Characterization of Neural Development : Linking Retinoic Acid Receptors to Cell Fate and Modelling Tumorigenesis in Brain Organoids

Aysis Koshy

THÈSE DE DOCTORAT

NNT : 2023UPASL119

université PARIS-SACLAY

Characterization of Neural Development: Linking Retinoic Acid Receptors to Cell Fate and Modelling Tumorigenesis in Brain Organoids

Caractérisation du Développement Neuronal : Le rôle des Récepteurs de L'acide Rétinoïque dans le contrôle du Destin Cellulaire et Modélisation de la Tumorigenèse dans des Organoïdes cérébraux

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577 Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Sciences de la vie et de la santé
Graduate School : Life Sciences and Health. Référent : Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche **Génomique métabolique** (Université Paris-Saclay, Univ Evry, CNRS, CEA), sous la direction de **Marco Antonio MENDOZA-PARRA, Docteur**

Thèse soutenue à Paris-Saclay, le 7 Décembre 2023, par

Aysis Maria KOSHY

Composition du Jury

Membres du jury avec voix délibérative

Laurence VANDEL Director of research, CNRS, GRed, Clermont Ferrand	Présidente & Rapporteure
Jonathan CORCORAN Professor, Kings College London, London	Rapporteur
Jean-Pierre MOTHET Director of research, CNRS, ENS-Paris-Saclay	Examinateur
Kristine SCHAUER Director of research, CNRS, Institut Gustave Roussy	Examinatrice

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Titre : Caractérisation du Développement Neuronal : Le rôle des Récepteurs de L'acide Rétinoïque dans le contrôle du Destin Cellulaire et Modélisation de la Tumorigenèse dans des Organoïdes cérébraux

Mots clés : Différenciation Cellulaire, Récepteurs nucléaires à l'Acide Rétinoïque, Organoïdes Cérébraux de Souris, RNAseq Unicellulaire, Glioblastome

Résumé : Le développement du système nerveux central dans l'embryon dépend d'une signalisation opportune et précise des molécules. L'acide rétinoïque est l'une de ces molécules bien caractérisées par son impact sur le développement du cerveau et des yeux. Sous sa forme métaboliquement active, l'ATRA (acide All Trans Retinoïque) se lie aux récepteurs de l'acide rétinoïque (RAR) et contrôle l'expression d'une panoplie de gènes participant à des évènements impliqués dans la maturation cellulaire ainsi qu'à l'apoptose. Le RAR existe sous trois isotypes – RARα, RARβ et RARγ. Au cours du développement embryonnaire, chaque isotype est présent à des endroits spatialement distincts, influençant la structuration et la maturation. L'état actuel de la recherche est limité à la corrélation entre un isotype RAR spécifique et un destin cellulaire particulier. Dans cette thèse, nous discutons des résultats qui indiquent la capacité de RARβ et RARγ à restaurer de manière synergique la spécialisation cellulaire en détournant les programmes génétiques contrôlés par RARα. Dans une approche RNAseq unicellulaire, nous sommes en mesure de visualiser plusieurs clusters uniques à l'activation de RARβ + RARγ au cours de la différenciation des cellules souches allant au délà de précurseurs neuronaux.

Dans le même ordre d'idées, l'étude du développement du tissu nerveux dans le contexte de maladies est pertinente pour comprendre les caractéristiques de la maladie et identifier les options thérapeutiques ciblées. Dans cet esprit, nous avons souhaité développer un modèle organoïde cérébral de souris (BORG) à partir de cellules ES de souris mutantes H3F3A (H3.3K27M et H3.3G34R) en tant que modèle de recherche *in vitro* rentable et reproductible. À ce jour, il n'existe pas de modèle BORG de souris modélisant le cancer à partir de cellules ES mutantes H3F3A, d'autant plus que ces modèles peuvent être bénéfiques pour étudier de manière rentable les événements de cancer précoces. Ici, nous avons étudié plusieurs méthodes de génération d'un BORG tumorigène en utilisant les cellules ES mutantes H3F3A ainsi qu'en incorporant des mutations cancéreuses connues telles que l'inactivation de TP53 et la surexpression de MYCN. Nous avons réussi à démontrer l'existence d'un BORG de souris tumorigène comme celui-ci, qui héberge une signature knock-out TP53.

Title : Characterization of Neural Development: Linking Retinoic Acid Receptors to Cell Fate and Modelling Tumorigenesis in Mouse Brain Organoids

Keywords : Cell Differentiation, Retinoic Acid Receptors, Mouse Brain Organoids, Single-Cell RNAseq, Glioblastoma

Abstract : The development of the Central Nervous system in the embryo depends on timely and precise signaling of molecules. Retinoic acid is one such molecule well characterized for its impact in brain and eye development. In its metabolically active form, ATRA (All Trans Retinoic acid) binds Retinoic acid receptors (RAR), and controls downstream gene expression attributed to cell maturation and apoptosis. The RAR exists as three isotypes – RARα, RARβ, & RARγ. During embryological development, each isotype is present in spatially distinct locations influencing patterning and maturation. The current state of research is limited to the correlation of a specific RAR isotype to a particular cell fate. In this thesis, we discuss findings that point to the ability of RARβ & RARγ to synergistically restore cell specialization by hijacking RARα-controlled gene programs. In a single-cell RNAseq approach, we are able to visualize several clusters unique to RARβ + RARγ activation during mouse stem cell differentiation beyond neuronal precursor stages.

In a similar vein, studying nervous tissue development in the context of diseases is relevant to understanding disease characteristics and identifying targeted therapy options. With this in mind, we wanted to develop a mouse brain organoid (BORG) model from H3F3A mutant (H3.3K27M and H3.3G34R) mouse ES cells as an *in vitro* research model that is cost effective and reproducible. To date, there doesn't exist a mouse BORG model that models cancer from H3F3A mutant ES cells, especially since these models can be beneficial to study early cancer events cost-effectively. Here, we have investigate multiple methods of generating a tumorigenic BORG by using the H3F3A mutant ES cells as well as incorporating known cancer mutations such as TP53 knockout and MYCN overexpression. We successfully managed to show proof of a tumorigenic like mouse BORG that harbors a TP53 knockout signature.

Acknowledgements

In the course of my time spent at the Genoscope Institute where I first interned as a Master student and subsequently as a PhD student, I have met and cherished my relationships with many.

First and foremost, I would like to express my gratitude to my supervisor Dr. Marco Mendoza-Parra, for his constant guidance and support in my work. I have been fortunate to receive his mentorship that has helped me to grow as a researcher.

I would also like to thank my team members, Dr. Bruno Colombo, Aude Bramoulle, Antoine Aubert, Ariel Galindo, Gwendoline Lozachmeur, Maria Ferri, Peggy Azzolin, and former team members, Elodie Mathieux, Francois Studer, Julien Mollet, for being on this journey with me. I am grateful for the friendship we have built and the good memories we have working together. I have gained a lot of knowledge in our interactions about our research but also the French culture! I especially want to thank Aude Bramoulle who helped me often to feed my organoid babies and in various routine cell culture tasks, and Ariel Galindo who has performed the single cell transcriptome data processing.

I would like to thank the PhD students that I built friendships with and the staff at Genoscope, for welcoming me and helping me whenever I needed it.

I want to thank the Jury members- Dr. Jonathan Corcoran and Dr. Laurence Vandel as the rapporteurs and Dr. Kristine Schauer and Dr. Jean-Pierre Mothet, as examiners and Dr. Maxence Wisztorski for accepting to be on my thesis defense panel and review my work.

Dear Mamma and Pappa, I would not be here if it were not for your support. You'll pushed me to realize more than I thought I was capable of, and I cannot thank you'll enough. Being away from both of you for so long has been difficult, but your laughter and love kept me going. To Pappa, I constantly hope to make you proud and follow in your footsteps of building my life with integrity and curiosity. Mamma, you've been a constant source of comfort, especially through all my doubts and worries. From accompanying me all the way to France to sending me tasty recipes to try at home, to the daily morning inspirational quotes, I will never be able to thank you enough for checking up on me regularly. Sela, your presence grew immensely in my life this past year and I will forever be grateful. As I reflect on the times I sat

down to write my thesis, I am constantly reminded of your efforts to encourage me and feed me; it fills me with delight to no end, you're the best sister anyone could ask for.

While this part may be comprehensive, it certainly isn't exhaustive and I want to thank all who had a role to play during my time as a PhD student.

Contents

List of Figures

List of Supplementary Figures

List of Abbreviations

ADH- Alcohol Dehydrogenase

Alpha-SMA- Alpha Smooth Muscle Actin

ATRA- All Trans Retinoic Acid

BDNF- Brain-Derived Neurotrophic Factor

BMP- Bone Morphogenetic Protein

BMS641 RARβ Agonist (Beta)

BMS753- RARα Agonist (Alpha)

BMS961 RARγ Agonist (Gamma)

BORG Brain Organoids

CNP- Caudal Neuropore

CNS- Central Nervous System

CSC- Cancer Stem Cell

Ct- Cycle Threshold

DIPG- Diffuse Intrinsic Pontine Glioma

DMSO- Dimethyl Sulfoxide

DR- Direct Repeats

EB- Embryoid Body

EC-Embryonic Carcinoma Cells

EGF- Epidermal Growth Factor

ES- Embryonic Stem Cells

ETOH- Ethanol

FGF- Fibroblast Growth Factor

FL- Fore Limb

FNP- Frontonasal Process

GBM- Glioblastoma

GFP- Green Fluorescent Protein

GRN- Gene Regulator Network

GRP- Gene Regulatory Program

GSC- Glioma Stem Cells

HD- Homeobox Domain

hESC- Human Embryonic Stem Cells

hiPSC- Human Induced Pluripotent Stem Cells

HL- Hind Limb

hPSC- Human Pluripotent Stem Cells

KO- Knockout

LGE- Lateral Ganglionic Eminence

LIF- Leukemia Inhibitory Factor

LMC- Lateral Motor Column

MEF- Mouse Embryonic Feeder Cells

MES- Mouse Embryonic Stem Cell

MESC- Mouse Embryonic Stem Cells

MSN- Medium Spiny Neuron

MWOL Mouse Embryonic Stem Cell Without LIF

NI- Neural Induction

NPC- Neural Progenitor Cells

NSC- Neural Stem Cells

OPC- Oligodendrocyte Precursors

pGBM- Pediatric Glioblastoma

pHGG- Pediatric High Grade Gliomas

pLGGs- Pediatric Low Grade Gliomas

PRC2- Polycomb Repressive Complex 2

Pv- Prevertebrae

RA- Retinoic Acid

RAR- Retinoic Acid Receptor

RARE- Retinoic Acid Receptor Element

RXR- Retinoic X Receptor

RXRE- Retinoic X Receptor Element

SAG- Smoothened Agonist

SC- Stem Cells

SHH- Sonic Hedgehog

SVZ- Sub Ventricular Zone

TF- Transcription Factor

VITA- Vitamin A

VZ- Ventricular Zone

WT- Wild Type

Introduction

Retinoid driven Neurogenesis

Chronological hallmarks of the developing nervous system in the mouse embryo

The Nervous system can be classified into parts: The Central Nervous system (CNS) comprising the brain and the spinal cord, and the Peripheral Nervous system (PNS) comprising the nerves, ganglia, plexuses that emerge from the cranium and spinal cord.

The first step giving rise to the Nervous system, is the development of the primitive streak [1] at Embryonic Day 6.5 (E6.5). This occurs when the cells start accumulating between the epiblast and the visceral endoderm, giving rise to the mesoderm/ primitive streak. This is the beginning of gastrulation.

At E8.0, there is continued deepening of the neural groove, and the neural fold has delineated into the outer surface ectoderm and inner neuroectoderm [2,3]. This is also the time when the neural folds start to first fuse and close into the neural tube at the somites 4 and 5, extending rostrally and caudally. This is also the time when the embryo shifts, turning its dorsal side from facing the inside (concave) to facing outwards (convex).

E9.0, is when the caudal neuropore (an opening that exists between 2 closure points extending towards each other along the neural tube) closes up completely. The prosencephalon, mesencephalon, and rhombencephalon are now detectable, and there is a clear distinction between the neuroectoderm (future CNS) and the ectoderm (future skin) [3]. The volume of the prosencephalon increases drastically, and the telencephalic vesicles begin as two buds.

By E10, the telencephalic walls are thickening and differentiating. The neuroectoderm has also started to evolve into the inner ventricular zone (VZ), an intermediate zone (mantle layer), and an outer marginal zone. The mantle layer forms the gray matter and the marginal layer, the white matter. The neuroepithelial progenitor cells of the VZ will give rise to the greater part of the neural cell populations in the CNS and the ependymal cells of the ventricular system [2]. Radial glia are also produced in the first division of neural stem cells in the VZ, extending their processes across the cortex. The ganglionic eminences (medial, lateral, and caudal) emerge from the neuroepithelium over the next few days, with the medial eminence

providing the GABAergic (g-aminobutyric acid-containing) interneurons and oligodendrocytes; while in the spinal cord the motor and sensory neuron differentiation has been initiated [2,3].

E12 to E15 is marked by radical processes. The choroid plexus is visible, and olfactory nerves are detectable as they move to the olfactory cortex, and subsequently the olfactory lobes will become more distinct. The primordia of the pons and cerebellum in the hindbrain have started to enlarge and demonstrate early stages of differentiation, with the myelencephalon forming into the medulla oblongata [2]. Famously, the cortical layers of the cerebrum also start to stratify, resulting from migration of neuroblasts from the ependymal layer radially into the overlying marginal zone, in an inside-out fashion. The excitatory glutaminergic cortical projection neurons arise from this process, while simultaneously neuroblasts from the medial eminence is moving through the cerebral cortex to form the GABAergic inhibitory interneurons. In the spinal cord, the sensory neuron columns present dorsally are more pronounced compared to ventral motor neuron columns [2].

Between E16 to E18, the cerebral cortex has become well-structured forming all 6 layers - the marginal zone, cortical plate, cortical subplate, intermediate zone, subventricular zone (SVZ), and VZ- and this differentiation continues up to a few weeks after birth before stabilizing. At this point, axonal projections from the spinal cord make their way through the spinal nerves to the autonomic- specific branches [2,3].

Signaling during nervous system development

During neural development, the embryo receives a multitude of signals from various sources, and the combination of these cues in a spatially and temporally distinct manner, is what drives the formation of distinctive nervous system regions.

The sources of these signals can be broadly classified into intrinsic regulation and extrinsic signaling-including diffusible molecules, cell to cell interactions, cell to extracellular matrix interactions, blood vessels and cerebrospinal fluid.

Intrinsic regulation

These signals arise from within the cell, such as the innate knowledge of progenitors to first pass through neurogenesis before commencing with gliogenesis and subsequent specialized neuronal cells.

- Transcription factors such as basic-helix-loop-helix (bHLH), SOX family, estrogen receptors, peroxisome proliferator activated receptor γ (PPARγ) and nuclear receptor co-repressor (N-CoR), can mediate such activities.
- Epigenetic regulation at the level of histones and DNA modifications can affect chromatin accessibility, and therefore gene expression. Histone methyltransferases, histone acetyltransferases (HATs) and deacetylases (HDACs) work to modulate the genetic material, influencing the activation and repression of genes [4].
- Small RNAs, namely, miRNA, siRNA, can also influence gene expression and cell specialization.

Extrinsic signaling

Diffusible signals or extrinsic factors are small molecules that are produced at specific regions within the embryo and their concentrations dilute as they move further away from the source, thereby forming a gradient. This gradient is what helps achieve the unique cell signatures at each segment of the embryo. The main types of extrinsic signals observed are:

Growth factors

These include the Bone morphogenetic proteins (BMPs), that belong to the Transforming growth factor beta (TGF-β) family. BMP2 and BMP4 in particular, play important roles in CNS development, namely, their inhibition by noggin in the center of ectoderm germ layer supports the initial neural plate formation that proceeds to become the neural tube [5]. This gradient is essential for the formation of the dorso-ventral patterning in the embryo. Additionally, fibroblast growth factors (FGFs), and insulin growth factors (IGFs) are also involved at the level of SMAD phosphorylation, negatively regulating this pathway to further inhibit BMP downstream events.

Wnt signaling pathway

The Wnt glycoprotein binds Frazzled receptor on the plasma membrane, along with coreceptors (canonical pathway) such as lipoprotein receptor-related protein (LRP)-5/6, and receptor tyrosine kinase (RTK), recruits Disheveled, which inactivates destruction complex comprising of adenomatosis polyposis coli (APC), protein phosphatase 2A (PP2A), glycogen synthase kinase 3 (GSK3) and casein kinase 1α (CK1α) [6]. This allows for the accumulation of β-catenin, which in turn, localizes to the nucleus along with T-cell factor/lymphoid enhancing factor (TCF/LEF), activating Wnt-responsive genes. This has an effect on the regulation of the Neural stem cell/precursor cell cycle, maintaining progenitor populations in the Ventricular zones [5].

Morphogens

Notably, Sonic hedgehog (Shh), that binds Patched, a transmembrane receptor protein, reverses the inhibition of Smoothened, a G-protein coupled receptor. As a result, downstream signaling events are regulated, exhibiting an effect on transcriptional mediators such as Gli proteins.

Retinoic acid (RA) is another morphogen derived from Vitamin A after it has been converted from its alcohol form (retinol) to retinaldehyde via alcohol dehydrogenases (ADH) and retinol dehydrogenases (RDH), and then to retinoic acid via retinaldehyde dehydrogenases (RALDH) [7]. RA has a role to play in a slew of nervous tissue developmental processes such as hindbrain anteroposterior patterning, dorso-ventral patterning of the spinal cord, spinal cord motor neuron differentiation, formation of early somites, optic cup formation and much more [7].

Retinoic acid

Retinoic acid is a metabolic product of Vitamin A, not inherently produced by the mammalian body, and thus ingested from external sources in the form of carotenoids and retinyl esters from plants and animals respectively. Vitamin A was first discovered in 1913 [8], and numerous studies testified to the necessity of this factor in early development, particularly in brain and optical development [9,10]. The identification of the chemical structure can be attributed to Paul Karrer (1937, Nobel Prize in Chemistry) while the discovery of all-trans-retinal as a crucial component in rhodopsin and thus subsequently necessary for vision is credited to

George Wald (1967, Nobel Prize in Physiology and Medicine) [11]. Vitamin A is taken up by the body and since it is fat-soluble, gets stored in the liver primarily but can also be found in extrahepatic sites such as the lungs, bone marrow, and kidneys [12].

Retinoic acid uptake and metabolism

RBP4 (retinol-binding protein 4) is a protein found in the plasma capable of binding retinol and transporting it through the bloodstream to target cells containing the signaling receptor and transporter of retinol (STRA6) receptor [13]. Once retinol enters the cytoplasm, it binds RBP1 (retinol-binding protein 1), to be stored before its transformation into All-trans retinoic acid (AtRA).

Retinol is converted to retinaldehyde via alcohol dehydrogenases (ADH) and retinol dehydrogenases (RDH), and then to retinoic acid via retinaldehyde dehydrogenases (RALDH) [7]. The ADH family of enzymes can be ubiquitously expressed in embryonic and adult tissue (ADH5) or restricted to certain tissues (ADH1, ADH7) [14]. Knockout studies suggest that ADH enzymes are most likely involved in curbing retinol toxicity than participating in RA synthesis [15,16]. RDH enzymes, on the other hand, have a more pronounced effect in optical structures [17]. The RALDH family of enzymes consists of 3 enzymes- RALDH1, RALDH2 and RALDH3. Their distribution is site specific and the rate limiting step in RA synthesis. RALDH2, the earliest one to be expressed, first found in the primitive streak and mesodermal cells, is then localized to the somitic and lateral mesoderm, posterior heart tube and rostral forebrain [18]. It is responsible for RA availability upto ~E8.5, after which RALDH1 and RALDH3 contribute to RA synthesis in the eyes and the olfactory system.

Retinoic acid resembles fatty acids due to the presence of a beta-ionone ring, a polyunsaturated side chain, and a polar end group [19], allowing for it to enter the cell easily through the plasma membrane. Now, in either an autocrine or paracrine fashion, retinoic acid diffuses into the neighboring cells, binds cellular retinoic acid-binding protein II (CRABPII), and moves into the nucleus. This is the point at which retinoic acid can now bind the retinoic acid receptors already bound to the DNA. After this action, retinoic acid moves out of the nucleus and is degraded by the CYP26 class of the cytochrome P450 (CYP450) superfamily of enzymes.

Retinoic acid receptors

Retinoic acid binds Retinoic Acid Receptor/Retinoid X Receptor (RAR/RXR) heterodimers, which are usually bound to retinoic acid response elements (RAREs) or retinoic X response elements (RXREs) in the promoter or enhancer regions of target genes [20]. RAREs usually consist of two direct repeats of either the hexameric sequence (A/G)G(G/T)TCA or the more relaxed motif (A/G)G(G/T)(G/T)(G/T)(G/C)A, separated by 1, 2, or 5 bp (DR1, DR2, and DR5 RAREs, respectively) [21]. This constitutive binding allows for the recruitment of corepressor complexes maintaining target gene repression (Figure 1).

The RAR and RXR receptors each exist as isotypes, RARα/β/γ and RXRα/β/γ, encoded by different genes, and each of them has its own isoforms, issued from alternative splicing and/or alternative promoter usage. While RXR receptors are only able to bind 9-cis retinoic acid (9-cis RA), RAR receptors can bind with high affinity AtRA and 9-cis RA.

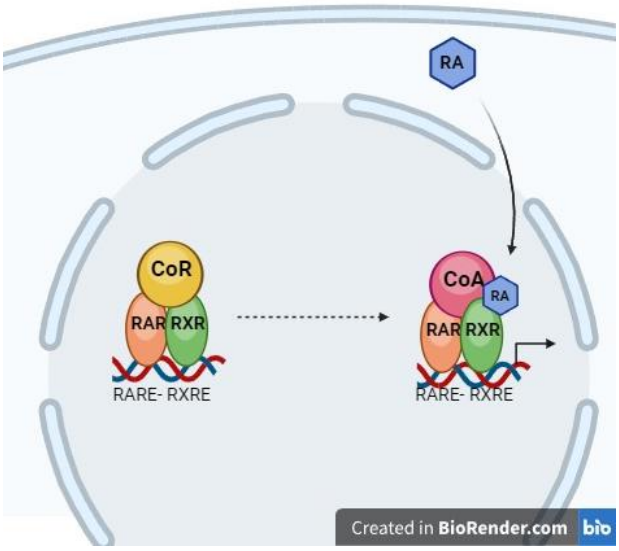


Figure 1. Retinoic acid signaling

RAR-RXR nuclear heterodimers usually exist bound to the DNA at RARE-RXRE elements. They are found along with Corepressor (CoR) complex that inhibits gene activity. When RA enters the nucleus, it binds the RAR-RXR complex, and the Corepressor complex is switched out for the Coactivator (CoA) complex, which allows the DNA to become accessible for target gene expression.

Retinoic acid receptor ligands

Each isotype receptor can be activated with specific ligands, that can be synthetic or naturally occurring.

Natural agonists for RXR receptors are docosahexaenoic acid, lithocholic acid, and methoprene acid[22,23]. Phytanic acid, honokiol are RXRα specific agonists, while danthron, derived from chinese rhubarb, inversely is an RXRα specific antagonist. There is an extensive list of pan-RXR synthetic agonists including, bexarotene, LG100268, AGN194204, LG 101506, HX 630, SR 11237. Those that are RXRα specific are peretinoin, AM-6-36, CD3254 [24].

Conversely, for RAR receptors, pan RAR agonists such as retinoic acid, luffariellolide [25], are widely studied while no known naturally occurring ligands exist that are isotype- specific or selective. In such an instance, several synthetic ligands have been designed like Am 580, Am 80, BMS753, AGN 193835, and AGN 193836, that are RARα specific. BMS641 and 2- thienyl substituted dihydronaphthalene retinoids 45-47 are modified ligands that are RARβ specified, while BMS961 and α-hydroxyacetamide-linked retinoids 54-57 have been described as potent for RARγ activity, and TTNPB as a pan-RAR agonist [23,26].

The capacity of each of the synthetic RAR agonists to regulate a specific transcriptional response has been described early on by Pierre Chambon's team [27–29], and more recently, these specific responses have been disentangled at the level of the various controlled gene programs [30–32].

The role of retinoic acid and retinoic acid receptors in embryogenesis

The distribution of the RARs in embryos differs from region to region. Numerous studies looked into the expression patterns of the RAR and RXR receptors, especially in early mouse embryogenesis [33–35], while a fair few also made comparisons with zebrafish [36–38] and xenopus models [39–42]. In these studies, the presence of these receptors was systematically addressed by taking into consideration the developmental stages described from a chronological point of view (e.g., pregastrulation, gastrulation, neurula, and organogenesis) [43].

Notably, while pregastrulation in mice and rat models does not reveal detectable levels of RAR/RXR gene expression, at gastrulation (E7.5), RARα and RARγ expression is ubiquitous

and diffuse, and RARβ expression was found mainly in the lateral regions of the embryo [35]. Similarly, RXR expression during pregastrulation is also not defined, with RAR and RXR expression often overlapping and omnipresent throughout the embryo.

In addition to the presence of RAR/RXR receptors, the driving differentiation force for site-specific cell specialization leading to development structures depends on the bioavailability of RA. Retinoic acid is first detected at the primitive streak phase, i.e., E7.5 of embryonic development in mice, confirmed by the expression of class IV ADH mRNA seen initially in the posterior region along the primitive streak [14,18]. It is then found to be localized close to the trunk, hindbrain, and optic vesicle by E8.5–E9.5 [14,44].

Postgastrulation is characterized by the regional patterning of the neurectoderm, the formation of the somites, the migration of cranial neural crest cells, and more. During this span of time, RAR expression becomes more distinctive. RARα is localized to the neuroectoderm, primarily at the rostral area. RARβ is specific to the neural tube, and the rostral regions of the mesodermal tissue. RARγ, on the other hand, occupies the opposite end, being found mostly in the regressing primitive streak, and in the caudal-most regions of the somite-axis [35], and almost completely absent from mesodermal tissue (Figure 2).

A schematic representation of the developing stages of mouse embryo demonstrating expression of RAR receptors can be seen in Figure 2. At E7.5, there is no detectable or very low signaling of RAR receptors.

By E8.5 to E13.5, there is an increased presence of the RARα (red line shading, Figure 2) in the neuroectodermal region, and the caudal most regions of the Caudal Neuropore (CNP), as well as certain regions of the hindbrain up to rhombomere 6-7. RARβ (green line shading) is highly expressed in early stage midbrain at E8.5, specific to the neural tube, and the rostral regions of the mesodermal tissue (green line shading; Figure 2) and then later on in the spinal cord [38–40]. RARγ, on the other hand, occupies the opposite end, being found mostly in the regressing primitive streak, and in the caudal-most regions of the somite-axis [28], and almost completely absent from mesodermal tissue , as illustrated in the schematic representation provided in Figure 2 (blue line shading).

By E13.5, RARα is majorly expressed in the corpus callosum, corpus striatum. At no point is RARγ (blue line shading) detected in the developing CNS. Its expression seems to be strongly

repressed in this tissue layer only appearing at the regions of frontonasal process (FNP), limb bud regions (HL, FL), and prevertebrae (Pv) from E9.5 onwards [40] (Figure 2).

Retinoic acid functionality also works in synchrony with complementary signals such as fibroblast growth factor (FGF-8), sonic hedgehog (SHH), and bone morphogenetic proteins (BMPs). Retinoic acid plays a role primarily in the growth and differentiation of the posterior structures, while in the anterior part of the embryo, its activity is regulated by the RA-degrading CYP450 enzymes CYP26A1 and CYP26C1, promoting the anterior-posterior patterning [45–47]. Retinoic acid patterns the anteroposterior axis by being synthesized in the posterior mesoderm. The presence of CYP26C1 in the anterior mesoderm prevents the influence of retinoic acid in these regions, leading to the emergence of the hindbrain from the neural plate. In the dorsoventral axis of the developing neural tube, retinoic acid is produced at the somite sites, along with SHH and BMP expressed ventrally and dorsally, respectively. FGF-8 expression is detectable at the posterior end of the extending neural tube, and these factor gradients predict the ultimate fate of specialized neurons that emerge such as interneurons, sensory neurons, and motor neurons [48–50].

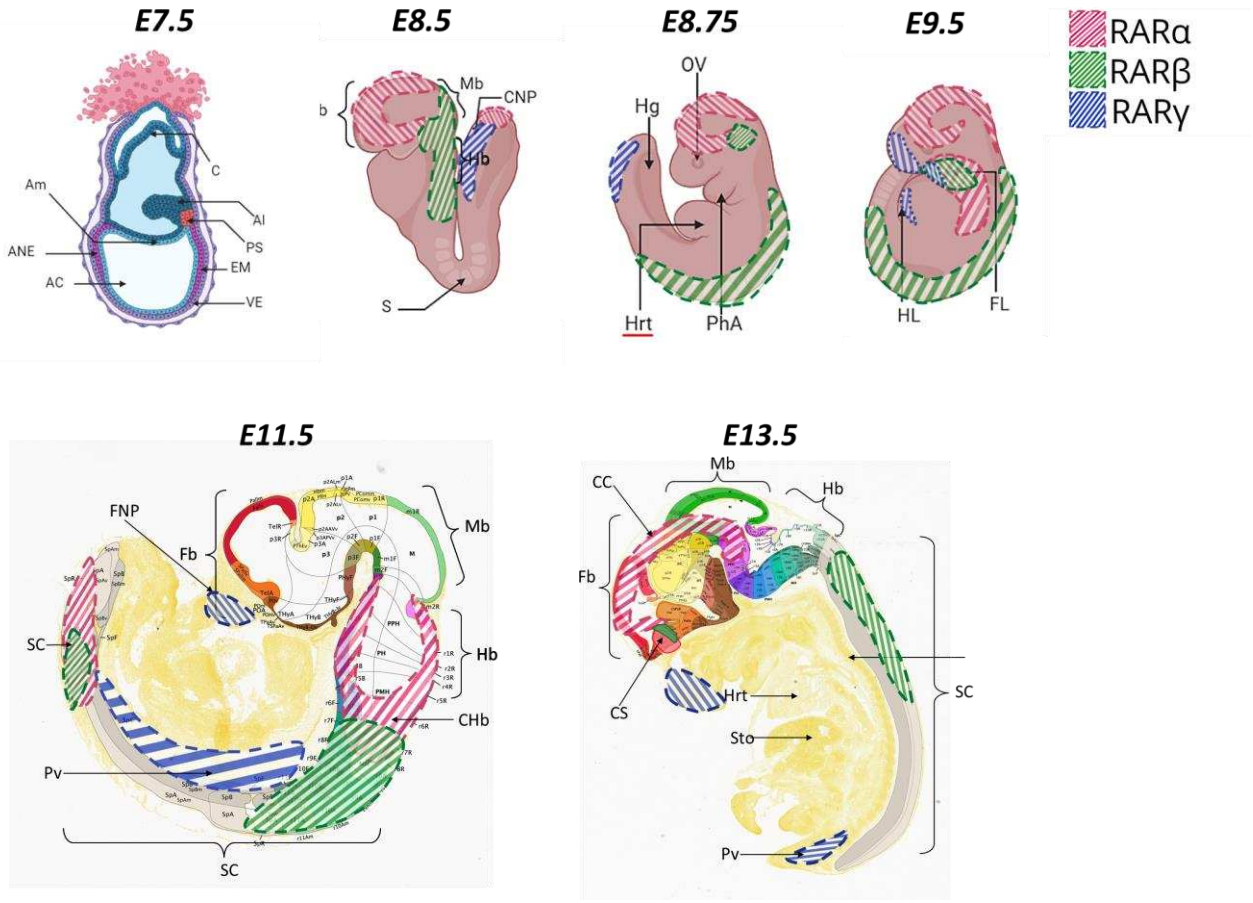


Figure 2. RAR isotype expression in mouse embryogenesis

Schematic representation of the expression of RARα (red line shading), RARβ (green line shading), RARγ (blue line shading) at different stages of mouse embryogenesis from E7.5 to E13.5, particularly in relation to CNS development. At E7.5, there is very low to undetectable levels of RAR expression observed. RARα is first seen at E8.5 lowly expressed in the neuroectoderm, near to the Caudal neuropore (CNP). By E11.5 it is more present in the posterior region of the spinal cord (SC), and in the Caudal Hindbrain (CHb) before being expressed primarily in the Corpus Striatum (CS) and Corpus Callosum (CC) at E13.5. RARβ (green line shading) is present mostly in the neural tube between E8.5-E9.5, with a more localized presence at the caudal region of the spinal cord (SC) in E11.5 onwards, while RARγ (blue line shading) is first seen in the regressing Primitive streak (PS; at E8.5), it is undetectable later on in the brain and spinal cord (SC) neuroepithelium. Its expression is limited to the Frontonasal process (FNP) and Branchial arches (limb region; HL, FL). Top panel created on Biorender; Bottom panel (E11.5, E13.5) images taken from Allen Developing Mouse Brain Reference Atlases (https://atlas.brain-map.org/) and overlayed with red/ green/ blue line shading to indicate RARα/β/γ expression respectively. References made from studies [8,37–40,87]

AC: Amniotic cavity, Al: Allantois, Am: Amnion, ANE: Anterior neuroectoderm, C: Chorion, CC: Corpus Callosum, CHb: Caudal Hindbrain, CNP: Caudal neuropore, CS: Corpus Striatum, EM: Embryonic mesoderm, Fb: Forebrain, FL: Forelimb, FNP: Frontonasal process, Hb: Hindbrain, Hf: Headfold, Hg: Hindgut, HL: Hindlimb, Hrt: Heart, Mb: Midbrain, OV: Optic vesicle, PhA: Pharyngeal Arch, PS: Primitive streak, Pv: Prevertebrae, Ri: Ribs, S: Somites, Sto: Stomach, VE: Visceral endoderm.

Neuronal cell specialization studies in animal models

Beyond its major role in patterning, retinoic acid also plays a role in neuronal differentiation. In addition to its direct action in the regional patterning of the neurectoderm, there have been several studies that have discussed the importance of retinoic acid for the generation of motor neurons and the ventral progenitors [48,51,52], subsets of GABAergic or dopaminergic neurons [53–55], as well as terminal neuronal differentiation in ventricular and subventricular zones [56].

Among these studies, the role of retinoic acid in the emergence of motor neurons during the development of the CNS in chick embryos was already studied at the end of the 1990s by Shanthini Sockanathan and colleagues [51].

Specifically, during CNS development, some proliferating stem cells either remain undifferentiated or they can leave the cell cycle to become progenitors and eventually terminally differentiated cell types like neurons or glia, in an inside-out fashion [57]. This means that the earlier born neurons are found inside while the late-born neurons feature more towards the apical side. In a similar way, medial lateral motor column (LMC) neurons, that project to ventral limb muscles, leave the cell cycle first and form the first layers, followed by lateral LMC neurons, projecting to dorsal limb muscles, that move past the medial LMC neurons to their endpoint. Evidence proved the effect of retinoids synthesized by neurons that influence the differentiation capacity of the lateral LMC neurons, but also the quantity, subtype identity, and timing of maturation at the level of the limbs. The longitudinal limb development, shown to be defined by a proximal source of retinoic acid and a distal source of FGFs, is concomitant with the motor neuron innervation process issued from the LMC, driven by the patterned expression of nearly two dozen individual Hox genes and the conserved cofactor forkhead box P1, which together control retinoic acid synthesis in limb-innervating motor neurons [58].

The combined and harmonious signaling of retinoic acid, but also FGF-8 and SHH, is vital in inducing the motor neuron identity as well as patterning of the

subsequent ventral neural structures. For example, oligodendrocyte transcription factor 2 (Olig2) expression—driven by retinoic acid and SHH action—marks the identity of motor neuron progenitors and is tightly linked to the expression of homeobox domain (HD) regulators. The expression of Olig2 is negatively regulated by the expression of NK2 homeobox 2 (Nkx2.2) (driven by SHH), while Olig2 negatively regulates the expression of the HD factor Pax6 [59].

Additionally, retinoic acid has been discovered to have a dose-dependent effect on the identities of differentiating p3 V3 neurons versus serotonergic neurons [55,60]. The p3 progenitors are the most ventral progenitor domain present in the spinal cord and hindbrain, both being equivalent populations. They give rise to the glutamatergic V3 interneurons in the spinal cord and serotonergic (5-HT) neurons in the hindbrain. Conclusively, these studies were able to show that higher levels of retinoic acid are present in the p3 [V3] domain, and via the retinoid acid receptor signaling, one of the downstream targets, Notch, has a direct role to play in the activity of achaete-scute homolog 1 (Ascl1). Lower levels of Ascl1 confirm a glutamatergic V3 interneuron fate from p3 progenitors, and the opposite is true for serotonergic fate commitment in the hindbrain.

In the zebrafish embryo, particularly in the posterior hindbrain, the medulla and the area postrema, noradrenergic (NA) neuronal differentiation is initiated by the action of retinoic acid [53]. Similarly, FGF-8 is the signal inducing noradrenergic specification in the locus coeruleus. In both cases, precursor neurons require the expression of transcription factor AP-2α, in order to terminally mature into NA neurons.

Retinoic acid is also responsible for GABAergic differentiation in mouse embryos at E14.5 in the subventricular zone of the basal ganglia, due to the expression of RALDH3 [54]. This is also seen in the lateral ganglionic eminence (LGE) where the same pathway leads to endogenous retinoic acid, required for GABAergic neuronal differentiation. By E18.5, the role of retinoic acid is to maintain GABAergic

differentiation in the LGE, primarily through the stimulation of 67 kDa glutamic acid decarboxylase (GAD-67) activity. But it is not yet clear which RAR/RXR receptors might be linked to this activity, since Gad67 does not show evidence of a canonical RARE element in its promoter region.

This being said, RARβ seems to be necessary for the development of striatonigral projection neurons. Indeed, studies in RARβ-ablated mouse embryos showed reduced levels of 65 kDa glutamic acid decarboxylase (GAD-65), dopamine D1 receptor, and substance P found in these neurons, at E16.5 [56]. At E18.5, there is a partial recovery of Gad67 expression, suggesting neurogenesis of some GABAergic neurons. This is also confirmed by the lower levels of Ascl1, a marker for neural progenitors and a GABAergic neuron determinant [61], concurrently with increased homeobox Meis gene expression, a marker for postmitotic neurons, seen in the ventricular and subventricular zones. This marked reduction in proliferation and substitution with premature maturation can be explained by the activity of FGF-3, a direct target of retinoid receptors [62], which, at E13.5, is the main FGF functioning in these zones. This interplay between retinoic acid and FGF-3 holds the key in balancing the populations and chronological events from neural progenitors and their proliferation to terminal neuronal differentiation.

The role of retinoic acid in *in vitro* cell specialization

There has been a conscientious effort to characterize the effect of RA independently and in combination with other growth factors and morphogens, by mimicking *in vivo* embryogenesis, *in vitro*.

Retinoic acid released from the somites and the Shh gradient produced by the floor plate and notochord [63–65], provide a rostral-caudal and ventral-dorsal gradient respectively, influencing the emergence of the ventral progenitor interneuron domains (p0-p3) and a progenitor motor neuron domain (pMN) arranged in the ventral-dorsal axis; which ultimately mature into the ventral interneuron classes and motor neurons [63,66,67].

V2a interneurons, found in the spinal cord and the respiratory centers of the hindbrain, are synthesized from mouse embryonic stem cells that get treated with low concentrations of Retinoic acid and high concentrations of Shh agonist (Pur) [68]. Additionally, in order to specifically choose the commitment of these cells for V2a interneuron production over V2b interneurons, Notch1 signaling is inhibited. This also has an auxillary benefit in preventing glial cell proliferation. A more dorsal phenotype of neural differnetiation can be achieved by switching to higher concentrations of RA and their positional identity gets determined within 48h of exposure [65].

Motor neuron generation is a frequently studied topic and almost always involves the introduction of the morphogens RA and Shh; with RA inducing a caudalizing effect (especially noticeable in embryoid bodies of neural progenitor cells), and Shh creating a motor neuron progenitor specialization [69].

Conversely, generation of cerebellar neurons requires a host of signaling molecules at particular timepoints, viz. the cerebellar organizers FGF-8 and RA, and subsequent treatment with dorsalizing molecules- Wnt, BMP6/7, GDF7, followed by Shh and JAG1 to induce proliferation of granule cell progenitors (GCPs) along with medium cultured by cerebellar glial cells [70].

Some studies looked at the least invasive methods of generation neural cell types. For example, Gaspard et al, 2009 utilized only cyclopamine, a Shh inhibitor, on neural progenitors with a forebrain identity, to generate cortical-like progenitors that develop into glutamatergic neurons with a pyramidal morphology [71]. Here, there is absolutely no interference from

external molecules like RA, FGF-8 etc, relying only on cyclopamine to change positional identity.

These same studies extended into the use of human pluripotent stem cells (hPSC) as a research model, paving the way for more clinically relevant and transferable knowledge on neuronal cell fate.

In the case of motor neuron generation, RA and Shh treatment are deemed necessary for Sox1+ neuroectodermal cells derived from human embryonic stem cells (hESC), preferably at low concentrations, with retinoic acid treatment occurring early on [72]. This is to avoid regional specificity that the cells acquire once they become Pax6+/Sox1+. Their effects can be seen in monolayer cultures at 2 weeks with extensive neurite outgrowth and the differentiated neurons having a positive profile for Islet1/2, 50% of those HB9+. These double positive marker cells confirm the generation of spinal neurons in the cell culture. As mentioned before, RA here is acting as the caudalizing factor while Shh is the ventralizing inducer. The role of RA is seen in its effect on the upregulation of Shh as well as Class II HD profile proteins (Olig2, Nkx2.2, Nkx6.1), essential for motor neuron specification. This protocol received an update in 2013, with the addition of SAG (Smoothened agonist) [73]. It generates a higher yield of motor neurons in a shorter time frame, of which there is a varying expression pattern of Hb9, ISL1, or double positive expression, mimicking *in vivo* distribution patterns.

Neuronal differentiation using Retinoic acid continues to be tested to determine the best possible conditions for generating neuronal cell types. Longer exposure times to RA, culturing cells as embryoid bodies especially during their exposure to Retinoic acid, culturing cells at higher cell densities, are factors shown to increase neuronal yield [74]. Higher cell densities promote cell crosstalk and have an effect on the upregulation of neurogenesis factors like Sox2, Neurod1, Pax6, ultimately influencing Tuj1+ neuronal formation.

Many of these studies provide protocols to generate a subset of neural identities from a starting embryonic stem cell culture with varying percentages, yield, heterogeneity and are heavily dependent on the concentrations of morphogens, growth factors and culture conditions.

The role of retinoic acid receptors in *in vitro* cell specialization

A number of studies have looked into the effects of molecules such as dimethyl sulfoxide (DMSO), and Retinoic acid, in their abilities to induce any of the 3 germ layers [75]. DMSO treatment on P19 aggregates, pushes the cells to differentiate into cardiac and muscle -like tissue. DMSO exposure at 0.5 - 1% has been shown to cause the outer cell layers of the P19 embryoid to mature into cardiac, skeletal, and epithelial cells, mostly expressing alpha-SMA [75–77]. Within 6 days, striated cardiac tissue is evident, exhibiting contractile movements. By 9-10 days of differentiation, skeletal tissue appears. Additionally, these developments are seen only in embryoid bodies and not when the cells have been seeded as a monolayer culture prior to DMSO treatment. What DMSO is to cardiac tissue, Retinoic acid is to Neurons and treatment of Retinoic acid on P19 aggregates at concentrations of 1μM is the ideal method of generating neurons within a 10-day span [78].

The link between RAR/RXRs and cell specialization has been explored previously in papers looking at the differentiation capacity of RA in P19 and F9 cell cultures [28,30,31,79]. In this chapter, we will explore the current state of research investigating the role of RARs and RXRs in stem cell differentiation.

Note: P19 and F9 embryonic carcinoma cell line

Modeling neuronal differentiation *in vitro*, driven by RA action, has been made possible with the help of the P19 embryonic carcinoma cells. This cell line was first derived by McBurney and Rogers in 1982, from the transplantation of a 7.5-day old embryo onto the testis of a mouse. The resulting teratocarcinoma could be cultured *in vitro*, growing rapidly without the need for irradiated mouse feeder cells [74]. They don't form tumors unless injected into neonates and provided multiple advantages for their use in studying cell differentiation early on. Their advantages lie in their ease of culture, multipotency, anchorage-independence and lack of contact inhibition [75]. This was beneficial to study the influence of externally introduced chemicals on a simplified *in vitro* neuronal development model.

The F9 cell line was also isolated in 1973 [76], by implanting an E6 mouse embryo in the testis of a mouse. It has been described as a nullipotent cell line due to its inability to differentiate spontaneously, and has otherwise been known to differentiate into the endodermal lineage by introduction of various compounds [77,78].

Understanding the role of RAR/RXR in cell specialization studies relied on the use of synthetic agonists, especially since RAR has no known naturally occurring agonists that are subtype specific. In the 90s, Taneja, Roy, and Chambon explored the functionalities of the RAR/RXR activity in embryonic carcinoma cells [27,28].

Specifically, they first demonstrate the activity potential of the receptors to their ligands. At very low concentrations, the synthetic agonists for each RAR, are not able to induce efficient gene expression. Similarly, RXR agonists, have no effect of their own, when introduced individually into the culture. However, the combination of any RAR synthetic agonist with a pan-RXR agonist (ex. BMS649) results in a synergistic effect with induction of RA-target genes [28]. This pointed to the strong capacity of the heterodimer nuclear receptor pair to work in sync. Additionally, the presence of RAR-specific ligands at appropriate concentrations was sufficient to induce differentiation in EC cells by activating the RAR/RXR heterodimer.

Furthermore, they discuss the functional redundancies of the RAR receptor subtypes in inducing certain RA-responsive genes in P19 and F9 EC cell differentiation. This redundancy is

cell type and promoter context-dependent. For example, RARγ is able to induce differentiation of F9 cells, while P19 cells can be induced to do so by either RARα or RARγ [28]. A particular RAR may have the capacity to induce a specific gene response but this is also dependent on the activity of the other receptor subtypes and its 'dominance' when all 3 receptors are active. This is visible in knockout studies where either of the receptors are able to take over the gene expression profiles to a certain extent [27,80]. And as mentioned before, this can vary from cell line, to promoter profile.

The role of RARγ in F9 differentiation was studied [30]. Compared to induction with ATRA in F9 cells, BMS961 (RARγ agonist) is able to recover 62% of the RXRα– RARγ target genes, in contrast to BMS753 (RARα agonist) and BMS641 (RARβ agonist) inducing 40 and 10% of the network respectively. These targets are either found to be de novo targets or previously occupied by different RXR-RAR heterodimers that get switched out in a sequential fashion according to ligand exposure. This also confirms the redundancy and specificity potential of each receptor subtype in cell differentiation.

In fact, RARγ KO cells no longer have the capacity to differentiate in response to RA, as demonstrated by the failed induction of CDXL, GAP43, STRA4, and STRA6 [80,81]. This is supported from mouse RARγ null mutants that develop abnormally with high mortality, inhibited growth and sterility [82]. However, partial recovery of differentiation potential and target gene activation can be seen when there is a re-expression of RARγ or overexpression of RARα [80]. In contrast, RARβ overexpression only weakly restores differentiation, visualized morphologically and from expression levels of differentiation markers, laminin B1 and collagen type IV. It should be noted that while RARα can mediate RARγ activity during RA mediated gene response, the levels of recovery can be influenced primarily due to the fact that there is an abundance of RARγ in F9 cells usually as compared to the other subtypes, and this recovery might be more indicative of a quantitative rather than qualitative virtue.

Nevertheless, multiple studies are able to confirm the functional redundancy, as well as cell fate specific influence of these receptors. Specifically, in F9 cells, ATRA and BMS961 are capable of inducing a cell differentiation towards the endodermal lineage, while only ATRA and BMS753 produce a neuronal cell fate in P19 EC cells [31]. While each receptor agonist might induce a different cell fate transition in different EC lines, they both also activate a

common core of programs that may have different temporal expression patterns, constituting more than 60% of the RA- induced differentially expressed genes.

Each cell fate transition pathway revealed commitment specific and common program specific TFs. Particularly in the neuronal cell fate commitment, apart from ASCL1, NR2F2, NR4A2 as known neuronal fate regulators activating more than 60% of the gene regulatory network, novel targets (GBX2, TAL2,, DMRT1, LHX2) were discovered to be responsible for >50% of network regulation [31].

Interestingly, in a proof of concept assay, using CRISPR/dCas9 activation of these TFs revealed that BMS961 was necessary to first induce an initial activation of the common program, that allows the transactivation of P19 specific neurogenesis Master regulators (GBX2, TAL2, DMRT1, LHX2) [31].

RA- induced differentiation in mouse embryonic stem cells, reconstitutes a combination response of what is observed in P19 and F9 differentiation. The differentiation is not homogenous and indicates the pluripotent and noncommitment of ES cells as compared to EC cells. 65 and 75% of the genes downregulated and upregulated respectively in P19 cells are recovered in ES differentiation and this is ~30% for the upregulated genes in F9 cells [31].

This provides a useful tool in the determination of a RAR subtype specific response to be studied in pluripotent non committed cells.

In 2017, Podleśny-Drabiniok and team highlights not just the role of RA in driving neurogenesis but also the additional layer of regulation by the retinoid receptors that infleunce cell fate [79]. Using individual synthetic RAR agonists, allowed to detect the neural cell lidentity achieved in P19 embryoid bodies over a period of 10 days. On exposure to ATRA, 88.5% of the cell population expressed Tubb3 (pan-neuronal marker) and 90% of these cells had an inhibitory phenotype (Gad65/67). The neurons also had the presence of marker Drd2, indicative of striatopallidal medium spiny neurons (MSNs). The authors also used synthetic agonists targeting specific RAR isotypes.

Out of all the RAR agonists, CD666 (RARγ agonist) treated cells were most successful in generating GABAergic neurons ie. 77%, and BMS641 (RARβ agonist), the least efficient ie. 28%. Interestingly, there was also a subpopulation of dopaminergic neurons confirmed by the

presence of dopamine transporter (DAT) and absence of noradrenaline transporter (NET), and all of them were GABAergic.

The activation of RARα and RARβ, gave rise to mainly Th+ dopaminergic neurons. And this was confirmed to be most likely due to the effect of RARα rather than RARβ; alternately, RARγ activation was solely responsible for inducing the production of Drd2+ striatopallidal medium spiny neurons. The consequence can be seen in markers expressed at 24h of exposure to the individual ligands or ATRA. In all cases, Meis1 and Meis2 expression was pronounced, indicating neurogenesis, while only ATRA and CD666, were able to cause an increase in Gsx2 and Ascl1, both determinants of GABAergic phenotype cells, primarily found in the Lateral Ganglionic Eminence (LGE). RARα and RARβ activativation on the other hand only mildly induced expression of these markers, instead favoring the induction of FGF-8 and En1, indicative of Dopaminergic fate. These programs were already set within 24h of exposure to the ligands, stipulating that commitment is made early on in differentiation.

Conclusion

While extensive work has been achieved in studying EC differentiation in relation to RAR/RXR isotype activation, there is a missing link in our knowledge about these receptors in ES cell differentiation. The link between specific RAR receptors and their ultimate cell fate influence still remains to be decoded as illustrated by Podleśny-Drabiniok and colleagues [79].

Furthermore, preliminary studies often rely on *in vitro* monolayer cultures owing to their simplicity and ease of deducing cause and effect. However, we are now approaching more complex and integrated model systems to study tissue development and disease. For example, Retinoic acid is very clearly demonstrated to be a crucial element for brain organoid development [83,84]. At the early stages of development, it directs cells to commit to the neuroectodermal pathway, whether it be for unpatterned whole brain organoid generation or specifically for cerebral, or dorsally patterned brain organoids. The current state of research is moving quickly to include single cell analysis [85–87], spatial transcriptomics and lineage tracing [88–90], and we are already witness to the possibility of growing embryo-like structures from stem cells (stembryos) [91,92]. Thus, the need to study the influence of each RAR in brain development and neural cell fate remains quite strong, possibly providing a means for targeted therapy in the coming future.

Dysregulations in neurogenesis

In the previous chapter, we focus on normal processes of brain development and studying cell fate driven by retinoid acid receptors. But equally interesting is the study of developmental processes that get modified. And this can be due to degenerative diseases or the occurrence of cancer.

In relation to brain development, we were interested in studying the characteristics of Glioblastoma, more specifically pediatric glioblastoma. Currently, there is very limited means of treatment for such cancer patients. Moreover, current research models rely on patient-derived samples and human iPSC lines carrying the Glioblastoma specific mutations to study this disease.

In the next chapter, I will discuss the current state of research in Glioblastoma modeling and propose the usage of a mouse brain organoid model to study mutation specific tumors.

Glioblastoma

Glioblastoma multiforme (GBM)

Gliomas are brain tumors originating from cancerous glial cells. There are four grades of WHO classified gliomas depending on the malignancy. Glioblastoma multiforme (GBM), denoted Grade IV, is considered the most aggressive form and is associated with poor prognosis, lower survival rates, and higher recurrence rates post-treatment [93].

Up until 2016, tumors arising from the Central Nervous system (CNS) were classified solely based on histological features, which included morphological characteristics such as microvascular proliferation, cell proliferation activity, and necrosis [94]. This was appreciated as the 'gold standard' of diagnosis, since it was accessible worldwide, even in countries with limited resources, easily performed, and didn't require costly laboratory equipment. However, with the growing research on molecular typing and its widespread applications, the diagnosis was subsequently modified to incorporate molecular markers as well.

Glioblastoma arises from astrocytes, the chief component of glial cells in the CNS. Although presenting with distinctive margins during imaging, they are diffusively infiltrative [95]. They appear as dusky red or yellowish pink tumors that are fragile and highly vascularized, usually characterized by thrombosed vessels. Histologically, four hallmark components observed are nuclear atypia, vascular endothelial cell proliferation, hypercellularity, and pseudo-palisading necrosis which are hypercellular zones that commonly surround necrotic centers. Additionally, one can also observe neovascularization and bizarre multinucleated cells.

Pediatric glioblastoma (pGBM)

CNS tumors rank as the second most common tumor in children after leukemia and the most common solid tumors with an incidence of 30 per million [96]. Depending on the definitions of age and WHO defined grades for pediatric glioblastoma (pGBM), incidence can vary from 3-15% [97]. In 2021, the classification of CNS tumors was updated to include molecularly distinct entities and genetic sequencing when providing a diagnosis. The highest incidence is tracked in individuals aged from 15-19 years, most likely due to the cumulative effects of mutations over time. Pediatric glioblastoma usually appears in the supratentorial brain with nearly 50% of these cases being located in the cerebrum. The median survival period

ranges from 13-73 months with a 5-year survival rate of less than 20%. To date, a gold standard for treatment hasn't been established. The most effective and widely followed protocol is surgical tumor resection in combination with oral temozolomide treatment [95].

Distinguishing pediatric glioblastoma from adult glioblastoma

pGBMs have signature variations from adult glioblastoma in terms of the mutations and corresponding prognoses which can significantly alter treatment modalities and effectiveness. The most common CNS tumors in adults are diffuse HGGs whereas only 10% of pediatric brain tumors can be assigned to this classification. Consequently, there are a lot fewer molecular markers that have been established. Glioblastoma cases in children are characterized by the lack of IDH mutations, absence of 1p/19q co-deletion [94], and fewer mutations of EGFR and/or PTEN. On the other hand, there is a high correlation with p53 [95], ATRX, and PDGFRA genetic alterations [98].

Even though molecularly, pGBM and adult GBM differ [99], they are comparable in respect of their hierarchical organization. In both cases, slow-cycling cancer stem cells occupy the apex of the proliferation cascade, giving rise to rapidly proliferating progenitor cells and non-multiplying cells. Due to their near quiescent state, treatment with radiation and mitotic inhibitors can be ineffective; this is particularly evident in pGBM cases, where the mutational burden in recurrences isn't as high as seen in adult cases, offering the possibility that pGBM CSC (cancer stem cell) populations might most likely be more latent [100]. However, other unknown factors might be playing a role in this.

Of interest, is the heritability of pGBM. It has been found that pGBM is a combination of de novo germline variants originating as mutational events in the parent's germ cells or at the level of the zygote and somatic SNVs in genes associated with hereditary syndromes [100]. One frequently occurring germline somatic variation is the subclonal deletions in ATRX. This has been shown to affect euchromatin levels genome-wide, leading to enhanced cell migration.

Types of pediatric glioblastoma

Typically, gliomas occurring in children can be separated into 2 categories: low-grade gliomas that are usually mutant in BRAF genes, affecting MAPK signaling downstream, and high-grade gliomas that usually feature mutations in the H3F3A locus [94], that is known as

variant 3 of Histone 3. In pLGGs (pediatric Low Grade Gliomas), the most commonly detected abnormalities are focused in BRAF (48%), FGFR1 missense (17.6%), NF1 (8.8%), and TP53 (5.6%) [98]. Alternatively, TP53 (49%), H3F3A (37.6%), ATRX (24.2%), NF1 (22.2%), and PDGFRA (21.7%) account for the most frequent mutations in pHGG (pediatric High Grade Gliomas).

Molecular basis of pediatric glioblastoma

Histones are proteins involved not only in protecting and packaging the DNA but are also involved in euchromatin and heterochromatin regulation and consequently transcription levels. The 5 different types of histones are well established, but Histone 3 exists as 8 different variants; Histone 3.3 being described as transcription-essential [101]. There are several sites on this protein that are post-translationally modified, and that in turn leads to downstream effects on differentiation, stem cell state maintenance, and more. Due to their evolutionarily highly conserved profile, diseases involving mutated histones are rare.

H3.3 is a variant that is constitutively expressed and diffusely present. H3.3 production is encoded for by the genes H3F3A and H3F3B, and the variant differs from other H3 variants by several amino acids that regulate chaperone protein affinity and chromatin expression [102].

The two most common mutations linked to the H3F3A locus are on the amino-terminal tail of the histone, H3K27M, and H3G34R/V where the lysine at 27th position is converted to methionine and the glycine at position 34 is converted to arginine/ valine respectively. The roles of these amino acids in regulating cell transition events have been charted out recently albeit not extensively. As reported by several studies, 30% of pGBM patients express the H3.3 mutations, either of K27M or G34R/V.

More frequently discussed is the K27M mutation in part due to its higher frequency. Thus, its role in repressing transcription and chromosome packaging is more well established.

According to Gielen et al, 2013, out of the 129 pediatric glioblastomas that they sequenced, 27.1% showed K27M mutation [99]. They are also exclusively expressed only in pediatric high-grade astrocytoma which makes them useful screening prognosis markers for diagnosis. It is also a more frequently encountered mutation in pGBM versus G34R/V (ratio of 6:1); G34R/V alternatively is also indicative of CNS-PNET tumors (Central Nervous System Primitive Neuroectodermal Tumor) [99]. The Schwartzentruber et al, 2012 study on driver

mutations in H3.3 complex identifies 31% of pGBM tumors positive for H3F3A mutations, most commonly at K27 and G34 to M and R/V respectively [103]. In 100% of the samples expressing G34R/V substitutions, mutations in ATRX and DAXX were also observed [94,103]. These genes encode subunits of a chromatin remodeling complex implicated in H3.3 incorporation at the pericentric heterochromatin and telomeres. The consequent demethylation that occurs at sub-telomeric ends might effectuate alternative lengthening of telomeres.

TP53 mutations occurred in 86% of samples having H3F3A and/or ATRX mutations [103]. Korshunov et al, 2015, showcased more than 50% of the pGBM cases profiled as having H3F3A mutations (43% for K27 and 15% for G34) [104]. Correspondingly, Pathak et al, 2015, revealed that loss of widespread H3K27me3 occurred in K27M and K27M/ATRX double mutants as well, >50% of the pGBM cases analyzed corresponded with H3K4/K9me3 loss, and that H3K27me3/H3K4me3 double-negative patients had a poorer survival outcome compared to H3K27me3 negative/H3K4me3 positive patients [105]. H3K27M mutations are routinely associated with poor clinical outcomes. These cases usually demonstrate oncogenic amplifications in genes such as PDGFRA, MYC or MYCN, CDK4 or CDK6 or cyclin D, ID2, and MET; and mutations in TP53 and PPMD1 [94].

The schematic in Figure 3, outlines molecular mutation subgroups commonly associated with pHGG and DIPG (Diffuse Intrinsic Pontine Glioma), H3F3A specific mutation sites, age at diagnosis, and survival rate specific to a mutation. Out of the 1067 samples profiled, 954 belonged to patients 18 and under at the time of diagnosis.

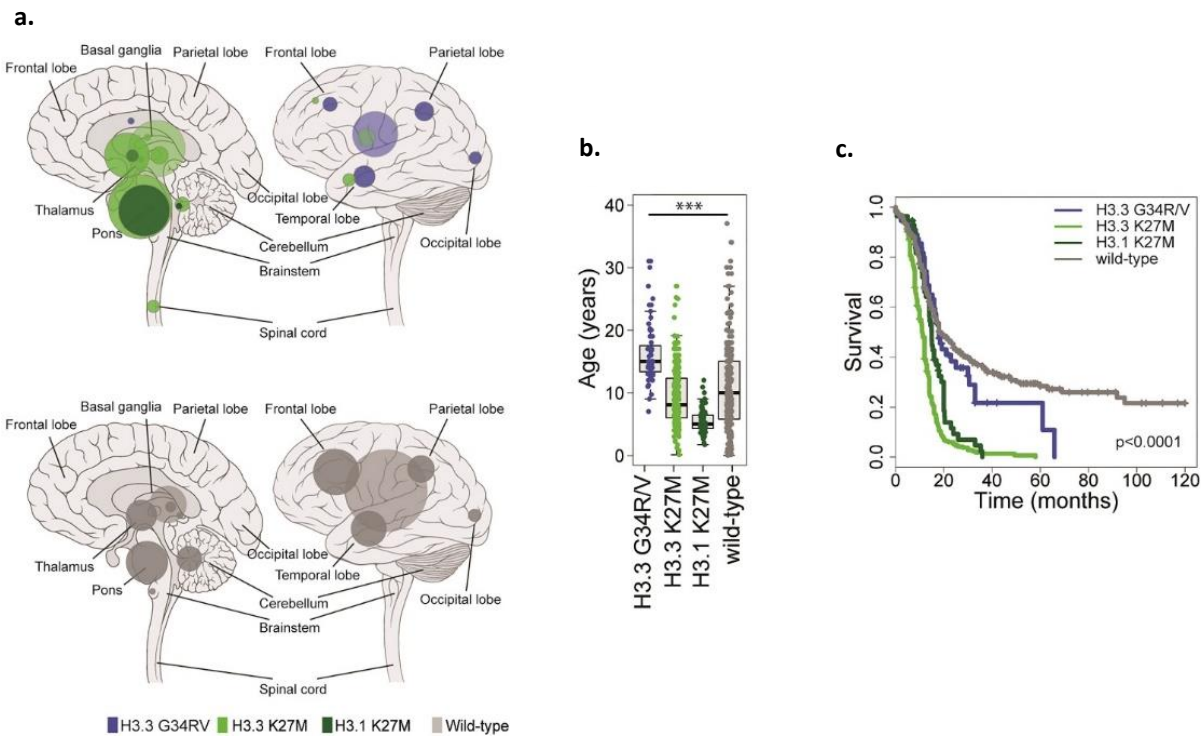


Figure 3. Molecular subgroups of pHGG

Adapted from Mackay A. et al, 2017 [106]

a. Anatomical location of all cases separated by histone mutation (top, n = 441) and histone WT (bottom, n = 314). Left, sagittal section showing internal structures; right, external view highlighting cerebral lobes. Blue, H3.3G34R/V; green, H3.3K27M; dark green, H3.1K27M. The radius of the circle is proportional to the number of cases. Lighter shaded circles represent a non-specific designation of hemispheric, midline, or brainstem.

b. Boxplot showing age at diagnosis of included cases, separated by histone mutation (n = 753). The thick line within the box is the median, the lower and upper limits of the boxes represent the first and third quartiles, and the whiskers 1.5× the interquartile range. ***Adjusted p < 0.0001 for all pairwise comparisons, t-test.

c. Kaplan-Meier plot of overall survival of cases separated by histone mutation, p-value calculated by the log-rank test (n = 693)

H3F3A driver mutations and their epigenetic role in pGBM

H3K27M mutation

H3K27 is normally a target of EZH2, becoming either di- or tri-methylated. The EZH2 is part of a larger complex known as PRC2 (Polycomb Repressive Complex 2). The methylated H3K27, H3K27me3, acts as a transcriptional repressor, inhibiting the association and binding of transcription complexes (Figure 4a). Since most of the H3.3 variants occupy positions of the genome involved in maintaining undifferentiated states, the mutated H3K27M can lead to

oncogenic profiles. It is hypothesized that the K27M mutation either causes a steric hindrance for the binding of the EZH2 or most probably competes for binding with EZH2, thereby sequestering and inactivating the PRC2 complex permanently and causing a global reduction in K27 methylation [107,108]. This results in increased gene expression at H3K27 regulated expression sites and at bivalent promoters (H3K27 and H3K4) (Figure 4b). In reversible K27M mutant studies, the uninhibited PRC2 restores methylation patterns of H3K27 at CpG islands and globally; likewise, in K27M mutant cell lines modified to express modified EZH2, H3K27me2/3 is restored by expanding it from existing PRC2 recruitment sites. This supports the theory that K27M might be inhibiting the spread of methylation repressive marks rather than retaining PRC2 at specific loci [107,108]. Therefore, conventional silencing profile doesn't extend across the genome, keeping it in an undifferentiated progenitor-like state, continuously dividing, time-dependent stockpiling of mutations, and subsequently giving rise to tumors. Although K27M has been deduced to be a driver mutation for pGBM, it is most likely not the only mutagenic event capable of propelling tumor formation [107]. The H3F3A mutations have been frequently shown to be capable of inducing cellular hyperproliferation but to give rise to voracious tumors, other driver mutations seem to be required such as TP53 mutations [100]. This corroborates with mouse studies, wherein H3K27M mutation coupled with P53 loss in Nestin progenitors, was able to generate proliferating ectopic clusters but not full-scale gliomas [109].

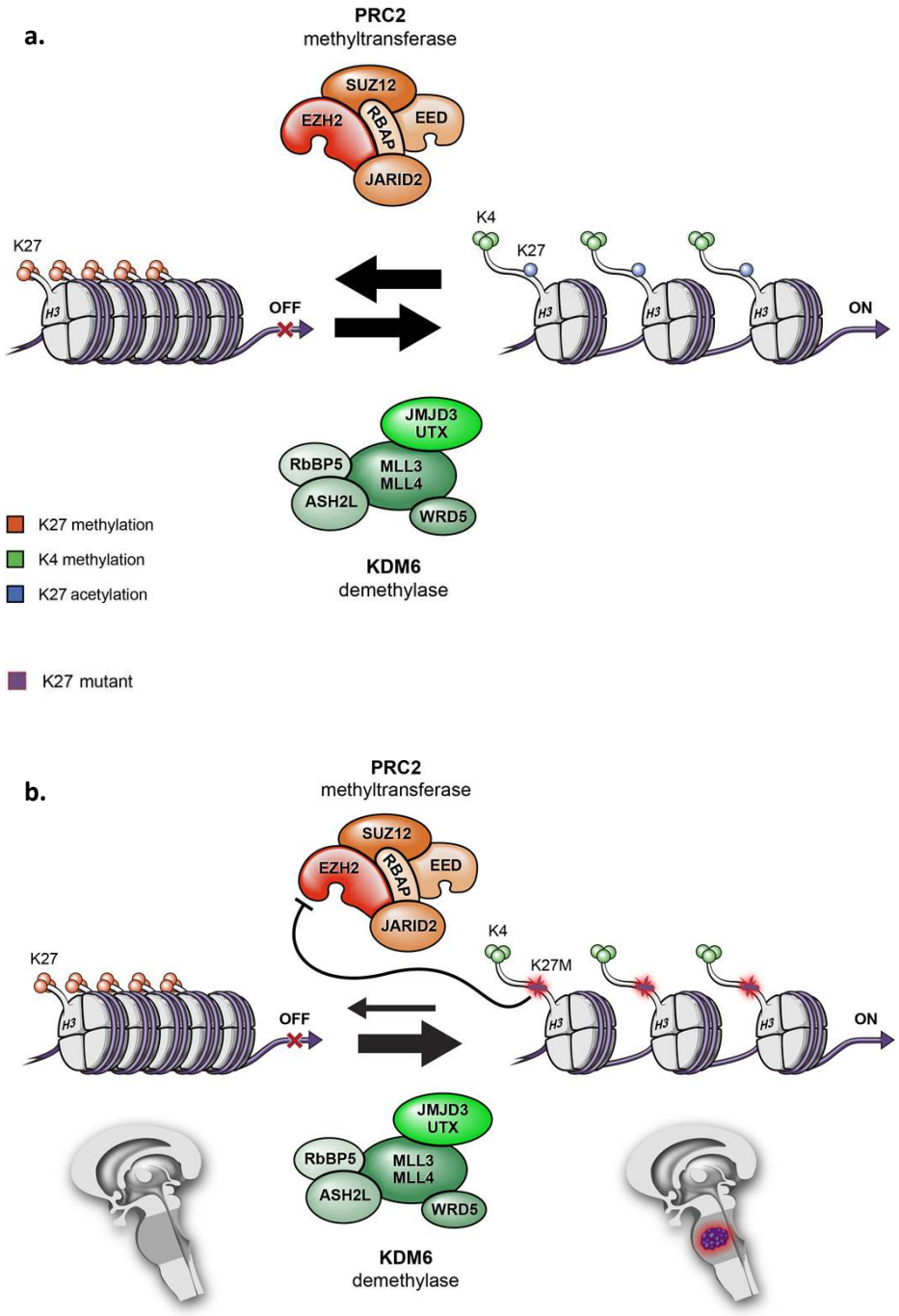


Figure 4. Histone modifications by PRC2 methyltransferase And KDM demethylase and model for global reduction of H3K27 methylation In K27M DIPG

Adapted from Lulla RR. et al, 2016 [110].

a. Histone proteins are modified by the PRC2 and KDM. PRC2 increases methylation of K27, which promotes a more compact and transcriptionally repressed chromatin state; while KDM removes methyl groups from K27 and increases methylation of K4 that, in combination, promotes an open and transcriptionally active chromatin state.

b. H3K27M mutant protein sequesters PRC2 and functionally inactivates it, leading to a global reduction of K27 methylation, thereby promoting an open chromatin structure that favors increased gene transcription.

H3G34R mutation

Glycine at position 34, though it doesn't undergo post-translational modifications, is in close proximity to K36, whose methylation pattern affects transcription. H3K36 trimethylation has an active role in DNA repair. The trimethylated amino acid interacts with the PWWP domain of MutSα, a protein involved in mismatch repair of DNA damage. MutSα being recruited to the site of damage initiates a mismatch repair process [111]. H3G34R/V/D converts the glycine residue into a bulky side chain residue that sterically inhibits protein interactions from occurring at the K36 position. These interactions involve the SETD2, an H3K36-specific trimethyl transferase, which recognizes the G33-34 motif. Similarly, NSD1/2 methyltransferase is also inhibited from binding at the site [108,111]. Furthermore, H3K36me3 is restricted from interacting with MutSα.

Sturm et al, 2012, distinguished 6 different subgroups of Glioblastoma based on global DNA methylation patterns- IDH, K27, G34, RTK I, mesenchymal, and RTK II [112]. The K27 subgroup showed a gene expression profile similar to the proneural subtype of the Verhaak glioblastoma profiles and was commonly derived from the brainstem. The G34 subgroup comparatively had a mixed expression profile and was found in the cerebrum, primarily the parietal and temporal lobes. Additionally, in the K27 subgroup, OLIG1/2 are highly differentially expressed, but are hypermethylated in the G34 subgroup, therefore showing lowest expression levels [112]; this imitates the scenario in embryonic stem cells, as a way of preventing neural lineage commitment (Figure 5). Although long thought that OLIG2 mediates p53 inactivation, this evidence indicates that there might be a secondary mechanism of p53 immobilization. FOXG1 expression in K27, on the other hand, is markedly abrogated (Figure 5).

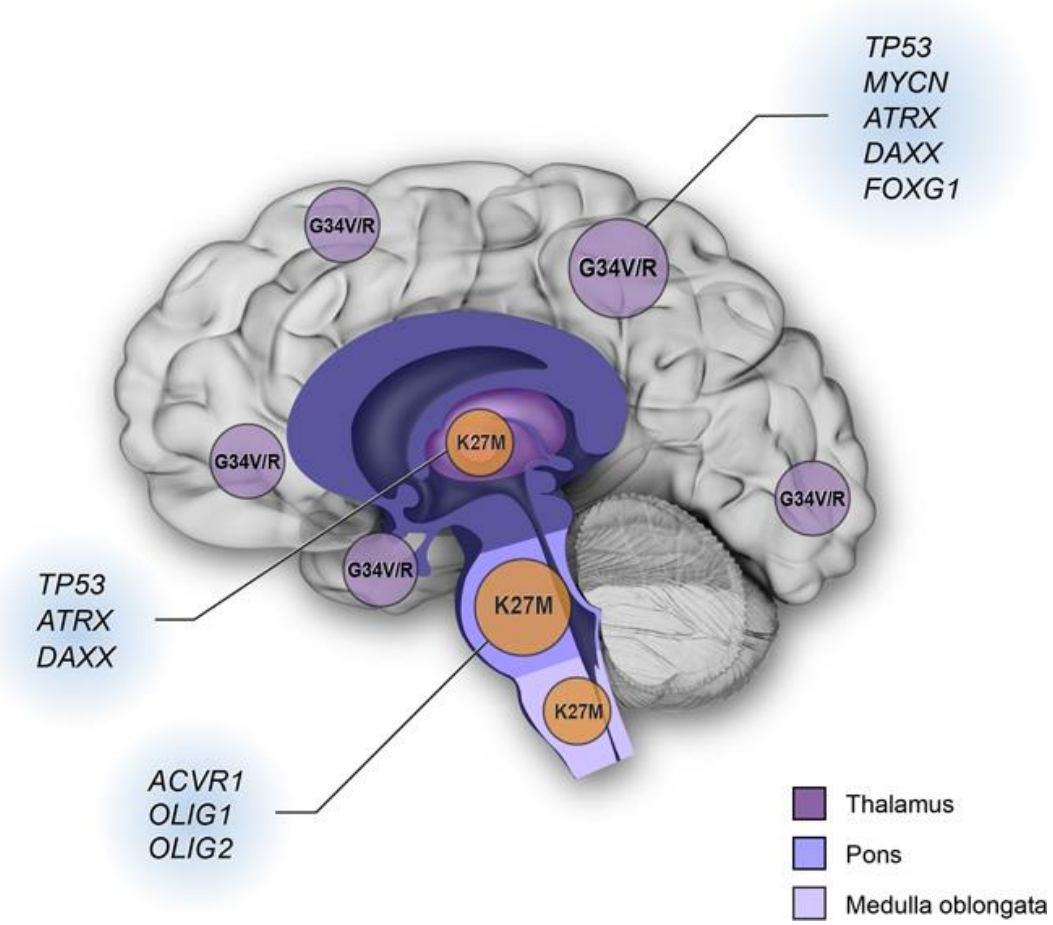


Figure 5. Neuroanatomic and gene associations with histone mutations

Adapted from Lulla RR. et al, 2016 [110]

K27M mutations are predominantly found in the tumors occurring in midline locations (thalamus, pons, and medulla oblongata). G34V or G34R (G34V/R) mutations are found in cerebral cortical tumors. Other gene alterations associated with histone gene mutations also occur in location-specific patterns. For example, TP53 mutations overlap with H3F3A mutations in cortical and thalamic tumors. ATRX and DAXX mutations are strongly associated with cortical G34V/R tumors and K27M mutant thalamic tumors, respectively. ACVR1 mutations are frequently present in histone H3.1 K27M mutant DIPG. OLIG1 and OLIG2 are highly expressed in K27M mutant tumors, whereas FOXG1 expression is predominantly found in G34V/R mutant tumors.

IDH mutations in pediatric glioblastoma

In contrast to the H3F3A mutations, IDH mutant glioblastomas are rare (6%), while H3F3A/IDH-wildtype tumors made up 36% of the cases [104]. H2K27M mutations are generally observed in children, G34R/V in adolescents, and IDH in young adults [112]. IDH1/2 enzymatically catalyze the oxidative carboxylation of isocitrate to α-KG, and reduce NADP to NADPH. The mutated IDH instead converts α-KG to 2-HG (2-hydroxyglutarate), which in turn affects the activation of histone lysine demethylases, increasing the methylation levels across the genome, especially in CpG islands, rendering the cell more prone to oncogenic transformations. It has also been hypothesized that 2-HG activates PHD3/EGLN which can cause reduced HIF quantities and enhanced astrocyte proliferation [94]. Moreover, diminished NADPH levels increase cell susceptibility to oxidative stress. This can be another initiator event for tumorigenesis but also means higher susceptibility to cytotoxic therapy. This can translate into better overall survival for these patients compared to IDH wild-type cases. Unfortunately, ~90% of glioblastomas are categorized as IDH-wildtype in pediatric cases, while they account for <5% of adult glioblastoma instances.

H3F3A and IDH mutations have been found to be mutually exclusive. Within the H3F3A/IDH-wildtype class, 3 subsets have been identified with distinct molecular signatures. These are the pGBM_MYCN (MYCN amplification, median OS of 14 months), pGBM_RTK1 (PDGFRA amplification, median OS of 14-44 months), and pGBM_RTK2 (EGFR amplification, median OS of 44 months) [113].

