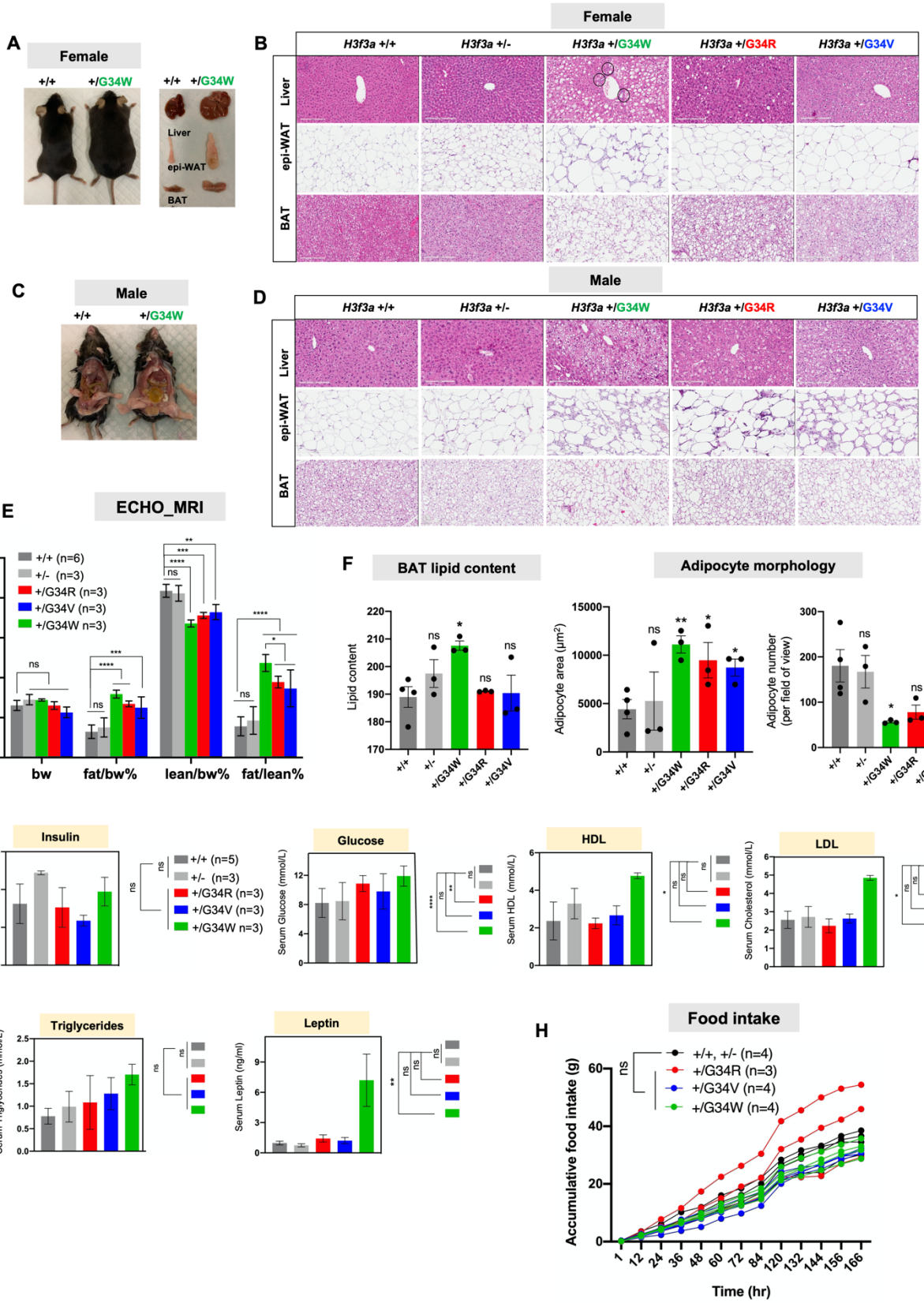

A
Female
+/+
+/G34W
Liver
epi-WAT
BAT
B
Female
H3f3a +/+
H3f3a +/-
H3f3a +/G34W
H3f3a +/G34R
H3f3a +/G34V
Liver
epi-WAT
BAT
C
Male
+/+
+/G34W
D
Male
E
ECHO_MRI
+/+ (n=6)
+/- (n=3)
+/G34R (n=3)
+/G34V (n=3)
+/G34W n=3)
bw
fat/bw%
lean/bw%
fat/lean%
F
BAT lipid content
Lipid content
Adipocyte morphology
Adipocyte area (μm²)
Adipocyte number (per field of view)
G
Insulin
Serum Insulin (ng/ml)
Glucose
Serum Glucose (mmol/L)
HDL
Serum HDL (mmol/L)
LDL
Serum Cholesterol (mmol/L)
Triglycerides
Serum Triglycerides (mmol/L)
Leptin
Serum Leptin (ng/ml)
H
Food intake
+/+, +/- (n=4)
+/G34R (n=3)
+/G34V (n=4)
+/G34W (n=4)
Accumulative food intake (g)
Time (hr)

Figure 3. 3 *H3f3a*+/G34 mutant mice have an obesity phenotype

A) A representative picture of WT (*H3f3a*+/+) and *H3f3a*+/G34W female mice at 5-month (C57BL/6J) fed a normal diet (left) and a representative picture of liver, epi-WAT, and BAT at that age.

B) Representative images of H&E staining of BAT, liver and epi-WAT sections from females at 5-month. Control (*H3f3a*+/+ and *H3f3a*+/-) and DKI-mice with the different G34 genotypes (*H3f3a*+/G34R, *H3f3a*+/G34W, and *H3f3a*+/G34V; C57BL/6J) fed a normal diet. Scale bar: 200 µm. Black circles indicate immune cell infiltration.

C) Representative picture of WT (*H3f3a*+/+) and *H3f3a*+/G34W male mice at 5-month (C57BL/6J) fed a normal diet.

D) Representative images of H&E staining of BAT, liver and epi-WAT sections from males at 5-month. Control (*H3f3a*+/+ and *H3f3a*+/-) and DKI-mice with the different G34 genotypes (*H3f3a*+/G34R, *H3f3a*+/G34W, and *H3f3a*+/G34V; C57BL/6J) fed a normal diet. Scale bar: 200 µm. Black circles indicate immune cell infiltration.

E) Body composition of 2-3 months old male mice (B6C3F1/Crl) on a normal diet as measured by Echo-MRI (*H3f3a*+/+ (n=6), *H3f3a*+/- (n=3), *H3f3a*+/G34R (n=3), *H3f3a*+/G34W(n=3), *H3f3a*+/G34V (n=3)). The data is presented as means ± SD. The statistical significance assessed using two-way ANOVA. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, **** *P* < 0.0001 and n.s.: non-significant. Bw: body weight.

F) BAT lipid content (left panel), adipocyte area (middle panel) (μm^2) and adipocyte cell number (right panel) in epi-WAT of male mice at 2-3 months (*H3f3a*+/+ (n=4); *H3f3a*+/- (n=3); *H3f3a*+/G34R (n=3); *H3f3a*+/G34W(n=3); *H3f3a*+/G34V (n=3)). The data is presented as means ± SEM. The statistical significance assessed using one-way ANOVA. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, **** *P* < 0.0001 and n.s.: non- significant compared to WT mice (+/+).

G) Serum insulin, glucose, HDL, LDL, and triglycerides and leptin of 2-3 months old male mice. (*H3f3a*+/+ (n=5); *H3f3a*+/- (n=3); *H3f3a*+/G34R (n=3); *H3f3a*+/G34W(n=3); *H3f3a*+/G34V (n=3)). The data is presented as means ± SEM. The statistical significance assessed using one-way ANOVA. **P* < 0.05, ***P* < 0.01, and n.s.: non- significant compared to control mice (+/+, +/-).

H) Seven-day cumulative food intake of female mice at mature adult age (5-6-month) (*H3f3a*+/+; +/- (n=4), H*3f3a* +/G34R (n=3), *H3f3a*+/G34W (n=4), *H3f3a*+/G34V (n=4)). The statistical significance assessed using one-way ANOVA. n.s.: non- significant compared to control mice (+/+, +/-).

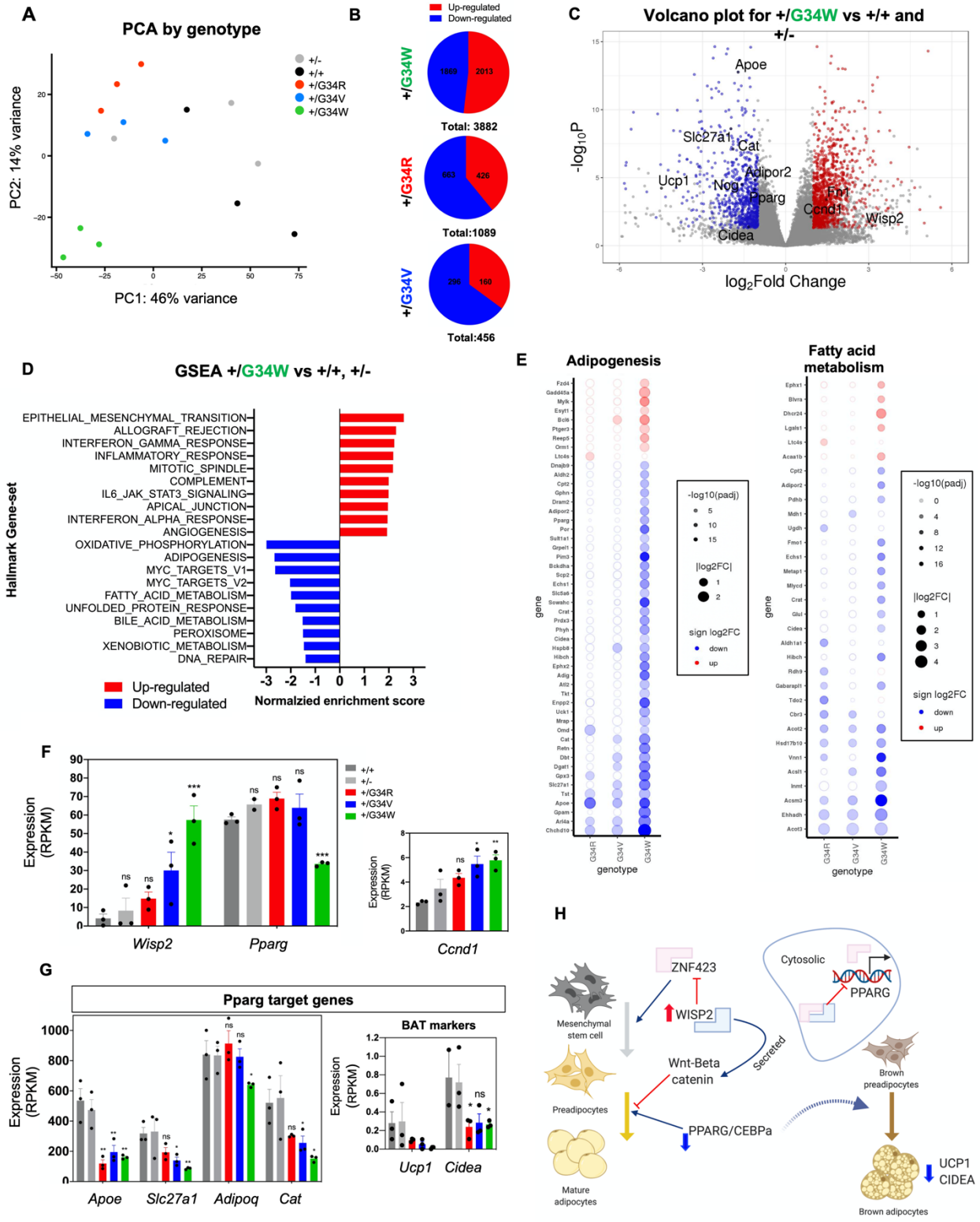

A
PCA by genotype
+/-
+/+
+/G34R
+/G34V
+/G34W
PC2: 14% variance
PC1: 46% variance
B
Up-regulated
Down-regulated
+/G34W
1869
2013
Total: 3882
+/G34R
663
426
Total:1089
+/G34V
296
160
Total:456
C
Volcano plot for +/G34W vs +/+ and +/-
Apoe
Slc27a1
Cat
Adipor2
Ucp1
Nog
Pparg
Fmt
Ccnd1
Wisp2
Cidea
-log10P
log2Fold Change
D
GSEA +/G34W vs +/+, +/-
EPITHELIAL_MESENCHYMAL_TRANSITION
ALLOGRAFT_REJECTION
INTERFERON_GAMMA_RESPONSE
INFLAMMATORY_RESPONSE
MITOTIC_SPINDLE
COMPLEMENT
IL6_JAK_STAT3_SIGNALING
APICAL_JUNCTION
INTERFERON_ALPHA_RESPONSE
ANGIOGENESIS
OXIDATIVE_PHOSPHORYLATION
ADIPOGENESIS
MYC_TARGETS_V1
MYC_TARGETS_V2
FATTY_ACID_METABOLISM
UNFOLDED_PROTEIN_RESPONSE
BILE_ACID_METABOLISM
PEROXISOME
XENOBIOTIC_METABOLISM
DNA_REPAIR
Hallmark Gene-set
Up-regulated
Down-regulated
Normalzied enrichment score
E
Adipogenesis
Fatty acid metabolism
gene
genotype
G34R
G34V
G34W
-log10(padj)
|log2FC|
sign log2FC
down
up
F
Expression (RPKM)
+/+
+/-
+/G34R
+/G34V
+/G34W
Wisp2
Pparg
Ccnd1
G
Pparg target genes
Apoe
Slc27a1
Adipoq
Cat
BAT markers
Ucp1
Cidea
H
Mesenchymal stem cell
Preadipocytes
Mature adipocytes
ZNF423
WISP2
Wnt-Beta catenin
Secreted
PPARG/CEBPa
Cytosolic
PPARG
Brown preadipocytes
Brown adipocytes
UCP1
CIDEA

Figure 3. 4 *H3f3a*+/G34W mice show common dysregulated pathways of the obesity phenotype

A) Principal Component Analysis based on gene expression of the 10K most variant genes.

B) Pie chart with the number of differentially expressed genes (DEGs) in epi-WAT tissue from *H3f3a*+/G34 mutated mice (G34V/W/R) versus control samples.

C) Volcano plot visualizes distribution of DEGs comparison of *H3f3a*+/G34W versus control (*H3f3a*+/+, and *H3f3a*+/-) samples. Fold change is indicated on log2-scale (LFC) on the x-axis, the adjusted p-value (FDR) is indicated on log10-scale on the y-axis. The values highlighted in blue and red have LFC of ≥ 1, FDR ≤ 0.05, and base mean ≥ 50, and show downregulated and upregulated DEGs respectively. Labeled genes do not necessarily meet highlighting criteria.

D) Gene set enrichment analysis (GSEA) of RNA-seq of epi-WAT samples in *H3f3a*+/G34W versus control (*H3f3a*+/+, and *H3f3a*+/-) mice at 8-weeks-old, indicating significant hallmark gene sets (FDR < 0.05). Red: upregulated genes; Blue: downregulated genes. The normalized enrichment score and significance was computed using the GSEA software (https://www.gsea-msigdb.org/gsea/index.jsp).

E) Bubble plot illustrates changes in the expression levels of adipogenesis and fatty acid metabolism in *H3f3a*+/G34 mutants relative to control samples. y axis: gene symbols. Log2FC >1, and FDR <0.05. Only significant genes meeting a criteria of absolute log2FC > 1 and FDR < 0.05 in at least one group are shown, significance is denoted by a filled bubble.

F-G) Bar graphs depicting expression levels of *Wisp2, Pparg* (F, left panel)*, Ccnd1* (F, right panel)*, Apoe, Slc27a1, Adipoq, Cat* (G, left panel)*, Ucp1, Cidea* (G, right panel) in *H3f3a*+/G34 mutants and control samples. The statistical significance assessed using one-way ANOVA. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 and n.s.: non- significant compared to WT mice (+/+). The gene expression levels are reported in median-of-ratios normalized read counts. The significance was assessed using the RNA Seq analysis software DESeq2.

H) A schematic illustrating a potential impairment G34W-mediated adipogenesis through down-regulation of the adipogenesis master regulator Pprg and up-regulation of the adipogenesis protein regulator Wisp2.

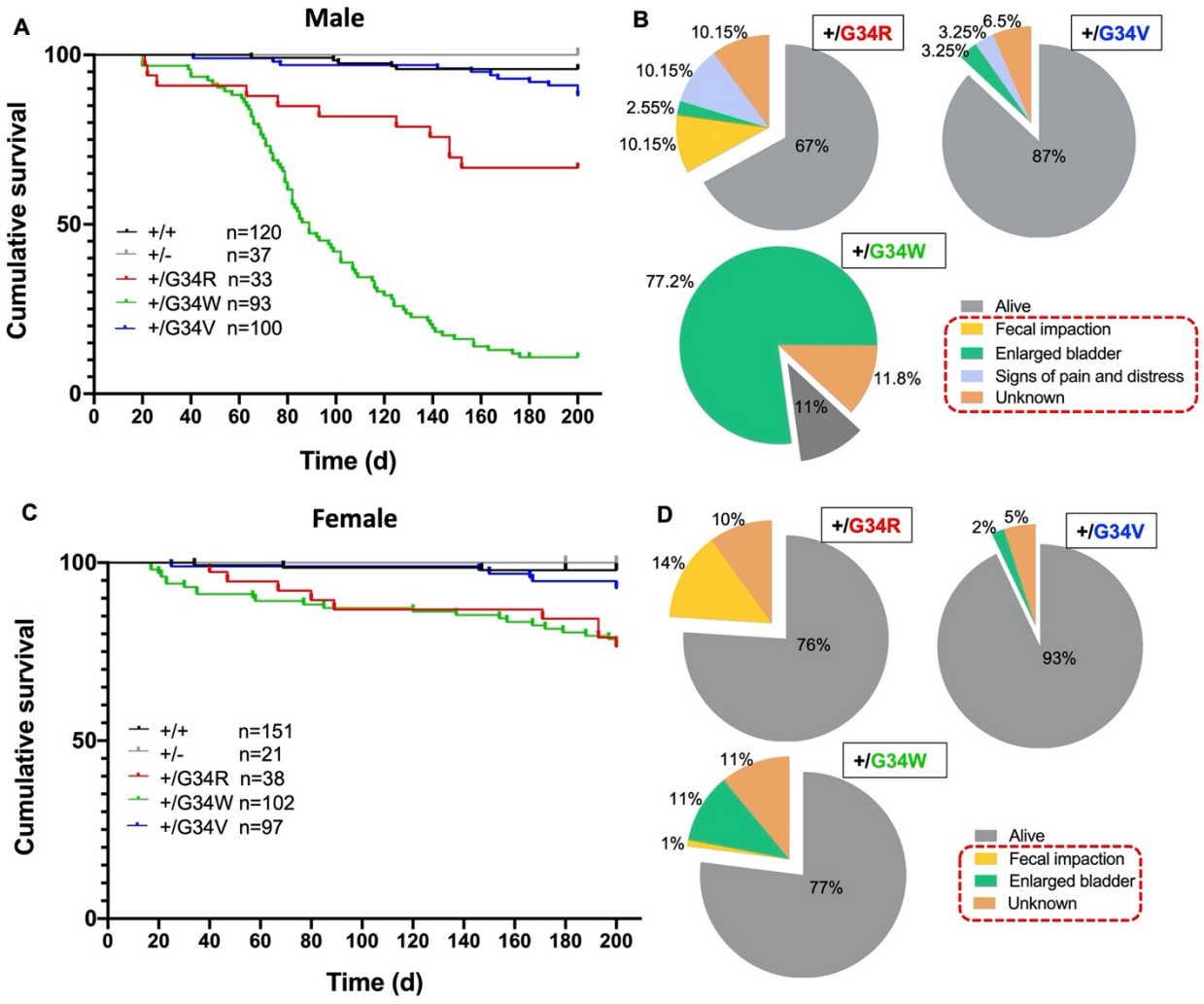


Figure 3. 5 Overall survival rates for the *H3f3a*+/G34 mutants and control mice

A and C) Kaplan-Meier curves for cumulative survival rates of *H3f3a*+/G34 mutants and control (WT and KO-het) in males (A) and female (C) mice.

B and D) Pie charts showing the causes of mortality of *H3f3a*+/G34 mutants and control (WT and KO-het) in males (B) and female (D) mice. Some *H3f3a*+/G34 mutants were found dead without an obvious reason, these mice were classified as unknown.

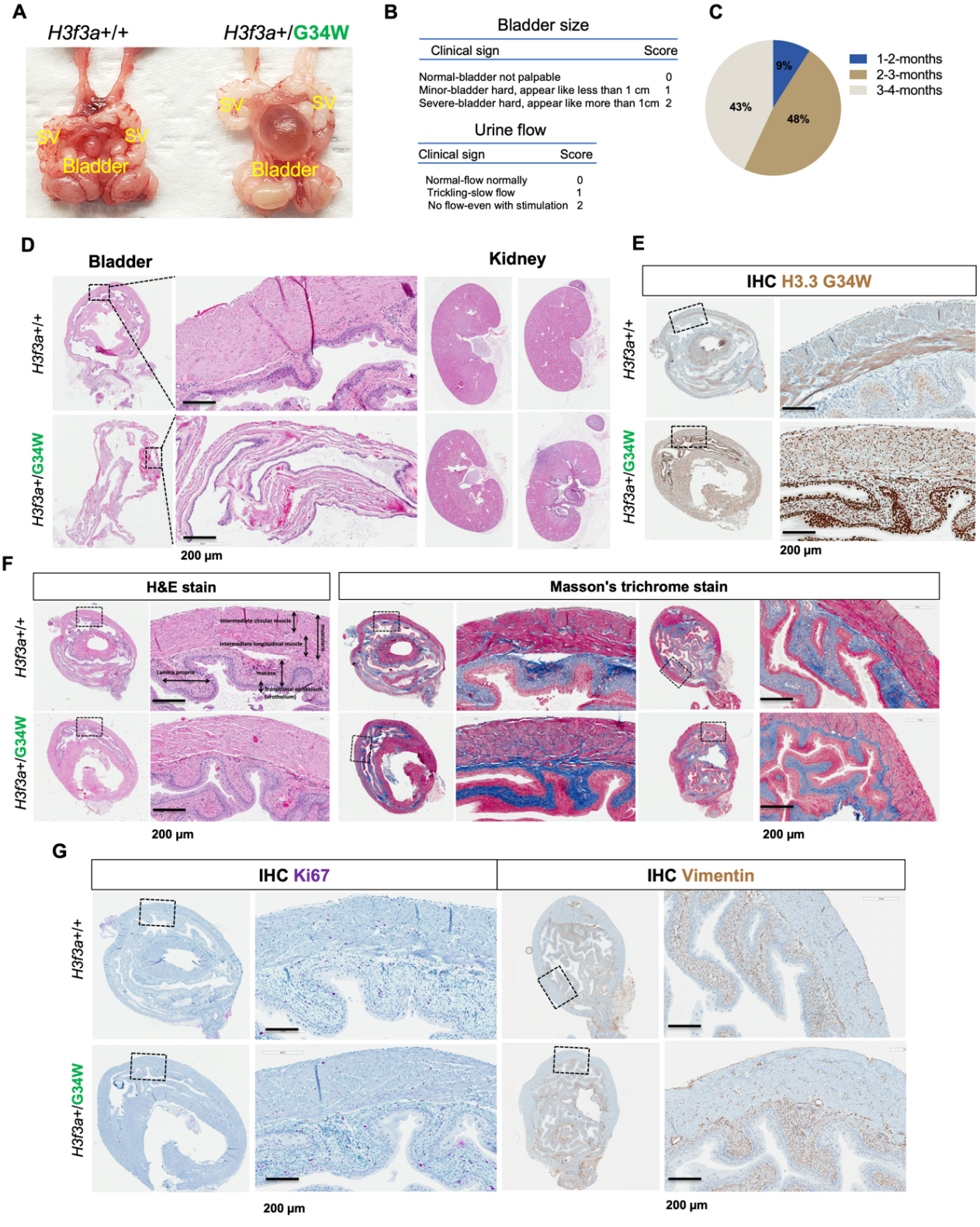

A
H3f3a+/+
H3f3a+/G34W
SV
SV
SV
SV
Bladder
Bladder
B
Bladder size
Clinical sign
Score
Normal-bladder not palpable 0
Minor-bladder hard, appear like less than 1 cm 1
Severe-bladder hard, appear like more than 1cm 2
Urine flow
Clinical sign
Score
Normal-flow normally 0
Trickling-slow flow 1
No flow-even with stimulation 2
C
9%
43%
48%
1-2-months
2-3-months
3-4-months
D
Bladder
Kidney
H3f3a+/+
H3f3a+/G34W
200 μm
E
IHC H3.3 G34W
H3f3a+/+
H3f3a+/G34W
200 μm
F
H&E stain
Masson's trichrome stain
H3f3a+/+
H3f3a+/G34W
Intermediate circular muscle
Intermediate longitudinal muscle
muscularis
Lamina propria
mucosa
Transitional epithelium (urothelium)
200 μm
200 μm
G
IHC Ki67
IHC Vimentin
H3f3a+/+
H3f3a+/G34W
200 μm
200 μm

Figure 3. 6 The mortality of male *H3f3a+/*G34W mice is associated with an enlarged bladder

A) A representative image of the bladder with the urogenital tracts from a control *H3f3a*+/+ (left panel) and *H3f3a+/*G34W (right panel). Notice in the latter the enlarged bladder filled with urine. SV is the seminal vesicle.

B) A scoring system based on bladder size and ability to urinate (Urine flow) in *H3f3a+/*G34W mice. Euthanasia is recommended when the score form both columns is 2.

C) Pie chart demonstrating the age distribution of *H3f3a+/*G34W mice who developed a score of 2 for both bladder size and urinating and were euthanized.

D) Representative images of H&E staining of a bladder (left panels), an enlargement of the bladder wall (middle panels) and a kidney (right panels) of a control mouse (*H3f3a+/+)* and from a *H3f3a+/*G34W male littermate that died with a severely distended bladder. The middle panel in the G34W mice shows a luminal distension, with an associated thinning of the bladder wall. In the right panel, in the G34W mouse the kidney shows signs of kidney hydronephrosis.

E) Immunohistochemistry (IHC) H3.3 G34W staining of a horizontal section of a control mouse (WT, *H3f3a*+/+) top panels, and *H3f3a+/*G34W (bottom panels) mice.

F-G) Histopathology of the urinary bladders *H3f3a*+/+ and *H3f3a+/*G34W mice. F) Representative images of H&E (left) and Masson's Trichrome-stained cross sections of the bladder tissues (right) of male *H3f3a*+/+ and *H3f3a+/*G34W mice at 3-weeks old. The Masson Trichrome stains bladder muscle red and collagen blue. G) IHC analysis of cross sections of bladder tissues of male *H3f3a*+/+ and *H3f3a+/*G34W mice at 3-weeks old. The left panels are stained with an antibody for Ki67, while the right panels show staining for vimentin (n=3 each genotype). The image scale bar of the enlarged picture: 200 µm.