To summarize the molecular profile of pediatric glioblastoma cases, in a comprehensive molecular meta-analysis of 1067 pHGG and DIPG samples (45), a clear distinction between the tumor profiles of gliomas observed in different age groups and corresponding to different prognoses was made (Figure 6). It provides clear and extensive profiling of the driver mutations associated with a certain age group, common mutation profiles for each driver mutation, and corresponding prognoses.

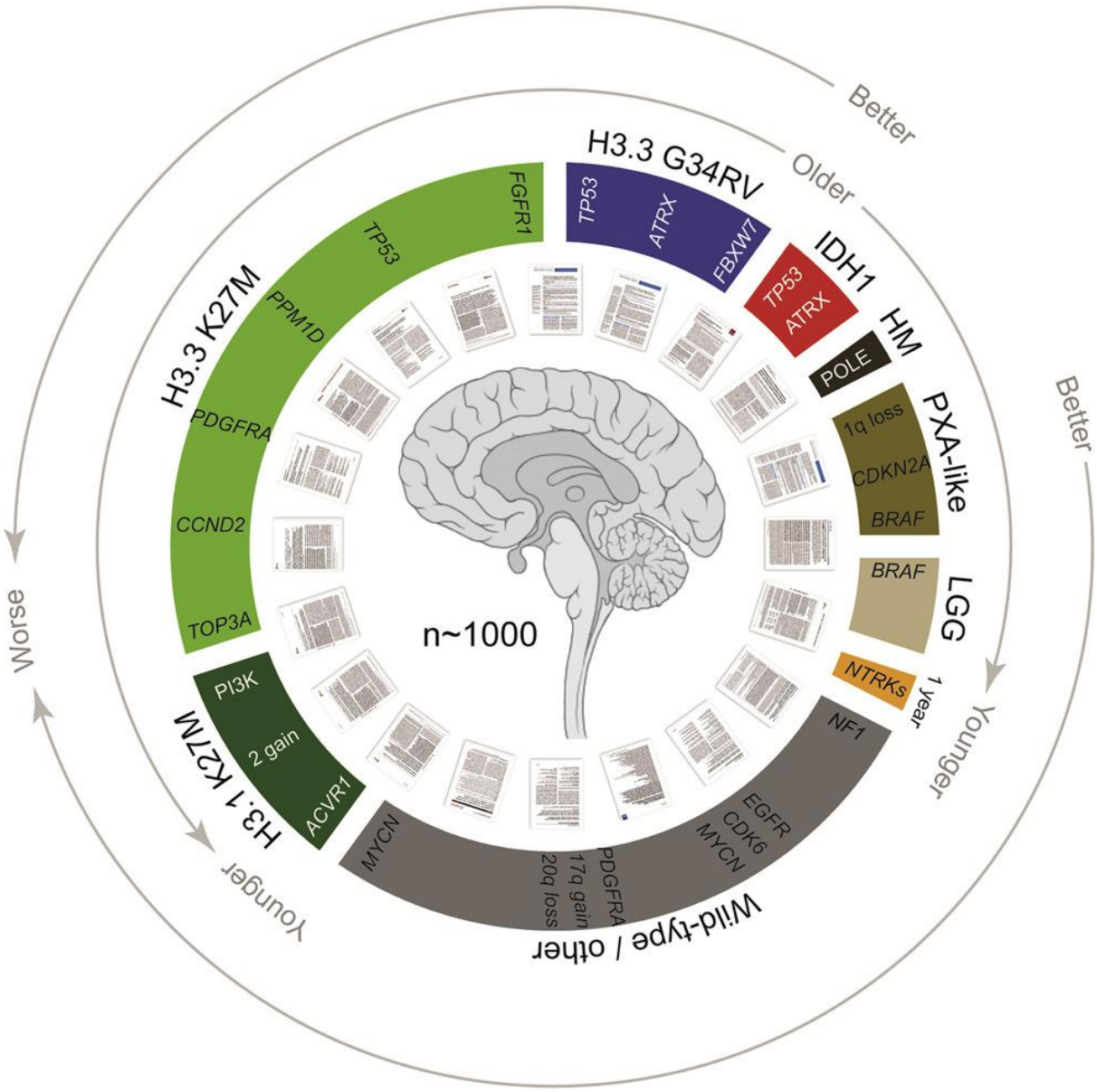


Figure 6. Summary of mutation profiles and prognoses related to pediatric high-grade glioma and diffuse intrinsic pontine glioma

Adapted from Mackay A. et al, 2017 [106]

Organoids as *in vitro* clinical research models

Organoids are three-dimensional, self-organizing, tissue cultures that reconstitute either a whole organ or a part of it. They are derived from stem cells, essentially differing from spheroid cultures that are free-floating undefined tissue cultures composed of cancerous tissue. They, offer several advantages over conventional 2D cell culture systems and have been steadily growing in popularity in research studies since 2014. For instance, they can recapitulate more realistically the *in vivo* environment, and the chronological development and maturation process of the organ. Organoids have now been generated for a glut of organs including the brain, intestine, stomach, kidney, lung, heart, and more [114].

Brain organoids have afforded important roles in several clinical applications such as disease models, cancer research, drug screening, and regenerative therapy.

Even with the relative success of organoids in basic research, there are still quite a few challenges encountered in culturing and maintaining these systems, first and foremost being the lack of reproducibility [114,115]. Since organoids typically develop in suspended Matrigel droplets without external support or molds, they give rise to unique structures in each individual culture. Additionally, even if nutritional and medium input, temperatures, and culturing procedures are kept standardized, there still exists variation between batches, cellular compositions, and architecture. Moreover, there is the complication that not all the cells that are induced to differentiate will move to the neuroectodermal commitment [116]. This can pose a problem in obtaining pure brain organoid cultures and also affect interpretation.

The second limitation experienced is the absence of vasculature. Once the organoids grow to a few millimeters in diameter, they can no longer efficiently transport oxygen via diffusion to the innermost cores. This leads to necrotic and underdeveloped cores. This can partly explain why most organoid models don't survive past 9-12 months of culturing and don't grow more than 5-7 millimeters in diameter. While this problem has been partially circumvented by using spinning bioreactors, more effective solutions are still being studied and have also shown fruitful results, such as using hollow fiber constructs [117], cocultures with endothelial cells [118], and genetic modification of genes such as ETV2 [119].

Thirdly, is the complete tissue maturation. Most of the early development and differentiation of the organoid is well-formed but fails to continue similarly as time progresses;

with some structures showing distinct architectures while other parts remain undifferentiated. As seen in 3D cultures established by Lancaster and Knoblich [83], the organoids don't achieve full maturity, and continue to retain a population of stem cell/neural stem cells.

Current state of research in brain organoids modeling glioblastoma

Over the past decade, organoids have been used extensively in research to model normal developmental processes but also to study altered cell states as seen in cancer, developmental disorders, diseases etc. Similarly, Glioblastoma has been a widely researched disease recapitulated *in vitro* in brain organoid systems using various methodologies.

This is evidenced in the study by Bian et al, 2018, where they developed neoplastic cerebral organoid (neoCOR), where CRISPR/Cas9 and/ or transposon constructs is used to introduce oncogenes or disrupt tumor suppressor genes in the hiPSC (human induced pluripotent stem cells) derived brain organoids. Particularly, they focused on genetic mutations that gave rise to visible over proliferation by tracking GFP tagged constructs, and delineated the differences between Myc overexpressed Brain organoids versus Glioblastoma signature brain organoids [120]. Primarily, Myc overexpressed brain organoids tend to feature overabundance of Sox2 and CD99, while Glioblastoma group organoids have a higher prominence of S100β and GFAP (astrocytes) expression.

In a similar fashion, CRISPR/Cas9 was used to introduce a HRasG12V mutation in a TP53 locus, in hESC (human embryonic stem cell) derived brain organoids that were 4 months old [121]. Amazingly, within 4 months of electroporation, the organoid showed an increased presence of the tdTomato-positive HRASG12V construct in up to 86% of the organoid. This type of an experimental construct, while it doesn't mimic commonly seen glioblastoma mutations, serves as an important way of visualizing tumor growth and expansion.

A significant cell population often found in glioblastoma patients are GSCs (Glioma stem cells). As exhibited in multiple studies [122–124] GSCs frequently remain inaccessible to tumor resection and develop resistance to currently approved chemotherapy. In order to be able to study them *in vitro*, the GLICO model was developed that involves hESC- or iPSC-derived brain organoids that are cocultured with patient derived GSCs and monitored closely to track invasion and proliferation [125]. Within 1 week, GFP+ GSCs can be observed showing a diffuse infiltrating edge, reminiscent of *in vivo* tumor morphology, and more than 20% of these tumors

show a KI67+ (proliferation marker) staining. Surprisingly, comparing the growth of GSCs in a 3D brain-like environment, demonstrates the increased sensitivity of the GSCs to drug and radiation toxicity versus when grown in monolayer conditions. This recapitulates another methodology for studying patient specific pathologies.

There have also been efforts to directly utilize clinical glioblastoma samples to generate organoid tissue in a fully defined, serum-free medium [126]. These GBOs (Glioblastoma organoids) are cut into 1mm pieces and directly placed in an orbital shaker, providing fully formed organoids in less than 2 weeks. This is in stark contrast to other protocols, that are iPSC-derived, that require a minimum of 1 month to start showing markers of cell maturity. Since these samples come directly from patients, certain structures such as blood vessels and CD31+ vasculature remain, recapitulating almost identically the tumor microenvironment in a patient. Although, the organoids were routinely dissected to prevent a necrotic core, it was still possible to visualize a hypoxic core, with reducing expression of KI67 from the outer rim to the core.

Mouse organoid models

In very limited cases, has mouse brain organoids been developed successfully. But the benefits to the development of such a clinical model is the shorter division cycles, leading to a shorter time scale to achieve maturity, useful for protocol optimization with lower production costs [127]. There is also a dearth of already readily available *in vivo* mouse experimental data that can be cross examined for comparison of pathophysiology, development markers, drug toxicity assays and more. Earlier protocols have been developed to use self-aggregating mouse embryonic stem cells (MESC) to develop into structures that resemble cortical structures and the optic cup [128,129]. More recently, mouse brain organoids resembling brain structures with a dorsal forebrain phenotype were developed [130]. In a span of 5-6 weeks, they produced mature cortical organoids, that are highly reproducible and robust. The Neural stem cells derived from the subgranular zone (SGZ) of mouse brain embryos (E14.5), are cultured over a period of 1 month while being exposed to epidermal growth factor (EGF), basic fibroblast growth factor (bFGF), and brain-derived neurotrophic factor (BDNF). In the first week of development, there is a high expression of KI67 and this switches to an identity of Neural stem cells later on characterized by Nes, Pax6, and Sox2 positive staining [130]. At 1 month,

these organoids have a higher expression of mature neurons (NeuN, Map1b, Snap25) and astrocytes (Gfap).

Thus, there is the possibility to model Glioblastoma in mouse brain organoids specifically. It provides benefits such as cheaper reagent costs, possible quicker timelines and a means to verify protocols before investing in human iPSC derived brain organoids for disease studies, as highlighted before.

Scope of thesis

This thesis is divided into two sections. The first section is focused on Retinoid driven neurogenesis. The second section will address the possibility to generate mouse brain organoids that recapitulate pediatric glioblastoma.

Section 1- Retinoid driven neurogenesis

There is a gap in literature that does not fully explain the connection between neuronal cell fate and individual retinoic acid receptor subtypes (RARα, RARβ, RARγ). In this section, the aim is to decipher if there exists the possibility of cell fate processes unique to each individual retinoic acid receptor. Subsequently, the goal is to use Single cell transcriptomics to visualize the distinct neuronal cell types generated from exposure to synthetic retinoid ligands during ES cell differentiation and more precisely determine the cell fate acquisition unique to RARβ and RARγ coactivation.

Section 2- Development of Mouse brain organoids that recapitulate H3.3 mutated pediatric glioblastoma

Current research has a focus on using patient derived samples and human stem cell lines to study diseases. However, there is a benefit to developing mouse *in vitro* models that can successfully recapitulate such diseases.

Here, we will work with H3.3 mutant cell lines – H3.3K27M, H3.3G34R- to develop mouse brain organoids that demonstrate tumorigenic characteristics such as proliferation, and overgrowth.

Affiliation with Nanotumor consortium

This project is part of a nationwide cooperation of research projects spanning 13 laboratories across France- known as the Nanotumor consortium (www.nanotumor.fr), issued from the Program Federateur Aviesan 2019, whose end goal is the construction of a multiresolution tumor cell atlas. The organization is divided into 4 working projects, each aimed at investigating different aspects. Our particular study belongs to the Working Project 1 aimed at isolating, characterizing the protein composition, and determining the structure of

key molecular complexes involved in cancer onset and progression. Under this target, our lab is focused on objective 1.2, which is the development of H3.3-mutated mouse organoid models for pediatric studies.

Results

Section 1- Retinoid driven Neurogenesis

Synergistic activation of RARβ and RARγ nuclear receptors restores cell specialization during stem cell differentiation by hijacking RARα-controlled programs

From previous studies we already knew the significant role of RARγ in inducing an endodermal fate in F9 EC cells, and RARα in inducing a neuronal fate in P19 EC cells [31]. And there is evidence of functional redundancy as well as specificity of each receptor – RARα, RARβ, RARγ. But these studies were performed in a short term period and following recent research that has shown that in long term culture, specific receptor isotypes can give rise to specialized neuronal subtypes [79], we decided to address the role of the individual activation of these receptors but also their combined role.

In order to decipher the role of each individual receptor, a 2D neuronal differentiation assay was conducted on P19 EC cells using synthetic agonists for each receptor isotype as well as combinations of the agonists for 10 days. Specifically, being aware of the fact that RARβ activation on its own shows no visible change in differentiation studies of EC cells [80], and that RARγ cannot induce a neuronal phenotype in P19 EC cells [31], we combined the ligands BMS641+BMS961 to verify if there is a boosted recovery of neuronal differentiation as compared to using only BMS961.

The following paper highlights this study in detail, where following this experimental setup, we also conduct similar experiments in each receptor isotype knockout cell lines. The neuronal differentiation capacity of the cells is visualized with immunocytochemistry analysis, and their transcriptomes analyzed to determine recovery potential of neuronal differentiation markers when using BMS641+961 synergistically compared to the positive control (ATRA).

The major discovery is that RARβ and RARγ synergistic activation leads to a higher recovery levels of markers associated with neuronal cells, and that the gene programs activated by these receptors are able to better recapture the neuronal signature when they are activated in RARα KO cells. This suggests an inhibitory effect of RARα in wildtype conditions that might no longer suppress the activity of RARβ and RARγ. Intersecting the gene programs that get activated in the presence of BMS753 in all conditions, with those that are activated in

the presence of BMS641+961 in all conditions, reveals a subset of programs of 235 genes that are termed as 'inhibited programs by the unliganded RARα'. Finally, from the gene programs we can trace a gene regulatory network that provides a master regulatory index that reveals the top 22 transcriptions factors regulating this pathway.

We extrapolated our findings in MESC cells (E14) where they were neuronally differentiated for a longer time period of 15 days; and neuronal differentiation was confirmed with gene expression analysis and immunocytochemistry assays. Additionally, we looked at the gene expression profiles of the Top 22 transcription factors, and were able to find as compared to our positive control (ATRA treatment), BMS641+961 treated cells express 10 of the TFs compared to BMS753 treated cells expressing only 5. Here too we noticed a central role of the TF Prdm8, a histone lysine methyltransferase that appeared to centrally control several downstream TFs that can be linked to have a role in neurogenesis.

Contribution

I am the co-first author of this study, shared with Elodie Mathieux. I performed the neuronal differentiation studies in MESC (E14) cells and the validations with gene expression studies and immunostaining assays. The study was published in the journal Life Science Alliance on November 29, 2022. The online version of this article can be found here https://doi.org/10.26508/lsa.202201627.

Research Article

TRANSPARENT PROCESS OPEN ACCESS Life Science Alliance

Synergistic activation of RARβ and RARγ nuclear receptors restores cell specialization during stem cell differentiation by hijacking RARα-controlled programs

Aysis Koshy[1,*], Elodie Mathieux[1,*], François Stüder[1], Aude Bramoulle[1], Michele Lieb[2], Bruno Maria Colombo[1], Hinrich Gronemeyer[2], Marco Antonio Mendoza-Parra[1]

How cells respond to different external cues to develop along defined cell lineages to form complex tissues is a major question in systems biology. Here, we investigated the potential of retinoic acid receptor (RAR)–selective synthetic agonists to activate the gene regulatory programs driving cell specialization during nervous tissue formation from embryonic carcinoma (P19) and mouse embryonic (E14) stem cells. Specifically, we found that the synergistic activation of the RARβ and RARγ by selective ligands (BMS641 or BMS961) induces cell maturation to specialized neuronal subtypes, and to astrocytes and oligodendrocyte precursors. Using RAR isotype knockout lines exposed to RAR-specific agonists, interrogated by global transcriptome landscaping and in silico modeling of transcription regulatory signal propagation, revealed major RARα-driven gene programs essential for optimal neuronal cell specialization and hijacked by the synergistic activation of the RARβ and RARγ receptors. Overall, this study provides a systems biology view of the gene programs accounting for the previously observed redundancy between RARs, paving the way toward their potential use for directing cell specialization during nervous tissue formation.

DOI 10.26508/lsa.202201627 | Received 23 July 2022 | Revised 10 November 2022 | Accepted 14 November 2022 | Published online 29 November 2022

Introduction

The potential of all-trans retinoid acid (ATRA) to induce differentiation of embryonic stem and embryonic carcinoma (EC) cells is well established (Soprano et al, 2007; Niederreither and Dollé, 2008). ATRA is a ligand for the three retinoic acid receptors (RARα, RARβ, and RARγ), and major medicinal chemistry efforts have resulted in the synthesis of ligands that are selective for each RAR isotype (de Lera et al, 2007; Álvarez et al, 2014). Multiple studies, including ours, demonstrated that P19 stem cells differentiate into neuronal precursors when treated with ATRA or the RARα-specific agonist BMS753, but they do not progress in differentiation when treated with the RARβ-specific agonist BMS641 or the RARγ-specific agonist BMS961 (Mendoza-Parra et al, 2016a).

Here, we have investigated the neuronal lineage–inducing potential of individual and combined subtype–specific retinoids in the two-dimensional monolayer culture of P19 EC and mouse embryonic stem cells (mESCs) (E14). We observe that, in addition to ATRA and RARα agonists, the combination of RARβ and RARγ agonists triggers a complex differentiation process generating a variety of neuronal subtypes, oligodendrocyte precursors and GFAP (+) astrocytes. This synergistic effect has been decorticated on the grounds of the RAR/RXR-driven gene programs, and the use of RAR subtype–deficient cells, which were instrumental for revealing the specificity of each of the synthetic ligands. Finally, we reveal that the RARβ+γ synergy, which involves a defined set of gene programs controlled by key master players, is antagonized in the presence of RARα, suggesting that an asynchronous activation of the various RARs leads to impaired neuronal specialization.

Results

Synergistic activation of RARγ and RARβ induces neuronal cell specialization in P19 embryonic stem cells

Using the well-established monolayer culture for efficient morphological P19 cell differentiation (Monzo et al, 2012; Mendoza-Parra et al, 2016a), we observed that after 10 d of treatment, ATRA or the RARα agonist BMS753 induced not only neuronal precursors, as revealed by immunofluorescence using the neuronal marker tubulin β-3 (TUBB3), but also mature neurons, as revealed by the microtubule-associated protein 2 (MAP2) (Fig 1A and B). In contrast, treatment with the RARβ-specific ligand BMS641 or the RARγ agonist BMS961 did not lead to neuronal differentiation. We only observed

[1]UMR 8030 Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry-val-d'Essonne, University Paris-Saclay, Évry, France [2]Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France

Correspondence: mmendoza@genoscope.cns.fr
*Aysis Koshy and Elodie Mathieux contributed equally to this work

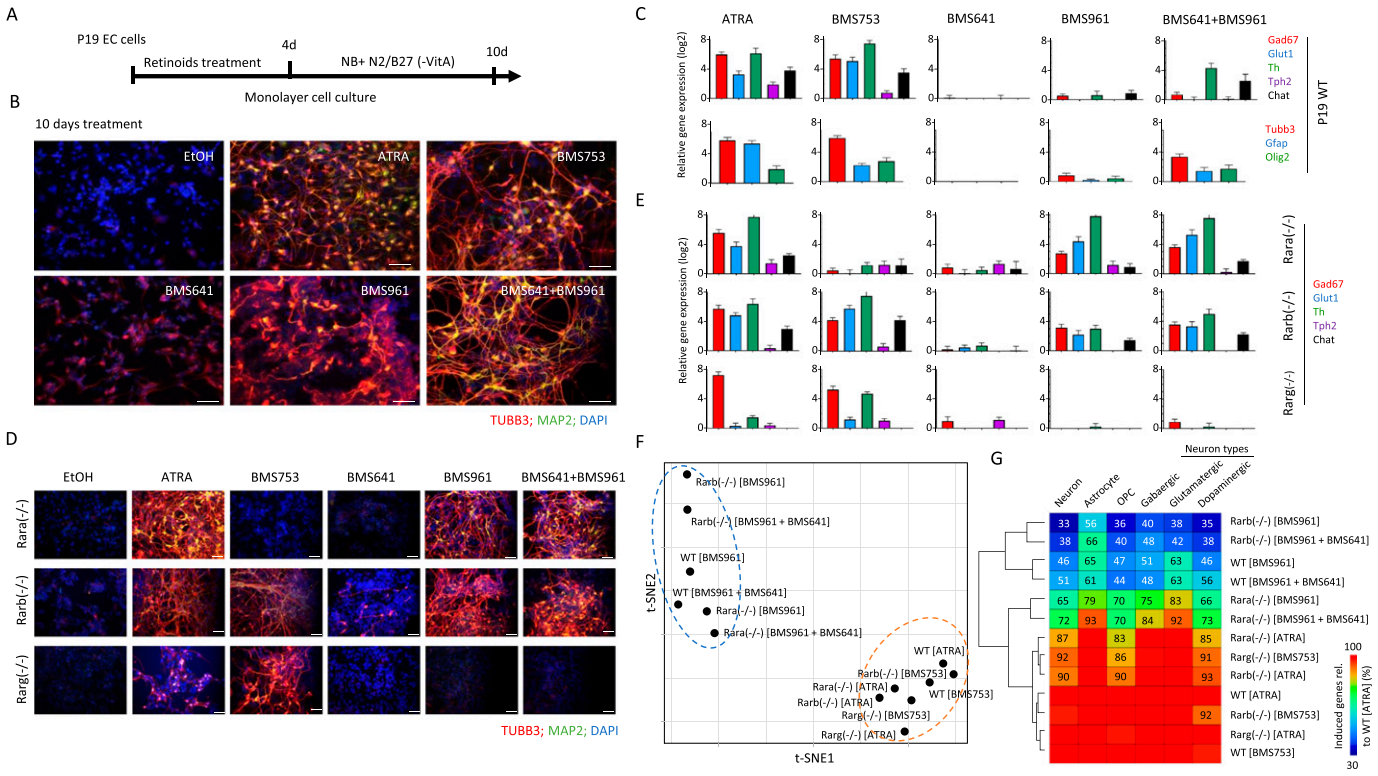


Figure 1. Synergistic activation of the RARγ and RARβ induces neuronal cell specialization in P19 embryonic stem cells.
(A) Schematic representation of the P19 cell differentiation assay. P19 cells cultured on monolayer are exposed to retinoids during 4 d to induce cell fate commitment; then, they are cultured for six more days on a synthetic medium (Neurobasal, NB) complemented with N2 and B27 (without vitamin A) supplements. **(B)** Immunofluorescence micrograph of WT P19 cells after 10 d of culture in presence of either ethanol (EtOH: vehicle control), all-trans retinoic acid, the RARα agonist BMS753, the RARβ agonist BSM641, the RARγ agonist BMS961, or the combination of RARβ and RARγ agonists. Cells were stained for the neuronal precursor marker TUBB3 (red) and the marker for mature neurons MAP2 (green). Nuclei were stained with DAPI (blue). **(C)** Top panel: RT–qPCR revealing the mRNA expression levels of gene markers associated with GABAergic (Gad67), glutamatergic (Glut1), dopaminergic (Th), or cholinergic (Chat) neuronal subtypes in samples treated with the indicated RAR agonists. Bottom panel: RT–qPCR mRNA gene expression levels of the glial fibrillary acidic protein (Gfap), the oligodendrocyte transcription factor 2 (Olig2) and the neuronal precursor marker Tubb3. **(D)** Immunofluorescence micrograph of P19 Rar-null mutant cells after 10 d of treatment with the aforementioned RAR agonists. **(E)** RT–qPCR mRNA expression levels of gene markers associated with the aforementioned neuronal subtypes assessed on P19 Rar-null mutant cells. **(F)** t-Distributed stochastic neighbor embedding analysis of differential gene expression readouts assessed on global transcriptomes performed on WT or Rar-null cells treated with specific agonists (10 d). Differential gene expression has been assessed relative to the ethanol-treated control sample (fold change levels >4). **(G)** Fraction of up-regulated genes (fold change levels >4) associated with markers corresponding to specialized cells relative to those observed on the gold-standard WT all-trans retinoic acid–treated sample. Fraction levels higher than 95% are only displayed with the heatmap color code (red).

neuronal-like cells presenting short neurite outgrowth structures, devoid of MAP2 immunostaining. Surprisingly, the combination of these two synthetic agonists (BMS641 and BMS961) restored neuronal differentiation presenting neurite outgrowth characteristics as similar as those observed on ATRA- or BMS753-treated samples (Fig 1B).

Neuronal maturation has been further supported by RT–qPCR assays revealing significant transcript levels associated with markers for GABAergic (*Gad67*), glutamatergic (*Glut1*), dopaminergic (*Th*), or cholinergic (*Chat*) neuronal subtypes in samples treated with ATRA and BMS753 (Fig 1C). Combined exposure to RARβ+γ agonists (BMS641 and BMS961) presented significant expression levels only for the markers *Th* and *Chat*, suggesting not only a partial neuronal subtype differentiation in comparison with ATRA or BMS753 treatment, but also the necessity of a more comprehensive strategy (global transcriptomes) to evaluate the cell specialization success. In addition to neuronal cell specialization, RT–qPCR assays also revealed significant expression levels of the glial fibrillary acidic proteins (*Gfap*) and oligodendrocyte transcription factor 2 (*Olig2*) genes, indicative of the presence of astrocytes and oligodendrocyte precursors, both in ATRA and BMS753 treatment and in the combination of BMS641 and BMS961 agonists (Fig 1C).

Although the combined exposure to RARβ+γ agonists (BMS641 and BMS961) led to morphological neuronal cell specialization, the evaluated markers present systematic lower levels than those observed in ATRA or BMS753. As this could be due to a potential inhibitory effect of non-liganded RARα, we engineered P19 cells deficient for each of the RARs using the CRISPR/Cas9 technology. Surprisingly, the absence of the expression of either RARα, RARβ, or RARγ receptor directly affected the expression of the non-deleted RARα and the RARβ receptors, notably by preserving their induction after 96 h of treatment (Fig S1). Furthermore, P19 Rara(−/−) cells gave rise to mature neurons when treated not only with ATRA, but also with the RARγ agonist BMS961 or the combination of RARβ+γ ligands (BMS641 and BMS961), as revealed by TUBB3/MAP2 immunostaining (Fig 1D) and the high expression of transcripts

associated with neuronal subtypes *Gad67*, *Glut1*, and *Th* (Fig 1E), *Tubb3* and the oligodendrocyte marker *Olig2* (Fig S2). The observed enhanced neuronal differentiation in the presence of the RARγ agonist in P19 Rara(−/−) relative to WT cells is in agreement with previous studies on the functional redundancy of RAR subtypes during endodermal (F9) and neuronal (P19) cell differentiation (Roy et al, 1995; Taneja et al, 1995, 1996). Neuronal differentiation was also observed in RARβ-deficient cells treated with not only ATRA or the RARα agonist (BMS753), but also the RARγ agonist (BMS961) and the RARβ+γ agonist combination (BMS641 and BMS961) (Fig 1D and E). Finally, the Rarg(−/−) cells entered neuronal differentiation only in the presence of ATRA or BMS753, in agreement with our earlier finding that the RARα-dependent gene program directs the neuronal cell fate of P19 cells (Mendoza-Parra et al, 2016a).

Although neuronal differentiation performance driven by RARβ+γ agonist treatment on WT, and Rara- and Rarb-deficient cells was evaluated by immunostaining and RT–qPCR assays targeting few marker genes, we reasoned that a comprehensive strategy could reveal potential differences among these multiple conditions. Global transcriptome assays performed on WT or RAR subtype–deficient cells treated with specific agonists during 10 d revealed between 1,340 and 2,250 up-regulated genes (fold change levels >4 relative to the ethanol control) allowing to query for cell specialization signatures and their corresponding divergencies between samples (Fig S3). Indeed, a t-distributed stochastic neighbor embedding analysis of their differentially expressed genes revealed two major groups. The first comprises WT or RAR subtype–deficient cells treated with either ATRA or the RARα agonist (BMS753) (Fig 1F). The second group gathers samples treated with the RARγ (BMS961) or the combination of RARβ and RARγ agonists (BMS641 and BMS961). This second group displays significant disparities among their components, with the transcriptomes of RARβ+γ agonist–treated Rara(−/−) cells being closer to group 1 than the others, in line with the observed neuronal differentiation (Fig 1F).

To further understand this classification through better characterization of the cell specialization signature during these differentiation conditions, we have collected an ensemble of gene markers associated with neurons (1,352 genes), astrocytes (501 genes), and oligodendrocyte precursors (OPCs: 501 genes), and stratified on GABAergic (318 genes), glutamatergic (311 genes), and dopaminergic (513 genes) neuronal subtypes (Table S1) (Hook et al, 2018; Tasic et al, 2018; Voskuhl et al, 2019). By comparing the number of up-regulated genes in the WT ATRA treatment with these comprehensive lists of markers, we have revealed that ~30% of them are associated with markers corresponding to specialized cells (632 from 2,158 up-regulated genes from which 401 are associated with neurons, 100 with astrocytes, and 131 with oligodendrocyte precursors) (Fig S3). Considering the up-regulated genes associated with specialized cells in the WT ATRA condition as the gold standard for optimal cell differentiation, we have revealed that most of the ATRA- or BMS753-treated samples presented similar amounts of genes associated with specialized cells. Indeed, the Rara(−/−) mutant sample treated with ATRA recapitulated ~83% of the gold-standard up-regulated genes associated with oligodendrocyte precursors, ~87% for neuronal-associated gene markers, and ~85% for the dopaminergic neuronal subtype; a similar behavior is observed for the Rarb(−/−) mutant treated with ATRA or the Rarg(−/−) mutant treated with the BMS753 ligand (Fig 1G). In contrast, samples treated with the RARγ agonist BMS961 give rise to cell marker levels of only ~30% in the context of the Rarb(−/−) mutant, and to ~70% in the context of the Rara(−/−). Importantly, treating the Rara(−/−) mutant with the combination of the RARβ+γ agonist (BMS641 and BMS961) leads to >70% of the gold-standard levels associated with neuronal cells, and even more than 90% for astrocyte or the glutamatergic neuronal cell type, demonstrating that the use of a synergistic RARβ+γ agonist treatment on RARα subtype–deficient cells leads to enhanced restoration of cell specialization during P19 stem cell differentiation.

P19 differentiation driven by the combination of RARβ and RARγ agonists presents a delayed expression of cell specialization markers

To assess the temporal evolution of gene expression during RAR ligand–induced neuronal differentiation and to associate specific gene programs with the appearance of neuronal cell subtypes, we generated global transcriptomes after 2, 4, and 10 d of treatment of WT P19 cells. This has been performed for samples treated with either the pan-agonist ATRA—as a gold-standard treatment—the RARα-specific agonist BMS753, or the combination of RARβ and RARγ agonists (BMS641 and BMS961), shown to induce neuronal cell specialization.

Differential gene expression through the aforementioned timepoints —assessed during ATRA treatment— was classified into 14 relevant gene co-expression paths, defined herein as a group of genes with similar temporal changes of expression levels (Fig 2A). For instance, Path 1 (1,195 genes) corresponds to genes up-regulated (fold change >2 relative to vehicle at d0) after 2 d of treatment and remained overexpressed until day 10. Paths 2 (725 genes) and 4 (582 genes) comprise late up-regulated genes, induced only at d4 and d10, respectively (Fig 2A). A gene ontology term analysis performed over the first seven gene co-expression paths—associated with up-regulated genes at least at one timepoint—reveals that Path 1 comprises genes involved in neuronal differentiation, nervous system development, or axon guidance, and Path 2, in axonogenesis or axon development, whereas Path 4 is associated with chemical synaptic transmission, synaptic vesicle budding, or synapse organization, in agreement with the time of induction along the neuronal differentiation lineage and subtype specification (Fig 2B).

As expected, the up-regulated gene co-expression paths of WT P19 cells contained also markers indicative of the various specialized cells, as revealed by comparison with the aforementioned collection resource (Table S1). As illustrated in Fig 2C, 194 genes of the early-responsive gene co-expression Path 1 corresponded to neuronal markers, whereas 65 genes corresponded to astrocytes and other 68 genes corresponded to oligodendrocyte precursors (OPCs). The intermediate-responsive Path 2 presented 107 genes associated with neurons, 32 with astrocytes, and 35 with OPCs, whereas the late-responsive Path 4 presented 93 neuronal markers, 12 genes associated with astrocytes, and 30 others associated with OPCs. These kinetics indicate that neuronal differentiation precedes glial cell emergence, in agreement with previous findings in

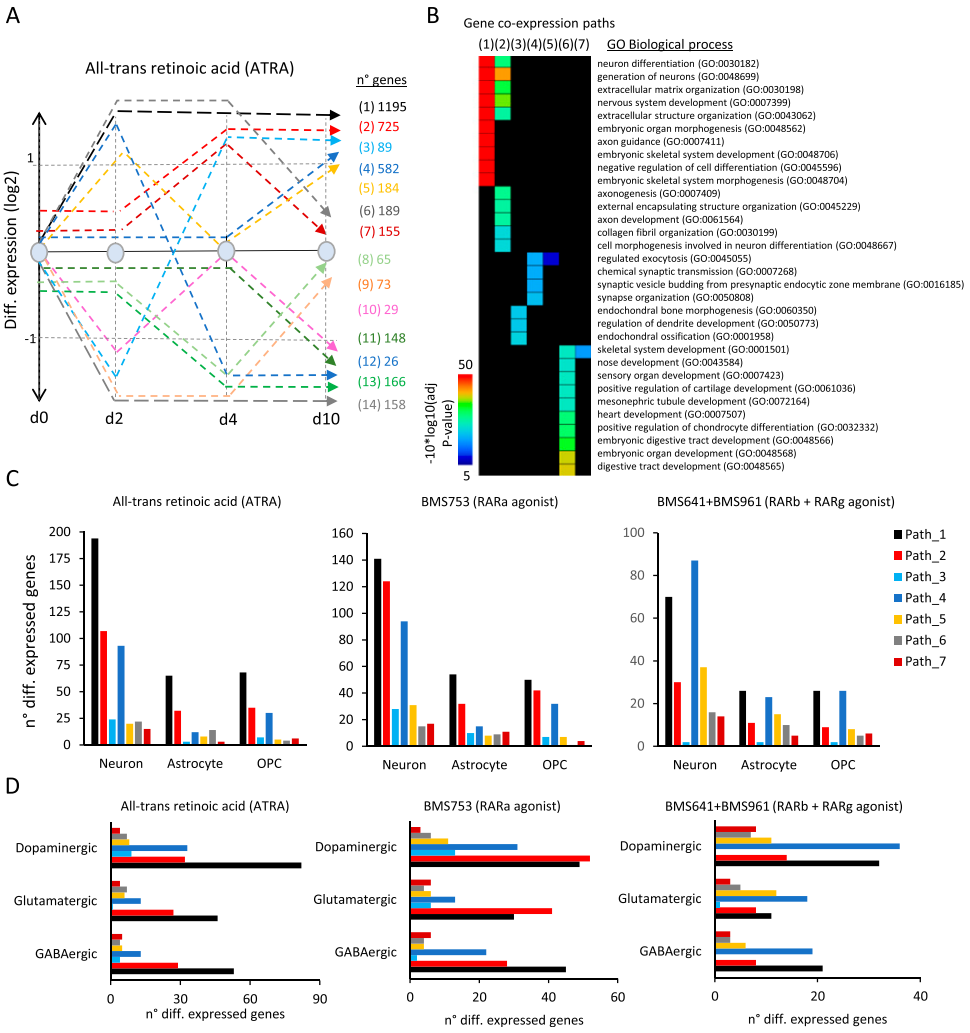


Figure 2. Temporal gene co-expression analysis during cell specialization driven by retinoid treatment.
(A) Stratification of the temporal transcriptome profiling during WT P19 cell differentiation driven by ATRA treatment. Transcriptomes were assessed on samples collected at 2, 4, and 10 d of treatment. Dashed lines correspond to groups of differentially co-expressed genes (gene co-expression paths; fold change levels >2). The numbers of genes composing each of the co-expression paths are displayed (right). **(B)** Gene ontology analysis on gene co-expression paths displayed in (A) associated with up-regulated events. **(C)** Number of genes per co-expression path corresponding to neuronal, astrocyte, or oligodendrocyte precursor cell types assessed during ATRA (left panel), BMS753 (middle panel), and BMS641 + BMS961 (right panel) treatment. **(D)** Similar to (C) but corresponding to dopaminergic, glutamatergic, and GABAergic neuronal subtypes.

in vivo and in vitro mammalian systems (reviewed in Miller & Gauthier [2007] and Hirabayashi & Gotoh [2010]). All other paths presented less than 25 genes corresponding to the aforementioned cells, indicating that early (Path 1)-, middle (Path 2)-, and late (Path 4)-responsive co-expression paths are the most relevant for describing cell specialization (Fig 2C and D).

Although WT P19 cells treated with the RARα agonist BMS753 presented relatively similar transcriptome kinetics, the number of markers associated with specialized cells retrieved on Path 1 was lower than that observed on the gold-standard ATRA treatment (141 genes associated with neurons, ~54 genes with astrocytes, and 50 genes with OPCs). This observation for Path 1 has been further enhanced on cells treated with RARβ and RARγ agonists (BMS641 and BMS961), including in addition a significant reduction in the number of gene markers associated with specialized cells on Path 2. In contrast, the number of gene markers observed on the late-responsive Path 4 remained rather unchanged (87 genes associated with neuronal markers, 23 with astrocytes, and 26 with OPCs) (Fig 2C). This observation suggests that although treatment with the combination of the RARβ+γ agonist gives rise to specialized cells, their differentiation process is delayed over time relative to that observed under the ATRA treatment. This is also supported by the fact that cells under the RARβ+γ agonist treatment present gene markers associated with specialized neurons preferentially found on the late-responsive Path 4 (Fig 2D).

Reconstruction of gene regulatory networks (GRNs) involved in cell specialization driven by retinoid treatment

Our previous work has shown that ligand binding of retinoid receptors triggers a cascade of events, which leads to the dynamic activation of other transcription factors (TFs), which then regulate their cognate targets. This cascade of transcription regulatory events can be reconstructed by integration of transcription factor–target gene (TF-TG) databases in the temporal transcriptome analysis (Cahan et al, 2014; Mendoza-Parra et al, 2016a). This way, GRNs can be reconstructed and master regulator genes deduced (Cholley et al, 2018).

Herein, we have reconstructed a master GRN from the integration of the temporal transcriptomes assessed on WT P19 cells treated with the pan-agonist ATRA, covering 10 d of cell treatment, with TF-TG annotations (CellNet database [Cahan et al, 2014]). This master GRN, composed of 1,156 nodes (genes) and 17,914 edges

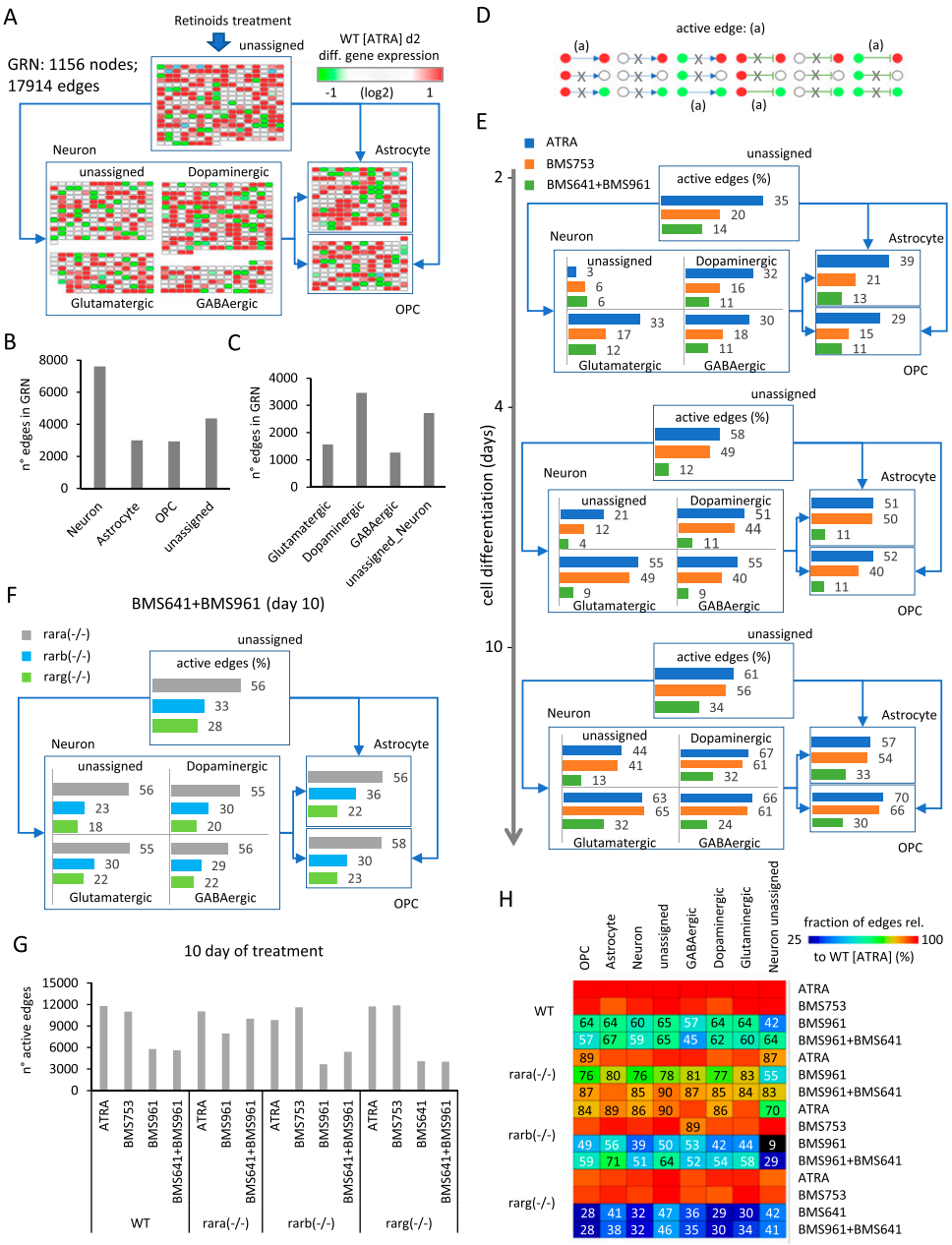


Figure 3. Active gene regulatory wire reconstruction during cell specialization driven by retinoids.
(A) Structure of the reconstructed gene regulatory network (GRN) displaying differentially expressed genes stratified into four major groups: neuronal cell markers (582 nodes), astrocytes (161 nodes), oligodendrocyte precursors (OPCs: 133 nodes), and a fourth group composed of genes not retrieved in none of the previous classifications (unassigned: 280 nodes). Nodes associated with the neuronal group have been further stratified on dopaminergic (214 nodes), glutamatergic (111 nodes), GABAergic (87 nodes), or unassigned (170 nodes) neurons. For illustration purposes, all edges were removed and replaced by simplified connectors (blue arrows). The color code associated with nodes reflects the differential gene expression levels in WT P19 cells after 2 d of all-trans retinoic acid (ATRA) treatment. **(B, C)** Number of edges interconnecting nodes retrieved on each of the aforementioned groups. **(D)** Scheme illustrating all potential types of node states (active: red; repressed: green; and unresponsive: white) and their inter-relationships defined by the illustrated edges (positive regulation: arrow connector; negative regulation: t-shaped connector). "Active edges" (a) correspond to transcriptionally relevant node/edge relationships and are conserved during the analytical processing of the GRN illustrated in (A). **(E)** Temporal transcription evolution of the reconstructed GRN during WT P19 cell differentiation. Illustrated barplots correspond to the fraction of active edges (as defined in (D)) relative to the total edges (displayed in (B, C)) issued from the treatment with either the ATRA, the RARα-specific agonist BMS753, or the combination of RARβ and RARγ agonists (BMS641 + BMS961). **(F)** Barplots corresponding to the fraction of active genes after 10 d of treatment with the RARβ and RARγ agonists (BMS641 + BMS961) of P19 Rar-null mutant lines. **(G)** Number of total active edges retrieved in GRNs issued from 10 d of treatment with the indicated retinoids and over the different P19 lines. Notice that the number of active edges on the Rara(−/−) line treated with the combination of BMS641 + BMS961 agonists leads to similar levels to those observed on the gold-standard WT line treated with the pan-agonist ATRA. **(H)** Fraction of active edges (relative to the WT line treated with ATRA) associated with markers corresponding to the classification of specialized cells retrieved in (A) relative to those observed on the gold-standard WT ATRA–treated sample. Fraction levels higher than 90% are only displayed with the heatmap color code (red).

(TF-TG relationships), was stratified based on the presence of nodes (genes) associated with a given type of specialized cell type, as described in the aforementioned collection resource (Table S1) (Hook et al, 2018; Tasic et al, 2018; Voskuhl et al, 2019) (Fig 3A). Specifically, the GRN has been first stratified into four major groups: neuronal cell markers (582 nodes; 7,611 edges), astrocytes (161 nodes; 3,007 edges), oligodendrocyte precursors (OPC: 133 nodes; 2,929 edges), and a fourth group composed of genes not retrieved in none of the previous classifications (unassigned: 280 nodes; 4,367 edges) (Fig 3A and B). Nodes associated with the neuronal group have been further stratified on dopaminergic (214 nodes; 3,457 edges), glutamatergic (111 nodes; 1,568 edges), GABAergic (87 nodes; 1,264 edges), or unassigned (170 nodes; 2,717 edges) neurons (Fig 3A and C).

One of the major advantages of working with reconstructed GRNs is the fact that the relevance of the system can be challenged by the coherence of the interconnected players. In this case, we define an "active edge" as a set of two nodes being differentially responsive and interconnected with a transcription regulatory relationship (active or repressed regulation) coherent with the gene expression status of the interconnected nodes (e.g., active genes require to be

interconnected by active transcription regulatory edges; Fig 3D). Hence, during differentiation of WT P19 cells treated with the pan-agonist ATRA, the fraction of active edges passes from ~30% to ~50% and finally ~60% when evaluating readouts at 2, 4, and 10 d of treatment associated with specialized cell types (Fig 3E). Interestingly, WT P19 cells treated with the RARα agonist (BMS753) present a lower number of active edges after 2 d of treatment (21% for astrocytes, 15% for OPCs, and ~17% for the specialized neurons), but recovered similar levels than those observed with the ATRA treatment in the late time-points. In contrast, the use of the combination of the RARβ+γ agonist (BMS641 + BMS961) raises the levels of active edges to barely ~30% after 10 d of treatment (Fig 3E). Although this poor performance is also observed in P19 lines deficient for the RARβ or the RARγ receptor, the Rara(−/−) P19 mutant line revealed a significant gain in the number of active edges (56% for astrocytes, 58% for OPCs, and ~55% for specialized neurons) (Figs 3F and S4).

Indeed, although GRNs corresponding to WT P19 cells treated with BMS961 and BMS641 agonists present ~half of active edges observed on ATRA treatment conditions (5,628 edges), P19 Rara(−/−) mutant cells under the same retinoids' treatment lead to GRNs presenting 10,044 edges interconnecting responsive genes (Fig 3G). The comparison between the number of active edges under different conditions and that observed on the gold-standard P19 WT ATRA treatment reveals that the Rara(−/−) mutant treated with BMS961 and BMS641 agonists leads to a recovery of >80% for all cell-specialized groups (Fig 3H). In contrast, the same treatment on WT P19 cells reaches levels of ~60% for most of the groups, with the exception of the GABAergic neuronal subtypes where only 45% of edges observed on ATRA treatment are recovered.

Overall, the reconstructed GRN describing cell type specialization during retinoid-driven cell differentiation reveals the fraction of reactivated edges by the synergistic activation of the RARβ and RARγ receptors, notably in the Rara mutant line.

Enhanced restoration of neuronal cell specialization driven by RARγ and RARβ receptors requires to bypass gene programs controlled by the RARα receptor

A deeper analysis of the reconstructed GRN revealed a twofold increase in the number of active edges for the P19 Rara(−/−) mutant relative to the WT line when treated with RARβ+γ agonists (BMS641 and BMS961). Such enhanced performance could be explained by a functional redundancy of RAR subtypes, as previously demonstrated during the early phases of endodermal (F9) and neuronal (P19) cell differentiation (Roy et al, 1995; Taneja et al, 1995, 1996). Specifically, we speculate that in the absence of the RARα receptor, the synergistic activation of the RARβ and RARγ receptors could drive the activation of RARα-specific programs. Similarly, such RARα-specific programs might remain "inhibited" in WT cells (for instance, because of the unliganded binding of the RARα receptor) despite the combined exposure to RARβ+γ agonists.

To address this hypothesis, we first identified the RARα agonist (BMS753)–specific programs, corresponding to the common active edges between WT, Rarb(−/−), and Rarg(−/−) mutant lines treated with this synthetic ligand (Fig 4A). The obtained 9,328 active edges were then intersected with those observed on WT or Rara(−/−) mutant lines treated with RARβ+γ agonists, to reveal those programs commonly activated by both treatments (3,806 active edges), and those specifically activated by the RARα agonist (BMS753) but inhibited in the WT line despite of the combined exposure to RARβ+γ agonists (3,830 active edges).

A close look at the "common" and "inhibited" programs revealed that, despite their distinct number of edges, most of the genes composing the "common" program are also part of the "inhibited" program (414 from 452 genes), and this observation is also conserved for the involved TFs (161 shared TFs; Fig 4B). This observation suggests that gene expression for the "common" and "inhibited" programs is differentially controlled by other molecular factors in addition to TF regulation. To address this hypothesis, we have evaluated their promoter epigenetic status (defined by the repressive mark H3K27me3 and the active mark H3K4me3), their chromatin accessibility (revealed by FAIRE-sequencing assays), and their transcriptional response (revealed by the enrichment of the RNA Polymerase II) after 2 d of ATRA treatment (GSE68291 [Mendoza-Parra et al, 2016a]). This enrichment analysis (relative to those observed on EtOH vehicle treatment) revealed that genes being either specifically "inhibited" in WT (235 genes), shared between the "inhibited" and the "common" programs (414 genes), or specifically associated with the "common" programs (38 genes) are preferentially repressed at 48 h of ATRA treatment, as revealed by the enrichment of the H3K27me3 modification (Fig 4C and D).

Genes associated with the "inhibited" program are induced between the 4 and 10 d of ATRA treatment, or the synthetic RARα agonist (BMS753), but they remain unresponsive in presence of the combined exposure to RARβ+γ agonists. This being said, the combined exposure to RARβ+γ agonists leads to their gene induction on the P19 Rara(−/−) mutant line (Fig 4E).

With the aim of confirming the role of the RARα receptor on the predicted "inhibited" program, we have performed an enrichment analysis on their associated 219 FAIRE sites (Fig 4F), by comparing them with ChIP-seq binding sites collected from the public domain. Specifically, we have used a collection of more than 40,000 public mouse ChIP-seq datasets, collected as part of our NGS-QC database (https://ngsqc.org/), among which 71 ChIP-seq public profiles correspond to RXR or RAR TFs (Mendoza-Parra et al, 2013, 2016b; Blum et al, 2020). This analysis revealed that ~21% of the FAIRE sites associated with the "inhibited" program were enriched for RARα binding sites, and ~20% for pan-RXR sites, further supported by a motif analysis revealing the enrichment of the RARα primary motif (Fig 4G and H), confirming their transcriptional response driven by the RARα/RXR heterodimer.

With the aim of summarizing this information within a gene regulatory wiring, we have first assembled the FAIRE- and H3K27me3-associated "inhibited" programs into a GRN, complemented by edges issued from public ChIP-seq binding sites associated with various RARs and RXRα receptors, and with the RXRα primary motif discovery. This summarized "inhibited RXR/RAR" GRN is composed of 85 nodes and 160 edges, on which each node has been highlighted on the basis of their promoter epigenetic status (Fig 5). To further enhance the relevance of master TFs within this "inhibited" network, we have computed their master regulatory index by simulating transcription regulatory cascades over the complete reconstructed GRN (described in Fig 3A; TETRAMER [Cholley et al, 2018]). A ranking of the TFs on the basis of their master regulatory index allowed to identify a set of 22 TFs able to

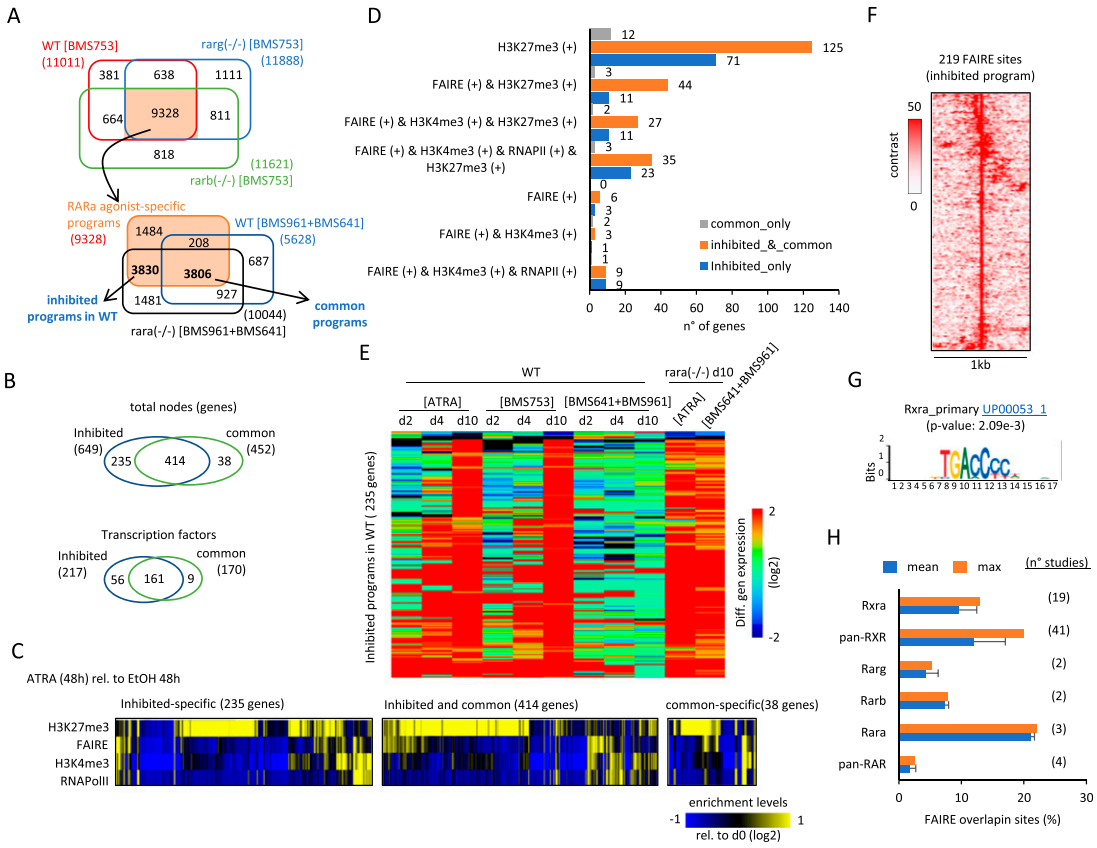


Figure 4. Identification of a subset of active edges remained inhibited by the unliganded RARα receptor in P19 WT cells during the synergistic activation of the RARβ and RARγ.
(A) Top panel: Venn diagram revealing the RARα-specific programs corresponding to the common active edges retrieved in WT, Rarg(−/−), and Rarb(−/−) P19 lines treated with the RARα agonist BMS753 (9,328 active edges). Bottom panel: Venn diagram revealing the "common programs" (3,806 active edges) driven by the RARα agonist and those responding to the synergistic activation of the RARβ + RARγ receptors; and a subset of active edges specifically driven by the RARα agonist BMS753 (3,830). This last subset is defined herein as "inhibited programs by the unliganded RARα," because they remain unresponsive on WT cells treated with the combination of RARβ + RARγ ligands (BMS641 + BMS961), but they are reactivated on the Rara(−/−) line. **(B)** Top Venn diagram: comparison between the number of genes retrieved in the "inhibited" and the "common" programs highlighted in (A). Bottom Venn diagram: comparison between the number of transcription factors retrieved in the "inhibited" and the "common" programs highlighted in (A). **(C)** Heatmap illustrating the promoter epigenetic status (repressive mark H3K27me3 and the active mark H3K4me3), the chromatin accessibility (FAIRE), and the transcriptional response (RNA Polymerase II) of genes specific to the "inhibited" or the "common" program, and those shared between these two programs after 2 d of ATRA treatment. **(D)** Number of genes presenting the indicated promoter epigenetic combinatorial status in the conditions illustrated in (C). **(E)** Heatmap displaying the differential expression levels for genes associated with the "inhibited"-specific program at different time-points and retinoid treatment. Notice that although most of these genes remained unresponsive when treated with the combination of RARβ + RARγ ligands (BMS641 + BMS961) in WT cells, they are up-regulated on the Rara(−/−) line. **(F)** Open-chromatin FAIRE sites retrieved on the promoters of the "inhibited programs by the unliganded RARα." **(G)** Motif analysis performed on the FAIRE sites presented in (F), revealing the enrichment of the RXRα primary motif. **(H)** Binding site enrichment analysis performed on the aforementioned FAIRE sites, by comparing with 71 RXR or RAR ChIP-seq publicly available profiles (NGS-QC Generator database: https://ngsqc.org/). Blue bars correspond to the mean fraction of binding sites, and orange bars correspond to the highest fraction assessed over the indicated number of studies.

regulate more than 70% of the ATRA-driven gene programs (Fig 5A and B). 15 of them present a transcriptionally active signature after 2 d of treatment, as revealed by their FAIRE-associated promoter status, whereas the remaining seven are rather repressed (H3K27me3-associated promoter status) (Fig 5C). Interestingly, this last group of TFs is composed of players like the T-box family member *Tbx18* (known to be regulated by retinoic acid during somitogenesis [Sirbu & Duester, 2006]), the transcriptional repressors *Hic1* (known to have a role in neural differentiation and tumor suppression in the central nervous system [Rood & Leprince, 2013]), *Hes6* (known to promote cortical neuronal differentiation through the repression of the TF *Hes1* [Gratton et al, 2003]), the mesoderm-specific factor *Tcf21*, the RNA binding protein *Csdc2*, the nuclear factor IC (*Nfic*; known to regulate cell proliferation and differentiation in the central nervous system notably by modulating the expression of the miR-200b [Huang et al, 2021]) and *Zfp827* (also known as ZNF827) recently shown to negatively regulate neuronal differentiation through the expression of its circular RNA (Hollensen et al, 2020). Importantly, all these repressed factors appear interconnected within the reconstructed GRN and preferentially associated with the RARβ or RARγ receptors (Fig 5D).

Among the FAIRE-associated factors, several of them present a RXRα binding site, including the zinc-finger TF *Tshz3*—recently described as a "hub" gene involved in early cortical development (Caubit et al, 2016); *Nfix*, recently shown to drive astrocytic maturation within the developing spinal cord (Matuzelski et al, 2017); and

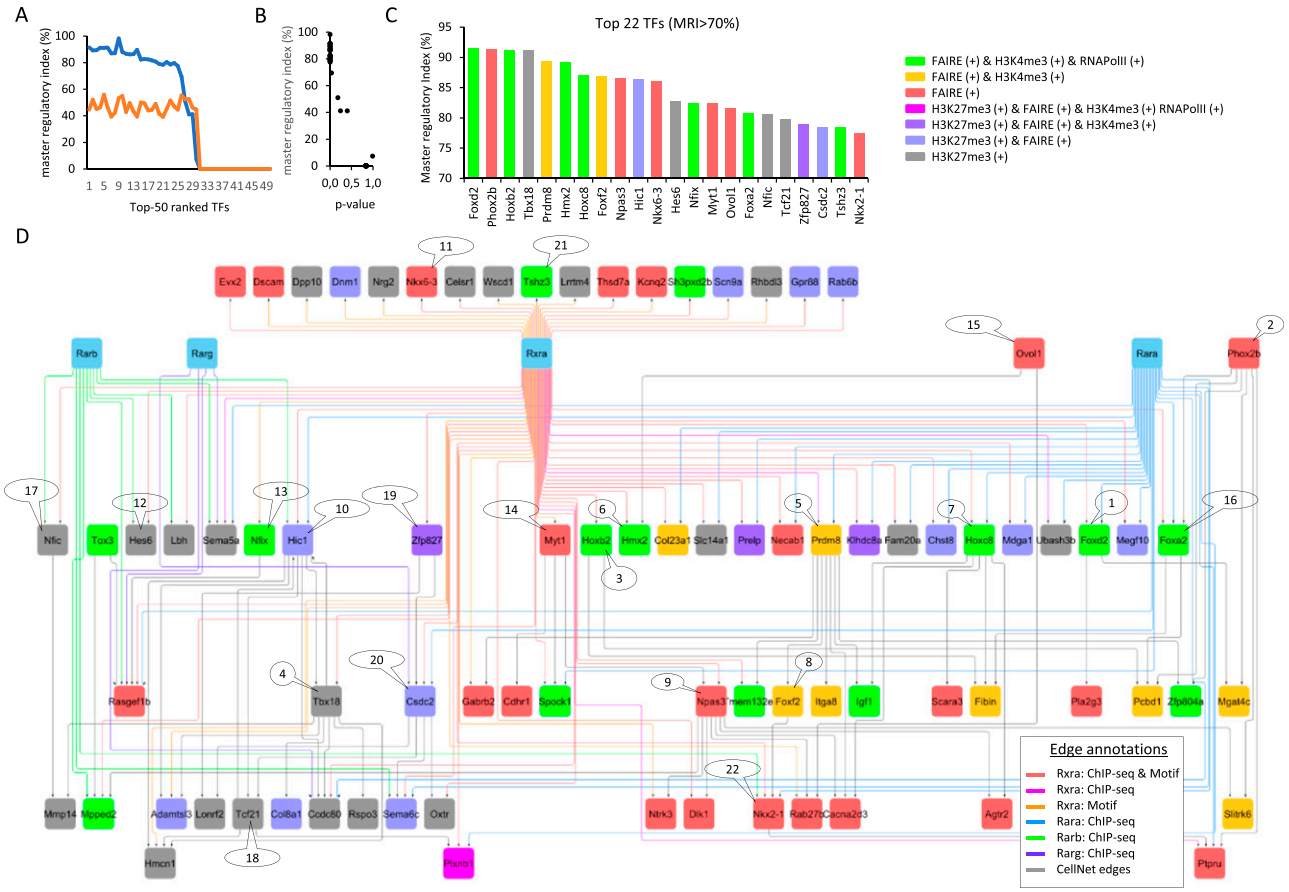


Figure 5. Gene regulatory view of the master players inhibited by the unliganded RARα receptor during neuronal cell specialization.
(A) Top 50 transcription factors retrieved within the "inhibited program" ranked on the basis of the fraction of downstream controlled genes within the reconstructed GRN (Fig 4A) (blue line). The orange line corresponds to the fraction of downstream controlled genes predicted on a randomized GRN (master regulatory index [MRI] computed as described in TETRAMER: Cholley et al, 2018). **(B)** Confidence associated with the TFs' ranking. Notice that a MRI > 70% presents the most confident *P*-values. **(C)** Transcription factors (22) presenting a MRI > 70% and colored on the basis of their promoter epigenetic combinatorial status. **(D)** Gene co-regulatory view of the 22 TFs, illustrating their most relevant (co-)regulated players and including their known relationships with RXRα and RAR nuclear receptors.

the homeodomain TFs *Hoxb2*, *Nkx6-3* (involved in the development of the central nervous system; the homeobox family factor), or *Nkx2-1*—known to control GABAergic interneurons and oligodendrocyte differentiation, and more recently described as driving astroglial production by controlling the expression of the *Gfap* (Minocha et al, 2017). Similarly, the homeobox factor *Hmx2*, the myelin transcription factor 1 (*Myt1*), or the Neuronal PAS domain protein 3 (*Npas3*) does present proximal RXRα binding sites, whereas other players like the forkhead TFs *Foxa2*, *Foxd2*, or *Foxf2*, the homeobox factor *Hoxc8*, or the histone lysine methyltransferase factor *Prdm8* do present in addition a previously described proximal RARα binding site (Fig 5D). Interestingly, *Prdm8* appears as a central player within the reconstructed GRN. Indeed, *Prdm8* controls seven other factors known to be highly expressed in nervous tissue: *Gabrb2* (the β2 subunit of the $GABA_A$ receptors, known to play a crucial role in neurogenesis and synaptogenesis [Barki & Xue, 2022]), the transmembrane protein *Tmem132e*, *Zfp804a* (known to regulate neurite outgrowth and involved in neuronal migration [Deans et al, 2017]), *Igf1* (insulin-like growth factor 1; synthesized by dopamine neurons [Pristerà et al, 2019]), *Foxf2* (known to be expressed in neural crest cells leading to pericytes [Reyahi et al, 2015]), the integrin alpha 8 (*Itga8*) known to regulate the outgrowth of neurites of sensory and motor neurons, and the calcium voltage-gated channel auxiliary subunit alpha2/delta3 (*Cacna2d3*), known to be essential for proper function of glutamatergic synapses notably on the auditory brainstem (Bracic et al, 2022). As a whole, this highlighted *Prdm8* regulome appears as a critical player for controlling neuronal differentiation and specialization, in agreement with previous reports on mouse and human differentiation systems (Ross et al, 2012; Inoue et al, 2015; Cypris et al, 2020). Furthermore, our data indicate that *Prdm8* and all other factors composing the illustrated regulome in Fig 5 are driven by the RARα binding sites but can be controlled by the RARβ and RARγ receptors in the absence of the RARα receptor.

RARγ- and RARβ-driven cell specialization programs retrieved in P19 ECs are also observed during differentiation of mESCs

To explore the relevance of the restored cell specialization capacity in P19 EC cells driven by the synergistic action of RARγ and RARβ

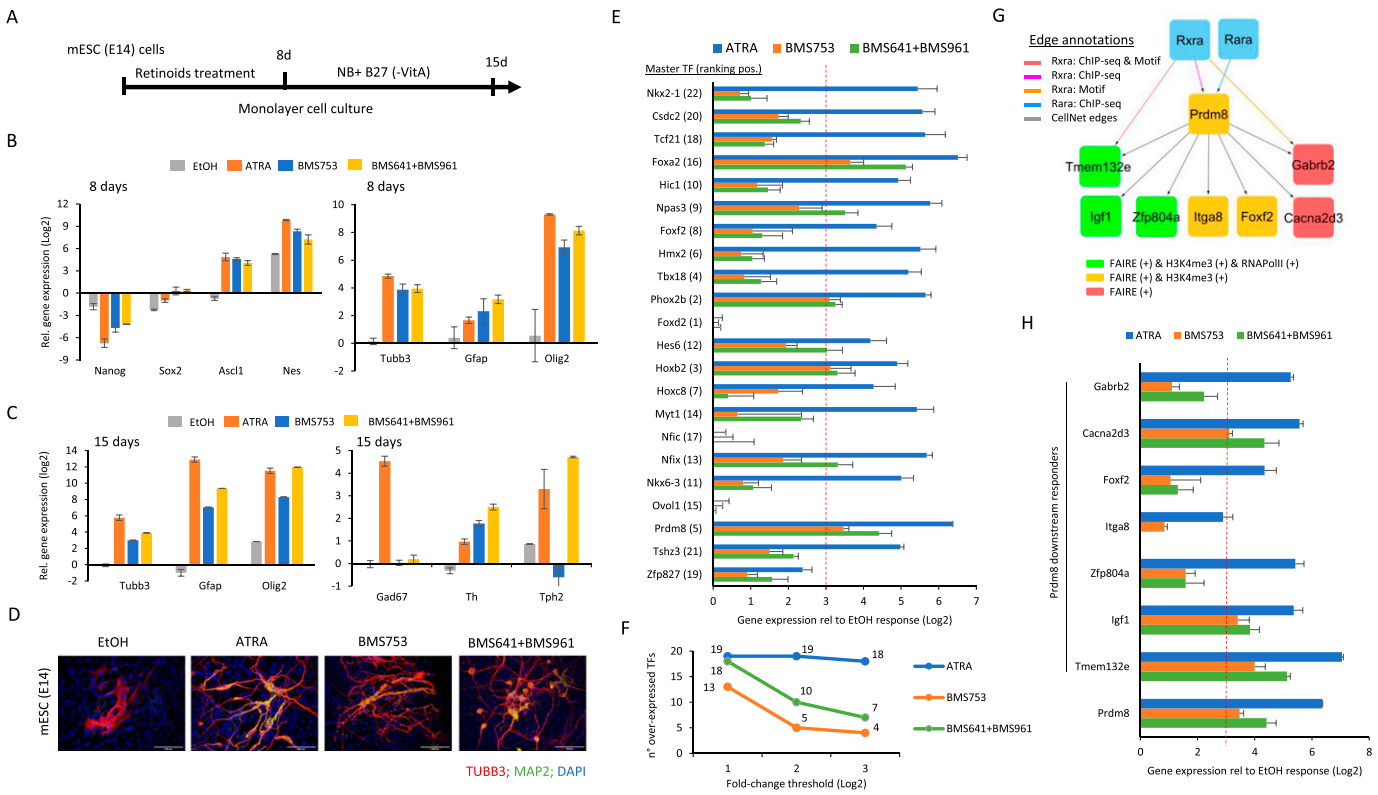


Figure 6. RARγ- and RARβ-driven cell specialization programs retrieved in P19 ECs are also reactivated during differentiation of mouse embryonic stem cells.
(A) Schematic representation of mouse ES (E14) cell differentiation assay. mES cells cultured on monolayer are exposed to retinoids during 8 d to induce cell fate commitment; then, they are cultured for seven more days on a synthetic medium (Neurobasal, NB) complemented with B27 (without vitamin A) supplements. **(B)** RT–qPCR after 8 d of differentiation, revealing the mRNA expression levels of gene markers associated with the stem cell markers (Nanog and Sox2), neuronal precursors (Ascl1, Nestin, and Tubb3), and glial cells (astrocyte-related: Gfap; oligodendrocyte-related: Olig2). **(C)** RT–qPCR after 15 d of differentiation, revealing the mRNA expression levels of gene markers associated with the neuronal marker Tubb3, the glial cell–related markers Gfap and Olig2, and the neuronal subtype markers Gad67 (GABAergic), Th (dopaminergic), and Tph2 (serotonergic). **(D)** Immunofluorescence micrograph of mES cells after 15 d of culture in the presence of either ethanol (EtOH: vehicle control), all-trans retinoic acid, the RARα agonist BMS753, or the combination of RARβ and RARγ agonists. Cells were stained for the neuronal precursor marker TUBB3 (red) and the marker for mature neurons MAP2 (green). Nuclei were stained with DAPI (blue). **(E)** RT–qPCR mRNA expression levels measured in mES cells (15 d of differentiation) corresponding to the top 22 master TFs identified in P19 cells (Fig 5). Differential gene expression is expressed relative to the expression levels observed in the presence of the ethanol control sample after 15 d of differentiation. The dashed red line demarcates a fold change threshold value of 3. **(F)** Number of overexpressed TFs under the various retinoid treatments computed at three different fold change thresholds (Log_2). **(G)** Subset of the "inhibited program" retrieved in P19 cells, revealing the gene co-regulatory network associated with Prdm8. **(H)** RT–qPCR mRNA expression levels measured in mES cells (15 d of differentiation) corresponding to the different downstream targets of Prm8 revealed in P19 cells. The dashed red line demarcates a fold change threshold value of 3.

agonists, we have extended this study to the use of WT mESCs. Specifically, we have cultured mESCs in monolayer in presence of either the pan-agonist ATRA, the RARα agonist BMS753, or the combination of the RARβ-specific ligand BMS641 and the RARγ agonist BMS961. After 8 d of treatment, retinoid treatment has been replaced by Neurobasal-complemented medium for other 7 d to promote cell specialization (Fig 6A).

RT–qPCR assays performed after 8 d of mESC treatment with the aforementioned retinoids revealed the decrease in expression of stem cell markers (*Nanog* and *Sox2*)—confirming cell differentiation commitment—and the gain in the expression for neuronal precursor markers such as *Ascl1*, Nestin (*Nes*), and *Tubb3* (Fig 6B). Furthermore, gene expression induction of *Gfap* and *Olig2* was observed, indicative of the presence of glial cells in addition to neuronal commitment at this stage of differentiation. After 15 d of treatment, gene expression levels for glial cells (*Gfap* and *Olig2*) and the neuronal precursor marker *Tubb3* appeared enhanced, and neuronal subtype markers such as *Gad67* (GABAergic), *Th* (dopaminergic), or *Tph2* (serotonergic) were also detected (Fig 6C).

Like in the case of P19 cells, in addition to the differentiation response observed in presence of ATRA or the BMS753 ligand, the synergistic treatment of mESCs with the RARβ-specific ligand BMS641 and the RARγ agonist BMS961 led to neuronal differentiation, as revealed not only by the aforementioned RT–qPCR assays, but also by immunofluorescence using the neuronal precursor marker TUBB3 and the mature neuronal marker MAP2 (Fig 6D).

Finally, to evaluate whether the synergistic action of RARγ- and RARβ-specific agonists in mESCs uses the same gene programs revealed in P19 EC cells, we have performed RT–qPCR assays targeting the gene expression of major master TFs (Fig 5). As illustrated in Fig 6E, ATRA treatment of mESCs gives rise to significant overexpression of 18 over 22 master TFs revealed in P19 EC cells (fold change [log_2] ≥ 3 rel. to stem cell state; Fig 6F). The use of the RARα agonist (BMS753) leads to a strong response of at least four of the

master players (fold change [$\log_2$] ≥ 3: *Prdm8*, *Hoxb2*, *Phox2b*, and *Foxa2*), and up to 13 master TFs when considering lower fold change levels (fold change [$\log_2$] ≥ 1). Interestingly, the combination of RARβ+γ agonists (BMS641 + BMS961) gives rise to strong significant overexpression of the four master TFs driven by the RARα agonist (BMS753) and three other players (*Nfix*, *Hes6*, and *Npas3*), reaching up to 18 TFs activated when considering less stringent fold change levels (Fig 6E and F).

Among all responsive master players during P19 and mESC (E14) differentiation driven by the combination of RARβ+γ agonists (BMS641 + BMS961), the histone lysine methyltransferase factor *Prdm8* appeared as a major hub. Indeed, the summarized "inhibited RXR/RAR" GRN revealed in P19 EC cells up to seven factors under the direct control of *Prdm8* (Fig 6G). Interestingly, during mouse ES cell differentiation most of these downstream factors are also strongly reactivated: six of them under ATRA treatment and three of them (*Cacna2d3*, *Igf1*, and *Tmem132e*) in the presence of either the RARα agonist (BMS753) or the combination of RARβ+γ agonists (BMS641 + BMS961) (log fold change ≥ 3; Fig 6H).

Overall, the use of the combination of RARβ+γ agonists (BMS641 + BMS961) for inducing neuronal differentiation and specialization in mES (E14) cells revealed the reactivation of the same gene programs initially retrieved in P19 EC cells.

Discussion

How cells respond to different signals to develop along defined cell lineages is a key open question to understand physiological cell differentiation, leading to the formation of not only organs and tissues, but also events such as in vitro cell reprogramming and even tumorigenesis. In this study, we specifically address the role of retinoids in activating major gene regulatory wires driving neuronal cell lineage and notably cell specialization. Previously, we have dissected the major retinoid-driven gene regulatory programs, leading to neuronal precursor formation, notably by evaluating the relevance of the activation of the RARα nuclear receptor in P19 cells with the synthetic agonist BMS753 (Mendoza-Parra et al, 2016a).

Although we have also shown in our previous study that activation of RARβ or RARγ nuclear receptors by their cognate BMS641 or BMS961 synthetic agonists is insufficient to promote neuronal differentiation, others reported that in long-term culture conditions, which included embryoid body formation, RARγ-specific ligand could induce the formation of GABAergic neurons, whereas RARα induced dopaminergic neurons (Podleśny-Drabiniok et al, 2017). In this study, we addressed neuronal differentiation in long-term culture conditions, but we have kept a monolayer culture strategy because it is known that cell–cell contact interactions retrieved on either two-dimensional (monolayer) or three-dimensional cultures could lead to different outcomes, as highlighted by the cellular complexity observed on cerebral organoid cultures (Lancaster et al, 2013). We have shown herein that activation of the RARβ receptor does not lead to mature neurons, nor to other specialized cells, whereas activation of RARγ nuclear receptors gives rise to lower yields of cell specialization than that observed when using the pan-agonist ATRA or the RARα-specific agonist BMS753. Surprisingly, their synergistic activation gave rise to high yields of maturation, including specialized neuronal subtypes, and to other glial cells.

Previous studies demonstrated a redundancy for the activation of certain genes by distinct RXRs/RARs, and notably on Rar-null mutant lines, suggesting that in the absence of a given RAR nuclear receptor, the remaining isotypes could compensate for such dysfunction (Roy et al, 1995; Taneja et al, 1996; Chiba et al, 1997). Similarly, a synergistic 24 h activation by combining RAR isotype agonists has been attempted with P19 embryoid bodies in a recent study, suggesting that the synergistic activation of RARα and RARβ agonists might lead to TH + dopaminergic neurons, whereas RARγ and RARβ (or RARγ and RARα) might have a preference to induce Drd2+ neuronal subtypes (Podleśny-Drabiniok et al, 2017). Altogether, these studies clearly highlight the redundancy between RAR isotypes, as is further supported by Rar-null mutant experiments illustrated here. Indeed, we clearly demonstrate that Rara KO cells present an enhanced cell specialization yield relative to the WT situation. Furthermore, we have decorticated the gene programs that are inhibited by the potential action of the unliganded RARα receptor, notably by observing their activation on the Rara-null line via the synergistic action of the RARβ and RARγ agonists (BMS641 + BMS961). Among them, we have revealed the inhibition of *Prdm8*, a member of the family of histone methyltransferases, shown to play a role in the development of brain structures, notably by its capacity to regulate the transition from multipolar to bipolar morphology of cortical neurons (reviewed in Leszczyński et al [2020]).

Finally, we have expanded this study to neuronal cell specialization in mouse ES cells, notably by revealing the reactivation of the same gene regulatory programs retrieved in P19 EC cells, strongly suggesting for a general mechanism driven by the synergistic action of the RARβ and RARγ agonists (BMS641 + BMS961).

In summary, this study provides a systems biology view of the gene programs behind the previously observed redundancy between RARs, paving the way for their potential use for directing cell specialization during nervous tissue formation.

Materials and Methods

Cell culture

P19 cells were grown in DMEM supplemented with 1 g/l glucose, 5% FCS, and 5% delipidated FCS. P19 EC cells were cultured in a monolayer on gelatin-coated culture plates (0.1%). For cell differentiation assays, ATRA was added to plates to a final concentration of 1 μM for different exposure times. For treatment with RAR subtype–specific agonists, cells were incubated with BMS961 (RARγ-specific; 0.1 μM), BMS753 (RARα-specific; 1 μM), and/or BMS641 (RARβ-specific; 0.1 μM). After 4 d of treatment with either of the aforementioned retinoids, the medium was replaced by Neurobasal medium (ref: 21103049; Thermo Fisher Scientific) supplemented with N2 (ref: 17502048; Thermo Fisher Scientific) and B27 devoid of vitamin A (ref: 12587010; Thermo Fisher Scientific) and cultured for six more days.

Embryonic stem cells (E14) were grown in DMEM supplemented with 4.5 g/l glucose and GlutaMAX-I (ref. 11594446; Thermo Fisher Scientific), 15% FBS-ES, 5 ng/ml LIF recombinant mouse protein (ref.

15870082; Thermo Fisher Scientific), 1% penicillin–streptomycin, 1% MEM-NEAA, and 0.02% β-mercaptoethanol. E14 cells were cultured in a monolayer on poly-D-lysine–coated culture plates (0.1%). For cell differentiation assays, ATRA was added to plates to a final concentration of 1 μM. For treatment with RAR subtype–specific agonists, cells were incubated with BMS753 (RARα-specific; 1 μM) or BMS641 + BMS961 (RARβ + RARγ-specific; 0.1 μM each). After 8 d of treatment with either of the aforementioned retinoids, the medium was replaced by Neurobasal medium (ref. 11570556; Thermo Fisher Scientific) and B27 devoid of vitamin A (ref. 11500446; Thermo Fisher Scientific) and cultured for seven more days.

Immunohistochemical staining

After 10 d of induced differentiation, cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences), followed by 3 × 5 min washes in PBS. Cells were permeabilized (Triton 0.1% in PBS; 15 min at room temperature) and blocked (10% heat-inactivated FCS in PBS for P19 cells; 0.1% Triton and 1% BSA in PBS for E14 cells during 1 h at room temperature). Cells were washed 3 × 5 min in permeabilization buffer, then incubated with the primary antibodies anti-β III tubulin/anti-TUBB3 (ab14545; Abcam) or anti-MAP2 (ab32454). After 1 h of incubation at room temperature (for E14 cells, primary antibody incubation was done overnight at +4°C), cells were washed 3 × 10 min with permeabilization buffer followed by incubation with a secondary antibody (donkey anti-mouse IgG [H + L] antibody Alexa 555; Invitrogen A-31570; donkey anti-rabbit IgG [H + L] antibody Alexa 488; Invitrogen A-21206) and/or DAPI (D3571; Invitrogen). After 1 h at room temperature, cells were washed for 3 × 10 min in permeabilization buffer twice with Milli-Q water and finally mounted on microscope slides.

RT–qPCR and RNA sequencing

Total RNA was extracted from P19 and E14 cells treated with either ATRA- or RAR-specific agonists, using the TRIzol RNA isolation reagent (ref: 15596026; Thermo Fisher Scientific) or RNeasy Mini Kit (ref. 74104; Qiagen). 0.5–1 μg of the extracted RNA was used for reverse transcription (HIGH CAPACITY CDNA RT; ref: 4368814; Applied Biosystems). Transcribed cDNA was diluted fivefold and used for real-time quantitative PCR (QuantiTect SYBR Green Kit; ref: 204145; Qiagen). RNA-sequencing libraries were produced with the NEBNext Ultra II RNA Library Prep Kit for Illumina (E7770). Libraries were sequenced within the French National Sequencing Center, Genoscope (150-nt pair-end sequencing; NovaSeq Illumina).

Primary bioinformatics analyses

Fastq files were qualified with the NGS-QC Generator tool (Mendoza-Parra et al, 2013, 2016b). Reads from fastq files were mapped to the mouse mm9 reference genome using Bowtie 2.1.0 under default parameters. Mapped reads were associated with known genes with featureCounts. RNA-seq analyses were done with the DESeq2 R package. t-Distributed stochastic neighbor embedding analysis was performed with the R package Rtsne. Heatmap matrix display was generated with MeV 4.9.0. Gene ontology analyses were performed with DAVID Bioinformatics Resources (https://david.ncifcrf.gov/).

Collection of gene markers associated with specialized cells

Gene markers associated with neurons (1,352 genes), astrocytes (501 genes), and oligodendrocyte precursors (OPCs: 501 genes) were collected from the supplementary material (Dataset_S02) of Voskuhl et al (2019). Gene markers associated with GABAergic (318 genes) and glutamatergic (311 genes) neuronal subtypes were collected from Table S9 of Tasic et al (2018). Gene markers for dopaminergic (Th+) (513 genes) neuronal subtypes were collected from Table S2 of Hook et al (2018). This assembled collection is available in our supplementary material (Table S1).

Promoter epigenetic status analysis and RAR/RXR enrichment

P19 epigenetic readouts assessed for the repressive mark H3K27me3 and the active mark H3K4me3, the chromatin accessibility—revealed by FAIRE-sequencing profiling—and the transcriptional response (RNAPII) after 2 d of ATRA treatment were collected from our previous published study (GSE68291 [Mendoza-Parra et al, 2016a]). Normalized enrichment levels at gene promoter regions, relative to EtOH control profiles, were used for predicting their epigenetic combinatorial status. Enrichment heatmaps and mean density plots for FAIRE readouts at gene promoter regions (±500 bp) were obtained with seqMINER 1.3.4.

FAIRE site motif analysis has been performed with the MEME Suite 5.4.1. RXR/RAR enrichment on FAIRE sites was inferred by comparing them with > 40,000 mouse ChIP-seq binding sites collected from the public domain, as part of our NGS-QC database (https://ngsqc.org/). Among all mouse ChIP-seq collected data, 71 ChIP-seq public profiles correspond to RXR or RAR TFs (Mendoza-Parra et al, 2013, 2016b; Blum et al, 2020).

GRN reconstruction

Temporal transcriptomes issued from ATRA treatment were integrated with a collection of TF-TG relationships (CellNet [Cahan et al, 2014]) with the help of our previously developed Cytoscape App, TETRAMER (Cholley et al, 2018). TETRAMER has been also used for identifying the top 22 master TFs (master regulatory index > 70%). Gene co-regulatory wire visualization has been performed with Cytoscape 3.8.2.

Data Availability

All RNA-sequencing datasets generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE204816.

Supplementary Information

Supplementary Information is available at https://doi.org/10.26508/lsa.202201627

Acknowledgements

We thank all the members of our laboratory and the Genoscope sequencing platform for discussion. This work was supported by the institutional bodies CEA, CNRS, and Université d'Evry-Val d'Essonne. E Mathieux was supported by Genopole Thematic Incentive Actions funding (ATIGE-2017); A Koshy by the "Fondation pour la Recherche Medicale" (FRM; funding ALZ-201912009904); F Stüder by the funding 2OI9-L22 from the Institut National du Cancer (INCa); and A Bramoulle by the funding 2020-181 (INCa).

Author Contributions

A Koshy: formal analysis, investigation, and methodology.
E Mathieux: formal analysis, investigation, and methodology.
F Stüder: data curation, software, and formal analysis.
A Bramoulle: resources and methodology.
M Lieb: resources and methodology.
BM Colombo: writing—review and editing.
H Gronemeyer: supervision, funding acquisition, and writing—review and editing.
MA Mendoza-Parra: conceptualization, formal analysis, supervision, funding acquisition, and writing—original draft.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Álvarez R, Vaz B, Gronemeyer H, de Lera ÁR (2014) Functions, therapeutic applications, and synthesis of retinoids and carotenoids. *Chem Rev* 114: 1–125. doi:10.1021/cr400126u

Barki M, Xue H (2022) GABRB2, a key player in neuropsychiatric disorders and beyond. *Gene* 809: 146021. doi:10.1016/j.gene.2021.146021

Blum M, Cholley P-E, Malysheva V, Nicaise S, Moehlin J, Gronemeyer H, Mendoza-Parra MA (2020) A comprehensive resource for retrieving, visualizing, and integrating functional genomics data. *Life Sci Alliance* 3: e201900546. doi:10.26508/lsa.201900546

Bracic G, Hegmann K, Engel J, Kurt S (2022) Impaired subcortical processing of amplitude-modulated tones in mice deficient for cacna2d3, a risk gene for autism spectrum disorders in humans. *eNeuro* 9: ENEURO.0118-22.2022. doi:10.1523/eneuro.0118-22.2022

Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ (2014) CellNet: Network biology applied to stem cell engineering. *Cell* 158: 903–915. doi:10.1016/j.cell.2014.07.020

Caubit X, Gubellini P, Andrieux J, Roubertoux PL, Metwaly M, Jacq B, Fatmi A, Had-Aissouni L, Kwan KY, Salin P, et al (2016) TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons. *Nat Genet* 48: 1359–1369. doi:10.1038/ng.3681

Chiba H, Clifford J, Metzger D, Chambon P (1997) Specific and redundant functions of retinoid X receptor/retinoic acid receptor heterodimers in differentiation, proliferation, and apoptosis of F9 embryonal carcinoma cells. *J Cell Biol* 139: 735–747. doi:10.1083/jcb.139.3.735

Cholley P-E, Moehlin J, Rohmer A, Zilliox V, Nicaise S, Gronemeyer H, Mendoza-Parra MA (2018) Modeling gene-regulatory networks to describe cell fate transitions and predict master regulators. *NPJ Syst Biol Appl* 4: 29. doi:10.1038/s41540-018-0066-z

Cypris O, Eipel M, Franzen J, Rösseler C, Tharmapalan V, Kuo C-C, Vieri M, Nikolić M, Kirschner M, Brümmendorf TH, et al (2020) PRDM8 reveals aberrant DNA methylation in aging syndromes and is relevant for hematopoietic and neuronal differentiation. *Clin Epigenetics* 12: 125. doi:10.1186/s13148-020-00914-5

de Lera AR, Bourguet W, Altucci L, Gronemeyer H (2007) Design of selective nuclear receptor modulators: RAR and RXR as a case study. *Nat Rev Drug Discov* 6: 811–820. doi:10.1038/nrd2398

Deans PM, Raval P, Sellers KJ, Gatford NJF, Halai S, Duarte RRR, Shum C, Warre-Cornish K, Kaplun VE, Cocks G, et al (2017) Psychosis risk candidate ZNF804A localizes to synapses and regulates neurite formation and dendritic spine structure. *Biol Psychiatry* 82: 49–61. doi:10.1016/j.biopsych.2016.08.038

Gratton M-O, Torban E, Jasmin SB, Theriault FM, German MS, Stifani S (2003) Hes6 promotes cortical neurogenesis and inhibits Hes1 transcription repression activity by multiple mechanisms. *Mol Cell Biol* 23: 6922–6935. doi:10.1128/mcb.23.19.6922-6935.2003

Hirabayashi Y, Gotoh Y (2010) Epigenetic control of neural precursor cell fate during development. *Nat Rev Neurosci* 11: 377–388. doi:10.1038/nrn2810

Hollensen AK, Thomsen HS, Lloret-Llinares M, Kamstrup AB, Jensen JM, Luckmann M, Birkmose N, Palmfeldt J, Jensen TH, Hansen TB, et al (2020) circZNF827 nucleates a transcription inhibitory complex to balance neuronal differentiation. *Elife* 9: e58478. doi:10.7554/elife.58478

Hook PW, McClymont SA, Cannon GH, Law WD, Morton AJ, Goff LA, McCallion AS (2018) Single-cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for sporadic Parkinson disease. *Am J Hum Genet* 102: 427–446. doi:10.1016/j.ajhg.2018.02.001

Huang W, Huang T, Liu Y, Fu J, Wei X, Liu D, Ma W, Gu H, Yuan Z (2021) Nuclear factor I-C disrupts cellular homeostasis between autophagy and apoptosis via miR-200b-Ambra1 in neural tube defects. *Cell Death Dis* 13: 17. doi:10.1038/s41419-021-04473-2

Inoue M, Iwai R, Yamanishi E, Yamagata K, Komabayashi-Suzuki M, Honda A, Komai T, Miyachi H, Kitano S, Watanabe C, et al (2015) Deletion of Prdm8 impairs development of upper-layer neocortical neurons. *Genes Cells* 20: 758–770. doi:10.1111/gtc.12274

Lancaster MA, Renner M, Martin C-A, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA (2013) Cerebral organoids model human brain development and microcephaly. *Nature* 501: 373–379. doi:10.1038/nature12517

Leszczyński P, Śmiech M, Parvanov E, Watanabe C, Mizutani K-I, Taniguchi H (2020) Emerging roles of PRDM factors in stem cells and neuronal system: Cofactor dependent regulation of PRDM3/16 and FOG1/2 (novel PRDM factors). *Cells* 9: 2603. doi:10.3390/cells9122603

Matuzelski E, Bunt J, Harkins D, Lim JWC, Gronostajski RM, Richards LJ, Harris L, Piper M (2017) Transcriptional regulation of Nfix by NFIB drives astrocytic maturation within the developing spinal cord. *Dev Biol* 432: 286–297. doi:10.1016/j.ydbio.2017.10.019

Mendoza-Parra M-A, Van Gool W, Mohamed Saleem MA, Ceschin DG, Gronemeyer H (2013) A quality control system for profiles obtained by ChIP sequencing. *Nucleic Acids Res* 41: e196. doi:10.1093/nar/gkt829

Mendoza-Parra M-A, Malysheva V, Mohamed Saleem MA, Lieb M, Godel A, Gronemeyer H (2016a) Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis. *Genome Res* 26: 1505–1519. doi:10.1101/gr.208926.116

Mendoza-Parra MA, Saleem M-AM, Blum M, Cholley P-E, Gronemeyer H (2016b) NGS-QC generator: A quality control system for ChIP-seq and related deep sequencing-generated datasets. In *Statistical Genomics: Methods and Protocols*. Mathé E, Davis S (eds.). pp 243–265. New York, NY: Springer.

Miller FD, Gauthier AS (2007) Timing is everything: Making neurons versus glia in the developing cortex. *Neuron* 54: 357–369. doi:10.1016/j.neuron.2007.04.019

Minocha S, Valloton D, Arsenijevic Y, Cardinaux J-R, Guidi R, Hornung J-P, Lebrand C (2017) Nkx2.1 regulates the generation of telencephalic astrocytes during embryonic development. *Sci Rep* 7: 43093. doi:10.1038/srep43093

Monzo HJ, Park TIH, Montgomery JM, Faull RLM, Dragunow M, Curtis MA (2012) A method for generating high-yield enriched neuronal cultures from P19 embryonal carcinoma cells. *J Neurosci Methods* 204: 87–103. doi:10.1016/j.jneumeth.2011.11.008

Niederreither K, Dollé P (2008) Retinoic acid in development: Towards an integrated view. *Nat Rev Genet* 9: 541–553. doi:10.1038/nrg2340

Podleśny-Drabiniok A, Sobska J, de Lera AR, Gołembiowska K, Kamińska K, Dollé P, Cebrat M, Krężel W (2017) Distinct retinoic acid receptor (RAR) isotypes control differentiation of embryonal carcinoma cells to dopaminergic or striatopallidal medium spiny neurons. *Sci Rep* 7: 13671. doi:10.1038/s41598-017-13826-x

Pristerà A, Blomeley C, Lopes E, Threlfell S, Merlini E, Burdakov D, Cragg S, Guillemot F, Ang S-L (2019) Dopamine neuron-derived IGF-1 controls dopamine neuron firing, skill learning, and exploration. *Proc Natl Acad Sci U S A* 116: 3817–3826. doi:10.1073/pnas.1806820116

Reyahi A, Nik AM, Ghiami M, Gritli-Linde A, Pontén F, Johansson BR, Carlsson P (2015) Foxf2 is required for brain pericyte differentiation and development and maintenance of the blood-brain barrier. *Dev Cell* 34: 19–32. doi:10.1016/j.devcel.2015.05.008

Rood BR, Leprince D (2013) Deciphering HIC1 control pathways to reveal new avenues in cancer therapeutics. *Expert Opin Ther Targets* 17: 811–827. doi:10.1517/14728222.2013.788152

Ross SE, McCord AE, Jung C, Atan D, Mok SI, Hemberg M, Kim T-K, Salogiannis J, Hu L, Cohen S, et al (2012) Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron* 73: 292–303. doi:10.1016/j.neuron.2011.09.035

Roy B, Taneja R, Chambon P (1995) Synergistic activation of retinoic acid (RA)-responsive genes and induction of embryonal carcinoma cell differentiation by an RA receptor alpha (RAR alpha)-RAR beta-or RAR gamma-selective ligand in combination with a retinoid X receptor-specific ligand. *Mol Cell Biol* 15: 6481–6487. doi:10.1128/mcb.15.12.6481

Sirbu IO, Duester G (2006) Retinoic-acid signalling in node ectoderm and posterior neural plate directs left–right patterning of somitic mesoderm. *Nat Cell Biol* 8: 271–277. doi:10.1038/ncb1374

Soprano DR, Teets BW, Soprano KJ (2007) Role of retinoic acid in the differentiation of embryonal carcinoma and embryonic stem cells. *Vitam Horm* 75: 69–95. doi:10.1016/S0083-6729(06)75003-8

Taneja R, Bouillet P, Boylan JF, Gaub MP, Roy B, Gudas LJ, Chambon P (1995) Reexpression of retinoic acid receptor (RAR) gamma or overexpression of RAR alpha or RAR beta in RAR gamma-null F9 cells reveals a partial functional redundancy between the three RAR types. *Proc Natl Acad Sci U S A* 92: 7854–7858. doi:10.1073/pnas.92.17.7854

Taneja R, Roy B, Plassat JL, Zusi CF, Ostrowski J, Reczek PR, Chambon P (1996) Cell-type and promoter-context dependent retinoic acid receptor (RAR) redundancies for RAR beta 2 and Hoxa-1 activation in F9 and P19 cells can be artefactually generated by gene knockouts. *Proc Natl Acad Sci U S A* 93: 6197–6202. doi:10.1073/pnas.93.12.6197

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, et al (2018) Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563: 72–78. doi:10.1038/s41586-018-0654-5

Voskuhl RR, Itoh N, Tassoni A, Matsukawa MA, Ren E, Tse V, Jang E, Suen TT, Itoh Y (2019) Gene expression in oligodendrocytes during remyelination reveals cholesterol homeostasis as a therapeutic target in multiple sclerosis. *Proc Natl Acad Sci U S A* 116: 10130–10139. doi:10.1073/pnas.1821306116

Conclusion

We demonstrate that P19 neural cell differentiation is recovered by the combined acitivity of RARβ and RARγ activity and are able to retrieve a gene regulatory network (GRN) that displays top transcription factors controlling the network and key central players that influence downstream TFs.

To make a contrast between what we have published and what has been determined in Podleśny-Drabiniok et al, 2017 [79], both studies provide proofs for the role of individual receptors distinctly influencing terminal cell fate but there are observations that do not align with each other. Notably, in contrary to our findings, cells treated with a RARγ agonist (CD666) were most successful in generating GABAergic neurons ie. 77%, and BMS641 (RARβ agonist), the least efficient ie. 28%. More specifically, they demonstrate that the combination of RARβ and RARγ agonists do not yield a different outcome than when treating the cells with only RARγ agonist.

But there are several variances in the methodologies that can explain our discordant findings. For instance, different RARγ selective agonists were used in these studies- BMS961 versus CD666- and a potential cross-activation of the RARβ program by the agonist CD666 remains plausible, as discussed by Million et al, 2001 [131]. Furthermore, the experimental setup used by us follows monolayer differentiation conditions, while Podleśny-Drabiniok and team differentiated P19 EC cells in embryoid bodies (EBs) during exposure to the synthetic agonists prior to seeding in monolayer conditions for maturation. The main takeaway from our study is the synergistic activity of RARγ and RARβ in recovering the neuronal differentiation potential that is not possible when each of these receptors are activated individually, while in the case of Podleśny-Drabiniok, RARγ activation leads to a striatopallidal-like neuronal population.

Comparison of RARα versus RARβ + RARγ activation in neuronal cell specialization

Following the work we published in the previous paper [132], there was evidence of a synergistic effect of RARβ and RARγ being able to recover the cell specialization process in stem cells particularly when RARα expression was abrogated. This prompted further investigation at the single cell resolution of the differences seen between BMS753 versus BMS641+961 induced cell specialization in MESC cells in a neuronal differentiation capacity during early, middle and late timepoints.

The MESC cells (Carlin) were differentiated (Materials and Methods: 2D Neuronal differentiation protocol) for a period of 4, 8 and 16 days according to the schematic (Figure 7). Cells collected at 4 days (early timepoint) were exposed to the ligands for 4 days; cells collected at 8 days (middle timepoint) were exposed to the ligands for 8 days without any refreshment of media; and cells collected at 16 days (late timepoint) were exposed to ligands for 8 days before switching to a Neurobasal media with a B27 supplement (without vitamin A) for 8 additional days with a half medium change every alternate day (Figure 7).

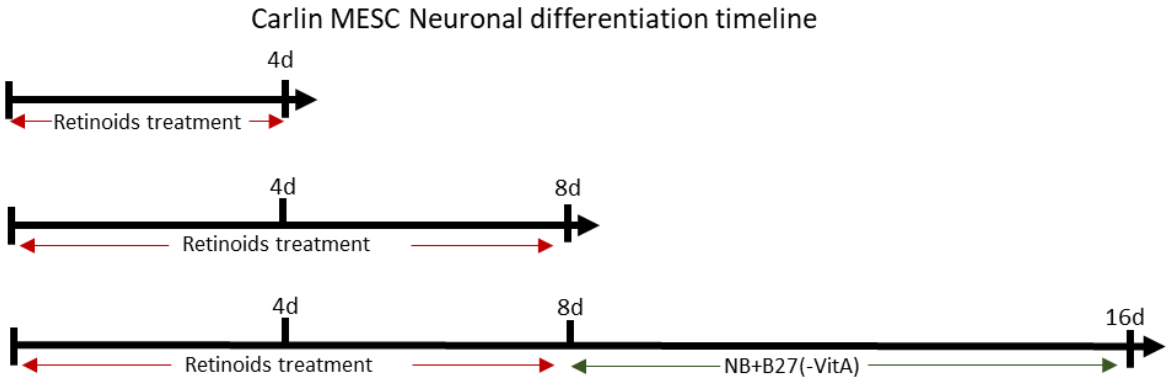


Figure 7. 2D Neuronal differentiation timeline in MESC Carlin cells

collected at 3 different timepoints (Day 4, 8, 16) with the duration of retinoids treatment and post retinoid treatment. NBM: Neurobasal medium; B27(-VitA): B27 supplement without Vitamin A.

Note: Carlin mouse embryonic stem cell line

The Carlin mouse embryonic stem cell line was developed by Sarah Bowling and team [133] as a tool to trace lineage and interrogate transcriptome profiles simultaneously. The system relies on a Doxycycline-inducible Cas9 activation that binds any of the 10 guide RNAs that can target 1 of 10 target sites in a molecular barcode/array known as CARLIN (CRISPR Array Repair LINeage tracing) (Supplementary Figure 1). This array is inherited and mutable based on the doxycycline saturation and frequency of pulsing. The mutated array can be sequenced and compared to trace and determine cells at the beginning of that lineage.

This cell line was kindly provided by the team of Sarah Bowling, but the CARLIN array has not been exploited for the purposes of this thesis, notably due to problems in reading the CARLIN cassette in an accurate manner; an aspect that is in contrast to what has been described in their original article.

Brightfield imaging of the differentiated cells demonstrate that at 4 days, there is no morphological distinction between the different ligands used (Figure 8). At day 8, there are neural precursors and/or early neurons with short neurite structures visible more prominently in ATRA treated cells (positive control) versus BMS753- and BMS641+961-treated cells (Figure 8). At day 16, Ethanol treated cells (negative control) have no visible neuronal cells, while ATRA, BMS753- and BMS641+961-treated cells all show neurons with small condensed cell bodies and extending dendritic structures and axonal filaments (Figure 8).

This expression of neuronal markers can be observed in the immunocytochemistry analysis of day 8 and day 16 differentiated cells (Figure 9). At day 8, cells exposed to ATRA, BMS753 and BMS641+961 ligands have Tubb3 marker (immature neurons) but not Map2 (mature neurons). By day 16 however, the neurons in these cultures have reached maturity and express Map2 (Figure 9).

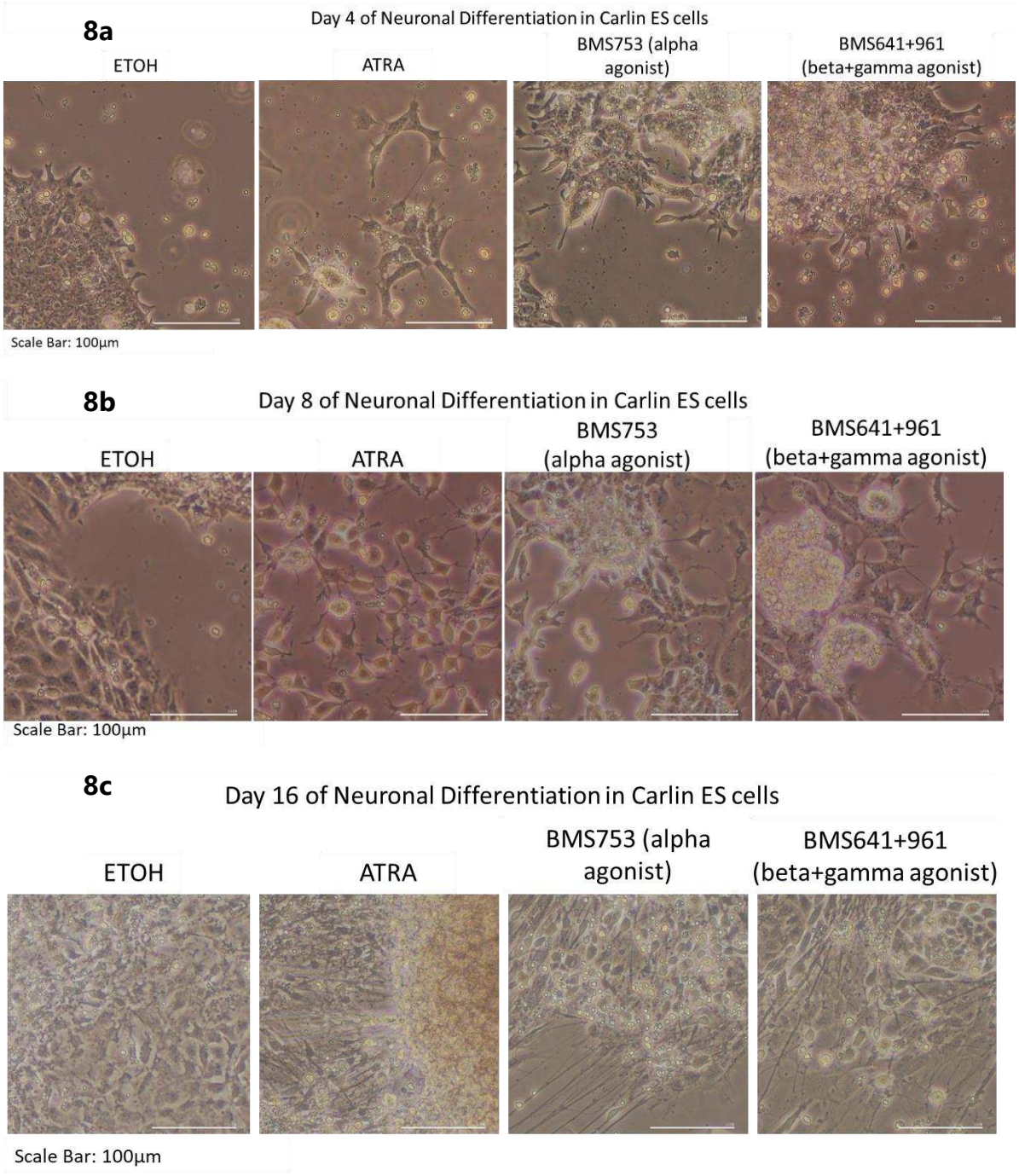


Figure 8. Brightfield images of MESC Carlin cells

at a) Day 4, b) Day 8, and c) Day 16 of Neuronal differentiation and treated with ligands (ETOH: ethanol, ATRA, BMS753: RARα agonist, BMS641+961: RARβ + RARγ agonist).

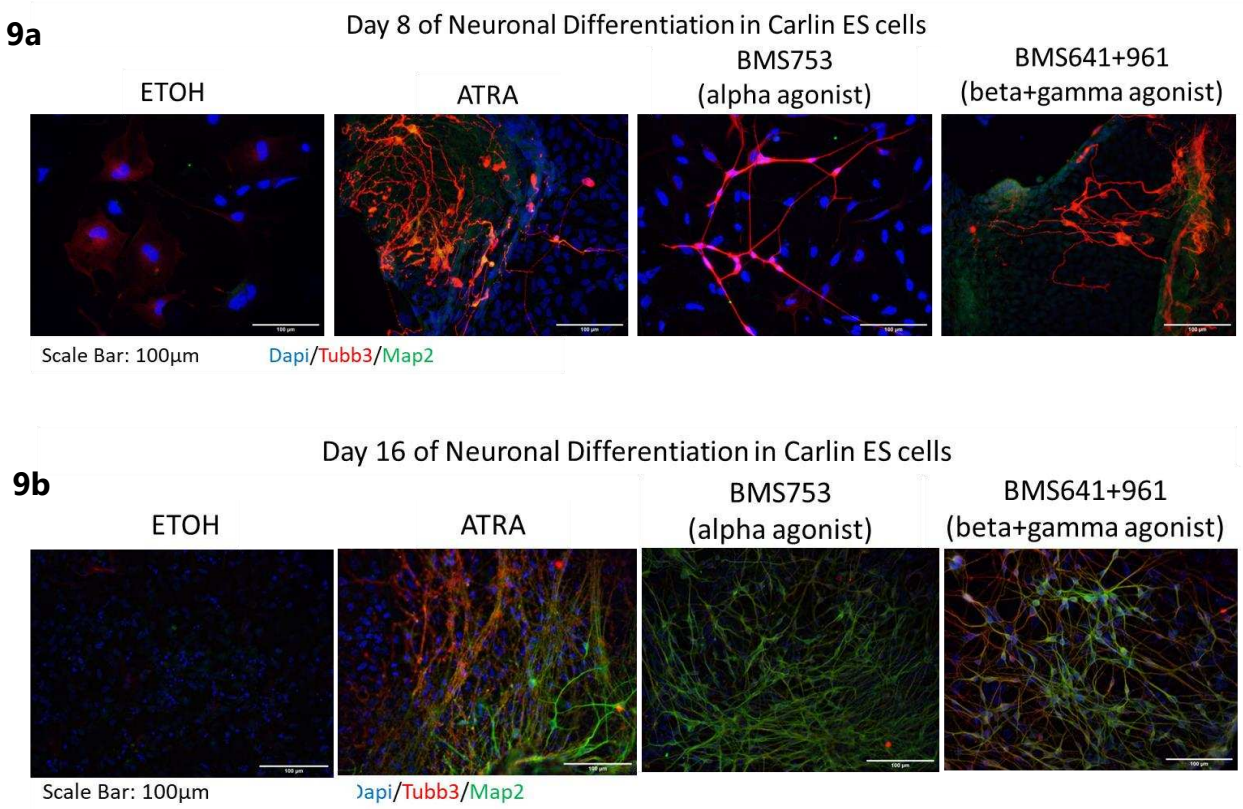


Figure 9. Immunocytochemistry analysis of neuronally differentiated MESC Carlin cells

at a) Day 8 and b) Day 16 with Dapi (Nuclei), Tubb3 (Immature neurons) and Map2 (Mature neurons).

Gene expression analysis at mid and late timepoint of differentiation reveals that at day 8, neurogenesis markers (Ascl1, Nes) are strongly upregulated at this stage and there is a an early presence of Astrocytes (Gfap) and Oligodendrocyte precursor (Olig2) markers appear earlier for BMS753 and BMS641+961 conditions (Figure 10). The cells also express mature neurons markers (Gad67, Glut1, Th, Tph2) although they are not visible in the immunocytochemistry assay (Map2), likely since the protein expression may have not yet been at detectable levels. At day 16, neurogenesis and immature neuron markers remain the same or have increased expression, pointing to expanding neural progenitor population (Figure 10). Gfap positive cells also emerge in these differentiated cells, as well as the expression of mature neuronal subtypes.

Comparing BMS753 versus BMS641+961 treated conditions, at day 8, Ascl1 and Gad67 are more strongly expressed in BMS641+961 treated cells. Ascl1 positive cells are known to be precursor markers for Gad67 positive neuronal cells. At day 16, BMS753 treated cells display a

high expression for Dopaminergic neurons (Th). These results are in line with the findings of Podleśny-Drabiniok et al, 2017.

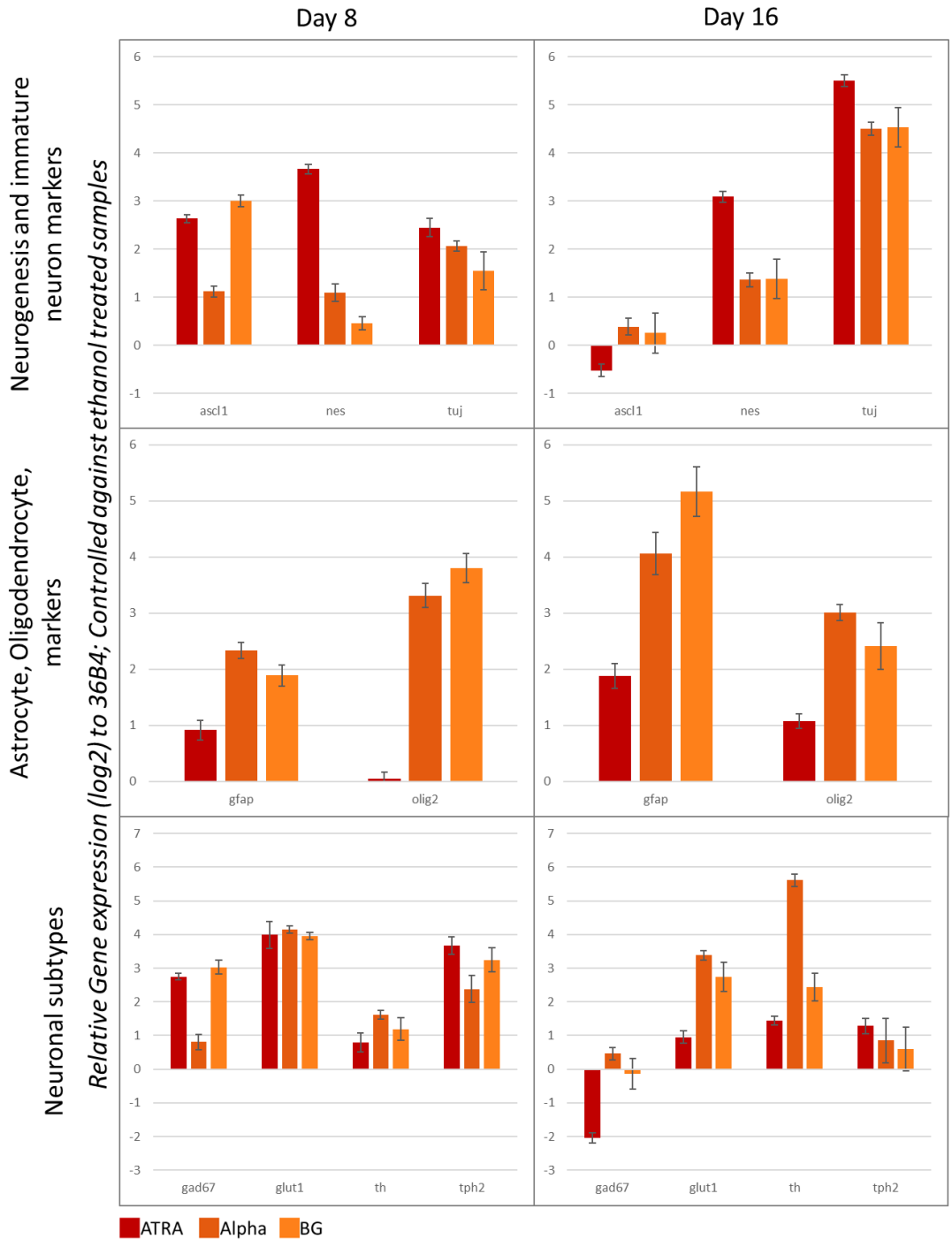


Figure 10. Gene expression analysis of neuronally differentiated MESC Carlin cells

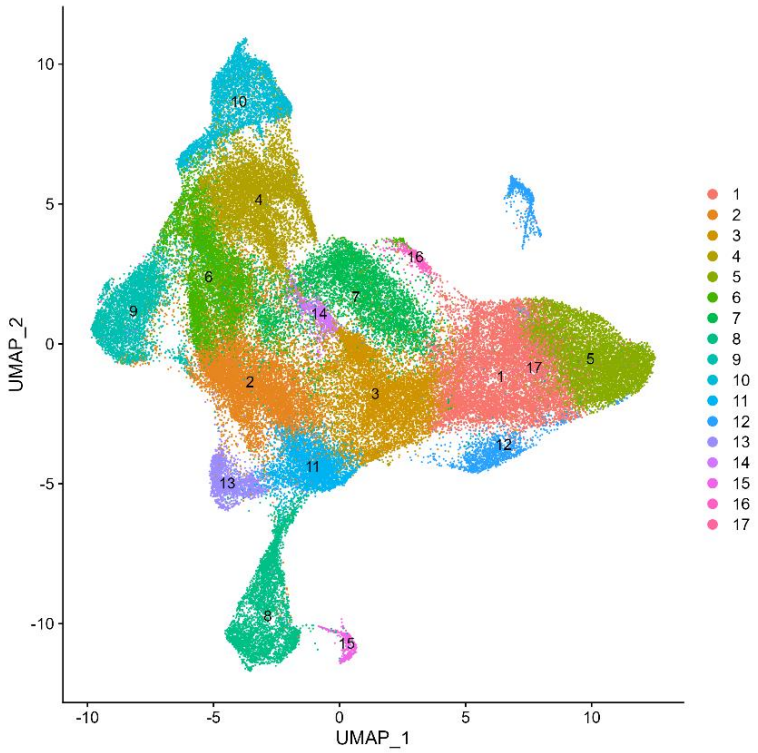
at Day 8 and Day 16 treated with ligands (ATRA, alpha: BMS753 (RARα agonist), BG: BMS641+961 (RARβ + RARγ agonist)). Housekeeping gene: 36B4; Neurogenesis: Ascl1, Nes; Immature neurons: Tuj; Astrocytes: Gfap; Oligodendrocyte precursors: Olig2; Neuronal subtypes: Gad67, Glut1, Th, Tph2.

Single cell RNA sequencing of 2D neuronally differentiated MESC

Following the neuronal differentiation assays, there were 9 different samples – Day 4, Day 8, and Day 16 timepoints with the cells having being treated with ATRA, BMS753, or BMS641+961. The library was prepared from the cells using the 10X Genomics Chromium Single cell 3′ Reagents kit v3 (Materials and Methods: Single cell RNA sequencing). The data was normalized, filtered and aligned as mentioned in Materials and Methods: Single cell RNA sequencing (Supplementary Figure 2).

In Figure 11a we have a UMAP representation of all the 9 samples and a visualization individually corresponding to each condition (Supplementary Figure 3). Each point represents a cell and they are plotted based on the similarities of the transcriptomes of each cell (Figure 11a). These cells were clustered and annotated in Seurat, to reveal 17 distinct clusters (Figure 11b).

11a



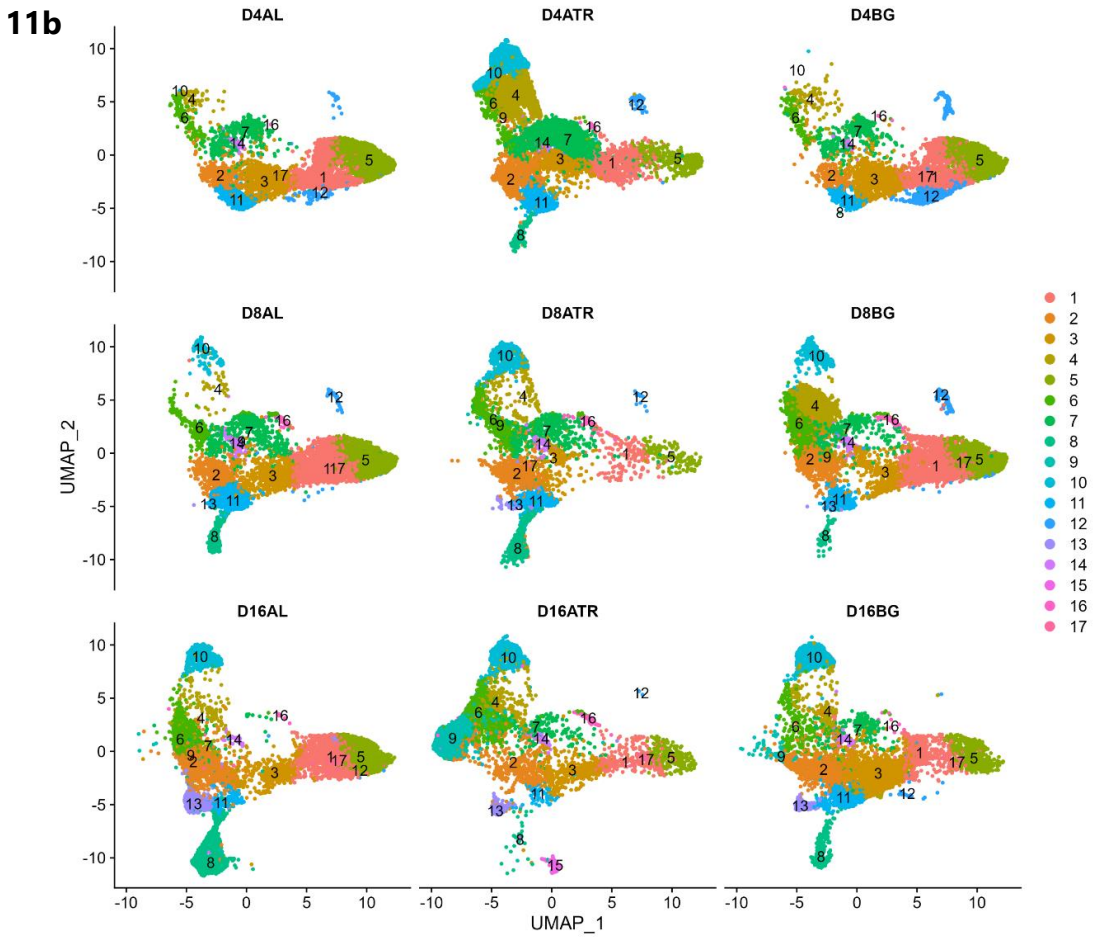


Figure 11. UMAP projection of the 9 different treated conditions of neuronally differentiated cells clustered and annotated numerically

presented as a a) compiled representation and b) individual mapping for each condition.

A table format of the number of cells present in each cluster in each condition can be seen in Figure 12a, and these have been converted to percentages of cells per cluster (Figure 12b). In the Day 4 Beta + Gamma condition, cluster 12 is strongly expressed. At Day 16, Alpha agonist treated cells show a strong expression for cluster 2, 5, 6, 8, 13, 17, while Beta + Gamma treated cells show a strong expression for cluster 3 and 11.

12a

Cluster	D4AL	D4ATR	D4BG	D8AL	D8ATR	D8BG	D16AL	D16ATR	D16BG
1	1475	780	1357	2112	174	1691	1462	273	822
2	272	1099	179	539	797	1176	3033	621	1392
3	818	1631	916	607	236	482	455	438	2603
4	68	3998	117	49	220	2739	233	374	274
5	1203	380	1041	1829	133	1481	1204	224	510
6	135	333	127	200	503	2662	1972	1037	225
7	470	3428	391	519	500	338	64	207	231
8	0	163	1	316	849	71	2787	53	319
9	0	2	0	1	11	1	49	4229	163
10	1	1570	2	152	613	357	518	606	606
11	423	496	283	1119	334	275	173	76	493
12	113	54	1293	69	11	306	3	2	24
13	0	0	0	3	37	4	1453	186	172
14	60	96	19	158	86	155	39	116	161
15	0	0	0	0	0	0	0	405	0
16	1	12	4	43	49	83	8	161	10
17	1	0	2	4	1	4	184	27	6

12b

Cluster	D4AL	D4ATR	D4BG	D8AL	D8ATR	D8BG	D16AL	D16ATR	D16BG
1	14.54	7.69	13.37	20.82	1.71	16.67	14.41	2.69	8.10
2	2.99	12.07	1.97	5.92	8.75	12.91	33.30	6.82	15.28
3	9.99	19.92	11.19	7.42	2.88	5.89	5.56	5.35	31.80
4	0.84	49.53	1.45	0.61	2.73	33.93	2.89	4.63	3.39
5	15.03	4.75	13.00	22.85	1.66	18.50	15.04	2.80	6.37
6	1.88	4.63	1.77	2.78	6.99	37.00	27.41	14.41	3.13
7	7.64	55.76	6.36	8.44	8.13	5.50	1.04	3.37	3.76
8	0.00	3.58	0.02	6.93	18.62	1.56	61.13	1.16	7.00
9	0.00	0.04	0.00	0.02	0.25	0.02	1.10	94.91	3.66
10	0.02	35.48	0.05	3.44	13.85	8.07	11.71	13.69	13.69
11	11.52	13.51	7.71	30.47	9.10	7.49	4.71	2.07	13.43
12	6.03	2.88	68.96	3.68	0.59	16.32	0.16	0.11	1.28
13	0.00	0.00	0.00	0.16	1.99	0.22	78.33	10.03	9.27
14	6.74	10.79	2.13	17.75	9.66	17.42	4.38	13.03	18.09
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00
16	0.27	3.23	1.08	11.59	13.21	22.37	2.16	43.40	2.70
17	0.44	0.00	0.87	1.75	0.44	1.75	80.35	11.79	2.62

Figure 12. Counts of cell per cluster in each single cell experimental condition

a) Table representing number of cells attributed to each condition per cluster; b) Table representing percentage of cells per cluster in all the conditions and a heatmap scale highlighting higher percentages highest to lowest.

Each of the conditions were interrogated using Loupe Browser (10x Genomics Loupe Browser 4.1.0) to visualize the gene expression in each dataset based on cell marker lists. The list of genes for each marker has been listed in Supplementary Figure 4.

At day 4 and day 16, we see the presence of endoderm and mesoderm lineage in the cell population, implying that we have multiple differentiation pathways in these cells, irrelevant of the ligand they have been exposed to (Figure 13). This is similar to what is seen in Mendoza-Parra et al, 2016, where RA-induced MESC differentiation resulted in both endodermal and ectodermal differentiation programs [31]. Unsurprisingly at day 16, these clusters inhabit different regions of the UMAP projection map. Interrogating stem cell versus neural progenitor marker profiles shows an overlap in expression at day 4, but we can also see that stem cell expression is stronger in the right half of the UMAP projection comprising clusters 1, 5, 17, and 12, while Neural progenitors occupies regions on the left side of the UMAP making up clusters 2, 3, 4, 6, 7, and 11 located separately on the UMAP (Figure 14a). This expression is more distinct as well as dense by day 16; particularly, stem cell features heavily occupy cluster 5 and cluster 13 in BMS753 treated cells, while Neural progenitors has a presence in clusters 2, 3, 6, 9, and 13 (Figure 14b).

The expression profile for Neurons and oligodendrocyte precursors (OPCs) at day 4 is indistinguishable, and for astrocytes weakly existent (Figure 15a), and at day 16, cluster 8 heavily profiles neuronal markers and cluster 13 for astrocytes while OPCs are more present in cluster 5, 3, 11 (Figure 15b). Finally, at the early differentiation timepoint we don't see a presence of mature neuronal markers, but at day 16, we find clusters 8, 13, and 3 being highlighted for expression of GABAergic and glutamatergic neurons primarily (Figure 16).

The number of cells present in cluster 8 at day 16 BMS753 treatment is eight times more than those present in BMS641+961 treatment (Figure 12a). In fact, cluster 3 and cluster 11, in day 16 BMS641+961 treatment, has five and three times a greater number of cells respectively, than BMS753 treatment. And both of these clusters in day 16 BMS641+961 treatment capture the signature for glutamatergic neurons along with cluster 8.

Another observation is that by day 16, in both treatments, we can see the robust appearance of new clusters such as cluster 10 that seems to have an endodermal signature (Figure 11b, Figure 13b), possibly one that can also express mature endoderm markers, cluster

8 with a neuronal signature and more specifically GABAergic and Glutamatergic neurons, and cluster 13 that has a mixed profile showing features for stem cells, progenitors, glutamatergic neurons suggesting an intermediate transition state for this cluster (Figure 11b, Figure 14b, Figure 16b).

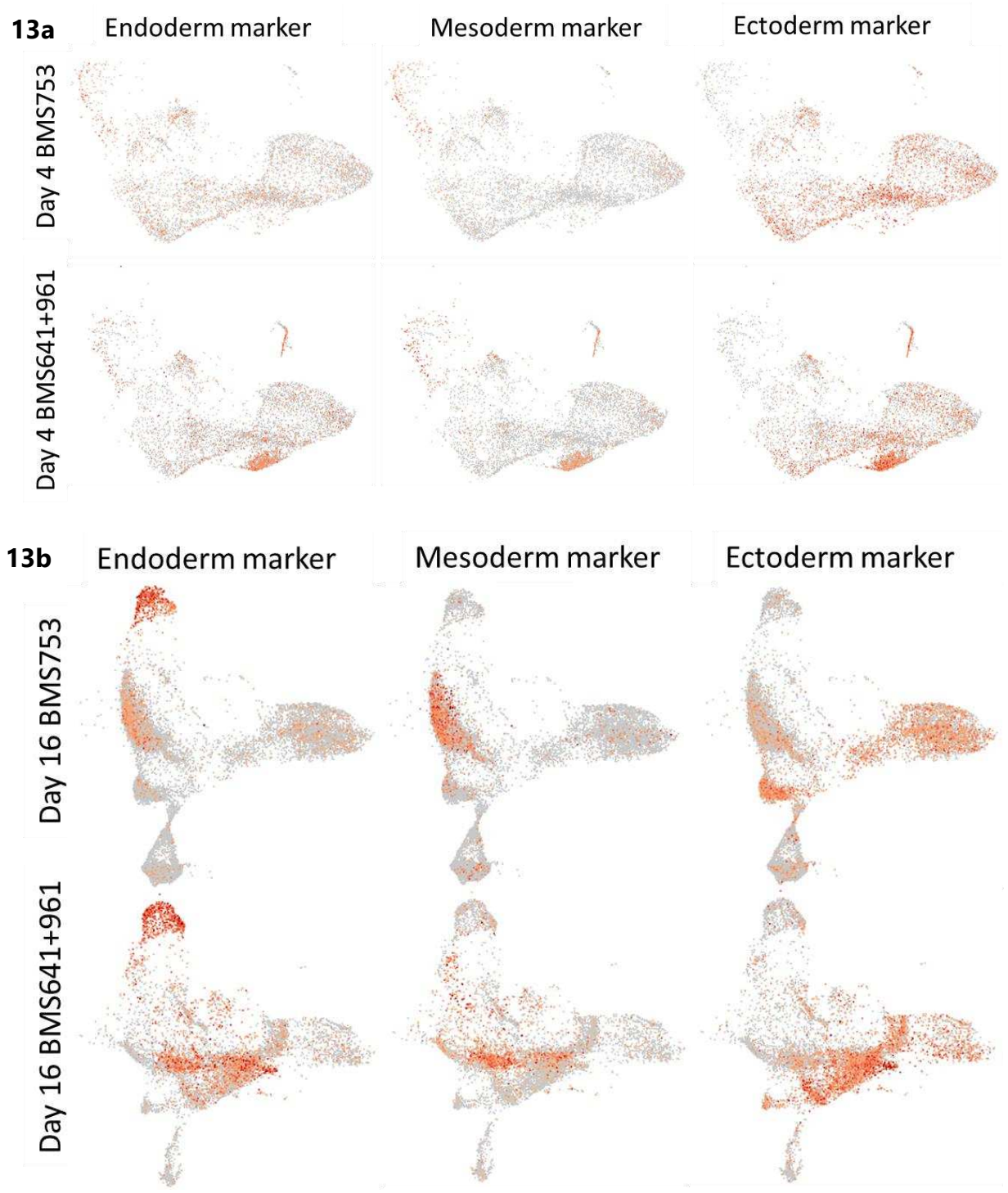


Figure 13. Gene expression distribution for Endoderm, Mesoderm, and Ectoderm markers

for BMS753 treated and BMS641+961 treated MESC Carlin cells at a) Day 4 and b) Day 16. Endoderm Markers (GATA6, GATA4, EOMES, SOX7, SOX17, FOXA2), Mesoderm Markers (MESP1, WNT3A, BMP4, FOXF1), Ectoderm Markers (PAX6, OTX2, SOX1).

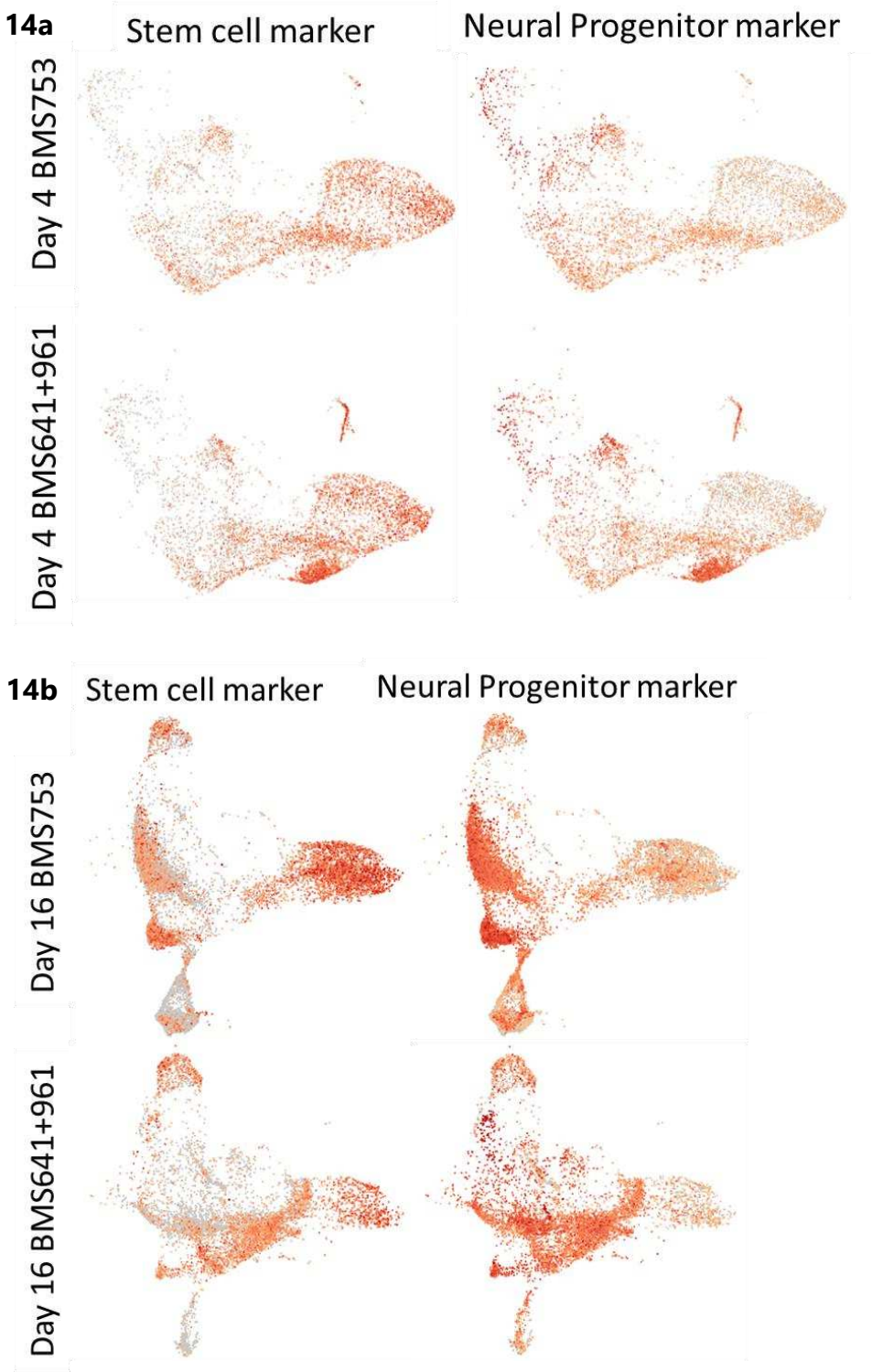


Figure 14. Gene expression for Stem cell and Neural Progenitor markers

for BMS753 treated and BMS641+961 treated MESC Carlin cells at a) Day 4 and b) Day 16. Stem Cell Markers (NANOG, SOX2, KLF4), Neural Progenitor Markers (PAX6, NES, NEUROD1, VIM, ASCL1, NEUROG2).

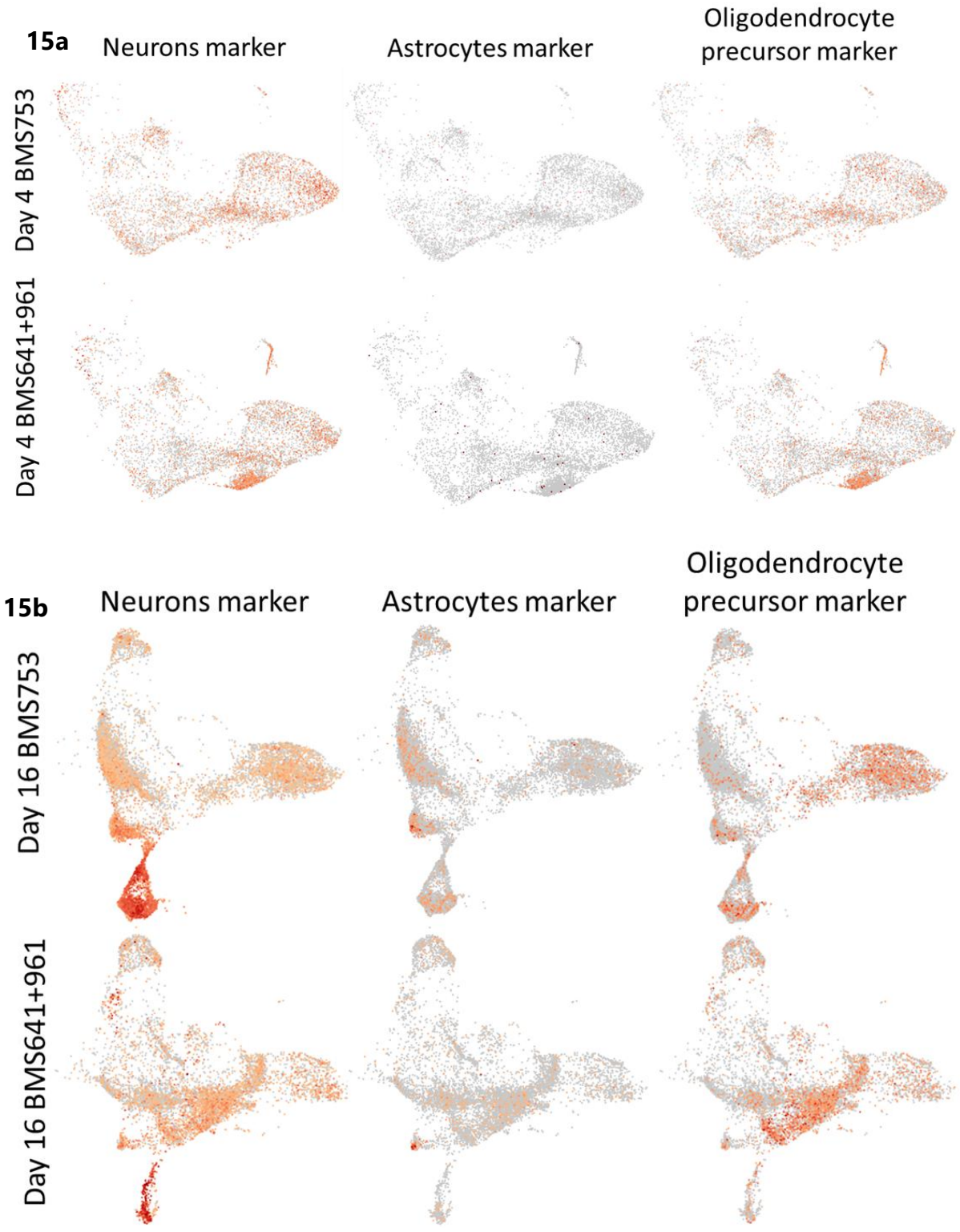


Figure 15. Gene expression for Neurons, Astrocytes, and Oligodendrocyte precursor markers

for BMS753 treated and BMS641+961 treated MESC Carlin cells at a) Day 4 and b) Day 16. Neuron Markers (TUBB3, MAP2, DCX, TH, TPH2, ENO2), Astrocyte Markers (S110B, AQP4, GFAP), Oligodendrocyte precursor Markers (MBP, MOG, SOX10, OLIG1, OLIG2).

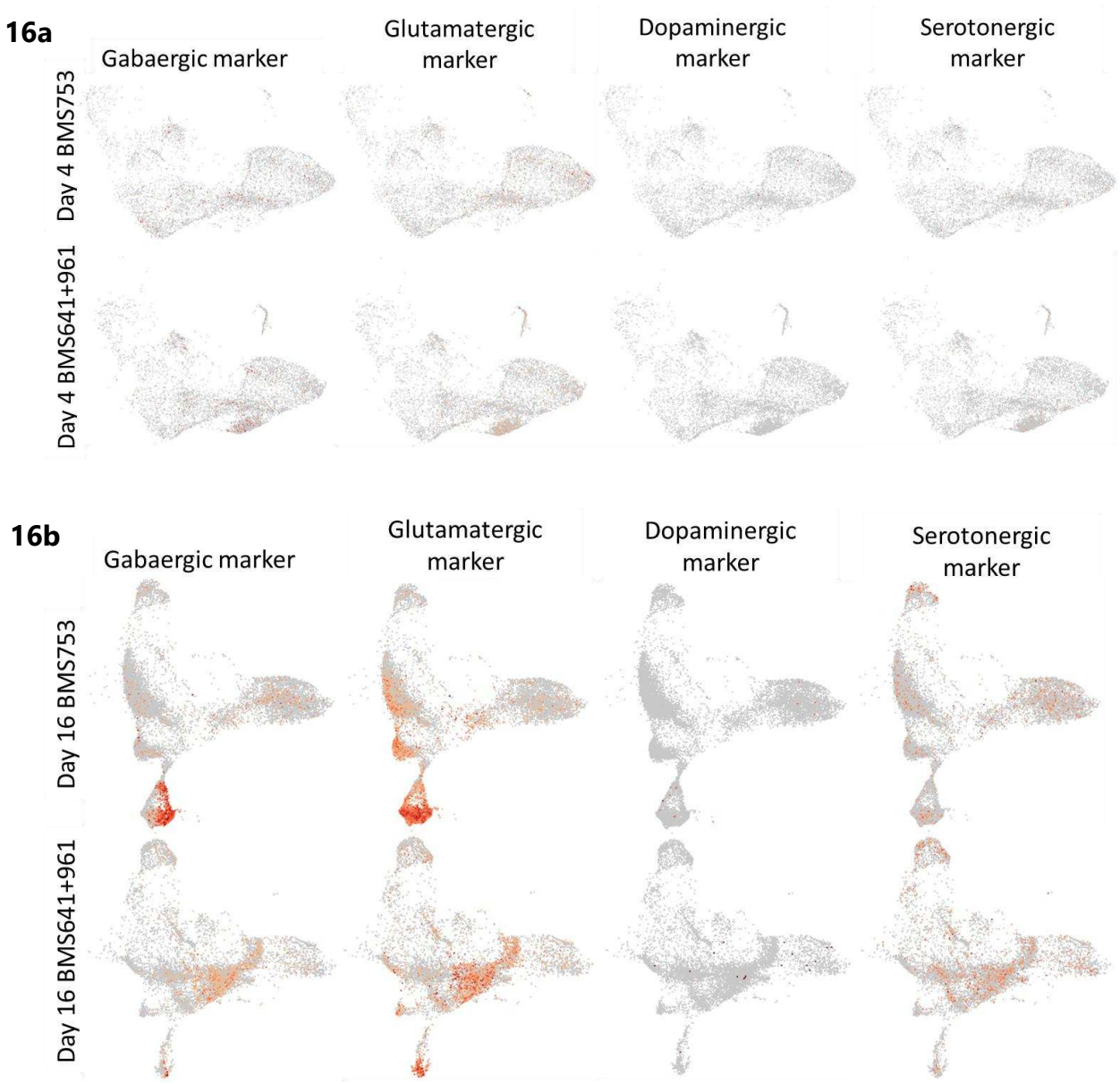


Figure 16. Gene expression for Gabaergic, Glutamatergic, Dopaminergic, and Serotonergic neuronal markers

for BMS753 treated and BMS641+961 treated MESC Carlin cells at a) Day 4 and b) Day 16. Gabaergic neurons Markers (GAD1, GAD2), Glutamatergic neurons Markers (GRIN1, GRIN2B), Dopaminergic neurons Markers (TH), Serotonergic neurons Markers (TPH2, SERTM1).

With a focus on the differences between BMS753 and BMS641+961 treated samples at endpoints of differentiation, I analyzed the top 20 genes from cluster 3 and cluster 11, that have a larger number of cells annotated to them in BMS641+961 versus BMS753 (Figure 12a). Genes associated with each cluster have been ranked according to Log 2 Fold change values (Supplementary Figure 5). These genes were then fed into the EnrichR Gene ontology tool

(https://maayanlab.cloud/Enrichr/) [133–135]. Specifically we were interested in the cell identities and compared the list of genes to the CellMarker and Panglao Databases (Figure 17). Cluster 3 doesn't have a distinct cell type but shows a possible progenitor/ interneuron phenotype contributed by genes LDHA, HMGB2, PCLAF, DUT, but also a mesodermal lineage from the blood cell expression genes – RAN, NPM1 (Figure 17a). Cluster 11 on the other hand, with CellMarker database has a possible identity of Neural progenitors from genes such as MDK, HMGB3, PLCAF, DACH1, FAT3, ADGRV1, CRABP1 (Figure 17b); and with the Panglao database, FAT3, DACH1, NPAS3, HES5, CCND2 suggest an identity of immature neural cells such as radial glia, interneurons, precursor cells and even mature neurons (Cajal Retzius cells)

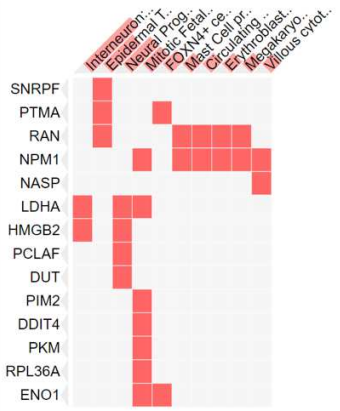
17a

Cluster 3: Top 20 genes (CellMarker DB)

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Mitotic Fetal Germ cell:Fetal Gonad	0.000001987	0.0001272	16.19	212.60
2	Natural Killer T (NKT) cell:Fetal Kidney	0.000008972	0.0002871	7.96	92.51
3	Neural Progenitor cell:Embryonic Prefrontal Cortex	0.00001999	0.0004264	30.58	330.92
4	Epidermal Transit Amplifying cell:Foreskin	0.0001048	0.001676	39.00	357.38
5	Interneuron:Embryonic Prefrontal Cortex	0.0004628	0.005924	73.89	567.33
6	Regulatory T (Treg) cell:Liver	0.0006924	0.007386	11.84	86.17
7	Exhausted CD8+ T cell:Liver	0.002102	0.01768	13.54	83.49
8	FOXN4+ cell:Lung	0.002795	0.01768	28.72	168.87
9	Gonadal Mitotic Phase Fetal Germ cell:Fetal Gonad	0.002893	0.01768	12.07	70.53
10	Erythroblast:Blood	0.003849	0.01768	24.28	135.02

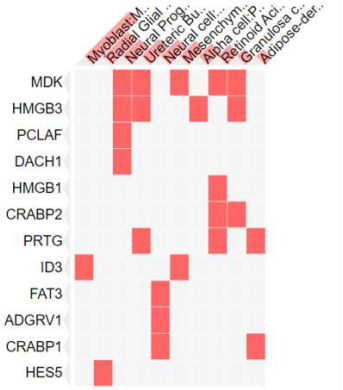
(Figure 17c).

17b



Cluster 11: Top 20 genes (CellMarker DB)

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Neural Progenitor cell:Embryonic Prefrontal Cortex	0.00001999	0.001439	30.58	330.92
2	Retinoid Acid Signaling-Responsive Fetal Germ cell:Fetal Gonad	0.00005905	0.002126	22.98	223.78
3	Ureteric Bud cell:Fetal Kidney	0.0001262	0.002340	36.55	328.16
4	Neural cell:Embryoid Body	0.0001300	0.002340	36.17	323.68
5	Granulosa cell:Fetal Gonad	0.0007905	0.01138	19.20	137.12
6	Endothelial cell:Fetal Gonad	0.001109	0.01210	17.02	115.83
7	Mesenchymal cell:Liver	0.001176	0.01210	45.20	304.86
8	Adipose-derived Stromal cell:Abdominal Adipose Tissue	0.003930	0.01301	24.02	133.05
9	Adipose Progenitor cell:Abdominal Adipose Tissue	0.003930	0.01301	24.02	133.05
10	Retinal Progenitor cell:Retina	0.004179	0.01301	23.26	127.40



17c

Cluster 11: Top 20 genes (Panglao DB)

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Radial Glia Cells	0.0001916	0.004598	31.59	270.40
2	Neural Stem/Precursor Cells	0.0004462	0.005354	23.49	181.20
3	Immature Neurons	0.008049	0.06439	16.46	79.35
4	Cajal-Retzius Cells	0.01290	0.07740	12.80	55.67
5	Osteocytes	0.09361	0.1822	10.79	25.55
6	Trophoblast Stem Cells	0.09452	0.1822	10.68	25.19
7	Purkinje Fiber Cells	0.09725	0.1822	10.36	24.14
8	Cholinergic Neurons	0.09725	0.1822	10.36	24.14
9	Kidney Progenitor Cells	0.09997	0.1822	10.06	23.16
10	Adrenergic Neurons	0.1009	0.1822	9.96	22.85

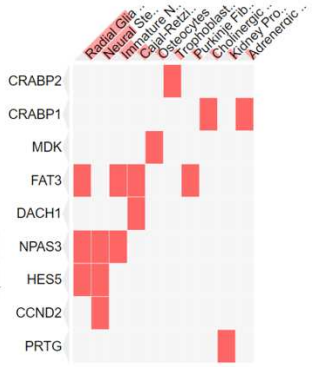


Figure 17. Tabular and Clustergram representation of the Gene ontology analysis performed on the top 20 genes

in a) cluster 3 compared to CellMarker database b) cluster 11 compared to CellMarker database and c) cluster 11 compared to Panglao database.

Interestingly, FAT3 is a gene that is involved in axonal development in the CNS, in the embryological development facilitating axonal tract formation [136] and it also has a possible role in modulating the morphology of microglia [137]. DACH1 is a gene involved in transcriptional processes and more specifically plays a role in determining cell fate in the eye and limbs and is expressed primarily in the neuroepithelia and radial glia, and also in striatal medium spiny neurons [138,139]. NPAS3 also is largely expressed in the ventricle zone by radial glia involved in their migration during neocortex formation, and possibly in the transition of stem cells to a differentiated fate [140], and disruptions in its expression have been seen to cause abnormal synaptic activity between astrocytes and neurons as well as impaired astrocyte differentiation [141]. HES5 acts inversely by promoting stemness and maintaining a neural stem cell population, and times the cortical layer formation as well as gliogenesis activation [142]. CCND2 is a cell cycle regulator, that has been implicated in megalencephaly, cortical malformations where it is found to exist in its variant form [143].

Conclusion

The initial annotation interpretation using Loupe Browser of the datasets generated from Seurat reveals that at early differentiation (day 4), as expected, the mature neuronal markers don't appear distinctly but we can already see the expression of neurogenesis and OPC features at this stage. Moreover, we discover an endodermal and mesodermal signature that continues to be present in our cultures at late timepoints. This is an expected off-target activity of the synthetic Retinoid ligands since we don't use any other growth factors or inhibitors in the medium to prevent this fate. Unfortunately, we don't know to what extent this can interfere with neuronal differentiation and the yield of mature postmitotic neurons in our culture.

Additionally, we can visualize the appearance of cluster 8 with a phenotype corresponding to neural progenitors and mature neuronal cells with a GABAergic and/or Glutamatergic signature, cluster 10 with a possible endodermal signature, and cluster 13 most likely corresponding to a neural fate with cells expressing genes related to neural stem cells and mature neurons.

3D cell differentiation of MESC to study cell fate specialization (Brain organoids)

In the previous section we looked at the context of retinoid receptor activity in monolayer cell cultures. However, cells cultured in a 2D versus 3D structures interact differently. In order to realize the complexity of neuronal differentiation regulated by retinoid receptor subtypes in a 3D format, we utilized the technology of brain organoid generation that is sustainable for a longer period of time (months) and traced their development in relation to the synthetic retinoid ligands (ATRA, BMS753, and BMS641+961).

The brain organoids were generated using Mouse brain organoid protocol version 2, and grown for a period of 3 months (Figure 18). The organoids were cultured as usual till day 9, in maturation medium. At this point, medium is generally changed to Maturation medium+ Vitamin A, which is the positive control. The other conditions utilized maturation medium + ligand (ATRA at 1µM, BMS753 (Alpha) at 1µM, and BMS641+961 (Beta + Gamma) at 0.1µM). The ligands are added fresh to the organoids every time the medium is refreshed.

An immunofluorescence staining was performed at 1, 2, and 3-month old brain organoids staining for Sox2 (neural stem cells), Tubb3 (Immature neurons), Gfap (Astrocytes), Gad67 (GABAergic neurons) (Figure 19, Figure 20, Figure 21, Supplementary Figure 6). All the organoids show a presence of these neural lineage markers indicating a heterogenous population of neural cell types but also that a large percentage of the cells in the organoid are able to differentiate into the neuroectodermal pathway.