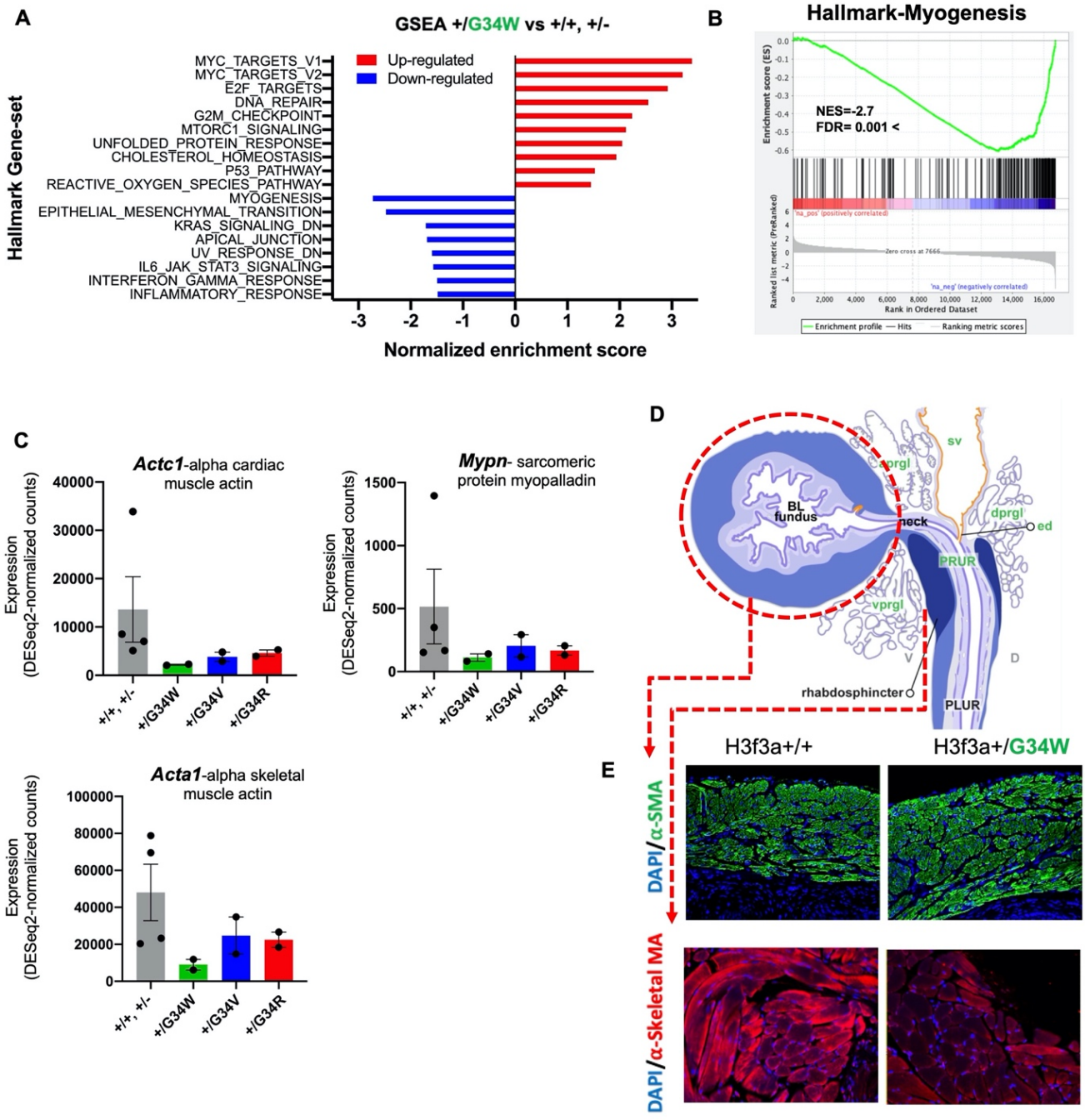

A
GSEA +/G34W vs +/+, +/-
Up-regulated
Down-regulated
Hallmark Gene-set
MYC_TARGETS_V1
MYC_TARGETS_V2
E2F_TARGETS
DNA_REPAIR
G2M_CHECKPOINT
MTORC1_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
CHOLESTEROL_HOMEOSTASIS
P53_PATHWAY
REACTIVE_OXYGEN_SPECIES_PATHWAY
MYOGENESIS
EPITHELIAL_MESENCHYMAL_TRANSITION
KRAS_SIGNALING_DN
APICAL_JUNCTION
UV_RESPONSE_DN
IL6_JAK_STAT3_SIGNALING
INTERFERON_GAMMA_RESPONSE
INFLAMMATORY_RESPONSE
-3 -2 -1 0 1 2 3
Normalized enrichment score
B
Hallmark-Myogenesis
Enrichment score (ES)
NES=-2.7
FDR= 0.001 <
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 7666
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
C
Actc1-alpha cardiac muscle actin
Expression (DESeq2-normalized counts)
Mypn- sarcomeric protein myopalladin
Acta1-alpha skeletal muscle actin
+/+, +/-
+/G34W
+/G34V
+/G34R
D
BL
fundus
neck
sv
vprgl
dprgl
ed
PRUR
vprgl
V
D
rhabdosphincter
PLUR
E
H3f3a+/+
H3f3a+/G34W
DAPI/α-SMA
DAPI/α-Skeletal MA

Figure 3. 7 Striated muscle defects and urethral obstruction lead to enlarged bladder and renal dysfunction in *H3f3a+/*G34W mice

A) Gene set enrichment analysis (GSEA) of RNA-seq data of bladder tissues from *H3f3a+/*G34W (n=2) mice compared with control (*H3f3a*+/+ (n=2), and *H3f3a*+/- (n=2)) mice at age P7, indicating significant hallmark gene sets (FDR < 0.05). Red: upregulated hallmark gene-sets and Blue: downregulated hallmark gene-sets. The normalized enrichment score and significance were computed using GSEA[190] .

B) Enrichment plot indicating strong negative correlation (blue) between differentially expressed genes (from bladders of *H3f3a+/*G34W vs control mice) and myogenesis target genes in the GSEA hallmark gene sets (3. 7A).

C) Bar graphs of expression levels of *Acta1*, *Actc1 and Mypn* in control *H3f3a*+/+ and *H3f3a*+/- (n=4), *H3f3a+/*G34W (n=2), *H3f3a+/*G34V (n=2) and *H3f3a+/*G34R (n=2) bladder tissues at P7. Gene expression levels are reported in median-of-ratios normalized read counts.

D) Schematics illustrate the sagittal section of mouse bladder pelvic urethra (image has been taken from Georgas et al. 2015[191]. Bladder (BL), prostatic urethra (PRUR), seminal vesicle (SV), prostate glands (anterior, dorsal, ventral).

E) Representative images of immune stained mouse bladder sections for α-SMA and (green) α-skeletal MA (red) antibodies of *H3f3a*+/+ and *H3f3a+/*G34W mice. DAPI (blue) was used for nuclear staining.

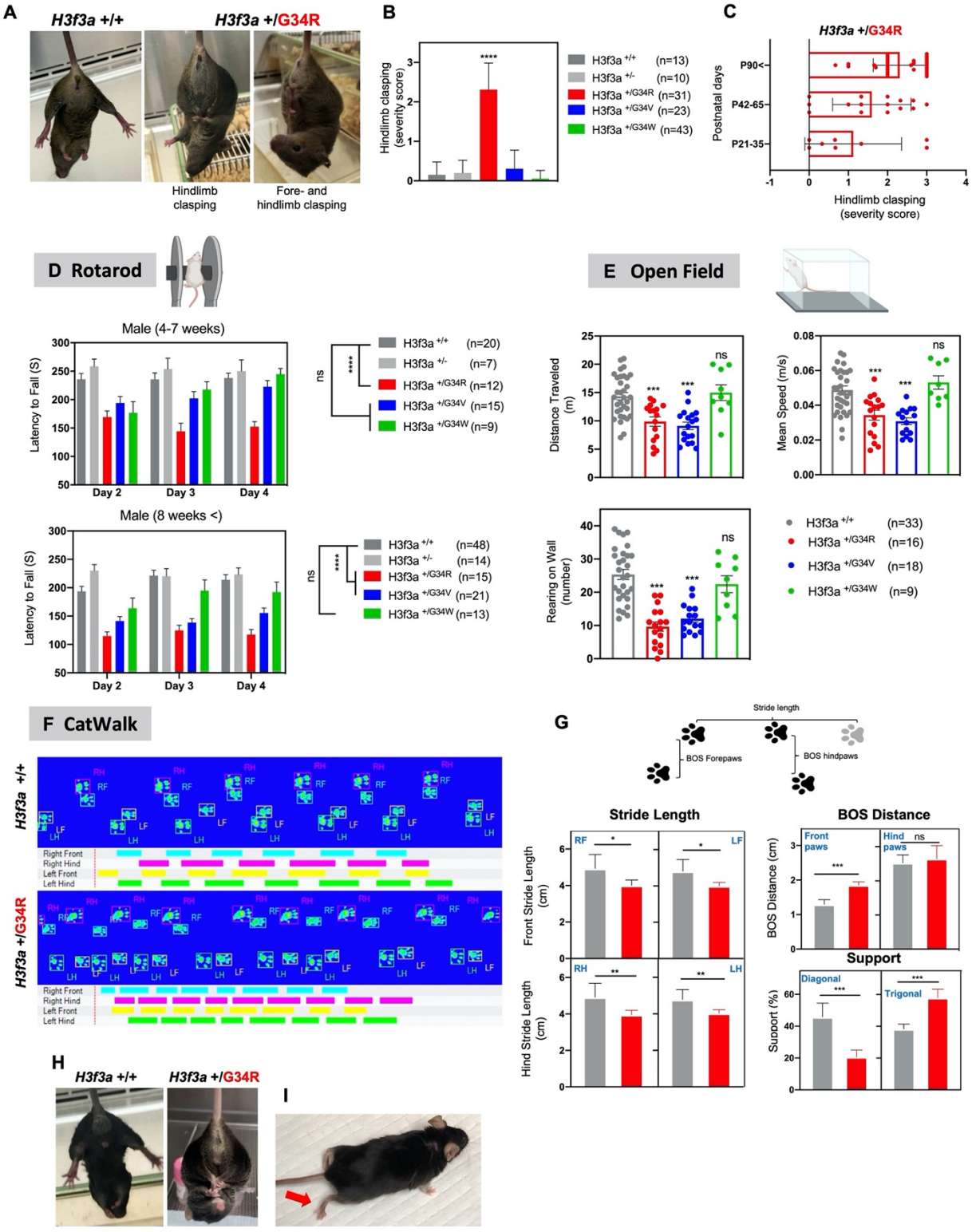

A
H3f3a +/+
H3f3a +/G34R
Hindlimb clasping
Fore- and hindlimb clasping
B
Hindlimb clasping (severity score)
H3f3a +/+ (n=13)
H3f3a +/- (n=10)
H3f3a +/G34R (n=31)
H3f3a +/G34V (n=23)
H3f3a +/G34W (n=43)
C
H3f3a +/G34R
Postnatal days
P90<
P42-65
P21-35
Hindlimb clasping (severity score)
D Rotarod
Male (4-7 weeks)
Latency to Fall (S)
Day 2
Day 3
Day 4
H3f3a +/+ (n=20)
H3f3a +/- (n=7)
H3f3a +/G34R (n=12)
H3f3a +/G34V (n=15)
H3f3a +/G34W (n=9)
Male (8 weeks <)
H3f3a +/+ (n=48)
H3f3a +/- (n=14)
H3f3a +/G34R (n=15)
H3f3a +/G34V (n=21)
H3f3a +/G34W (n=13)
E Open Field
Distance Traveled (m)
Mean Speed (m/s)
Rearing on Wall (number)
H3f3a +/+ (n=33)
H3f3a +/G34R (n=16)
H3f3a +/G34V (n=18)
H3f3a +/G34W (n=9)
F CatWalk
H3f3a +/+
H3f3a +/G34R
Right Front
Right Hind
Left Front
Left Hind
G
Stride length
BOS Forepaws
BOS hindpaws
Stride Length
Front Stride Length (cm)
Hind Stride Length (cm)
BOS Distance
BOS Distance (cm)
Front paws
Hind paws
Support
Support (%)
Diagonal
Trigonal
H
H3f3a +/+
H3f3a +/G34R
I

Figure 3. 8 *H3f3a*+/G34R mice show abnormalities in motor coordination

A) Representative images of WT and *H3f3a*+/G34R littermates exhibiting various degrees of limb clasping.

B) Graph scoring in the hindlimb clasping test for all mice at a young adult age (P30-90).

C) Hindlimb clasping reflex was scored in *H3f3a*+/G34R mice at age P21-35, P42-65, P90-death. Data from *H3f3a*+/+ control mice is zero and is not shown in this graph.

D) Rotarod test, a rotating rod on which the mice have to balance at increasing speed. Upper graph: time (latency to fall) that the mice were able to stand on the rod of all mice from age 4-7 weeks. Lower graph: time (latency to fall) that the mice were able to stand on the rod of all mice from age 8 weeks and older. The data is presented as means ± SD. The statistical significance assessed using two-way ANOVA. **** $P < 0.0001$ and n.s.: non- significant.

E) Open field test. The graphs show the total distance traveled by the mice, the average speed of the mice and the rearing activity against the wall of the open field. The result of each mouse is plotted individually and as means ± SD. Student's *t*-test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, **** $P < 0.0001$ and n.s.: non- significant compared to WT mice (+/+).

F) Representative digitized mice footprints and associated step cycles of each foot as recorded during the CatWalk test. The upper part shows the results of the control (*H3f3a*+/+) mice; the lower part shows the result of the *H3f3a*+/G34R mice.

G) CatWalk measurements of stride length, base of support of front-paws and hind-paws and diagonal and triagonal support. The graphs show the static data from control and *H3f3a*+/G34R mice at 8W of age.

H) Representative images of *H3f3a*+/G34R and WT littermate exhibiting limb clasping in *H3f3a*+/G34R mice at 12-months-old.

I) Representative image of a 12-month-old *H3f3a*+/G34R mouse demonstrating abnormal gaiting.

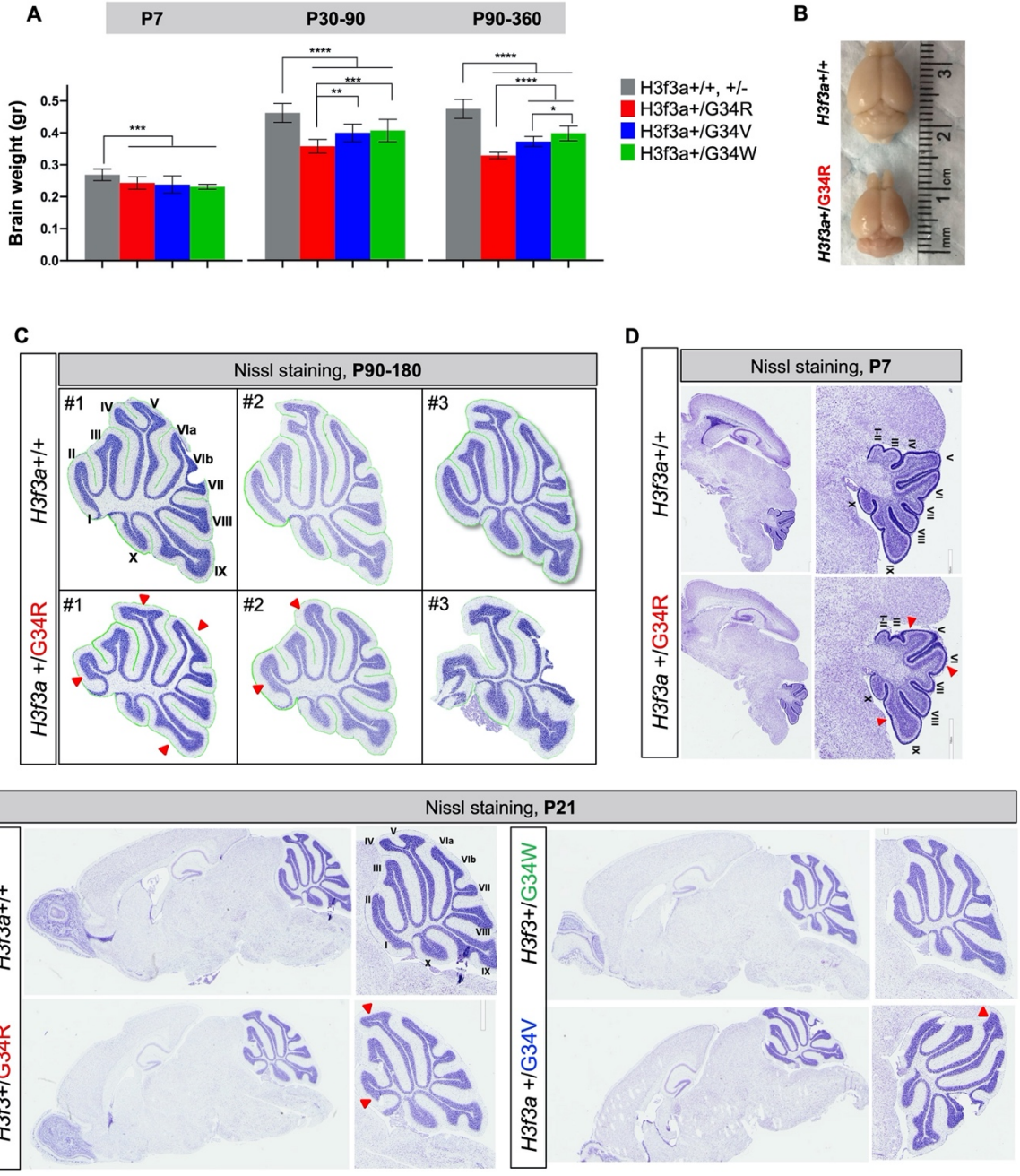

A
P7
P30-90
P90-360
Brain weight (gr)
0.0
0.1
0.2
0.3
0.4
0.5
***
****
***
**
****
****
*
H3f3a+/+, +/-
H3f3a+/G34R
H3f3a+/G34V
H3f3a+/G34W
B
H3f3a+/+
H3f3a+/G34R
C
Nissl staining, P90-180
H3f3a+/+
H3f3a +/G34R
#1
#2
#3
I
II
III
IV
V
VIa
VIb
VII
VIII
IX
X
D
Nissl staining, P7
H3f3a+/+
H3f3a +/G34R
E
Nissl staining, P21
H3f3a+/+
H3f3+/G34R
H3f3+/G34W
H3f3a +/G34V

Figure 3. 9 *H3f3a+/*G34R mice have defects in cerebellar foliation patterning and a progressive shrinking of brain

A) Brain weight of mice at various ages ranging from P7 to P360. For each age group at least 5 brains per genotype were used. The data is presented as means ± SD. The statistical significance assessed using one-way ANOVA. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, **** $P < 0.0001$ and n.s.: non- significant.

B) A representative image of brains of *H3f3a*+/+ and *H3f3a+/*G34R mice.

C-E) Representative images of midsagittal brain slices of control and *H3f3a+/*G34 mutant mice at different ages and stained with Cresyl Violet (Nissl Staining). I-X are the different lobes in control animals. Alterations in foliation patterning can be seen as a result of the G34 mutations and over time (C-E) at P7 (D), P21 (E) and P90-180 (C). Red arrowheads indicate poorly developed lobules or abnormal fissures in the cerebellum of mutant animals.

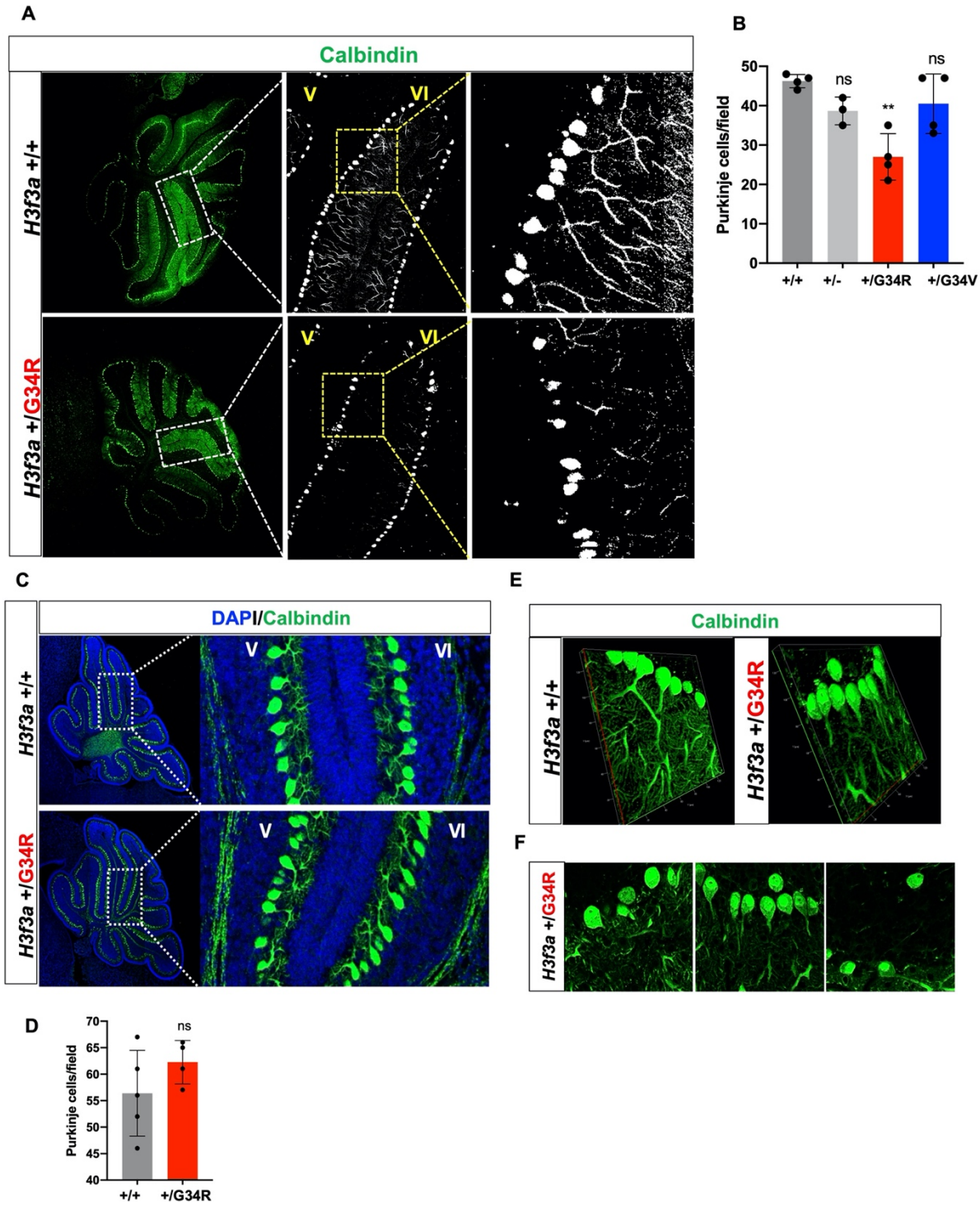

A
Calbindin
H3f3a +/+
H3f3a +/G34R
V
VI
B
Purkinje cells/field
ns
**
ns
+/+
+/-
+/G34R
+/G34V
C
DAPI/Calbindin
H3f3a +/+
H3f3a +/G34R
V
VI
D
Purkinje cells/field
ns
+/+
+/G34R
E
Calbindin
H3f3a +/+
H3f3a +/G34R
F
H3f3a +/G34R

Figure 3. 10 *H3f3a*+/G34R mice have partial to severe Purkinje cells loss in cerebellar lobules

A) Representative confocal images of cerebellar para-sagittal sections from 12-months-old mice of control and G34R mice. The immunofluorescence staining with a Calbindin-D28K antibody shows the Purkinje cells. The left, middle and right images magnifications are at 10X, 20X and 40X. The white hatched areas in the images were used for quantification of the number of cells.

B) Quantification of Calbindin-positive cell numbers in control (*H3f3a*+/+; n=4) and KO-het (*H3f3a*+/-; n=3) and mutants (*H3f3a*+/G34R; n=4) and *H3f3a*+/G34V; n=4). The data is shown as mean ± SD *p<0.05; **p<0.01, ns (not significant), comparing mutants to control *H3f3a*+/+.

C) Representative confocal images of cerebellar para-sagittal sections from P7 mice. The immunofluorescence staining with Calbindin-D28K antibody. Calbindin antibody (green) shows the Purkinje cells, while DAPI (blue) is used to visualize nuclei. The left and right images magnifications are at 10X, 40X. The white hatched areas in the images are used to perform a quantification.

D) Quantification of the Calbindin-positive cell numbers in *H3f3a*+/+ (n=5) and *H3f3a*+/G34R (n=4). The data are shown as mean ±SD, ns (not significant).

E) 3-dimensional confocal image of Purkinje cells labelled with Calbindin-D28K at mature adult age (3-6 Months).

F) Representative confocal images of immunofluorescence staining of Purkinje cells with Calbindin-D28K antibody in *H3f3a*+/G34R cerebellar sections at mature adult age (3-6 months).

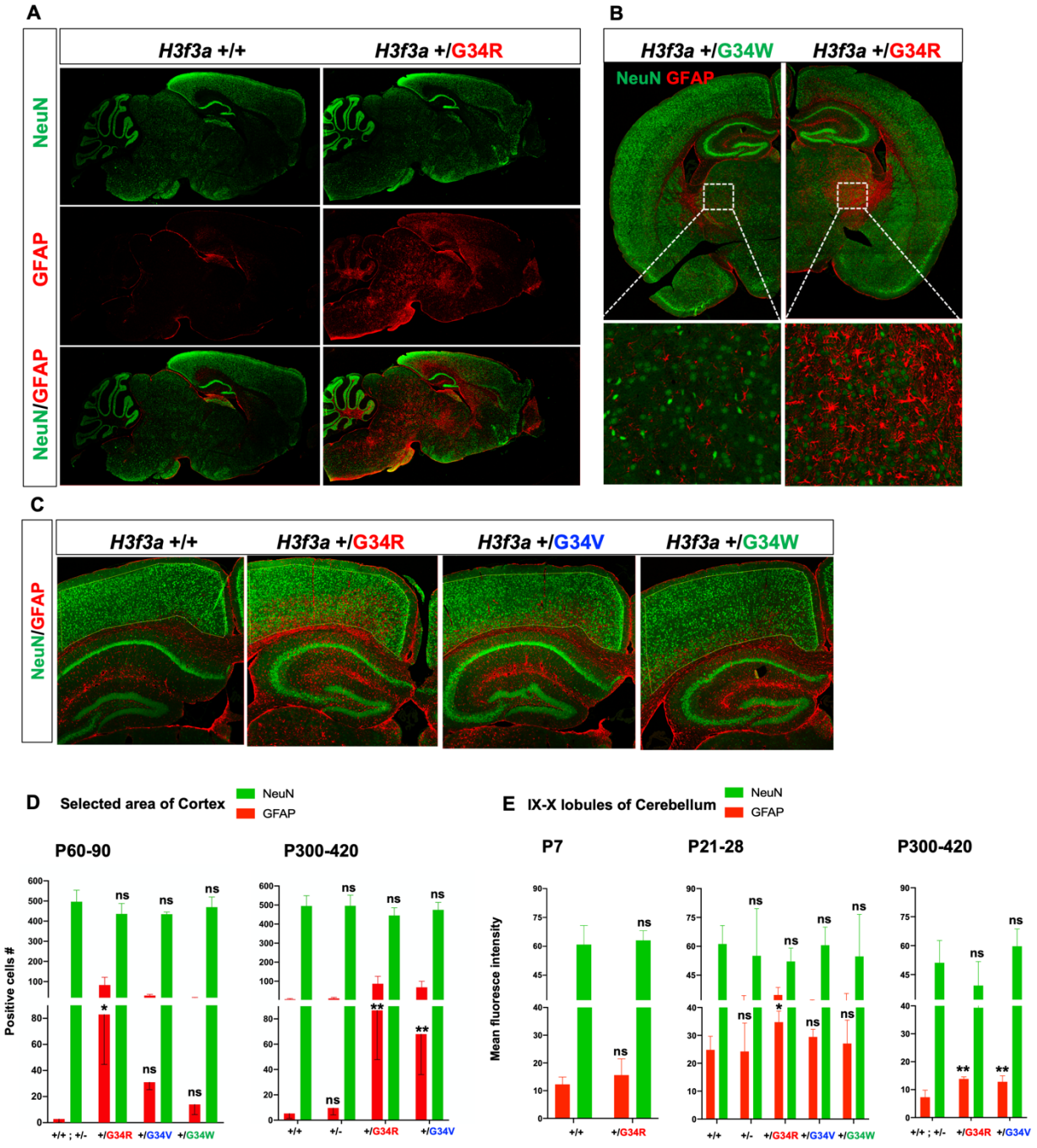

A
H3f3a +/+
H3f3a +/G34R
NeuN
GFAP
NeuN/GFAP
B
H3f3a +/G34W
H3f3a +/G34R
NeuN GFAP
C
H3f3a +/+
H3f3a +/G34R
H3f3a +/G34V
H3f3a +/G34W
NeuN/GFAP
D
Selected area of Cortex
NeuN
GFAP
P60-90
P300-420
Positive cells #
+/+ ; +/-
+/G34R
+/G34V
+/G34W
+/+
+/-
E
IX-X lobules of Cerebellum
NeuN
GFAP
P7
P21-28
P300-420
Mean fluoresce intensity

Figure 3. 11 GFAP+ cells are increased in the brain of H3f3a+/G34R animals

A-C) Representative confocal images of double immunofluorescence staining of neurons (NeuN) and astrocytes (GFAP) in the whole brain sagittal (A) and coronal (B) sections from 2-3-months old of control and *H3f3a*+/G34 mice. Scale bars: 1 mm (A&B). Under coronal image there is a higher magnification of the striatal area scale bars: 100 µm (B). Higher magnification images of coronal sections for the cortex area which used to perform quantification. Scale bars: 500 µm (C).