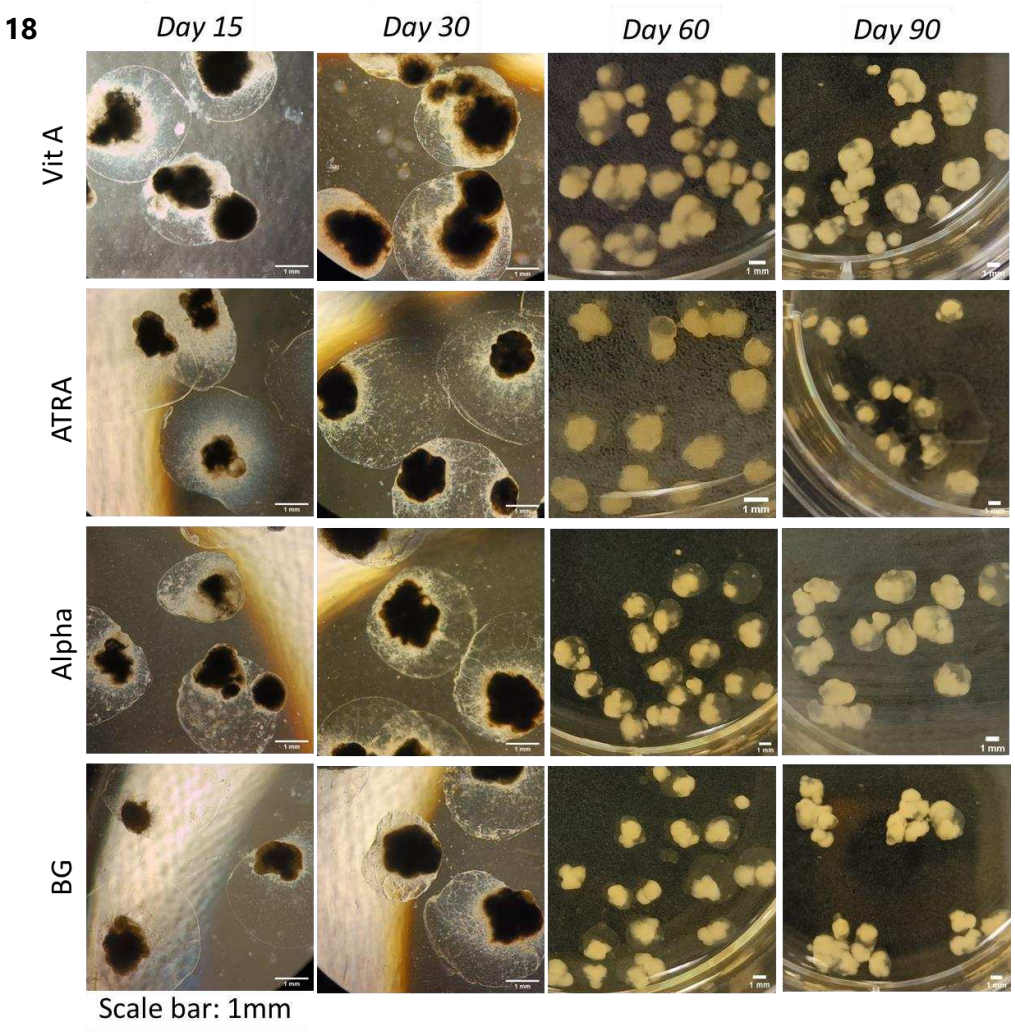


Figure 18. Brightfield images of 3D Brain organoids derived from Carlin MESC at Day 15, 30, 60, 90 of growth

19a

19b

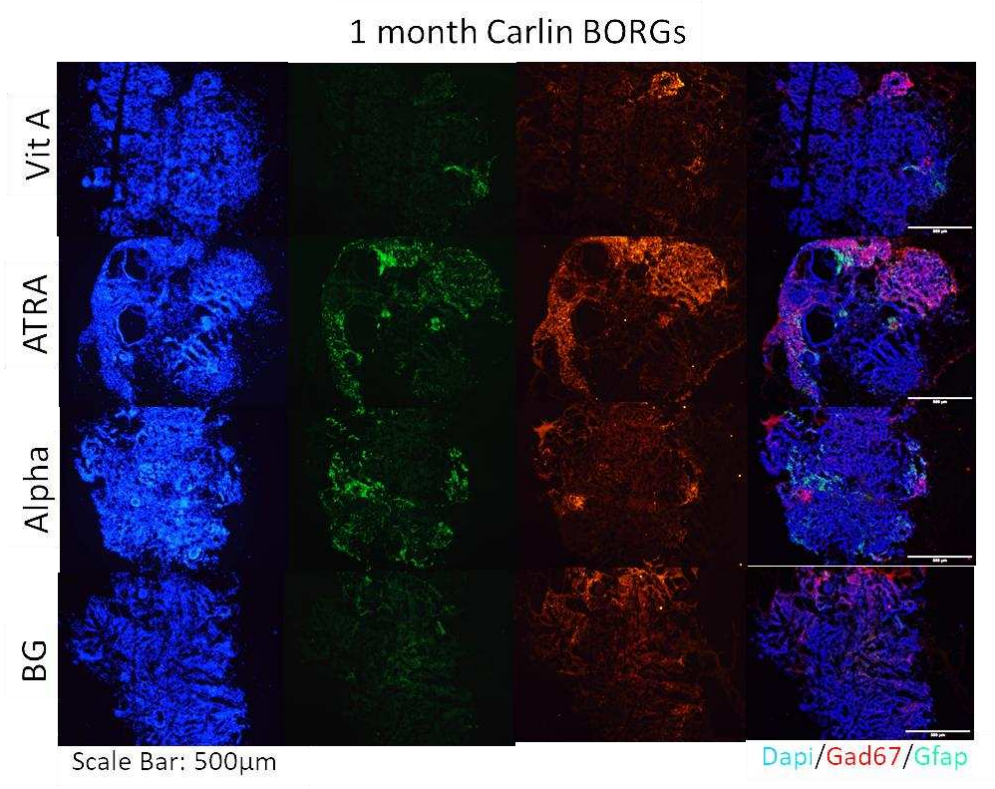
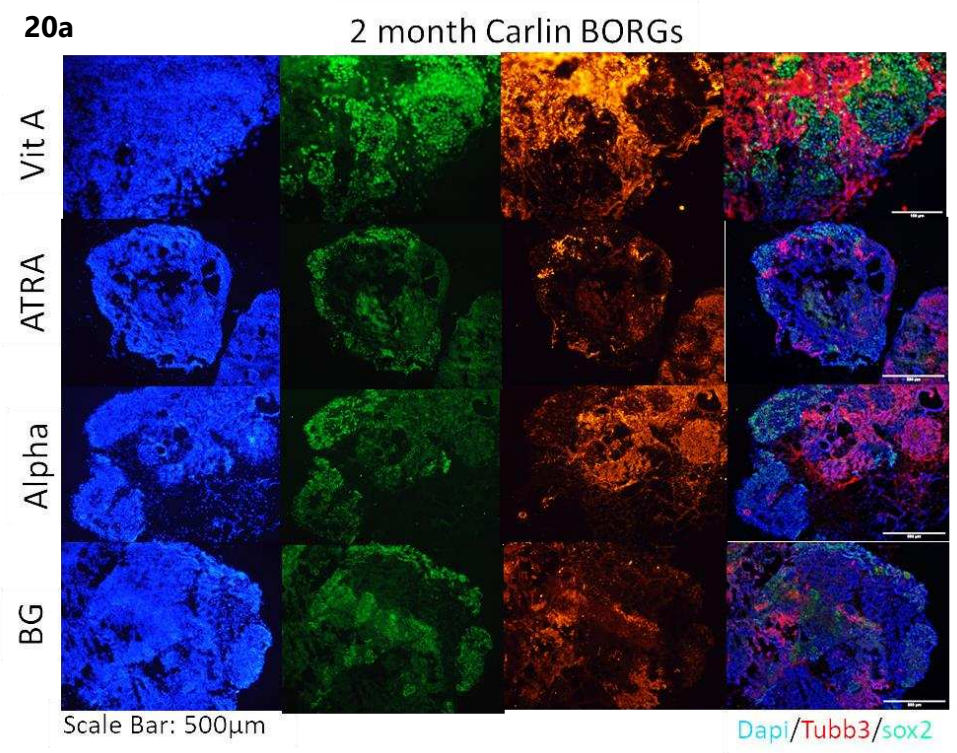


Figure 19. Immunofluorescence staining of 3D Carlin BORGs at 1 month

Stained with a) Sox2 (neural stem cells), Tubb3 (immature neurons), b) Gad67 (Gabaergic neurons), Gfap (astrocytes), and Dapi (nuclei).



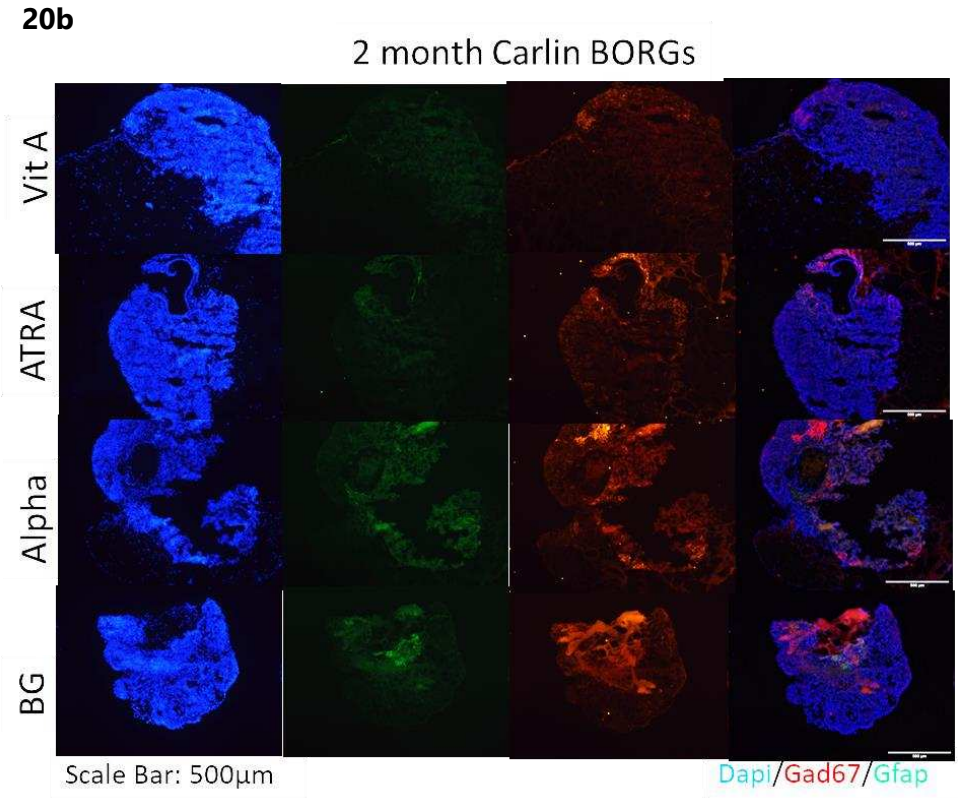
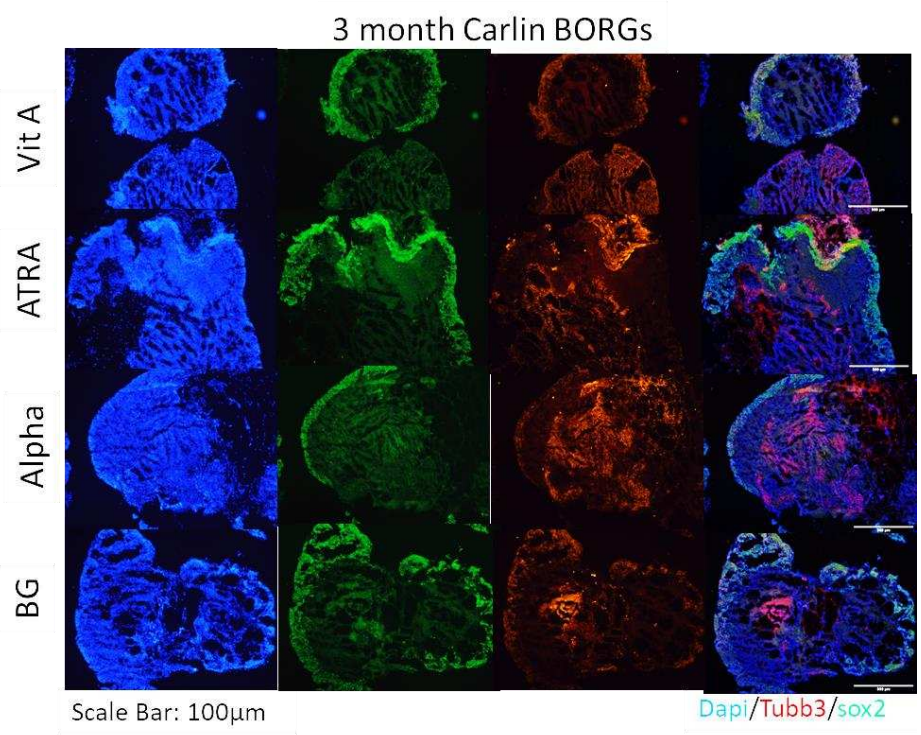


Figure 20. Immunofluorescence staining of 3D Carlin BORGs at 2 months

Stained with a) Sox2 (neural stem cells), Tubb3 (immature neurons), b) Gad67 (Gabaergic neurons), Gfap (astrocytes), and Dapi (nuclei).

21a



21b

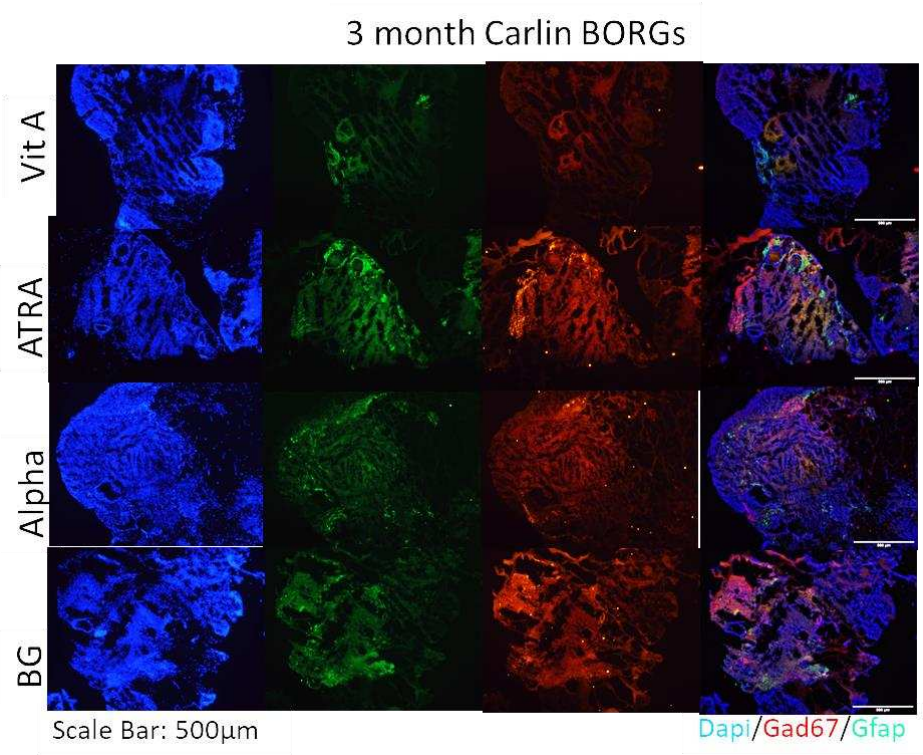


Figure 21. Immunofluorescence staining of 3D Carlin BORGs at 3 months

Stained with a) Sox2 (neural stem cells), Tubb3 (immature neurons), b) Gad67 (Gabaergic neurons), Gfap (astrocytes), and Dapi (nuclei).

We also did a comparison of the gene expression profiles targeting markers for pluripotency, neurogenesis, astrocytes, OPCs, and mature neuronal subtypes over a period of 3 months (Figure 22). We find that most pluripotency factors remain down throughout the differentiation timeline, except for Klf4 that is highly expressed in all conditions. Since Vitamin A and ATRA treated organoid are the positive controls, we mainly focused on differences between the Alpha (BMS753) and BG (BMS641+961) condition. In terms of neurogenesis, we find that Nes is strongly expressed up to 2 months in Alpha while there is a steady decline for BG treatment. Tuj (Tubb3) is also strongly expressed in Alpha BORGs at 15 days of growth compared to BG where they remain below Log2 fold change levels of 2. Interestingly, the presence of Th (Dopaminergic neurons) and Tph2 (Serotonergic neurons) positive cells is more robust for Alpha BORGs than BG (Figure 22).

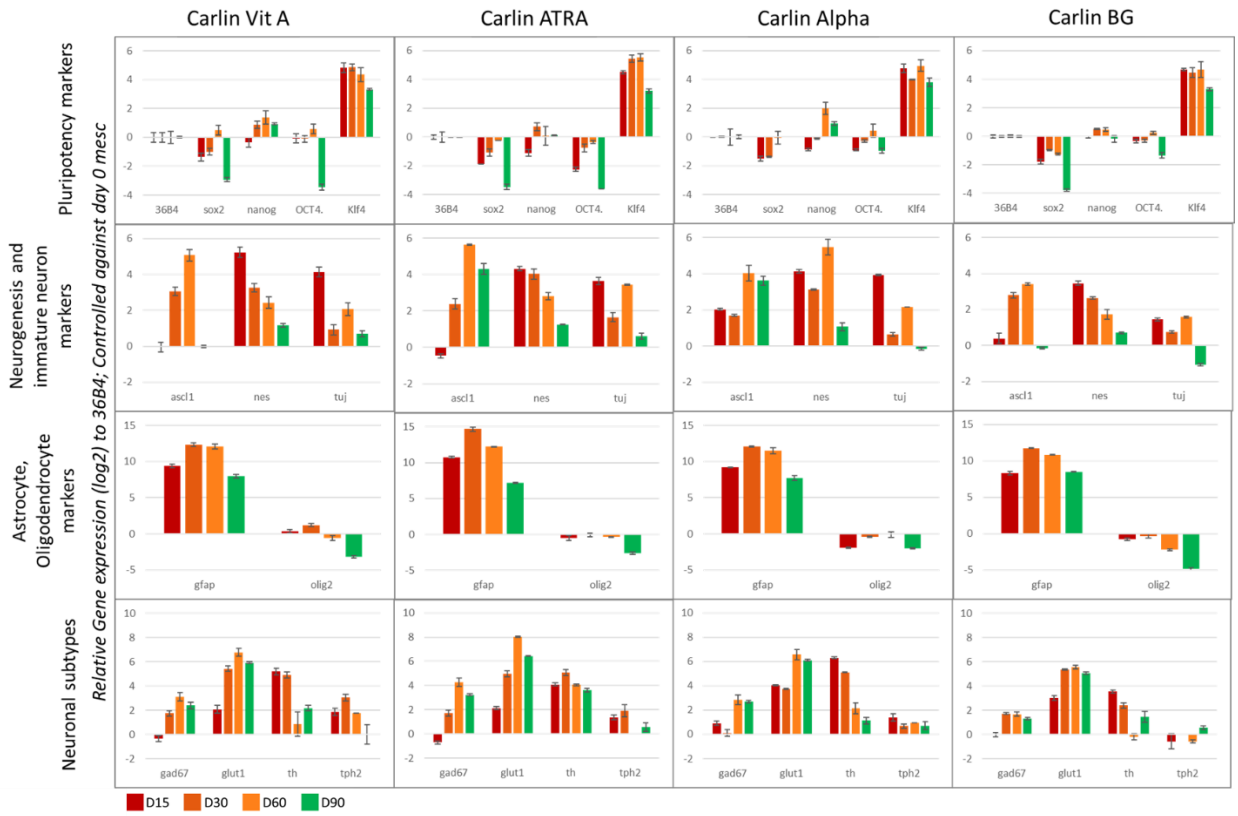


Figure 22. Gene expression analysis of brain organoids at Day 15, 30, 60, 90.

Conclusion

Utilizing the technology of organoids, we sought to develop brain organoids in conditions where they are exposed to individual receptor ligands during the maturation process in opposition to the control treatment of Vitamin A. Retinoids in any form are essential for brain development [83,119], and there is ample evidence of the development defects that arise in a Vitamin A deficient diet [9,10]. Numerous RAR knockout studies also corroborate findings of the necessity of certain receptors during embryological development [48,51–56], and this was the motivation to expand the duration of our neuronal differentiation studies and introduce the complexity of a 3-dimensional structure in addition to the monolayer cultures.

To this end, we are able to find that the treatment of individual receptor ligands are still able to produce BORGs that develop in the neuroectodermal differentiation pathway. However, we are not able to observe striking differences between the ligands treatments. This can possibly be attributed to the redundancy of retinoid receptors in development, and a more

in-depth comparison would require an approach such as single cell RNA sequencing and spatial transcriptomics to elucidate cell fate acquisition attributed to receptor isotype activation in long term cultures.

Discussion & Perspectives

In Section 1, we discuss the findings of our publication, highlighting the observance that RARβ and RARγ synergistic activity demonstrates an enhanced cell specialization recovery in RARα KO cells compared to WT cells.

Consequently, we set out to verify the neuronal cell specialization ability of RARβ and RARγ, compared to RARα using Single Cell RNA sequencing approach. In our preliminary analysis, we can make assumptions as to the identity of the certain cell clusters such as 8 corresponding to Gabaergic and/or Glutamatergic neuronal phenotype, cluster 10 to endodermal phenotype, and cluster 13 to a transitory neural fate phenotype comprising stem cells, glial cell and mature neuronal markers.

Since we are interested in elucidating the specificity of RARβ and RARγ coactivation on neural cell fate, we focused on cluster 3 and 11 that represent a larger number of cells compared to RARα activation at day 16 of differentiation and found that for cluster 11, there seems to be an identity related to progenitor cells that are in a transition state comprising radial glia, interneurons, precursor and mature neuron cells.

Further analysis will require defining more comprehensive and exhaustive list of verified markers for each cell cluster type in order to annotate our clusters and determine, if present, a contrasting cell fate signature that is unique to the coactivation of RARβ and RARγ receptors.

Additionally, we extended our differentiation experiments to 3 months of culture to elucidate these same differences in a 3D brain organoid cell culture system. While we don't see the striking differences with a qCPR gene expression profile, it would be necessary to analyze these samples using a single cell transcriptomics or even spatial transcriptomics and epigenomics approach. This type of analysis is already gaining momentum where study of neurodevelopment disorders or distinct neuronal subtypes benefits from such an approach [86,144]. There is much to be mined from the organizational regions that form in a brain organoid that can be exploited to study long term cell differentiation. While currently, we are working on this aspect, it is too early to publish our findings at the time of the deposit of this thesis.

Section 2- Development of Mouse brain organoids that recapitulate H3.3 mutated pediatric glioblastoma

2D Neuronal Differentiation Capacity

The cell lines used in our Glioblastoma studies have been kindly provided by Dr. Ali Hamiche, IGBMC, Strasbourg as part of the Nanotumor consortium.

To verify the neuronal differentiation capacity of our cell lines, we subjected them to the 2D Neuronal differentiation protocol, driven by retinoic acid treatment as described in the Methods section. All 3 cell lines- H3.3 WT, H3.3K27M, H3.3G34R (hereafter referred to as WT Glio, K27M, G34R respectively)- showed the capacity to neuronally differentiate (Figure 23); more often, the G34R cell line was more robust in neuronal production, and required the presence of Retinoic acid to be induced to form neurons (Figure 24).

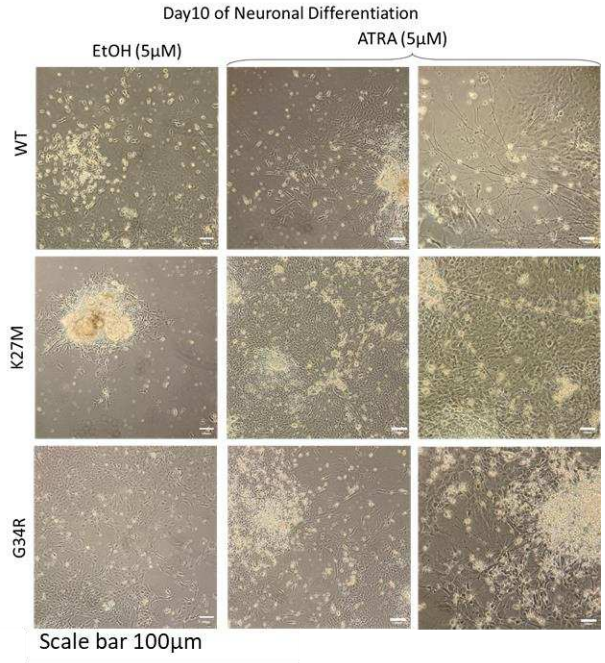


Figure 23. Brightfield images of neuronally differentiated H3.3 wildtype and mutant cell MES cells (WT Glio, G34R, K27M)

at a) Day 4, b) Day 8, and c) Day 16 of Neuronal differentiation and treated with ligands (ETOH: ethanol, ATRA, BMS753: RARα agonist, BMS641+961: RARβ + RARγ agonist).

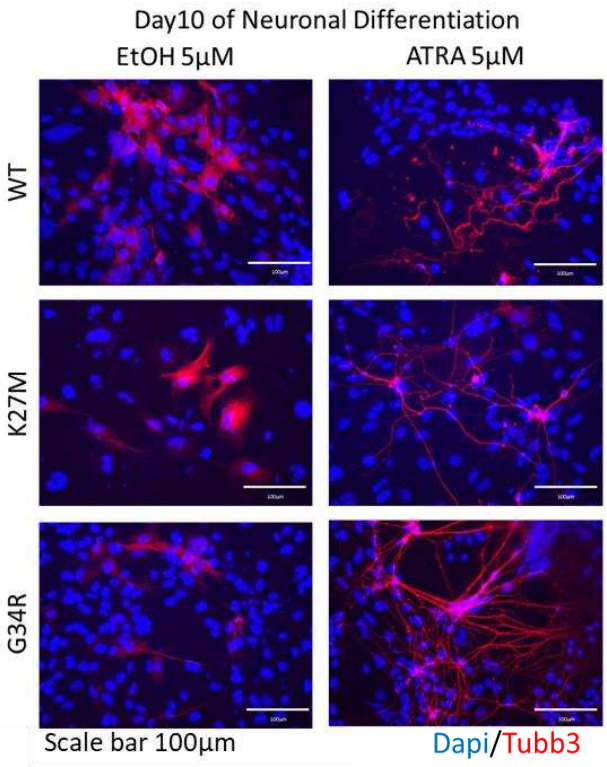


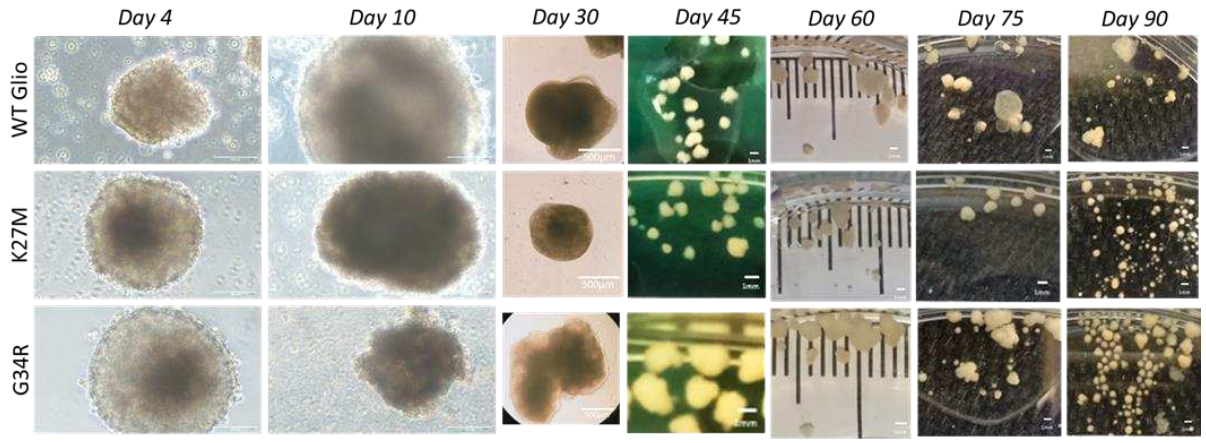
Figure 24. Immunocytochemistry staining of neuronally differentiated H3.3 wildtype and mutant cell MES cells (WT Glio, G34R, K27M)

with immature neurons (Tubb3) after 10 days of 2D neuronal differentiation.

Mouse Brain organoid generation with H3.3 mutant cell lines

The first batch of brain organoids were developed using the protocol described in Lancaster et al, 2013, for human induced pluripotent stem cells (hiPSC) (Material and methods: Mouse brain organoid generation version 1) with an initial 10-day period of Neuronal commitment and induction, before the EBs (embryoid bodies) generated were submerged in Matrigel droplets and transferred to maturation medium. Their growth over a period of 3 months overall did not show significant differences (Figure 25a), however there was an initial larger size seen in G34R organoids from Day 30- Day 45 (Figure 25b, Supplementary Figure 7).

25a



Scale bar: 100µm (Day 4 & 10), 500µm (Day 30), 1mm (Day 45, 60, 75, 90)

25b

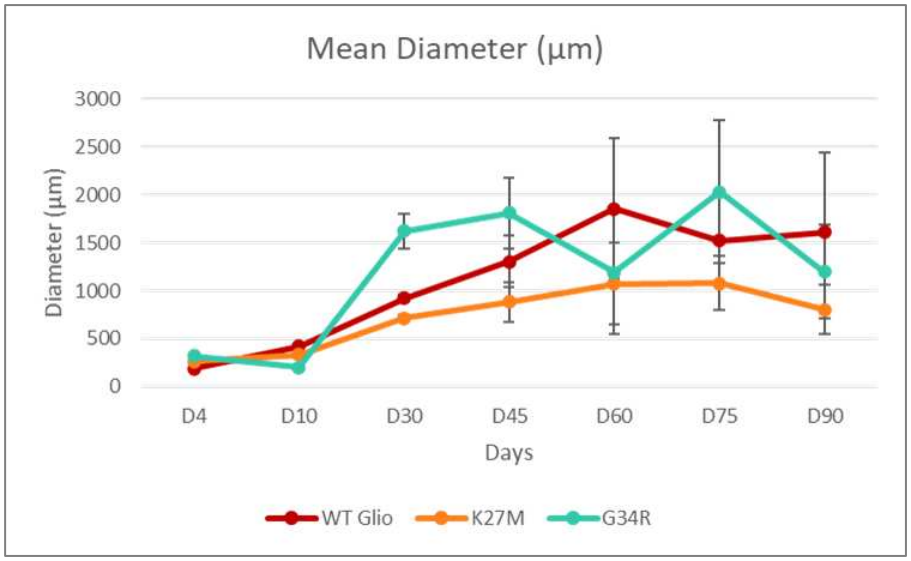


Figure 25. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1)

a) Brightfield images of mouse brain organoids derived from MESC of WT Glio, G34R, K27M; b) Plot of mean diameter of brain organoids generated for the 3 cell lines

An immunofluorescence staining of all 3 brain organoids at 2 months of age, revealed that they were successfully differentiated into the neuroectodermal pathway by the presence of marker Gad67 (Figure 26, Figure 27, Figure 28). There continues to exist a niche of neural stem cells (Sox2) in all the brain organoids, as well as proliferation spots (KI67). Additionally, we observed a robust population of astrocytes (Gfap) as well (Figure 26, Figure 27, Figure 28).

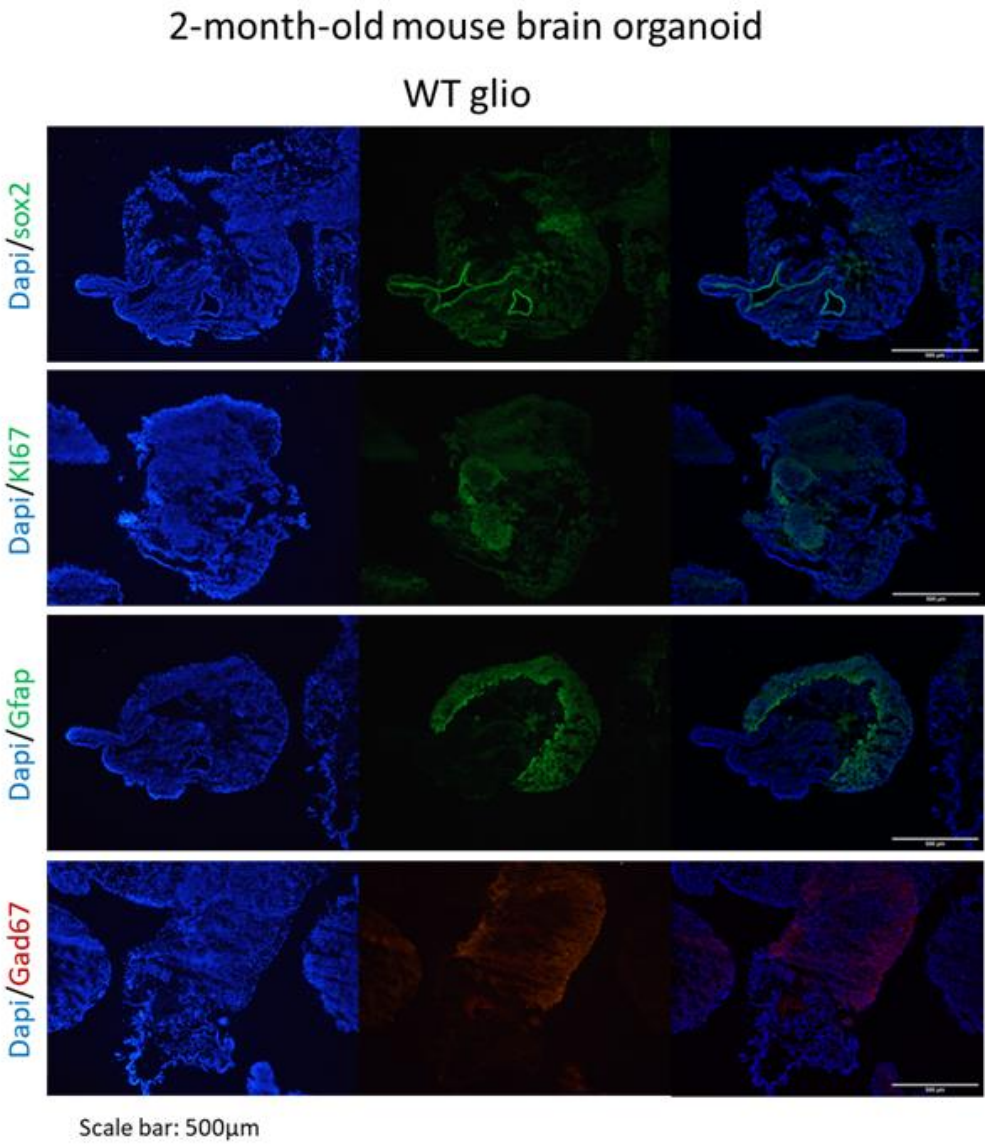


Figure 26. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1) immunofluorescence assay

2 month old WT Glio brain organoids with markers Sox2 (Neural stem cells), KI67 (proliferation), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons)

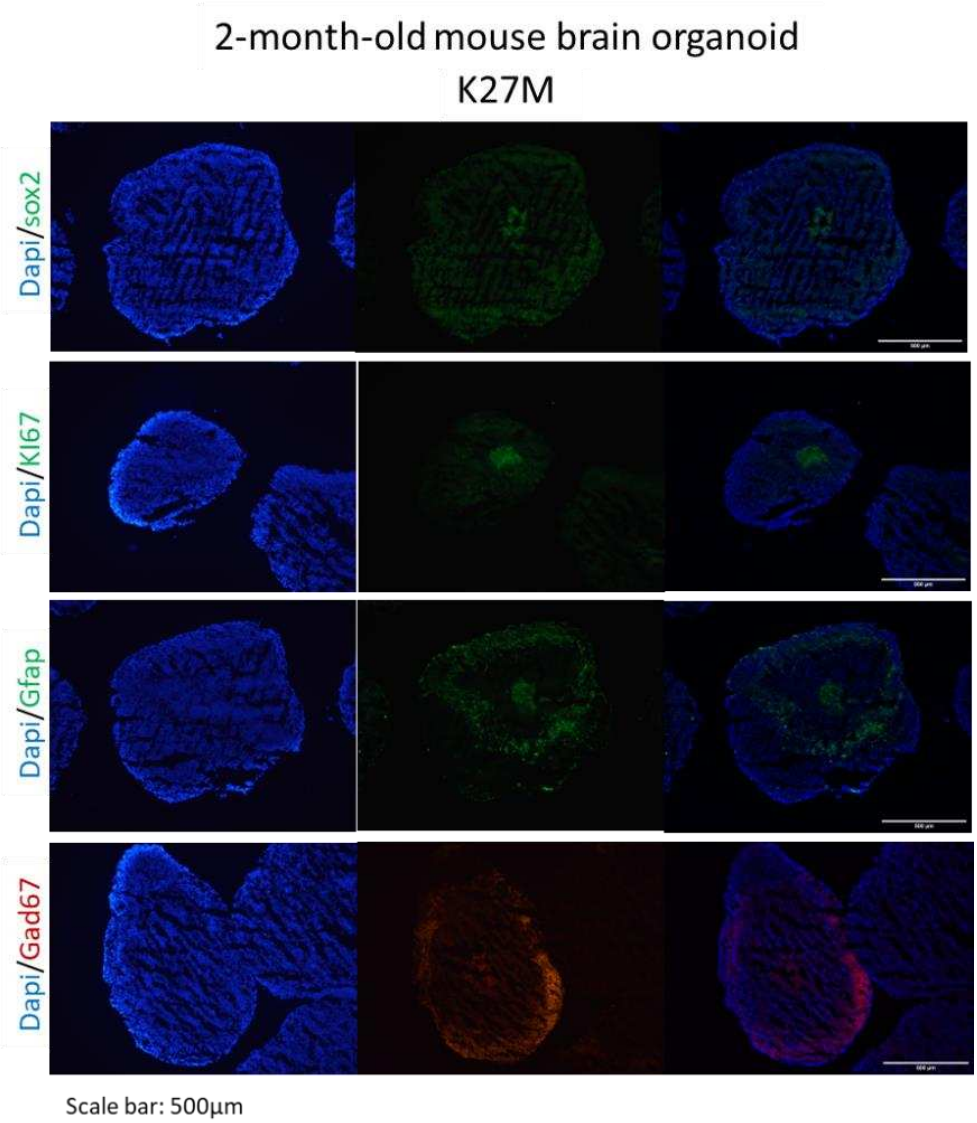


Figure 27. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1) immunofluorescence assay

2 month old WT Glio brain organoids with markers Sox2 (Neural stem cells), KI67 (proliferation), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons)

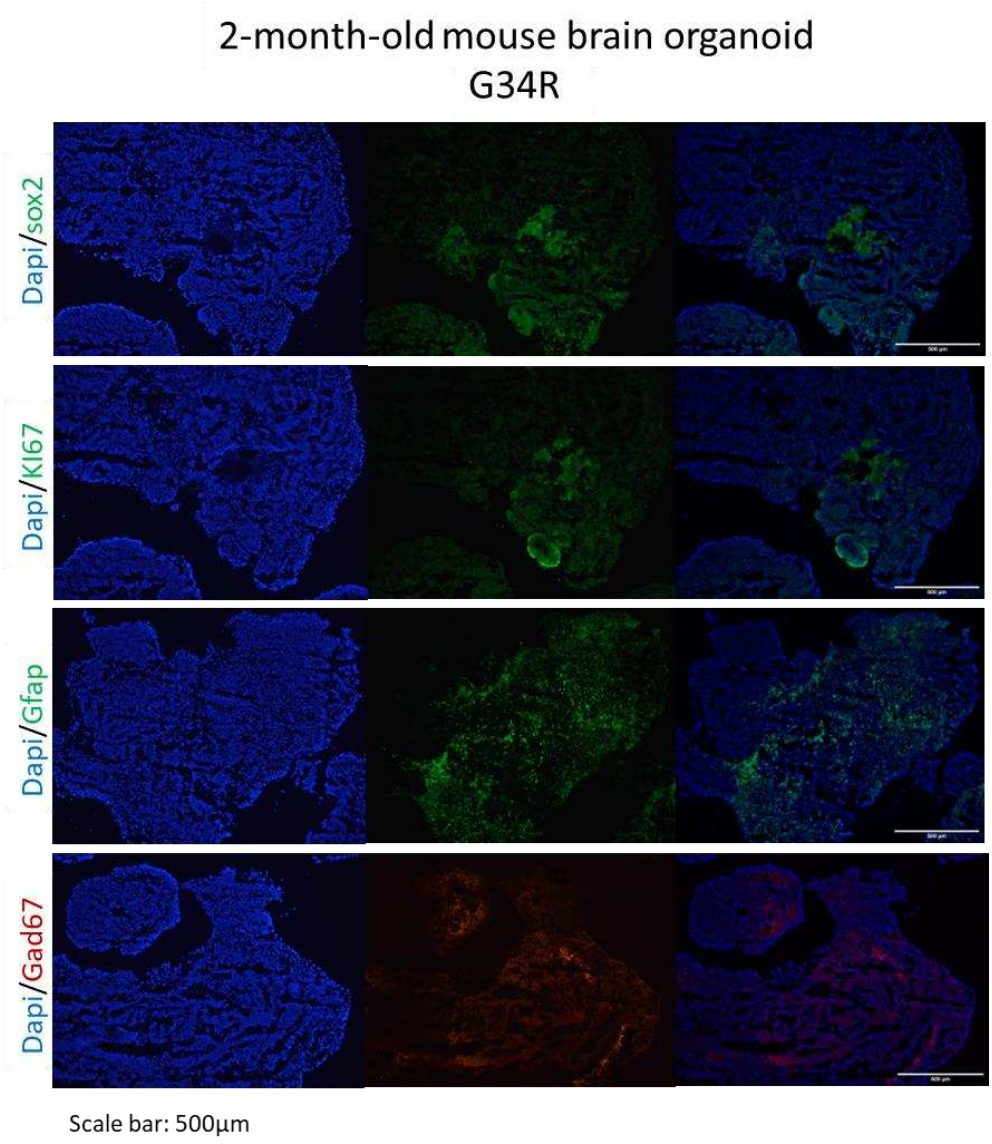


Figure 28. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1) immunofluorescence assay

2 month old G34R brain organoids with markers Sox2 (Neural stem cells), KI67 (proliferation), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons)

We validated our findings by performing gene expression analysis for a variety of markers at different timepoints for all the organoids. G34R organoids retain a neural stem cell niche (Sox2) throughout their growth, while for K27M organoids, there is a small presence at 3 months, and for WT Glio organoids, we also observed a regain of stemness by 3 months (Figure 29). This regain of stemness profile was more robust in G34R organoids, compared to the other 2 cell lines. All other markers for pluripotency (Nanog, Oct4, Klf4) remain down. All the 3 cell lines also have show strong expression for markers of neurogenesis and immature

neuron markers that peak around 1 month of age, indicating a strong neural precursor population expanding and the receding as they probably move to become more mature (Figure 29). We also notice a good population of astrocytes (Gfap) and surprisingly, a population of microglia as well (Iba1, Trem2) (Figure 29). Since these cells emerge from mesodermal lineage, they are evidence of the protocol not being 100% penetrative. There are likely niches of cells in the initial stages that don't get exposed in enough quantities to the initial ligands and escape into the mesodermal/endodermal fate. Interestingly, the expression for Oligodendrocyte precursor cells (OPCs) (Olig2), was very low in WT Glio and K27M organoids but expressed at high levels in G34R organoids. This is in contrast to studies where it has been found that Olig1/2 expression tends to be high in K27M mutant and not in G34R mutant tumors [112].

There is also confirmed expression of mature/postmitotic neuronal markers (Gad67- Gabaergic, Glut1- Glutamatergic, Th- Dopaminergic, Tph2- Serotonergic, Chat- Cholinergic). Primarily, G34R organoids have a higher expression profile for Gabaergic and Glutamatergic neurons, and all 3 cell lines show presence of markers for Dopaminergic and Serotonergic neurons, while none show a strong expression for cholinergic neurons (Figure 29). While none of the cell lines show a comparable difference in expression of KI67, there is a higher expression seen for CD133 (Cancer stem cell) in K27M and G34R organoids. Additionally, CD44 (Cancer stem cell) is stronger for WT Glio and K27M organoids but diminishes over time in G34R organoids (Figure 29). We also analyzed the expression patterns of cell cycle markers that tend to show variations in Glioblastoma, and observed a slight increase in P53 and P21 expression over time, indicating that ell cycle regulation is not altered. However the overexpression of Pdgfra in K27M organoids, is consistent with clinical observation of Pdgfra amplification in Glioblastoma tumors [94].

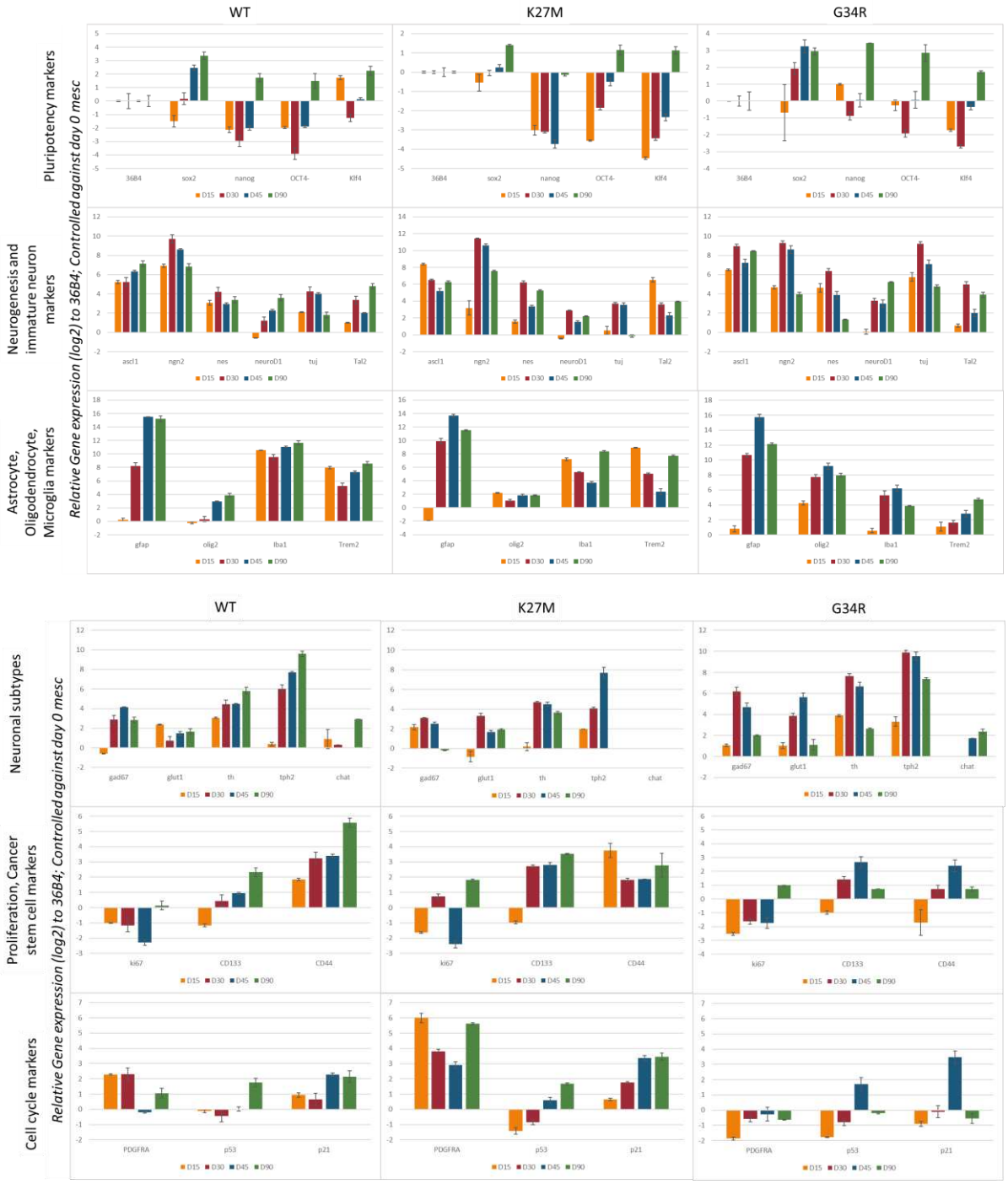


Figure 29. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1) gene expression analysis

at Day 15, 30, 45, 90.

Conclusion

There are some patterns observed in the organoid derived from H3.3 mutant cell lines that are suggestive of an oncogenic profile, but there is no overt expression of a tumor overgrowth, excessive proliferation, and expansion, that is characteristic of Glioblastoma. The other concern was the presence of mesodermal lineage cells (microglia), that made us reconsider our methods of culturing organoids.

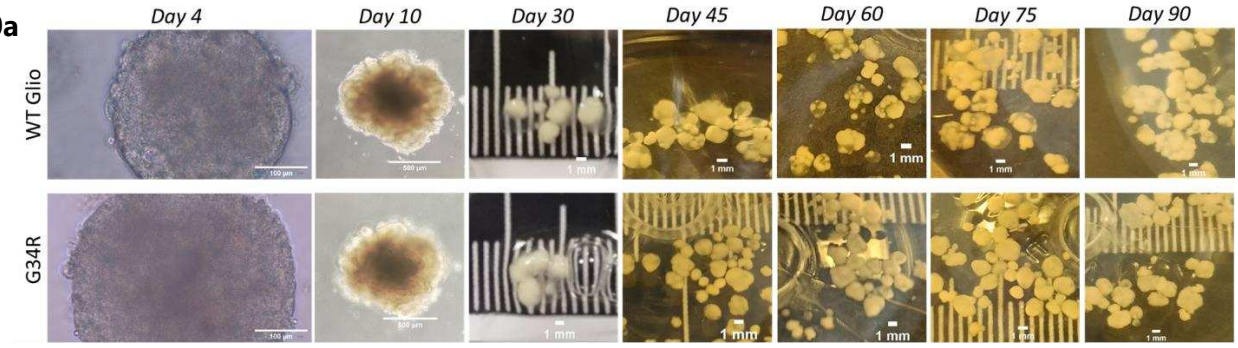
We know that mouse cells have shorter cell cycles and inferring from literature, we discovered that mouse brain organoids might be better manufactured when they are induced for a period of 6 days rather than 10 days before being embedded in Matrigel [83]. We therefore utilized the protocol from Lancaster et al, 2013, to produce a new batch of organoids with shorter neuroectodermal induction time frames.

Mouse Brain organoid generation with H3.3 mutant cell lines using a modified protocol

This batch of brain organoids were developed using the protocol described in Lancaster et al, 2013, (Material and methods: Mouse brain organoid generation version 2) with an initial 6-day period of Neuronal commitment and induction, before the EBs generated were submerged in Matrigel droplets and transferred to maturation medium (Figure 30a). As visible in Day 10, there is an abundance of neuroectodermal folds reminiscent of the cortical development of the brain, that was absent in the previous batch of organoids (Figure 25a).

Their growth over a period of 3 months was monitored but did not show size differences between the 2 cell lines- WT Glio and G34R (Figure 30b, Supplementary Figure 8). We have continued the rest of our experiments comparing only WT Glio to G34R and eliminating K27M, to be able to optimize our protocols and economize on time and analysis.

30a



Scale bar: 100 µm (Day 4), 500µm (Day 10), 1mm (Day 30, 45, 60, 75, 90)

30b

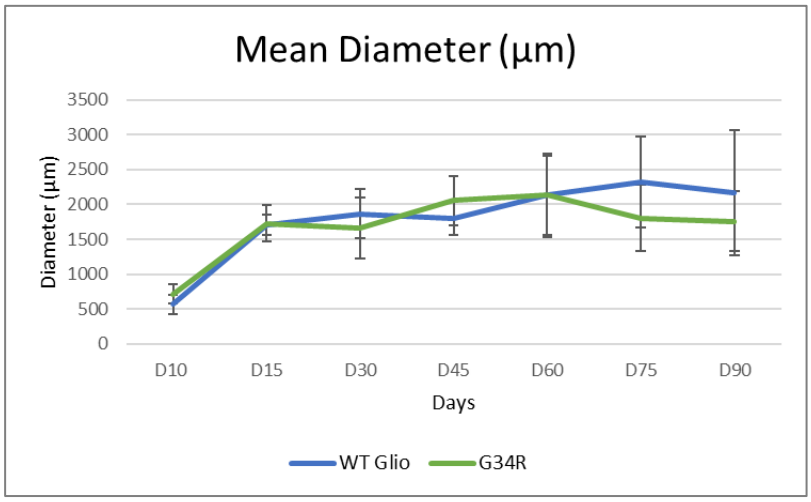
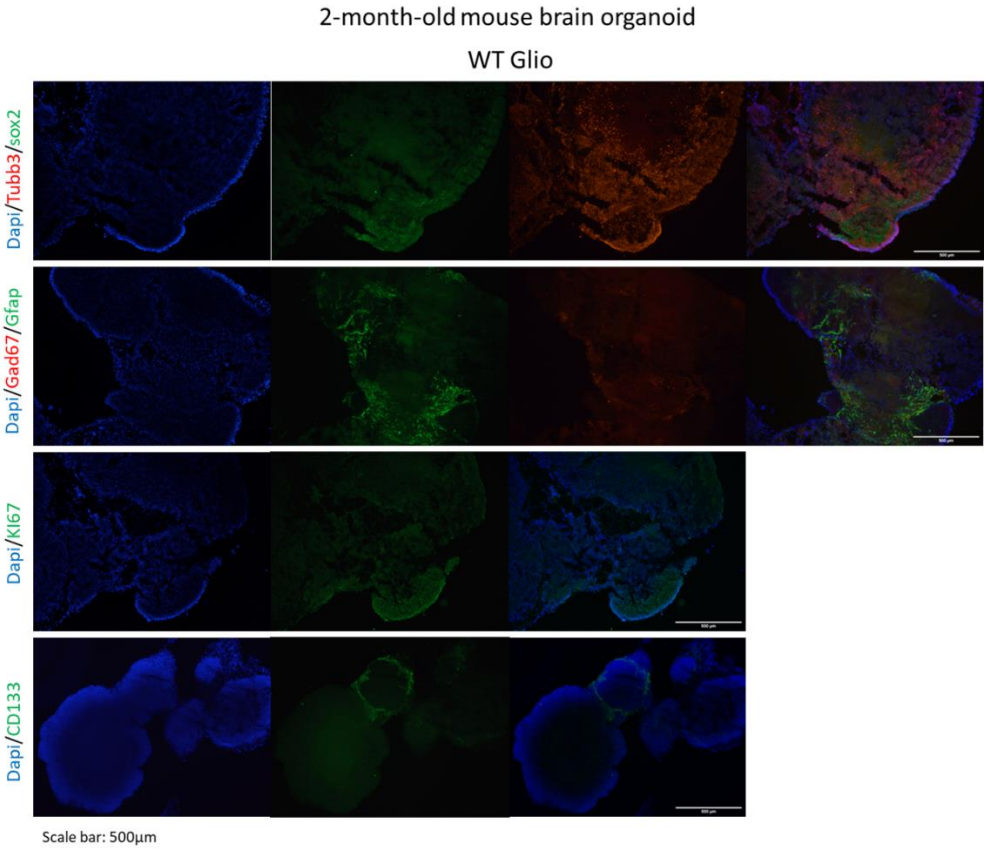


Figure 30. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 2) size measurement

a) Brightfield images of mouse brain organoids derived from MESC of WT Glio, G34R; b) Plot of mean diameter of brain organoids generated for both cell lines;

At 2 months of age, we were able to visualize a robust population of immature neurons (Tubb3), in close proximity to neural stem cell niches (Sox2) and strong expression of mature neural cell types (astrocytes and Gabaergic neurons (Gfap, Gad67)), localized to the outer regions of the organoids (Figure 31, Supplementary Figure 9). There is a dispersed expression of KI67 all over the organoids, and CD133 tends to be localized in patches in both organoids as well. This represents a difficulty in the inference of the presence of a cancer stem cell niche, since CD133 expression is seen in both cell lines.

31a



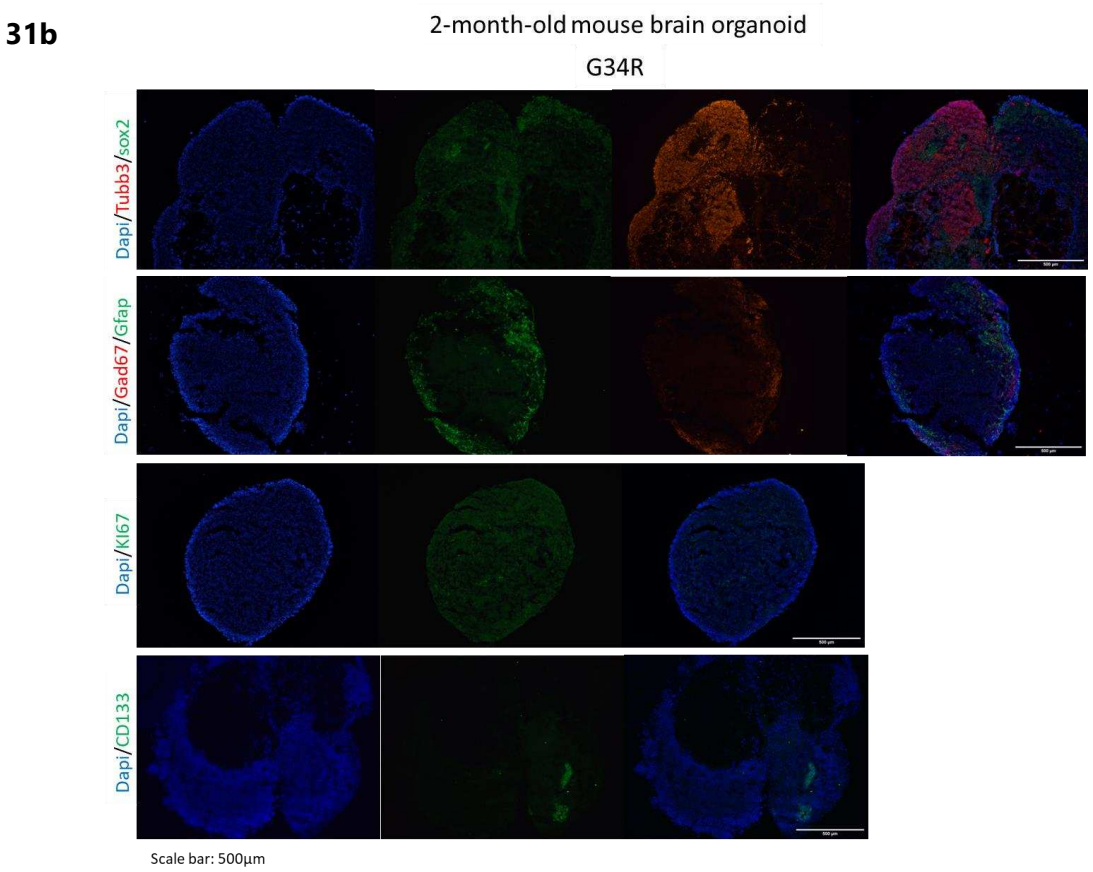


Figure 31. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 2) immunofluorescence staining

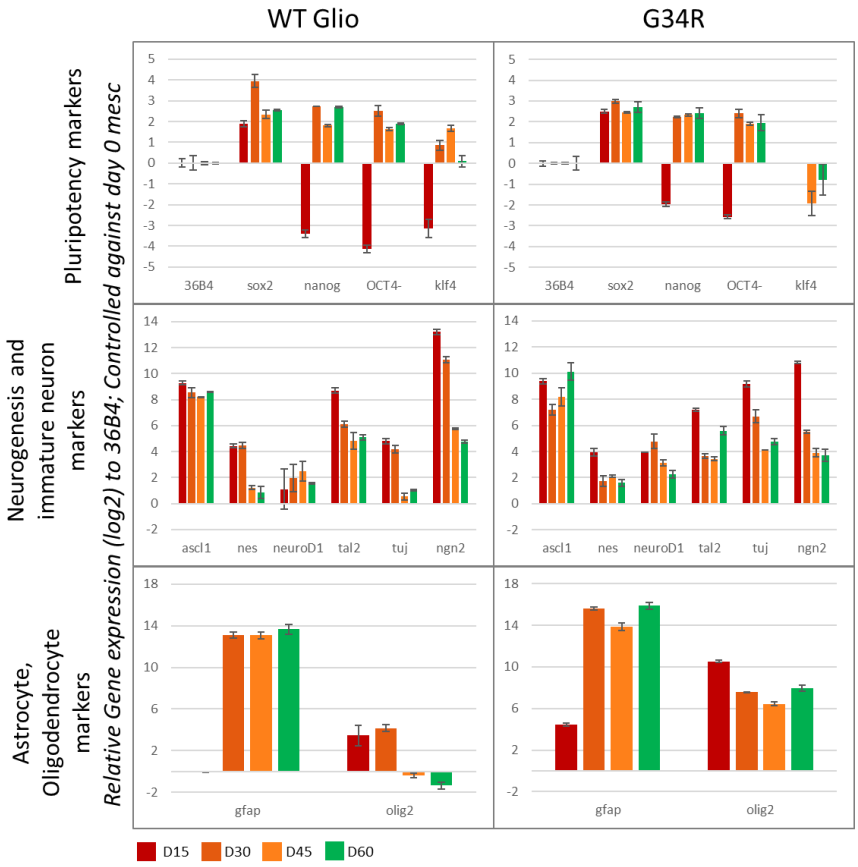
2 month old brain organoids for a) WT Glio, b) G34R with markers Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell)

Analysis of the gene expression profiles reveals an initial dip in the pluripotency markers that reemerges from 1 month onwards (Figure 32). The population of cells expressing neural precursor and immature neuronal markers has a high expression generally at 2 weeks and this reduces as the organoids mature. Interestingly, while Gfap levels are comparable in both organoids, it's the high expression of Olig2 in G34R organoids that provides a hint of tumorigenic influence of the H3.3G34R mutation.

All the markers of mature neuronal markers are strongly expressed, indicative of a robust neuroectodermal commitment, possibly due to the version 2 protocol (Figure 32). The expression of Pdgfra dips over time in WT Glio organoids and increases for G34R organoids while P53 expression is expressed at a basal level for WT Glio and strongly repressed initially

before returning to basal levels for G34R organoids. P21 as well decreases for WT Glio but remains moderately expressed for G34R organoids.

KI67 is not prominent in G34R and CD133 expression falls for WT Glio organoids but continues to be expressed in G34R mutant organoids (Figure 32). There could be an indication of a disruption in the P53 pathway and maintenance or reemergence of cancer stem cell expression in G34R-derived organoids.



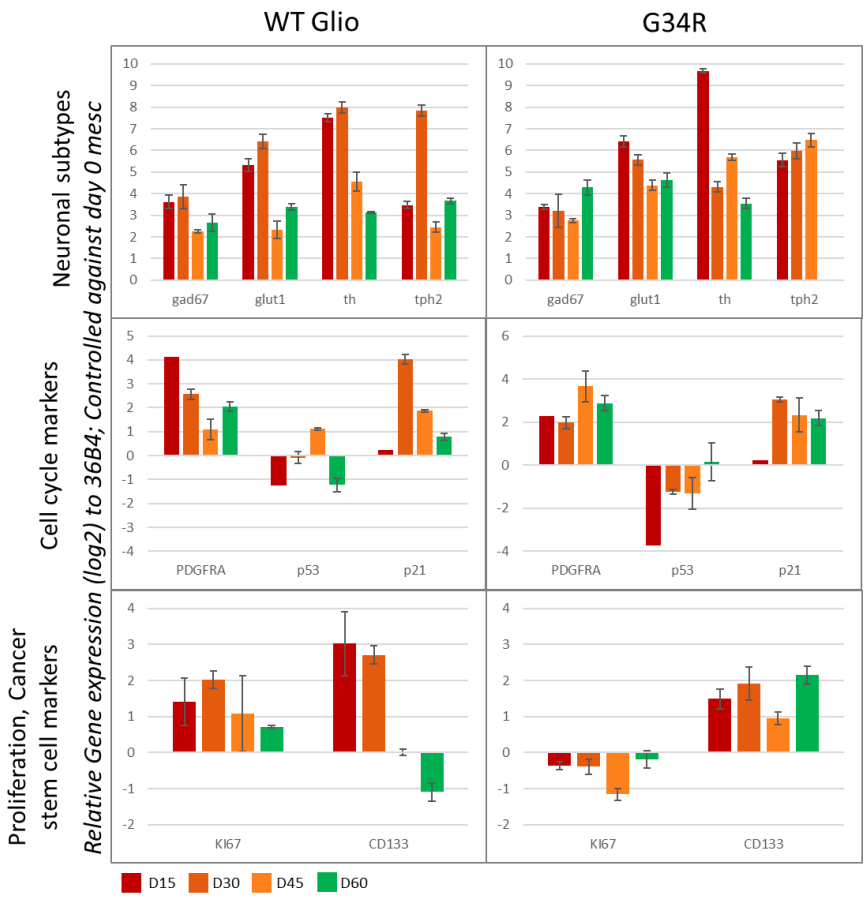


Figure 32. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 2) gene expression analysis

Day 15, 30, 45, 60.

Conclusion

The modified protocol provided a more reproducible protocol for the generation of mouse brain organoids, which show visible neural folds (Figure 30a) at day 10 of brain organoid generation. Moreover, there continues to exist the possibility of an influence of the G34R mutation on certain gene expression profiles such as Olig2, P53 and CD133. To elucidate an in-depth view into the expression profiles of our organoids, we performed a bulk transcriptomics comparing 2 month WT Glio and G34R mouse brain organoids.

Bulk transcriptomics analysis of WT Glio and G34R brain organoids

We sequenced 2-month-old mouse brain organoids derived from the WT Glio and G34R cell lines and performed a gene ontology analysis on subsets of upregulated and downregulated genes (Figure 33).

From the analysis of the common upregulated genes in both cell lines, we were able to confirm the neuronal differentiation of the organoids, showing signatures of dorsal striatum, amygdala, olfactory bulb, and cerebral cortex along with presence of astrocytes, glial cells, and neurons. Analyzing specifically the upregulated genes in G34R versus WT Glio organoids, we find a significance for the P53 pathway.

Conclusion

We know from clinical representation studies that glioblastoma patients with a H3.3 mutant phenotype usually also frequently display mutation in the TP53 [98,100,103,109,110]. Additionally, with the organoids generated till this time point we weren't able to notice an overexpansion of tissue that would be indicative of tumor infiltration and proliferation. This can be due to the fact that H3.3 mutations, while being described as driver mutations, require additional mutational profiles in order to express a tumor morphology. With this is mind, we proceeded to create a TP53 knockdown model in the G34R cell line, to unlock a possible restrictive cell cycle regulation point that might be inhibiting the earlier stages of tumorigenesis.

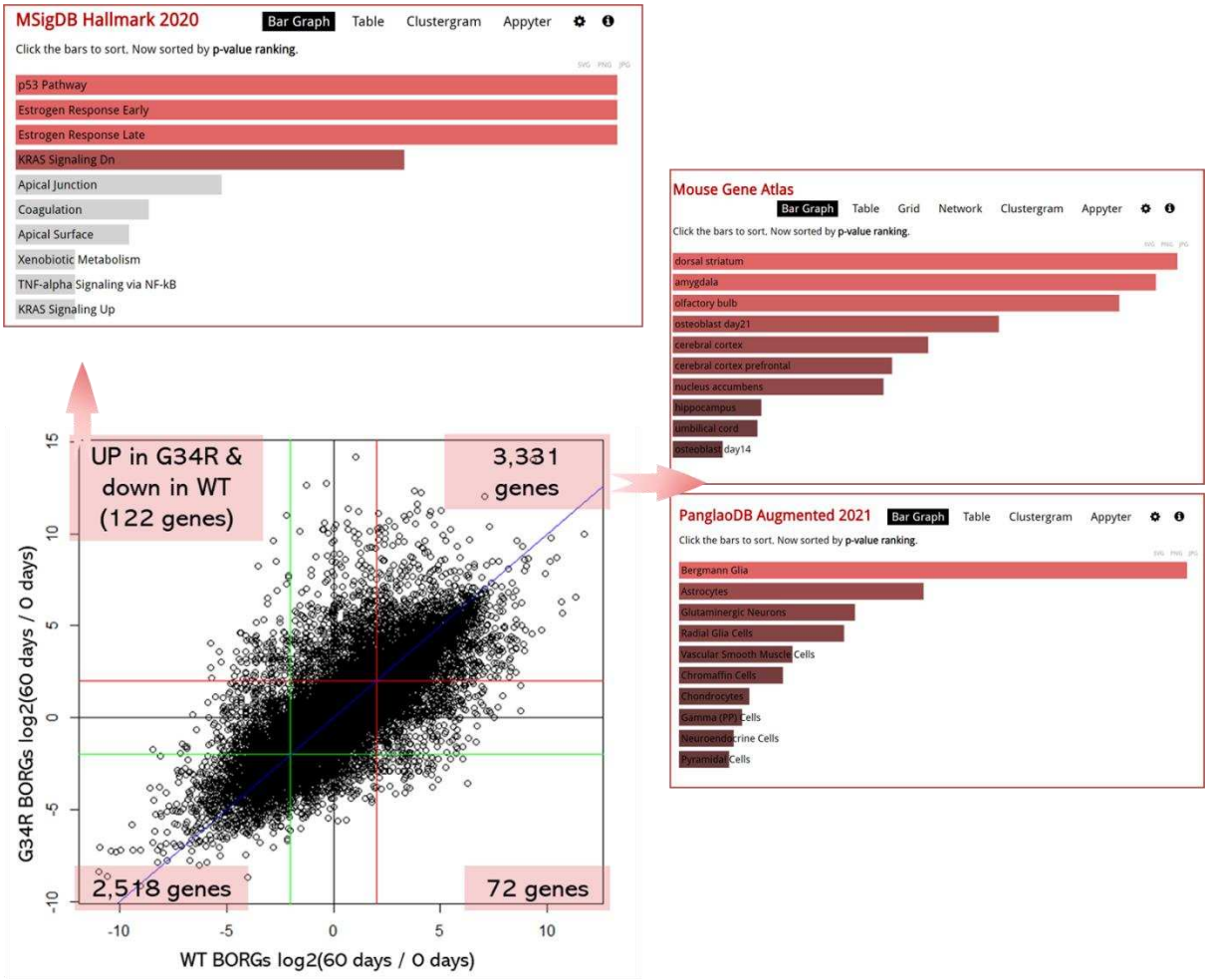


Figure 33. Bulk Transcriptome analysis of WT Glio versus G34R 2 month-old brain organoids

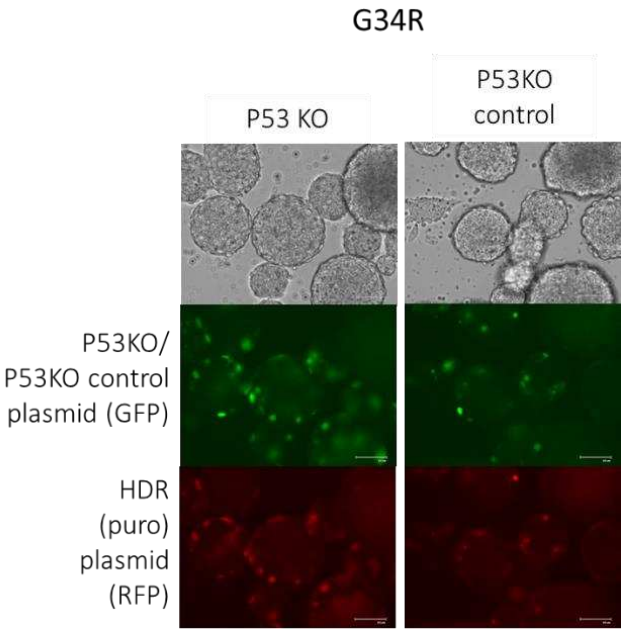
with a Gene ontology analysis of the upregulated genes in G34R organoids only, and for common upregulated genes in organoids.

TP53 Knockout of G34R cell line

The G34R cell line was selected to be transfected with a P53 KO plasmid to compare the tumorigenic effect of a G34R P53KO versus untransfected G34R- derived brain organoids. The cells were transfected using lipofectamine 2000 transfection solution, 1 day after cells were seeded at 100,000 cells/well and reached at least 50% confluency. The protocol is described in detail in Material and Methods: Trp53 plasmid Lipofectamine transfection.

Several trial experiments were performed to test different values of the plasmid to lipofectamine ratio, such as 1:8, 1:6, and different quantities of Optimem solution but these were inefficient transfection experiments or too toxic causing the cells to die within 1 day (data not shown). We successfully managed to create TP53 knockout cells in G34R cells with a ratio of Plasmid to Lipofectamine 2000 as 1:4, and keeping the total volume of the transfection solution to less than 50µl in a total volume of DMEM basal medium of 1ml. 24H later, the transfected cells can be visualized with a P53KO/control construct (GFP tag) and a HDR construct (RFP tag) (Figure 34a). The cells were subjected to a puromycin selection for 3 days at 7µg/ml, before being verified for a TP53KO profile (Figure 34b), and then amplified and banked.

34a



34b

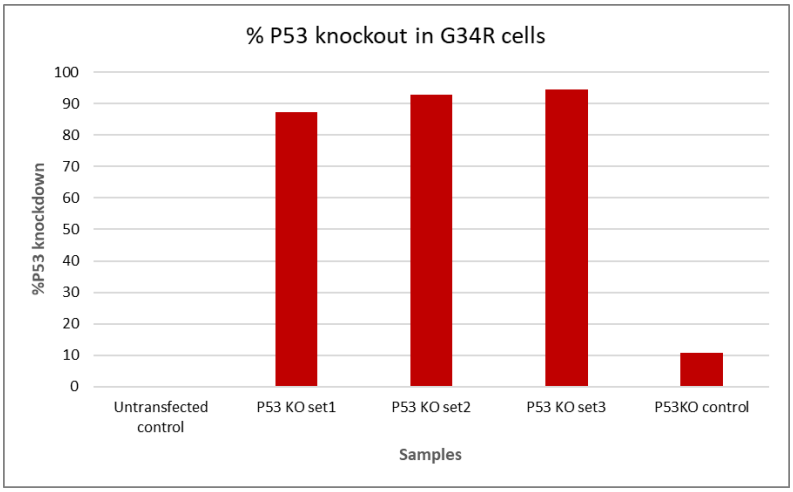


Figure 34. TP53 Knockout of G34R cell line

a) Verification of TP53 transfection in G34R MESC cells with P53KO and control plasmid (GFP), HDR plasmid (RFP); b) Validation of TP53 knockout in G34R MESC cells (3 sets of transfected cells) and the untransfceted control and negative control with P53KO control plasmid.

G34R P53KO brain organoid

The G34R cell line along with its isogenic P53KO and the control plasmid transfected cell line were generated using the Mouse brain organoid generation version 2 for a period of 3 months (Figure 35). We begin to notice a size difference starting from Day 30 of culture, with the G34R P53KO BORGs generally appearing to be significantly larger than the control organoids (Figure 36, Supplementary Figure 10). At the end of 3 months of growth, the P53KO BORGs are 0.5mm larger than the control BORGs.

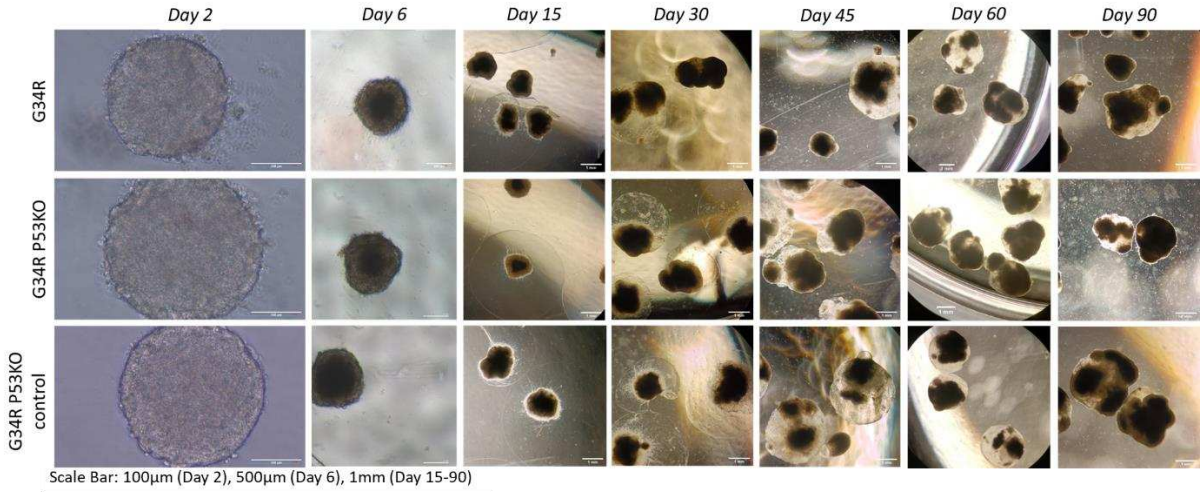
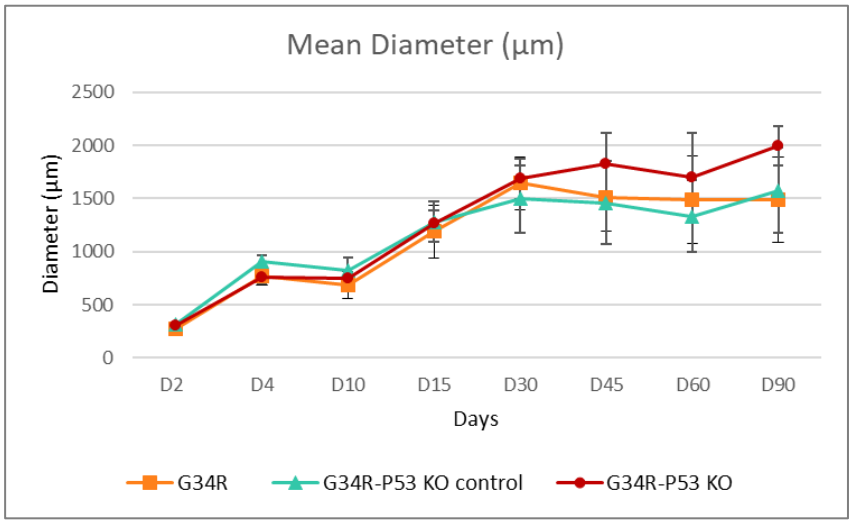


Figure 35. G34R P53KO brain organoid brightfield images
derived from MESC of G34R, G34R P53KO, and G34R P53KO control cells

36a



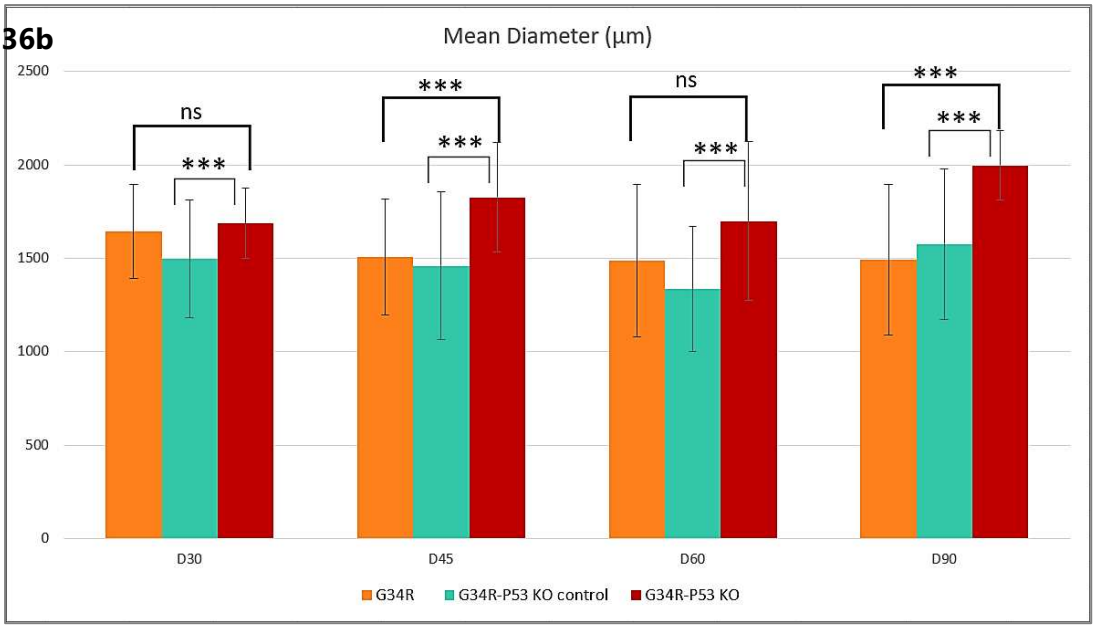
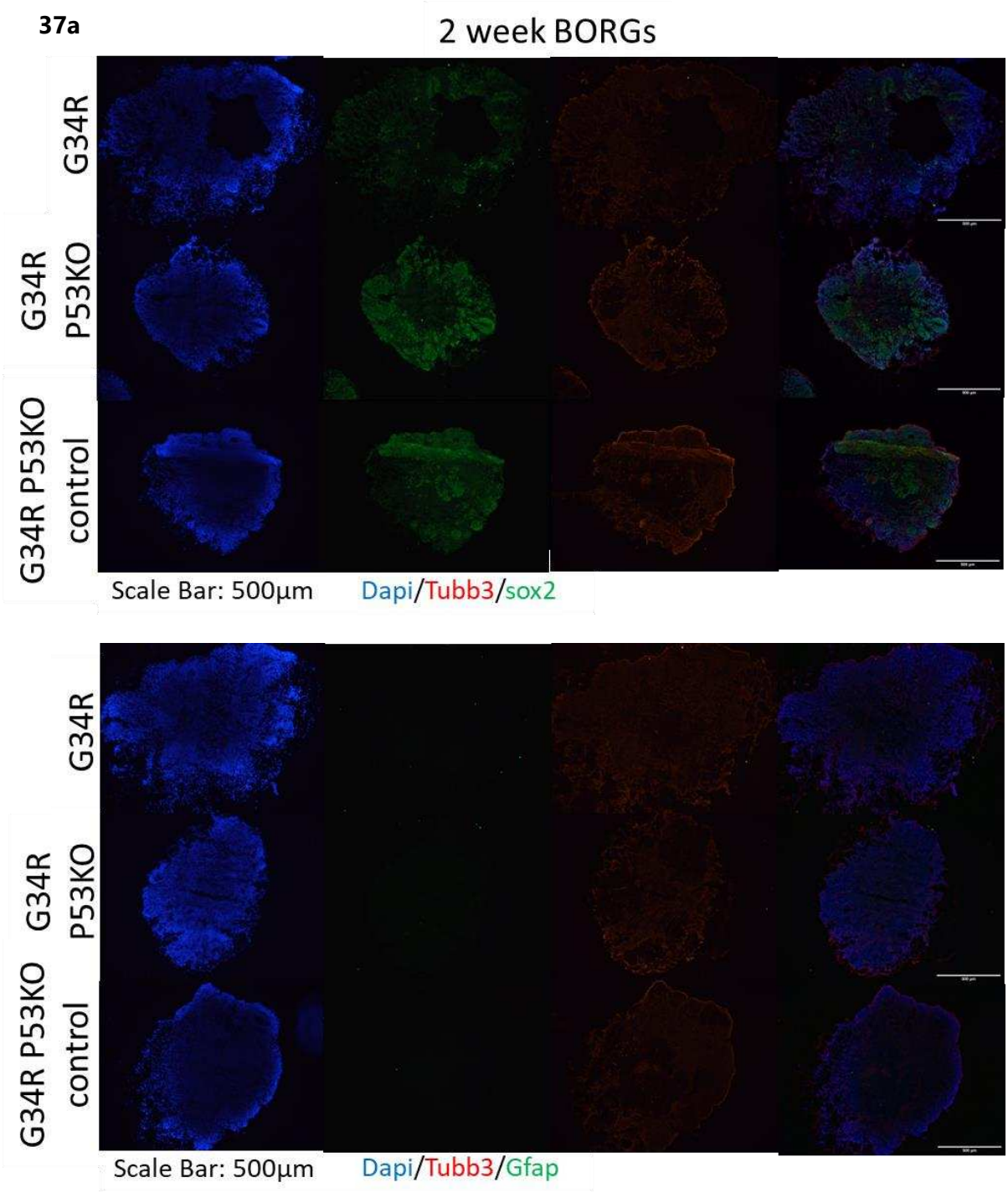


Figure 36. G34R P53KO brain organoid size measurement

a) Plot of mean diameter of brain organoids generated for all 3 cell lines; b) Barplots of mean diameter measurements of G34R, G34R P53KO, and G34R P53KO control brain organoids from Day 30 to Day 90 where data is presented as Two sample, Equal Variance t-Test, ns=no significance, ***p<0.001;

An immunofluorescence staining of 2-week old brain or ganoids reveals that there is a robust Sox2 expression coupled with moderate staining for Tubb3 (Figure 37a, Supplementary Figure 11). At this stage, we don't see any expression of Gfap, which is expected since astrocytes appear later on in development (Figure 37a). There is also a dispersed expression of KI67 and spotted regions of CD133 expression in the organoids (Figure 37b, Supplementary Figure 11). At 1 month, Tubb3 expression has reduced considerably, while Sox2 persists (Figure 38a, Supplementary Figure 12). There is a strong expression for Gfap and Gad67 by this time point, and KI67 and CD133 expression is unchanged (Figure 38b, Supplementary Figure 12). By 2 months, Sox2 expression is more visible to the edges of the organoid and not as strongly expressed, while Gfap expression can be seen in tracts around the outer rims of the organoids (Figure 39a, Supplementary Figure 13). KI67 staining is visible in spots all over the organoid and there are patches of CD133 positive cells in all 3 cell line derived organoids (Figure 39b, Supplementary Figure 13). In 3 month old organoids, we can find reemergence of a robust signal for Sox2 expression all over the organoids, and a moderate Ctip2 (postmitotic neurons) signal (Figure 40a, Supplementary Figure 14). KI67 expression is still present all over the

organoids, and slightly more dense in the edges, while CD133 expression is present as always in niches (Figure 40b, Supplementary Figure 14).



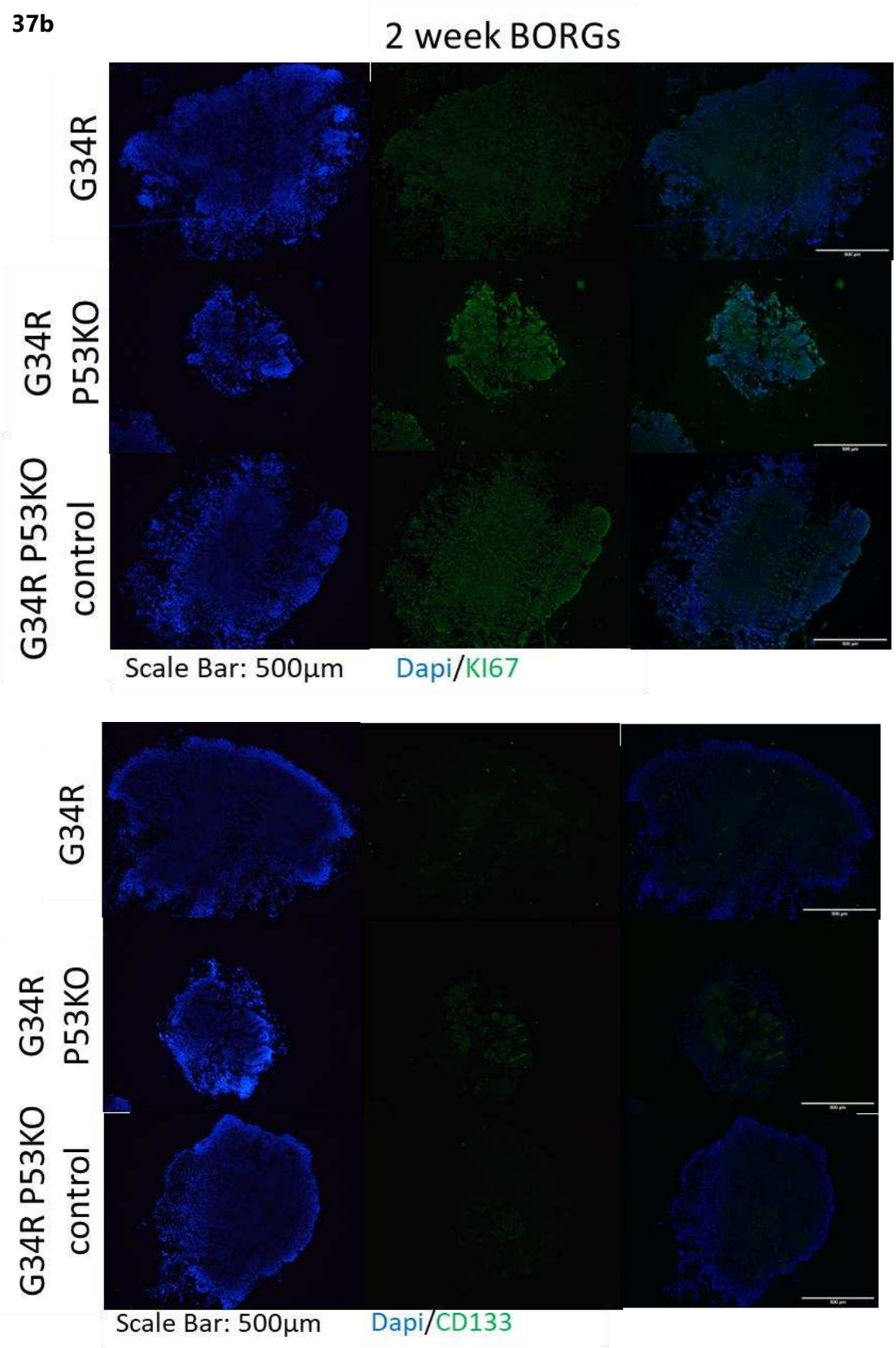
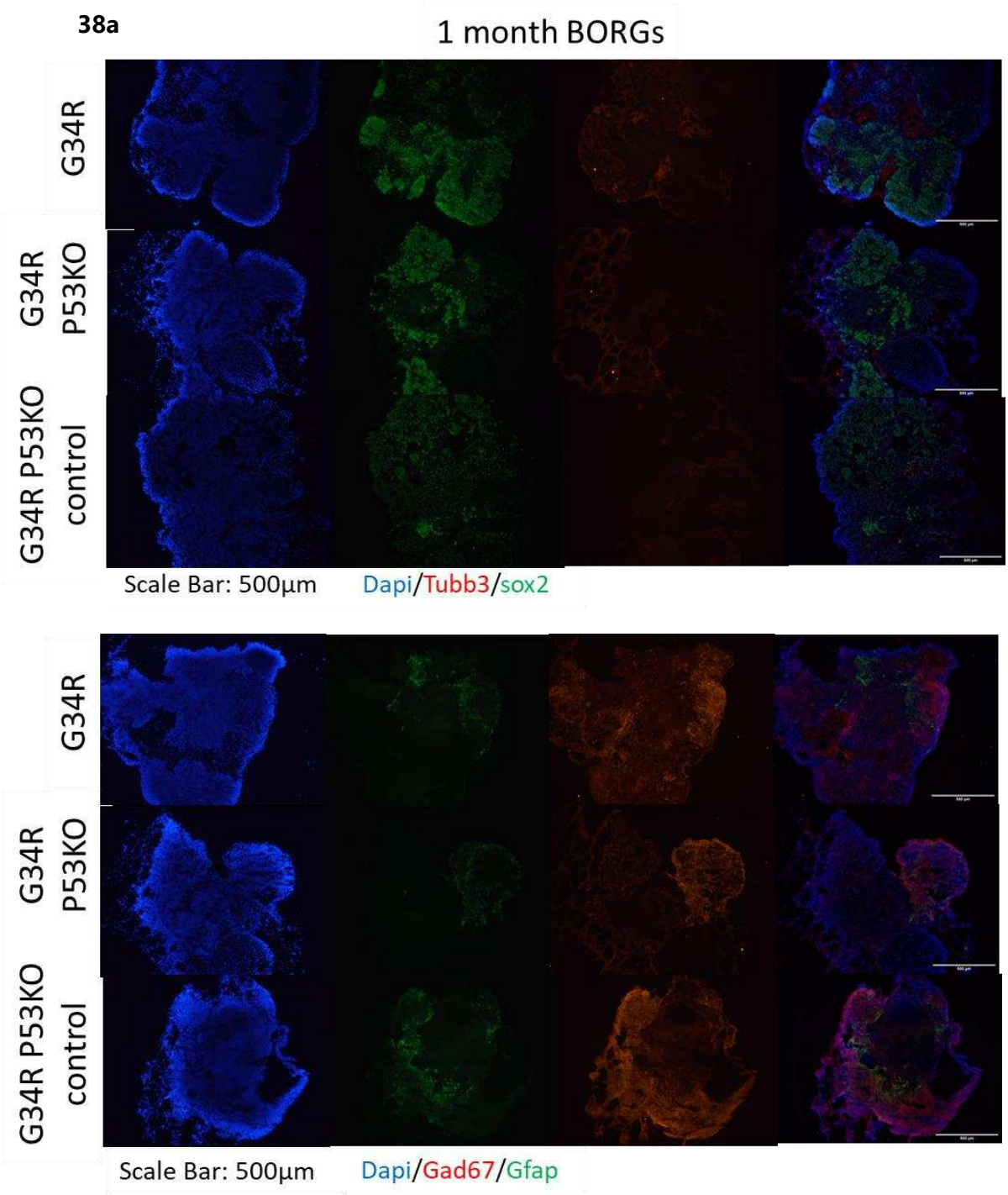


Figure 37. G34R P53KO brain organoid immunofluorescence staining

2 week old brain organoids for all 3 cell lines, with markers a) Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), b) KI67 (proliferation), CD133 (Cancer stem cell)

38a


1 month BORGs
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 500µm
Dapi/Tubb3/sox2
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 500µm
Dapi/Gad67/Gfap

38b

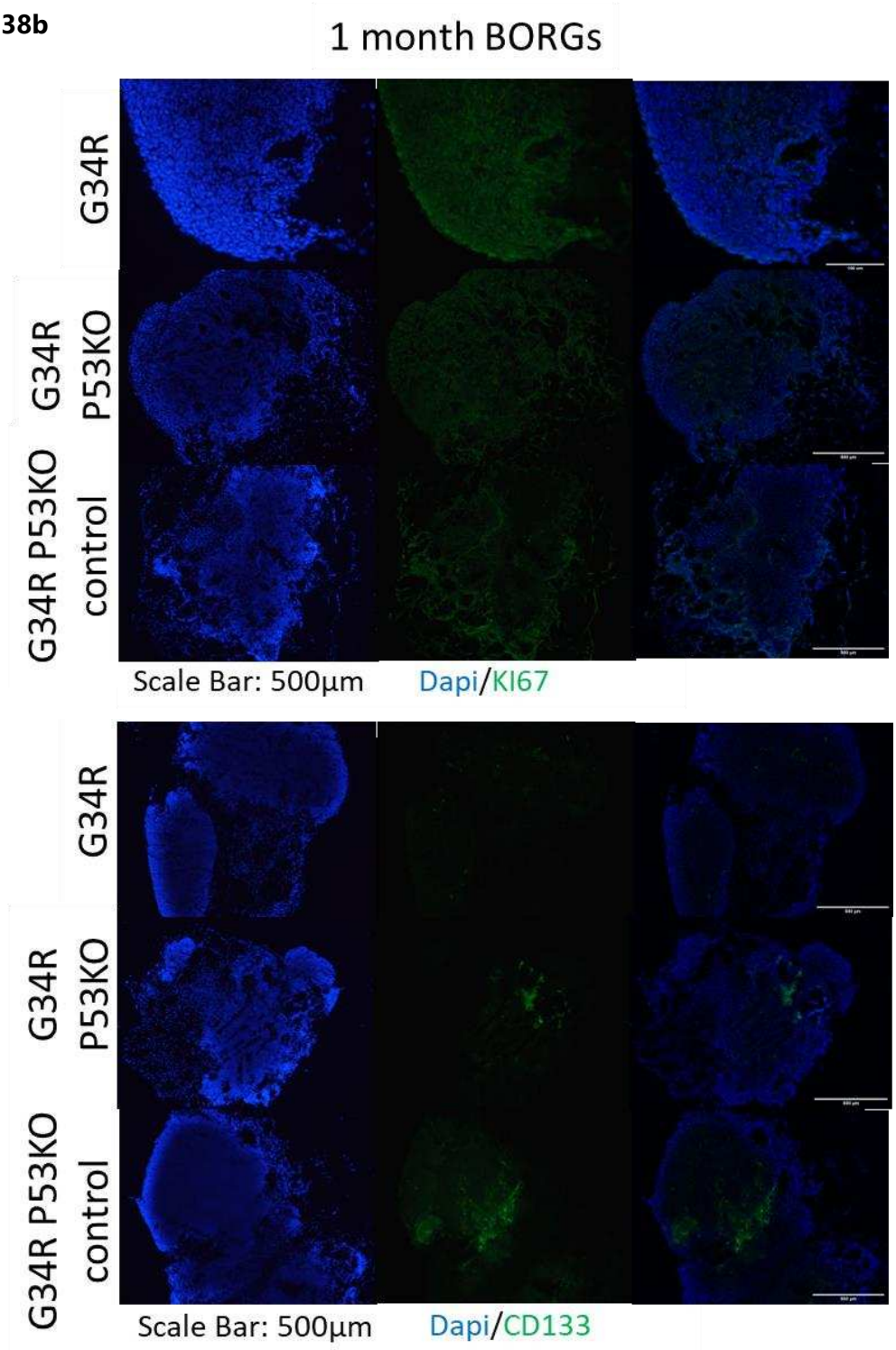
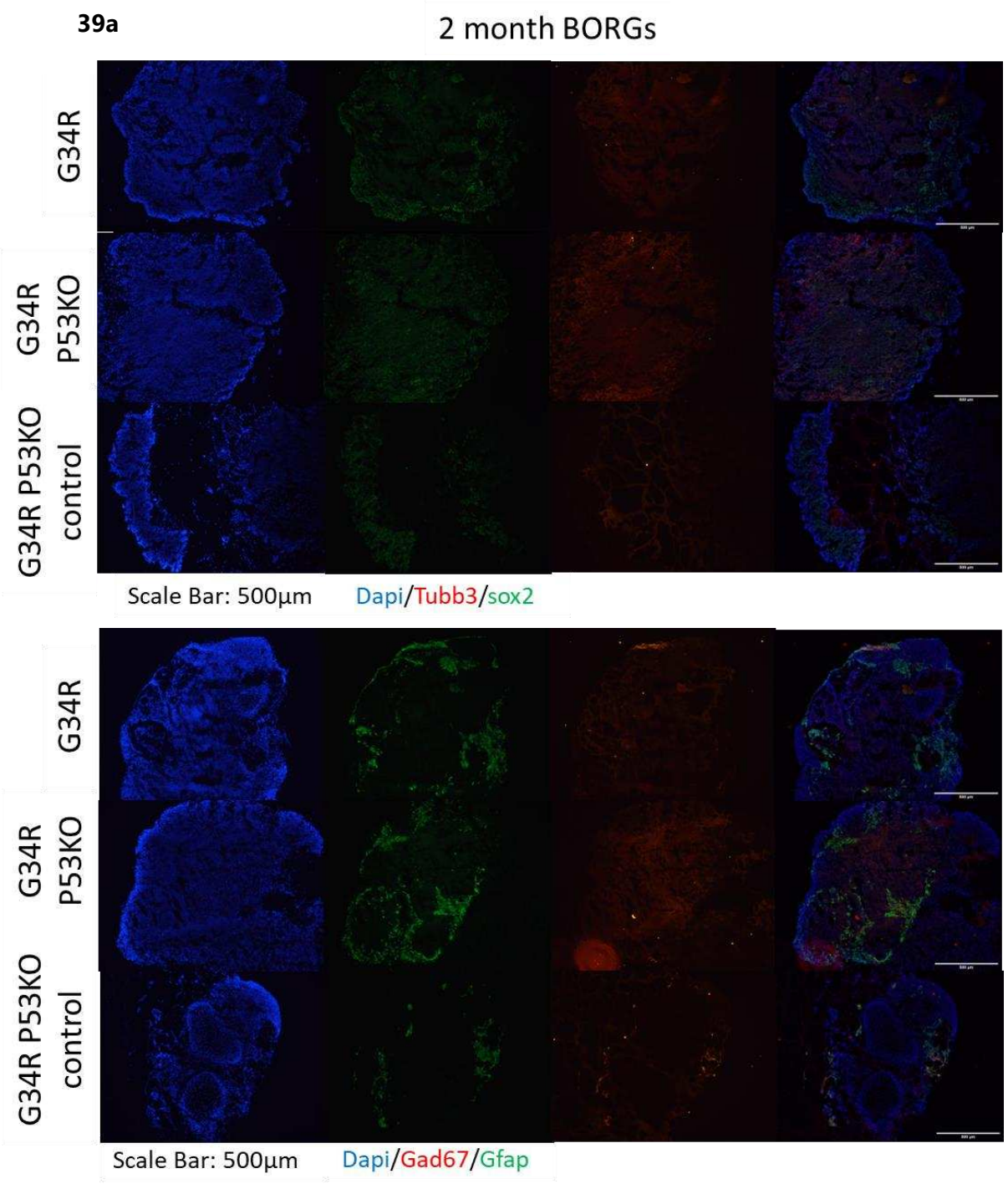


Figure 38. G34R P53KO brain organoid immunofluorescence staining

1 month old brain organoids for all 3 cell lines, with markers a) Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), b) KI67 (proliferation), CD133 (Cancer stem cell)

39a


2 month BORGs
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 500µm
Dapi/Tubb3/sox2
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 500µm
Dapi/Gad67/Gfap

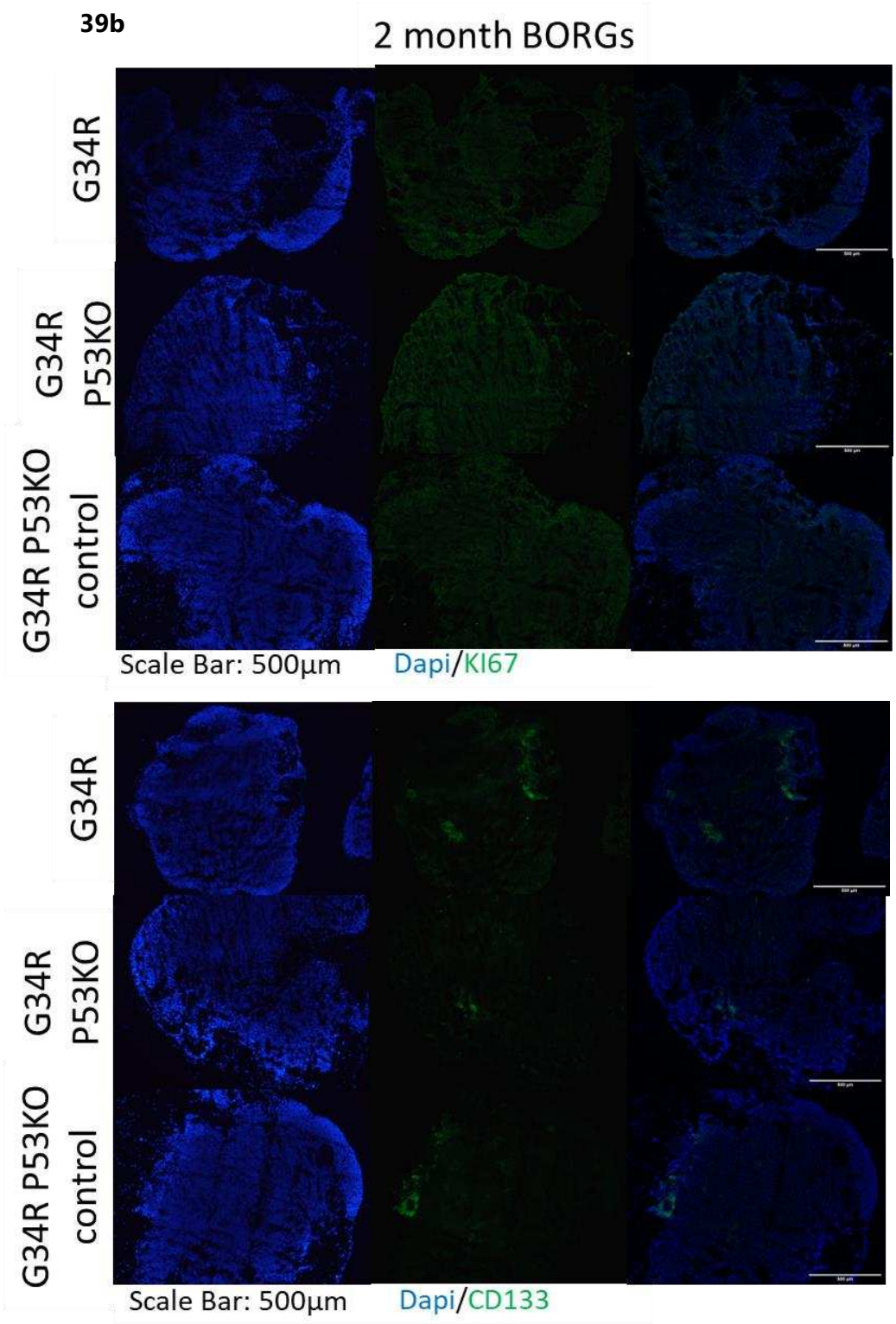
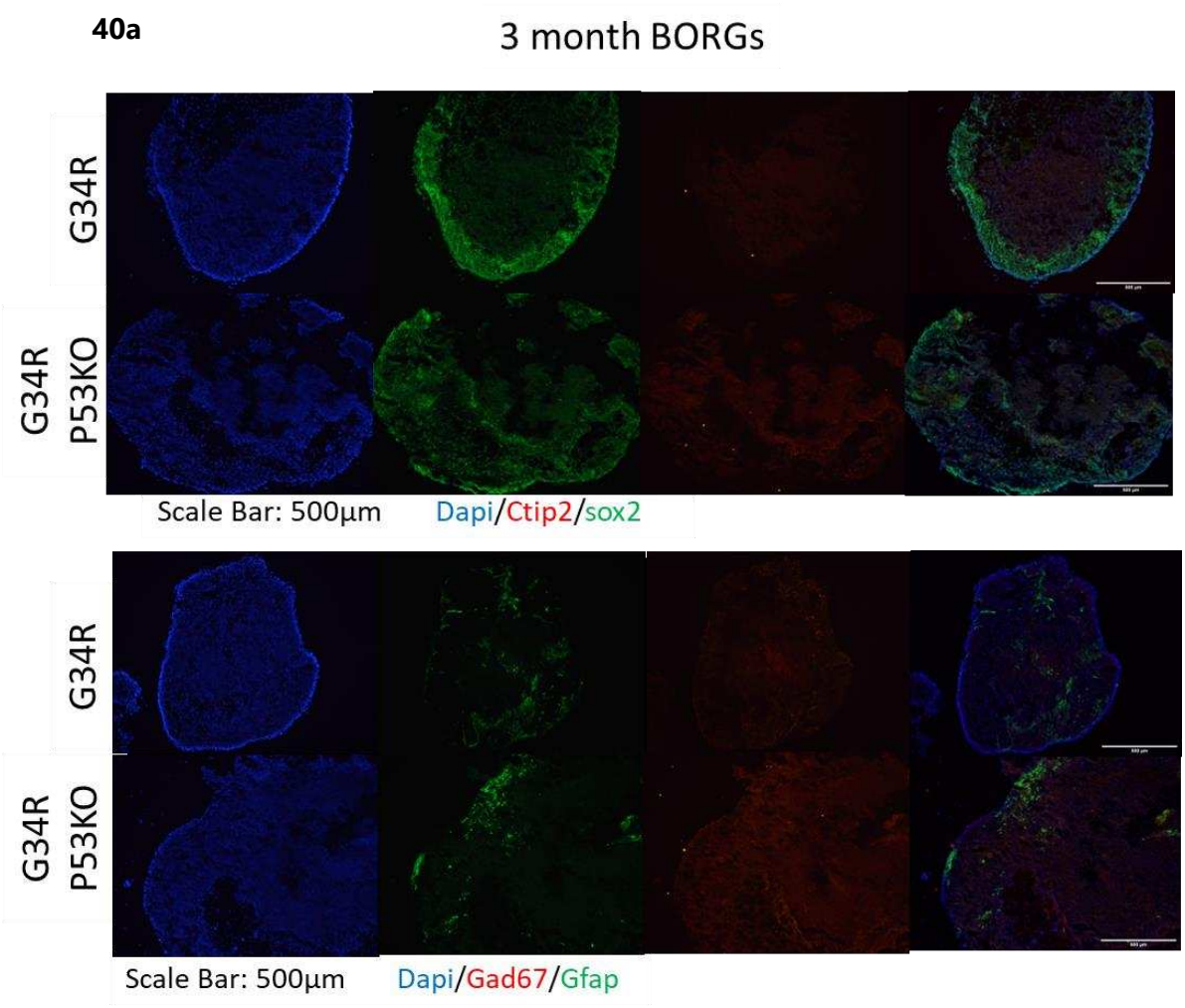


Figure 39. G34R P53KO brain organoid immunofluorescence staining

2 month old brain organoids for all 3 cell lines, with markers a) Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), b) KI67 (proliferation), CD133 (Cancer stem cell)

40a


3 month BORGs
G34R
G34R P53KO
Scale Bar: 500µm
Dapi/Ctip2/sox2
G34R
G34R P53KO
Scale Bar: 500µm
Dapi/Gad67/Gfap

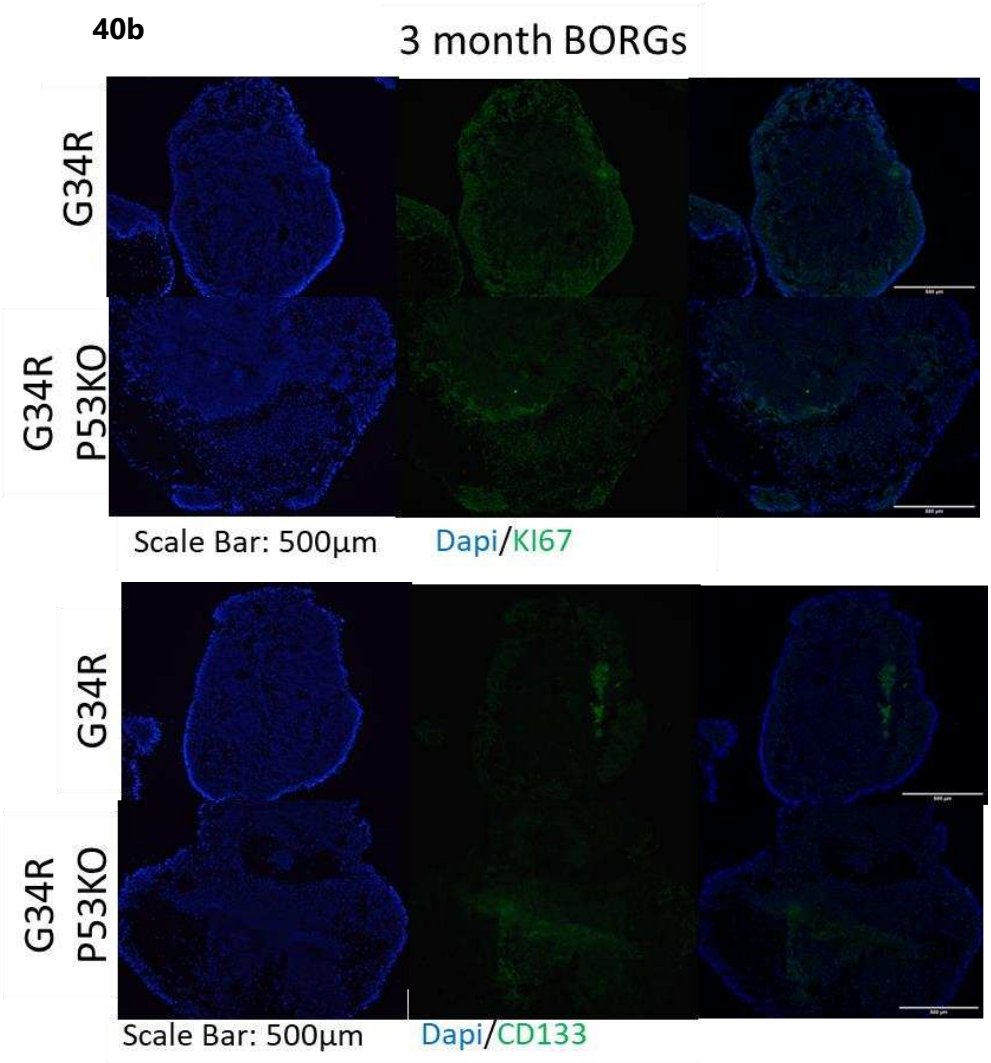
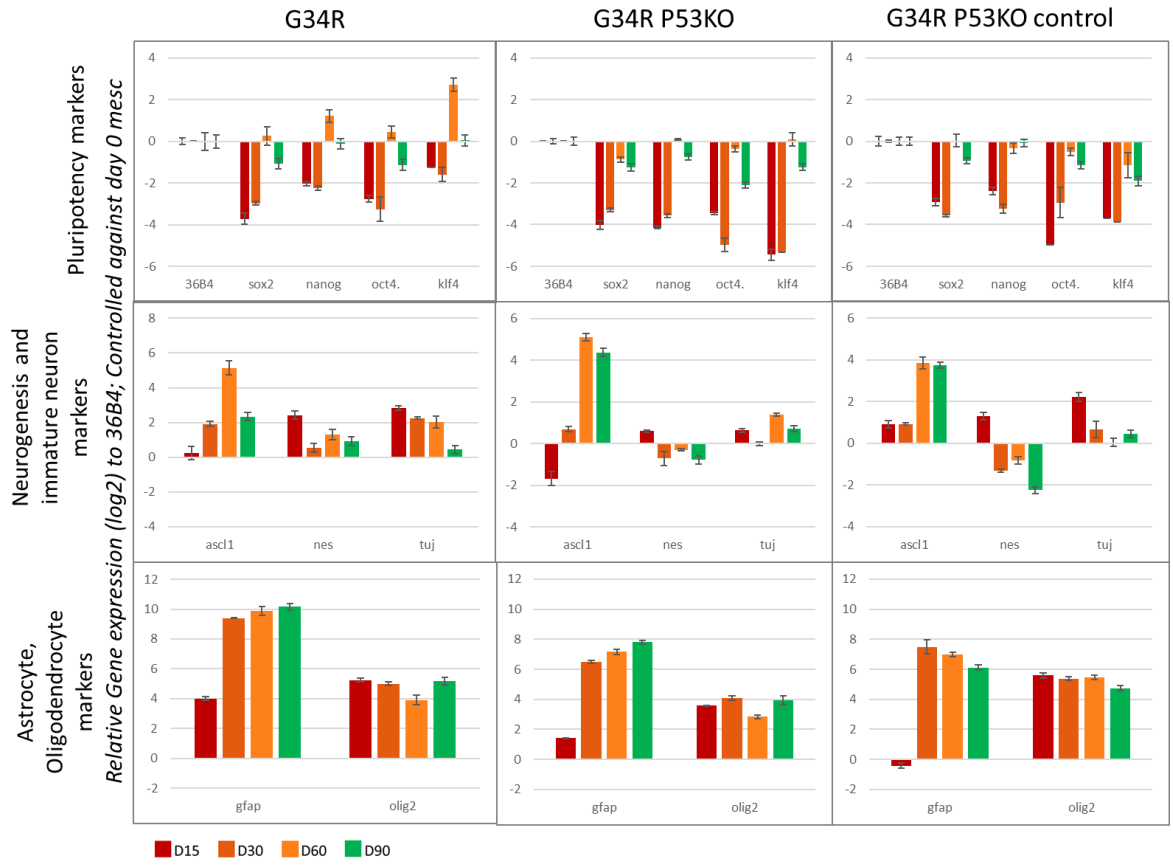


Figure 40. G34R P53KO brain organoid immunofluorescence staining

3 month old brain organoids for all 3 cell lines, with markers a) Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), b) KI67 (proliferation), CD133 (Cancer stem cell)

We confirmed the immunostaining analysis with gene expression assays over a period of 3 months (Figure 41). We noticed that while pluripotency markers were strongly repressed early on as the cells are in a differentiation state, there is a shift towards expression levels close to as seen in MESC cells. All 3 cell lines show a population of neurogenesis positive cells in the organoids. While Gfap expression overall for the transfected cell lines (P53KO and P53KO control) are lower than the non transfected cell line, Olig2 expression is visibly lower for G34R P53KO organoids. Interestingly, the markers for Neuronal subtypes reveals that in G34R P53KO organoids specifically, we no longer see a serotonergic (Tph2) expression compared to the controls. Comparing KI67 and CD133 expression over the 3 cell lines doesn't show any

differences, and there is a transient over expression of Pdgfra seen in 2 month G34R and G34R P53KO organoids, that gets repressed later on.



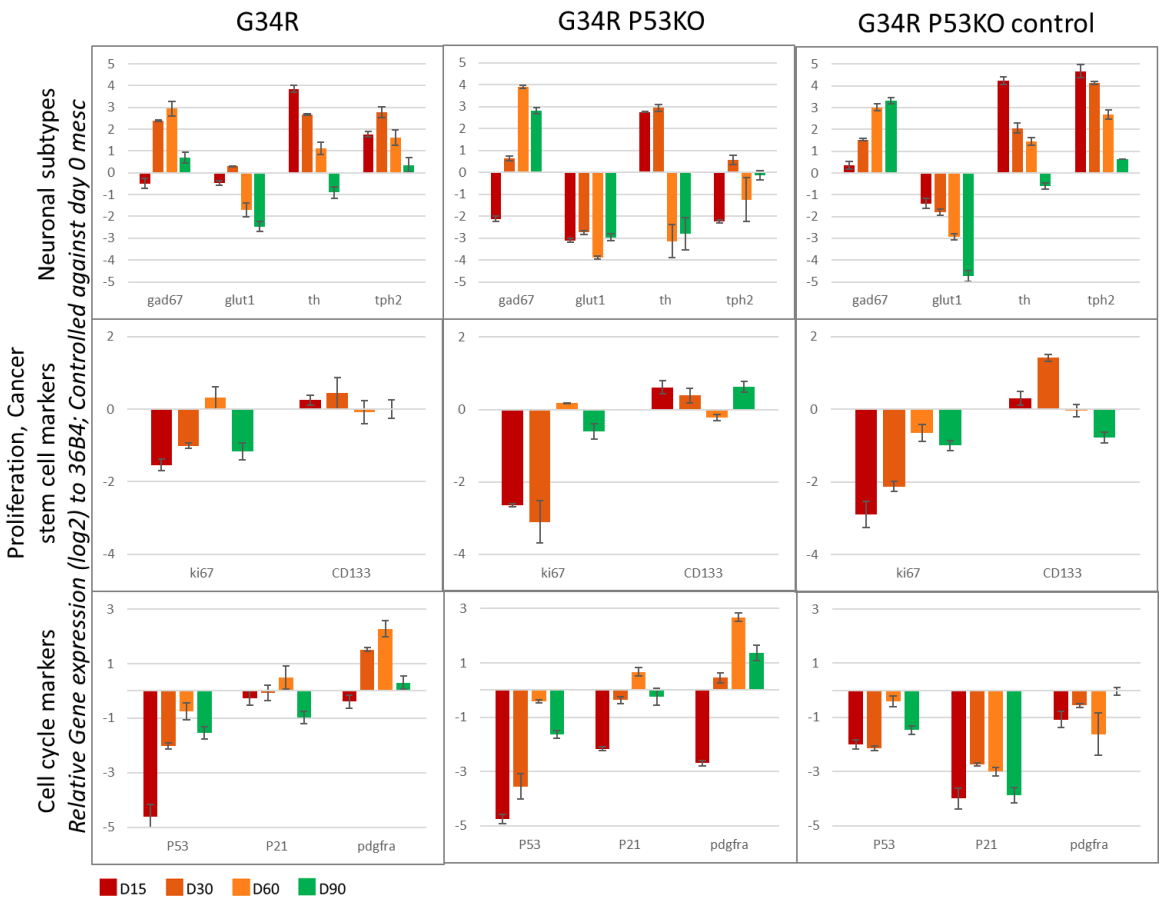


Figure 41. G34R P53KO brain organoid gene expression analysis
Day 15, 30, 60, 90.

Conclusion

We were able to successfully produce TP53KO G34R cells, that are confirmed to be knockout cell lines. We used these cell lines to generate mouse brain organoids for a period of 3 months and can find confirmation of neuronally differentiated organoids (Figure 37, Figure 38, Figure 39, Figure 40) and some indications of a tumorigenic effect, as seen with the reduced repression of pluripotency markers, and variations in Olig2 and Pdgfra expression (Figure 41). More importantly, we find an significantly increased size difference in G43R P53KO BORGs compared to G34R and G43R P53KO control BORGs starting from 1 month of growth (Figure 36).

Apart from TP53 aberrations in Glioblastoma patients, there are also other mutational profiles seen in the Glioblastoma patients, for example Myc amplification. This oncogene

amplification introduced via electroporation in early Brain organoids, showed an expansive proliferation into the rest of the organoid by 1 month [120]. Thus, we proceeded to generate organoids that were electroporated with a Myc over expression plasmid and observe their growth.

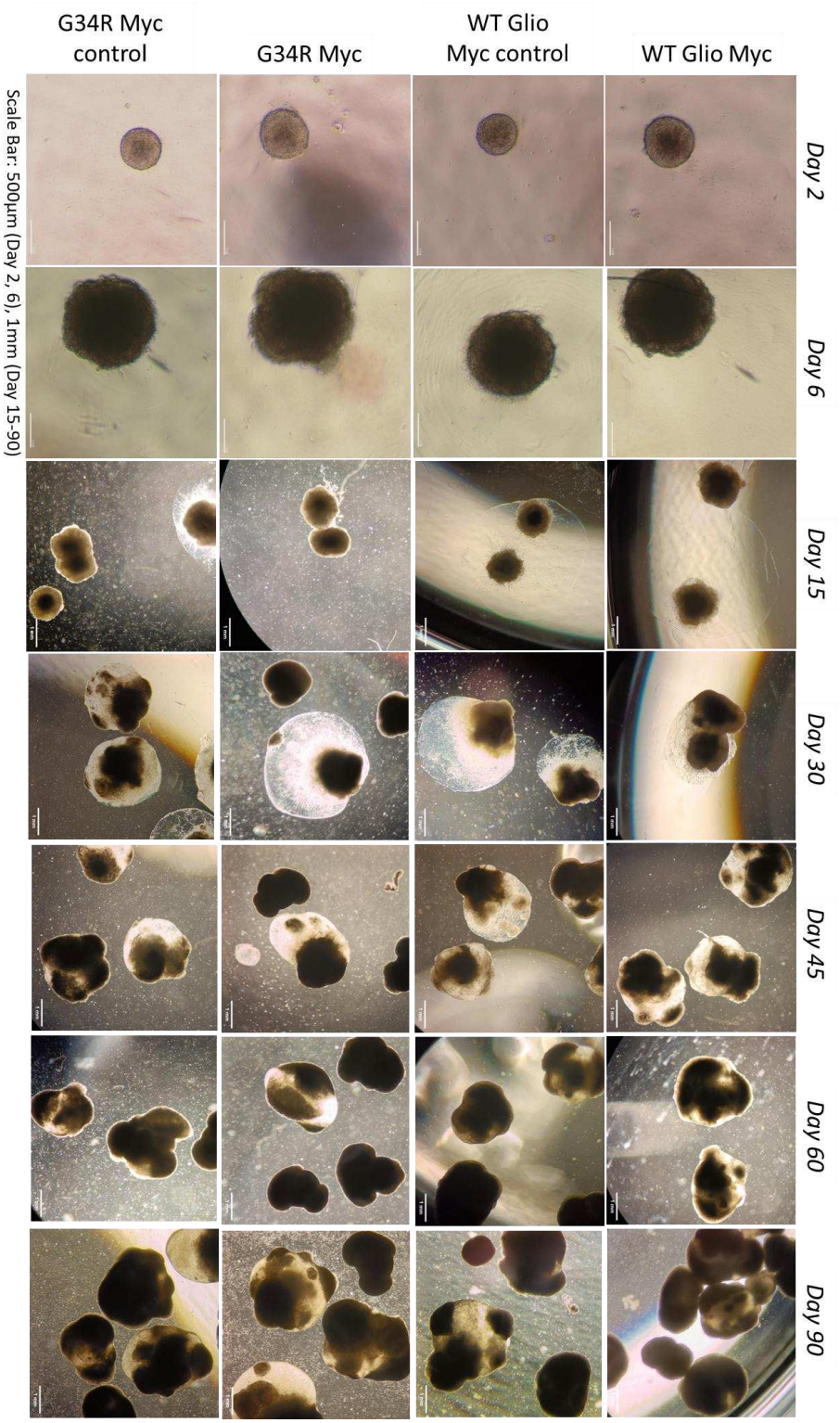
Brain organoids derived from Myc overexpression transfected MESC

The WT Glio and G34R cell line were both transfected with a Myc overexpression or a Control plasmid (Material and methods: Myc plasmid Lipofectamine transfection).

These cells are a heterogenous population, that have not been clonally selected for only Myc positive cells. They were differentiated using the mouse brain organoid generation version 2 protocol, and their sizes tracked over a period of 3 months (Figure 42a). The Myc transfected and Myc control transfected brain organoids reach a maximum mean diameter of 2-2.5 mm over the 3 month period and with no apparent overexpanded tissue (Figure 42b, Supplementary Figure 15).

We analyzed gene expression data at different timepoints (Day 15 - 90) and observed a weaker repression of pluripotency factors in Myc transfected organoids (WT Glio and G34R) and moreover a high expression of stem cell marker Oct4 in G34R Myc organoids, throughout its lifespan (Figure 43). Neurogenesis markers have a higher expression earlier on before reducing over time, as the organoid matures. Astrocyte (Gfap) marker expression appears at 1 month and remains relatively stable, while OPC (Olig2) expression is conversely quite low in expression for G34R Myc transfected organoids. In G34R organoids there is an abrogated expression of Gabaergic neurons (Gad67) and Dopaminergic neurons (Th) (Figure 43). In G34R Myc BORGs, the expression for cell cycle markers is repressed considerably, but surprisingly Myc expression is also downregulated.

42a


G34R Myc control
G34R Myc
WT Glio Myc control
WT Glio Myc
Scale Bar: 500µm (Day 2, 6), 1mm (Day 15-90)
Day 2
Day 6
Day 15
Day 30
Day 45
Day 60
Day 90

42b

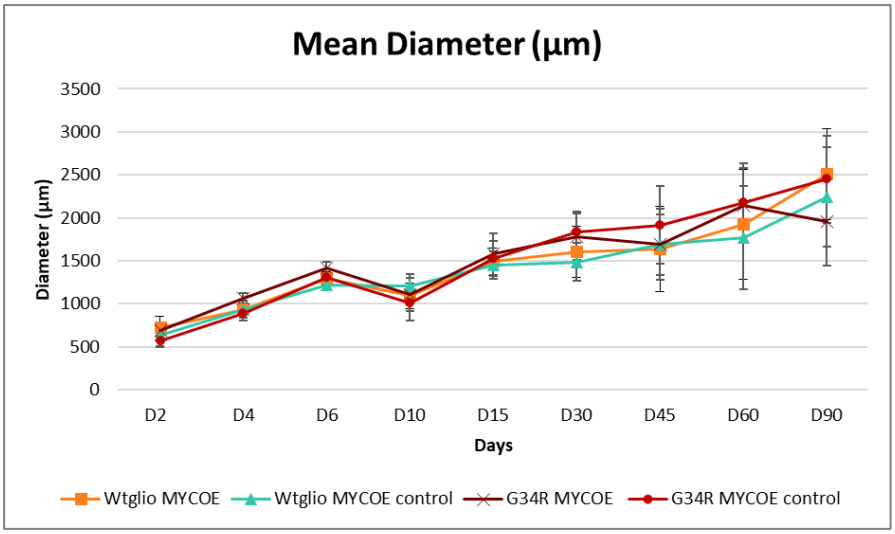
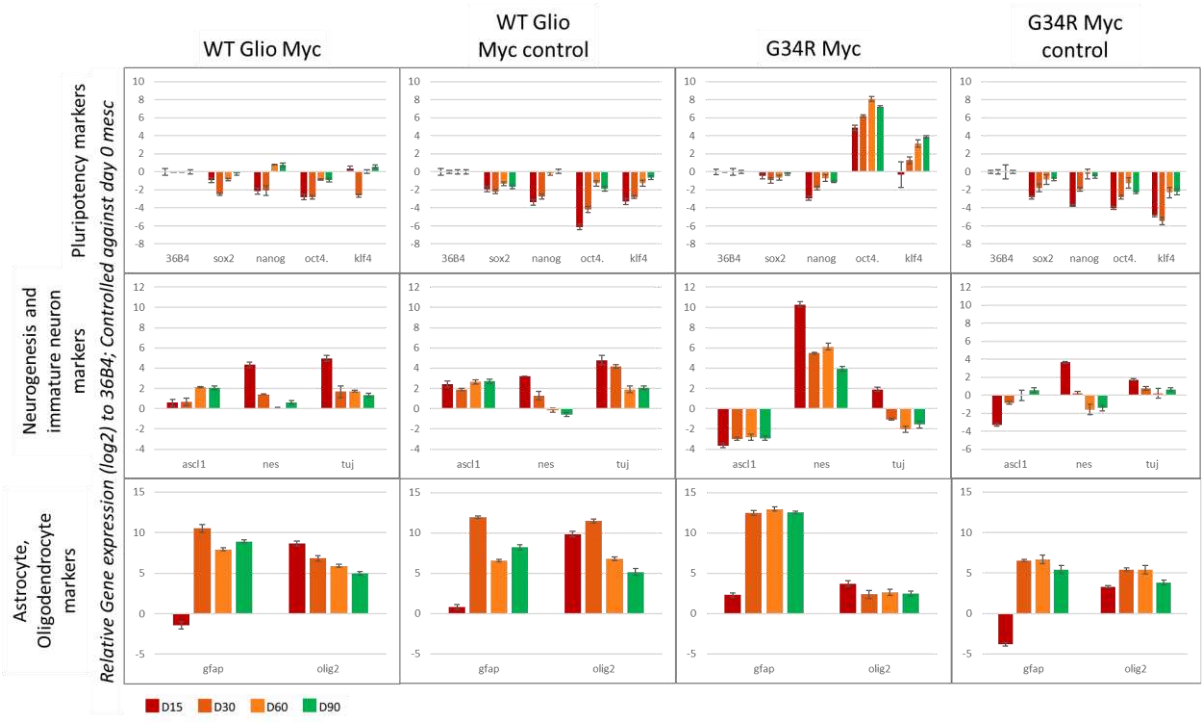


Figure 42. Brain organoids derived from Myc overexpression transfected MESC

a) Brightfield images of mouse brain organoids derived from MESC of WT Glio and G34R cells transfected with Myc or Myc control plasmids; b) Plot of mean diameter of brain organoids generated for all 4 cell lines



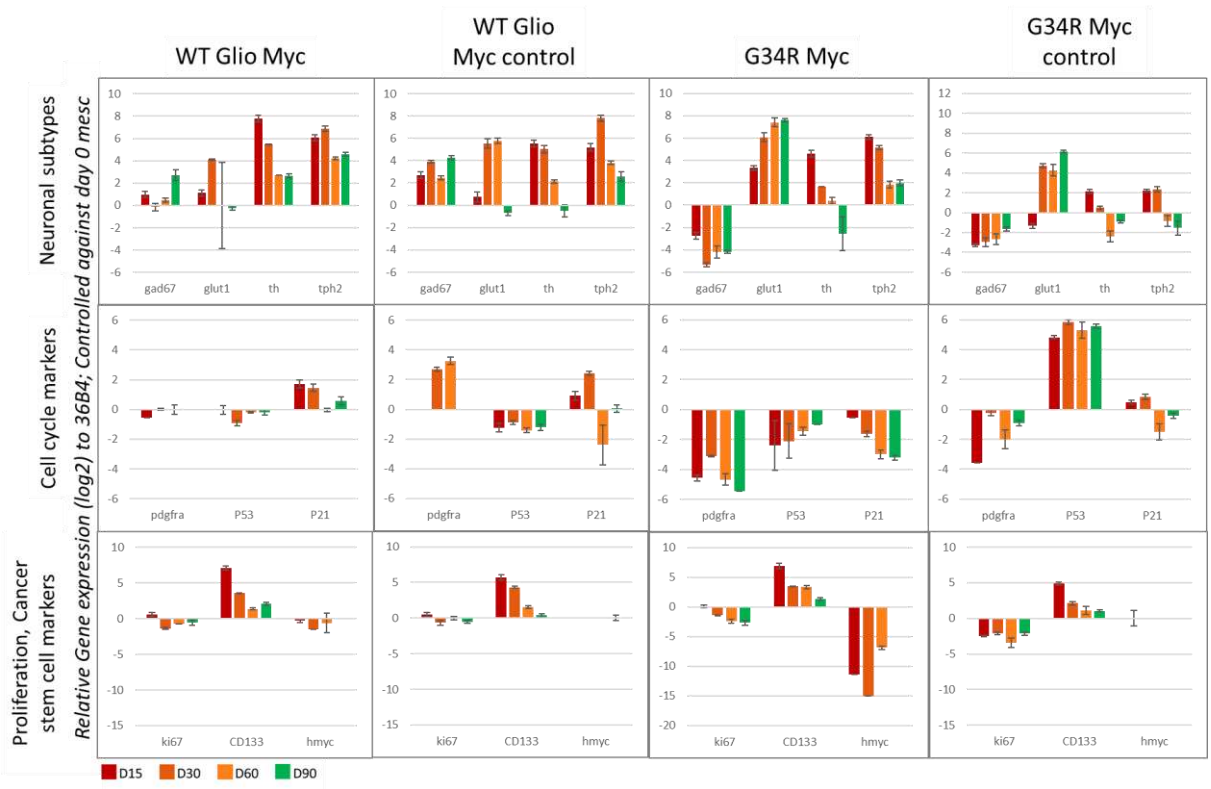


Figure 43. Brain organoids derived from Myc overexpression transfected MESC gene expression analysis

Day 15, 30, 60, 90.

Conclusion

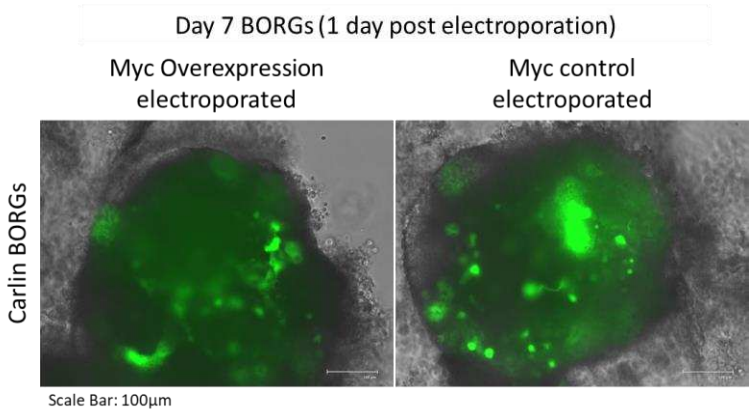
Myc transfected cells that have been differentiated into brain organoids don't show a strong tumor phenotype. There are differences in gene expression levels when comparing to G34R Myc BORGs to the other conditions, indicating an early tumor phenotype that possibly can't overcome the normal cell growth regulation points. In order to mimic the phenotype of Myc electroporation in organoids, we followed the protocol of Bian et al, 2018 [120], and electroporated brain organoids at an early stage of development, instead of electroporating the MESC cells.

Myc electroporated Brain organoids

In Bian et al, 2018, tumorigenic profile was replicated in human brain organoids, by introducing plasmids containing a Myc overexpression gene or commonly associated genetic profiles of Glioblastoma [120]. The growth of the electroporated cells was monitored in these cells using the tracking of GFP tag in the plasmids over a period of 1 month revealing that certain genetic phenotypes contributed to an overexpansion and infiltration compared to others.

In a similar manner, we replicated the study in our mouse brain organoids derived from Carlin MESC cells, and at Day 6 of differentiation, electroporated EBs with a Myc or Myc control plasmid (Materiams and Methods: Myc plasmid Electroporation). They were monitored primarily with immunofluorescence assays and gene expression levels instead of a GFP tag since the fluorescence tag gets diffused after 3 to 4 days of electroporation and doesn't appear distinct from the rest of the organoid (Figure 44a). The organoids have a comparable growth profile over the course of 3 months, not showing any differences (Figure 44b), and an immunostaining profile of Carlin Borgs electroporated with Myc and Myc control plasmids, show expression for Sox2 (neuronal stem cells), Tubb3 (immature neurons), Gfap (Astrocytes), Gad67 (GABAergic neurons), confirming a Neurogenic profile (Figure 45, Figure 46, Figure 47, Supplementary Figure 16).

44a



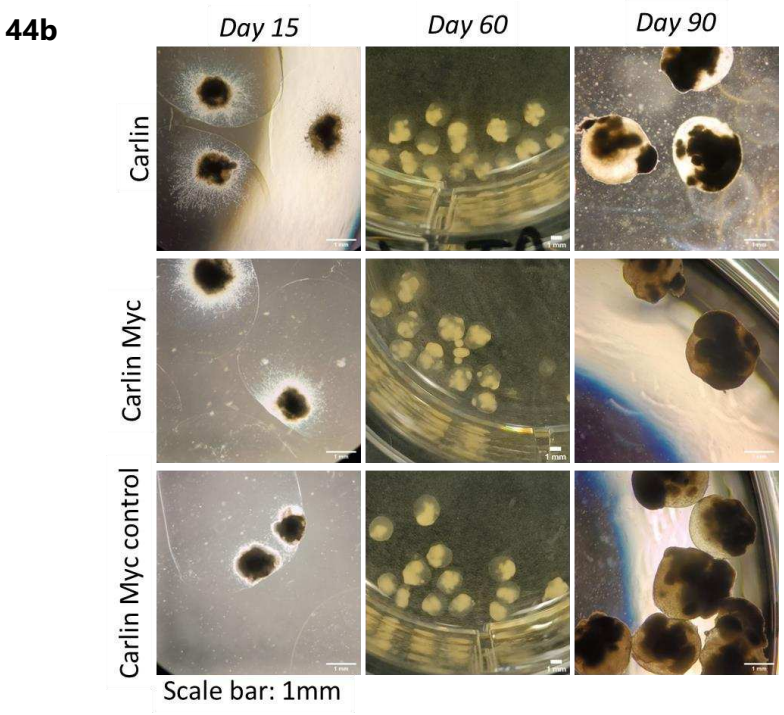


Figure 44. Myc electroporated Brain organoids

a) GFP tag visualization in BORGs 1 day post electroporation confirming successful electroporation of Myc plasmid; b)Brightfield images of mouse brain organoids derived from MESC of Carlin cell line from Day 15 to Day 90, and subjected either to no electroporation or electroporation with Myc or Myc control plasmid

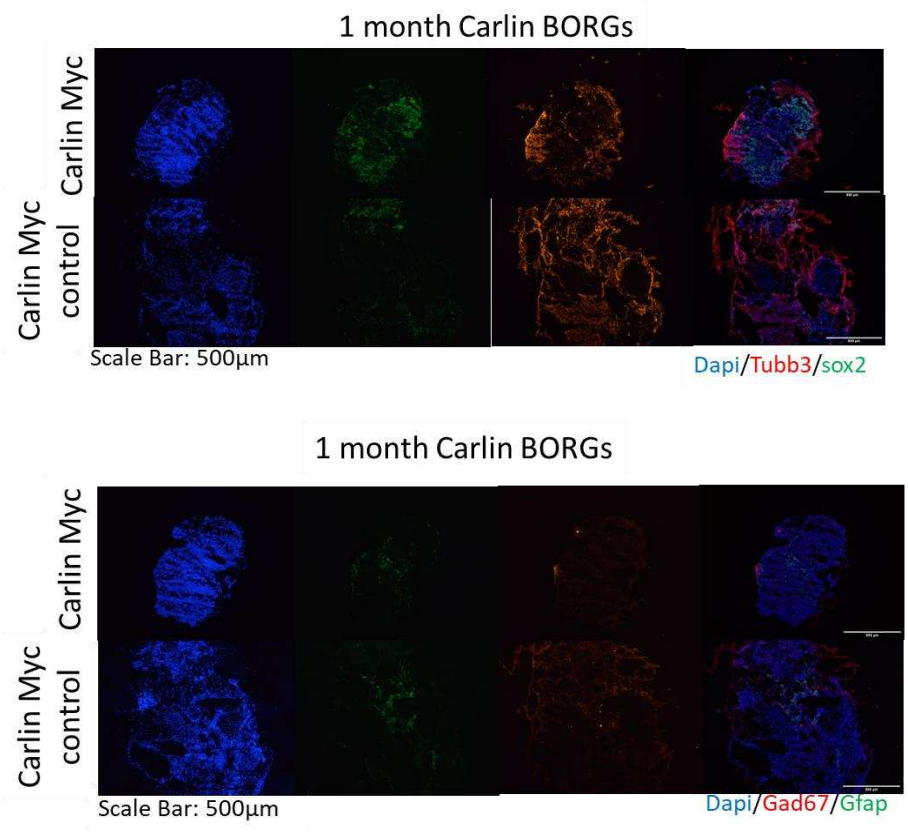


Figure 45. Myc electroporated Brain organoids immunostaining of 1 month BORGs

Staining with antibodies against Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes),and Gad67 (GABAergic neurons)

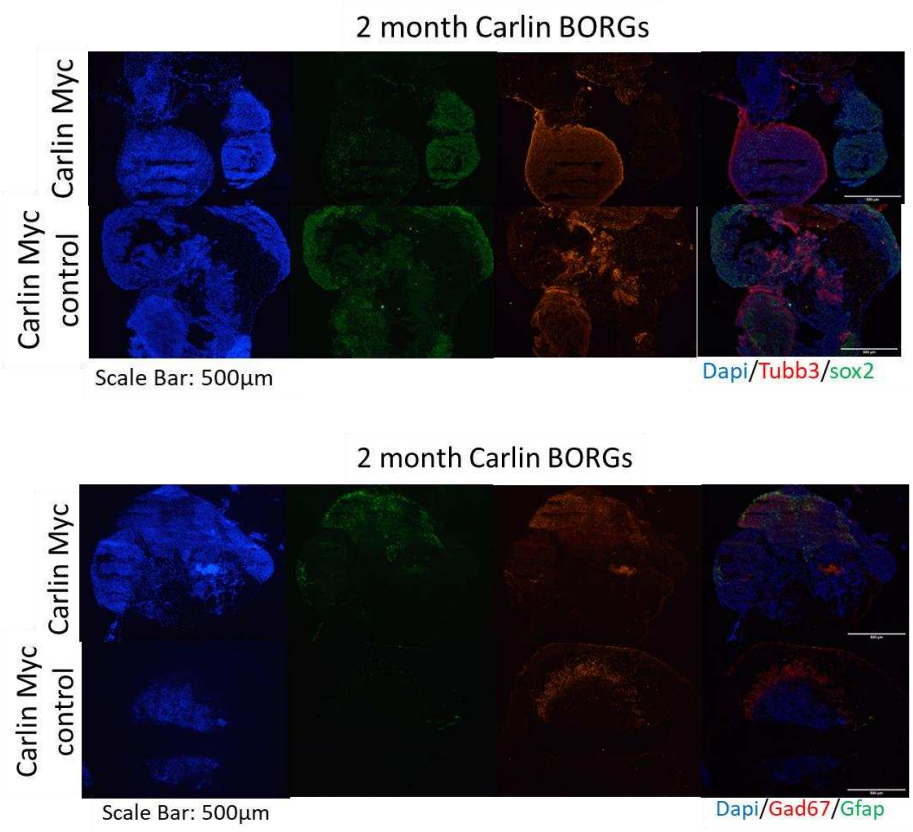


Figure 46. Myc electroporated Brain organoids immunostaining of 2 month BORGs

Staining with antibodies against Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes),and Gad67 (GABAergic neurons)

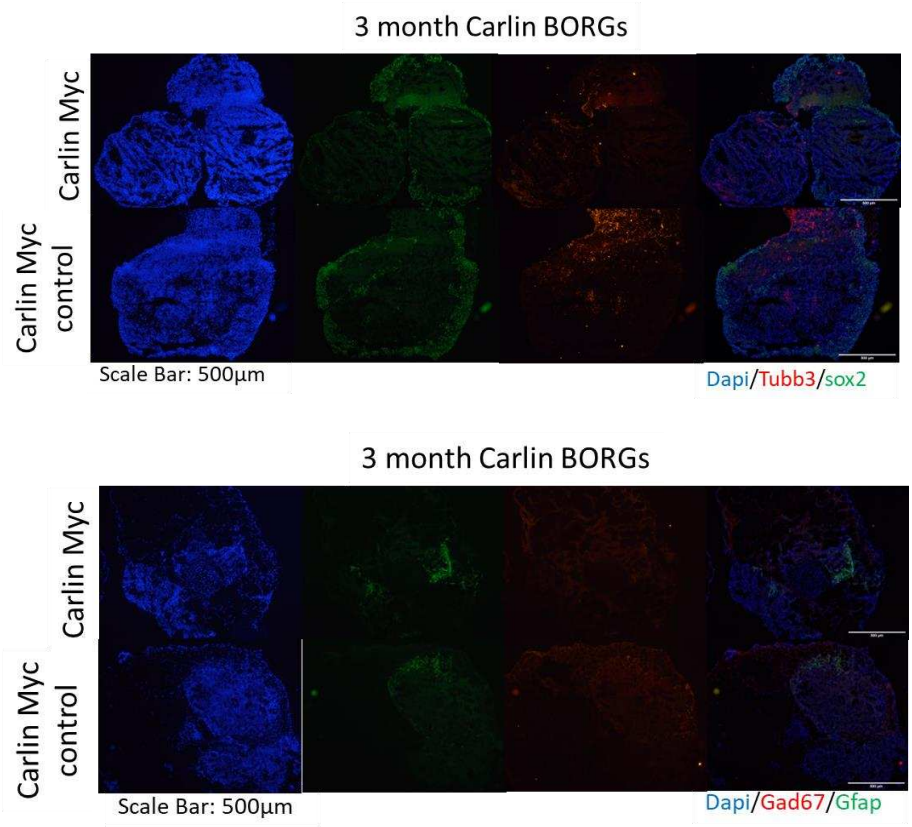


Figure 47. Myc electroporated Brain organoids immunostaining of 3 month BORGs

Staining with antibodies against Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes),and Gad67 (GABAergic neurons)

We assayed the organoids for gene expression levels and for pluripotency we observe an increase over time, for neruogenesis markers there is a sustained high level of expression, for astrocytes (Gfap) there is also a consistently high level of expression while for oligodendrocytes it remains lower in comparison (Figure 48). All the mature neuronal subtypes are strongly expressed in all organoids, suggesting a good maturation of all the 3 types of organoids; while cell cycle, proliferation and cancer stem cell markers remain consistent and similar across the 3 types of organoids (Figure 48). Since Hmyc is not expressed and therefore not quantifiable with qPCR analysis at Day 0, the Ct values (Cycle threshold) of 36B4 (housekeeping gene) and Hmyc (Human Myc gene) are compared (Figure 49). Unsurprisingly, we don't see an expression for Hmyc at Day 0, as the organoids are electroporated at Day 6 of differentiation; but we observe a Ct value of 25-27 from 1 to 3 month old BORGs.

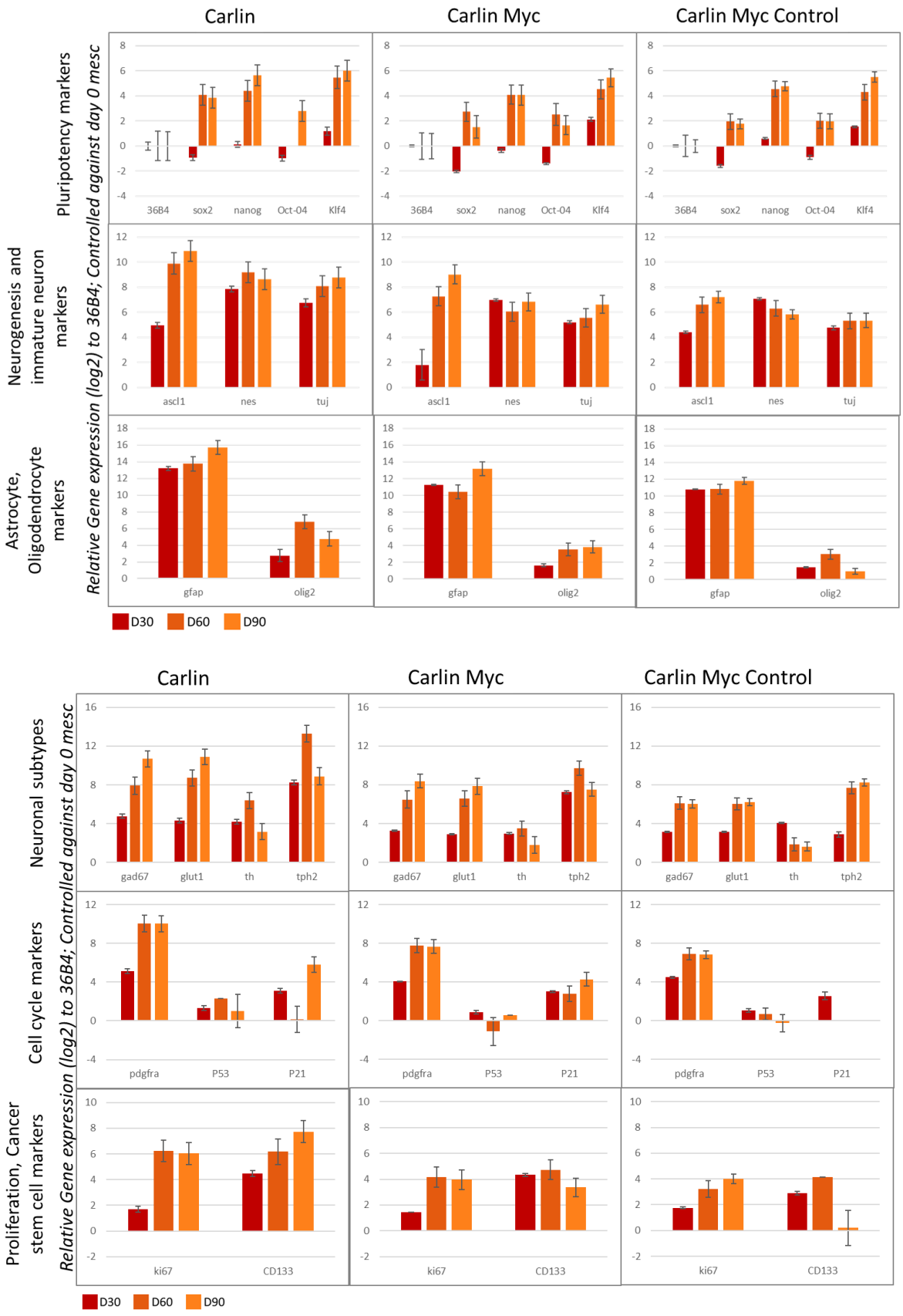


Figure 48. Myc electroporated Brain organoids gene expression analysis

Day 30, 60, 90

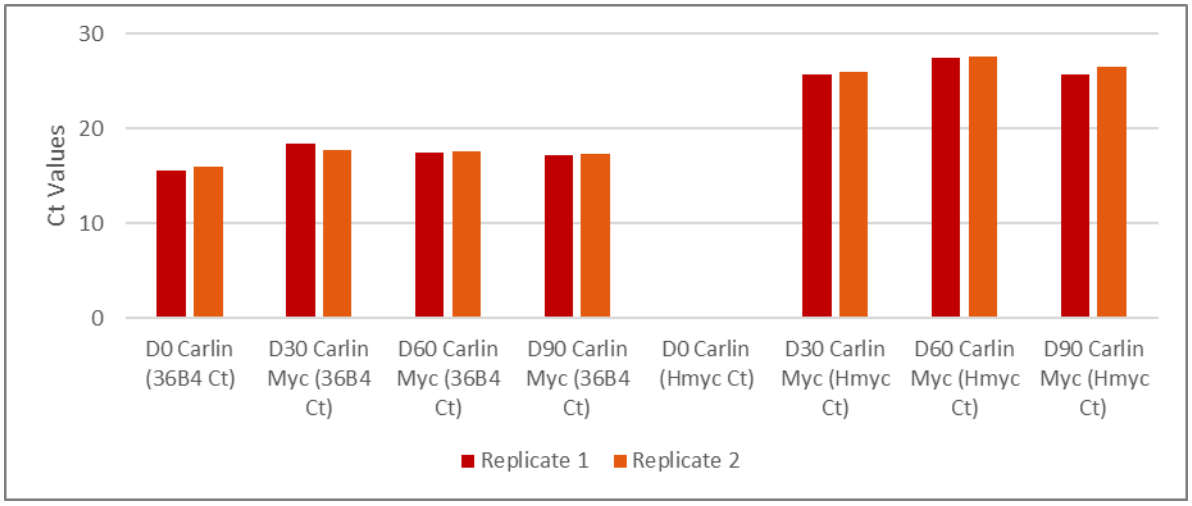


Figure 49. Myc electroporated Brain organoids Myc gene expression profile

36B4 (Housekeeping gene) and Hmyc plasmid gene Ct Value (Cycle threshold) barplots for Myc overexpression/ Myc control electroporated Carlin BORGs from Day 0 to Day 90 of differentiation.

Conclusion

We can confirm the electroporation of the organoids with Myc plasmid, but arent able to view a tumorigenic invasion. Using gene expression profiles and immunostaining confirms a neuronal differentiation and presence of neural stem cell niches and postmitotic neurons as well as the presence of proliferation (KI67) positive cells. Thus, to push the tumor invasion, we combined the P53KO cells with Myc overexpression profile and observe their growth.

Myc electroporated TP53KO Brain organoids

Using the WT Glio cell lines that have been modified with a P53KO plasmid, we performed the same experimentation as seen in the previous chapter. We studied 3 different cell lines – WT Glio, WT Glio P53KO, and WT Glio P53KO control- that had either been electroporated with Myc overexpression or Myc control plasmids at day 6 of organoid differentiation (Figure 50) and compared them with non electroporated brain organoids (a total of 9 conditions).

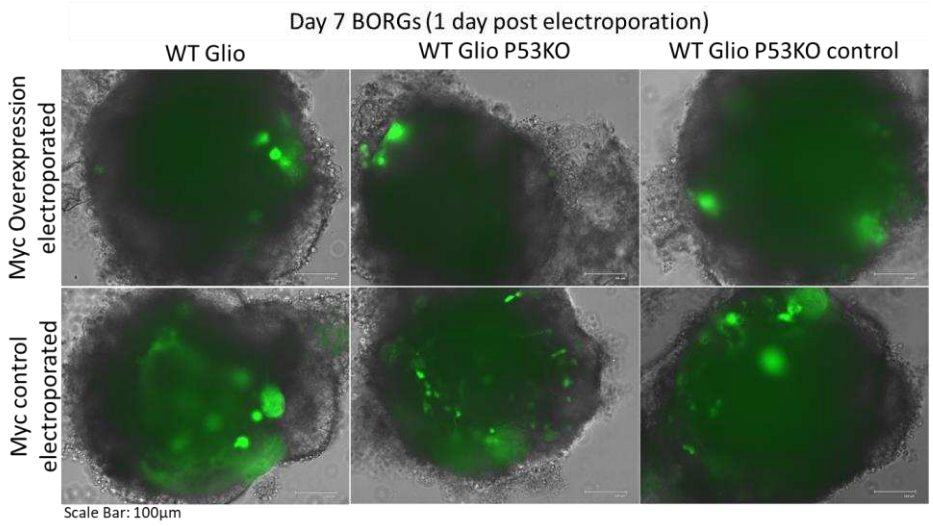


Figure 50. Myc electroporated TP53KO Brain organoids GFP tag visualization in BORGs 1 day post electroporation confirming successful electroporation of Myc plasmid

These BORGs were cultured for a period of 1 month and the sizes of all 9 conditions of BORGs were tracked (Figure 51). We find a significant difference in sizes seen at 1 month of growth and a statistical plot of of their mean diamaeters reveals that the sizes of WT Glio P53KO BORGs are significantly bigger than their WT Glio and WT Glio P53KO control counterparts, when comparing against the same Myc electroporation profile (Figure 52). Additionally, the Myc electroporated BORGs of WT Glio and WT Glio P53KO control BORGs are significantly larger than their non electroporated counterparts; this is not the case for WT Glio P53KO where the difference is insignificant.