D) Quantitative analysis of GFAP and NeuN positive cells counted in selected area of the cortex of brains from P60-90 and P300-420 old control and mutant mice. Data are shown as mean ±SD. The statistical significance assessed using two-way ANOVA. *$p<0.05$; **$p<0.01$, ns (not significant), comparing mutants to *H3f3a*+/+ areas. For each genotype at least n=4 mouse was quantified.

E) Quantitative analysis of the mean fluorescence intensity of GFAP and NeuN staining in IX-X lobules of cerebellum from P7, P21-28, and P300-420 old control and mutant mice. Data are shown as mean ±SD. The statistical significance assessed using two-way ANOVA. *$p<0.05$; **$p<0.01$, ns (not significant), comparing mutants to *H3f3a*+/+ areas. For each genotype at least n=4 mouse was quantified.

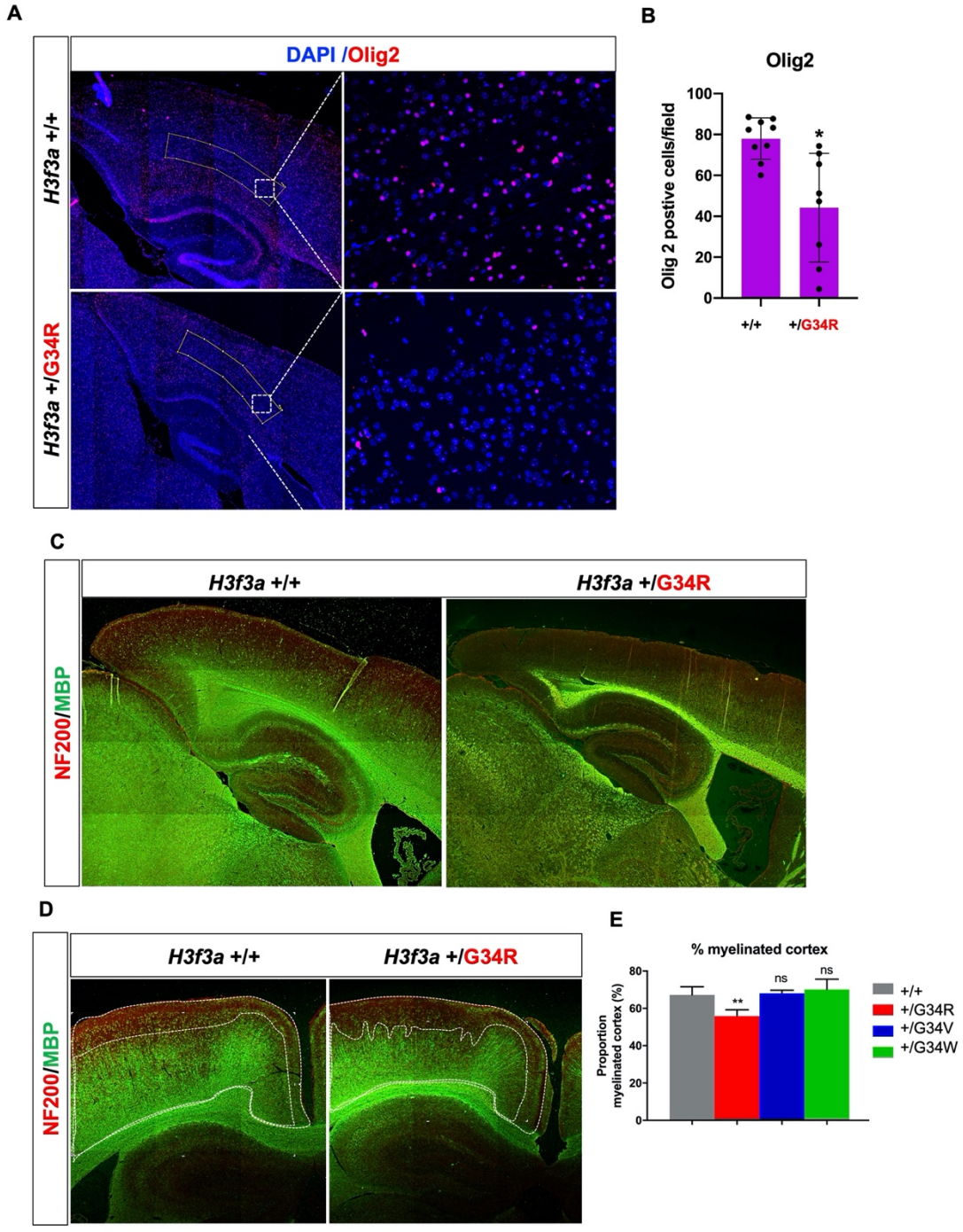

A
DAPI /Olig2
H3f3a +/+
H3f3a +/G34R
B
Olig2
Olig 2 postive cells/field
100
80
60
40
20
0
*
+/+
+/G34R
C
H3f3a +/+
H3f3a +/G34R
NF200/MBP
D
H3f3a +/+
H3f3a +/G34R
NF200/MBP
E
% myelinated cortex
Proportion myelinated cortex (%)
80
60
40
20
0
**
ns
ns
+/+
+/G34R
+/G34V
+/G34W

Figure 3. 12 *H3f3a*+/G34R mice show reduced Olig2+ cells and impaired myelination by aging

A) Representative confocal images of immunofluorescence staining of oligodendrocytes cells with Olig2 antibody in cortex sagittal sections from adult mice. Scale bars: 1 mm. Higher magnification images scale bars: 500 µm. Yellow hatched areas in the images used to perform a quantification.

B) Quantitative analysis of Olig2 positive cells counted in the selected area of the cortex of brains. *H3f3a*+/+ (n=9) and *H3f3a*+/G34R (n=8). Data are shown as mean ±SD *p<0.05.

C, D) Representative confocal images of immunofluorescence staining of axonal marker NF200 (red) and myelin marker MBP (green) at 12-months-old (C) and 2-3 months-old (D) of *H3f3a*+/+ and *H3f3a*+/G34R mice.

E) Quantitative analysis of the portion of myelinated cortex of control and mutant mice at 2-3-months-old Data represent mean ± SD. Student's *t*-test. $^{**}P < 0.01$ and n.s.: non- significant compared to WT mice (+/+).

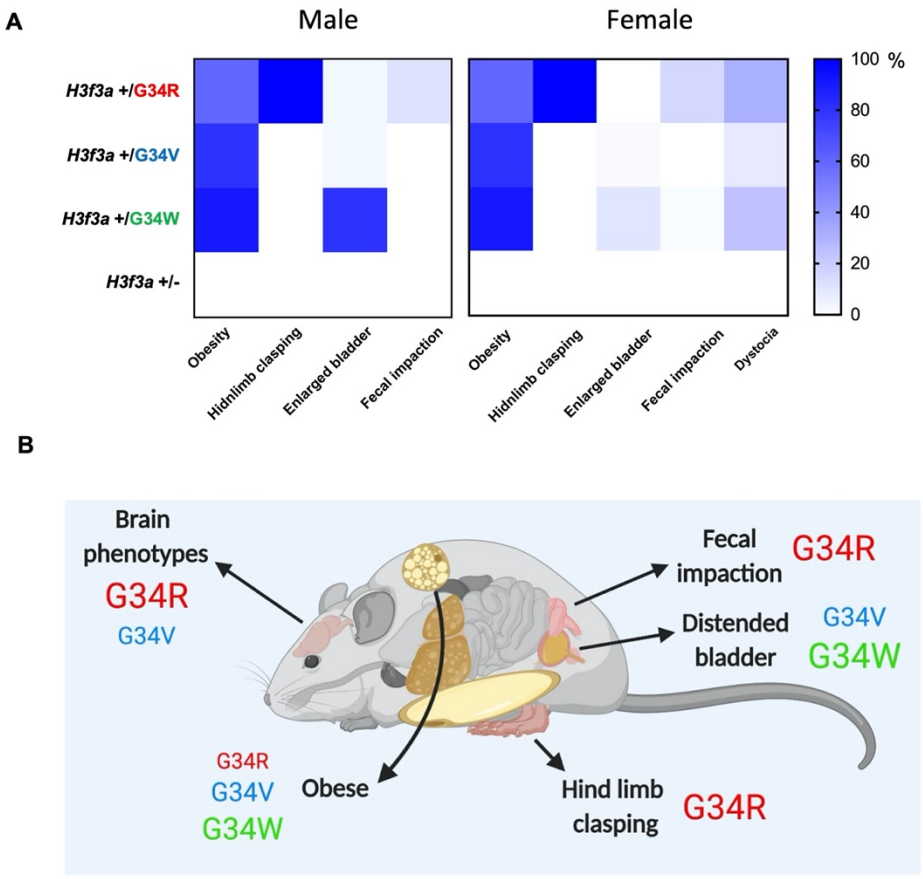


Figure 3. 13 Phenotype–tissue associations with G34 histone mutations

A) A heatmap of the different major phenotypes across *H3f3a* mutant mice.

B) A cartoon of a mouse with the effect of each G34 mutation (genotype) on the different organ (phenotype). The font size shows the severity of the phenotype seen for each genotype.

Chapter IV: H3.3 G34W promotes tumor growth and impedes differentiation of osteoblast-like progenitors in GCTB

In Chapter III, we showed that H3.3 G34 mutations caused distinct phenotypes in G34R or in G34W mutant mice and found that the phenotypes in G34W mutant mice are predominantly restricted to MSC-related cell lineages and tissues (e.g. adipose and muscle tissues). In this chapter, we will explore the role of the H3.3 G34W mutation in the mesenchymal tumor, GCTB. For this, we use isogenic GCTB cell lines to study to the effects of this mutation on the transcriptome and epigenome of these cells, as well as on tumorigenicity.

4.1. Correction of the H3.3G34W oncohistone mutation in human GCTB cell lines by CRISPR/Cas9 gene editing

To study the effects of H3.3 G34W mutation in an isogenic context, we used CRISPR/Cas9 to correct the G34W mutation to wildtype in established primary GCTB cell lines (Fig. 4.1A). GCTB contains three histologically different cell types: multinucleated, osteoclast-like giant cells; monocytic round macrophage-like cells; and spindle-shaped, fibroblast-like stromal cells[121,122]. Stromal cells are the only compartment that carry the H3.3 G34W mutation (Fig. 4.1B). Stromal cells remain the only compartment after low passages of GCTB tumor cells in culture[122]. However, these cells cannot be maintained in culture after several passages, which complicates the study of the role of the G34W mutation in neoplastic stromal cells. We therefore used retroviral infection of human telomerase reverse transcriptase (hTERT) to establish immortalized (ImGCT) cell lines from three patient tumors that match the typical GCTB profile in terms of patient age, tumor location (long bones) and G34W mutation status (Table 4.1).

We then used CRISPR/Cas9 with sgRNA to target the *H3F3A* G34W-mutation in these three ImGCT cell lines (Table 4.2). We successfully generated clones carrying loss-of-function indels on the H3.3G34W-mutant allele (G34W-KO), as well as clones where the G34 mutation was corrected to wild-type (G34-WT) (Fig. 4.1C-E).

4.2. G34W mutation promotes proliferation and colony formation

We noted that cell morphology and size were distinct between the H3.3G34W and CRISPR-edited cells, with spindle-like H3.3G34W cells transforming into larger edited cells (Fig. 4.2A-B). Furthermore, we performed proliferation and colony forming ability assays in culture for two isogenic cell lines (ImGCT-4072 and ImGCT-3504). Correction of the mutation significantly reduced cell proliferation (Fig. 4.2C-D) and colony forming ability (Fig. 4.2E-G), highlighting the proliferative potential conferred by H3.3G34W *in vitro*.

4.3. G34W is required for tumor formation and promotes osteoclast recruitment in GCTB

We next investigated H3.3G34W's effect on tumorigenicity *in vivo*. Orthotopic tibial and subcutaneous tissue injections of two out of three parental ImGCT lines (ImGCT-4072 and ImGCT-6176) in immunodeficient mice produced tumors after 2 and 10 months respectively (Table 4.2). To identify the effect of the G34W mutation on tumorigenicity, we injected luciferase-tagged unedited and parental H3.3G34W and edited (G34W-KO and G34-WT) ImGCT-4072 cells into tibial and subcutaneous tissue and monitored tumor growth weekly by bioluminescent imaging (Fig. 4.3.1A). Injection of the H3.3G34W cells, resulted in increased bioluminescence signal over time (Fig. 4.3.1A-C) and subsequent formation of aggressive tumors (Fig. 4.3.1D-E). However, tibial and subcutaneous injections of the edited clones did not yield tumors in mice for up to a year post-injection (Fig. 4.3.1F-G), underscoring an enduring role for the H3.3G34W mutation in GCTB tumorigenesis.

We compared the features of tumors in our xenograft model to that of GCTB. Unlike GCTBs which rarely metastasize, the tumors in our model were aggressive and prone to disseminate to the lungs, adjacent soft tissue and abdominal cavity (Fig. 4.3.1H-I). Orthotopic tibial xenograft tumors exhibited osteolytic properties typical of GCTBs, but with more bone destruction and extrusion into soft tissue, thereby resembling high-grade osteosarcoma (Fig. 4.3.2A-B).

A more detailed investigation revealed that murine osteoblast and osteoclast cells were recruited to H3.3G34W-positive stromal cells adjacent to bone tissue, characteristic of a reactive bone-remodeling process (Fig. 4.3.2A). We also observed eponymous giant multinucleated osteoclast cells within the neoplastic stromal cell mass extruding outside the bone (Fig. 4.3.2B). Moreover, we observed two distinct histological compartments within the xenograft tumor: one undifferentiated and highly proliferative (increase the numbers of Ki-67-positive cells), while the other resembled differentiated GCTB stromal cells (Fig. 4.3.2B). Although sparsely distributed, giant cells (TRAP+) were most abundant in differentiated regions of the tumor (Fig. 4.3.2B). Furthermore, we classified the H3.3G34W subcutaneous tumors as resembling high-grade undifferentiated sarcomas based on several immunohistochemical (IHC) markers (Fig. 4.3.2C-D).

Taken together, these results show that H3.3G34W is required for tumor formation and that giant osteoclasts were recruited to xenograft tumors.

4.4. H3.3G34W is associated with global epigenetic remodeling in addition to *in cis* epigenetics effects

H3.3G34 mutations are proposed to modify the epigenome through a loss of H3K36me3[39] and a gain of H3K27me3[40] *in cis* on the mutant histone. We therefore quantified in four primary *H3F3A+/G34*W GCTB cell lines the abundance of H3K27 and H3K36 methylation using histone mass spectrometry (hMS) (Table 4.2).

The H3.3G34W mutation accounts for 25.7% of the total H3.3 and contributes only 2.6% of the total H3 pool (Fig. 4.4A). However, on mutant H3.3G34W histones compared to wild-type H3.3 histones, we observed a significant loss of H3.3K36me3 and H3.3K27me1 (marks known to co-occur in actively transcribed gene bodies) (Fig. 4.4B)[43]. Moreover, the mutant H3.3G34W containing histones also exhibited a significant increase in H3.3K36me2 (Fig. 4.4B). This was not found in previous studies[39,40] and potentially reflects the unique cellular context of GCTB stromal cells. We also observed a significant gain of the repressive H3.3K27me2/3 marks on the H3.3G34W histones (Fig. 4.4B). The latter could be explained by the fact that H3K36me3-depleted nucleosomes are a better substrate for PRC2 than H3K36me3-enriched ones[62,192].

To determine whether the G34W mutation impacts histone incorporation into chromatin, and to assess genome-wide epigenetics effects of the G34W mutation, we profiled H3.3, H3.3G34W, H3K36me3 and H3K27me3 by ChIP-sequencing in the isogenic G34W and edited lines.

We found that H3.3 and H3.3G34W showed highly similar genome-wide deposition patterns (Fig. 4.4C), which indicates that H3.3-specific histone chaperones load H3.3G34W in the same regions as wildtype H3.3. Furthermore, principal component analysis (PCA) confirmed that both H3K27me3 and H3K36me3 show distinct genome-wide deposition profiles in G34W compared to edited lines (Fig. 4.4D). This led us to speculate about a global epigenetic remodeling process initiated by the G34 mutation. Moreover, in two isogenic GCTB cell lines, we observed a genome-wide redistribution of H3K27me3: in the presence of G34W, the loss of H3K27me3 occurred preferentially in intergenic regions ($P<0.001$; $\chi 2$ test), whereas gain of the mark was enriched equally in promoter, intragenic and intergenic regions (Fig. 4.4E-F). This redistribution of H3K27me3 could result from recruitment of the PRC2 complex to the new H3K36me3-depleted areas in genic regions and away from poor substrates in intergenic areas.

These findings of the global epigenetic disruption due to G34W mutation, in concert with cell morphological, functional and tumorigenic differences, are highly suggestive of a H3.3G34W-mediated transformation of transcriptional programs, cell identity, and differentiation status.

4.5 H3.3G34W is associated with dysregulation of extracellular matrix organization and muscle contraction pathways

Because our results described above suggest that G34W mediates global epigenetic changes that may impact transcriptional programs and cell identity, we studied the effects of H3.3G34W on genes associated with tumorigenicity and stromal cell identity. For this we profiled the whole transcriptome of the three G34W-mutant and -edited isogenic GCTB cell lines (ImGCT-4072, ImGCT-3504 and ImGCT-6176).

The transcriptomic profiles between mutant G34W and edited lines were distinct in all three Im-GCT isogenic cell lines, as can be seen in the PCA plots (Fig. 4.5A). Pathway enrichment

analysis of differentially expressed genes (DEGs) in two out of three isogenic GCTB cell lines showed enrichment of extracellular matrix (ECM) organization genes in G34W cells compared to edited cells (Fig. 4.5B). Some of the important genes in this pathway that are significantly up-regulated in G34W cells are the collagen VI encoding genes (*COL6A1*, and *COL6A3*), and elastin coding gene (*EMILIN-2*) (Fig. 4.5C). These fibrous proteins, in addition to their structural role in the ECM, also play an important role in cell adhesion and regulation of cell differentiation[193].

Moreover, in two out of three isogenic Im-GCT cell lines pathways related to actin filament processes and muscle contraction were strongly depleted in G34W cells (Fig. 4.5B). Furthermore, intersection of commonly dysregulated genes across the three Im-GCT isogenic cell lines revealed 27 common DEGs, which were all downregulated in G34W cells compared to edited cells (Fig. 4.5D). Pathway enrichment analysis of these common DEGs showed enrichment of pathways such as actin filament-based process and muscle contraction, with related genes including *TNNT2*, *MYL1*, and *LMOD1* (Fig. 4.5E-F). *TNN2*- troponin T2, is a tropomyosin protein which regulates muscle contraction, and *MYL1*- myosin light chain 3, is required for proper skeletal muscle function (formation/maintenance of myofibers)[194], and *LMOD1*- leiomodin 1, is a member of the actin filament nucleator family[195].

Notably, we found in Im-GCTB isogenic cell line, genes with a chromatin state consistent with the *in cis* effect of G34W mutation (H3K36me3 loss and H3K27me3 gain) were enriched in pathways related to actin-myosin contractile functions (such as striated muscle thin filament, myofilament and sarcomere pathways) (Fig. 4.5G-H). Other downregulated genes in these pathways coupled with epigenetic changes are *ACTC1*- alpha cardiac muscle actin and *TNN11*- Troponin I, slow skeletal muscle, associated with actin filament-based process/myosin-actin interaction (Fig. 4.5I). By contrast, up-regulated genes encoding extracellular matrix components did not show a consistent epigenetic trend (Fig. 4.5J). In addition, we also found that pathways related to muscle contraction were depleted in a previously published cohort of primary G34W cell lines compared to WT cell lines[122].

Together, the expression signature of H3.3G34W cells suggest that H3.3G34W potentially downregulates actin myosin contractile filament genes.

4.6.1. GCTB stromal cells comprise of *SPP1*+ and *ACTA2*+ sub-populations

GCTB stromal cells with the G34W mutation are not well defined and have been alternately described as possessing properties of mesenchymal stem cells, fibroblasts, or pre-osteoblasts based on *in vitro* differentiation assays and histological studies[125]. To better define this stromal component, we performed single-cell RNA-sequencing (scRNA-seq) on four G34W GCTB tumors (Table 4.2). Clustering analysis based on similarity to expression profiles from cell types within the Human Primary Cell Atlas and known specific markers (data not shown) clustered GCTB cells into lymphoid, myeloid, endothelial, and putative stromal cells (Fig. 4.6.1A). In contrast to other clusters which contained cells from all patient tumors, the putative stromal cell clusters segregated by patient, indicating increased variation within this compartment likely reflecting divergent clonal evolution. Furthermore, these clusters were significantly enriched for the *H3F3A* G34W mutation (Fig. 4.6.1B) and displayed the highest enrichment score for a transcriptomic G34W signature derived from our independent isogenic model (Fig. 4.6.1C). They also showed similar pathway enrichment as isogenic G34W stromal cells (e.g. ECM organization, extracellular structure organization) (Fig. 4.5B). We therefore conclude that these clusters represent the neoplastic stromal compartment of primary GCTB tumors as analyzed at single cell resolution.

As expected, the mononuclear stromal component was detected in all of our samples. To query potential neoplastic stromal subpopulations shared by all samples and to reduce the impact of technical and biological inter-tumoral stromal cell heterogeneity we used Harmony, a software package to integrate multiple data sets (https://portals.broadinstitute.org/harmony/index.html). Clustering analysis identified four clusters within the stromal compartment (Fig. 4.6.1E-G), with *SPP1*-osteopontin and *ACTA2*-smooth muscle actin as the most discriminative markers. Based on the expression modules of genes correlated with *SPP1* or *ACTA2* (Fig. 4.6.1E), three stromal populations were present in all GCTB tumors (Fig. 4.6.1F). S1 cells showed expression of genes associated with osteoblast and chondrocyte functions (e.g. *SPP1*, *IBSP*-integrin binding sialoprotein, *MMP13*-matrix metallopeptidase 13), whereas S3 cells expressed markers of contractile cells (e.g. *ACTA2*-alpha smooth muscle actin, *TAGLN*-transgelin, *POSTN*-

periostin). S2 had fewer differentially expressed genes and was considered an intermediate between both states (Fig. 4.6.1G).

We further confirmed *SPP1*-osteopontin (OPN) and *ACTA2*-smooth muscle actin (SMA) expression by IHC in two tibial xenografts and two patient tumors (Fig. 4.6.1H). Together, these IHC results support and validate our scRNA-seq findings. Importantly, we provide a cohesive explanation for disparate IHC findings showing simultaneous expression of osteoblastic and myofibroblast-like markers in GCTBs. Using scRNA-seq, we show that osteoblast and myofibroblast-like markers are expressed in distinct stromal cell populations linked by a differentiation trajectory.

4.6.2. GCTB stromal cells resemble osteoprogenitors with features of contractile cells

Lineage inference analysis using Slingshot [162] and Velocyto [163], to identify developmental trajectories in scRNA-seq data revealed two potential trajectories among the identified four clusters: namely from S1A to S1B cells, or from S1A to S3 cells (Fig. 4.6.2A). This could suggest a continuum of stromal states within G34W-mutant cells, either by transitioning between the early S1A to the S1B states, or by following a trajectory from a progenitor S1A state towards a myofibroblast-like S3 state.

To identify the cell of origin of the neoplastic stromal cells, we next mapped the transcriptional profiles of the S1-S3 populations to two recently published scRNA-seq murine bone-marrow stroma Baryawno and Tikhonova datasets[166,167]. We found that the S1, S2, and S3 populations strongly mapped to a specific kind of osteoprogenitor in both datasets (Fig. 4.6.2B-C). Notably, S1 and S2 populations mapped more strongly than S3 cells to the less-differentiated “osteoprogenitor” or “preosteoblasts" reference subsets[166,167](Fig. 4.6.2B-C). which is consistent with their earlier position in the temporal trajectory. These findings indicate that GCTB neoplastic stromal cells resemble osteoprogenitors or a closely related cell type, with some cells (S3) progressing to display contractile features.

To validate the myofibroblast differentiation potential of GCTB stromal cells, we induced differentiation in isogenic GCTB lines (G34W and edited) and in control human mesenchymal stromal cells (hMSCs). After induction, GCTB cells expressed high levels of myofibroblast markers including stress fiber-associated protein calponin 1 (*CNN1*) (Fig. 4.6.2D-E) and alpha smooth muscle actin (*ACTA2*) (Fig. 4.6.2F). Notably, while both G34W and edited cells were able to differentiate into myofibroblasts, the edited cells already exhibited higher expression of calponin 1 at baseline relative to the G34W lines (Fig. 4.6.2D-E). This is consistent with the transcriptomic data that shows an enrichment of actin-myosin contractile genes in the edited cells.

Together, these findings reinforce that neoplastic stromal cells transition from S1 to S3, while the epigenetic reprogramming mediated by G34W appears to impede differentiation into a myofibroblast-like state.

4.7. G34W *ACTA2*+ stromal cells secrete factors promoting ECM remodeling and are association with myeloid cells.

The formation of giant multinucleated histone wild type osteoclasts is an essential contributor to the pathogenesis of GCTB tumors and is thought to be mediated by G34W-dependent secretion of factors such as RANKL (*TNFSF11*)[128]. In contrast to previous findings[122], we did not find that RANKL nor its decoy receptor OPG (*TNFRSF11B*) were differentially expressed between G34W and edited clones in three isogenic GCTB cell lines (Fig. 4.7A). We therefore conclude that the RANK/RANKL signalling, while required for osteoclastogenesis, is independent of G34W. We therefore investigated whether G34W-mediated aberrant tumor microenvironment (TME) interactions could contribute to the giant cell phenotype.