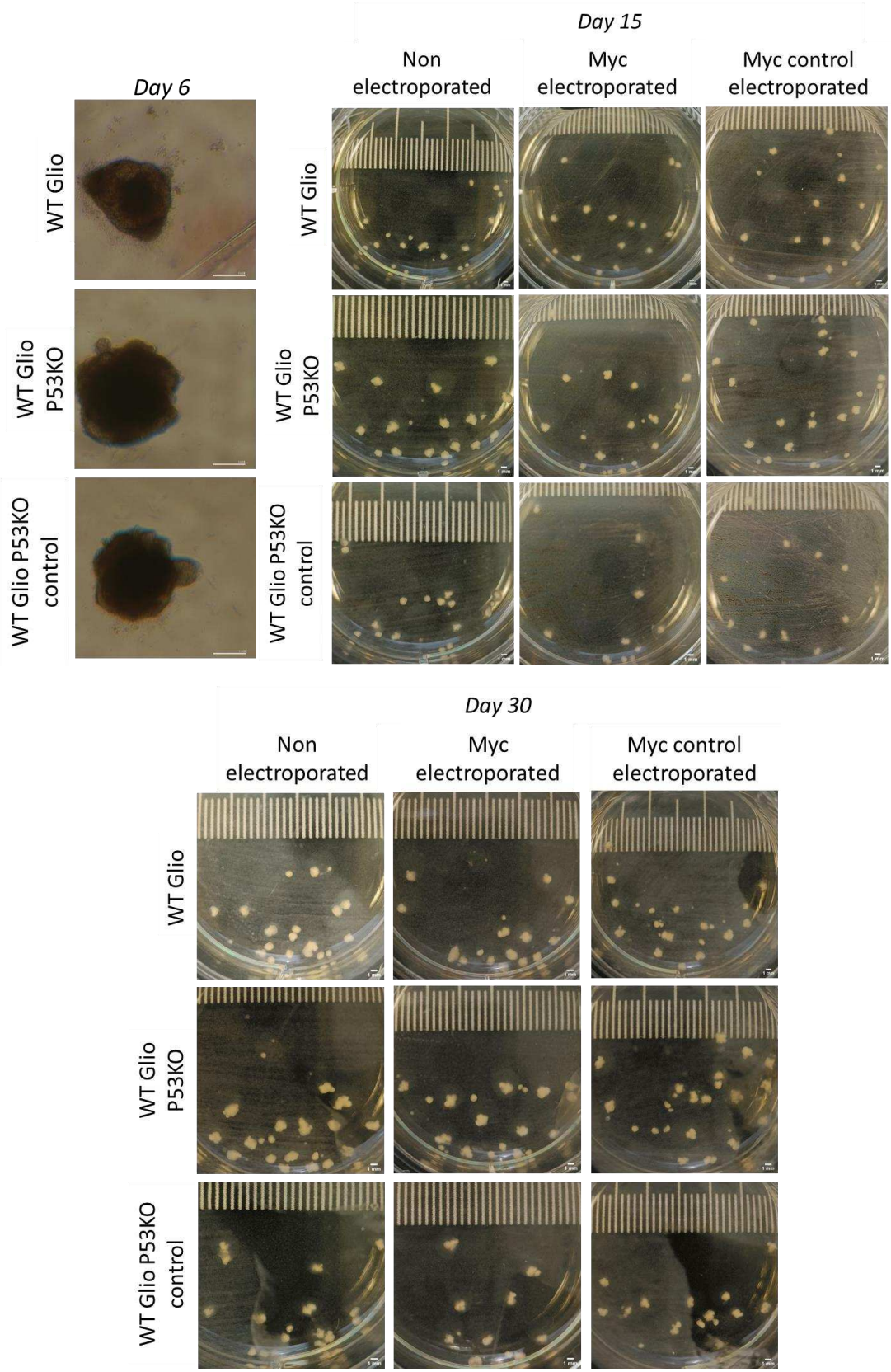
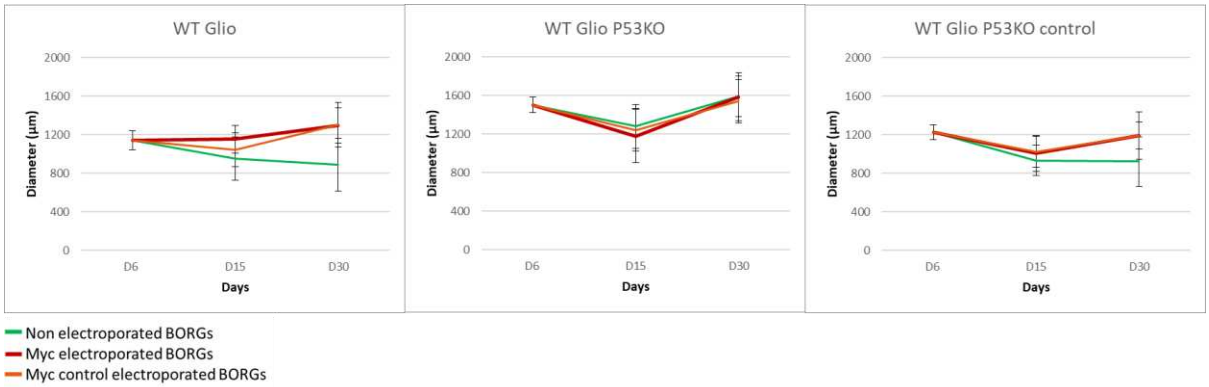


Figure 51. Myc electroporated TP53KO Brain organoids

Brightfield images of mouse brain organoids derived from MESC of WT Glio, WT Glio P53KO and WT Glio P53KO control, that are subsequently either electroporated with Myc overexpression or Myc control plasmids or not electroporated

52a



52b

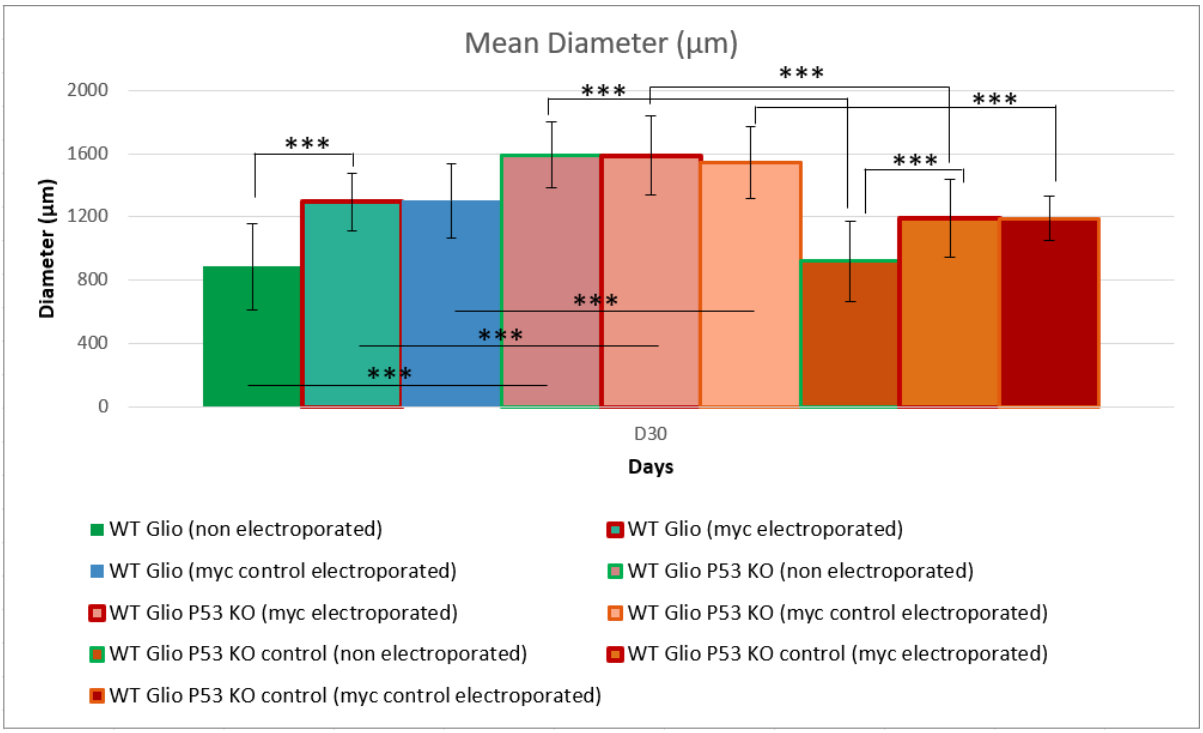
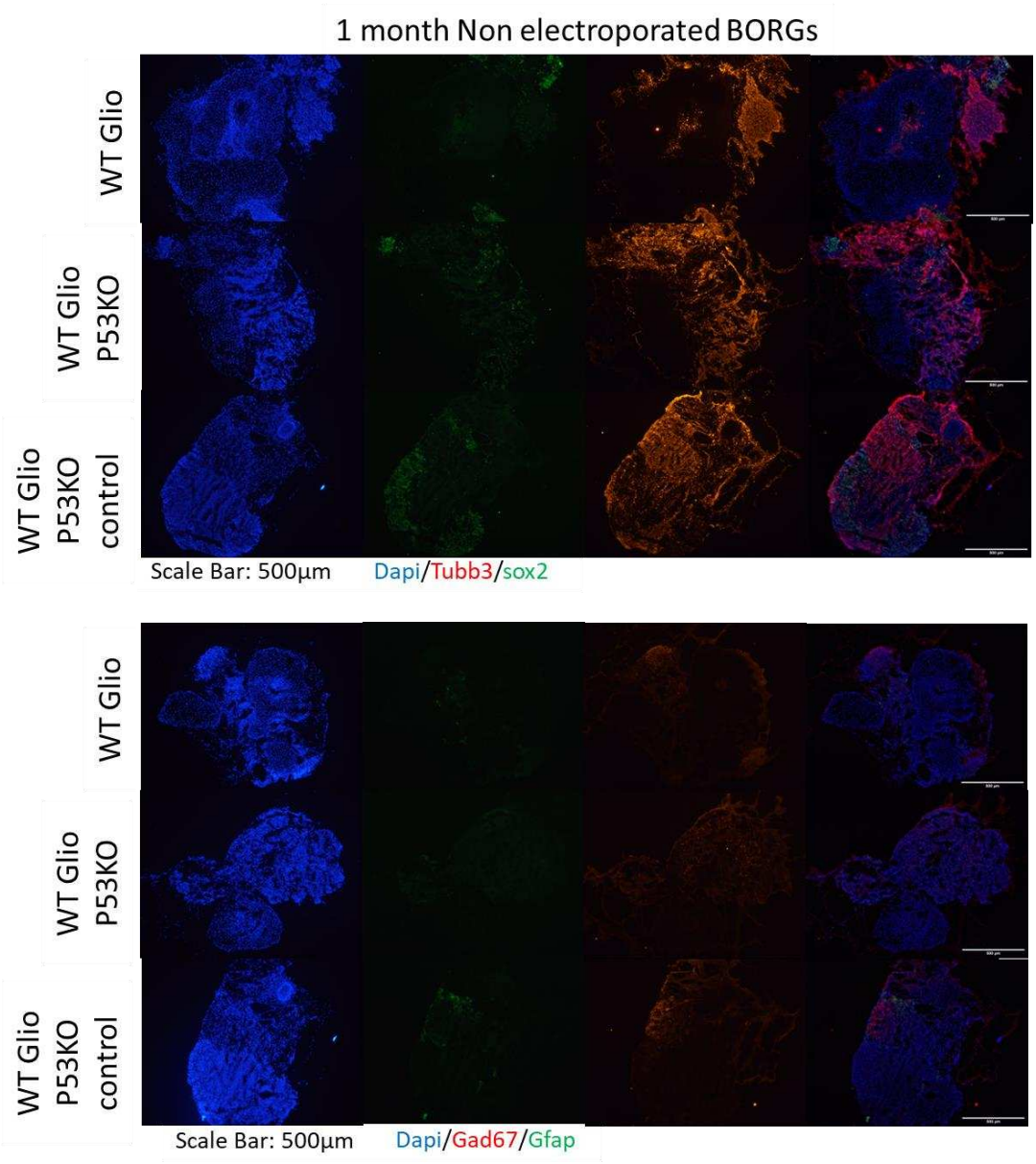


Figure 52. Myc electroporated TP53KO Brain organoids size measurements

a) Graphs of Mean diameter measurements of the sizes of the organoids over 30 days of culture; b) Barplots of mean diameter measurements of Day 30 WT Glio, WT Glio P53KO and WT Glio P53KO control brain organoids as non electroporated, Myc or Myc control electroporated conditions where data is presented as Two sample, Equal Variance t-Test, ***p<0.001, ns=no significance applies to the other means tested but not demonstrated in plot to keep the image clear

An immunostaining was subsequently performed at 1 month for all 9 conditions, and we find comparable levels of Sox2 (neuronal stem cells), Tubb3 (immature neurons), Gfap (Astrocytes), Gad67 (GABAergic neurons) across all samples (Figure 53, Figure 54, Figure 55, Supplementary Figure 17, Supplementary Figure 18, Supplementary Figure 19); while KI67

(proliferation) displayed a similar intensity of staining for all BORGs, there tended to be more regions of CD133 positive cells in Myc and Myc control electroporated BORGs, especially in the P53KO cell lines (Supplementary Figure 17, Supplementary Figure 18, Supplementary Figure 19).



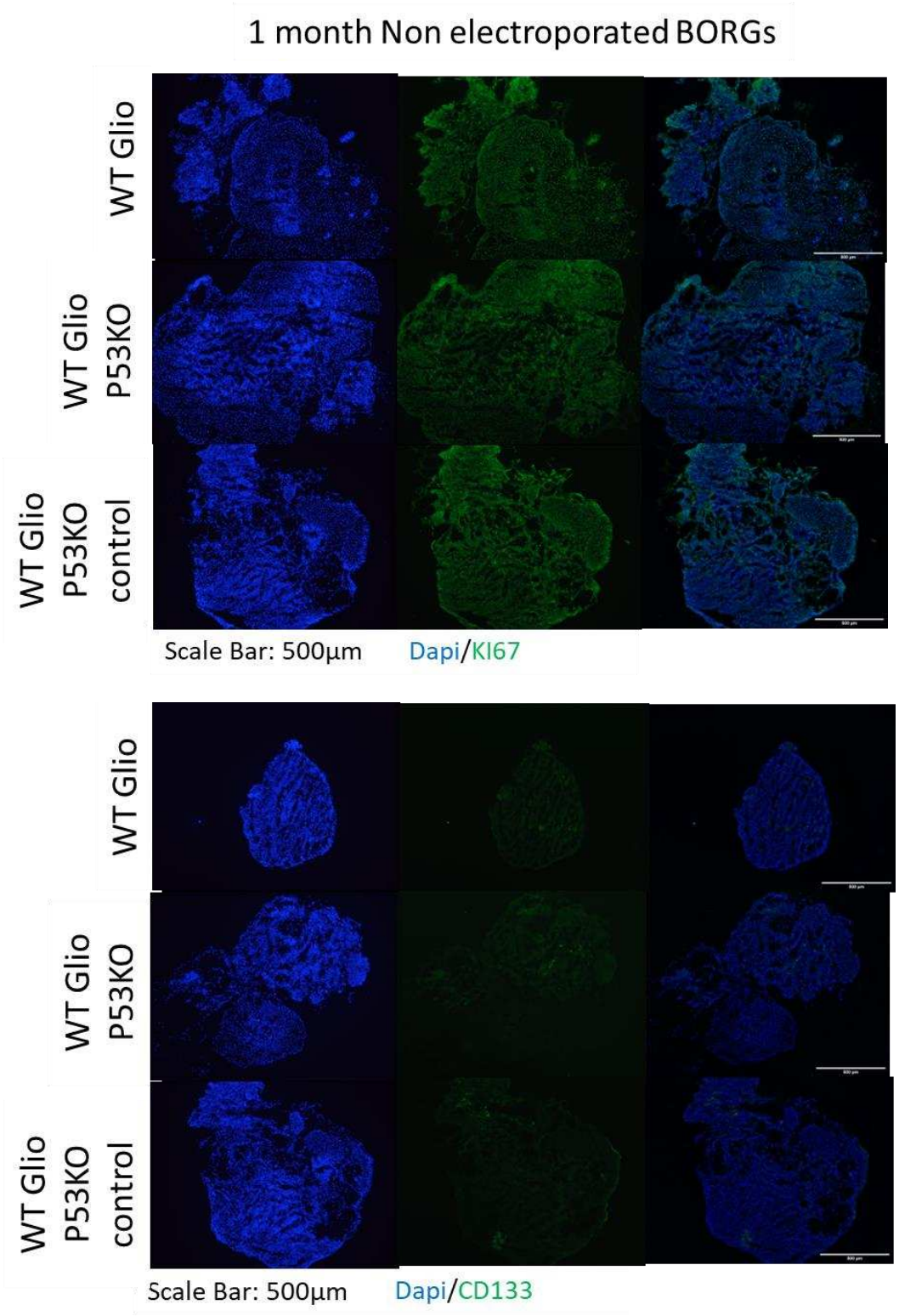
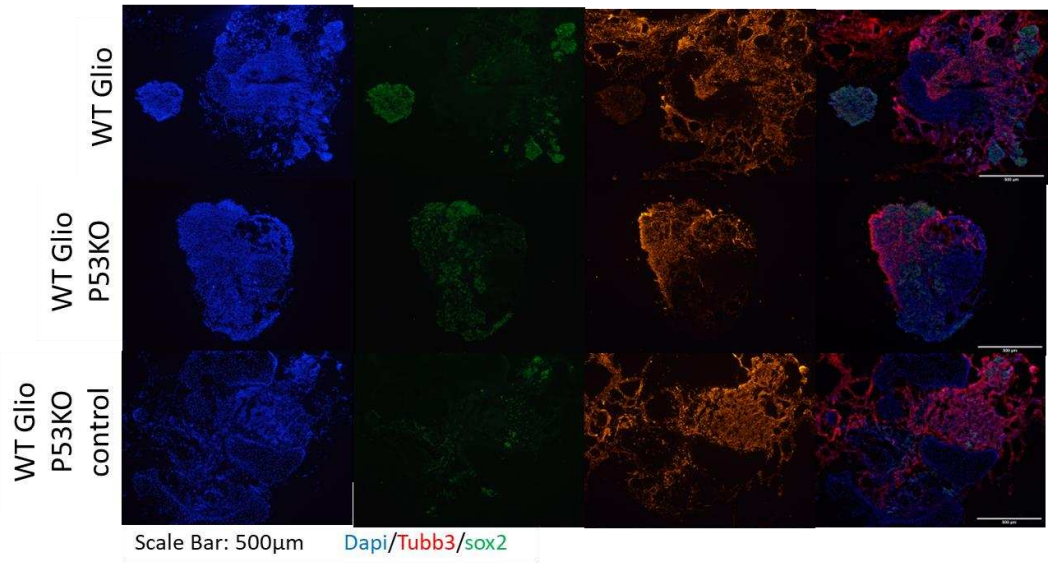


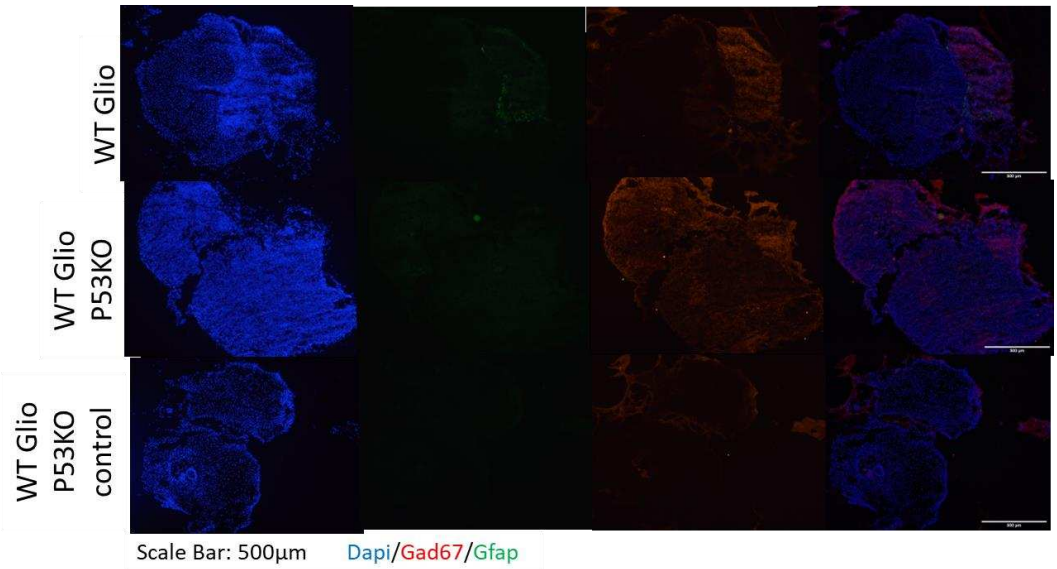
Figure 53. Myc electroporated TP53KO Brain organoids: immunostaining of 1 month non-electroporated BORGs

Immunofluorescence staining of 1 month old brain organoids with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (Gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell)

1 month Myc electroporated BORGs



Scale Bar: 500µm Dapi/Tubb3/sox2



Scale Bar: 500µm Dapi/Gad67/Gfap

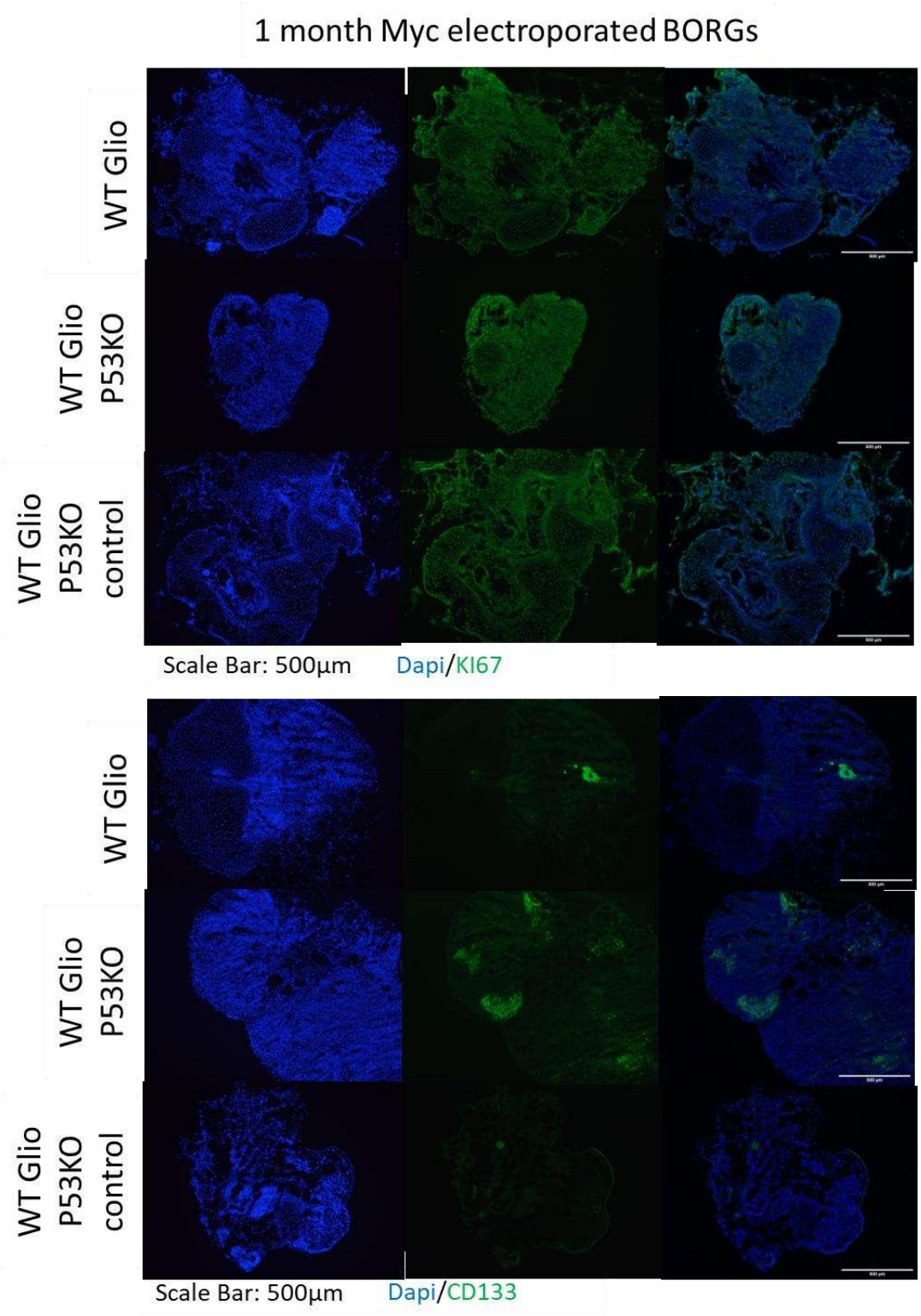
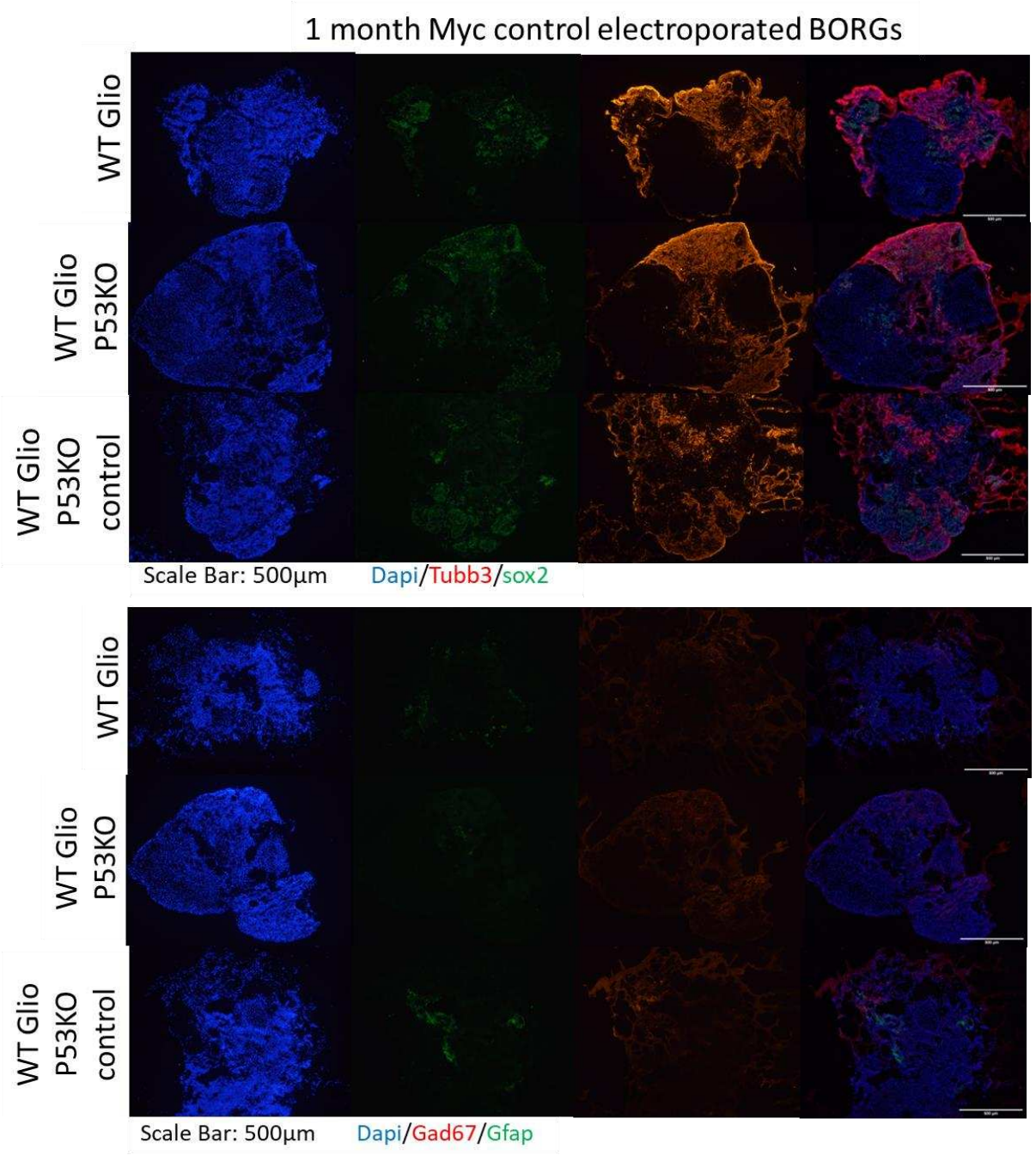


Figure 54. Myc electroporated TP53KO Brain organoids: immunostaining of 1 month myc-electroporated BORGs

Immunofluorescence staining of 1 month old brain organoids with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (Gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell)


1 month Myc control electroporated BORGs
WT Glio
WT Glio P53KO
WT Glio P53KO control
Scale Bar: 500µm
Dapi/Tubb3/sox2
WT Glio
WT Glio P53KO
WT Glio P53KO control
Scale Bar: 500µm
Dapi/Gad67/Gfap

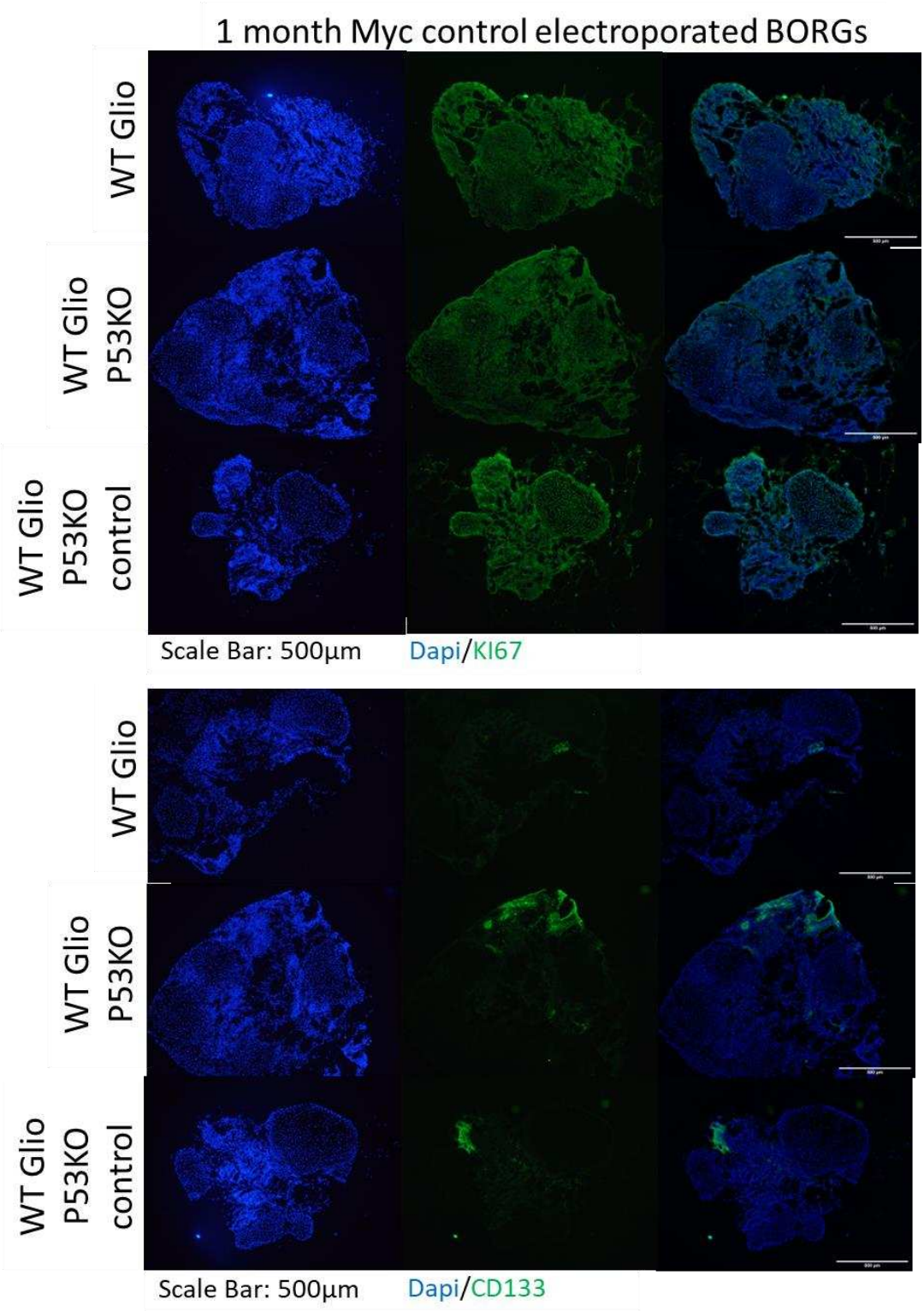


Figure 55. Myc electroporated TP53KO Brain organoids: immunostaining of 1 month myc control-electroporated BORGs

Immunofluorescence staining of 1 month old brain organoids with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (Gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell)

A comprehensive analysis of the gene expression levels of the 9 different conditions revealed that Nanog expression increased significantly at 1 month for WT Glio and WT Glio P53KO control BORGs; while interestingly, expression of Oct4 and Klf4 is no longer repressed at 1 month in WT Glio P53KO BORGs compared to their controls that continue to be repressed (Figure 56). The expression for Nes and Tuj (neurogenesis and immature neuron markers) remain down in WT Glio P53KO BORGs, irrespective of Myc electroporation status. While Gfap levels increase substantially across the board for all 9 conditions at 1 month, in WT Glio P53KO BORGs specifically Olig2 expression has increased from Day 6 to Day 30 16-fold, while in control organoids they remain steady. The expression of GABAergic and Glutamatergic neurons (Gad67, Glut1) is not seen in these organoids, while instead there is a strong expression for the Dopamine and Serotonergic neurons (Th, Tph2) (Figure 56). Notably in WT Glio P53KO BORGs, there was an increase in the expression of KI67 in Myc and Myc control electroporated conditions, and not for the non electroporated BORGs; while CD133 expression as well remains the same or increases over time in P53KO BORGs but dips in expression for the other controls. In order to trace the expression levels of Hmyc, we compared their Ct values with the control house keeping gene (36B4) and plotted their values at Day 0 and Day 30, only for the Myc electroporate BORGs (Figure 57). Primarily, we are able to observe the expression of Hmyc in Day 30 Myc electroporated BORGs in all 3 cell lines (Ct Values ranging from 27-35) (Figure 57); interestingly, the Ct values for Day 30 WT Glio and WT Glio P53KO control BORGs are lower than that of WT Glio P53KO.

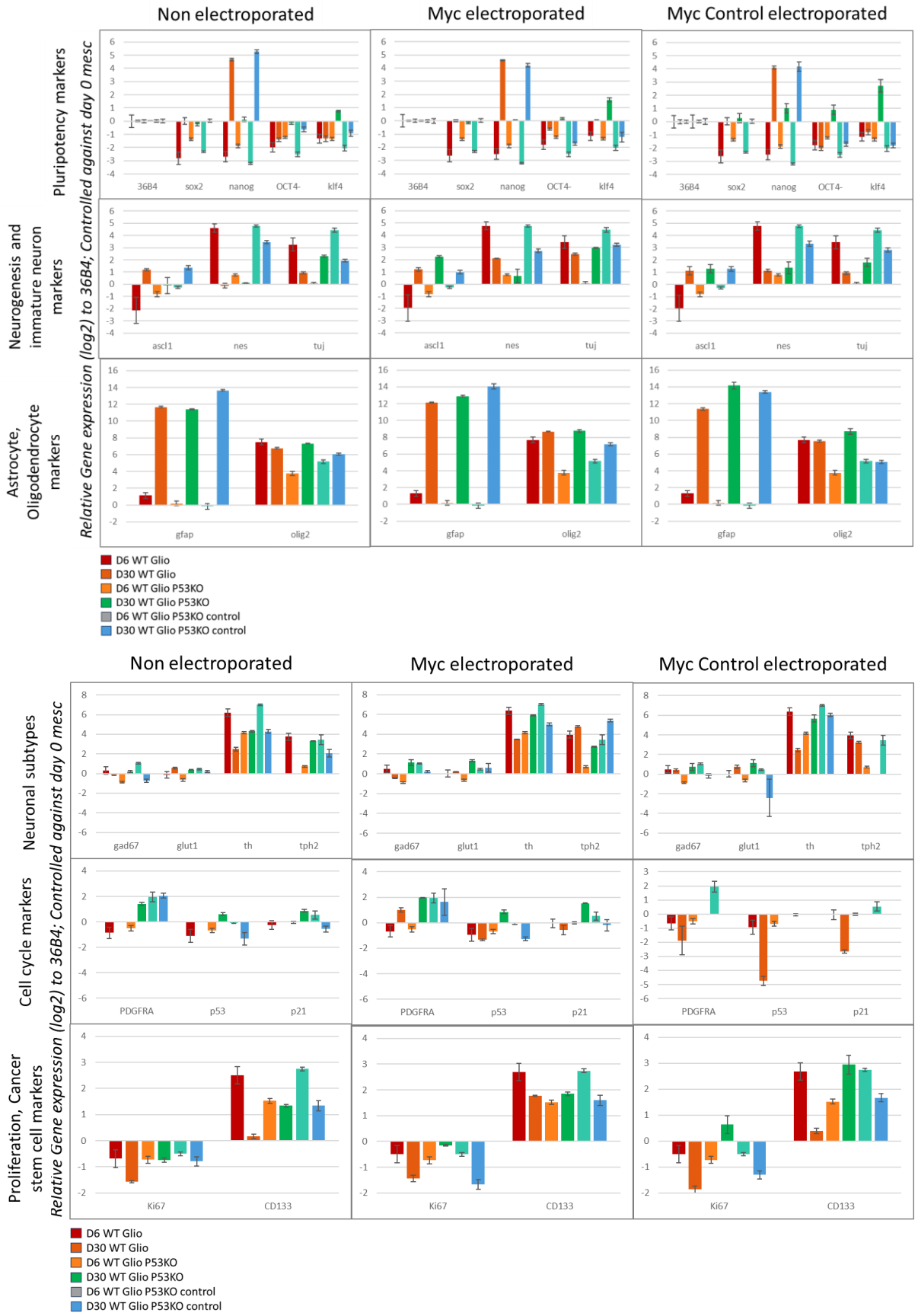
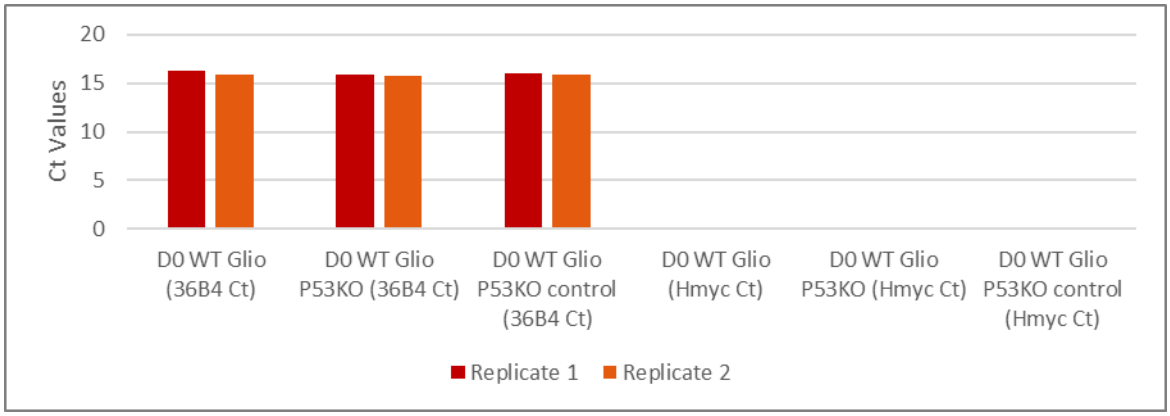


Figure 56. Myc electroporated TP53KO Brain organoids gene expression analysis

Day 6, 30

57a



57b

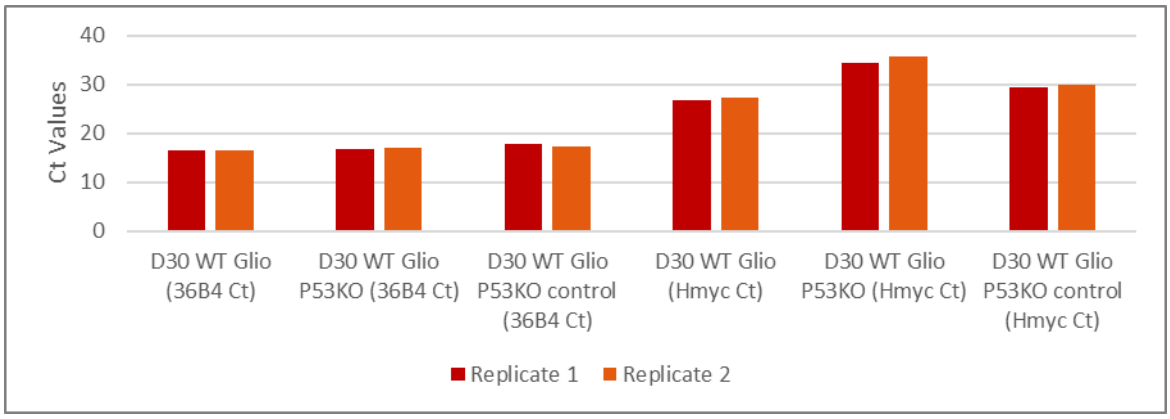


Figure 57. Myc electroporated TP53KO Brain organoids Myc gene expression analysis

36B4 (Housekeeping gene) and Hmyc plasmid gene Ct Value (Cycle threshold) barplots for WT Glio, WT Glio P53KO and WT Glio P53KO control BORGs, from a) Day 0 MESC and b) Day 30 Myc electroporated BORGs.

Conclusion

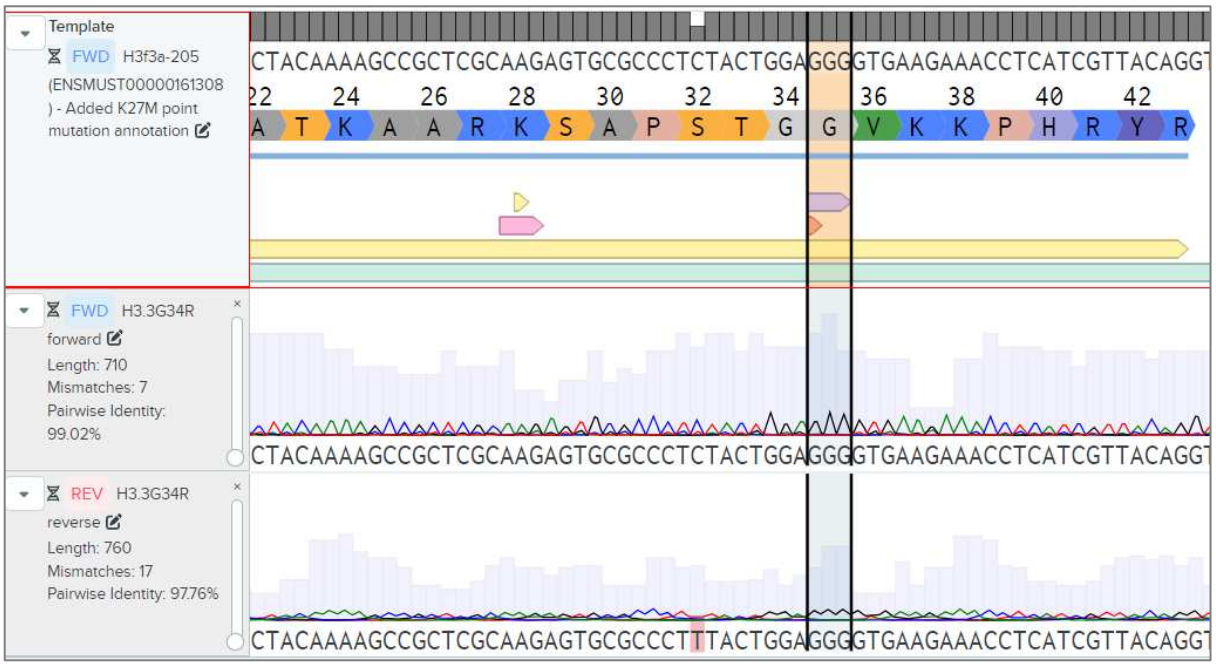
In this batch of organoids we are able to see a clear size difference between P53KO BORGs compared to the controls and Myc electroporated BORGs to the non electroporated BORGs except in the case of P53KO cell line (Figure 52b). This partially validates that P53KO cell lines are moving into a hyperproliferative state that can attribute to their larger sizes. Additionally, we are able to infer from the gene expression data that there is a reemergence of specific pluripotency factors (Nanog, Oct4, Klf4), an affected neurogenesis (Nes, Tuj), overexpression of OPC markers (Olig2), and maintenance/ upregulation in P53KO BORGs over the period of 1 month.

The data derived from this experiment demonstrates the capability of P53KO cell line to specifically create a tumorigenic profile.

Verification of H3.3 Mutation in MESC lines

Following the results from previous chapters, we had doubts as to the performance of our cell lines in recapturing a Glioblastoma phenotype and decided to sequence our mutant cell llines– H3.3G34R and H3.3K27M- to verify the presence of a heterogenous mutation in the H3.3 locus. Unfortunately, from the chromatograms, we noticed that both our mutant cell lines – H3.3G34R and H3.3K27M- don't carry the mutations (Figure 58). The G34R mutation is a glycine to arginine amino acid conversion and thus requires to be a conversion from GGA, GGT, GGC, GGG to CGA, CGT, CGC, CGG respectively. Likewise a K27M mutation would be a lysine to methionine conversion that is highlighted with a change of AAA, AAG to ATG. The regions selected in the chromatograms correspond to the amino acids after taking into account the amino acid and therefore have been highlighted as G35 and K28. We were unable to verify if the cells were possibly cross contaminated during routine cell culture, lost the mutation, or were not present initially in the first stocks of cells received. Therefore at this point, it is difficult to make a comment on the effect of G34R mutations on tumor profile.

58a



58b

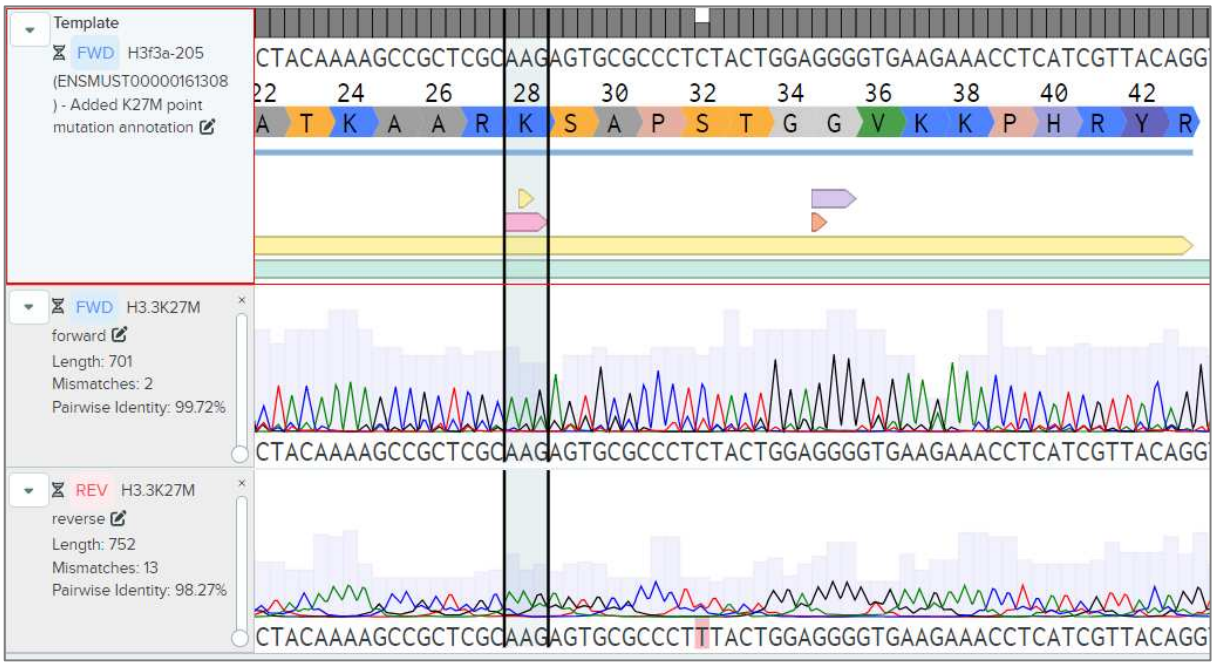


Figure 58. Verification of H3.3 Mutation in MESC lines

Chromatogram screenshots of H3.3 locus with regions for a) G34R and b) K27M mutation points highlighted. Benchling [Biology Software]. (2023). Retrieved from https://benchling.com.

Discussion & Perspectives

In Section 2, the aim of the thesis was to develop mouse brain organoids recapitulating *in vivo* characteristics of Pediatric Glioblastoma from H3.3 mutant phenotypes. Specifically two mutations, H3.3K27M and H3.3G34R mutant MESC lines would be grown in the lab as a means to easily study neurodevelopment disorders and diseases in an easily accessible, short term cell culture system. This would also contribute to the aspiration of the Working project 1 (Objective 1.2) in the Nanotumor Consortium which is the development of H3.3-mutated mouse organoid models for pediatric studies.

In line with this goal, we first optimized our mouse brain organoid generation protocols to generate reproducible and higher quality un-patterned whole brain organoids. On discovering the relevance of P53 signature from bulk RNA transcriptomics, and also as a frequently co occurring mutation, we incorporated a P53KO signature into our G34R mutant cell lines to visualize a possible overgrowth and are able to see a significant increase in G34R P53KO organoids compared to the control conditions after 1 month of culture. We also exploited the Myc overexpression phenotype as observed in Bian et al, 2018 to discover if this supports an overgrowth in our organoids. We manipulated the cells as MESC to introduce Myc overexpression plasmid and in another experimentation, manipulated the organoids, to electroporate Myc plasmid early in growth. Unfortunately, in either conditions, we don't observe overgrowth of organoids, nor strikingly different gene expression profiles over time.

However, by combining the P53KO cell lines with a Myc plasmid electroporation, we can observe a significant increase in the size of organoids that have a P53KO signature. These organoids are 300-800 µm larger in size than their control counterparts. Additionally, BORGs treated with a Myc overexpression plasmid show 300-400 µm increase in sizing compared to the non electroporated counterparts, except for P53KO BORGs.

In an attempt to verify the mutational profile of our cell lines however, we noticed that both the mutant cell profiles appear to be wildtype cell lines. Continuing from this, we still cant make assumptions as to what point we have been dealing with non-mutated or contaminated cell lines.

To conclude, I have generated a working protocol of P53KO signature BORGs that can demonstrate significant overgrowth within 1 month of culture. From this point, it would be

beneficial to redo our work in confirmed H3.3 mutated cell lines, in order to compare with their isogenic control lines. Specifically, using P53KO in confirmed H3.3 mutant BORGs, we should be able to generate a tumorigenic H3.3 mouse BORG, that can be further analyzed at a single cell resolution and spatially using already established tools in the team such as MULTILAYER [145].

Conclusion

Section 1- Retinoid driven neurogenesis

Here, I first explained the findings of our recent publication [132], that demonstrates the ability of RARβ and RARγ conjointly to hijack the role of RARα in recovering neuronal cell specialization in P19 EC cells.

Subsequently, we replicate our strategies in mouse embryonic stem cells using single cell RNA sequencing technology to track the differences in cell fate acquisition when cells are treated with synthetic agonists – ATRA (pan RAR agonist), BMS753 (RARα agonist), BMS641+961 (RARβ + RARγ agonist). To this end, we have identified 2 clusters at the late differentiation timepoint that are present more abundantly in RARβ + RARγ agonist versus RARα agonist treated cells. However, the identity of only one cluster (cluster 11) has been partially deduced and requires further efforts for annotation after we have been able to define relevant gene identities for each cell type. Furthermore, with the presence of an endodermal and mesodermal signature in our cultures, it will be necessary to either incorporate or filter these signatures from the single cell transcriptomics analysis.

We also generated BORG cultures by maturing them in exposure to either BMS753 or BMS641+961, instead of conventionally used Vitamin A (positive control). These BORGs can produce neurogenesis and mature neuronal signatures, equivalent to the expression seen in positive controls. Therefore, if synthetic retinoid ligands are able to recapture brain organoid progression, the possible redundancy at the level of the RAR activation needs to be explored using spatial transcriptomics and epigenomics, which is underway in the team.

Section 2- Development of Mouse brain organoids that recapitulate H3.3 mutated pediatric glioblastoma

We begin with the verification of the neuronal differentiation capacity of our wildtype and isogenic mutant cell lines followed by the first mouse brain organoid generation protocol. This protocol is optimized to manufacture more reproducible BORGs and introduce commonly found mutation signatures in the BORGs during their development such as P53KO and Myc overexpression studies.

The major result is the development of a working protocol of a P53KO signature BORGs that can demonstrate significant overgrowth within 1 month of culture. From this point, it would be beneficial to redo our work in confirmed H3.3 mutated cell lines, in order to compare with their isogenic control lines. Specifically, using P53KO in confirmed H3.3 mutant BORGs, we should be able to generate a tumorigenic H3.3 mouse BORG, that can be further analyzed at a single cell resolution and spatially using already established tools in the team such as MULTILAYER [145]. The goal would be the setup of a mouse BORG model that can be used to study mutational profiles related to neurodegenerative diseases and cancer, primarily as an initial experimental setup that is cost effective and verifiable before proceeding to hiPSC derived cultures.

Further Scientific Contributions

In 2021, the Sysfate team published the development of the MULTILAYER tool, that allows the analysis of spatial transcriptomics in tissue and organ samples in multiple sections [145]. To complement the publication, we published a protocol detailing the use of MULTILAYER especially from the point of view of a Biologist.

Contribution

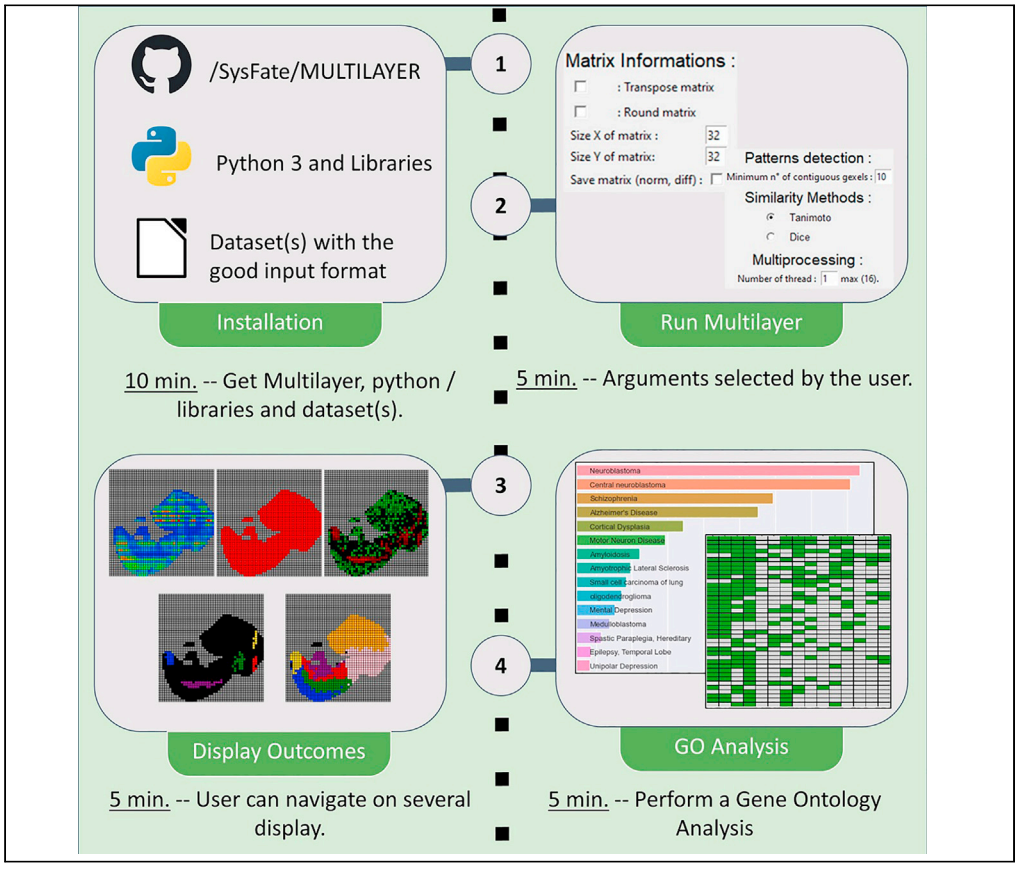
I am a co-first author of this paper, shared with Julien Moehlin. I evaluated the use of the tool and contributed to the writing, review and editing of a biologist- friendly protocol. The study was published in STAR protocols on December 17, 2021. The online version of this article can be found here https://doi.org/10.1016/j.xpro.2021.100823.

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for using MULTILAYER to reveal molecular tissue substructures from digitized spatial transcriptomes



Spatially resolved transcriptomics (SrT) allows us to explore organ/tissue architecture from the angle of the gene programs involved in their molecular complexity. Here, we describe the use of MULTILAYER to reveal molecular tissue substructures from the analysis of localized transcriptomes (defined as gexels). MULTILAYER can process low- and high-resolution SrT data but also perform a comparative analysis within multiple SrT readouts.

Julien Moehlin, Aysis Koshy, François Stüder, Marco Antonio Mendoza-Parra

mmendoza@genoscope.cns.fr

Highlights

MULTILAYER reveals spatial transcriptomes (ST) into biologically relevant substructures

MULTILAYER provides a user-friendly environment for ST processing

MULTILAYER has been used on a variety of data, including high-resolution ST

A "Visium data convertor" and a "high-resolution data compressor" are also available

Moehlin et al., STAR Protocols *2*, 100823
December 17, 2021
https://doi.org/10.1016/j.xpro.2021.100823

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for using MULTILAYER to reveal molecular tissue substructures from digitized spatial transcriptomes

Julien Moehlin,[1,2,3] Aysis Koshy,[1,2] François Stüder,[1] and Marco Antonio Mendoza-Parra[1,4,*]

[1]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry, University Paris-Saclay, 91057 Évry, France

[2]These authors contributed equally

[3]Technical contact

[4]Lead contact

*Correspondence: mmendoza@genoscope.cns.fr
https://doi.org/10.1016/j.xpro.2021.100823

SUMMARY

Spatially resolved transcriptomics (SrT) allow researchers to explore organ/tissue architecture from the angle of the gene programs involved in their molecular complexity. Here, we describe the use of MULTILAYER to reveal molecular tissue substructures from the analysis of localized transcriptomes (defined as gexels). MULTILAYER can process low- and high-resolution SrT data but also perform comparative analyses within multiple SrT readouts.

For complete details on the use and execution of this protocol, please refer to Moehlin et al., 2021.

BEFORE YOU BEGIN

Timing: < 30 min

This section includes the minimal hardware requirements, the installation procedures, as well as the format of the files to be processed by MULTILAYER.

Hardware

Local-Memory: a minimum of 8GB required

Downloading MULTILAYER toolkit

1. The Multilayer toolkit can be downloaded from https://github.com/SysFate/MULTILAYER (Figure 1A). This is done by clicking in the green 'Code' tab and downloading the zip folder. It contains the data set used as examples in the paper, the gene ontology databases, the tutorial in a pdf format, the MULTILAYER tool, MULTILAYER compressor tool, "Visium Converter" tool and a ReadMe file.
2. After downloading the zip folder, extract all files.

Anaconda Python Platform

MULTILAYER is functional on all operating systems (Windows, Linux and Mac OSX) with Python 3. Python version 3.8 is recommended. A simplified strategy to run Python on any operating systems is to use Anaconda.

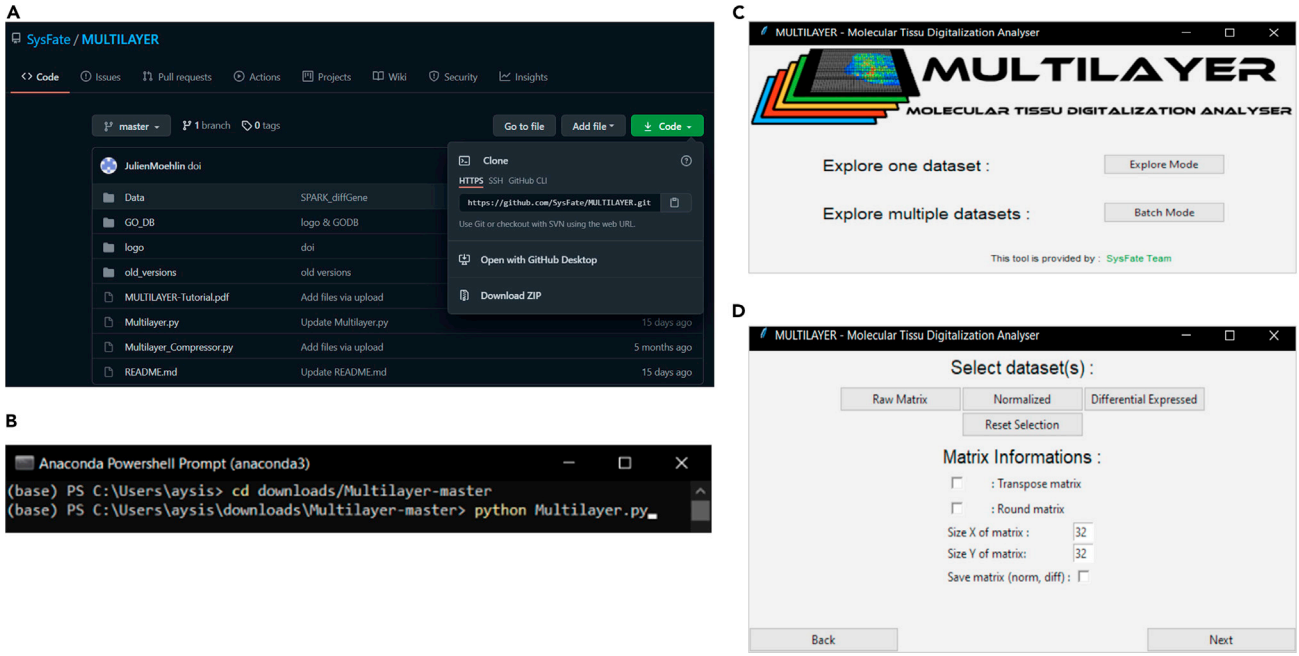


Figure 1. Download and Launching MULTILAYER

(A) The MULTILAYER toolkit is available at https://github.com/SysFate/MULTILAYER.

(B) The PowerShell prompt console with the written function to launch MULTILAYER.

(C) The MULTILAYER application home page providing either an "explore mode" for the analysis of a single SrT dataset, or the "batch mode" dedicated for processing multiple files.

(D) The MULTILAYER interface allowing to upload either raw, normalized or differentially-expressed SrT matrix data. In addition, users can transpose and/or round off the grid coordinates, as well as select the grid size if required.

3. Anaconda can be downloaded from https://www.anaconda.com/products/individual according to individual computer specifications.
4. Once Anaconda is installed, open up the Anaconda Navigator to launch the PowerShell Prompt console.
5. Before using MULTILAYER for the first time, it is necessary to install the following dependencies as following:

```
> pip install numpy

> pip install matplotlib

> pip install pandas

> pip install scipy

> pip install scikit-learn

> pip install seaborn

> pip install networkx

> pip install python-louvain

> pip install pillow
```

Launching MULTILAYER

6. To launch MULTILAYER, you have to first access the files by tracing the correct path from the PowerShell Prompt Console (Figure 1B). For e.g., If the Multilayer.py is located in the Multilayer-master folder, which is located in the downloads folder, type:

```
> cd downloads\Multilayer-master
```

7. To launch the Multilayer tool, type

```
> python3 Multilayer.py
```

or

```
> python Multilayer.py
```

(This will depend on which version of python has been installed.)

Data collection

Spatially resolved transcriptomics (SrT) promotes the transition of histological studies towards a data science context, notably by the generation of digitized maps of tissue architecture. MULTILAYER takes advantage of such a digitized view to reveal molecularly-defined spatial signatures, with the use of agglomerative clustering over contiguous gene expression elements, herein defined as gexels (in analogy to pixels retrieved on digital images). Considering the recent heterogeneity of platforms available for generating SrT data (e.g., use of DNA arrays (Rodriques et al., 2019; Ståhl et al., 2016) or microfluidic channels (Liu et al., 2020)), users might refer to the corresponding platforms for the preprocessing steps (spatial barcodes demultiplexing, read counts alignment, gene/transcript read counts association, etc). In all cases, the spatial transcriptome output requires to be in the format of a matrix composed of columns associated to spatial coordinates and rows to gene identifiers.

8. Depending on the SrT platform in use, such a spatial matrix is generated within the corresponding analytical pipeline or it can be obtained as following:
   a. In cases in which the structure of the matrix is transposed (i.e., columns are associated to gene identifiers and rows to spatial coordinates) or where the provided coordinates are not as round numbers; MULTILAYER provides options to handle these types of situations. Within the "Select dataset(s)" panel, user can opt to transpose and/or round off the entry matrix (Figure 1D). At this stage MULTILAYER is not able to recognize the format of the dataset in use.
   b. In cases in which the data is available as a three column dataframe (spatial coordinates, gene identifiers and read counts), the ad-hoc module called "MULTILAYER compressor", allows to convert it into the required matrix. In addition, this module is able to decrease the resolution of the available data by agglomerating contiguous gexels defined by a user-provided compression factor (refer to section "processing high-resolution SrT data").
   c. In cases in which the data is issued from the commercial solution "Visium", the ad-hoc module, called "VisiumConverter", allows to generate the required matrix. For this, users need to download the 10×Genomics proprietary tool spaceranger from here: https://support.10xgenomics.com/spatial-gene-expression/software/downloads/latest and proceed with the following Visium generated files:
      i. a matrix in h5 format.
      ii. the spatial imaging data containing the tissue_positions_list.csv file.

iii. the feature barcodes matrix files containing the corresponding features.tsv.gz file.

On the PowerShell Prompt Console, users can covert h5 matrix files generated by Visium to csv format as following:

```
> spaceranger mat2csv cellmatrix_HDF5.h5 out_file_matrix.csv
```

Where cellmatrix_HDF5.h5 is the h5 matrix files to convert to the csv format. For a detailed description of these steps, users can follow the spaceranger dedicated tutorial here: https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/output/matrices

The obtained csv matrix can be processed by our ad-hoc module, called "VisiumConverter" (available at https://github.com/SysFate/MULTILAYER) module as follows:

```
> python visiumConverter.py -m out_file_matrix.csv -p spatial/tissue_positions_list.csv -g
raw_feature_bc_matrix/features.tsv.gz -o matrix_multilayer.tsv –compressor
```

Where -m defines the matrix to be converted; -p: the tissue positions information, -g: the features and -o: the matrix compatible with MULTILAYER. Additionally, the "–compressor" option generates a three-column format (spatial coordinates / Gene ID / read counts) file, which can be processed by "MULTILAYER compressor" in case users need to decrease gexels density prior to MULTILAYER processing. Since new Visium DNA arrays present an interstitial printed spot, the matrix generated with our "VisiumConverter" tool tends to stretch the maps on the y-axis, without a real impact on the spatial information, as illustrated on Figure 2.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Whole_mouse_embryo raw matrix (DBiT-Seq)	Liu et al., 2020	https://github.com/SysFate/MULTILAYER/tree/master/Data/Whole_mouse_embryo
High_resolution_brain raw matrices (Slide-seq)	Rodriques et al., 2019	https://github.com/SysFate/MULTILAYER/tree/master/Data/High_resolution_brain
Prostate_cancer raw matrices (ST)	Berglund et al., 2018	https://github.com/SysFate/MULTILAYER/tree/master/Data/Prostate_cancer
Software and algorithms		
MULTILAYER	Moehlin et al., 2021	https://github.com/SysFate/MULTILAYER
MULTILAYER compressor	Moehlin et al., 2021	https://github.com/SysFate/MULTILAYER
Visium Converter	This article	https://github.com/SysFate/MULTILAYER

STEP-BY-STEP METHOD DETAILS

Herein we describe Step-by-step methods for analyzing spatial read counts, from the loading of raw data, to its normalization, differential gene expression detection, gene co-expression pattern mapping and finally revealing the spatial communities corresponding to biologically relevant tissue substructures. To illustrate these various steps, we use as an example, the analysis of the Whole Mouse Embryo data generated by Liu and colleagues (Liu et al., 2020).

Open data on MULTILAYER

Timing: 10 min

1. Use Explore mode for this data. This is used for the analysis of one sample. (Figure 1C).

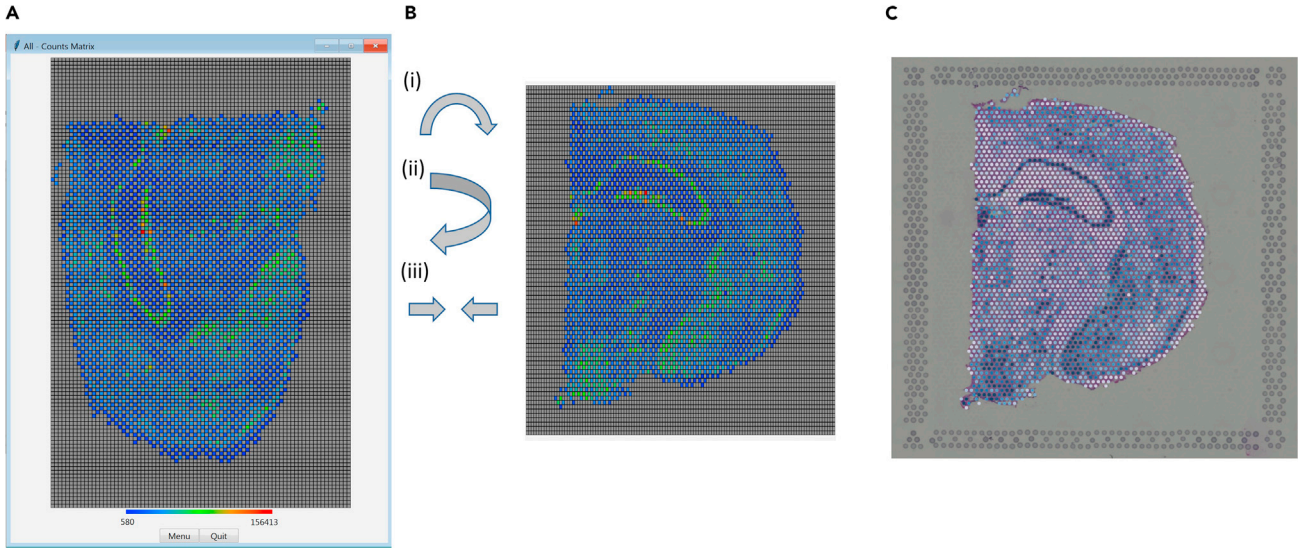


Figure 2. MULTILAYER display for datasets generated by the commercial platform Visium

A dataset corresponding to the analysis of a mouse brain section (Coronal) has been downloaded from 10×Genomics https://www.10xgenomics.com/resources/datasets/mouse-brain-section-coronal-1-standard. As indicated on "Data Collection" section, we have converted the corresponding cell matrix HDF5 (filtered) to csv with spaceranger; then we have used our "visiumConverter" module to generate a MULTILAYER compatible matrix.
(A) Raw count matrix corresponding to the coronal mouse brain section, displayed within the MULTILAYER platform. Notice that the image seems stretched on the y-axis. This is due to the fact that Visium DNA arrays present interstitial printed spots.
(B and C) By applying an image rotation of 90°; a flipping conversion and a stretch of the image on the new x-axis we can obtain a view identical to that provided by the commercial platform (displayed in C).

2. Provide your dataset by clicking Raw matrix and select the appropriate data; in this case Folder 'Multilayer-master', Folder 'Data', Folder 'Whole_mouse_embryo', File 'GSM4096262_0725cL.tsv'. (Figure 1D).

   Note: A .tsv file (table separated values) should be provided as raw matrix. The correct format is a matrix with gene names as rows, coordinates as columns (coordinates should be in the format 'XxY') (see data collection). It is necessary to enter a raw matrix. Optionally, if you provide your own normalized matrix and / or differential expressed matrix MULTILAYER will use them. If you provide only raw matrix, you can save those generated by MULTILAYER by checking the option for 'Save Matrix (norm, diff)'.

3. Select option to transpose matrix.

   Note: Transposing the dataset is useful when you have a matrix with gene names as columns and coordinates as rows (see data collection). When it is not performed (but required), the following error comes up – AttributeError: 'NoneType' object has no attribute 'columns'.

4. Optional setting to round off the matrix; default is unchecked.

   Note: MULTILAYER requires to round out spatial coordinates for the analysis. For this, the tool needs an integer as X & Y (coordinates) which will get rounded off. For eg: a given coordinate '3.36×13.94' will be converted to '3×14'.

5. Indicate maximum matrix size or leave as default option 32 × 32.
6. Click Next.
7. Indicate threshold for up and down regulated differential gene expression; default range is 1 to −1 (in log2).

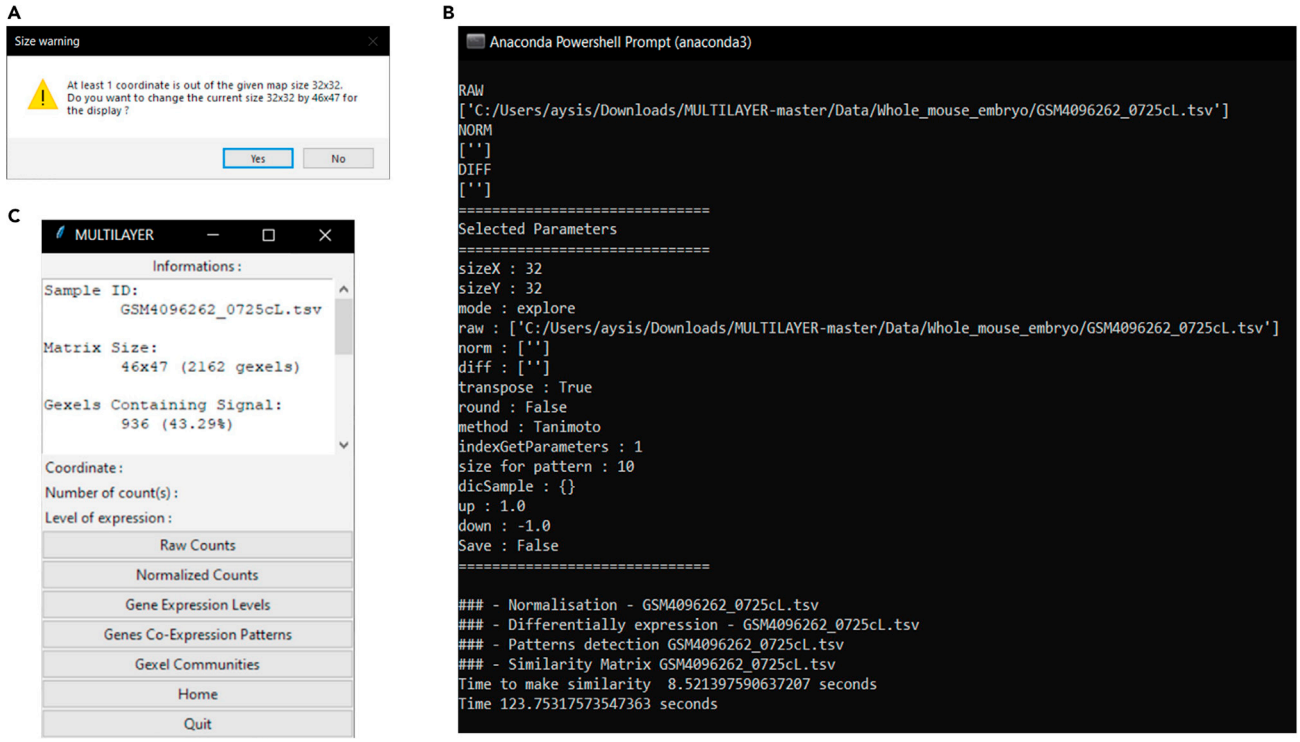


Figure 3. Processing whole_mouse_embryo data on MULTILAYER

(A) When the size matrix defined by the user doesn't comply with the grid size retrieved within the data, a size warning error is displayed. User can decide whether or not to accept the proposed grid size modification. This option does not impact the computing steps, but only the matrix display.