To identify ligands secreted by G34W stromal cells, we profiled the secreted proteome by isolating Golgi apparatus of isogenic GCTB lines and performing mass spectrometry (MS) (Fig. 4.7B-C). We identified significantly differentially secreted proteins between G34W and edited lines (Fig. 4.7B-C). Ligand-receptor gene interactions were inferred between stromal cells and the myeloid cell compartment in scRNA-seq data based on curated protein interaction databases[168] (Fig. 4.7D). Filtering these predicted ligand-receptor interactions for proteins that were

differentially secreted by the G34W lines, we observed stromal cell-specific expression of ECM ligands such as collagens (*COL6A1/3, COL5A2)* and proteoglycans (*BGN*) that are predicted to interact with specific integrin receptors on monocyte/macrophage (*ITGB2*) and osteoclast (*ITGAV*) cells (Fig. 4.7D). Expression of collagen VI (*COL6A1/3*) and BGN was further confirmed using immunohistochemistry in primary GCTB tumors as well as in orthotopic xenograft tumors (Fig. 4.7E). Interestingly, when intersecting the secretome and genes expressed in S1 to S3 populations, mainly the S3 cells were enriched for expression of G34W-specific secreted proteins (Fig. 4.7F). Indeed, 6 of the 24 genes specifically expressed by S3 cells are found in the ECM secretome, suggesting that S3 stromal cells are responsible for G34W-driven ECM remodeling.

Taken together, our results indicate that G34W containing stromal cells resemble osteoblast progenitors. These cells give rise to a myofibroblast-like cellular population (S3) that secretes ECM remodeling proteins that promote bone destruction by acting locally on myeloid cells in the TME. Therefore, G34W drives both components of GCTB pathogenesis: it sustains the neoplastic transformation of mononuclear stromal cells, as shown by our *in vivo* orthotopic model, and simultaneously enables the recruitment and formation of pathological giant osteoclasts that largely contribute to the morbidity of this tumor.

4.8. Overexpression of H3.3G34W in CRISPR-edited lines reverts some of the epigenetic, transcriptomic, and tumorigenic effects

Lastly, we overexpressed (O/E) hemagglutinin (HA)-tagged H3.3G34W and H3.3WT constructs using a lentiviral system in two ImGCT-4072 edited clones (Fig. 4.8A). There was a significant reduction in both tumor penetrance (2/7 mice) and latency of tumor formation upon implanting cell lines O/E H3.3 G34W (9-10 months) compared to G34W lines (2-3 months) (Fig. 4.8B-D), suggesting the CRISPR-editing promoted a level of terminal differentiation in edited lines. However, the pathology of tumors overexpressing H3.3 G34W corresponded to that of the parental G34W tumor, that is they are both characterized by a diffuse proliferation of G34W-positive undifferentiated and primitive round to spindle cells, with the presence of rare, multinucleated osteoclasts (Fig. 4.8C insert). In addition, one tumor was observed in edited lines O/E H3.3 WT with further delayed (by ~2 months) which displayed very different

histopathological features compared to mutant tumors. This tumor mainly extruded outside bone and proliferated in the adjacent soft tissue compartment (intact bone structure).

To address whether reintroduction of the mutation reverts the effects caused by CRISPR-editing, we profiled the transcriptomes of edited lines O/E H3.3 G34W and H3.3 WT by RNA-seq. Edited lines O/E H3.3 G34W showed transcriptional similarity to their parental edited lines, likely due to insufficient time for epigenetic changes to take effect, or due to irreversible terminal differentiation after CRISPR-editing (Fig. 4.8E).

To address whether G34W reintroduction reverts the epigenetic effects caused by CRISPR-editing, we assessed H3K27me3 distribution between edited lines O/E H3.3G34W and H3.3 WT. Investigation of H3K27me3 levels showed that the edited lines O/E H3.3G34W and H3.3 WT are separated first by clone (PC1: 53% of variance), thereby retaining the epigenetic features of their parental edited clone. However, PC2 (32%) separates by genotype. Although differences in the O/E model were more modest than between parental and edited lines, we were still able to observe a shift of H3K27me3 from intergenic to promoter and genic regions (Fig. 4.8F).

This reversion of redistribution of H3K27me3 in edited lines where H3.3G34W is reintroduced strongly suggests a G34W-specific effect as opposed to one driven merely by clonal variation.

Table 4. 1 Sample characteristics: tumor metadata

Patient ID	Genotype	Sex	Age	Tumor location	Metastases	Recurrence after surgery
P-4072	H3.3 G34W	M	22	R tibia	N	Y
P-3504	H3.3 G34W	M	30	L femur	N	Y
P-6176	H3.3 G34W	M	28	R femur	-	-
P-3474	H3.3 G34W	M	42	R femur	N	N
P-5071	H3.3 G34W	F	29	R radius	-	-
P-6028	H3.3 G34W	M	31	R femur	-	-
P-5874	H3.3 G34W	M	44	L tibia	-	-
P-6027	H3.3 G34W	M	31	R ulna	-	-
P-4671	H3.3 G34W	F	49	L tibia	-	Y
P-4724	H3.3 G34W	F	32	L sacrum	-	-
P-4779	H3.3 G34W	F	34	L tibia	-	-
P-2611	H3.3 G34W	-	-	-	-	-

Table 4. 2 Sample characteristics: assays

Patient ID	Immortaliz tion (Cell Line)	Xenograft (Cell Line)	RNA-seq (Cell Line)	ChIP-seq (Cell Line)	scRNA-seq (Tumour)	snRNA-Seq (Xenografts)	Histone Mass Spec (Cell Line)	Secretome Mass Spec (Cell Line)
P-4072	Y	Y	Y	Y	-	Y	Y	Y
P-3504	Y	N	Y	Y	-	-	-	-
P-6176	Y	Y	Y	-	-	Y	-	-
P-3474	-	-	-	-	-	-	-	-
P-5071	-	-	-	-	Y	-	-	-
P-6028	-	-	-	-	Y	-	-	-
P-5874	-	-	-	-	Y	-	-	-
P-6027	-	-	-	-	(#)	-	-	-
P-4671	-	-	-	-	Y (*)	-	-	-
P-4724	-	-	-	-	-	-	Y	-
P-4779	-	-	-	-	-	-	Y	-
P-2611	-	-	-	-	-	-	Y	-

* scRNA-seq of P-4671 was performed on the recurrence.
# scRNA-seq of P-6027 was excluded due to failing to meet single cell quality inclusion criteria, see methods.

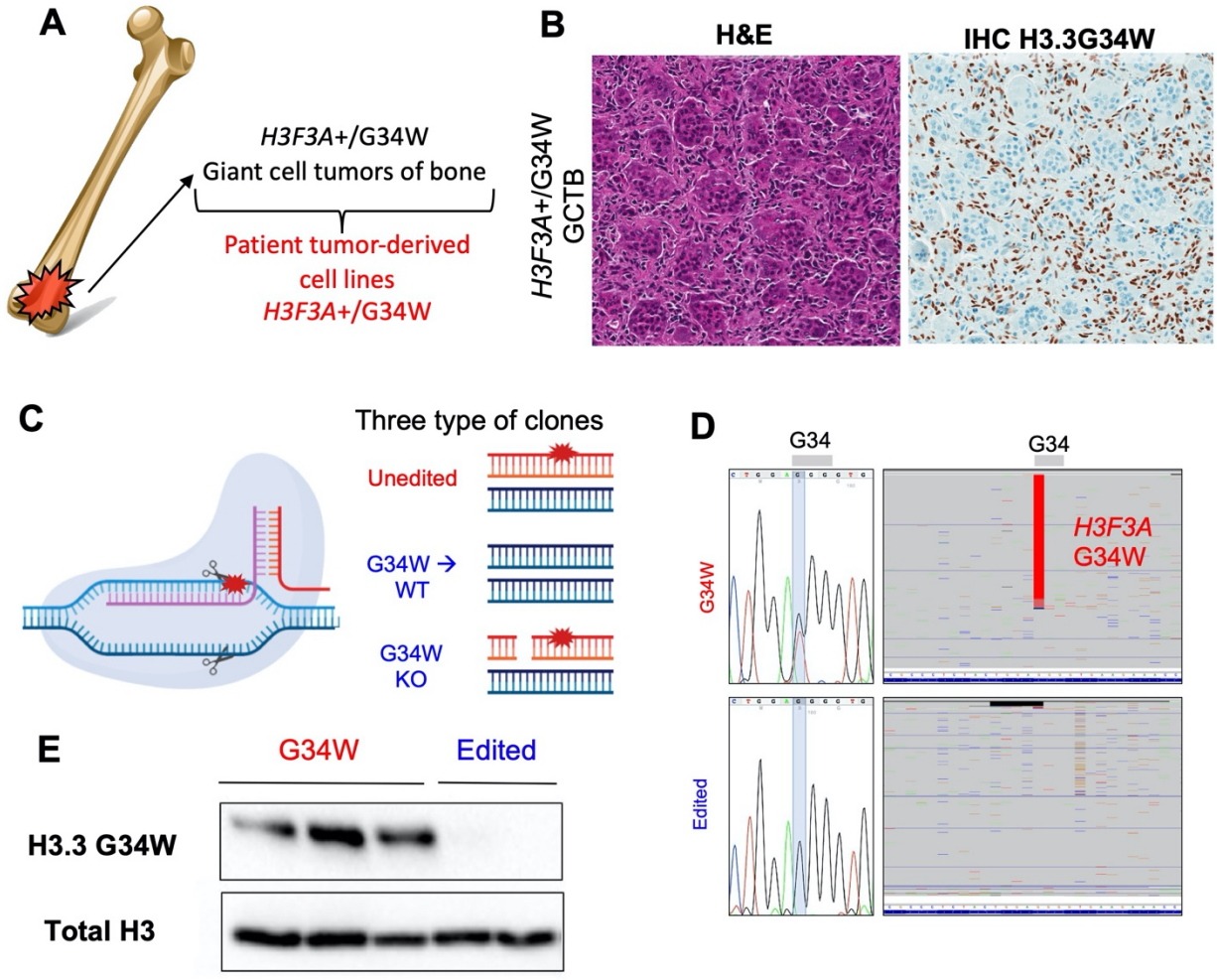


Figure 4. 1 Correction of H3.3G34W oncohistone mutation in GCTB cell lines by CRISPR/Cas9 gene editing

A) Schematic representation of bone with *H3F3A*+/G34W giant cell tumor of bone (GCTB).

B) Representative H&E and IHC G34W staining of a typical patient GCTB.

C) Schematic illustrating CRISPR/Cas9 strategy of targeting G34W mutation to derive edited clones (G34-WT and G34W-KO).

D) Sanger and MiSeq sequencing of representative ImGCT-4072 G34W and repair to WT clones.

E) G34W immunoblotting of ImGCT-4072 G34W (Parent: n=1; Clone: n=2) and edited (repair to WT: n=2) clones.

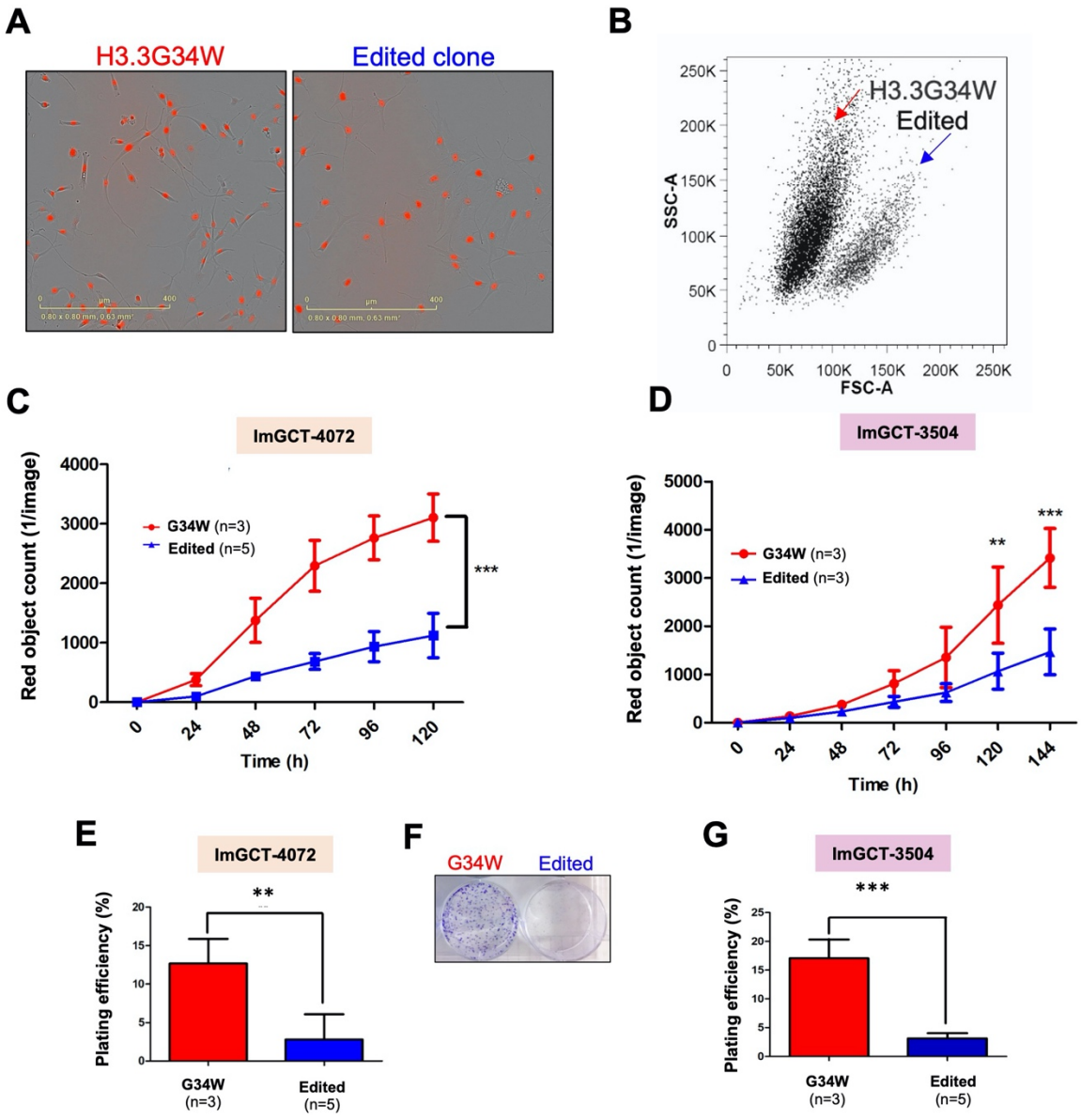

A
H3.3G34W
Edited clone
B
H3.3G34W
Edited
SSC-A
FSC-A
C
ImGCT-4072
Red object count (1/image)
G34W (n=3)
Edited (n=5)
***
Time (h)
D
ImGCT-3504
Red object count (1/image)
G34W (n=3)
Edited (n=3)
**
***
Time (h)
E
ImGCT-4072
Plating efficiency (%)
**
G34W (n=3)
Edited (n=5)
F
G34W
Edited
G
ImGCT-3504
Plating efficiency (%)
***
G34W (n=3)
Edited (n=5)

Figure 4. 2 G34W mutation promotes proliferation and colony formation

A) Representative Incucyte imaging illustrating morphological differences between ImGCT-4072 H3.3G34W and edited cells labeled with the nuclear NucLight stain.

B) FACS Analyses of ImGCT-4072 H3.3G34W and edited cell Size.

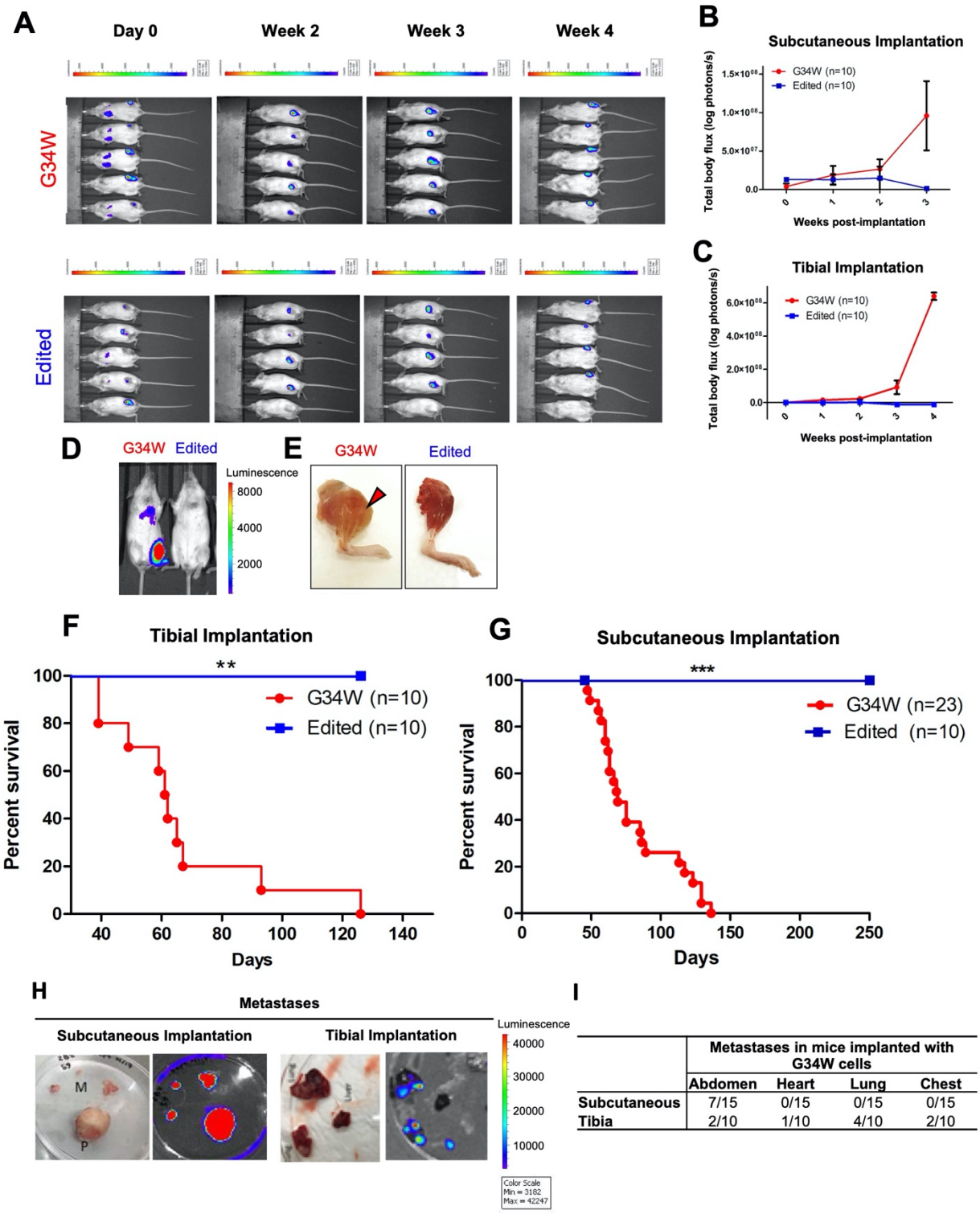
C) G34W lines (Parent: *n*=1; Clone: *n*=2) of ImGCT-4072 proliferate faster than edited lines (Repair to WT: *n*=3; G34W-KO: *n*=2), as measured using the IncuCyte live-cell analysis system for 5 consecutive days. Data are presented as mean red object count ± SD from five technical replicates per line. Statistical significance assessed using two-way ANOVA based on averaged observations from biological replicates (independent CRISPR clones, labeled).

D) G34W lines (Parent: *n*=1; Clone: *n*=1) of ImGCT-3504 proliferate faster than edited lines (Repair to WT: *n*=3), as measured using the IncuCyte live-cell analysis system for 5 consecutive days. Five technical replicates were measured per line. Data are presented as mean red object count ± SD from biological replicates (independent CRISPR clones). Statistical significance assessed using two-WAY ANOVA.

E) G34W lines (Parent: *n*=1; Clone: *n*=2) of ImGCT-4072 exhibit increased colony formation relative to edited lines (Repair to WT: *n*=3; G34W KO: *n*=2), as measured by manual counting of colonies stained with crystal violet after 3 weeks (bottom panel). Two technical replicates were counted per line, and data are presented as an average of biological replicates (independent CRISPR clones, labeled). Statistical significance assessed using Student's *t*-test.

F) Representative images of G34W (left) and edited (right) cell colonies in culture plates.

G) G34W lines (Parent: *n*=1; Clone: *n*=1) of Im-GCT-3504 exhibit increased colony formation relative to edited lines (Repair to WT: *n*=5), as measured by manual counting of colonies stained with crystal violet after 3 weeks. Two technical replicates were counted per line, and data are presented as an average of biological replicates (independent CRISPR clones). Statistical significance assessed using Student's *t*-test.



	Metastases in mice implanted with G34W cells			
	Abdomen	Heart	Lung	Chest
Subcutaneous	7/15	0/15	0/15	0/15
Tibia	2/10	1/10	4/10	2/10

Figure 4. 3. 1 G34W is required for tumor formation and promotes osteoclast recruitment in GCTB

A) Bioluminescence Xenogen IVIS 200 imaging of mice injected orthotopically with ImGCT-4072 G34W and edited luciferase-tagged lines at the time points indicated.

B-C) Quantitative analysis of *in vivo* Xenogen IVIS 200 bioluminescence signal intensity over time of mice injected subcutaneously (B) or orthotopically (C) showing increasing signal in ImGCT-4072 G34W (Parent: n=1; Clone: n=1) relative to edited (Repair to WT: n=2) lines.

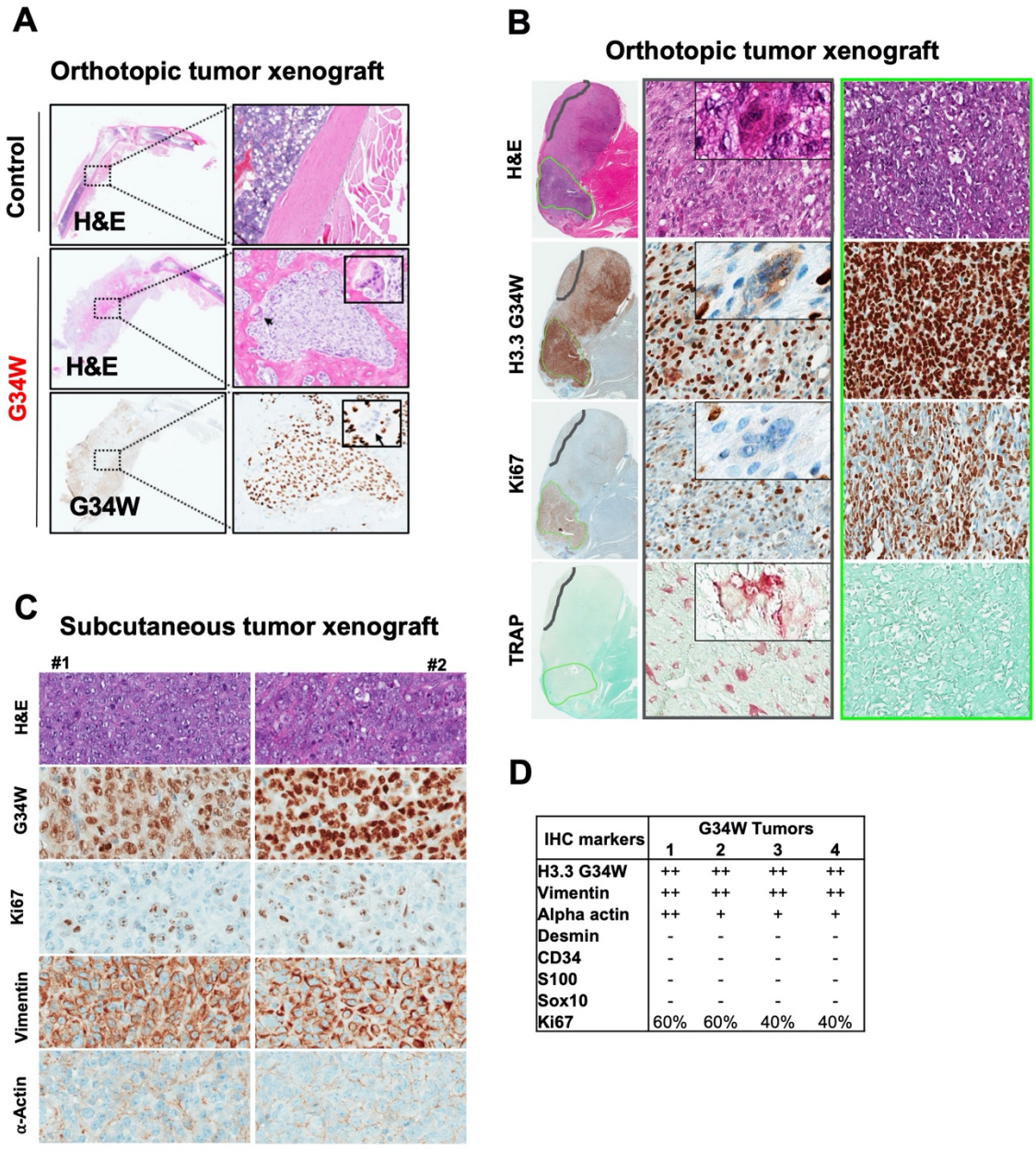
D-E) Representative images of mice tibias implanted with G34W (left) and edited (right) (D) cells and after skin removal at the time of sacrifice (E).

F) Kaplan-Meier survival curve for orthotopic tibial implantation of ImGCT-4072 G34W (Parent: n=1; Clone: n=2) and edited lines (Repair to WT: n=2) in NRG mice (10 mice per group) illustrates the dependence of tumor formation on the presence of G34W mutation.

G) Kaplan-Meier survival curve for subcutaneous implantation of ImGCT-4072 G34W (Parent: n=1; Clone: n=2, total 23 mice), and edited lines (Repair to WT: n=2, total 10 mice), illustrates the dependence of tumor formation on the presence of G34W mutation.

H) Representative bioluminescence imaging of organs with metastases from subcutaneous and orthotopic tibial injections of ImGCT-4072 G34W lines.

I) Table listing the number of mice implanted with ImGCT-4072 G34W cells with metastases to listed organs.



D

IHC markers	G34W Tumors 1	2	3	4
H3.3 G34W	++	++	++	++
Vimentin	++	++	++	++
Alpha actin	++	+	+	+
Desmin	-	-	-	-
CD34	-	-	-	-
S100	-	-	-	-
Sox10	-	-	-	-
Ki67	60%	60%	40%	40%

Figure 4. 3. 2 G34W is required for tumor formation and promotes osteoclast recruitment in GCTB

A) Representative H&E and G34W IHC of decalcified legs derived from tibial implantation of ImGCT-6276 G34W parental cells illustrating the osteolytic effect of G34W stromal cells relative to a control lateral leg. Inset features reactive G34W-negative osteoclasts observed at the interface between G34W-positive neoplastic stromal cells and normal bone.

B) H&E, TRAP (tartrate-resistant acid phosphatase) H3.3G34W and Ki67 IHC for a representative tibial xenograft tumor derived from implantation of ImGCT-4072 G34W parental cells. The 20X (70 µm) magnified area illustrates a histological compartment (middle panel) with differentiated stromal cells and abundant TRAP+ osteoclasts (an example of giant multinucleated osteoclast is featured in the inset). Right panel illustrates a histological compartment with undifferentiated stromal cells, high Ki67 staining, and absence of TRAP+ osteoclasts.