(B) As MULTILAYER processes the data, the Anaconda prompt interface displays a listing of the different steps taking place.

(C) Once it is finished, the processed data is available via the MULTILAYER interface.

8. Indicate minimum number of contiguous Gexels; default is 10.

   Note: In Pattern detection, the tool uses agglomerative clustering on gexels (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html) to determine a contiguous pattern. It considers a pattern if the number of gexels is equal to or higher than this value. Default value is 10. This value is to be adapted to the size of the tissue, and the resolution of the data (number of total gexels within the matrix). An empirical evaluation of the minimal number of contiguous gexels is required on the basis of the pattern's detection performance (as illustrated in Figure 10).

9. Indicate similarity methods- Tanimoto or dice; default is Tanimoto.

   Note: Similarity methods calculate the similarity between patterns.

- Tanimoto: the similarity is calculated as A Union B / A Intersection B.

- Dice- the similarity is calculated as 2*(A Union B) / A Intersection B.

10. Indicate multiprocessing i.e., number of threads; default is 1.
11. Click next.
12. If the indicated matrix size is inaccurate and/or data is detected in points outside the matrix size suggested, a dialog box will appear indicating a size warning and ask if you want to open the file in a suggested size; here for e.g., 46 × 47. Click yes (go with MULTILAYER's recommendation) (Figure 3A).

13. The MULTILAYER interface allows at this stage to access to the various types of processed data (Raw counts, Normalized counts, Gene expression levels, Gene co-expression patterns, gexel communities) (Figures 3B and 3C).

Visualize raw and normalized data on MULTILAYER

Timing: <10 min

14. On the MULTILAYER interface, select the type of data you want to view (Raw counts, or Normalized counts).

Note: In contrast to Raw counts, the Normalized matrix provides a view in which all read counts were corrected by using a quantile normalization strategy (as described in (Moehlin et al., 2021)). Briefly, global read counts normalization is essential to correct technical bias affecting local read count levels across the tissue section of interest. Hence, all downstream steps (differential gene expression, gene co-expression patterns detection, spatial communities' detection) are issued from the normalized data. In cases in which the user might prefer to use their own normalization strategy, they can upload the normalized and corresponding raw matrix at the beginning of the analysis. In such a situation, MULTILAYER will not apply the quantile normalization correction, but instead it will use the provided normalized data for all downstream steps.

15. In both cases, the Menu button can be used to search for specific genes.
16. When a gene is selected, users can indicate the minimum or maximum read count threshold to enhance the display; as well as to convert it to log2 (Figure 4).

Visualize differential gene expression data on MULTILAYER

Timing: <10 min

17. On the MULTILAYER interface, select "Gene Expression levels" (Figure 4A).
18. By clicking the "Menu" button, a panel displaying a list of the upregulated genes ranked by the number of associated gexels is displayed (Figure 5A).
19. Similarly, the list of ranked downregulated genes is available by clicking on "Show repressed genes".
20. To visualize the gene expression location, users can either enter the gene name in the search panel, or double-click on the ranked list.
21. To enhance the display, users can modify the heatmap intervals (within the "threshold panels"), and also modify the differential expression threshold (threshold expression panels).

Visualize gene co-expression patterns

Timing: <10 min

22. On the MULTILAYER interface, select "Gene co-expression patterns" (Figure 4A).
23. A dialog box will appear requesting the minimum number of gexels to consider as part of a gene expression pattern (10 gexels by default) (Figure 5B).

Note: The minimum number of contiguous gexels for defining a gene expression pattern depends on the size of the tissue under study, its complexity as well as the resolution of the SrT data.

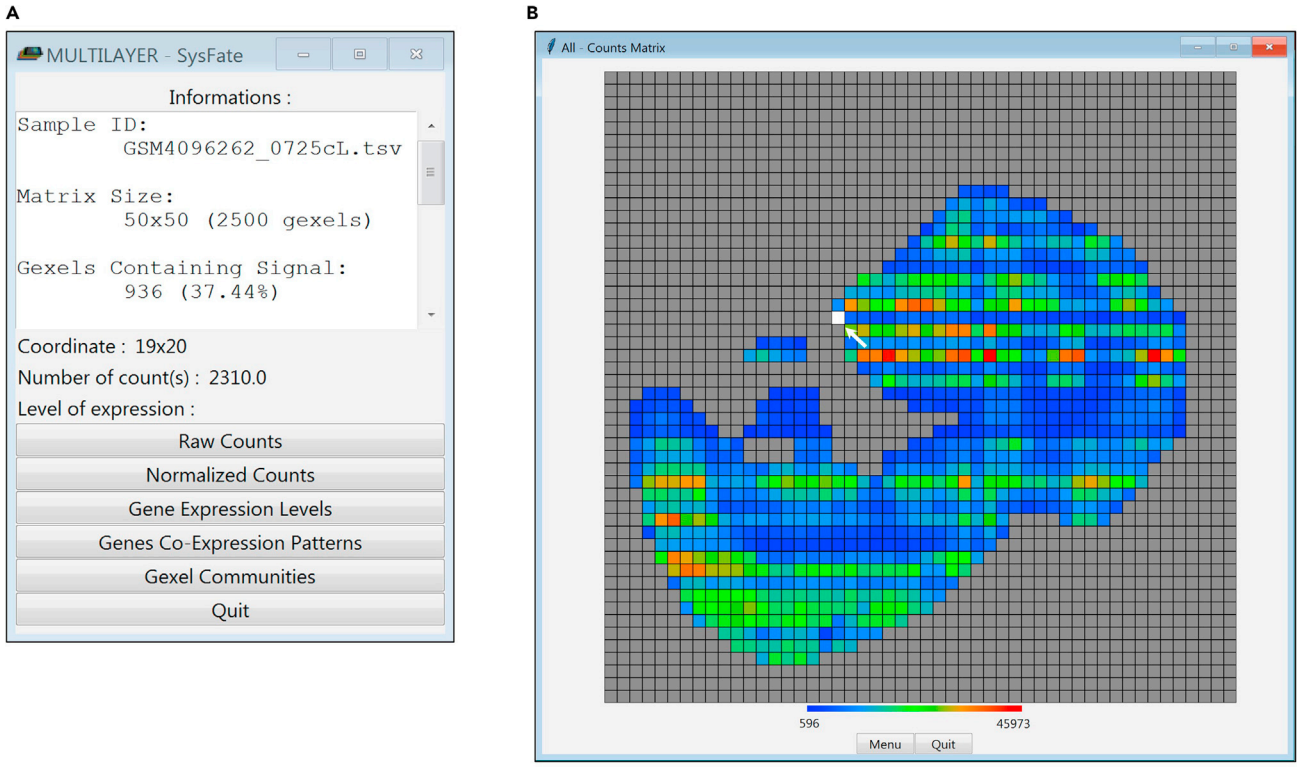


Figure 4. Visualizing a Whole Mouse Embryo data with MULTILAYER

(A) MULTILAYER interface providing general information, including the dataset sampel ID, matrix size, and the fraction of gexels presenting read counts.

(B) Raw count matrix displayed when selecting "raw counts" on the MULTILAYER interface. Each read count matrix (raw, Normalized, etc.), is accompanied by a menu bar dedicated to query for genes of interest. Read count levels per gexel are displayed as a heatmap. Hovering over a gexel on the image (indicated by white square and the pointer arrow) gives information of that particular area on the MULTILAYER interface (coordinate, number of counts or level of expression when visualizing the gene expression level matrix). The heatmap refers to the total read counts retrieved per gexel.

24. After clicking on "Run patterns", a matrix will be displayed. Like in the previous cases, the Menu button displays a panel in which gene patterns are ranked by the number of occupied gexels (Figure 5B).
25. To visualize the gene pattern location, users can either enter the gene name in the search panel, or double-click on the ranked list.
26. In some cases, multiple patterns (represented by different colors) can be displayed for a given gene (Figure 5B).

Note: Multiple patterns per gene potentially correspond to biologically relevant events, notably if they are separated by several gexels (excluding a technical discontinuity).

27. Select a pattern by clicking on one of their associated colored gexels.
28. Select the minimal gene co-expression similarity (in percent).
29. By clicking on "similarity", MULTILAYER will compute spatial gene co-expression signatures for the selected pattern. Gene co-expression is displayed by a spread of gexels from the selected pattern (corresponding to other gexels associated to co-expressed genes), as well as a heatmap corresponding to the gene co-expression similarity; i.e., from red meaning 100% of co-expression similarity (defining the location of the initial gene query), till dark blue corresponding to the least co-expression similarity level. (Figure 6A). A comprehensive list of the co-expressed genes

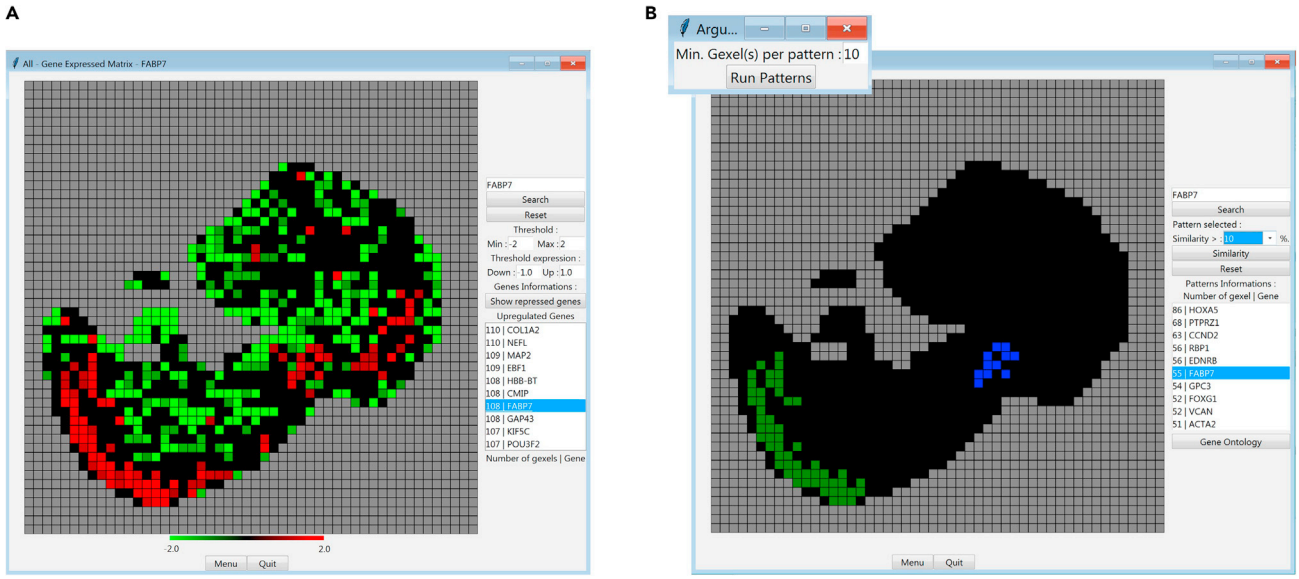


Figure 5. Revealing the differential gene expression within digitized tissue sections

(A) By selecting "Gene Expression levels" on the MULTILAYER interface, users can access a matrix view in which all precomputed differentially expressed genes are ranked on the basis of the number of their associated gexels. A double-click on the gene FABP7, retrieved on the ranked list, allows to visualize the over-expression (in red) or repressed (in green) signature of this gene over the whole tissue. The number at the left side of the Gene name (in this case "108") indicates the number of gexels within the whole tissue presenting a differential expression for that corresponding gene. The heatmap refers to the differential expression levels (log2) retrieved per gexel.

(B) when selecting "Gene co-expression patterns" on the MULTILAYER interface, a dialog box appears, allowing to select the minimal number of gexels per gene expression pattern. Like above, users can access a list of ranked genes, this time on the basis of the number of gexels per pattern. A double-click on the gene FABP7 reveals this time a cleaner view relative to (A), in which two distinct patterns (green and blue) complying with the minimal number of gexels per pattern criteria (at least 10) is displayed. For downstream analyses, users have to select a pattern of interest by clicking on one of their associated colored gexels.

and their corresponding similarity with the query gene is displayed within the Powershell prompt console (Figure 6B).

30. Click on "Gene Ontology" to access a dialog panel to select a GO terms database.
31. Select the adequate GO terms, then press "Run" to access to the Gene Ontology outcome, available either as a Barplot format or a gene vs GO terms heatmap (Figure 6C and 6D).

Note: In case users might require to interrogate a particular GO terms database; users can add the corresponding information to the folder "GO_DB" folder provided together with the MULTILAYER tool. For this purpose, users might adjust their GO terms database of interest to the format associated to the already available collections.

Visualize spatial communities defining tissue substructures

Timing: <10 min

While with the "Gene co-expression patterns" function we can query for a gene of interest and visualize the list of other genes that are spatially co-expressed, the "Gexel Communities" function allows to screen for all gene co-expression patterns over the whole tissue, and to then classify them within spatial communities. This is performed as follows:

32. When selecting "Gexel Communities" on the MULTILAYER interface a dialog box allows users to select the gene co-expression similarity threshold (in percentage), and also the possibility to perform 15 iterative runs and include the gene co-expression similarity levels as a "weight"

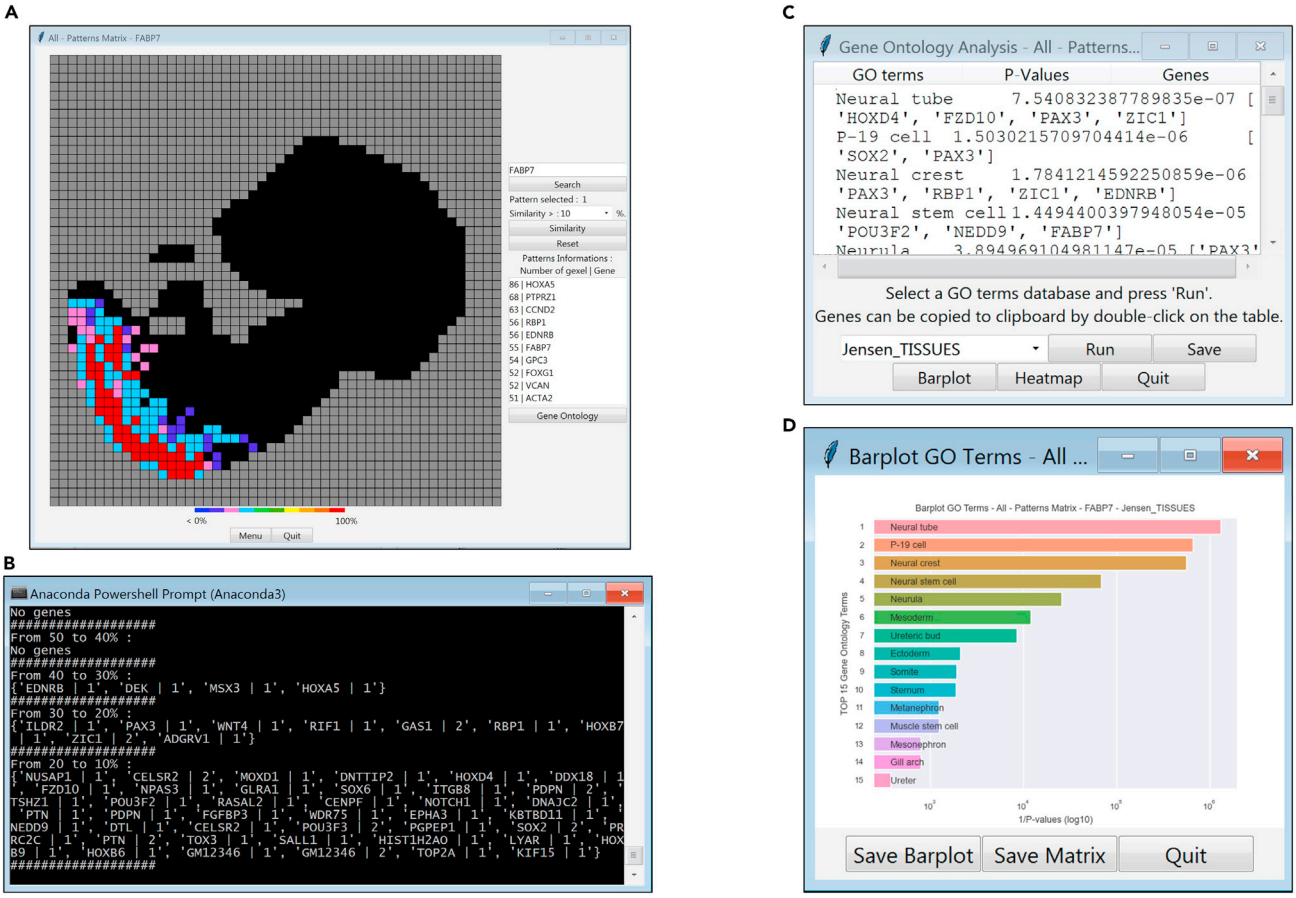


Figure 6. A functional analysis of co-expressed genes revealed by MULTILAYER

(A) Gene co-expression analysis for the spatial pattern associated to the gene FABP7. After selecting one of the FABP7 patterns shown in 5b (by clicking on one of the colored gexels associated to the pattern of interest), a gene co-expression similarity (minimal threshold: 10%) is computed. The illustrated matrix, displays gexels occupancy by spatial co-expressed genes with FABP7 (heatmap: similarity level with the query gene).

(B) A comprehensive list of the identified co-expressed genes is displayed within the PowerShell prompt.

(C) By clicking on the "Gene Ontology" button in (A), a new dialog panel allows to select for a pertinent GO terms database (here Jensen TISSUES). By pressing the Run button, MULTILAYER assesses the enrichment for the pertinent GO terms and displays their ranking on the basis of their confidence.

(D) Enriched GO terms associated to the co-expressed genes associated to FABP7 displayed as a Barplot.

parameter within the gene regulatory networks used for their spatial communities partitioning (Figure 7A).

Note: Louvain algorithm is performed over multiple iterations (default value: 15 times) and the most frequent outcomes are retained. This can be seen in the PowerShell prompt console.

33. By clicking the "Run Communities" button, MULTILAYER displays a matrix where the spatial communities are highlighted (color-coded gexels) (Figure 7A).

Note: When all communities are displayed, it's possible for communities to overlap. If there are overlaps between communities, the tool will color the gexel as part of the community with the highest similarity.

34. By choosing one community, the over-expressed genes colocalized in that region are listed (Figure 7B).

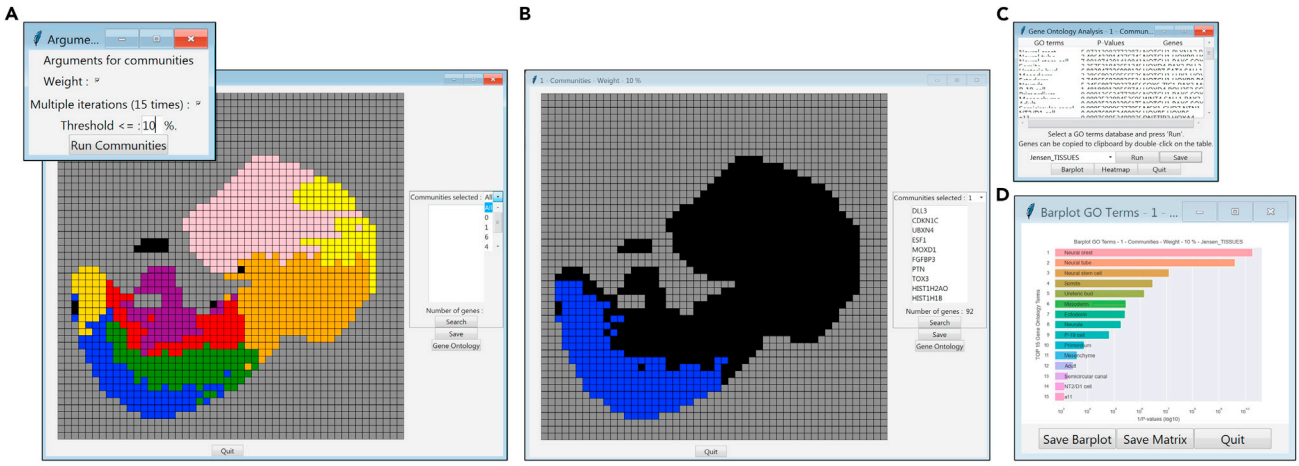


Figure 7. Revealing spatial tissue substructures by spatial communities partitioning

(A) Dialog panel defining the minimal gene co-expression similarity to detect gene co-expression Communities. By running this process, a matrix view in which all inferred gene co-expression communities within the tissue section is displayed. Note the presence of a scrolling bar (right side) by which users can access each of the communities.

(B) When selecting community "1", the list of the co-expressed genes and their corresponding spatial location is displayed.

(C) By clicking on Gene Ontology button in (B), a dialog panel allows to select for a pertinent GO terms database (here "Jensen_TISSUES").

(D) Barplot display of the enriched GO terms associated to the community "1" displayed in (B).

35. Similar to the gene expression analysis described on step 4, by clicking on the "Gene Ontology" button users can access the Gene ontology analysis (Figure 7C and 7D).

Note: A video illustrating the aforementioned steps is available here: https://www.youtube.com/watch?v=zByIdsUyJPg

Processing high-resolution SrT data

Step-by-step method of analyzing high resolution data. In this particular case, datasets issued from the technology Slide-seq (Rodriques et al., 2019) is used (Hippocampus Brain section). High resolution data are, in general, stored as a three-column format, instead of a gene/coordinate matrix. Prior to MULTILAYER processing, high-resolution files need to be condensed into larger gexels, thus reducing the number of gexels to display, but simultaneously gain on read counts per gexel.

Compress data with MULTILAYER compressor

Timing: 10 min

36. In order to compress the file, it is necessary to open the MULTILAYER compressor tool and also indicate the correct directory path to the data.
37. In order to access the compressor tool, type in the PowerShell prompt console:

```
> python3 Multilayer_compressor.py -i input.tsv -o output.tsv -cx 60 -cy 60
```

or

```
> python Multilayer_compressor.py -i input.tsv -o output.tsv -cx 60 -cy 60
```

For e.g.,

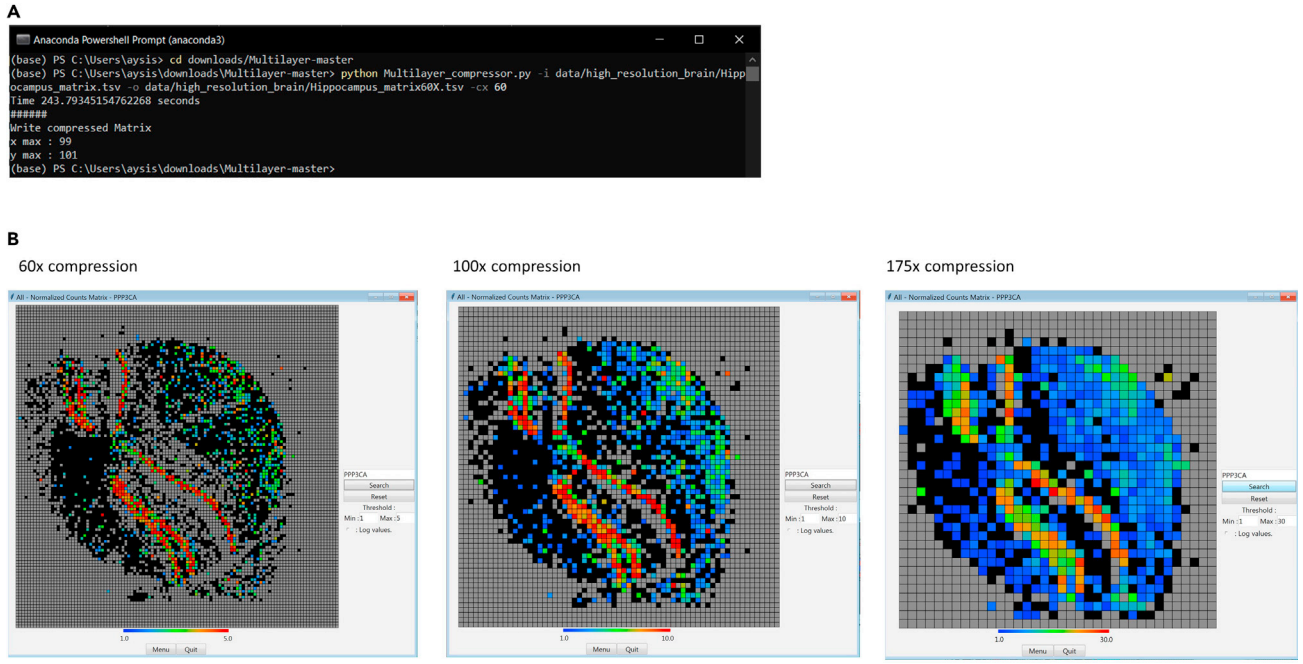


Figure 8. Processing high-resolution hippocampus SrT with MULTILAYER

(A) The PowerShell prompt console displaying the function to use the MULTILAYER compressor tool. The "-cx" parameter defined to "60", indicates a compression factor of 60 folds; i.e., the aggregation of proximal gexels such that the size of the matrix is reduced by 60 folds.

(B) High-resolution hippocampus spatial transcriptomics data compressed by a factor of 60×; 100× and 175× respectively and analyzed by MULTILAYER. Display of the spatial expression of the gene PPP3CA for these three compression factors demonstrates a conservation of its spatial signature. The heatmap refers to the total read counts retrieved per gexel.

```
> python Multilayer_compressor.py -i data/high_resolution_brain/Hippocampus_matrix.tsv
-o data/high_resolution_brain/Hippocampus_matrix60X.tsv -cx 60
```

(Figure 8A)

Note: The format of the input data has to be in 3 columns, where the first column correspond to barcode coordinates (XxY) with header 'bc'; the second column to genes with header 'gene'; and the third column to the gene counts with header 'count'.

The Multilayer_Compressor.py has several arguments: The function -i indicates the input matrix, -o indicates the output matrix, -cx and -cy indicates the compression factor (if -cx and -cy are identical only -cx needs to be defined). By typing

```
> python Multilayer_compressor.py -h
```

into the PowerShell console, you get access to the help section of the compressor.

38. The data will be stored in your output directory and can now be opened using the MULTILAYER tool.

Visualize compressed data on MULTILAYER

Timing: minutes

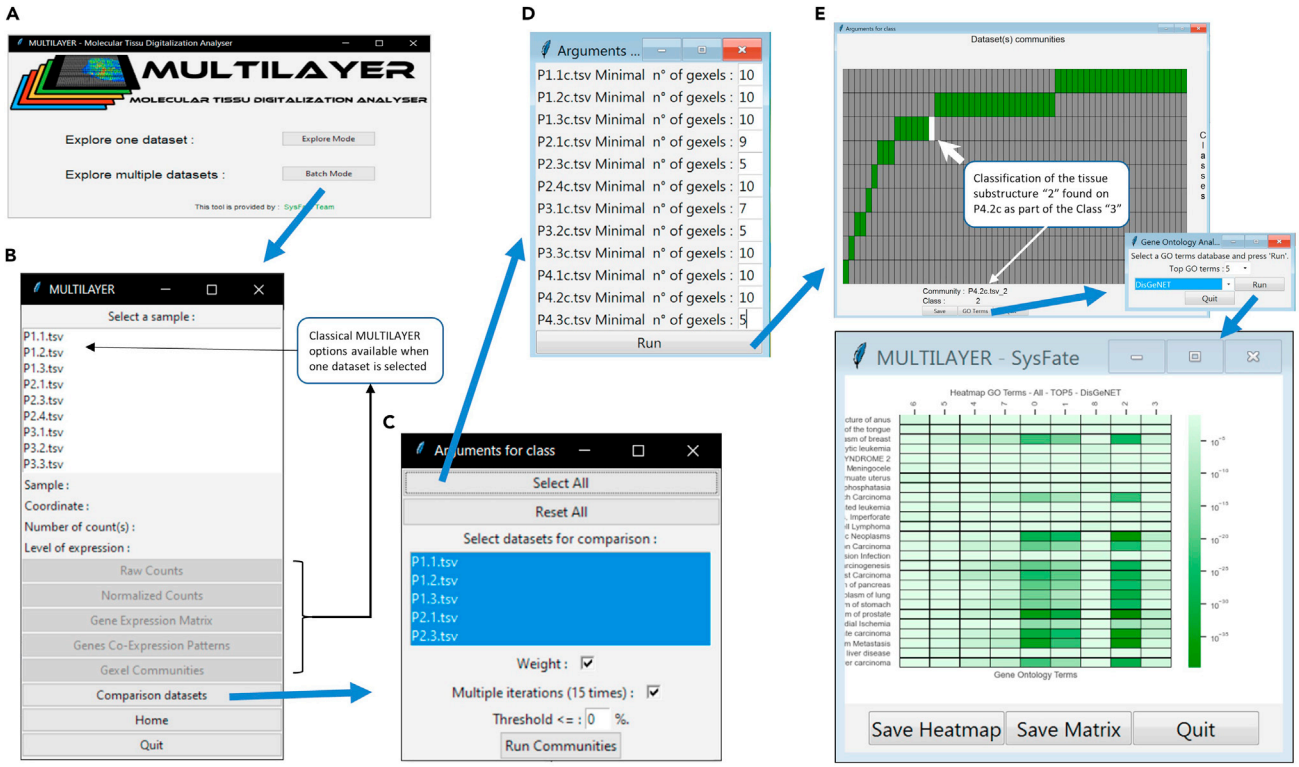


Figure 9. Use of the "batch-mode" available on MULTILAYER for processing several datasets at once

(A) MULTILAYER main interface allowing the selection of a folder containing all datasets to process.

(B) Once all datasets are processed (in this case a set of twelve prostate sections), MULTILAYER provides a similar interface than that used for the "explore mode", giving access to all different analytical panels when one dataset is selected. In addition, a "Comparison datasets" button is displayed.

(C) This "Comparison datasets" module, allows us to select the samples of interest for processing. Furthermore, it allows us to define the parameters for the spatial community's detection.

(D) For selected datasets, users can also define the minimal number of gexels for patterns detection.

(E) Panel illustrating the classification of all spatial communities detected per analyzed datasets. The interactive panel allows us to visualize the spatial tissue substructure ID and its corresponding class. This information can also be downloaded as a table for further analyses (e.g., graphs display). Furthermore, this panel allows us to access the gene ontology analysis as described for the "explore" mode.

39. After being compressed, datasets can be visualized on MULTILAYER tool as described above in the example of Whole Mouse Embryo (Figure 8B).

Note: A video illustrating the aforementioned steps is available here: https://www.youtube.com/watch?v=ww4x2aP6ENA

Processing multiple SrT data at once

Step-by-step method of using the batch-mode option. To illustrate this module, we have used twelve prostate cancer datasets generated by Berglund et al. (2018).

Open data on MULTILAYER

Timing: 10 min

40. Use Batch mode for this data (Figure 9A).

Note: The Batch mode is useful for analyzing multiple datasets from the same sample.

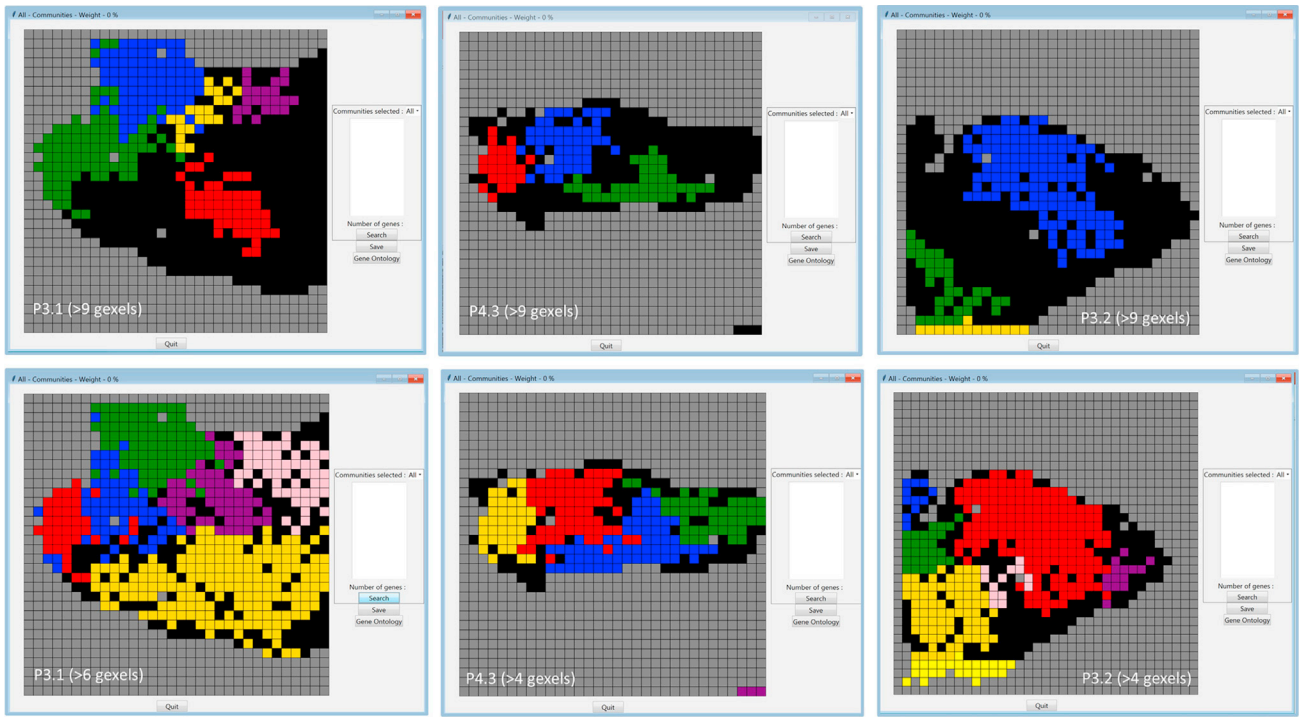


Figure 10. Spatial communities revealed over prostate sections as computed with different minimal number of gexels for pattern detection

Three of the twelve prostate sections analyzed in Figure 9, are displayed (P3.1, P4.3 and P3.2). When using a minimal number of gexels of 10 (i.e., >9) for pattern detection, several gexels within the tissues are not associated to any of the inferred spatial communities (black-colored gexels; Top panels). By decreasing the minimal number of gexels criteria, it is possible to maximize the number of gexels associated to any of the inferred spatial communities (Bottom panels).

41. Provide your dataset by clicking Raw matrix and select the appropriate folder where all datasets to process are available (in this case the "Prostate Cancer" datasets). All .tsv files in the selected folder will be considered as the input matrix.
42. All values are kept as default values.
43. Click next.

Note: Since multiple files are being processed, this step takes ~6 min.

44. You are now able to visualize digitally, the raw counts, normalized counts etc. for individual samples as in the example for Whole Mouse Embryo or a comparison dataset for several samples at once on the basis of their associated spatial communities (Figure 9B).

Compare multiple spatial communities on MULTILAYER

Timing: minutes

45. On the Multilayer interface, select "comparison datasets" and select all the files to be compared ("Select All" button) or make your own selection (Figure 9C).
46. Default values for community's detection are kept constant for all datasets (weight is checked, multiple iterations (15 times) is checked, threshold is 0%). Furthermore, users can adjust the minimal number of gexels for patterns detection (Figure 9D). This option is of major interest in case users have previously identified the minimal number of gexels (Figure 10).

Note: Users can empirically evaluate the minimal number of contiguous gexels for the analysis, by evaluating the fraction of gexels that are associated to at least one of the inferred spatial communities. As illustrated in Figure 10, by decreasing the minimal number of gexels from 10 to 5, the spatial communities retrieved within the displayed prostate tissue sections managed to incorporate most of the gexels of the tissue within the inferred patterns.

47. Run communities and save the .tsv file.

Note: When saving the files for the comparison analysis, the tool saves 2 files (ex. save.tsv, save_filter.tsv).

48. An Interactive heatmap comes up with all the dataset communities. The x-axis represents all the dataset communities, and the y-axis represents all the classes. A class is a group of 1 or several communities. (Figure 9E).
49. For gene ontology analysis, click GO terms.
50. Select the number of GO terms and database. In this example, top GO terms =5, database= DisGeNET.
51. This gives a heatmap of the GO terms that can be saved as a plot or matrix (Figure 9E).

EXPECTED OUTCOMES

MULTILAYER provides an intuitive platform for processing SrT data issued from a variety of technological platforms (e.g., DNA-array based methodologies (Rodriques et al., 2019; Ståhl et al., 2016), microfluidic channels (Liu et al., 2020)). Furthermore, thanks to additional modules, MULTILAYER can also analyze high-resolution data (MULTILAYER compressor module), as well as those issued from the commercial platform "Visium" (Visium converter module).

As outcome, MULTILAYER processes SrT maps as a digital image, in which users can visualize gene expression signatures, reveal gexel patterns resulting from agglomerative clustering, and even infer biologically relevant tissue substructures.

Considering the future needs for processing multiple datasets on a comparative manner, MULTILAYER also provides a "batch-mode", allowing to process multiple samples at once, as well as compare them, notably by classifying the assessed tissue substructures per dataset into groups that might reveal common biological functions.

LIMITATIONS

The current version of MULTILAYER reveals spatial gene co-expression signatures, but it is not able to infer their related co-regulatory relationships. Furthermore, it is not yet adapted for integrating single-cell transcriptome data. MULTILAYER is currently in progress to be updated with these two aspects. From a technical angle, MULTILAYER is limited with its visual functionalities for datasets represented within matrices > 250 gexels per side. In such a situation, the use of the MULTILAYER compressor module is strongly recommended.

TROUBLESHOOTING

The most common problem when running MULTILAYER is the format of the input dataset.

Problem 1

MULTILAYER fails to process the input dataset because the matrix has spatial coordinates as rows and genes as columns.

Potential solution

Use the "Transpose" option at the time of the data upload.

Problem 2
MULTILAYER fails to process the input dataset because the matrix has decimal values within the gene counts.

Potential solution
Use the "Round" option at the time of the data upload.

Problem 3
MULTILAYER fails to process the input dataset despite taking care of transposing the matrix or rounding off the read counts.

Potential solution
The provided data might miss values for the read counts. Make sure to fill empty cells or replace "NAs" by a value ("0" by default).

Problem 4
MULTILAYER fails to process the input dataset, notably because it has utilized all the available RAM of the computer.

Potential solution
High-resolution data might require large amounts of RAM for processing. Either use the "MULTILAYER Compressor" module to reduce the matrix size, or increase the RAM of your system.

Problem 5
Multilayer fails to process when the gene identifiers are not unique within the provided dataset.

Potential solution
To avoid this issue, we recommend using unique identifiers, for instance corresponding to transcripts (e.g., EnsemblID). Since the gene ontology analyses requires access to the gene symbol ID, an optimal solution is to concatenate both gene symbol and transcript ID into a single identifier separated by a "_" (e.g., 'geneSymbol_ensemblID'). MULTILAYER is able to separate such a gene ID structure to recognize the first part of the identifier as the corresponding gene symbol to be used for the gene ontology processing when required.

Problem 6
MULTILAYER processing does not provide gene patterns and/or spatial communities.

Potential solution
The processed data does not have sufficient differential expression data, or the minimal number of contiguous gexels is not satisfied. To evaluate such potential reasons, decrease the minimal number of contiguous gexels and/or decrease the differential gene expression threshold.

Problem 7
The predicted spatial communities do not cover the whole surface of the tissue (i.e., a significant part of the tissue displays black gexels, not associated with any of the predicted communities).

Potential solution
The minimal number of contiguous gexels for patterns detection is too high. Evaluate multiple lower thresholds to identify empirically the least number of contiguous gexels that allows to predict spatial communities within the whole (maximum) surface of the tissue.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Marco Antonio Mendoza-Parra (mmendoza@genoscope.cns.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The datasets and code used during this study are available at https://github.com/SysFate/MULTILAYER

ACKNOWLEDGMENTS

This work was supported by the institutional bodies CEA, CNRS, and Université d'Evry-Val d'Essonne. J.M. was supported by Genopole Thematic Incentive Actions funding (ATIGE-2017); A.K. by the "Fondation pour la Recherche Medicale" (FRM; funding ALZ°201912009904); and F.S. by the funding 2OI9-L22 from the Institut National du Cancer (INCa).

AUTHOR CONTRIBUTIONS

Conceptualization, J.M. and M.A.M.-P.; protocol elaboration, A.K., J.M., and M.A.M.-P.; software development, J.M.; "Visium converter" development, F.S.; scientific evaluation, M.A.M.-P.; writing, review, and editing, J.M., A.K., and M.A.M.-P.; funding acquisition, M.A.M.-P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat. Commun. *9*, 2419.

Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell *183*, 1665–1681.e18.

Moehlin, J., Mollet, B., Colombo, B.M., and Mendoza-Parra, M.A. (2021). Inferring biologically relevant molecular tissue substructures by agglomerative clustering of digitized spatial transcriptomes with multilayer. Cell Systems *21*, 694–705. https://doi.org/10.1016/j.cels.2021.04.008.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Materials and Methods

Materials

Cell Lines

The CARLIN MESC cell line was kindly provided by Sarah Bowling [146]. The system relies on a Doxycycline-inducible Cas9 activation that binds any of the 10 guide RNAs that can target 1 of 10 target sites in a molecular barcode/array known as CARLIN (CRISPR Array Repair LINeage tracing).

The three cell lines- H3.3WT, H3.3G34R, H3.3K27M- were kindly provided by the lab of Dr. Ali Hamiche, Functional genomics, and cancer, IGBMC (Institut de Genetique et de Biologie Moleculaire et Cellulaire), Strasbourg. These cell lines have been generated by the Phenomin-Institut Clinique de la Souris, Alsace, France. To introduce mutations, plasmid vectors containing loxP flanked Neomycin Resistance gene (NeoR) were transfected into mouse embryonic stem cell lines (MESC) C57BL/6NCrl S3 via electroporation. Successfully transfected MESC cells were then selected based on neomycin resistance. Identified clones were then submitted to the screening process with PCR, ddPCR aneuploidy, Southern blot, and Giemsa staining. These cell lines, after Cre-mediated excision, express the H3F3A protein with the associated mutation for that cell line along with an HA-tag, except for wild type. The three cell lines in this paper have been abbreviated to WT Glio-wild type, K27M-with K27M mutation in H3.3 gene, G34R- with G34R mutation in H3.3 gene.

MYC plasmid construct

The Myc overexpression plasmid (MYCOE or MYC) and Control plasmid (MYC CONTROL) were kindly provided by the team of Jürgen A. Knoblich, Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna, Austria [120]. For the overexpression constructs, based on the Sleeping Beauty Transposase System, the CMV promoter from pCMV(CAT)T7-SB100 (Addgene cat. No.: 34879) [147] was replaced with CAG promotor from pCAGEN (Addgene cat. No.: 11160) [148]. IRDR-R and IRDR-L sequences from pT2/LTR7-GFP (Addgene cat. No.: 62541) [149] were cloned into pCAGEN to produce pCAG-GS/IR. cDNAs used for overexpression were amplified from human cDNA and cloned into the MCS of pCAG-GS/IR. With the help of sleeping beauty transposase SB100X (pCAG-SB100X), CAG-GFP and

CAG-oncogenes get integrated into the genome of cells in organoids. Both plasmid constructs contain a GFP tag.

Reagents

1. DMEM F12 10x500ml (Thermo Fisher, cat. no.11330057)
2. DMEM W/ GLUTAMAX-I, PYR,4.5G GLU 500ML (Thermo Fisher, cat. no.31966021)
3. D-MEM (LG)W/NA PYR.(CE) 500ML (Thermo Fisher, cat. no.31885023)
4. Neurobasal MED SFM 500ML (Thermo Fisher, cat. no.21103049)
5. Penicillin Streptomycin SOL 100ML (Thermo Fisher, cat. no.15140122)
6. B-27 Supplement W/O Vita (50X) 10 ML (Thermo Fisher, cat. no.12587010)
7. B 27 Supplement 10ML (Thermo Fisher, cat. no.17504044)
8. N2 Supplement 5ML (Thermo Fisher, cat. no.17502048)
9. GlutaMAX supplement (Thermo Fisher, cat. no.35050-038)
10. MEM NEAA (Thermo Fisher, cat. no.11140050)
11. Embryonic stem-cell FBS, qualified, USDA-approved regions (Thermo Fisher, cat. no.10439024)
12. Embryonic stem-cell FBS, qualified, US origin (Thermo Fisher, cat. no.16141079)
13. LIF Recombinant mouse protein (Thermo Fisher, cat. no.A35933)
14. PBS, Dubeccos W/O CA,MG(1X)(CE) (Thermo Fisher, cat. no.14190094)
15. Poly-D-Lysine 100 ML (Thermo Fisher, cat. no.A3890401)
16. Beta mercaptoethanol (Thermo Fisher, cat. no.31350010)
17. Corning™ Matrice qualifiée CSEh Matrigel™ 354277 5mL (Thermo Fisher, cat. no.11573560)
18. CF-1 MEF MITC-Treated 2M (Thermo Fisher, cat. no.A34958)
19. Opti Mem I 100ML (Thermo Fisher, cat. no.31985062)
20. Trypan Blue solution (Thermo Fisher, cat. no.15250061)
21. Lipofectamine 2000 reagent 0.3 ML (Thermo Fisher, cat. no.11668030)
22. Dimethyl sulfoxide (Sigma, cat. no.D2438-50ML)
23. Accutase solution (Sigma, cat. no.A6964-100ML)
24. Heparin (Sigma, cat. no.H3149-10KU)
25. Sucrose (Sigma, cat. no.S7903-250G)
26. Insulin Human, Recombinant expressed (Sigma, cat. no.I2643-50MG)

27. SB 431542 (Tocris cat. no. 1614)
28. BMS753, BMS641, BMS961 (Lab of Hinrich Gronemeyer, IGBMC Strasbourg)
29. Mouse Embryonic Stem Cell Nucleofector Kit (Lonza cat. no.VPH-1001)
30. Plasmide CRISPR/Cas9 KO p53 (m2) (Santa Cruz cat. no.sc-423509-KO-2)
31. Plasmid HDR p53 (m2) (Santa Cruz cat. no.sc-423509-HDR-2)
32. Plasmide de Contrôle CRISPR/Cas9 (Santa Cruz cat. no.sc-418922)
33. XL10-Gold Ultracompetent Cells (Agilent cat.no. 200314)
34. S.O.C. Medium (Thermo Fisher cat.no. 15544034)
35. QIAGEN Plasmid Midi Kit (25) (Qiagen cat. no. 12143)
36. Donkey anti-Mouse IgG (H+L) Secondary Antibody, Alexa Fluor 555 (Thermo Fisher, cat. no. A-31570)
37. Donkey anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 488 (Thermo Fisher, cat. no.A-21206)
38. ProLong™ Diamond Antifade Mountant with DAPI (Thermo Fisher, cat. no.P36962)
39. BSA 50mg/mL UltraPure (Thermo Fisher, cat. no.AM2616)
40. KIT, HIGH CAPACITY CDNA RT (Thermo Fisher, cat. no.4368814)
41. O.c.t. compound 125 ml (Dutscher cat. no. 2581596)
42. QuantiTect SYBR Green PCR Kit(1000x50µl) (Qiagen cat. no. 204145)
43. RNeasy Mini Kit (50) (Qiagen cat. no. 74104)
44. Triton X100 solution (Sigma, cat. no.93443-100ML)
45. Anti-Sox2 Antibody Produced in RABBIT (Sigma, cat. no.AB5603)
46. Anti-KI67 Antibody Produced in RABBIT (Sigma, cat. no.SAB5700770-100UL)
47. Anti-beta III Tubulin (Tuj) [2G10] - Neuronal Marker (Abcam cat. no. ab78078)
48. Anti-GAD67 antibody [K-87] (ab26116) (Abcam cat. no. ab26116)
49. CD133 Rabbit recombinant multiclonal [RM1029] (Abcam cat. no. ab284389)
50. GFAP Recombinant antibody [EPR1034Y] (Abcam cat. no. ab68428)
51. MAP2 antibody - Neuronal Marker (Abcam cat. no. ab32454)
52. Dual Index Kit TT Set A 96 rxns (10X Genomics 1000215)
53. Chromium Single Cell 3' GEM kit v3 (10X Genomics PN-1000094)
54. Chromium Single Cell 3' Library kit v3 (10X Genomics PN-1000095))
55. Chromium Single Cell 3' Gel Bead kit v3 (10X Genomics PN-1000093)
56. Chromium Next GEM Chip G Single Cell Kit (10X Genomics 1000127)

57. Dynabeads™ MyOne™ Silane (Thermo Fisher cat. no. 37002D)

Instruments

1. Lonza Nucleofector Transfection 2b Device
2. Stratagene Mx3005P qPCR Instrument
3. NanoDrop ND-1000 Spectrophotometer
4. 10X Genomics Chromium Controller

Oligo sequences used for RTqPCR assay and Sanger Sequencing

No.	Oligo Name	Sequence
1	36B4-F	AATCTCCAGAGGCACCATTG
	36B4-R	CCGATCTGCAGACACACACT
2	Sox2-F	TGCCAATATTTTTCGAGGAAAG
	Sox2-R	CGGTCCTAAATTTAACTGCAGAAT
3	NANOG-F	GCCTCCAGCAGATGCAAG
	NANOG-R	GGTTTTGAAACCAGGTCTTAACC
4	OCT4-F	GCATTAGCATCACCATGCAG
	OCT4-R	CCGGTTACAGAACCATCCTC
5	KLF4-F	CGGGAAGGGAGAAGACACT
	KLF4-R	GAGTTCCTCACGCCAACG
6	Ascl1-F	GCTCTCCTGGGAATGGACT
	Ascl1-R	CGTTGGCGAGAAACACTAAAG
7	Nestin-F	CTGCAGGCCACTGAAAAGTT
	Nestin-R	TCTGACTCTGTAGACCCTGCTTC
9	Tubb3-F	GGCAACTATGTAGGGGACTCAG
	Tubb3-R	CCTGGGCACATACTTGTGAG
10	GFAP-F	CAGCTTACGGCCAACAGTG
	GFAP-R	GCCTCAGGTTGGTTTCATCT
11	Olig2-F	AGACCGAGCCAACACCAG
	Olig2-R	AAGCTCTCGAATGATCCTTCTTT
12	Tal2-F	TGCGACAGCTACCTTGACTG
	Tal2-R	GCTCCCTGGTATTTGTGAAGAT
13	NeuroD1-F	CGCAGAAGGCAAGGTGTC
	NeuroD1-R	TTTGGTCATGTTTCCACTTCC
14	Neurog2-F	ACATCTGGAGCCGCGTAG
	Neurog2-R	CCCAGCAGCATCAGTACCTC
15	TPH2-F	GAGCTTGATGCCGACCAT
	TPH2-R	TGGCCACATCCACAAAATAC

No.	Oligo Name	Sequence
16	GLUT1-F	CTGGGGTCCTTGTGCAGTAT
	GLUT1-R	CCGAAGCTGCCATAGACATAG
17	GAD67-F	TGGAGATGCGAACCATGAG
	GAD67-R	GAAGGGTTCCTGGTTTAGCC
18	Th-F	AAGATCAAACCTACCAGCCG
	Th-R	TACGGGTCAAACTTCACAGAG
19	ki67-F	TCCTACCTTAAGACGGCGGA
	ki67-R	ACTTGCTCACACTCGATGCA
20	CD133-F	AAGAGCCATCCACCAGCATC
	CD133-R	GTTGCTTGTTTGCTGGAGGG
21	CD44-F	AGAAGAGCACCCCAGAAAGC
	CD44-R	TCTGAAACCACGTCTCCTGC
22	PDGFRA-F	ATCCACACATGCGTCCTTGT
	PDGFRA-R	TCGGACCACTGACAGAAAGC
23	TP53-F	TGCATGGACGATCTGTTGCT
	TP53-R	AAAGATGACAGGGGCCATGG
24	human MYC-F	TTCGGGTAGTGGAAAACCAG
	human MYC-R	CACCGAGTCGTAGTCGAGGT
25	IBA1- F	TGCCAGCCTAAGACAACCAG
	IBA1-R	GCTTCAAGTTTGGACGGCAG
26	TREM2-F	GGGTCACCTCTAGCCTACCA
	TREM2-R	GGATCTGAAGTTGGTGCCCA
27	CHAT-F	AAGCTTCCACGCCACTTTC
	CHAT-R	AGAGCCTCCGACGAAGTTG
28	TP21-F	GGTGATGTCCGACCTGTTCC
	TP21-R	CAGCAGGGCAGAGGAAGTAC
29	H3F3A_Seq_F	TGGCTGTTTTATCGCTCGGA
	H3F3A_Seq_R	ACTTTACCACAAATGAGGAGCTG

Methods

Medium Preparation

MES medium (250ml)

206ml DMEM (4,5 g/l glucose) w/GLUTAMAX-I, 37.5ml FCS ES-tested (15%) , 2.5ml Penicillin –Streptomycin , 2.5ml NEAA , 1ml LIF , 0.5ml beta mercaptoethanol, filtered using 0.22 um filter

MWOL medium (250ml)

220ml DMEM (4,5 g/l glucose) w/GLUTAMAX-I, 25ml FCS ES-tested (10%) , 2.5ml Penicilline –Streptomycine , 2.5ml NEAA , 0.5ml beta mercaptoethanol, filtered using 0.22 um filter

Neurobasal, B27 medium (50ml)

50ml Neurobasal medium, 0.5ml Penicilline –Streptomycine, 0.5 ml Glutamax, 1ml B27 without vitamin A

Neural Induction (NI) Medium (50ml)

50ml DMEM/F12, 0.5 ml N2 supplement, 0.5 ml Glutamax supplement, 0.5 ml MEM-NEAA, 0.5ml Penicilline –Streptomycine, 50 µl Heparin solution

Maturation Medium (150ml)

75 ml DMEM/F12, 75 ml Neurobasal, 3 ml B27 without vitamin A supplement, 1.5 ml Glutamax supplement, 1.5 ml Penicilline –Streptomycine, 750 µl N2 supplement, 750 µl MEM-NEAA, 37.5 µl insulin.

Maturation Medium VITA (150ml)

75 ml DMEM/F12, 75 ml Neurobasal, 3 ml B27 with vitamin A supplement, 1.5 ml Glutamax supplement, 1.5 ml Penicilline –Streptomycine, 750 µl N2 supplement, 750 µl MEM-NEAA, 37.5 µl insulin.

Cell Culture Maintenance

All stem cells were maintained in a 5% CO2 (carbon dioxide) incubator at 37°C. The MESCs received were feeder-dependent. One day before culturing MESCs, a six-well plate is coated with 1ml 0.1% gelatin and incubated for at least 30 minutes at 37°C. Then, 1ml vial of mitomycin – inactivated mouse embryonic fibroblast cells (MEFs) is thawed and mixed into 10 ml of warm MEF media. The conical tube is then centrifuged for 2 minutes, RT, at 1000 rpm. The supernatant is discarded and the pellet resuspended in 2ml of MEF media. The gelatin is aspirated and replaced with 300,000 cells/well in 2ml of MEF media, swirled, and shaken well to allow for good distribution and incubated at 37°C overnight.

Next day, MEFs are verified to have been properly attached and distributed all over the well. MESCs are removed from the liquid nitrogen tank and thawed in a 37°C water bath and mixed into 10ml mouse embryonic stem cell media (MES). The tube is then centrifuged for 2 minutes, RT, 1000 rpm, and the supernatant discarded. The pellet is now resuspended in 2ml of MES media. The MEF-coated plates should be checked for properly adhered MEF cells under a light microscope, and the supernatant aspirated. Add the MESCs in as many plates as required and fill the volume of each well up to 2ml with MES media.

Cell lines are passaged at 80-90% confluence. The spent medium is removed from well and replaced with 1ml trypsin/ accutase per well and incubated for 5 minutes at 37°C. Detached cells are now transferred to a conical tube with 5ml MES media; another rinse of the well with 1 ml MES media is performed and collected as well. Centrifugation of 2 minutes, RT, 1000 rpm. The supernatant is discarded and the pellet resuspended in 2ml of media. These cells were now passaged onto MEF-coated plates at a dilution of 1:3 or 1:6 depending on confluency percentage. A medium change is performed every day, to replenish nutrients.

Embryoid Body Formation

Embryoid bodies (EBs) can be formed by allowing the stem cells to grow in suspension without adhering to a surface. At the time of passaging the MESCs, the remaining cell culture from the passage is added to 10 ml of MES media in 10cm bacteriological plates, for each cell line. These are incubated at 37°C.

After 2 days, EBs are visible and they can either be replenished with a medium change or passaged. Medium change is done by transferring the suspension to a 15ml conical tube

using a wide mouth pipette (5 or 10ml). Let the EBs settle down for 5-7 minutes. The single cells will remain in suspension while the heavier EBs will settle down. Discard the supernatant, and resuspend in 10 ml of MES media, mixing gently and add to a new 10cm bacteriological plate. This is done every alternate day to make sure that single cells, dead cells, and MEFs have been filtered out from the suspension.

To passage EBs, perform transfer the suspension to a 15ml conical tube using a wide mouth pipette (5 or 10ml). Let the EBs settle down for 5-7 minutes. Aspirate the supernatant, and add 1ml trypsin/ accutase per well and incubate for 5-10 minutes at 37°C. Supplementary dissociation can provided by mechanically dissociating the EBs using a P1000 micropipette. The trypsin/ accutase is now diluted with 5ml MES media and centrifuged for 2 minutes, RT, 1000 rpm. The supernatant is discarded and the pellet resuspended in 1ml of media. These cells are now ready to be passaged into a new bacteriological plate at a dilution of 1:3 or 1:6 for renewed EB production or directly used for experimentation.

2D Neuronal Differentiation

Prior to starting a neuronal differentiation assay, it is necessary to confirm that MESC are in EB formation in bacteriological plates, with MES media; are able to form smooth round EBs with no signs of auto differentiation (cells attached to plate surface, no dark centers observed in the center of the EBs); large quantities of EBs present 2-3 days after each passage (cells are highly proliferative and healthy); and no MEF cells present in the culture (EBs should be passaged for at least 2-3 times till no visible attached cells seen in the plate).

On Day 0, in 6-well TC treated plates, a mixture of 1X PBS and Poly-D-Lysine (1:1) is added at 1ml/well and kept under the Biosafety Cabinet Hood at room temperature for 1H minimum. In case immunostaining assays are to be carried out, coverslips are placed in the wells before adding the Poly-D-Lysine: PBS mixture. The solution is then washed out with sterile ddH20 3 times for 5 minutes each and left to dry with the lid open under the Hood. Once the wells are completely dry, medium can be added.

The EBs are passaged as mentioned above and resuspended in 1ml of MES without LIF (MWOL) medium. This medium contains no LIF (Leukocyte inducing factor) and less serum to push the stem cells towards a differentiation process. Cells are plated at different densities depending on when they are collected. For collection at 4 days, cell seeding is 100,000/ well;

at 8-10 days, cell seeding is 20-40,000/ well. at 16 days, cell seeding is 10-20,000/ well. Cells are kept in MWOL medium. At this point cells can be collected for RNA collection and RT-qPCR analysis as reference control point.

On Day 1, ensure proper attachment of cells. Synthetic ligands are added at the following concentrations- ATRA, BMS753 (1µM); BMS641, BMS961, BMS641+961 (0.1µM). EtOH is added at the same volume as ATRA. The cells are kept in 2-4ml of MWOL medium+ ligands for 4-8 days respectively. On Day 4 or Day 8, medium is completely aspirated and changed to Neurobasal medium + B27 media and incubated for another 6-8 days. Half medium is refreshed every alternate day. At midpoint of 4-8 days, or end point of 10-16 days, cells are collected for RNA extraction and RT-qPCR analysis, or immunostaining assay.

Mouse brain organoid generation version 1

MESCs should be present in culture as EBs, as clean cultures, highly proliferative and devoid of MEF. On Day 0, The EBs are passaged as mentioned above and resuspended in 1ml of MWOL medium+ SB 431542 (1/1000). Cells are counted and placed in each well of P96 (ultra low attachment) plate at 2000 cells/ well at 150µl medium/ well and incubated for 2 days at 37°C. On Day 2, medium is refreshed with MWOL medium+ SB 431542 (1/1000) by taking out 75µl and adding 150µl fresh medium per well. On Day 4, medium is refreshed with MWOL medium only by taking out 120µl and adding 150µl fresh medium per well. On Day 6, transfer EBs to P24 (ultra low attachment) plate (4 EBs/ well) using a cut P200 tip, and medium is refreshed by completely aspirating out MWOL medium and adding 500µl NI medium/ well. On Day 8, add 500µl NI medium/ well.

On day 10/11, EBs are ready to be embedded in Matrigel GF reduced. The previous day (or 1H prior to use), thaw Matrigel GF reduced in ice at 4°C. Prepare parafilm dimple sheets (cut a square piece of parafilm and press dimples with thumb on an empty pipette box rack) or use silicon embedding sheets placed in a sterile bacteriological plate. Transfer 1 EB to 1 dimple using a cut P200 tip. Remove as much medium as possible surrounding the EBs. Add 20µl Matrigel GF reduced per dimple using cold pipette tips. To polymerize, incubate the dish 30-45 minutes at 37°C in the closed bacteriological plate.

3ml Maturation medium is added to P6 (ultra low attachment) plates and embedded EBs are transferred to the wells (10-15 per well) using a cut tip or a spatula. The EBs, now BORGs,

are incubated for 2-3 days undisturbed at 37°C. The medium is now changed completely to Maturation medium VITA or Maturation medium+ ligand of choice (ATRA, BMS753 (1µM); BMS641, BMS961, BMS641+961 (0.1µM); EtOH is added at the same volume as ATRA) and placed on an orbital shaker at 75-80 rpm. BORGs need to be fed every 2-3 days (twice a week) by removing 2/3rd of medium and refreshing with fresh medium.

Mouse brain organoid generation version 2

This protocol has been derived from Lancaster et al, 2013 [83].

MESCs should be present in culture as EBs. On Day 0, The EBs are passaged as mentioned above and resuspended in 1ml of MWOL medium+ SB 431542 (1/1000). Cells are counted and placed in each well of P96 (ultra low attachment) plate at 5000 cells/ well at 150µl medium/ well and incubated for 2 days at 37°C. On Day 2, medium is refreshed with MWOL medium+ SB 431542 (1/1000) by taking out 75µl and adding 150µl fresh medium per well. On Day 4, medium is refreshed by taking out 150-170µl MWOL medium and adding 170µl NI medium/ well. On day 6, EBs are ready to be embedded in Matrigel GF reduced. The previous day (or 1H prior to use), thaw Matrigel GF reduced in ice at 4°C. Prepare parafilm dimple sheets or use silicon embedding sheets placed in a sterile bacteriological plate. Transfer 1 EB to 1 dimple using a cut P200 tip. Remove as much medium as possible surrounding the EBs. Add 20µl Matrigel GF reduced per dimple using cold pipette tips. To polymerize, incubate the dish 30-45 minutes at 37°C in the closed bacteriological plate. 3ml Maturation medium is added to P6 (ultra low attachment) plates and embedded EBs are transferred to the wells (10-15 per well) using a cut tip or a spatula. The EBs, now BORGs, are incubated for 2-3 days undisturbed at 37°C. The medium is now changed completely to Maturation medium VITA or Maturation medium+ ligand of choice (ATRA, BMS753 (1µM); BMS641, BMS961, BMS641+961 (0.1µM); EtOH is added at the same volume as ATRA) and placed on an orbital shaker at 75-80 rpm. BORGs need to be fed every 2-3 days (twice a week) by removing 2/3rd of medium and refreshing with fresh medium.

RNA extraction and cDNA generation

RNA was extracted, precipitated, and solubilized to use for reverse transcription and finally quantitative PCR. Total RNA was extracted using TRIzol^{TM} reagent or Qiagen RNeasy

Mini Kit. The extracted RNA is analyzed using Nanodrop, and 0.5-1µg of RNA is used to perform Reverse transcription using High-Capacity cDNA Reverse Transcription Kits.

RT-qPCR Assay

The cDNA from Reverse transcription was amplified with gene primers listed in Materials. Quantification was done using the QuantiTect SYBR Green PCR Kit on Stratagene Mx3005 Real-Time PCR System (Agilent Technologies). Comparative Ct (ΔΔCt) method was used for performing relative quantitation of gene expression. SYBR green qPCR amplification program was as follows: 95°C for 15min; 95°C for 15sec, 60°C for 30sec, 72°C for 45sec (40 cycles); 95°C for 1min, 55°C for 30sec, 95°C for 30sec.

Immunofluorescence assay

Immunocytochemistry

At the end of the experiment, medium is aspirated completely and cells are fixed in PFA 4% for 20 minutes. This solution is washed out 3 times with 1X PBS for 5 minutes each and stored in 1X PBS till they are ready to be stained. For pre-blocking, 0.1% Triton X100 is diluted in 1X PBS and cells are incubated for 10 minutes at RT followed by a 5 minute PBS wash, 3 times. For blocking step, 1% BSA, 0.1% Triton X100 is diluted in 1X PBS and incubation is for 30 minutes at RT. For Primary antibody incubation, antibodies are added at the following dilution to blocking solution- Tubb3:1/5000, Sox2:1/500, Gfap:1/500, Gad67:1/1000, KI67:1/5000, CD133:1/500. Cells are incubated with primary antibodies in a humid chamber at 4°C overnight. The next day, cells are put through a 5 minute PBS wash, 3 times. For Secondary antibody incubation, antibodies are added at the following dilution to blocking solution- Donkey anti-mouse:1/1000, Donkey anti-rabbit: 1/1000. Cells are incubated with Secondary antibodies in a humid chamber at RT, for 1H, followed by a 5 minute PBS wash, 3 times. DAPI Prolong™ Diamond Antifade Mounting solution is added and coverslips with cells mounted on labelled glass slides.

Cryo-sectioning and Immunohistochemistry

BORGs are collected for immunostaining by submerging in 4%PFA, 6% sucrose, 1X PBS solution for 20 minutes, and washed 3 times with 1X PBS for 5 minutes each. The BORGs are then suspended in 30% sucrose, 1X PBS solution at 4°C overnight. The next day, ensure that

BORGs are settled to the bottom of the tube (indicating absorption of sucrose solution that has replaced the PBS in the BORG; this prevents breakage of tissue during freezing). BORG is then placed in plastic square mold and any excess sucrose solution aspirated out before submerging the tissue in OCT. This mold is placed in dry ice chamber for 10 minutes to completely freeze before being stored long term at -80°C. BORGs are cryo sectioned using a cryo-star. The thickness of each tissue section is maintained at 14-18µm and collected on glass slides which can be stored long term at -20°C.

For pre-blocking, 0.1% Triton X100 is diluted in 1X PBS and tissue is incubated for 10 minutes at RT followed by a 5 minute PBS wash, 3 times. For blocking step, 1% BSA, 0.1% Triton X100 is diluted in 1X PBS and incubation is for 30 minutes at RT. For Primary antibody incubation, antibodies are added at the following dilution to blocking solution- Tubb3:1/5000, Sox2:1/500, Gfap:1/500, Gad67:1/1000, KI67:1/5000, CD133:1/500. The tissue is incubated with primary antibodies in a humid chamber at 4°C overnight. The next day, tissue is put through a 5 minute PBS wash, 3 times. For Secondary antibody incubation, antibodies are added at the following dilution to blocking solution- Donkey anti-mouse:1/1000, Donkey anti-rabbit: 1/1000. The tissue is incubated with Secondary antibodies in a humid chamber at RT, for 1H, followed by a 5 minute PBS wash, 3 times. DAPI Prolong™ Diamond Antifade Mounting solution is added and coverslips mounted onto the glass slides.

MYC plasmid amplification (Mini and Midi prep)

On Day 1, 30µl of XL10-Gold ultracompetent *E. coli* cells are incubated with ~1µg of plasmid for 20 minutes on ice. The cells are heat shocked at 42°C, for 45 seconds, and placed back immediately on ice for 5 minutes. 100µl of SOC (Super Optimal broth with Catabolite repression) medium is added to the E. coli + plasmid mixture and placed in a shaker at 37°C for 1H. This is spread onto an LB agar plate containing an antibiotic (ampicillin) and incubated at 37°C overnight. On Day 2, an isolated colony is picked and inoculated in 10 or 100ml of LB broth + antibiotic (ampicillin 1µg/ml) in a 100 or 500ml conical flask respectively and incubated overnight at 37°C, on a shaker at 225 rpm.

On Day 3, 2ml of media is harvested and centrifuged at 8000 rpm for 2 minutes. The pellet is resuspended in 80% glycerol and can be stored long term at -80°C. This vial can be used directly to create overnight amplified cultures without retransformation.

The rest of the culture is harvested and centrifuged at 6000 rpm for 15 minutes at 4°C. The plasmid is collected following the Qiagen Plasmid Mini or Midi kit protocol and quantified using nanodrop.

Myc plasmid Electroporation

Electroporation protocol is followed according to Lonza Mouse Embryonic Stem Cell Nucleofector™ Kit protocol. Electroporation was performed on EBs that are at Day 6 of Mouse brain organoid generation version 2 protocol before they have been embedded. The EBs should be in NI medium. 5µg of plasmid (Myc or Myc control) are used per electroporation experiment, per cuvette, for 8-16 EBs each time. The mixture of all plasmid + nucleofector solutions are prepared individually per electroporation batch. First, the EBs are washed in 0.5-1ml of DMEM F12 medium to remove trace amounts of serum from medium that can interfere with electroporation. Remove all DMEM and resuspend the EBs in nucleofector solution according to protocol. Each batch of electroporation is prepared fresh to avoid exposing the EBs to nucleofector solution for a long time to avoid toxicity. The plasmid+ nucleofector solution is now mixed with the EBs and transferred to a sterile cuvette without introducing bubbles and the cuvette Is capped. The EBs are then electroporated in the Nucleofector 2b Device using the A023/A013 program. The EBs are recovered from the cuvette using the pipette provided in the kit and introduced back into fresh NI medium in ultra low attachment plates. The EBs are transferred back to 37°C overnight and embedded the next day.

Trp53 plasmid Lipofectamine transfection

Prior to transfection assay, the cells are tested to determine MIC (Minimum inhibitory concentration) of the relevant antibiotic (puromycin in case of P53KO plasmid) that kills all cells in culture within 3 days of exposure.

Fresh cells to be transfected are seeded in Poly-D-Lysine coated P12 TC-treated plates. The cells are seeded at density of 100K cells /well in MES medium. Cells show around 50% confluency the next day which is optimal. The ratio of plasmid to Lipofectamine solution has been tested to be most effective at 1:4. A mixture of plasmid + Optimem is prepared, and a separate solution of Lipofectamine 2000 + Optimem is prepared as mentioned in the table below.

Per well (row) added,

Ratio	Amount plasmid	TP53 KO /Control plasmid	HDR	OPTIMEM		Lipo2000	OPTIMEM	DMEM
1:4	1µg	0.5µg (5µl)	0.5µg (5µl)	15µl		4µl	21µl	1ml

Total volume of the transfection solution is 50µl. DMEM F12 used to avoid using serum as a base solution to avoid serum that can affect transfection efficiency. The cells are left in the transfection solution overnight, and the medium is refreshed to MES medium + antibiotic (puromycin (WT Glio: 3µg/ml, G34R: 7µg/ml)). The transfection efficiency can be visualized with the help of the GFP and RFP tags present on the gene plasmid and HDR plasmid respectively. The cells are kept in the same antibiotic solution for 3 days before being tested for efficient Knockout using qpcr analysis. Successfully transfected cultures are passaged onto MEF feeder cells and eventually as EBs for further experimentation. It is necessary to avoid keeping these cells in culture for too long to avoid buildup of any additional mutation due to the presence of the P53KO signature.

Myc plasmid Lipofectamine transfection

For transfection with Myc plasmid or its control, there was no detectable antibiotic found in order to use for selection; therefore, the cells were transfected twice in 2 consecutive days.

One day prior to transfection, cells were seeded at 100,000 cells/well concentration in 24-well TC treated plates in MES medium. The next day, cells showed 50% confluency. The ratio of plasmid to Lipofectamine solution has been tested to be most effective at 1:4. A mixture of plasmid + Optimem is prepared, and a separate solution of Lipofectamine 2000 + Optimem is prepared as mentioned in the table below.

Per well, reagents added were,

Ratio	Amount plasmid	Myc/Myc control	OPTIMEM		Lipo2000	OPTIMEM	DMEM
1:4	1µg	1µg (0.8µl)	14.2µl		3.2µl	21.8µl	1ml

The cells were kept in transfection solution + DMEM supplemented with 4.5 g/l glucose and GlutaMAX-I overnight. The next day, the cells were transfected again at a ratio of 1:4.

The cells were kept in transfection solution + DMEM supplemented with 4.5 g/l glucose and GlutaMAX-I overnight. The next day they were refreshed with MES medium. The cells were visualized for the GFP tag confirming successful transfection and collected and verified with RTqPCR for Myc expression. Successfully transfected cultures are passaged onto MEF feeder cells and eventually as EBs for further experimentation.

Library preparation for sequencing

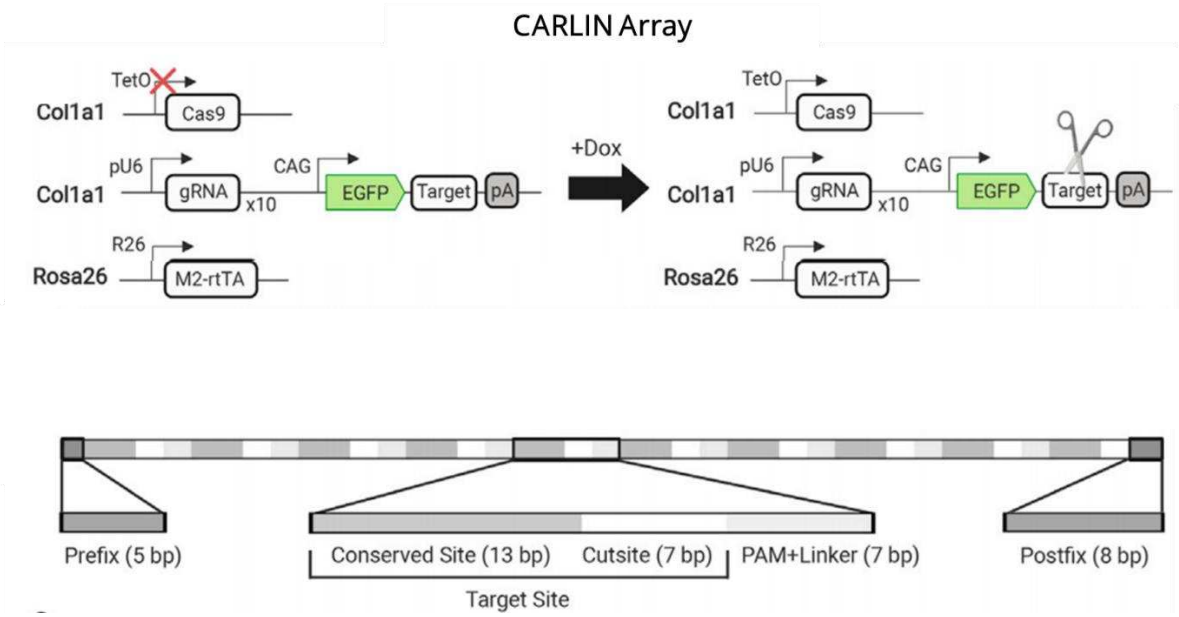
1ng of DNA was combined with 10µM of Forward and reverse primer (H3F3A_Seq_F, H3F3A_Seq_R), 12.5µl of Q5 Hot Start High Fidelity master mix, and completed to 25µl of reaction with ddH2O. PCR amplification program was as follows: 98°C for 30sec; 98°C for 10s, 65°C for 30sec, 72°C for 30sec (25 cycles); 72°C for 2min, 4°C for infinity. This was followed by a 0.8X cleanup with SPRI select beads and verified on Tapestation.

Single cell RNA sequencing and processing

The Chromium instrument and Single cell 3' Reagent Kit v3 was used to prepare individually barcoded single-cell RNA-Seq libraries following the manufacturer's protocol (10X Genomics) and performed by me. The target number of cells was 8000 per condition, where whole transcriptome libraries were prepared following the 10X V3.1 protocol according to the manufacturer's guidelines. Whole transcriptome libraries were sequenced on Illumina NextSeq500 using paired-end 150 cycles v3 kits (Read 1: 28 cycles; Index Read 1 (i7): 10 cycles; Read 2: 90 cycles).

The following data processing and generation of UMAP profiles was performed by Dr. Ariel Galindo (Post Doctoral fellow, Sysfate team). Single cell transcriptome data has been processed using Cell Ranger 7.0.1 for primary analyses, data alignment to the mouse reference genome mm10 2020A, read counts were normalized by SCTransform method, filtering cells on the grounds of doublets, more than 5% of mitochondrial content, novelty score higher than 0.8, and cells with less than 200 genes. These cells were processed downstream for dimensionality reduction analysis (UMAP) using Seurat (Version 4). 17 clusters were identified (KNN method) based on gene expression similarities (customized scripts in R, r-project).