C) H&E, and H3.3G34W, Ki67, Vimentin, Alpha actin IHC for representative subcutaneous xenograft tumors derived from injection of H3.3G34W parental and unedited ImGCT-4072 cells.

D) Table including the H3.3G34W, Ki67, Vimentin, Alpha actin, CD34, S100, Sox10 IHC scoring for subcutaneous xenograft tumors derived from injection of H3.3G34W parental and unedited ImGCT-4072 cells.

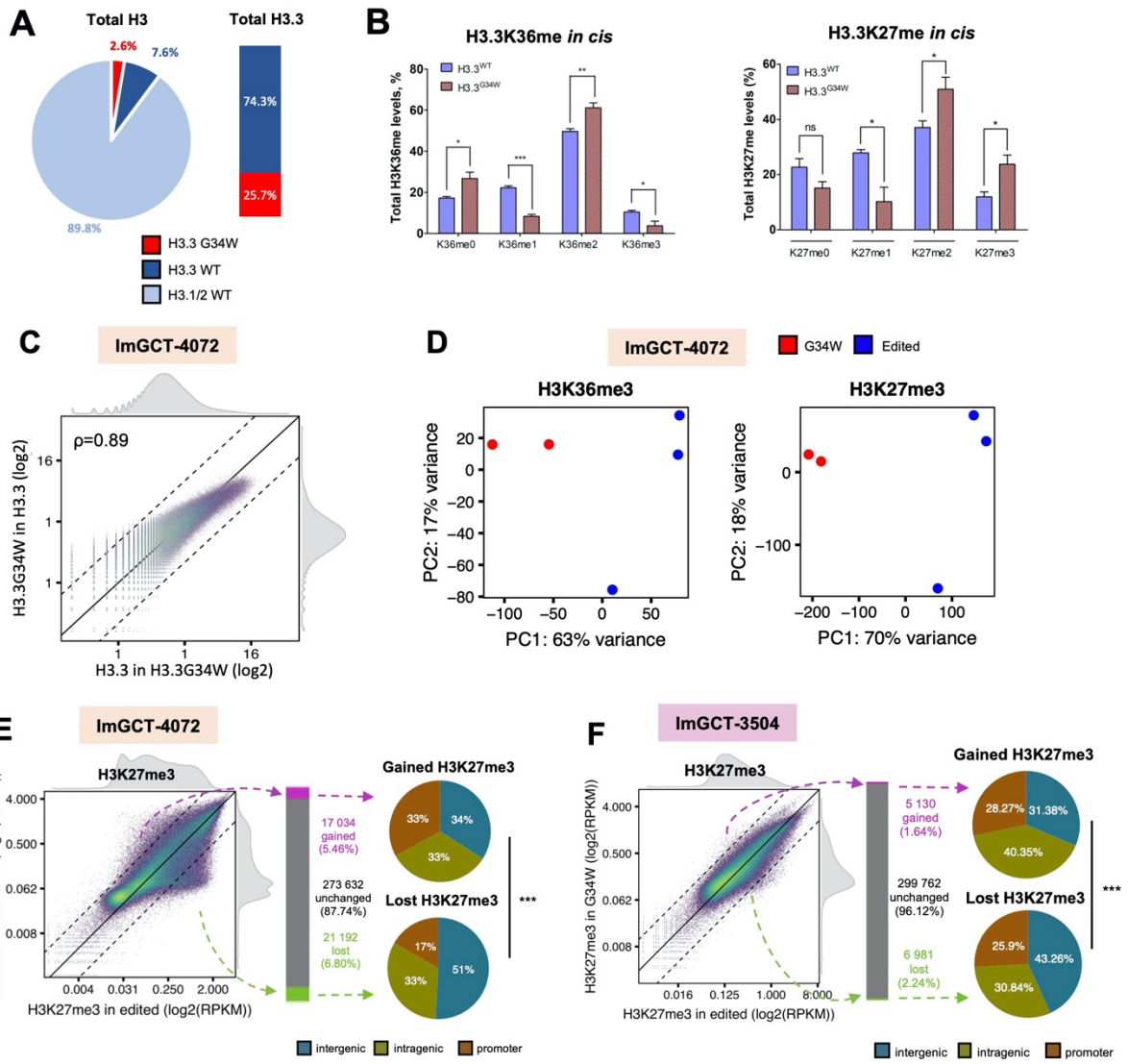

A
Total H3
Total H3.3
2.6%
7.6%
89.8%
74.3%
25.7%
H3.3 G34W
H3.3 WT
H3.1/2 WT
B
H3.3K36me in cis
H3.3K27me in cis
Total H3K36me levels, %
Total H3K27me levels (%)
H3.3WT
H3.3G34W
K36me0
K36me1
K36me2
K36me3
K27me0
K27me1
K27me2
K27me3
ns
C
ImGCT-4072
ρ=0.89
H3.3G34W in H3.3 (log2)
H3.3 in H3.3G34W (log2)
D
ImGCT-4072
G34W
Edited
H3K36me3
H3K27me3
PC2: 17% variance
PC1: 63% variance
PC2: 18% variance
PC1: 70% variance
E
ImGCT-4072
H3K27me3
H3K27me3 in G34W (log2(RPKM))
H3K27me3 in edited (log2(RPKM))
17 034 gained (5.46%)
273 632 unchanged (87.74%)
21 192 lost (6.80%)
Gained H3K27me3
Lost H3K27me3
33%
34%
33%
17%
51%
***
intergenic
intragenic
promoter
F
ImGCT-3504
H3K27me3
H3K27me3 in G34W (log2(RPKM))
H3K27me3 in edited (log2(RPKM))
5 130 gained (1.64%)
299 762 unchanged (96.12%)
6 981 lost (2.24%)
Gained H3K27me3
Lost H3K27me3
28.27%
31.38%
40.35%
25.9%
43.26%
30.84%
***
intergenic
intragenic
promoter

Figure 4. 4 H3.3G34W is associated with global epigenetic remodeling in addition to *in-cis*-acting epigenetics effects

A) Left: Pie chart illustrating relative abundance of G34W, and wild-type H3.3 and H3.1/2 histones by histone mass spectrometry in the ImGCT-4072 parental line. Right: stacked bar plot illustrating relative abundance of G34W and wildtype H3.3.

B) Histone mass spectrometry reveals *in cis* changes in methylation H3.3K36 and H3.3K27 on G34W compared to WT H3.3 peptides in G34W-mutant GCTB cell lines (n=4). *: P<0.05; **: P<0.01, ***: P<0.001, n.s.: non-significant. Significance was assessed using Student's t-test.

C) Scatterplot depicting changes in G34W (x-axis) and H3.3 (y-axis) ChIP abundance in the parental lines of ImGCT-4072. Color: point density. Solid line: no change in abundance. Dotted lines: two-fold change in abundance. Reads counted over 10kb bins, averaged per condition, normalized to RPKM and reported in $\log_2$ scale.

D) PCA reveals distinct H3K36me3 and H3K27me3 profiles between G34W (red; n=2) and edited lines (blue; n=3) from ImGCT-4072. Read counts were counted over 10kb genomic bins and normalized to RPKM.

E) Left: Scatterplot depicting changes in H3K27me3 in G34W (y-axis; n=2) and edited (x-axis; n=3) lines from ImGCT-4072. Color: point density. Solid line: no change in abundance. Dotted lines: two-fold change in abundance. Reads counted over 10kb bins, averaged per condition, normalized to RPKM and reported in log2 scale. Middle: Bar plot quantifying the number of 10kb bins with gained (purple), unchanged (grey) and lost (green) H3K27me3 in G34W relative to edited lines. Bins with above-median average H3K27me3 in either condition and with an absolute log2 fold-change (LFC) of H3K27me3 exceeding 1 were called gains and losses. Right: H3K27me3 is redistributed from intergenic to genic regions in G34W lines. Pie charts illustrate the proportion of 10kb bins gaining or losing H3K27me3 that overlap promoters, gene bodies and intergenic regions. ***: P<0.001. Significance was assessed using the $\chi 2$ test.

F) Left: Scatterplot depicting changes in H3K27me3 in G34W (y-axis; n=2) and edited (x-axis; n=3) lines from ImGCT-3504. Color: point density. Solid line: no change in abundance. Dotted lines: two-fold change in abundance. Reads counted over 10kb bins, averaged per condition, normalized to RPKM and reported in $\log_2$ scale. Middle: Bar plot quantifying the number of 10kb bins with gained (purple), unchanged (grey) and lost (green) H3K27me3 in G34W relative to edited lines. Bins with above-median average H3K27me3 in either condition and with an absolute $\log_2$ fold-change (LFC) of H3K27me3 exceeding 0.58 were called as gains and losses. Right: H3K27me3 is redistributed from intergenic to genic regions in G34W lines. Pie charts illustrating proportion of 10kb bins gaining or losing H3K27me3 that overlap promoters, gene bodies and intergenic regions. ***: P<0.001. Significance was assessed using $\chi 2$ test.

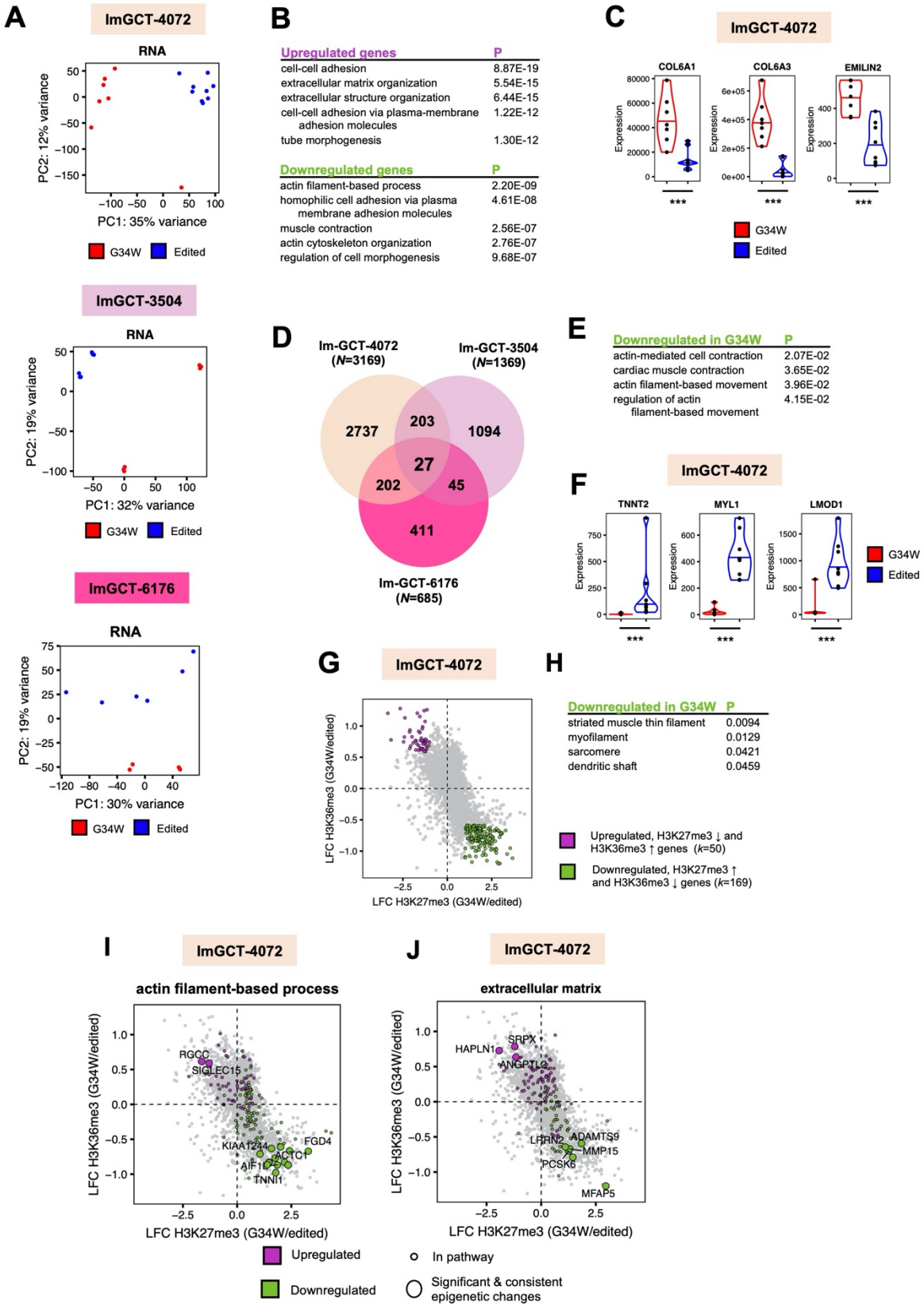

A
lmGCT-4072
RNA
PC2: 12% variance
PC1: 35% variance
G34W
Edited
lmGCT-3504
RNA
PC2: 19% variance
PC1: 32% variance
G34W
Edited
lmGCT-6176
RNA
PC2: 19% variance
PC1: 30% variance
G34W
Edited
B
Upregulated genes P
cell-cell adhesion 8.87E-19
extracellular matrix organization 5.54E-15
extracellular structure organization 6.44E-15
cell-cell adhesion via plasma-membrane adhesion molecules 1.22E-12
tube morphogenesis 1.30E-12
Downregulated genes P
actin filament-based process 2.20E-09
homophilic cell adhesion via plasma membrane adhesion molecules 4.61E-08
muscle contraction 2.56E-07
actin cytoskeleton organization 2.76E-07
regulation of cell morphogenesis 9.68E-07
C
lmGCT-4072
COL6A1
COL6A3
EMILIN2
Expression
G34W
Edited
D
lm-GCT-4072 (N=3169)
lm-GCT-3504 (N=1369)
lm-GCT-6176 (N=685)
2737
203
1094
27
202
45
411
E
Downregulated in G34W P
actin-mediated cell contraction 2.07E-02
cardiac muscle contraction 3.65E-02
actin filament-based movement 3.96E-02
regulation of actin filament-based movement 4.15E-02
F
lmGCT-4072
TNNT2
MYL1
LMOD1
Expression
G34W
Edited
G
lmGCT-4072
LFC H3K36me3 (G34W/edited)
LFC H3K27me3 (G34W/edited)
H
Downregulated in G34W P
striated muscle thin filament 0.0094
myofilament 0.0129
sarcomere 0.0421
dendritic shaft 0.0459
Upregulated, H3K27me3 ↓ and H3K36me3 ↑ genes (k=50)
Downregulated, H3K27me3 ↑ and H3K36me3 ↓ genes (k=169)
I
lmGCT-4072
actin filament-based process
RGCC
SIGLEC15
KIAA1244
FGD4
ACTC1
AIF1
TNNI1
LFC H3K36me3 (G34W/edited)
LFC H3K27me3 (G34W/edited)
J
lmGCT-4072
extracellular matrix
HAPLN1
SRPX
ANGPTL2
LRRN2
ADAMTS9
MMP15
PCSK6
MFAP5
LFC H3K36me3 (G34W/edited)
LFC H3K27me3 (G34W/edited)
Upregulated
Downregulated
In pathway
Significant & consistent epigenetic changes

Figure 4. 5 H3.3G34W is associated with dysregulation of transcriptional programs specifying multiple mesenchymal cell lineages

A) PCA reveals distinct transcriptomic profiles between ImGCT-4072 G34W (red; *n*=7) and edited lines (blue; *n*=9), ImGCT-3504 G34W (red; *n*=6) and edited lines (blue; *n*=6), and ImGCT-6176 G34W (red; *n*=4) and edited lines (blue; *n*=6). Read counts were counted over Ensembl genes, normalized using the median-of-ratios procedure and transformed using the variance-stabilizing transformation.

B) Pathway enrichment analysis of statistically significantly up- (purple) and down-regulated (green) genes between G34W and edited lines from ImGCT-4072. Pathway enrichment analysis was performed using g:Profiler. Top 5 statistically significantly enriched terms (GO:BP, term size<1000, *P*<0.05) are shown.

C) Boxplots depicting expression levels of *COL6A1, COL6A3, EMILIN2* in G34W (red) and edited (blue) lines from ImGCT-4072. *: *P*<0.05, ***: *P*<0.001, n.s.: non- significant. Gene expression levels reported in median-of-ratios normalized read counts. Significance was assessed using DESeq2.

D) Venn diagram illustrates the intersection of commonly dysregulated genes across the three ImGCT isogenic cell lines, affirmed 27 common DEGs which all were downregulated in G34W cells compared to edited cells.

E) Pathway enrichment analysis of statistically significantly down-regulated (green) genes between G34W and edited lines from intersection of ImGCT-4072, ImGCT-3504, and ImGCT-6176 DEGs (Fig. 4. 5. E). Pathway enrichment analysis was performed using g:Profiler. Top 4 statistically significantly enriched terms (GO:BP, term size<1000, *P*<0.05) are shown.

F) Boxplots depicting expression levels of *TNNT2, MYL1, LMOD1* in G34W (red) and edited (blue) lines from ImGCT-4072. *: *P*<0.05, ***: *P*<0.001, n.s.: non- significant. Gene expression levels reported in median-of-ratios normalized read counts. Significance was assessed using DESeq2.

G) Scatterplot illustrating genes with significant changes in expression, genic H3K36me3 and H3K27me3 in G34W lines compared to edited lines from ImGCT-4072. X-axis: log2 fold-change (LFC) of H3K27me3. Y-axis: LFC of H3K36me3. Grey: significantly differentially expressed genes; purple: upregulated genes with significant loss of H3K27me3 and gain of H3K36me3; green: downregulated genes with significant gain of H3K27me3 and gain of H3K36me3. Replicates: H3K27me3 (G34W *n*=2; edited *n*=3), H3K36me3 (G34W *n*=2; edited *n*=3), RNA (G34W *n*=7; edited *n*=9).

H) Pathway enrichment analysis of downregulated (green) genes with consistent changes in H3K36me3 and H3K27me3 between G34W and edited lines from ImGCT-4072. Pathway enrichment analysis was performed using g:Profiler. All statistically significantly enriched terms (*P*<0.05) are shown. No pathways were enriched among the upregulated genes.

I) Scatterplot illustrating epigenetic changes at significantly deregulated genes in actin filament-based process pathway in G34W lines compared to edited lines from ImGCT-4072. X-axis: log2 fold-change (LFC) of H3K27me3. Y-axis: LFC of H3K36me3. Grey: significant differentially expressed genes; purple/green: up- and downregulated genes in pathway; big circle: significant changes in genic H3K36me3 and H3K27me3. Replicates: H3K27me3 (G34W n=2; edited n=3), H3K36me3 (G34W n=2; edited n=3), RNA (G34W n=7; edited n=9). Thresholds for significance: H3K27me3 (baseMean>25; P<0.05; |LFC|>1) and H3K36me3 (baseMean>25; P<0.05; |LFC|>0.58), RNA (baseMean>50; P<0.05; |LFC|>1).

J) Scatterplot illustrates epigenetic changes at significantly deregulated genes in extracellular matrix pathways in G34W lines compared to edited lines from ImGCT-4072. X-axis: $\log_2$ fold-change (LFC) of H3K27me3. Y-axis: LFC of H3K36me3. Grey: significant differentially expressed genes; purple/green: up- and downregulated genes in pathway; big circle: significant changes in genic H3K36me3 and H3K27me3. Replicates: H3K27me3 (G34W n=2; edited n=3), H3K36me3 (G34W n=2; edited n=3), RNA (G34W n=7; edited n=9).

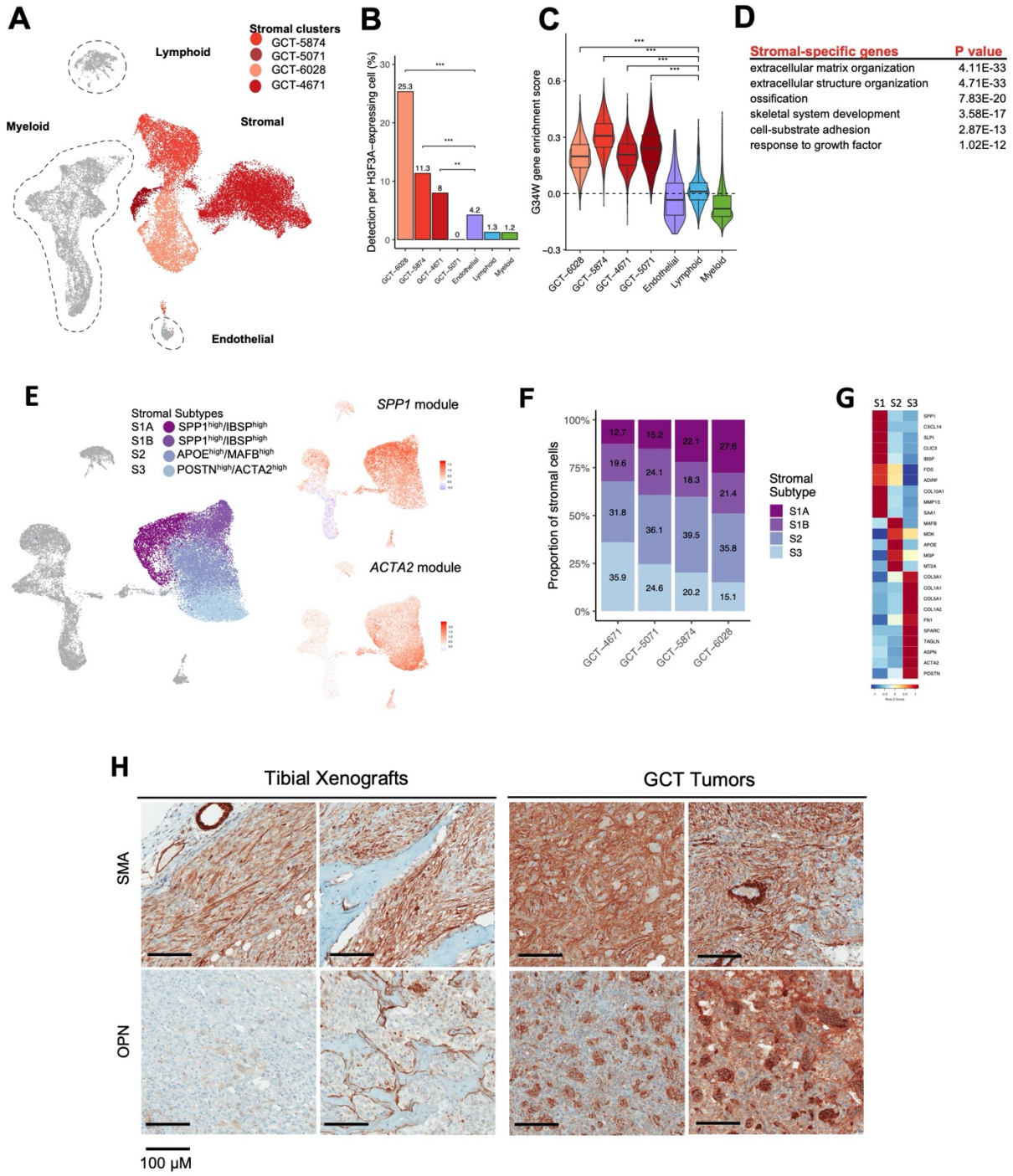

A
Stromal clusters
GCT-5874
GCT-5071
GCT-6028
GCT-4671
Lymphoid
Myeloid
Stromal
Endothelial
B
Detection per H3F3A–expressing cell (%)
25.3
11.3
8
0
4.2
1.3
1.2
GCT–6028
GCT–5874
GCT–4671
GCT–5071
Endothelial
Lymphoid
Myeloid
C
G34W gene enrichment score
GCT-6028
GCT-5874
GCT-4671
GCT-5071
Endothelial
Lymphoid
Myeloid
D
Stromal-specific genes
P value
extracellular matrix organization
4.11E-33
extracellular structure organization
4.71E-33
ossification
7.83E-20
skeletal system development
3.58E-17
cell-substrate adhesion
2.87E-13
response to growth factor
1.02E-12
E
Stromal Subtypes
S1A SPP1high/IBSPhigh
S1B SPP1high/IBSPhigh
S2 APOEhigh/MAFBhigh
S3 POSTNhigh/ACTA2high
SPP1 module
ACTA2 module
F
Proportion of stromal cells
100%
75%
50%
25%
0%
12.7
19.6
31.8
35.9
15.2
24.1
36.1
24.6
22.1
18.3
39.5
20.2
27.6
21.4
35.8
15.1
GCT–4671
GCT–5071
GCT–5874
GCT–6028
Stromal Subtype
S1A
S1B
S2
S3
G
S1 S2 S3
H
Tibial Xenografts
GCT Tumors
SMA
OPN
100 µM

Figure 4. 6. 1 GCTB stromal cells comprise *SPP1*+ and *ACTA2*+ sub-populations

A) UMAP plot displaying cell cluster annotation of GCTB tumor (n=4) scRNA-seq data. Myeloid, lymphoid and endothelial cell clusters are colored in grey, whereas stromal clusters are shown in shades of red based on the tumor of origin.

B) Bar graph showing G34W mutation detection in cells expressing at least one count of *H3F3A*. ***: P<0.0001, **: P<0.001. Significance was assessed using $\chi 2$ test. Detected G34W is significantly higher in the putative stromal cells of 3 out of 4 GCTB samples. G34W failed to be detected in any cells from one sample (GCT-5071) known to be G34W positive, likely due to the technical limitations of single-cell variant detection.

C) Boxplot displaying higher G34W enrichment scores for single cells in stromal clusters. The G34W enrichment score is derived from the average expression of differentially expressed genes (LFC>2) between isogenic G34W and edited Im-GCT-4072 lines. ***: P<0.0005, significance was assessed using a Wilcoxon rank sum test.

D) Pathway enrichment analysis of stromal-specific genes performed using g:Profiler shows similar enriched pathways as Fig. 5B. Top 6 statistically significant enriched GO terms (GO:BP, term size<1000, P<0.05) are shown.

E) Left: UMAP plot of Harmony integrated cell clusters reveals the 4 stromal subtypes S1A, S1B, S2, and S3. Right: Average expression of genes highly correlated with S1-specific *SPP1* gene (*SPP1* module), or with S3-specific *ACTA*2 gene (*ACTA2* module), shown on UMAP plot of Harmony integrated cell clusters.

F) Proportions bar graph showing the percent of stromal cells from each sample that belong to each stromal cell subtype.

G) Row-scaled heatmap showing average expression of differentially expressed genes that characterize each stromal subtype.

H) Representative IHC for osteopontin (SPP1) and alpha-SMA (ACTA2) in two GCTB patient and two PDOX tumors, showing comparatively more ACTA2+ cells than SPP1+ cells in PDOX tumors.

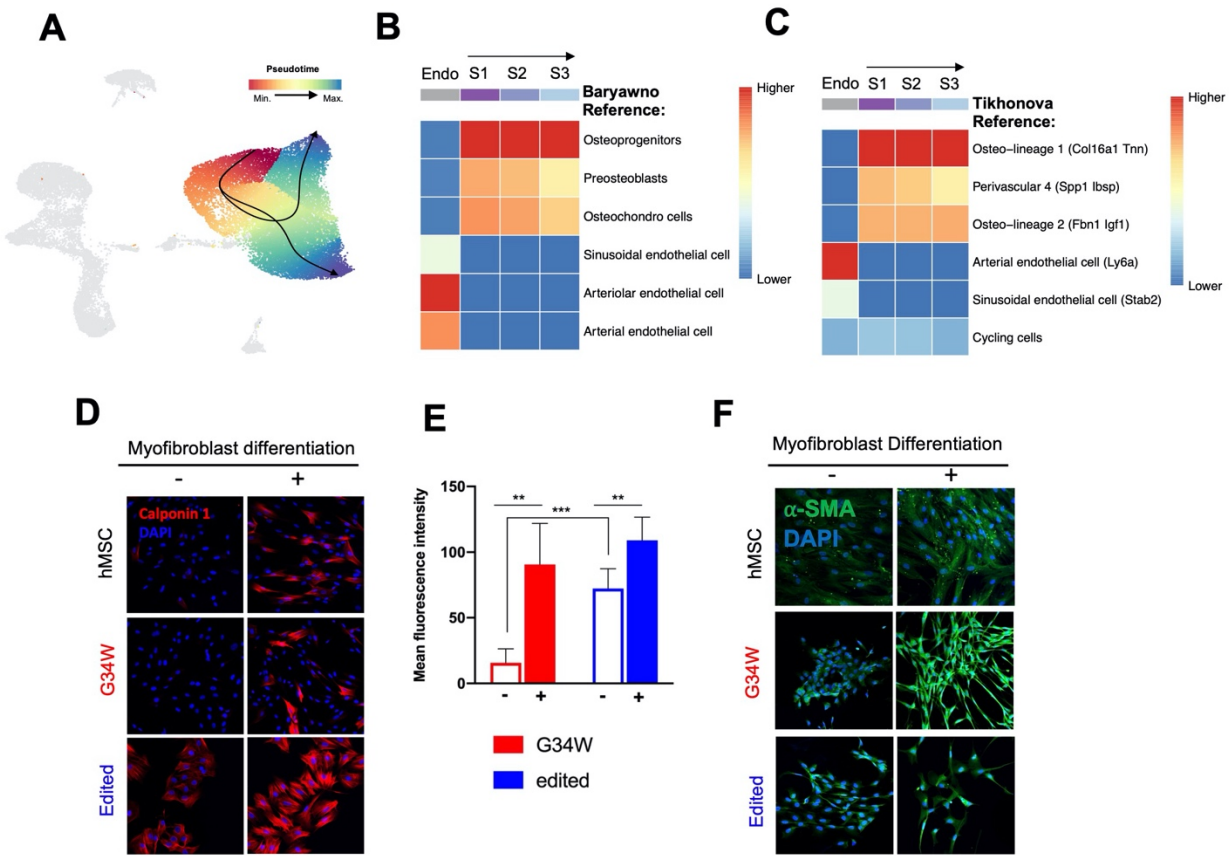

A
Pseudotime
Min.
Max.
B
Endo S1 S2 S3
Baryawno Reference:
Osteoprogenitors
Preosteoblasts
Osteochondro cells
Sinusoidal endothelial cell
Arteriolar endothelial cell
Arterial endothelial cell
Higher
Lower
C
Endo S1 S2 S3
Tikhonova Reference:
Osteo–lineage 1 (Col16a1 Tnn)
Perivascular 4 (Spp1 Ibsp)
Osteo–lineage 2 (Fbn1 Igf1)
Arterial endothelial cell (Ly6a)
Sinusoidal endothelial cell (Stab2)
Cycling cells
Higher
Lower
D
Myofibroblast differentiation
-
+
hMSC
G34W
Edited
Calponin 1
DAPI
E
Mean fluorescence intensity
150
100
50
0
**
***
**
- + - +
G34W
edited
F
Myofibroblast Differentiation
-
+
hMSC
G34W
Edited
α-SMA
DAPI

Figure 4. 6. 2 GCTB stromal cells resemble osteoprogenitors with features of contractile cells

A) Lineage inference by Slingshot showing neoplastic trajectories from S1A to S1B, and S1A to S3. Cells are coloured by pseudotime, with red cells occurring earlier than blue cells in the lineage trajectory.

B) SingleR classification of each stromal cell subtype (S1, S2, S3) and endothelial control based on Tikhonova et al. reference cell types[167]. Stromal cell subtypes most strongly resemble the Osteo-lineage 1 reference cell cluster (labelled as O1 (Col16a1 Tnn) in Tikhonova et al.).

C) SingleR classification of each stromal cell subtype (S1, S2, S3) and endothelial control based on Baryawno et al. reference cell types[166]. Stromal cell subtypes most strongly resemble the osteoprogenitors reference cell cluster (labelled as OLC-2 subtype 8_3 in Baryawno et al.).

D) Representative immunofluorescence images for the myofibroblast muscle marker calponin 1 in hMSCs and isogenic Im-GCT-4072 cells maintained in non-induced (-) or myofibroblast differentiation media (+) for 2 weeks.

E) Bar-plot quantifying the mean fluorescence intensity of calponin 1 staining in G34W (n=2 lines; three different fields each) and edited lines (n=2; four different fields each) maintained in non-induced (-) or myofibroblast differentiation media (+).

F) Representative immunofluorescence images for alpha smooth muscle actin (α-SMA) staining for hMSCs and isogenic ImGCT-4072 cells maintained in non-induced (-) or myofibroblast differentiation media (+) for 2 weeks.

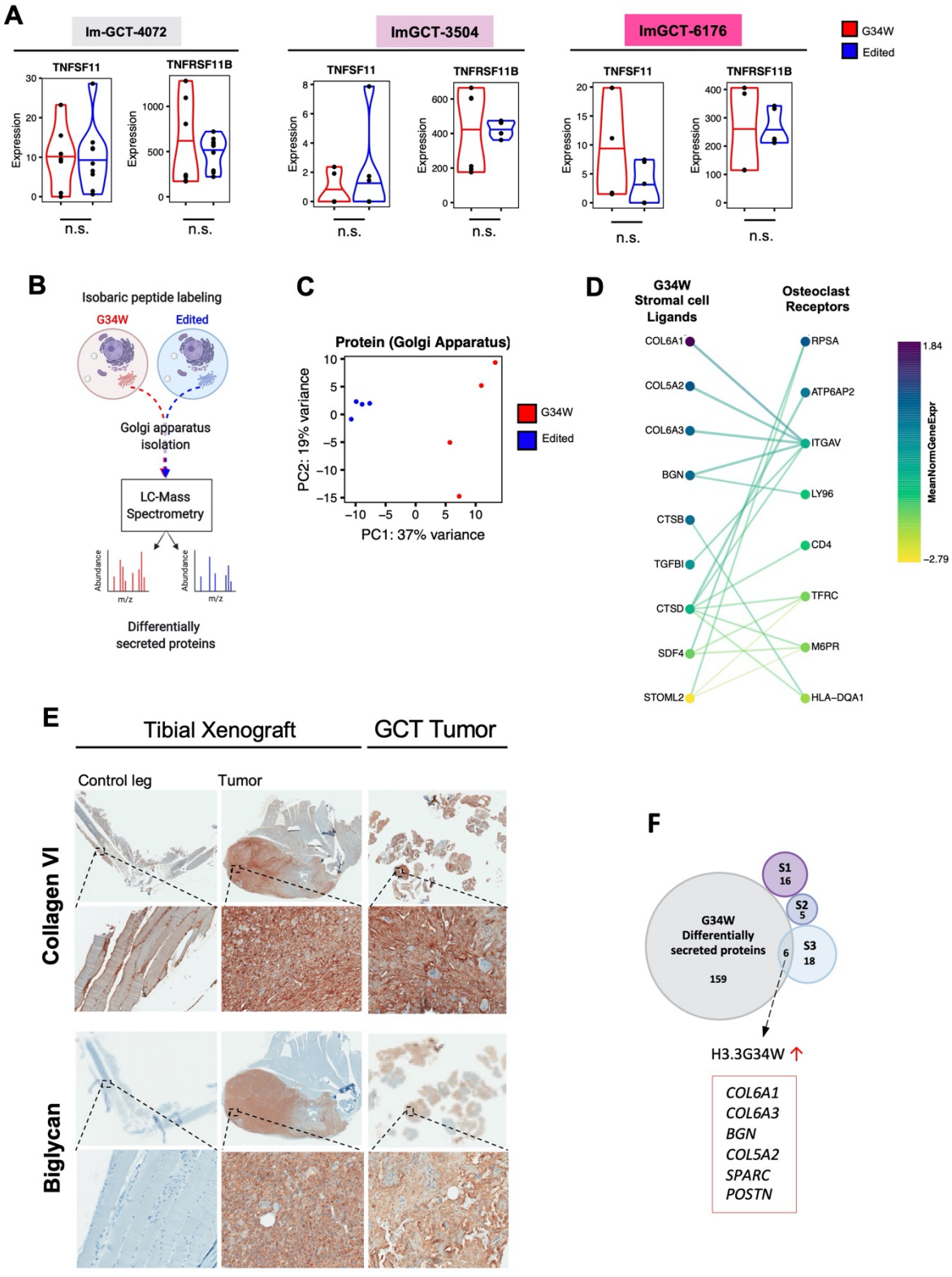

A
Im-GCT-4072
ImGCT-3504
ImGCT-6176
G34W
Edited
TNFSF11
TNFRSF11B
Expression
n.s.
B
Isobaric peptide labeling
G34W
Edited
Golgi apparatus isolation
LC-Mass Spectrometry
Abundance
m/z
Differentially secreted proteins
C
Protein (Golgi Apparatus)
PC2: 19% variance
PC1: 37% variance
G34W
Edited
D
G34W Stromal cell Ligands
Osteoclast Receptors
COL6A1
COL5A2
COL6A3
BGN
CTSB
TGFBI
CTSD
SDF4
STOML2
RPSA
ATP6AP2
ITGAV
LY96
CD4
TFRC
M6PR
HLA–DQA1
MeanNormGeneExpr
1.84
–2.79
E
Tibial Xenograft
GCT Tumor
Control leg
Tumor
Collagen VI
Biglycan
F
S1
16
S2
5
S3
18
G34W Differentially secreted proteins
159
6
H3.3G34W ↑
COL6A1
COL6A3
BGN
COL5A2
SPARC
POSTN

Figure 4. 7 G34W *ACTA2+* stromal cells secrete factors promoting ECM remodeling and association with myeloid cells

A) Boxplots depicting expression levels of *TNFSF11* (RANKL) and *TNFRSF11B* (OPG) in G34W (red) and edited (blue) lines from ImGCT-4072, ImGCT-3504 and ImGCT-6176 *: $P<0.05$, ***: $P<0.001$, n.s.: non- significant. Gene expression levels reported in median-of-ratios normalized read counts Significance assessed using DESeq2.

B) Schematic of Golgi apparatus isolation and mass spectrometry workflow to identify differentially secreted proteins between isogenic G34W (red) and edited (blue) cells.

C) PCA reveals distinct Golgi protein abundance profiles in G34W (red; $n=4$) and edited (blue; $n=4$) lines from Im-GCT-4072. Protein abundance profiles were derived from intensities and normalized to log2(1+x). $n=4$ technical replicates were used in each condition.

D) CCInx[168]predicted ligand-receptor interactions between GCT stromal cells (left) and osteoclast cells (right). Colors represent the mean normalized gene expression in each cell type. Only interactions between proteins differentially secreted in G34W cell lines by MS ($P<0.05$) and expressed by stromal cells are shown on the left and only genes differentially expressed ($P<0.05$) in osteoclasts (vs. non-myeloid cells) are shown on the right. P values were adjusted for multiple testing using FDR.

E) Representative IHC for Collagen type VI and Biglycan in a tibial xenograft tumor from implantation of Im-GCT-4072 G34W cells shown in Fig. 1G (middle panel) and patient GCT tumor (right panel). Stronger extracellular collagen VI and BGN staining is observed in G34W xenograft and GCT tumors relative to the bone matrix of a control contralateral mouse leg.

F) Venn diagram showing overlap of genes with significantly enriched expression in each stromal cell subtype, S1-S3 (Seurat Wilcox test, $P<0.05$, FDR corrected) and genes with significantly increased protein secretion in G34W cell lines by MS ($P<0.05$, FDR corrected). The 6 intersecting genes are highlighted.

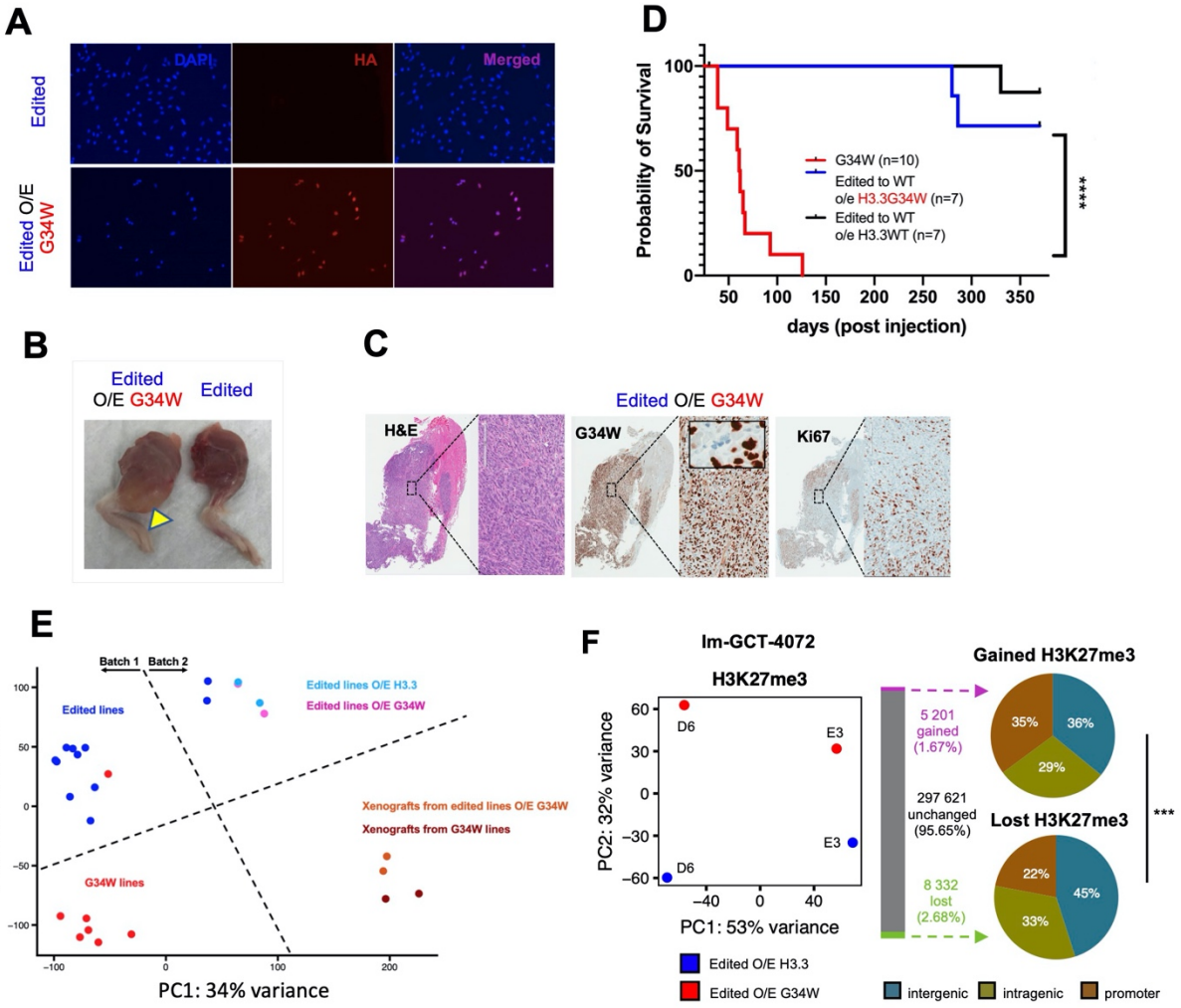

A
DAPI
HA
Merged
Edited
Edited O/E G34W
D
Probability of Survival
100
50
0
50
100
150
200
250
300
350
days (post injection)
G34W (n=10)
Edited to WT o/e H3.3G34W (n=7)
Edited to WT o/e H3.3WT (n=7)
****
B
Edited
Edited O/E G34W
Edited
C
Edited O/E G34W
H&E
G34W
Ki67
E
Batch 1
Batch 2
Edited lines
Edited lines O/E H3.3
Edited lines O/E G34W
Xenografts from edited lines O/E G34W
Xenografts from G34W lines
G34W lines
PC2: 18% variance
PC1: 34% variance
F
lm-GCT-4072
H3K27me3
D6
E3
E3
D6
PC2: 32% variance
PC1: 53% variance
Edited O/E H3.3
Edited O/E G34W
5 201 gained (1.67%)
297 621 unchanged (95.65%)
8 332 lost (2.68%)
Gained H3K27me3
35%
36%
29%
Lost H3K27me3
22%
45%
33%
***
intergenic
intragenic
promoter

Figure 4. 8 Overexpression of H3.3 G34W in CRISPR-edited lines reverts some of the epigenetic, transcriptomic, and tumorigenic effects

A) Immunofluorescence confirmation of overexpressed HA-tagged H3.3WT and H3.3G34W constructs in ImGCT-4072 edited clones.

B) Representative images of mice tibias implanted with edited cells O/E G34W and edited cells and after skin removal at the time of sacrifice.

C) Representative H&E, G34W and Ki67 IHC of decalcified legs derived from tibial implantation of ImGCT-4072 edited cells O/E G34W. Inset features G34W-negative osteoclasts.

D) Kaplan-Meier survival curve for orthotopic tibial implantation of ImGCT-4072 G34W (Parent: n=1, Clone: n=2; n=10 mice), edited O/E G34W (n=7 mice), edited O/E H3.3WT (n=7 mice) in NRG mice.

E) PCA of RNA-seq datasets of ImGCT-4072 G34, edited (WT), edited O/E H3.3WT or H3.3G34W cell lines, and xenografts from G34W and edited O/E H3.3 G34W lines. Read counts were counted over Ensembl genes, normalized using the median-of-ratios procedure and transformed using the variance-stabilizing transformation.

F) Left: PCA reveals changes in H3K27me3 in edited lines from ImGCT-4072 upon re-expression of H3.3G34W (red; n=2) compared to H3.3WT (blue; n=2). Clone-specific effects are captured in PC1 (53% of variance), whereas G34W-dependent effects are captured by PC2 (32% of the variance). Middle: Bar plot quantifying the number of 10kb bins with gained (purple), unchanged (grey) and lost (green) H3K27me3 in edited lines re-expressing H3.3G34W compared to H3.3WT. Bins with above-median average H3K27me3 in either condition and with an absolute log2 fold-change (LFC) of H3K27me3 exceeding 0.32 (1.25- fold change) were called as gains and losses. Right: Pie charts illustrating proportion of 10kb bins gaining or losing H3K27me3 that overlap promoters, gene bodies and intergenic regions. ***: P<0.001.

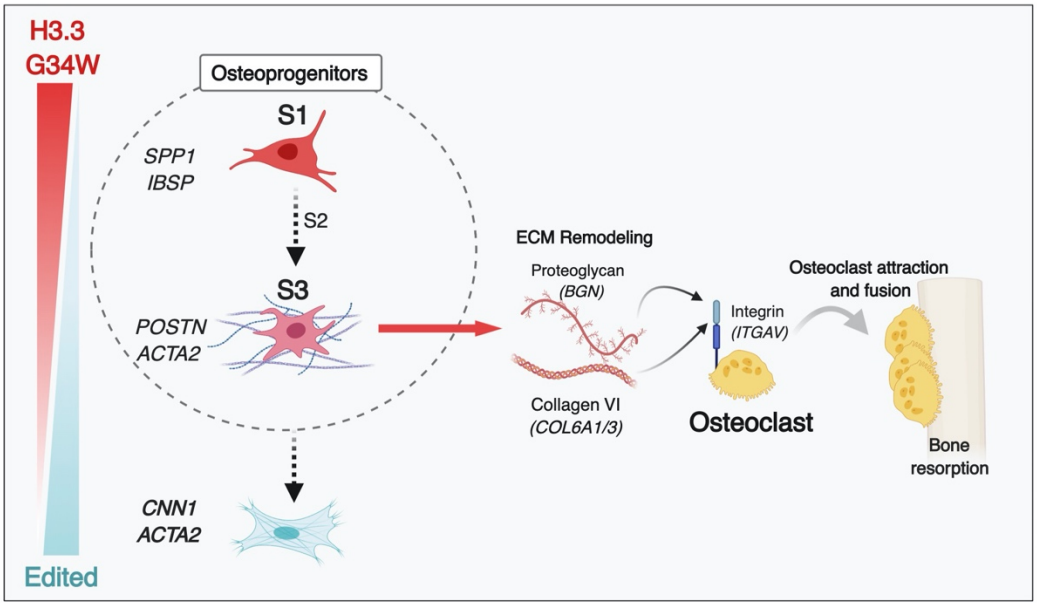


Figure 4. 9 Schematic illustrating G34W-mediated ECM remodeling by subpopulations of GCTB stromal cells.

Schematic illustrating G34W-dependent differentiation trajectory in stromal cells and interactions with osteoclasts in the bone TME.

Chapter V: Discussion

5.1. Novel DKI mouse models carrying germline histone H3.3 G34 mutations

Several studies have been performed to elucidate the mechanism of G34 mutations causing HGG and GCTB, but the mechanisms of the tissue specificity of tumors with histone H3.3G34 mutations, has not yet been elucidated. Modeling these point mutations, especially *in vivo* during organ development, is therefore critical to assess their involvement in induction of tumors. Creating DKI mouse models that express point mutations at the endogenous *H3f3a* locus enables us to determine developmental time points that are critical for tumorigenesis.

We found that DKI mice with either H3.3G34R or H3.3G34W mutations demonstrated remarkable differences in developmental phenotypes. The *H3f3a+/*G34W mutation mainly affected cells and tissues of mesodermal origin, while the *H3f3a+/G34*R mutation mainly affected cells and tissues of ectodermal origin. *H3f3a+/*G34W mice showed deficits in bladder structure and striated muscle fiber distribution, tissues that derive from the mesodermal layer during development. More specifically, *H3f3a+/G34*R mice showed defects in the cytoarchitecture of the cerebral cortex and cerebellum, which derive from the ectodermal layer. Mice with *H3f3a+/G34*R also showed severe neurological deficits such as ataxia phenotypes (hindlimb clasping, abnormal gait and tremor), and difficulties in motor learning and performance. We did not observe these phenotypes in *H3f3a+/G34*W mice, although we did find them to a lesser degree in *H3f3a+/G34*V mice. Moreover, all mice with G34 mutations suffered from obesity. Especially, the G34W mice developed a more severe obesity and at an earlier time than the *H3f3a+/G34*R/V mice.

The observed codon-specific phenotypes causing tissue-specific effects in our mice models are in line with the finding about tissue specificity of tumors in humans, namely G34R/V mutations are found in HGG in the brain, while the G34W mutation is exclusively found in specific tumors with mesodermal (mesenchymal) origin such as GCTB.

5. 2. Epigenetic-remodeling events in G34W GCTB

Using histone mass spectrometry, the G34W containing patient-derived GCTB cells showed a loss of H3K36me3 and concomitant gain of H3K27me3 on the mutated oncohistone.

This was also found in G34-oncohistone overexpression models of HEK293T and HeLa lines[39,40]. In contrast to these models, however, we also observed *in cis* gain of H3K36me2 on the H3.3G34W histone. The latter finding may underscore the differential impact on H3K36 modifiers by H3.3G34R/V and H3.3G34W mutations and explain their tissue-type specific nature.

Another example of differentially effects of the G34 mutations on the epigenome is on the specific H3.3K36me3 reader ZMYND11. It has been shown earlier that the binding of ZMYND11 is impaired by H3.3G34R/V[59]. However, this effect seems amino acid specific. Lim et al. used immunoprecipitations of histones coupled with mass spectrometry analysis and showed that in G34 mutations overexpression model of HEK293 cell lines, there are unique and common protein interactions between mutated G34R and G34W peptides. For example, ZMYND11 is a protein that interacts with G34W, but not with G34R mutations[122]. Moreover, cell-lineage specific effects of H3.3G34 mutations could be expected because of the expression and functions of each of the H3K36 methylation writers, erasers and readers is specific to neuronal cells compared to mesenchymal cell types. This in itself could explain the different mouse phenotypes and the cancers that occur with G34 mutations.

We also showed that at a genome wide level, H3K27me3 and H3K36me3 had distinct deposition profiles in the G34W cells compared to their isogenic edited (WT) cells. We found an equal gain of H3K27me3 in promoter, intragenic and intergenic genomic compartments and a preferential loss of H3K27me3 deposition in intergenic regions in G34W mutant cells. Isolating the progression of epigenetic-remodeling events in our CRISPR/Cas9 isogenic model is complex, but, Indeed, we posit that the earliest event following H3.3K36me3 loss on G34W histones is the deposition of H3K27me3 in active genic regions that are normally enriched for H3.3 as well as G34W (representing 25% of total H3.3). H3K36me3 loss in genic regions may create an improved substrates for the PRC2 complex and will result in redistribution of H3K27me3 from lower affinity intergenic regions to promoter and genic regions, thereby promoting gene silencing.

5.3. *H3f3a+/*G34R/V mouse phenotypes vs human brain tumor (HGGs) with G34R/V mutations

In pediatric HGGs, H3.3G34R/V mutations are found in tumors of the cerebral hemispheres, and specifically in the temporo-parietal cortex of adolescent and young adult patients[76]. Therefore, to precisely model H3.3G34R/V mutations in mice, one needs to express these mutations spatiotemporally during mouse brain development, at a specific time and in specific brain regions. For this tissue-specific or cell lineage-specific expression of these point mutations, more knowledge about the cell-of-origin of these HGGs is needed.

Our work with the DKI mouse models showed that in H3.3G34R mice, some regions of the brain such as the cortex, showed an increased number of GFAP positive cells (astrocyte marker), while Olig2 positive cells (oligodendrocyte marker) were depleted compared to WT littermates. Olig2 plays a pivotal role in development of oligodendrocytes and motor neurons and it is expressed in oligodendrocytes but not in mature astrocytes. This could indicate that in our G34 mutant mice, the differentiation of early NSPCs is biased towards differentiation to astrocytes at the expense of oligodendrocytes. Bjerke et al. showed that G34 brain tumors were enriched for a genetic signature of neural stem/progenitors (NSPCs). This could indicate that the cells of origin of G34 tumors have a neural progenitor identity and that G34 mutations may have happened early during forebrain development[102]. This aligns with our findings as well as previous studies have found that H3.3G34R/V mutant brain tumors have low levels of Olig1/2 expression[78,100]. Similarly, previous studies showed that G34-mutant HGG tumors are positive for GFAP and P53, but negative for ATRX and OLIG2[103,104]. Moreover, *in vitro* studies have shown that Olig2 has a negative regulatory effect on astrocyte differentiation[196,197].

In vivo studies have also shown that repression of Olig2+ cells in gray matter, but not in white matter of the cortex results in upregulation of GFAP[198]. We therefore postulate that in an early developmental stage of the brain, H3.3G34R/V mutations in neural progenitor cells result in the suppression of Olig2 expression, with a consequential increase in pro-astrocyte transcription factors followed by GFAP expression. This effect may result in a biased differentiation of NPCs towards more GFAP+ astrocytes at the expense of oligodendrocytes. An aberration in the

development of astrocyte precursors and oligodendrocyte precursor cells (OPCs) could consequently result in tumor formation[199].

However, although our data is in support of the theory that *H3f3a+/*G34R/V mutations favour the differentiation of NSPCs to astrocytes rather than oligodendrocytes, we cannot rule out the possibility that this phenotype may arise from a secondary activation of astrocytes at a later developmental stage due to a response to pathological conditions which induce the activation of astrocytes from a normal quiescence state into a hyperproliferative reactive state[200]. Upon activation, astrocytes also up-regulate the expression of GFAP, as a hallmark of reactivity[201], in addition to other intermediate filaments in astrocytes such as Vimentin[200], as we did observe in *H3f3a+/*G34R mouse brains as well. Reactive astrocytes have been shown to happen in pathological conditions such as brain trauma, stroke, inflammation, or in neurodegenerative diseases such as Parkinson's disease (PD), Alzheimer's disease (AD) and Huntington's diseases (HD), as well as amyotrophic lateral sclerosis (ALS)[202].

In addition, we observed demyelination of the cortex in aging H3.3G34R mutant mice as mature myelinating oligodendrocytes (MBP positive cells) are progressively lost in these mice. This cortex demyelination may play a role in the thinning of the cerebral hemispheres. In addition, we also found a loss of Purkinje neurons, the GABAergic inhibitory neurons, in the cerebellum of these animals. This loss of Purkinje neurons could be a potential underlying factor for the ataxic symptoms, deficits in motor learning and balance that we have identified in our H3.3G34R model.

5.4. *H3f3a+/*G34W mouse phenotypes vs human mesenchymal lineage tumors which carry G34W and K36M mutations

5.4.1. H3f3a+/G34W animals suffer morbid obesity

The H3.3G34 mutations in mice caused an obesity phenotype, with the *H3f3a+/*G34W mutation causing the most severe obesity phenotype. This obesity phenotype likely results from metabolic dysfunction and an impaired adipogenesis rather than a change in the food consumption patterns of the mice. In WAT tissues of *H3f3a+/*G34W mice we found that *Wisp2*, an inhibitor of adipogenesis, is significantly up-regulated, while *Pparg*, a master regulator of adipogenesis, is

down-regulated. The Wisp2 protein acts both in the cytosol (intracellularly) and as a secreted protein (extracellularly). In the cytosol, Wisp2 inhibits Znf423, an important protein in determining cell fate commitment of MSC to preadipocytes. Znf423 is also a transcriptional activator of *Pparg,* which together with *Cebpb* is a master regulator of adipogenesis, and in particular regulates the terminal differentiation of preadipocytes to adipocytes. Therefore, an increase in Wisp2 and subsequent changes impairs adipogenesis, which consequently causes hypertrophy of adipose cells[203] (Fig. 3.4H: model). Indeed, we found adipocyte hypertrophy in *H3f3a*+/G34W animals in addition to low grade inflammation and macrophage infiltration. All are signs of impaired adipogenesis and an unhealthy obesity.

In addition, in *H3f3a*+/G34W mice we found a whitening of BAT, which also is an indicator of obesity. Moreover, we found that the BAT marker genes (*Ucp1* and *Cidea*) are significantly downregulated in WAT tissues of these animals. BAT is a thermogenic organ, and plays its thermogenesis role through the Ucp1 protein, so it would be interesting to conduct further experiments to check the impairment of BAT function and its role in the development of the obesity phenotype in these mice.

Interestingly, other histone mutations were also found to have an effect on adipogenesis. For instances, Lu et al. showed in mouse MSCs (10T1/2 cell line) that overexpression (o/e) of the H3.3K36M mutation impedes MSC differentiation to chondrocytes, as well as to adipocytes and osteocytes[82]. However, they also found that MSCs with o/e G34W/L had little impact on the MSC differentiation into chondrocyte[82]. Furthermore, Zhuang et al. showed that H3.3K36M o/e and depletion of NSD2, an H3K36 mono/di-methyltransferase, inhibit preadipocyte differentiation to adipocytes by blocking PPARγ expression. Furthermore, in H3.3K36M expressed C2C12 cell line (mouse skeletal myoblasts), myogenesis is also inhibited[139]. H3K36M inhibits the H3K36 methyltransferases NSD1/NSD2 and SETD2, resulting in a reduction of the abundant H3K36me2/3 mark and a genome wide gain of the repressive H3K27me3 mark. Because we and others[39,192] have shown that the G34W mutation may act through its effects on methylation of the H3K36 histone marks, these results indicate the potential effect of epigenetics disturbances as result of histone mutations on the development of MSC-derived lineages.

5.4.2. High mortality in male *H3f3a+/*G34W animals

A major finding of our study is that male *H3f3a*+G34W mice have a higher mortality rate than control and female mice. The male mice also suffer from an enlarged bladder, that we speculate to be the main reason for the early demise of these mice. The level of blood creatinine, potassium, and urea nitrogen (BUN) is very high in these mice which all are indicators of disturbed kidney function (Table 3.2). The high potassium level (hyperkalemia) can lead to atrial fibrillation and sudden death.

With further investigation of the underlying cause for the enlarged bladders, we found that in bladder tissues of the male *H3f3a+/*G34W mice a set of genes important for myogenesis such as *Acta1*, *Actc1*, and *Mypn* were downregulated. These genes are associated with the actin myosin contraction apparatus in the striated muscles. The down regulation of these genes could therefore impede the actomyosin contraction of striated muscles in the urethral sphincter, which would result in reduced urine flow, as we observed in the *H3f3a*+G34W mice. The obstruction of this flow would also trigger an enlarged bladder in these mice. Our findings are supported by a study reporting a mouse model with a germline mutation in the *Acta1* gene (*Acta1* H40Y). In this study, the male mice also showed a high morbidity rate due to obstruction of the urethra, with a dilated bladder, urinary tract infection/inflammation, and hydronephrosis[178,179], all similar to the phenotypes we observed in our male *H3f3a+/*G34W mice.

5.4.3. G34W mutation in GCTB affects the expression of genes involved in muscle function

Previous reports suggest that GCTB stromal cells are early mesenchymal progenitors because they can differentiate *in vitro* to osteoblasts, chondroblasts and adipocytes, while *in vivo* they can form mineralized bone nodules [123,125,204]. We also found that early-passage GCTB stromal cells *in vitro* differentiate into osteoblasts, adipocytes and myofibroblast-like cells (data not shown). However, we also found that late-passage GCTB stromal cells lose their capacity to differentiate to osteoblasts and adipocytes (data not shown), while retaining the ability to express myofibroblast-like markers (**chapter 4: 4.6.2**). Furthermore, our experiments showed that in basic media, H3.3G34W GCTB cells have significantly lower expression of the contractile actomyosin

machinery than the edited (WT) cells. Using RNA-seq we found that removal of G34W led to a global effect on the transcriptome in three different G34W GCTB cells and their edited isogenic cell lines. Despite the fact that each of the cell lines derived from three patients with different genetic backgrounds and a large expected individual variability, we were able to identify common G34W dependent dysregulated pathways across all three cell lines. Pathway analysis of the differentially expressed genes revealed that pathways related to the actin filament processes and muscle contraction were strongly depleted in two out of the three G34W cell lines. In addition, the intersection of commonly dysregulated genes across all three isogenic cell lines revealed 27 genes, all of which were downregulated in the G34W cells compared to their edited WT isogenic cells. This group of 27 genes included genes associated with muscle contraction and suggest that G34W in neoplastic stromal cells represses genes associated with muscle contraction functions. This reinforces our finding in DKI G34W animals where the G34W mutation abrogated expression of striated muscle actomyosin related genes.

Altogether, we showed that in stromal cells of GCTB with the H3.3G34W mutation, genes related to actin and striated muscle contraction are down regulated compared to their isogenic edited (WT) cells. We proposed that this may be due to a gain of H3.3K27me3 and a loss of H3.3K36me3 at specific genes related to the actin filament pathway (**chapter 4: 4. 5**). In addition, previous studies have shown that Myogenin (Myog) is down regulated through PRC2 dependent deposition K27me3[205]. Myog is one of the master regulators of myogenesis and is involved in skeletal muscle development. *ACTA1* is one of the target genes of Myog transcription factor, and is significantly up-regulated in response to PRC2 (Suz12) depletion and elevated expression of Myog[205]. These observations further strengthen our finding that the H3.3G34W mutation prevents expression of skeletal muscle and actomyosin-related genes through inappropriate deposition of H3.3K27me3 histone mark.

5.5. G34W mutation in GCTB stalls the terminal differentiation into myofibroblasts

At single-cell resolution, G34W GCTB stromal cells comprise distinct cell populations related by a neoplastic lineage trajectory with *SPP1*+ (S1) tumor cells at the origin leading to

ACTA2+ myofibroblast progenitor (S3) tumor cells (Fig. 4.9). A terminally differentiated myofibroblast cell state is blocked by G34W, as suggested by lower expression of myofibroblast markers such as calponin 1 in G34W cells. This block can be overcome *in vitro* using strong differentiation factors (e.g. TGF-β1), an observation which may have implications for GCTB treatment, especially in recurring tumors or those where complete surgical resection is not feasible.

In summary, there are several reasons to support the hypothesis that S3 cells are myofibroblastic-like cells. In the scRNA-seq analyses of GCTB, we observed that the S3 population express markers that are well characterized as contractile myofibroblast markers such as alpha smooth muscle actin (*ACTA2*)[206,207]. In addition, our epigenetic/transcriptomic data showing that G34W actively and specifically represses genes associated with contractile function through H3K27me3, this led us to propose that G34W stromal cells may impair their ability to achieve terminal myofibroblast differentiation. After we induced myofibroblast differentiation *in vitro* with compounds such as the transforming growth factor-β1 (TGF-β1), the differentiation of fibroblasts into myofibroblasts is stimulated through expression of a contractile gene program and altered ECM features[208–210]. Notably, in *H3f3a*+/G34W mice we also noticed downregulation of myogenesis genes.

5.6. PDOX mouse model of GCTB reveals the central role of G34W mutation in tumorigenicity

With our CRISPR/Cas9 edited isogenic cell lines, we studied the tumorigenic effects of the G34W mutation *in vivo*. For this, we developed an orthotopic GCTB xenograft mouse model for GCTB, in which GCTB patient-derived cells with the G34W mutation were injected into mouse tibia. These GCTB patient-derived cells recruited murine multinucleated osteoclasts and formed a tumor which resembles GCTB. In contrast, the isogenic edited (WT) GCTB patient-derived cells did not form any tumor in both orthotopic and subcutaneous injections. Because the difference of the injected cells is only the G34W mutations, these results clearly show that G34W acts as the primary oncogenic driver in GCTB. In the literature, previous xenograft models relied on implantation of the full tumor in chick chorioallantoic membrane (CAM) and tibial nude mice models[211,212]. Fellenberg et al. showed that H3.3G34W siRNA knockdown in GCTB-derived cells

showed a significant reduction of tumor volume in CAM, but not complete inhibition of tumor formation[213]. This is in agreement with our results, and their incomplete inhibition of the tumor could be due to an incomplete siRNA knockdown. Although, these data suggested that H3.3 G34W mutation is a strong oncogenic driver in GCTB, DKI-mice with germline G34 mutations did not develop tumors. The reason for this is not totally clear, but it is possible that *in vivo* other potential cooperating mutations are needed for tumor initiation and progression. For example, mutations of *TP53* and the chromatin remodeler *ATRX*, both often co-occur with G34R/V mutations in pediatric HGG, could be needed for tumor initiation and progression in our DKI-mice with G34R/V mutations.

5.7. G34W mutation in GCTB contributes to extracellular matrix remodeling and osteoclastogenesis

scRNA-seq studies of several diseases where stroma plays a major role in its pathogenesis have identified *ACTA2*+ stromal cells expressing ECM ligands (*COL6A3, BGN*), which is consistent with a population of S3 cells. In GCTB tumorigenesis, the initiating factor in the oncogenic process is the G34W mutation which leads to a persistent activated progenitor state of the cells that are in capable of further transitioning toward a S3 myofibroblast-like cell. The presence of myeloid cells in the bone microenvironment niche enables the recruitment and syncytia of osteoclasts through a G34W-mediated ECM remodeling process enriched with these S3 *ACTA2*+ cells. Notably, secreted collagen VI (*COL6A1/3*) maintains the mechanical stiffness within the ECM, a function implicated in the activation of the mechanosensitive Ca^{2+} channel *TRPV4,* which is expressed on the plasma membrane of large osteoclasts and regulates terminal osteoclast differentiation[214,215]. We previously showed that H3 wild type giant cell lesions of the jaw (GCLJ), a disease that is histologically and radiologically similar to GCTB, carry *TRPV4* gain-of-function mutations[216], suggesting possible convergence of effects between G34W-mutant GCTB and *TRPV4*-mutant GCLJ on osteoclastogenesis and pathological features.

In conclusion, we show that G34W is necessary to drive the two major pathological features of GCTB: the destruction of bone and the maintenance of proliferating osteoprogenitors. G34W in a neoplastic stromal *ACTA2*+ population mediates the secretion of factors that recruit

osteoclasts within the bone tumor microenvironment (TME), resulting in destruction of bone. The global epigenetic remodeling process initiated by G34W's *in cis* effects helps maintain the neoplastic progenitor state. These epigenetic changes may be amenable to future therapies targeting the epigenome.

Chapter VI: Conclusions & Future Directions

6.1. Conclusion

Mice with germline H3.3G34 mutations vs isogenic GCTB-tumor derived cell lines

We generated DKI mice with the *heterozygous* germline G34R/V/W mutations and characterized the specific developmental phenotypes. We showed that each G34 mutation causing a different amino acid substitution results in tissue-specific phenotypes in the mice. To evaluate the several phenotypes, we used behavioural tests, histology and several molecular genomics tools to elucidate the role of G34 mutations on developmental processes in DKI-mice and tumorigenesis is isogenic GCTB models.

Importantly, we show that in DKI mice with germline G34W mutation, predominantly MSC-related cell lineages and tissues (e.g. adipose and muscle tissues) were affected. In G34W GCTB, we found that the gene signature of actomyosin in striated muscles is down regulated in G34W cells compared to edited (WT) cells, which is consistent with epigenetic remodeling events associated with the G34W oncohistone.

We thus hypothesize that each G34R/V/W mutation may have differential effects on readers, writers or erasers involved in the PTM of the H3.3K36 mark. This leads to a differential distribution of histone chromatin marks which accounts for the distinct phenotypes observed in our DKI mice and human G34R/V HGGs and G34W GCTB tumors.

6.2. Future directions

Our study on the DKI mice with G34 mutations has opened several new avenues of research.

We have derived neural stem cell (NSC) and mesenchymal stem cell (MSC) lines from all the different DKI mouse models that we have generated. Studying the differentiation potential and profile of the *H3f3a*+/G34R/V/W and the control H3f3a+/+ and Hf3a+/- NSCs and MSCs into different neural lineages (astrocyte, oligodendrocyte, and neuron) and mesenchymal lineages (chondrocytes, adipocytes, osteocytes, and myocytes) respectively *in vitro* will further elucidate the role of G343 mutations in the developmental processes of these neural and mesenchymal lineages.

Moreover, RNA-seq of NSCs and MSCs that are cultured in stem cell and differentiated media will help us to elucidate key developmental processes and core transcription factors which could explain the different phenotypes in mice. In addition, ChIP-seq for histone marks (e.g. K27me3, K27ac and K36me3) in the NSCs and MSCs will allow us to evaluate the epigenetic changes as consequences of G34 mutations in another isogenic model in a relevant cell-type context.

Importantly, finding difference in the proteins interacting with H3.3 G34R/V/W mutations will likely be very important to understanding the molecular pathways that are activated/inactivated by these mutations. Currently, we are collaborating with Dr. Eric Campos lab (University of Toronto, Canada) to perform BioID experiments to screen potential protein interactions with H3.3 G34R/V/W histones in mouse NSCs and MSCs cell lines, with the expectation of discovering differential binding of epigenetic writers, readers and erasers between G34R/V/W mutant histones. In addition, in collaboration with Dr. Benjamin Garcia lab (University of Pennsylvania School of Medicine, USA), we are in process of performing histone mass spectrometry to explore all the affected histone marks *in cis* or *in trans* in H3.3 G34R/V/W NSCs and MSCS.

Lastly, we have generated conditional, tissue-specific knock-in H3.3G34 mutant mice models by CRISPR/Cas9 genome editing. This allows us to precisely control their expression at different developmental time points and in specific cell types to model age and anatomical specificity of the tumors. To further explore the specific expression of H3.3G34 mutations in the central nervous system, we plan to breed conditional DKI H3.3G34 mutant mice with different Cre mouse lines such as Emx1- (dorsal forebrain), Foxg1- (forebrain), Gfap-cre (astrocytes) in order to express of the mutant histone in neuro-glial progenitors at different time-points to further study the origin of HGGs. In addition, we plan to breed conditional DKI H3.3G34 mutants with Adipoq-Cre to directly express H3.3G34 mutants in adipose tissue; Sox9-Cre or the Prx1-Cre for expression of the mutants in osteochondral progenitors; and Runx2-Cre and Osx-Cre in immature osteoblasts. Finally, to study muscle promoters, we plan to breed conditional DKI H3.3G34 mutants with Acta1-cre and Acta2-cre. Together, this will allow expression of each G34 mutation in specific cell lineages and will further understanding of the role of G34 mutations in tumor development for GCTB and other G34 containing cancers.

References

1. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).
2. Sidoli, S. & Garcia, B. A. Characterization of individual histone post-translational modifications and their combinatorial patterns by mass spectrometry-based proteomics strategies. *Methods Mol. Biol. Clifton NJ* **1528**, 121–148 (2017).
3. Aubert, Y., Egolf, S. & Capell, B. C. The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends Genet.* **35**, 645–657 (2019).
4. Noble, D. Conrad Waddington and the origin of epigenetics. *J. Exp. Biol.* **218**, 816–818 (2015).
5. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
6. Pan, M.-R., Hsu, M.-C., Chen, L.-T. & Hung, W.-C. Orchestration of H3K27 methylation: mechanisms and therapeutic implication. *Cell. Mol. Life Sci.* **75**, 209–223 (2018).
7. Baylin, S. B. & Jones, P. A. Epigenetic Determinants of Cancer. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
8. Workman, J. L. & Kingston, R. E. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579 (1998).
9. Harshman, S. W., Young, N. L., Parthun, M. R. & Freitas, M. A. H1 histones: current perspectives and challenges. *Nucleic Acids Res.* **41**, 9593–9609 (2013).
10. Peterson, C. L. & Laniel, M.-A. Histones and histone modifications. *Curr. Biol. CB* **14**, R546-551 (2004).
11. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A. & Landsman, D. Histone structure and nucleosome stability. *Expert Rev. Proteomics* **2**, 719–729 (2005).

12. Buschbeck, M. & Hake, S. B. Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat. Rev. Mol. Cell Biol.* **18**, 299–314 (2017).

13. Osley, M. A. The regulation of histone synthesis in the cell cycle. *Annu. Rev. Biochem.* **60**, 827–861 (1991).

14. Keith, K. C. *et al.* Analysis of Primary Structural Determinants That Distinguish the Centromere-Specific Function of Histone Variant Cse4p from Histone H3. *Mol. Cell. Biol.* **19**, 6130–6139 (1999).

15. Franklin, S. G. & Zweidler, A. Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature* **266**, 273–275 (1977).

16. Th, E., Ik, M. & C, L. The human histone H3 complement anno 2011. *Biochim. Biophys. Acta* **1809**, 577–586 (2011).

17. Brush, D., Dodgson, J. B., Choi, O. R., Stevens, P. W. & Engel, J. D. Replacement variant histone genes contain intervening sequences. *Mol. Cell. Biol.* **5**, 1307–1317 (1985).

18. Wells, D., Bains, W. & Kedes, L. Codon usage in histone gene families of higher eukaryotes reflects functional rather than phylogenetic relationships. *J. Mol. Evol.* **23**, 224–241 (1986).

19. Wells, D., Hoffman, D. & Kedes, L. Unusual structure, evolutionary conservation of non-coding sequences and numerous pseudogenes characterize the human H3.3 histone multigene family. *Nucleic Acids Res.* **15**, 2871–2889 (1987).

20. Bramlage, B., Kosciessa, U. & Doenecke, D. Differential expression of the murine histone genes H3.3A and H3.3B. *Differ. Res. Biol. Divers.* **62**, 13–20 (1997).

21. Yuen, B. T. K. & Knoepfler, P. S. Histone H3.3 mutations: a variant path to cancer. *Cancer Cell* **24**, (2013).

22. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res.* **21**, 421–434 (2011).

23. Tvardovskiy, A., Schwämmle, V., Kempf, S. J., Rogowska-Wrzesinska, A. & Jensen, O. N. Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nucleic Acids Res.* **45**, 9272–9289 (2017).

24. Schwartz, B. E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev.* **19**, 804–814 (2005).

25. Ng, R. K. & Gurdon, J. B. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat. Cell Biol.* **10**, 102–109 (2008).

26. Udugama, M. *et al.* Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res.* **43**, 10227–10237 (2015).

27. Zhang, H. *et al.* RPA Interacts with HIRA and Regulates H3.3 Deposition at Gene Regulatory Elements in Mammalian Cells. *Mol. Cell* **65**, 272–284 (2017).

28. Bano, D., Piazzesi, A., Salomoni, P. & Nicotera, P. The histone variant H3.3 claims its place in the crowded scene of epigenetics. *Aging* **9**, 602–614 (2017).

29. Kallappagoudar, S., Yadav, R. K., Lowe, B. R. & Partridge, J. F. Histone H3 mutations—a special role for H3.3 in tumorigenesis? *Chromosoma* **124**, 177–189 (2015).

30. Voon, H. P. J. *et al.* ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Rep.* **11**, 405–418 (2015).

31. Jang, C.-W., Shibata, Y., Starmer, J., Yee, D. & Magnuson, T. Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev.* **29**, 1377–1392 (2015).

32. Bachu, M. *et al.* A versatile mouse model of epitope-tagged histone H3.3 to study epigenome dynamics. *J. Biol. Chem.* **294**, 1904–1914 (2019).

33. Maze, I. *et al.* Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* **87**, 77–94 (2015).

34. 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* **11**, (2019).

35. Couldrey, C., Carlton, M. B. L., Nolan, P. M., Colledge, W. H. & Evans, M. J. A retroviral Gene Trap Insertion into the Histone 3.3A Gene Causes Partial Neonatal Lethality, Stunted Growth, Neuromuscular Deficits and Male Sub-fertility in Transgenic Mice. *Hum. Mol. Genet.* **8**, 2489–2495 (1999).

36. 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**, e1004964 (2015).

37. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).

38. Xu, Y.-M., Du, J.-Y. & Lau, A. T. Y. Posttranslational modifications of human histone H3: An update. *PROTEOMICS* **14**, 2047–2060 (2014).

39. Lewis, P. W. *et al.* Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science* **340**, 857–861 (2013).

40. Shi, L., Shi, J., Shi, X., Li, W. & Wen, H. Histone H3.3 G34 mutations alter histone H3K36 and H3K27 methylation in cis. *J. Mol. Biol.* **430**, 1562–1565 (2018).

41. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).

42. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32**, 491–502 (2008).

43. Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* **53**, 49–62 (2014).

44. Reinberg, D. & Vales, L. D. Chromatin domains rich in inheritance. *Science* **361**, 33–34 (2018).

45. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

46. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).

47. Wang, L. *et al.* A Role for Monomethylation of Histone H3-K27 in Gene Activity in Drosophila. *Genetics* **208**, 1023–1036 (2018).

48. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2**, 933–938 (2007).

49. Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev.* **29**, 238–249 (2015).

50. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).

51. Lavarone, E., Barbieri, C. M. & Pasini, D. Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat. Commun.* (2019) doi:10.1038/s41467-019-09624-w.

52. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126 (2012).

53. Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol. Cell* **70**, 371-379.e5 (2018).

54. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).

55. Li, L. *et al.* Dnmt3a-Mediated DNA Methylation Changes Regulate Osteogenic Differentiation of hMSCs Cultivated in the 3D Scaffolds under Oxidative Stress. *Oxid. Med. Cell. Longev.* **2019**, (2019).

56. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).

57. Sessa, A. *et al.* SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron* **104**, 271-289.e13 (2019).

58. Zaghi, M., Broccoli, V. & Sessa, A. H3K36 Methylation in Neural Development and Associated Diseases. *Front. Genet.* **10**, (2020).

59. Wen, H. *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508**, 263–268 (2014).

60. Guo, R. *et al.* BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation Decorated Chromatin to Regulated Pre-mRNA Processing. *Mol. Cell* **56**, 298–310 (2014).

61. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *EMBO Rep.* **16**, 1467–1481 (2015).

62. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2 mediated H3K27 methylation. *J. Biol. Chem.* jbc.M110.194027 (2011) doi:10.1074/jbc.M110.194027.

63. Pereira, J. D. *et al.* Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15957–15962 (2010).

64. Feng, X. *et al.* Polycomb Ezh2 controls the fate of GABAergic neurons in the embryonic cerebellum. *Dev. Camb. Engl.* **143**, 1971–1980 (2016).

65. Brennan, K. *et al.* NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci. Rep.* **7**, (2017).

66. Sotos, J. F., Dodge, P. R., Muirhead, D., Crawford, J. D. & Talbot, N. B. CEREBRAL GIGANTISM IN CHILDHOOD. A SYNDROME OF EXCESSIVELY RAPID GROWTH AND ACROMEGALIC FEATURES AND A NONPROGRESSIVE NEUROLOGIC DISORDER. *N. Engl. J. Med.* **271**, 109–116 (1964).

67. Žilina, O. *et al.* Patient with Dup(5)(q35.2-q35.3) reciprocal to the common Sotos syndrome deletion and review of the literature. *Eur. J. Med. Genet.* **56**, 202–206 (2013).

68. Battaglia, A. *et al.* Natural History of Wolf-Hirschhorn Syndrome: Experience With 15 Cases. *Pediatrics* **103**, 830–836 (1999).

69. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).

70. Näf, D. *et al.* Mouse models for the Wolf-Hirschhorn deletion syndrome. *Hum. Mol. Genet.* **10**, 91–98 (2001).

71. 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).

72. Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).

73. Tumiene, B. *et al.* Phenotype comparison confirms ZMYND11 as a critical gene for 10p15.3 microdeletion syndrome. *J. Appl. Genet.* **58**, 467–474 (2017).

74. Moskowitz, A. M. *et al.* A de novo missense mutation in ZMYND11 is associated with global developmental delay, seizures, and hypotonia. *Cold Spring Harb. Mol. Case Stud.* **2**, (2016).

75. Yates, T. M. *et al.* ZMYND11-related syndromic intellectual disability: 16 patients delineating and expanding the phenotypic spectrum. *Hum. Mutat.* **41**, 1042–1050 (2020).

76. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).

77. Wu, G. *et al.* Somatic Histone H3 Alterations in Paediatric Diffuse Intrinsic Pontine Gliomas and Non-Brainstem Glioblastomas. *Nat. Genet.* **44**, 251–253 (2012).

78. Sturm, D. *et al.* Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).

79. 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 Neuropathol. (Berl.)* **124**, 439–447 (2012).

80. Behjati, S. *et al.* Distinct H3F3A and H3F3B driver variants define chondroblastoma and giant cell tumour of bone. *Nat. Genet.* **45**, (2013).

81. Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat. Genet.* **49**, 180–185 (2017).

82. Lu, C. *et al.* Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844–849 (2016).

83. Lehnertz, B. *et al.* H3K27M/I mutations promote context-dependent transformation in acute myeloid leukemia with RUNX1 alterations. *Blood* **130**, 2204–2214 (2017).

84. Boileau, M. *et al.* Mutant H3 histones drive human pre-leukemic hematopoietic stem cell expansion and promote leukemic aggressiveness. *Nat. Commun.* **10**, (2019).

85. Pajtler, K. W. *et al.* Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta Neuropathol. (Berl.)* **136**, 211–226 (2018).

86. Fontebasso, A. M. *et al.* Epigenetic dysregulation: a novel pathway of oncogenesis in pediatric brain tumors. *Acta Neuropathol. (Berl.)* **128**, 615–627 (2014).

87. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444–450 (2014).

88. Buczkowicz, P. *et al.* Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nat. Genet.* **46**, 451–456 (2014).

89. Taylor, K. R. *et al.* Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. *Nat. Genet.* **46**, 457–461 (2014).

90. Fontebasso, A. M. *et al.* Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol. (Berl.)* **125**, 659–669 (2013).

91. Fontebasso, A. M., Liu, X.-Y., Sturm, D. & Jabado, N. Chromatin Remodeling Defects in Pediatric and Young Adult Glioblastoma: A Tale of a Variant Histone 3 Tail. *Brain Pathol.* **23**, 210–216 (2013).

92. Salloum, R. *et al.* Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. *Acta Neuropathol. Commun.* **5**, 78 (2017).

93. Boonsawat, P. *et al.* Elucidation of the phenotypic spectrum and genetic landscape in primary and secondary microcephaly. *Genet. Med.* **21**, 2043–2058 (2019).

94. 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. *Balk. J. Med. Genet. BJMG* **22**, 65–68 (2019).

95. Justin, N. *et al.* Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat. Commun.* **7**, 11316 (2016).

96. 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).

97. Weinberg, D. N., Allis, C. D. & Lu, C. Oncogenic Mechanisms of Histone H3 Mutations. *Cold Spring Harb. Perspect. Med.* **7**, a026443 (2017).

98. Harutyunyan, A. S. *et al.* H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Commun.* **10**, 1262 (2019).

99. 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).

100. Korshunov, A. *et al.* Histologically distinct neuroepithelial tumors with histone 3 G34 mutation are molecularly similar and comprise a single nosologic entity. *Acta Neuropathol. (Berl.)* **131**, 137–146 (2016).

101. Amorim, J., Santos, G., Vinagre, J. & Soares, P. The Role of ATRX in the Alternative Lengthening of Telomeres (ALT) Phenotype. *Genes* **7**, 66 (2016).

102. Bjerke, L. *et al.* Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN. *Cancer Discov.* **3**, 512–519 (2013).

103. Andreiuolo, F. *et al.* H3F3A-G34R mutant high grade neuroepithelial neoplasms with glial and dysplastic ganglion cell components. *Acta Neuropathol. Commun.* **7**, 78 (2019).

104. Cheng, Y., Bao, W. & Wu, Q. Cerebral hemispheric glioblastoma with PNET-like morphology and histone H3.3 G34 mutation in younger patients: Report of three rare cases and diagnostic pitfalls. *Indian J. Pathol. Microbiol.* **63**, 262–266 (2020).

105. 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).

106. Alcantara Llaguno, S. *et al.* Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model. *Cancer Cell* **15**, 45–56 (2009).

107. Friedmann-Morvinski, D. *et al.* Dedifferentiation of Neurons and Astrocytes by Oncogenes Can Induce Gliomas in Mice. *Science* **338**, 1080–1084 (2012).

108. Parsons, D. W. *et al.* An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science* **321**, 1807 (2008).

109. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).

110. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

111. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).

112. Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat. Cell Biol.* **12**, 853–862 (2010).

113. 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**, 684-700.e9 (2017).

114. Mohammad, F. *et al.* EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* **23**, 483–492 (2017).

115. 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* **346**, 1529–1533 (2014).

116. Larson, J. D. *et al.* Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression. *Cancer Cell* **35**, 140-155.e7 (2019).

117. Eckardt, J. J. & Grogan, T. J. Giant cell tumor of bone. *Clin. Orthop.* 45–58 (1986).

118. Manfrini, M. Giant Cell Tumor. in *Diagnosis of Musculoskeletal Tumors and Tumor-like Conditions: Clinical, Radiological and Histological Correlations - The Rizzoli Case Archive* (eds. Picci, P. et al.) 101–105 (Springer International Publishing, 2020). doi:10.1007/978-3-030-29676-6_22.

119. Cleven, A. H. G. *et al.* Mutation Analysis of H3F3A and H3F3B as a Diagnostic Tool for Giant Cell Tumor of Bone and Chondroblastoma. *Am. J. Surg. Pathol.* **39**, 1576–1583 (2015).

120. Ogura, K. *et al.* Highly recurrent H3F3A mutations with additional epigenetic regulator alterations in giant cell tumor of bone. *Genes. Chromosomes Cancer* **56**, 711–718 (2017).

121. Xu, L. *et al.* Soft tissue recurrence of giant cell tumor of the bone: Prevalence and radiographic features. *J. Bone Oncol.* **9**, 10–14 (2017).

122. Lim, J. *et al.* The histone variant H3.3 G34W substitution in giant cell tumor of the bone link chromatin and RNA processing. *Sci. Rep.* **7**, 13459 (2017).

123. James, I. E., Dodds, R. A., Olivera, D. L., Nuttall, M. E. & Gowen, M. Human osteoclastoma-derived stromal cells: correlation of the ability to form mineralized nodules in vitro with formation of bone in vivo. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **11**, 1453–1460 (1996).

124. Joyner, C. J., Quinn, J. M., Triffitt, J. T., Owen, M. E. & Athanasou, N. A. Phenotypic characterisation of mononuclear and multinucleated cells of giant cell tumour of bone. *Bone Miner.* **16**, 37–48 (1992).

125. Wülling, M., Delling, G. & Kaiser, E. The origin of the neoplastic stromal cell in giant cell tumor of bone. *Hum. Pathol.* **34**, 983–993 (2003).

126. Wittrant, Y. *et al.* RANKL/RANK/OPG: new therapeutic targets in bone tumours and associated osteolysis. *Biochim. Biophys. Acta* **1704**, 49–57 (2004).

127. Cowan, R. W. & Singh, G. Giant cell tumor of bone: a basic science perspective. *Bone* **52**, 238–246 (2013).

128. Noh, B.-J. & Park, Y.-K. Giant cell tumor of bone: updated molecular pathogenesis and tumor biology. *Hum. Pathol.* **81**, 1–8 (2018).

129. Cheng, Z. *et al.* A molecular threading mechanism underlies Jumonji lysine demethylase KDM2A regulation of methylated H3K36. *Genes Dev.* **28**, 1758–1771 (2014).

130. Zhang, Y. *et al.* Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci. Rep.* **7**, 43906 (2017).

131. Jani, K. S. *et al.* Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. *Proc. Natl. Acad. Sci.* **116**, 8295–8300 (2019).

132. Voon, H. P. J. *et al.* Inhibition of a K9/K36 demethylase by an H3.3 point mutation found in paediatric glioblastoma. *Nat. Commun.* **9**, 3142 (2018).

133. Nombela-Arrieta, C., Ritz, J. & Silberstein, L. E. The elusive nature and function of mesenchymal stem cells. *Nat. Rev. Mol. Cell Biol.* **12**, 126–131 (2011).

134. Ciuffreda, M. C., Malpasso, G., Musarò, P., Turco, V. & Gnecchi, M. Protocols for in vitro Differentiation of Human Mesenchymal Stem Cells into Osteogenic, Chondrogenic and Adipogenic Lineages. *Methods Mol. Biol. Clifton NJ* **1416**, 149–158 (2016).

135. Meyer, M. B., Benkusky, N. A., Sen, B., Rubin, J. & Pike, J. W. Epigenetic Plasticity Drives Adipogenic and Osteogenic Differentiation of Marrow-derived Mesenchymal Stem Cells. *J. Biol. Chem.* **291**, 17829–17847 (2016).

136. Musri, M. M., Gomis, R. & Párrizas, M. A chromatin perspective of adipogenesis. *Organogenesis* **6**, 15–23 (2010).

137. Herlofsen, S. R. *et al.* Genome-wide map of quantified epigenetic changes during in vitro chondrogenic differentiation of primary human mesenchymal stem cells. *BMC Genomics* **14**, 105 (2013).

138. Wang, L. *et al.* H3K36 trimethylation mediated by SETD2 regulates the fate of bone marrow mesenchymal stem cells. *PLOS Biol.* **16**, e2006522 (2018).

139. Zhuang, L. *et al.* Depletion of Nsd2-mediated histone H3K36 methylation impairs adipose tissue development and function. *Nat. Commun.* **9**, 1796 (2018).

140. Hemming, S. *et al.* EZH2 and KDM6A Act as an Epigenetic Switch to Regulate Mesenchymal Stem Cell Lineage Specification. *STEM CELLS* **32**, 802–815 (2014).

141. Wang, L., Jin, Q., Lee, J.-E., Su, I. -hsin & Ge, K. Histone H3K27 methyltransferase Ezh2 represses Wnt genes to facilitate adipogenesis. *Proc. Natl. Acad. Sci.* **107**, 7317–7322 (2010).

142. Yapp, C., Carr, A. J., Price, A., Oppermann, U. & Snelling, S. J. B. H3K27me3 demethylases regulate in vitro chondrogenesis and chondrocyte activity in osteoarthritis. *Arthritis Res. Ther.* **18**, (2016).

143. Sakurai, T., Watanabe, S., Kamiyoshi, A., Sato, M. & Shindo, T. A single blastocyst assay optimized for detecting CRISPR/Cas9 system-induced indel mutations in mice. *BMC Biotechnol.* **14**, 69 (2014).

144. Guyenet, S. J. *et al.* A Simple Composite Phenotype Scoring System for Evaluating Mouse Models of Cerebellar Ataxia. *J. Vis. Exp. JoVE* (2010) doi:10.3791/1787.

145. Scekic-Zahirovic, J. *et al.* Motor neuron intrinsic and extrinsic mechanisms contribute to the pathogenesis of FUS-associated amyotrophic lateral sclerosis. *Acta Neuropathol. (Berl.)* **133**, 887–906 (2017).

146. Brun, J. *et al.* Smooth Muscle-Like Cells Generated from Human Mesenchymal Stromal Cells Display Marker Gene Expression and Electrophysiological Competence Comparable to Bladder Smooth Muscle Cells. *PLoS ONE* **10**, (2015).

147. Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X. & Garcia, B. A. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J. Vis. Exp. JoVE* (2016) doi:10.3791/54112.

148. Sircar, K. *et al.* Isolation of human prostatic epithelial plasma membranes for proteomics using mirror image tissue banking of radical prostatectomy specimens. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **12**, 4178–4184 (2006).

149. Pfammatter, S. *et al.* Extending the Comprehensiveness of Immunopeptidome Analyses Using Isobaric Peptide Labeling. *Anal. Chem.* **92**, 9194–9204 (2020).

150. Harutyunyan, A. S. *et al.* H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Commun.* **10**, 1–13 (2019).

151. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat. Genet.* **51**, 1702–1713 (2019).

152. Bourgey, M. *et al. GenPipes: an open-source framework for distributed and scalable genomic analyses*. http://biorxiv.org/lookup/doi/10.1101/459552 (2018) doi:10.1101/459552.

153. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).

154. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

155. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

156. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

157. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

158. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).

159. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

160. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122 (2016).

161. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

162. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

163. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

164. Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinforma. Oxf. Engl.* **32**, 1241–1243 (2016).

165. Sievert, C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. (Chapman and Hall/CRC, 2020).

166. Baryawno, N. *et al.* A Cellular Taxonomy of the Bone Marrow Stroma in Homeostasis and Leukemia. *Cell* **177**, 1915-1932.e16 (2019).

167. Tikhonova, A. N. *et al.* The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).

168. Ximerakis, M. *et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).

169. Wilkinson, A. W. *et al.* SETD3 is an actin histidine methyltransferase that prevents primary dystocia. *Nature* **565**, 372–376 (2019).

170. Brunt, E. M., Janney, C. G., Di Bisceglie, A. M., Neuschwander-Tetri, B. A. & Bacon, B. R. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am. J. Gastroenterol.* **94**, 2467–2474 (1999).

171. Sheth, S. G., Gordon, F. D. & Chopra, S. Nonalcoholic steatohepatitis. *Ann. Intern. Med.* **126**, 137–145 (1997).

172. Murano, I. *et al.* Dead adipocytes, detected as crown-like structures, are prevalent in visceral fat depots of genetically obese mice. *J. Lipid Res.* **49**, 1562–1568 (2008).

173. Morrison, C. D. Leptin resistance and the response to positive energy balance. *Physiol. Behav.* **94**, 660–663 (2008).

174. Moreno-Viedma, V. *et al.* Common dysregulated pathways in obese adipose tissue and atherosclerosis. *Cardiovasc. Diabetol.* **15**, 120 (2016).

175. Nishimura, S. *et al.* ENPP2 Contributes to Adipose Tissue Expansion and Insulin Resistance in Diet-Induced Obesity. *Diabetes* **63**, 4154–4164 (2014).

176. Fang, L. *et al.* PPARgene: A Database of Experimentally Verified and Computationally Predicted PPAR Target Genes. *PPAR Research* vol. 2016 e6042162 https://www.hindawi.com/journals/ppar/2016/6042162/ (2016).

177. Grünberg, J. R. *et al.* CCN5/WISP2 and metabolic diseases. *J. Cell Commun. Signal.* **12**, 309–318 (2018).

178. Nguyen, M.-A. T. *et al.* Hypertrophy and dietary tyrosine ameliorate the phenotypes of a mouse model of severe nemaline myopathy. *Brain* **134**, 3516–3529 (2011).

179. Tinklenberg, J. *et al.* Treatment with ActRIIB-mFc Produces Myofiber Growth and Improves Lifespan in the Acta1 H40Y Murine Model of Nemaline Myopathy. *Am. J. Pathol.* **186**, 1568–1581 (2016).

180. Chan, C. *et al.* Myopathy-inducing mutation H40Y in ACTA1 hampers actin filament structure and function. *Biochim. Biophys. Acta* **1862**, 1453–1458 (2016).

181. Lalonde, R. & Strazielle, C. Brain regions and genes affecting limb-clasping responses. *Brain Res. Rev.* **67**, 252–259 (2011).

182. Seibenhener, M. L. & Wooten, M. C. Use of the Open Field Maze to Measure Locomotor and Anxiety-like Behavior in Mice. *J. Vis. Exp. JoVE* (2015) doi:10.3791/52434.

183. An Introduction to Gait Analysis - 4th Edition. https://www.elsevier.com/books/an-introduction-to-gait-analysis/whittle/978-0-7506-8883-3.

184. Liu, Y. *et al.* Quantitative gait analysis of long-term locomotion deficits in classical unilateral striatal intracerebral hemorrhage rat model. *Behav. Brain Res.* **257**, 166–177 (2013).

185. White, J. J. & Sillitoe, R. V. Development of the cerebellum: from gene expression patterns to circuit maps. *WIREs Dev. Biol.* **2**, 149–164 (2013).

186. Sl, W., A, K., Gd, O. & Al, J. Spatially restricted and developmentally dynamic expression of engrailed genes in multiple cerebellar cell types. *Cerebellum Lond. Engl.* **10**, 356–372 (2011).

187. Cendelin, J. From mice to men: lessons from mutant ataxic mice. *Cerebellum Ataxias* **1**, 4 (2014).

188. Schnitzer, J., Franke, W. W. & Schachner, M. Immunocytochemical demonstration of vimentin in astrocytes and ependymal cells of developing and adult mouse nervous system. *J. Cell Biol.* **90**, 435–447 (1981).

189. Raponi, E. *et al.* S100B expression defines a state in which GFAP-expressing cells lose their neural stem cell potential and acquire a more mature developmental stage. *Glia* **55**, 165–177 (2007).

190. 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).

191. An illustrated anatomical ontology of the developing mouse lower urogenital tract. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4440924/.

192. Shi, L., Shi, J., Shi, X., Li, W. & Wen, H. Histone H3.3 G34 Mutations Alter Histone H3K36 and H3K27 Methylation In Cis. *J. Mol. Biol.* **430**, 1562–1565 (2018).

193. Hynes, R. O. The extracellular matrix: not just pretty fibrils. *Science* **326**, 1216–1219 (2009).

194. Ravenscroft, G. *et al.* Bi-allelic mutations in MYL1 cause a severe congenital myopathy. *Hum. Mol. Genet.* **27**, 4263–4272 (2018).

195. Nanda, V. *et al.* Functional regulatory mechanism of smooth muscle cell-restricted LMOD1 coronary artery disease locus. *PLOS Genet.* **14**, e1007755 (2018).

196. Gabay, L., Lowell, S., Rubin, L. L. & Anderson, D. J. Deregulation of Dorsoventral Patterning by FGF Confers Trilineage Differentiation Capacity on CNS Stem Cells In Vitro. *Neuron* **40**, 485–499 (2003).

197. Fukuda, S., Kondo, T., Takebayashi, H. & Taga, T. Negative regulatory effect of an oligodendrocytic bHLH factor OLIG2 on the astrocytic differentiation pathway. *Cell Death Differ.* **11**, 196–202 (2004).

198. Cai, J. *et al.* A crucial role for Olig2 in white matter astrocyte development. *Development* **134**, 1887–1899 (2007).

199. Weng, Q. *et al.* Single-Cell Transcriptomics Uncovers Glial Progenitor Diversity and Cell Fate Determinants during Development and Gliomagenesis. *Cell Stem Cell* **24**, 707-723.e8 (2019).

200. Ben Haim, L., Carrillo-de Sauvage, M.-A., Ceyzériat, K. & Escartin, C. Elusive roles for reactive astrocytes in neurodegenerative diseases. *Front. Cell. Neurosci.* **9**, (2015).

201. Bignami, A. & Dahl, D. The Astroglial Response to Stabbing. Immunofluorescence Studies with Antibodies to Astrocyte-Specific Protein (gfa) in Mammalian and Submammalian Vertebrates. *Neuropathol. Appl. Neurobiol.* **2**, 99–110 (1976).

202. Kamphuis, W., Orre, M., Kooijman, L., Dahmen, M. & Hol, E. M. Differential cell proliferation in the cortex of the APPswePS1dE9 Alzheimer's disease mouse model. *Glia* **60**, 615–629 (2012).

203. Longo, M. *et al.* Adipose Tissue Dysfunction as Determinant of Obesity-Associated Metabolic Complications. *Int. J. Mol. Sci.* **20**, (2019).

204. Joyner, C. J., Quinn, J. M., Triffitt, J. T., Owen, M. E. & Athanasou, N. A. Phenotypic characterisation of mononuclear and multinucleated cells of giant cell tumour of bone. *Bone Miner.* **16**, 37–48 (1992).

205. Asp, P. *et al.* Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc. Natl. Acad. Sci.* **108**, E149–E158 (2011).

206. Ciuba, K. *et al.* Calponin-3 is critical for coordinated contractility of actin stress fibers. *Sci. Rep.* **8**, 17670 (2018).

207. Otranto, M. *et al.* The role of the myofibroblast in tumor stroma remodeling. *Cell Adhes. Migr.* **6**, 203–219 (2012).

208. Walker, G. A., Masters, K. S., Shah, D. N., Anseth, K. S. & Leinwand, L. A. Valvular myofibroblast activation by transforming growth factor-beta: implications for pathological extracellular matrix remodeling in heart valve disease. *Circ. Res.* **95**, 253–260 (2004).

209. Paranya, G. *et al.* Aortic valve endothelial cells undergo transforming growth factor-beta-mediated and non-transforming growth factor-beta-mediated transdifferentiation in vitro. *Am. J. Pathol.* **159**, 1335–1343 (2001).

210. Scharenberg, M. A. *et al.* TGF-β-induced differentiation into myofibroblasts involves specific regulation of two MKL1 isoforms. *J. Cell Sci.* **127**, 1079–1091 (2014).

211. Balke, M. *et al.* A short-term in vivo model for giant cell tumor of bone. *BMC Cancer* **11**, 241 (2011).

212. Xu, L., Wu, Z., Zhou, Z., Yang, X. & Xiao, J. Intratibial injection of patient-derived tumor cells from giant cell tumor of bone elicits osteolytic reaction in nude mouse. *Oncol. Lett.* **16**, 4649–4655 (2018).

213. Fellenberg, J. *et al.* Knock-down of oncohistone H3F3A-G34W counteracts the neoplastic phenotype of giant cell tumor of bone derived stromal cells. *Cancer Lett.* **448**, 61–69 (2019).

214. Zelenski, N. A. *et al.* Type VI Collagen Regulates Pericellular Matrix Properties, Chondrocyte Swelling, and Mechanotransduction in Mouse Articular Cartilage. *Arthritis Rheumatol. Hoboken NJ* **67**, 1286–1294 (2015).

215. Masuyama, R. *et al.* TRPV4-mediated calcium influx regulates terminal differentiation of osteoclasts. *Cell Metab.* **8**, 257–265 (2008).

216. Gomes, C. C. *et al.* TRPV4 and KRAS and FGFR1 gain-of-function mutations drive giant cell lesions of the jaw. *Nat. Commun.* **9**, 4572 (2018).

Appendix

Centre universitaire de santé McGill | McGill University Health Centre

Amendment to an approved study

Submit date: **2020-06-05 13:08**
Project's REB approbation date: **2017-05-11**
Project number: **MP-37-2017-3256**
Form status: **Approved**

Submitted by: **Anillo, Emma**
Nagano identifier: **ORCYD**
Form: **F1-60152**

Administration - REB

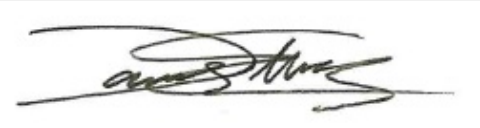
1. **MUHC REB Panel & Co-chair(s):**
   Cells, tissues, genetics & qualitative research (CTGQ)
   Co-chair: Marie Hirtle

2. **REB Decision:**
   Approved - REB delegated review

3. **Comments on the decision:**

   COVID-19 Given the impact of COVID-19 on the Quebec Health Care Network, including its impact on research activities, the implementation of the notification contained in this submission must be done in compliance with the specific guidelines in application at each participating institutions. At the MUHC, this may include having to submit an F3b form once the state of emergency has been lifted (see Nagano homepage).

4. **Date of the REB final decision & signature**
   2020-06-16
   Signature



James Ellasus
MUHC REB Coordinator
for MUHC Co-chair mentioned above

5. **Document(s)approved by the REB**
20200605 ORCYD framework v1.8_REBapproved.docx

A. General information

1. **Indicate the full title of the research study**

Oncology Repository for Children and Young Adults (ORCYD)

2. **If relevant, indicate the full study title in French**

CENTRE DE STOCKAGE ONCOLOGIQUE POUR ENFANTS ET JEUNES ADULTES (ORCYD)

3. **Indicate the name of the Principal Investigator in our institution (MUHC)**

Jabado, Nada

From which department is the principal investigator?

Pediatrics

Division

Hematology

4. **Are there local co-investigators & collaborators involved in this project?**

Yes

List all the local co-investigators & collaborators involved in the research study

Garzia, Livia

Turcotte, Robert E.

5. **For each participating centre part of the Québec health and social services network (RSSS), indicate the name of the external investigator**

Valérie Larouche

What is the name of the participating center(s)?

CHU-de-Québec

6. **Indicate the source(s) of funding for the research project (please check all that apply)**

- ☐ Study funded by private industry(e.g. pharmaceutical or biotech company)
- ☐ Investigator-initiated with industry funding study
- ☐ Study funded by a federal funding agency
- ☐ Research funded by a provincial funding agency
- ☐ Research funded by a federal ministry
- ☐ Study funded by a provincial ministry
- ☐ Research funded by US Federal funding (eg: NIH, VA) or other countries
- ☑ Research funding from a non-profit (e.g. charity, foundation, etc.)

  Indicate the name of the foundation:

  Montreal Children's Hospital Foundation

  Please attach any related files (contract, grant, budget, etc), if any, to the "study file" section.

- ☐ Internal funding (eg: investigator/departmental funds,
- ☐ Other source(s) (e.g. self-funded)
- ☐ No funding

B. Amendment

1. **This notification is concerning:**

   all participating institutions where the study is in progress

   Are there institutions for which ethics approval has been suspended or has not yet been granted?

   No

2. **Indicate the current study status at MUHC.**

   Study and recruitment in progress

3. **Please provide details about the amendment**

Please summarize the amendment(s)

Added a sub section under 5.2
Allowed windows to obtain consents - Exceptional circumstances

Please explain why these amendments are required

This section was edited to clarify and outline the allowed windows to obtain consents - Exceptional circumstances

First two exceptional circumstances described are already REB approved but we adding a third one to request a 10-15 weeks window to approach and consent patients-family's with particular religious practices. This window will allow research staff more flexibility to meet potential participants if needed.

This general situation was added so to avoid repeatedly REB submissions for special permission if the research team encounters a challenges to meet families due to religious practices in the future. However, we currently have a potential patient that meets this exceptional circumstance. Under such particular circumstances, we would like to request REB's permission to extend the allowed period to get consent for this particular case (PS-20-0985 (ATRT)) as they may request additional time to discuss it within their community.

Does this amendment require that participants be re-consented?

No

4. **What type of amendment is being requested**

Amendment to the protocol

Yes

Please attach amended protocol

20200605 ORCYD framework v1.8.docx

Version date

2020-06-05

Amendment to the investigator brochure

No

Amendment to the Information and Consent Form

No

New Information and Consent Form added

No

Amendment to documents used to recruit participants such as advertisements

No

Amendments to questionnaires or other documents participants fill out

No

New questionnaire or document to complete is being added

No

Changes to research team

No

Changes to study financing or budget

No

Other amendment, specify

No

5. **In our institution, how many participants will be affected by this amendment?**

0

How many participants have been recruited to date?

135

6. **Does this amendment require Health Canada approval?**

No

7. **Does the modification have an impact on other types of specific evaluations in the institution? (eg. use of resources, contract, budget, pharmacy, etc.)**

No

D. Signature

1. **I confirm that all information is complete & accurate.**

First & last name of person who completed the submission

Emma Anillo
2020-06-05 13:08

McGill

August 24, 2020

Animal Certificate

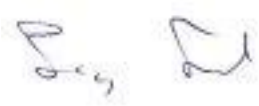
This is to certify that **Dr. Nada Jabado**, RI-MUHC Glen site, currently holds an approved **Animal Use Protocol #2010-5684** with McGill University and its Affiliated Hospital's Research Institutes for the following project:

Animal Use Protocol Title: Oncohistones in Cancer; Tackling Childhood Brain Cancer at the root to Improve survival and quality of life;Signature; Oncohistones: Role of Histone H3 Mutations in the Oncogenesis of Pediatric Cancers;Immuno-modulation to treat poor prognosis pediatric brain tumours.

Start date: July 1, 2020

Expiration date: June 30, 2021

McGill University and Affiliated Hospitals Research Institutes recognize the importance of animal research in our efforts to further our knowledge of natural processes, diseases and conservation. Research, educational and testing projects are conducted with full commitment to the wellbeing of the animal subjects. In order to limit animal use to meritorious research or educational projects, the institution relies on stringent peer review processes, along with assessment of ethical issues by the Animal Care Committee. McGill University recognizes that the use of animals in research, teaching and testing carries significant responsibilities. The institution will continue to develop and maintain guidelines and regulations, following the high standards established by the Canadian Council on Animal Care. It is committed to conducting the highest-quality research and to providing animals with the best care.



Suzanne Smith
Director, Animal Compliance Office
Office of Vice-Principal (Research and Innovation)
Suite 325, James Administration Building, McGill University
845 Sherbrooke Street West, Montreal, Quebec, Canada H3A 0G4
suzanne.smith@mcgill.ca