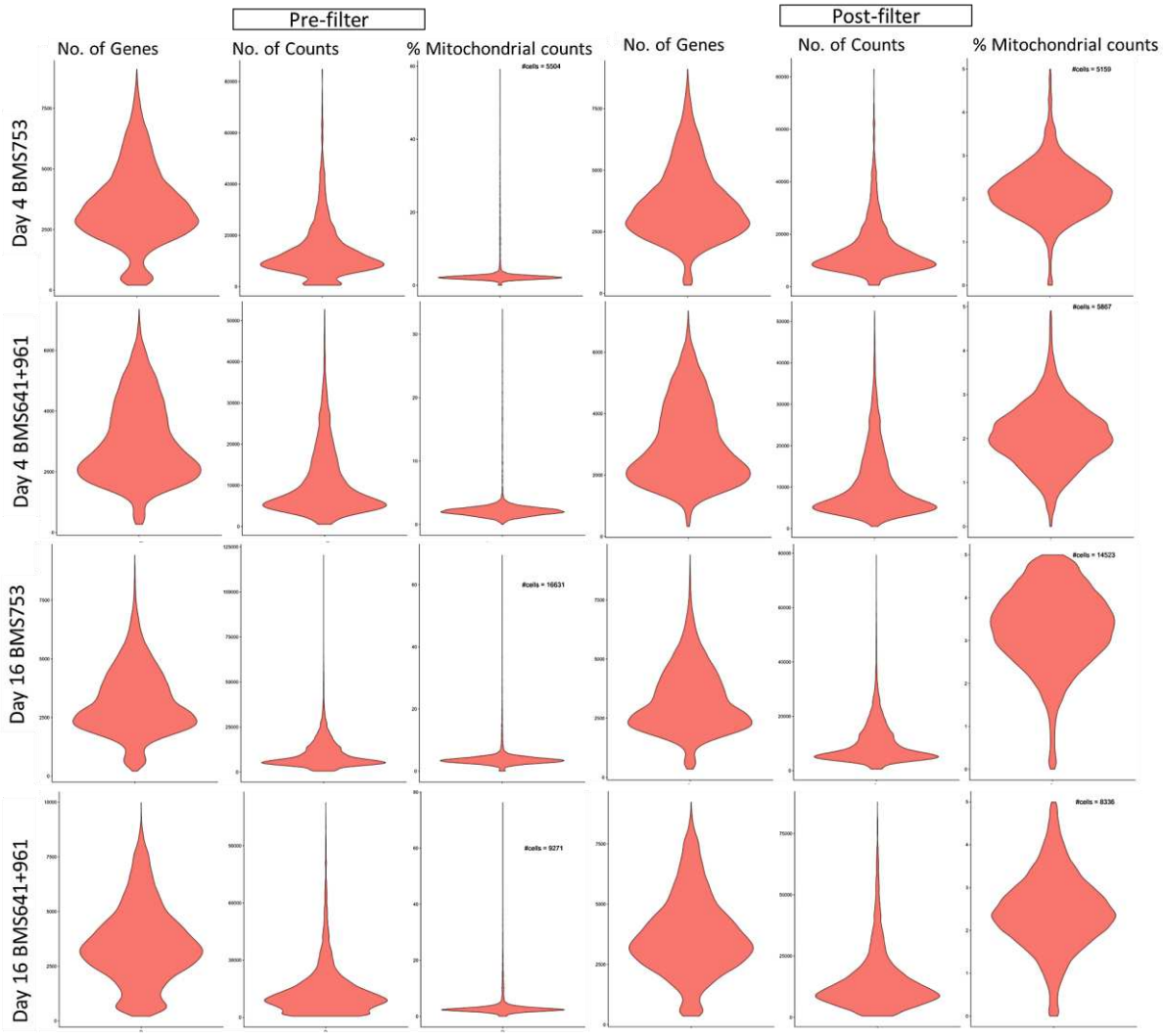
Supplementary Figures

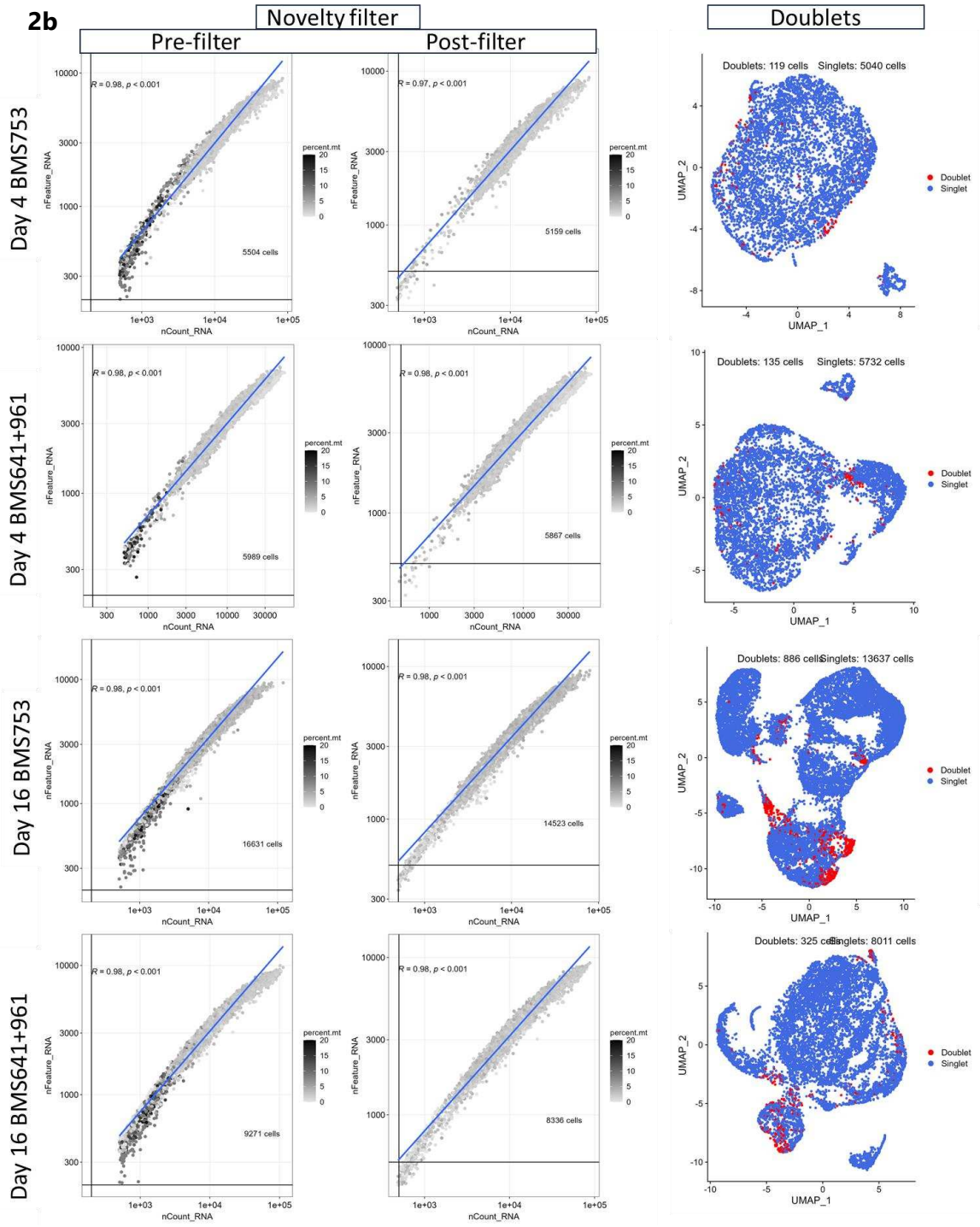


Supplementary Figure 1. CARLIN array

Adapted from Bowling S et al, 2020 [146]. a) Schematic of CARLIN system. Guides RNAs, target sites and inducible Cas9 components are contained within the Col1a1locus. The expression of each of the 10 gRNAs is driven by a separate U6 promoter (pU6). The CARLIN array sits in the 3'UTR of GFP and consists of 10 sites matching the gRNAs. The doxycycline (Dox) reverse tetracycline-controlled transactivator (rtTA) is contained within the Rosa26locus. b) In the CARLIN array, each target site is divided into a 13bp conserved site and 7bp cutsite. Consecutive target sites are interleaved by a 3bp protospacer adjacent motif (PAM) and 4bp linker sequence. There is a 5bp prefix motif upstream of the first target site and an 8bp postfix motif downstream of the last target site.

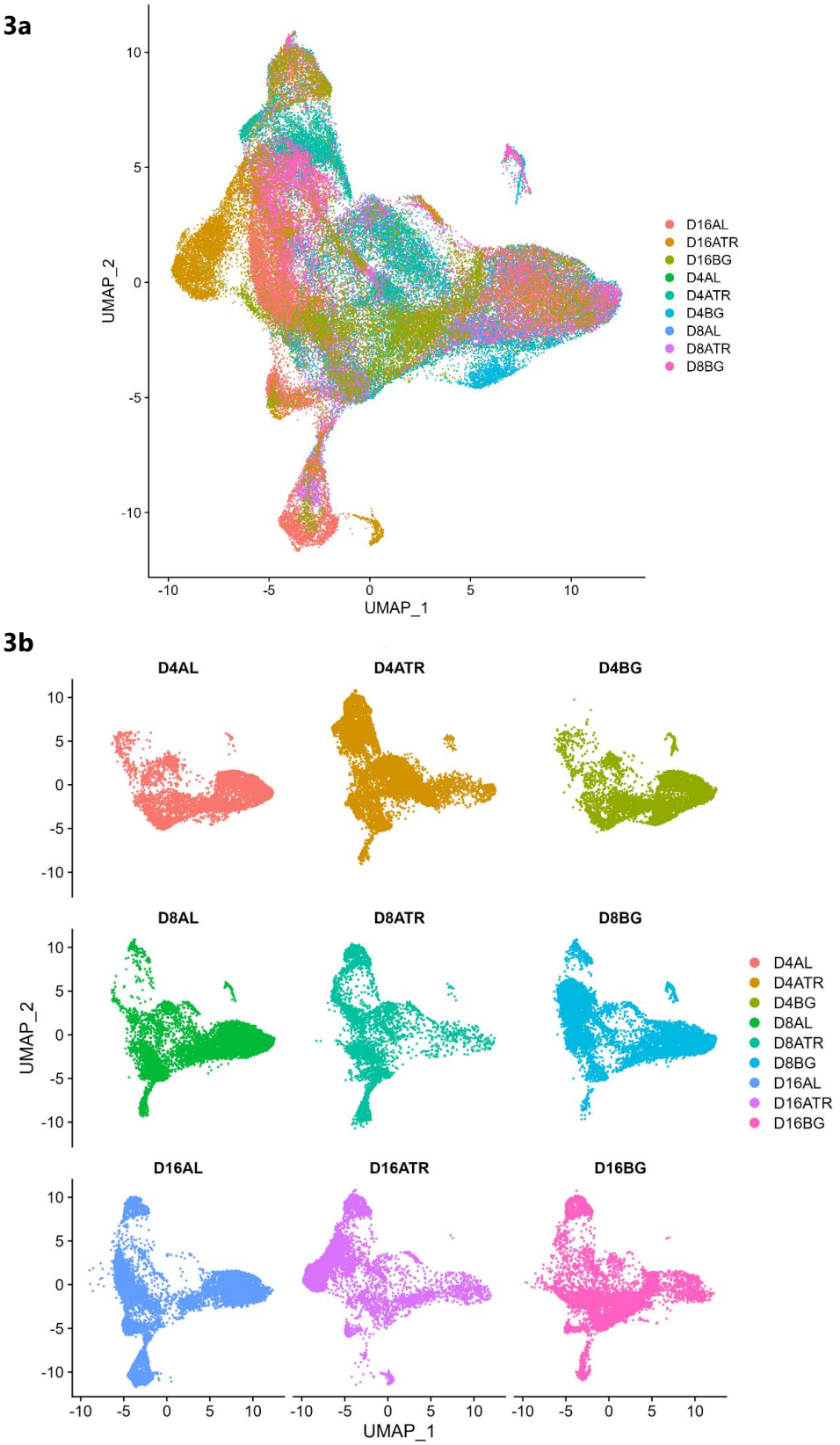
2a


Pre-filter
Post-filter
No. of Genes
No. of Counts
% Mitochondrial counts
No. of Genes
No. of Counts
% Mitochondrial counts
Day 4 BMS753
Day 4 BMS641+961
Day 16 BMS753
Day 16 BMS641+961
#cells = 5504
#cells = 5159
#cells = 5867
#cells = 16631
#cells = 14523
#cells = 9271
#cells = 8336



Supplementary Figure 2. Preprocessing filters used to select for good quality cells

with a) more than 200 genes, less than 5% mitochondrial content and b) Novelty Ratio greater than 0.8.



Supplementary Figure 3. UMAP projection

of all the 9 samples treated and collected at different timepoints a) compiled and b) separated individually. (D4, D8, D16: Day 4, Day 8, Day 16; ATR, AL, BG: ATRA, Alpha, Beta + Gamma).

Marker	Gene
Endoderm markers	GATA6, GATA4, EOMES, SOX7, SOX17, FOXA2
Mesoderm marker	MESP1, WNT3A, BMP4, FOXF1
Ectoderm marker	PAX6, OTX2, SOX1
Stem cell marker	NANOG, SOX2, KLF4
Neural Progenitor cells	PAX6, NES, NEUROD1, VIM, ASCL1, NEUROG2
Neurons	TUBB3, MAP2, DCX, TH, TPH2, ENO2
Astrocytes	S100B, AQP4, GFAP
Oligodendrocyte precursor cells	MBP, MOG, SOX10, OLIG1, OLIG2
Gabaergic neurons	GAD1, GAD2
Glutamatergic neurons	GRIN1, GRIN2B
Dopaminergic neurons	TH
Serotonergic neurons	TPH2, SERTM1

Supplementary Figure 4. List of genes for each cell type marker

used to interrogate the gene expression in each dataset of the UMAP projection of the clusters in Loupe Browser.

5a

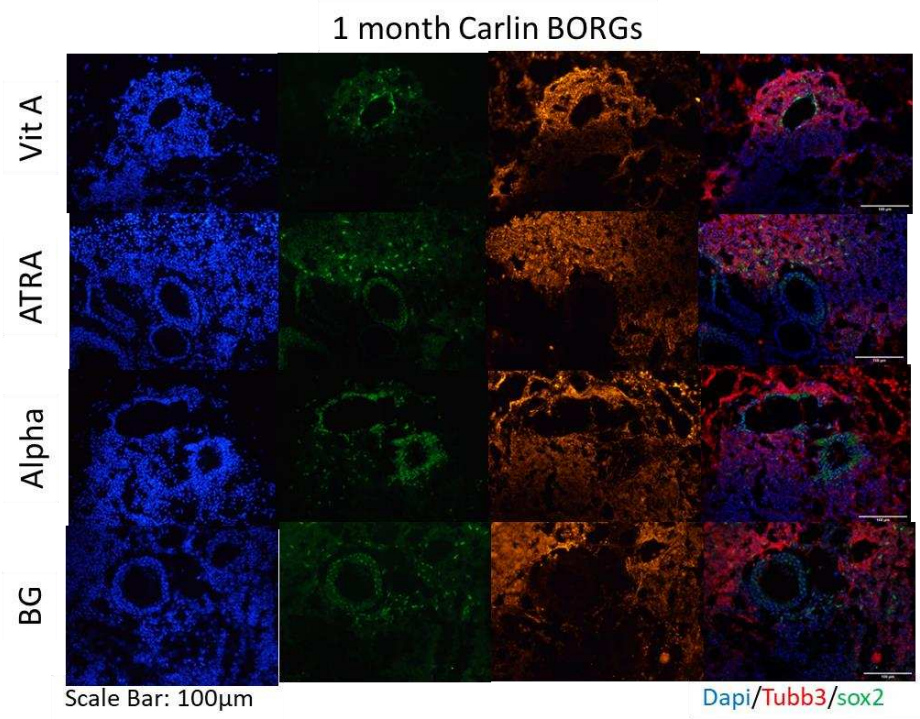
Cluster 3	Average Log2 (FC)
Hmgb2	0.70
Trh	0.75
Dab1	0.72
Ptprt	0.73
Mif	0.52
Ldha	0.52
Ptma	0.47
Pkm	0.46
H2afz	0.43
Pclaf	0.56
Eno1	0.44
Snrpf	0.42
Nasp	0.41
Ran	0.36
Npm1	0.35
Rpl36	0.35
Ddit4	0.49
Rpl36a	0.33
Pim2	0.48
Dut	0.35

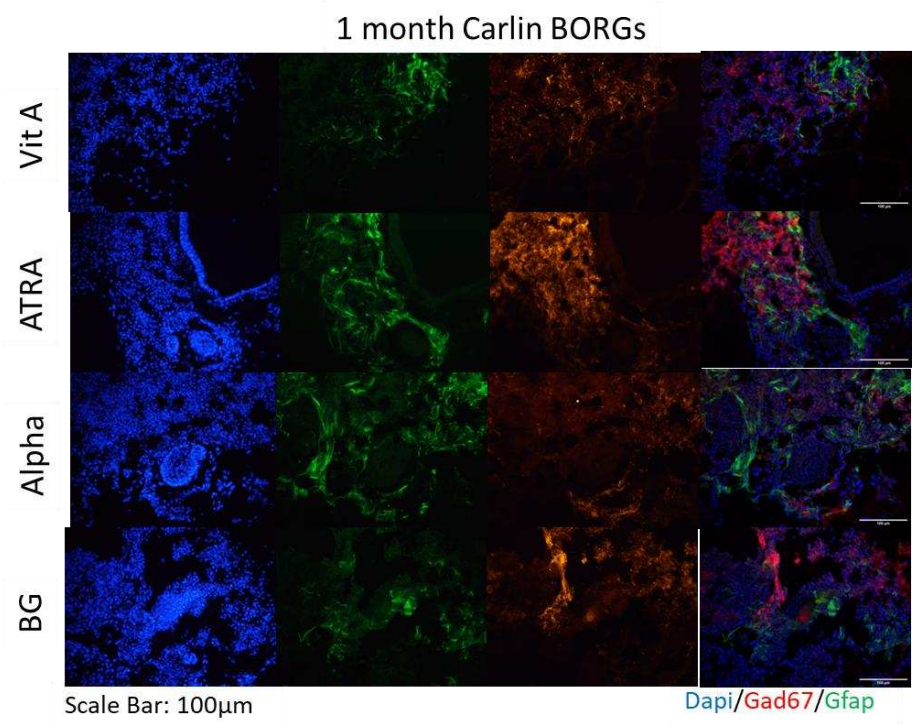
5b

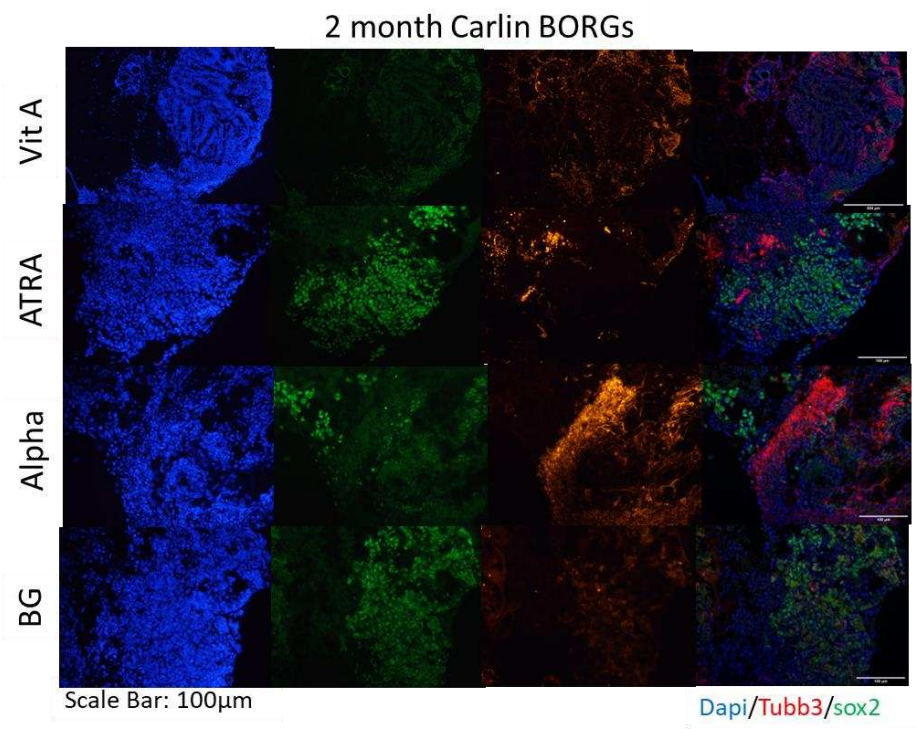
Cluster 11	Average Log2 (FC)
Crabp1	2.41
Crabp2	1.98
Adgrv1	1.64
Hist1h1b	1.34
Hist1h2ap	1.22
Pclaf	1.15
Npas3	1.12
Fat3	1.17
Id3	1.04
Ccnd2	0.99
H2afz	0.87
Hes5	1.17
Hmgb1	0.85
Prtg	1.06
Mdk	0.85
Hist1h2ae	0.95
Hmgb3	0.91
Hist1h1e	0.95
Hoxb5os	1.00
Dach1	0.95

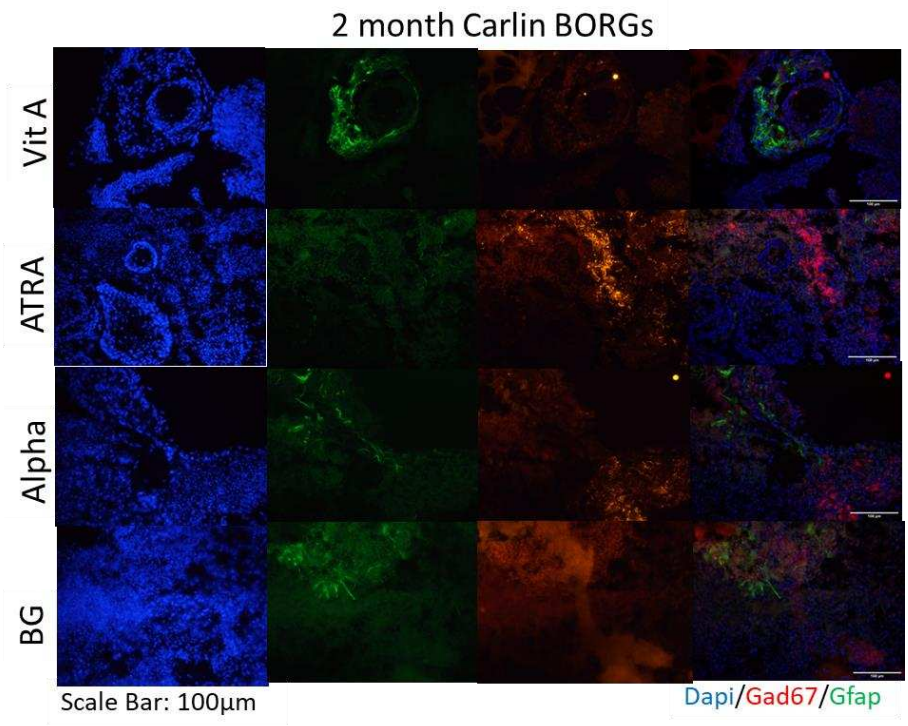
Supplementary Figure 5. Top 20 ranked genes

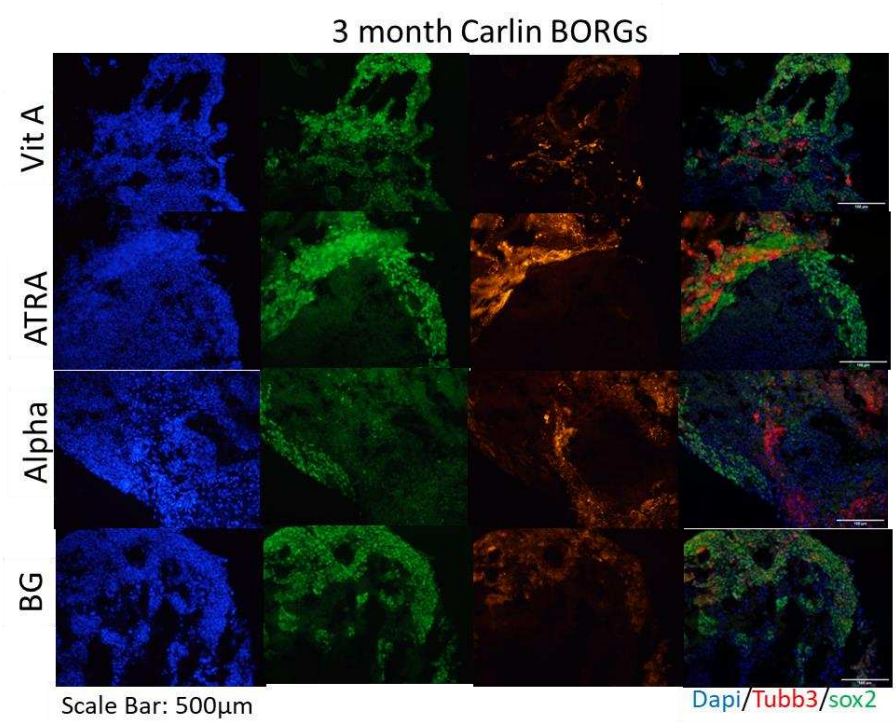
in a) cluster 3, and b) cluster 11; The genes were ranked according to Log2 Fold change levels in descending order. Adjusted P value<0.001.

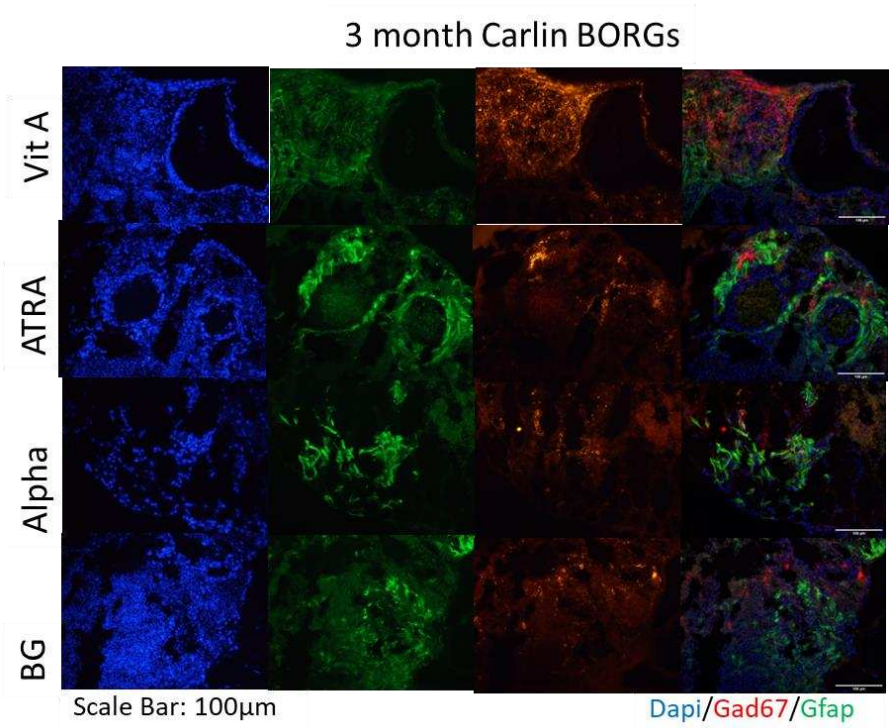

1 month Carlin BORGs
Vit A
ATRA
Alpha
BG
Scale Bar: 100µm
Dapi/Tubb3/sox2


1 month Carlin BORGs
Vit A
ATRA
Alpha
BG
Scale Bar: 100µm
Dapi/Gad67/Gfap


2 month Carlin BORGs
Vit A
ATRA
Alpha
BG
Scale Bar: 100µm
Dapi/Tubb3/sox2


2 month Carlin BORGs
Vit A
ATRA
Alpha
BG
Scale Bar: 100µm
Dapi/Gad67/Gfap





Supplementary Figure 6. 3D Brain organoids derived from MESC Carlin cells

treated with Vitamin A (Vit A), ATRA, BMS753 (RARα (alpha) agonist), BMS641+961 (RARβ + RARγ (beta+gamma; bg) agonist)**.** Immunoflueorescnece staining of Carlin BORGs with Sox2 (Neural stem cells), Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (Gabaergic neurons) at a) 1 month, b) 2 months c) 3 months of growth.

	WT Glio		K27M		G34R	
Days	*Mean Diameter (µm)*	*SD*	*Mean Diameter (µm)*	*SD*	*Mean Diameter (µm)*	*SD*
D4	188.747	4.17	263.416	1.678	318.334	5.111
D10	420.515	31.13	331.561	6.013	202.91	0.48
D30	921.713	36.109	713.348	32.341	1620.662	181.504
D45	1305.1	268.927	881.128	206.812	1808.052	368.258
D60	1856.056	732.529	1074.002	429.299	1193.387	640.777
D75	1522.946	466.647	1078.203	282.41	2028.997	745.796
D90	1608.267	829.334	803.863	257.857	1194.372	488.543

	WT Glio	K27M	G34R
Days	*Number of BORGs measured*		
D4	1	1	1
D10	1	1	1
D30	1	1	1
D45	13	24	15
D60	12	18	19
D75	13	16	11
D90	9	35	50

Supplementary Figure 7. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 1) size measurements

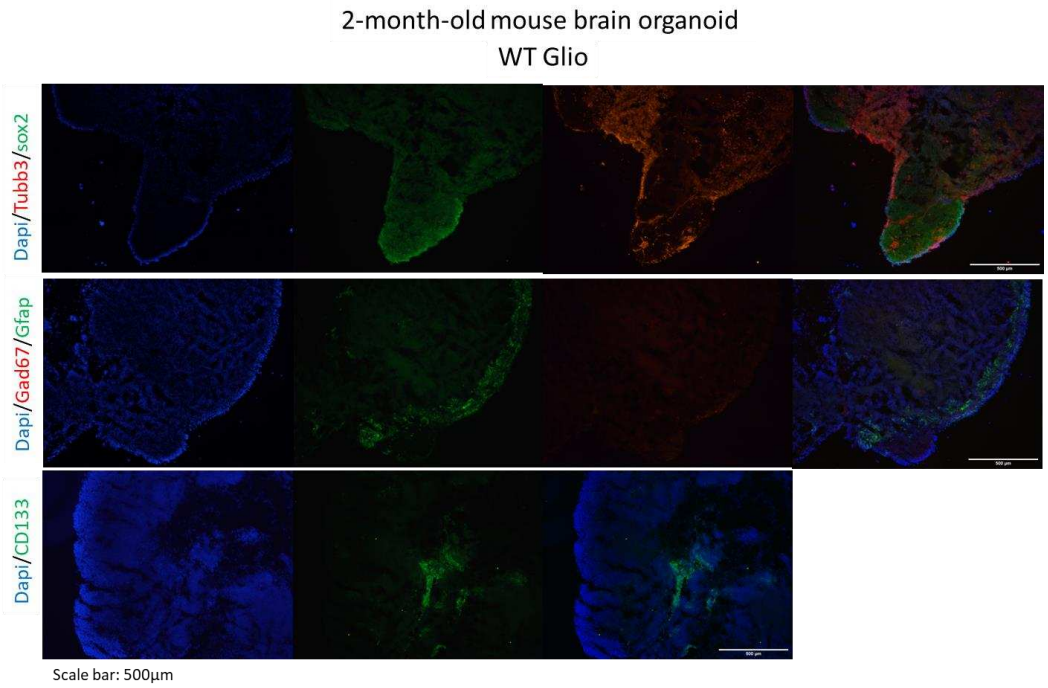
Diameter measurements of the brain organoids measured in a table format along with a summary table of the number of organoids measured.

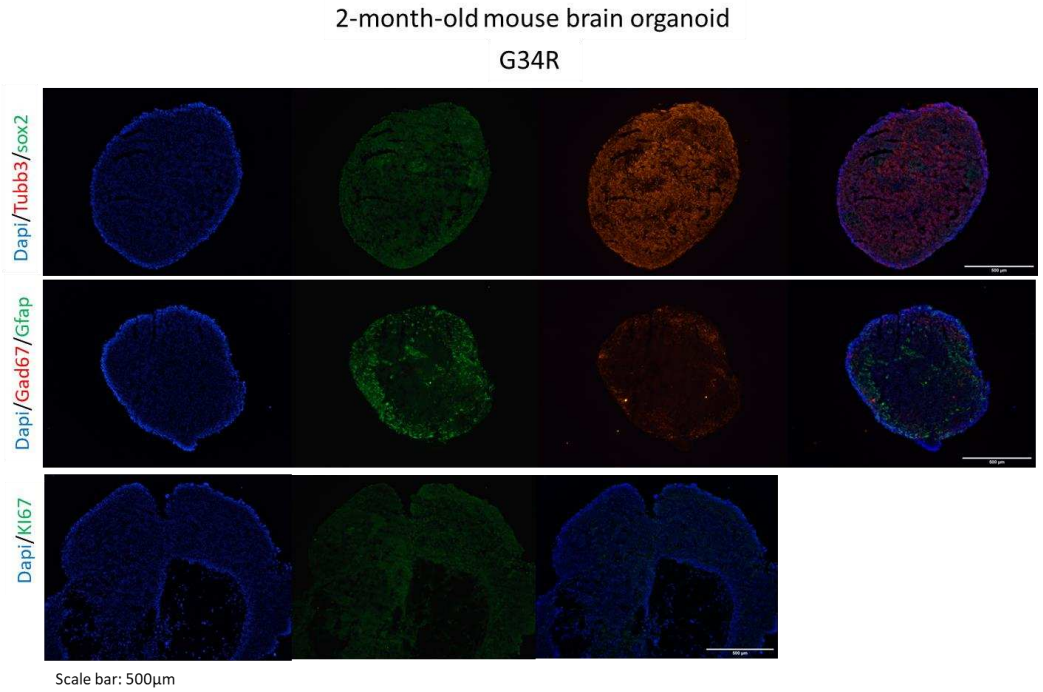
	WT Glio		G34R	
Days	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD
D10	568.231	139.906	715.883	137.262
D15	1705.513	149.038	1728.134	264.84
D30	1868.584	352.138	1656.316	438.722
D45	1799.395	240.606	2053.985	355.016
D60	2135.148	568.616	2130.527	596.527
D75	2325.469	657.083	1806.332	480.883
D90	2171.029	898.667	1759.673	428.361

	Number of BORGs measured	
Days	WT Glio	G34R
D10	9	15
D15	5	5
D30	38	35
D45	5	5
D60	35	49
D75	23	35
D90	27	41

Supplementary Figure 8. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 2) size measurements

Diameter measurements of the brain organoids measured in a table format along with a summary table of the number of organoids measured.





Supplementary Figure 9. H3.3 wildtype and mutant Brain Organoids (Mouse brain organoid generation version 2) immunohistochemistry assay

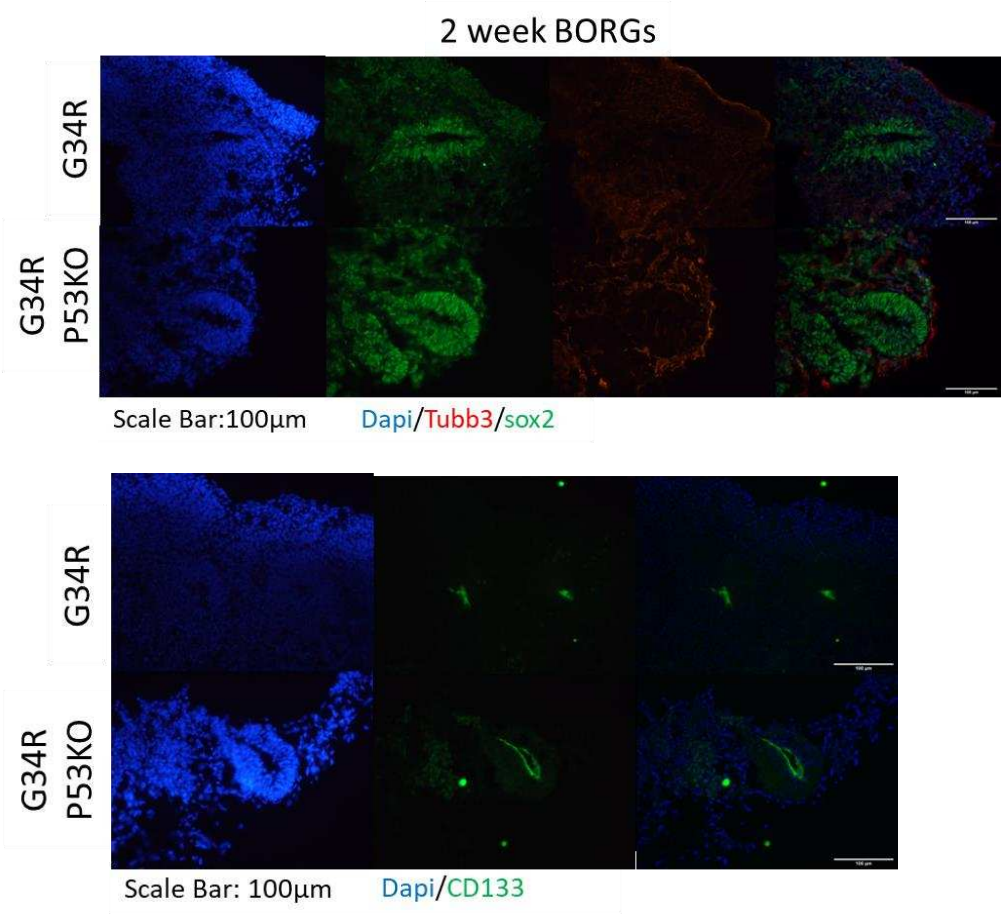
Mouse Brain organoid generation with H3.3 mutant cell lines using a modified protocol. Immunofluorescence staining of 2 month old brain organoids for both cell lines, with markers Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for a) WT Glio and b) G34R BORGs.

Days	G34R		G34R-P53 KO		G34R-P53 KO control	
	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD
D2	269.245	21.484	306.308	12.604	318.25	18.452
D4	770.885	84.927	756.472	54.226	910.73	53.216
D6	968.731	52.975	872.972	118.62	1153.403	97.16
D10	684.441	123.105	749.232	59.924	822.437	123.94
D15	1188.294	246.517	1267.356	126.423	1278.418	190.228
D30	1644	251	1687.598	190.183	1496.298	315.234
D45	1506.897	310.992	1825.073	292.851	1460.241	393.723
D60	1488.827	408.981	1699.909	422.67	1334.439	335.258
D90	1491.867	403.334	1996.999	186.138	1576.851	402.828

Days	Number of BORGs measured		
	G34R	G34R-P53 KO	G34R-P53 KO control
D2	38	38	38
D4	40	40	40
D6	38	38	40
D10	58	22	50
D15	55	35	56
D30	42	25	44
D45	46	20	40
D60	29	15	27
D90	68	14	54

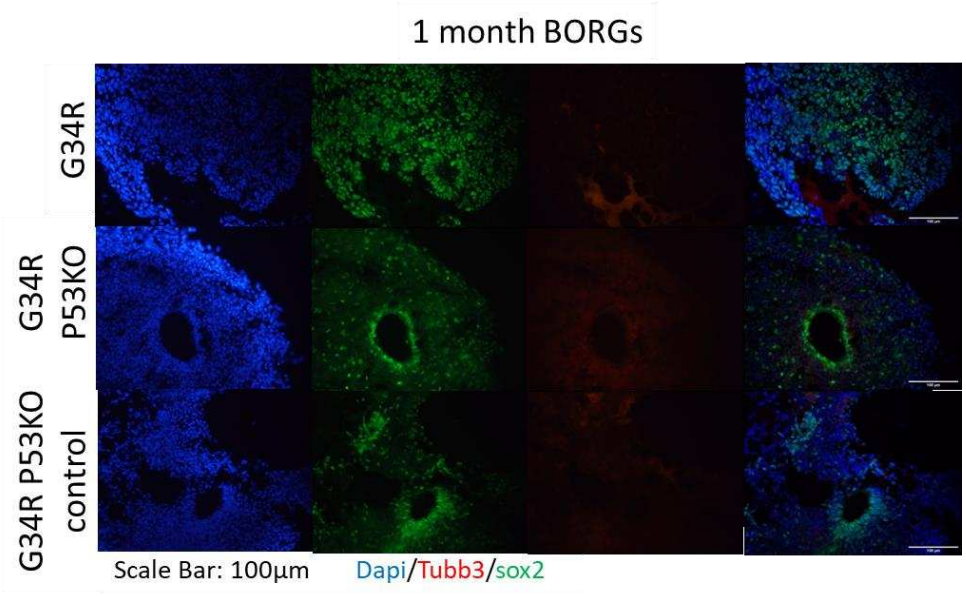
Supplementary Figure 10. G34R P53KO brain organoid size measurements

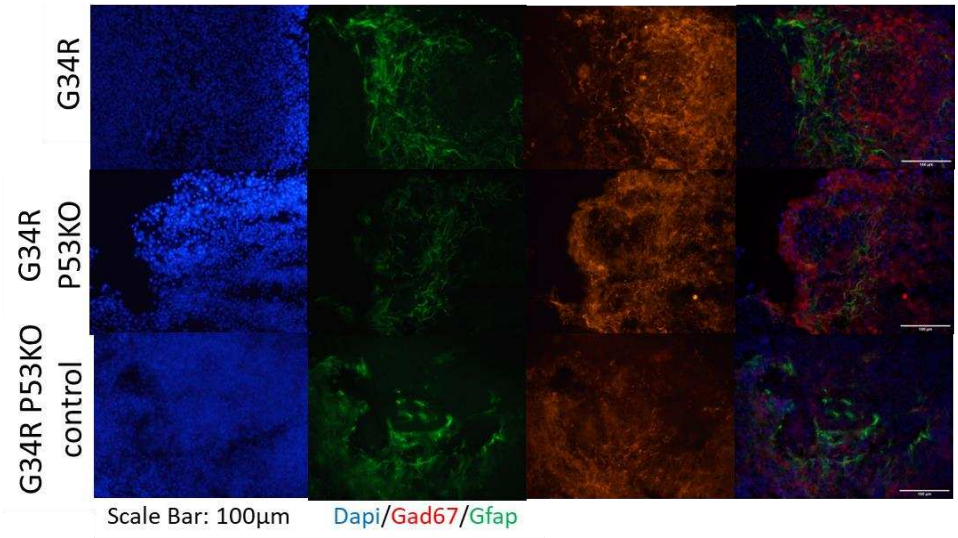
Diameter measurements of the brain organoids measured in a table format along with a summary of the number of organoids measured.

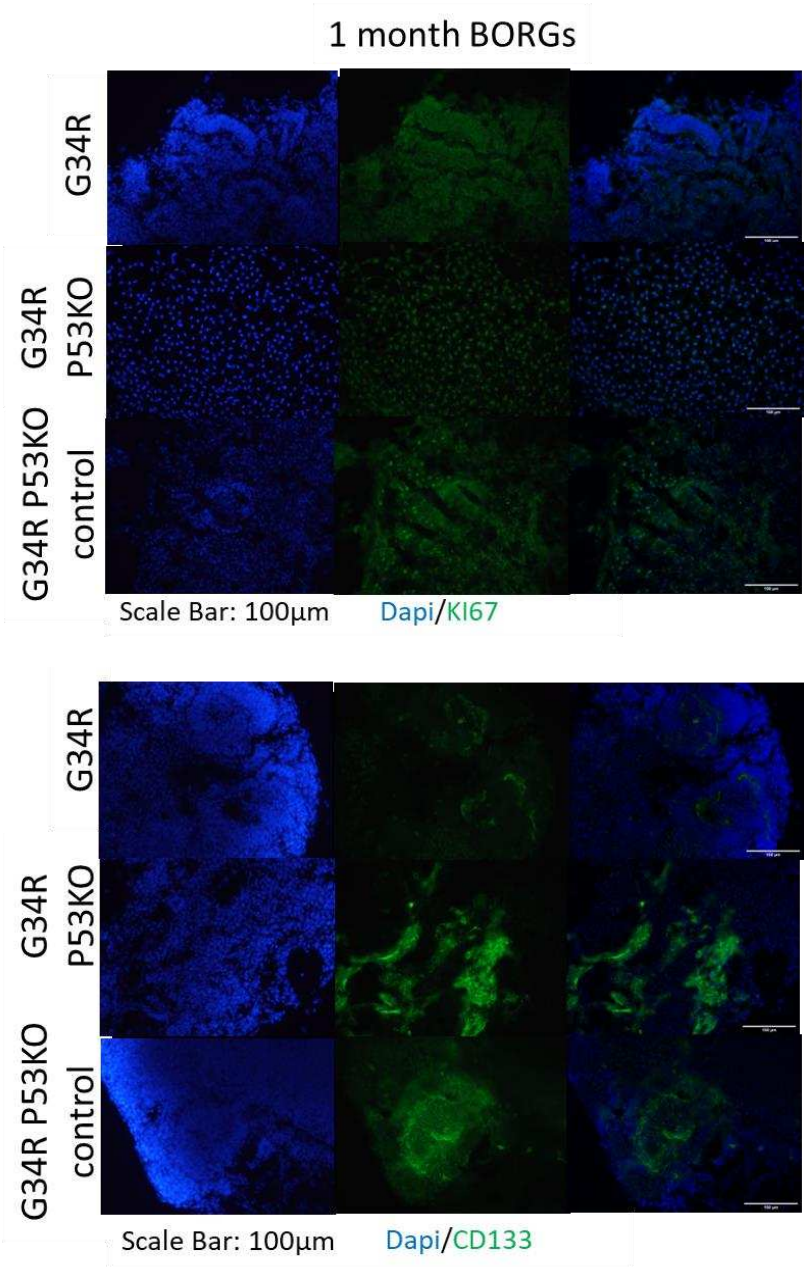


Supplementary Figure 11. G34R P53KO brain organoid immunohistochemistry assay of 2 week BORGs

Immunofluorescence staining with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (postmitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for G34R, G34R P53KO, G34R P53KO control BORGs

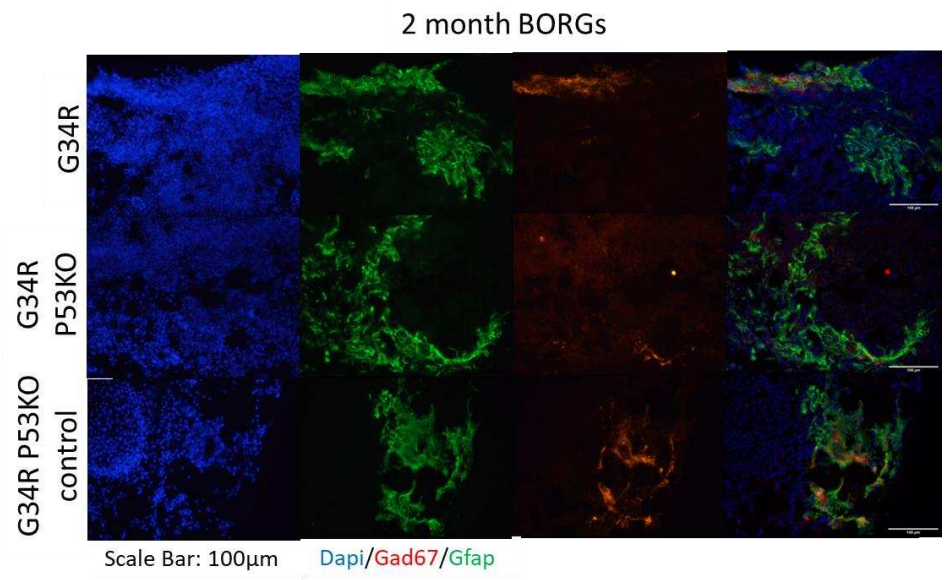

1 month BORGs
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 100µm
Dapi/Tubb3/sox2

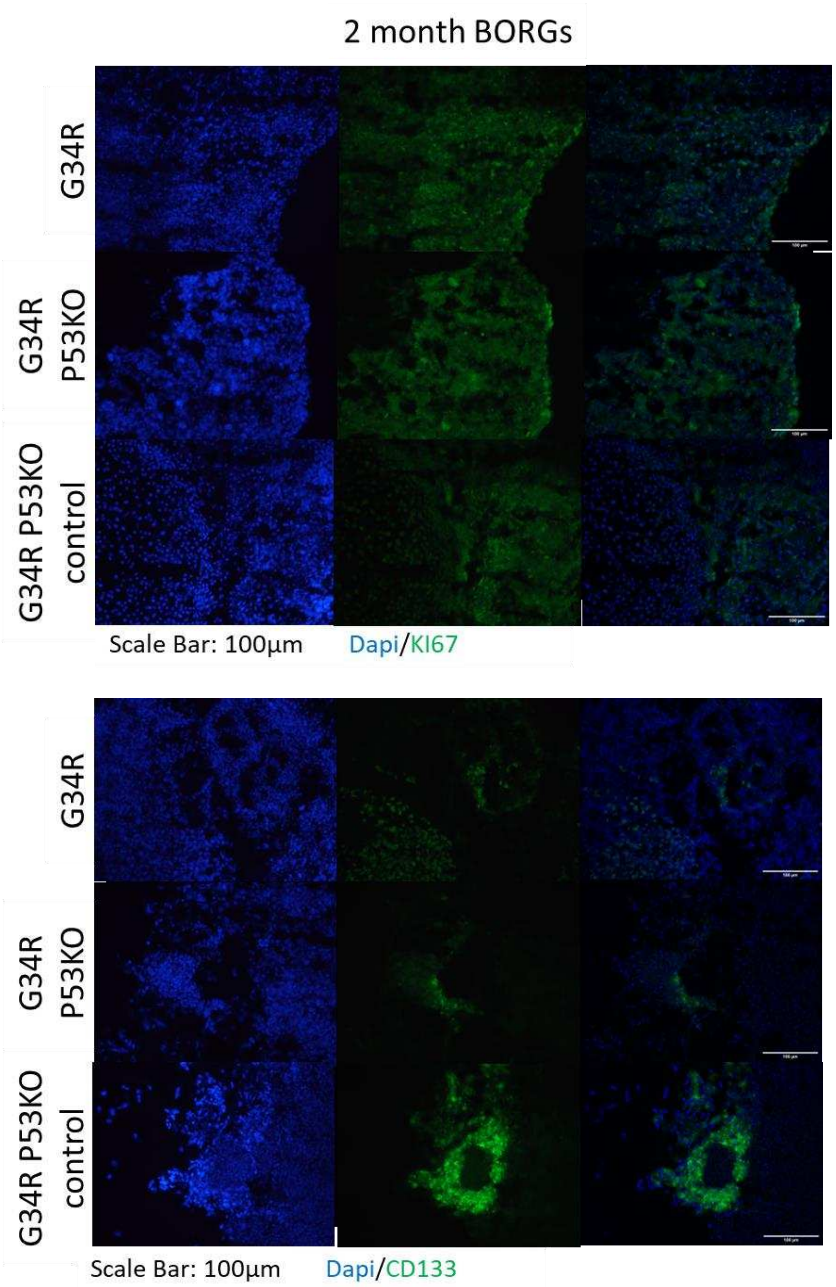

G34R
G34R P53KO
G34R P53KO control
Scale Bar: 100µm
Dapi/Gad67/Gfap



Supplementary Figure 12. G34R P53KO brain organoid immunohistochemistry assay of 1 month BORGs

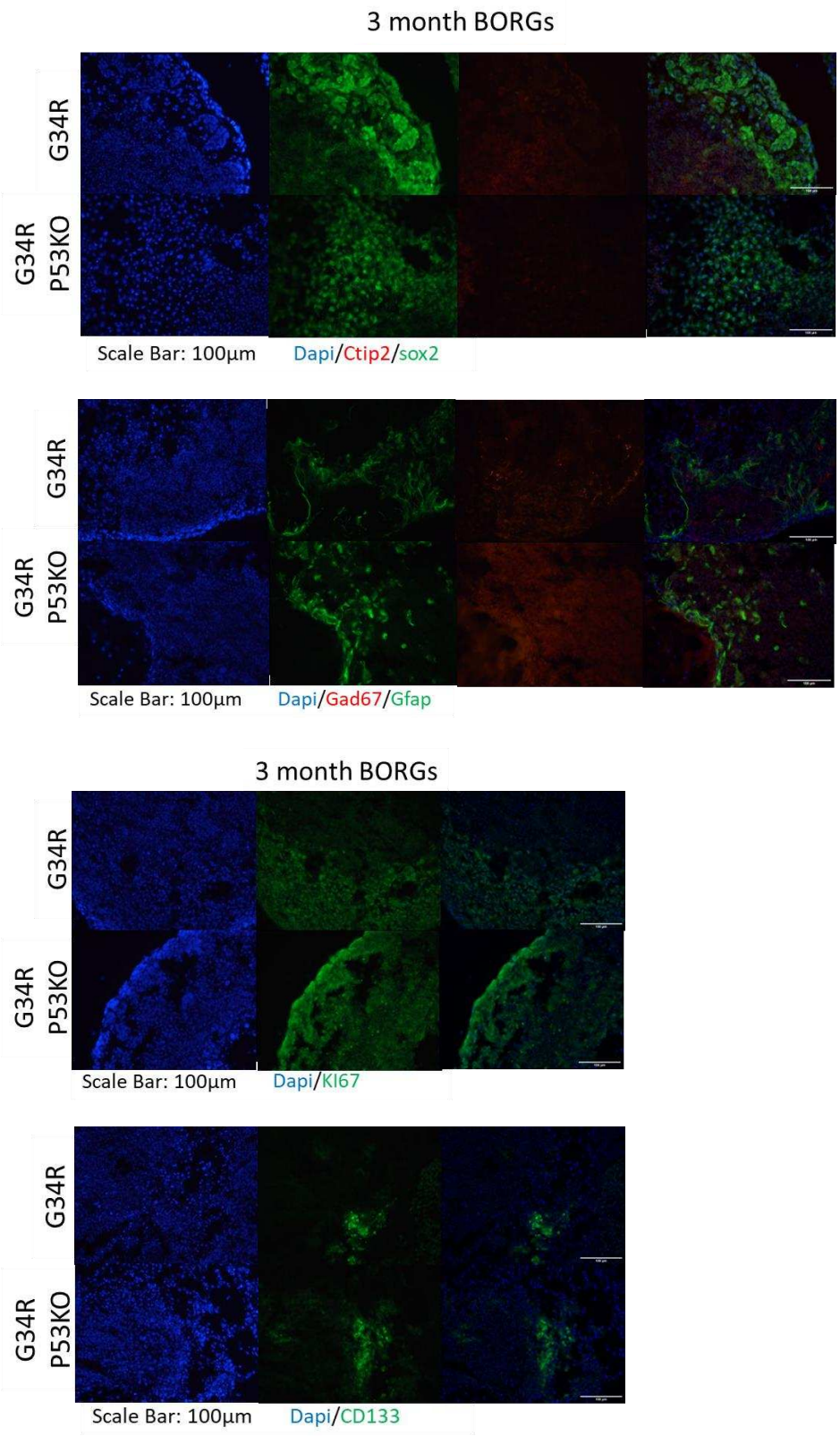
Immunofluorescence staining with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (postmitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for G34R, G34R P53KO, G34R P53KO control BORGs


2 month BORGs
G34R
G34R P53KO
G34R P53KO control
Scale Bar: 100µm
Dapi/Gad67/Gfap



Supplementary Figure 13. G34R P53KO brain organoid immunohistochemistry assay of 2 month BORGs

Immunofluorescence staining with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (postmitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for G34R, G34R P53KO, G34R P53KO control BORGs



Supplementary Figure 14. G34R P53KO brain organoid immunohistochemistry assay of 3 month BORGs

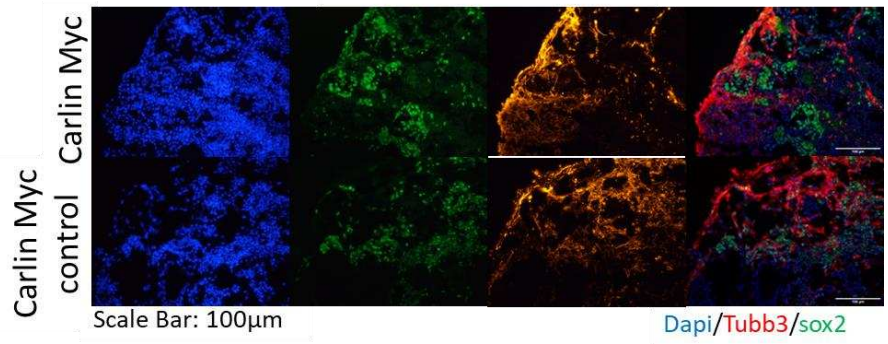
Immunofluorescence staining with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (postmitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for G34R, G34R P53KO, G34R P53KO control BORGs

Days	Wtglio Myc		Wtglio Myc control		G34R Myc		G34R Myc control	
	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD	Mean Diameter (µm)	SD
D2	717.034	136.537	631.207	135.342	689.654	75.441	570.781	59.974
D4	935.165	57.658	936.476	97.987	1063.919	55.713	889.282	88.723
D6	1292.697	78.63	1219.977	49.446	1418.918	73.675	1301.487	53.263
D10	1094.42	149.378	1202.231	145	1108.017	189.753	1010.713	201.777
D15	1488	164	1450	163	1577	244	1527	206
D30	1604.755	293.467	1486.146	223.056	1781.496	270.051	1830.754	240.424
D45	1636.275	496.817	1690.168	353.974	1692.919	410.819	1917.701	448.475
D60	1921.478	642.875	1770.329	597.969	2147.627	436.743	2179.725	459.114
D90	2510.488	526.882	2242.108	578.646	1953.724	507.318	2447.108	502.721

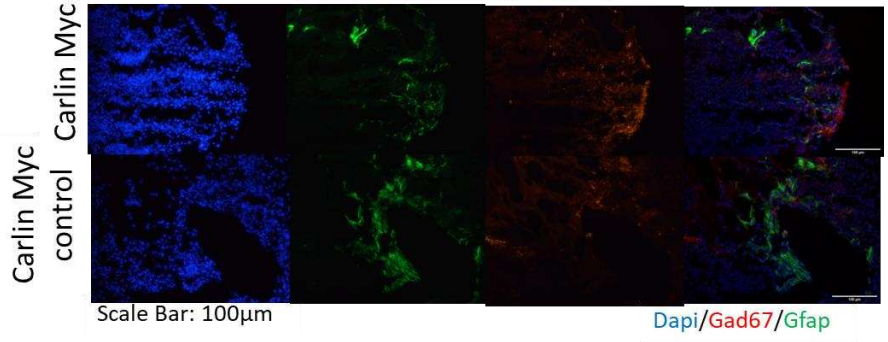
Supplementary Figure 15. Brain organoids derived from Myc overexpression transfected MESC size measurements

Diameter measurements of the brain organoids measured in a table format along with a summary of the number of organoids measured.

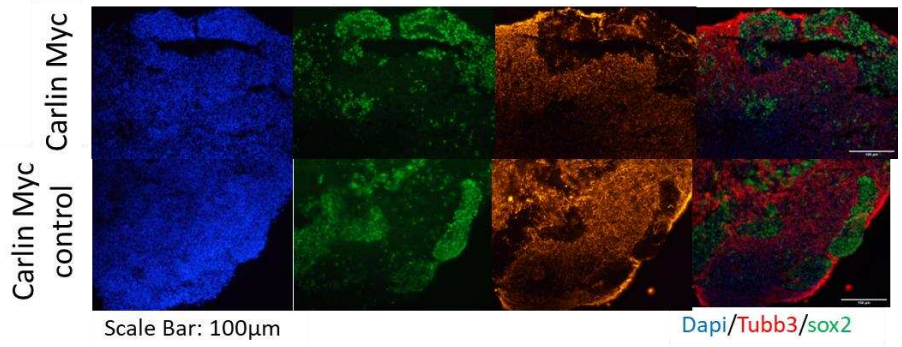
1 month Carlin BORGs



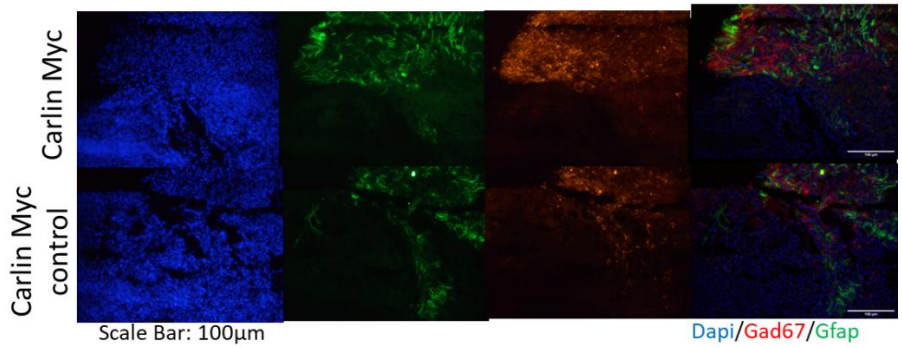
1 month Carlin BORGs

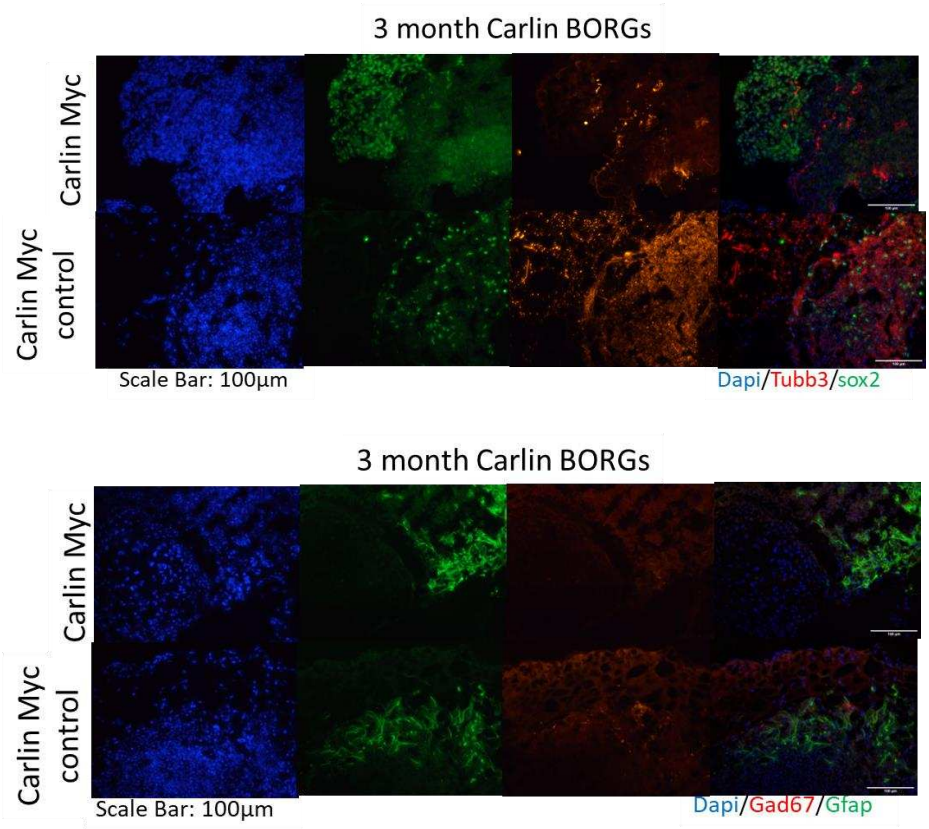


2 month Carlin BORGs



2 month Carlin BORGs

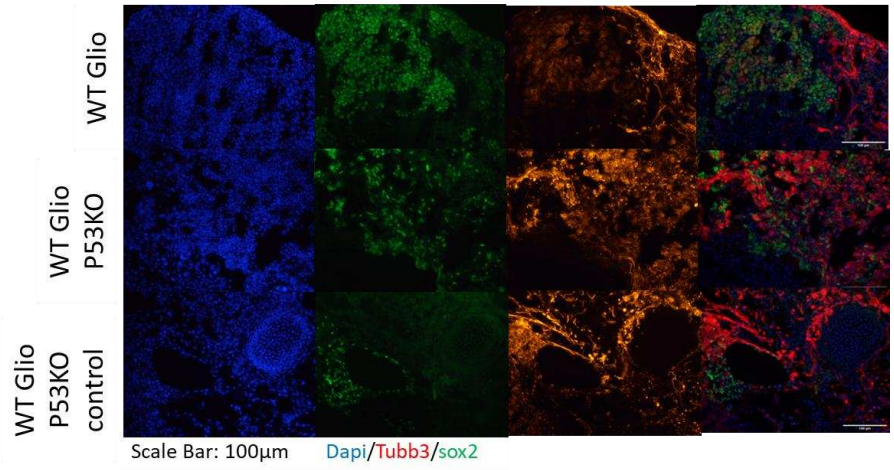




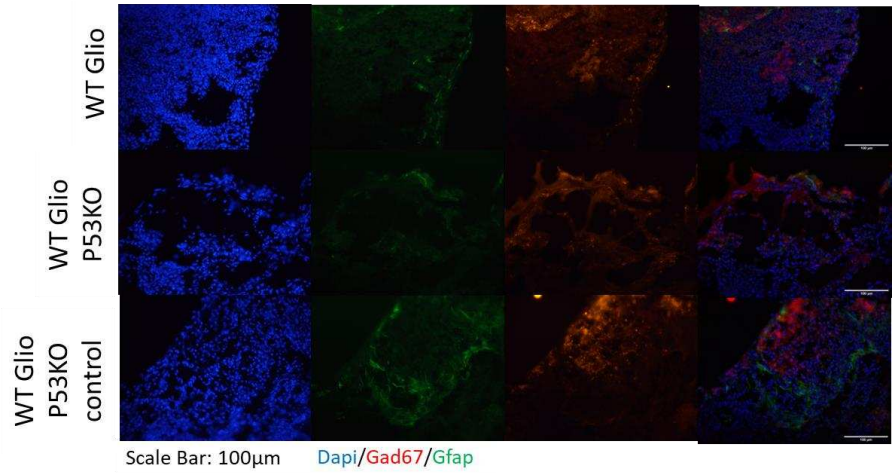
Supplementary Figure 16. Myc electroporated Brain organoids immunohistochemistry assay

Immunofluorescence staining, with markers Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), for Carlin BORGs with Myc and Myc control electroporation at 1 month (ii) 2 months and (iii) 3 months of differentiation.

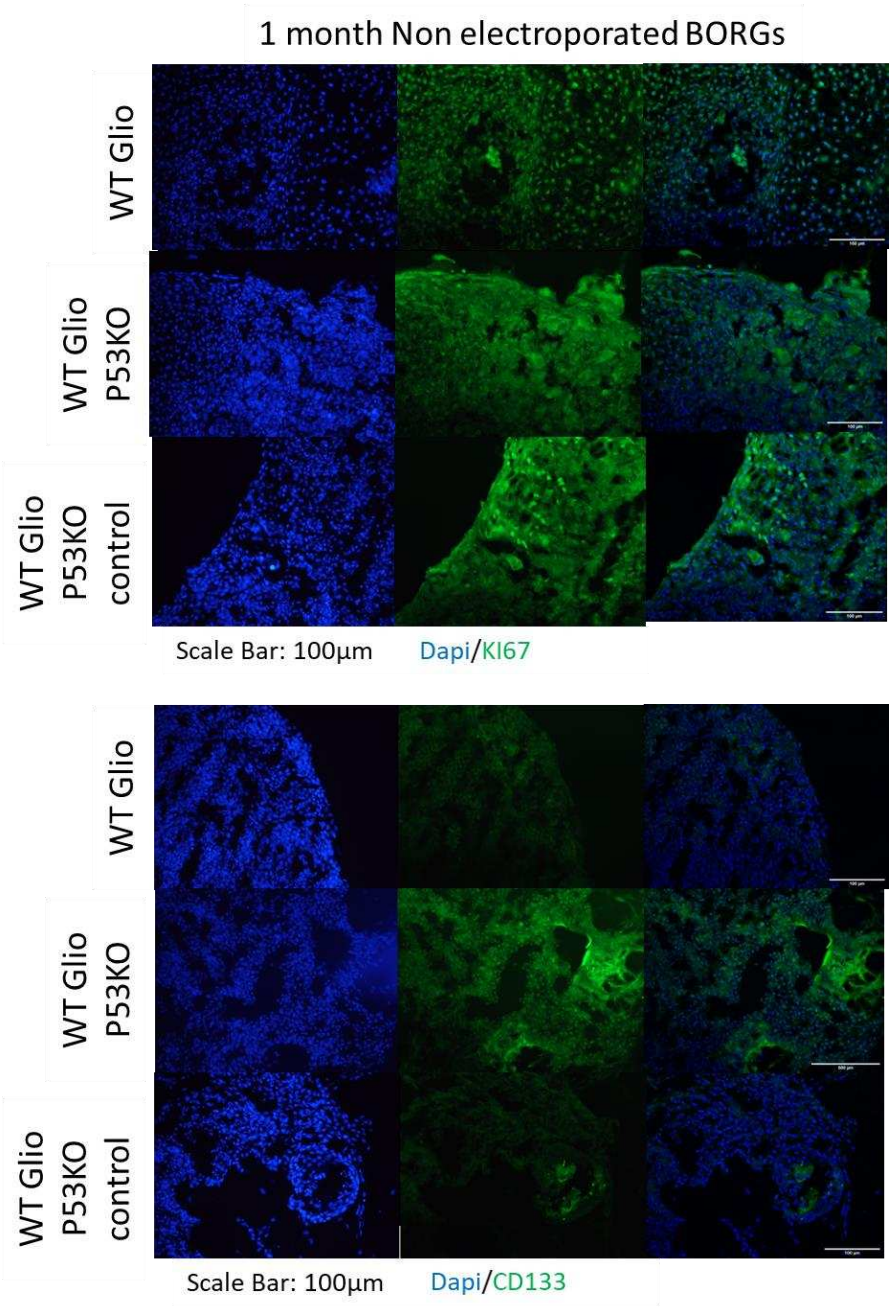
1 month Non electroporated BORGs



Scale Bar: 100µm Dapi/Tubb3/sox2

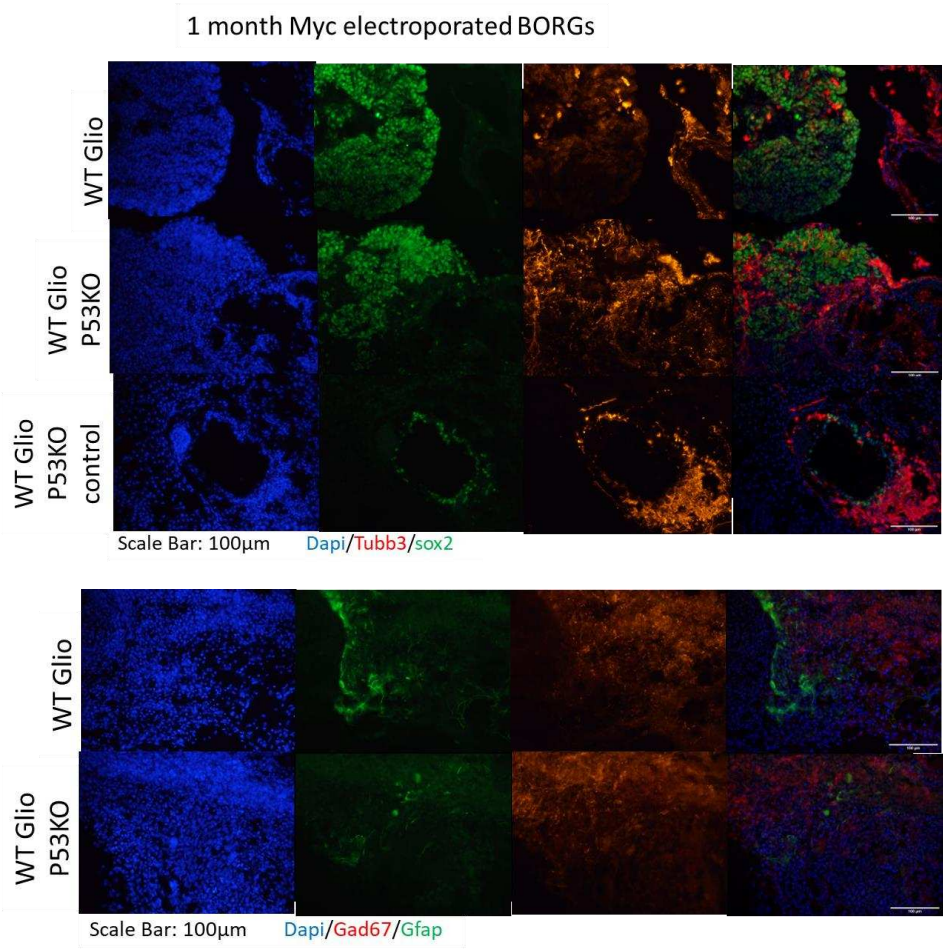


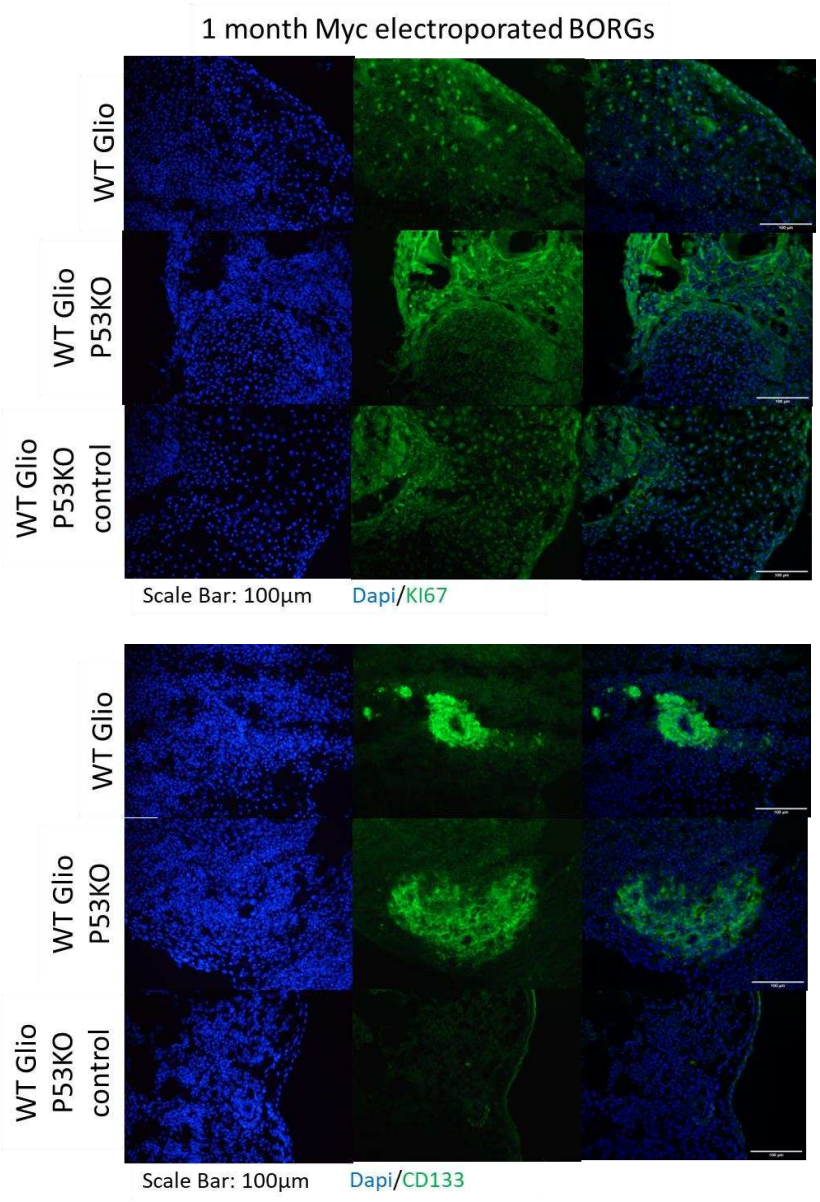
Scale Bar: 100µm Dapi/Gad67/Gfap



Supplementary Figure 17. Myc electroporated TP53KO Brain organoids immunohistochemistry assay of non- electroporated BORGs

Immunofluorescence staining, with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for 1 month old Brain organoids of WT Glio, WT Glio P53KO and WT Glio P53KO control cell lines

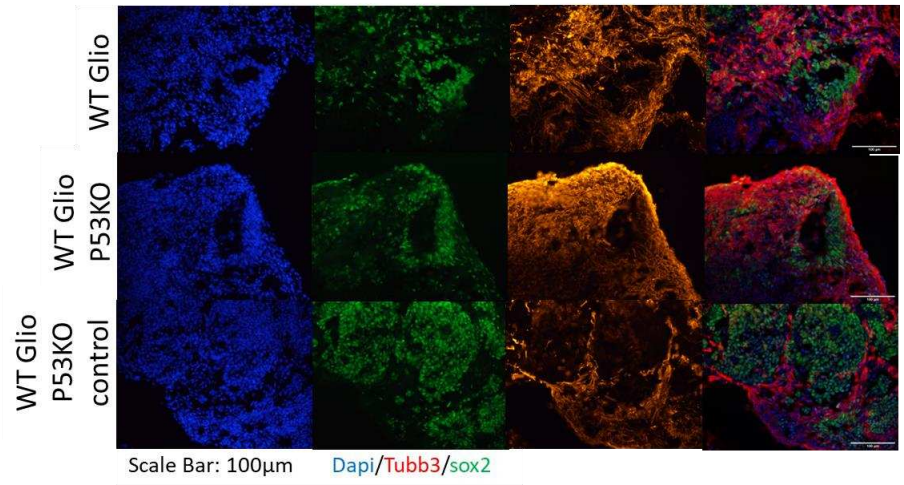

1 month Myc electroporated BORGs
WT Glio
WT Glio P53KO
WT Glio P53KO control
Scale Bar: 100µm
Dapi/Tubb3/sox2
WT Glio
WT Glio P53KO
Scale Bar: 100µm
Dapi/Gad67/Gfap



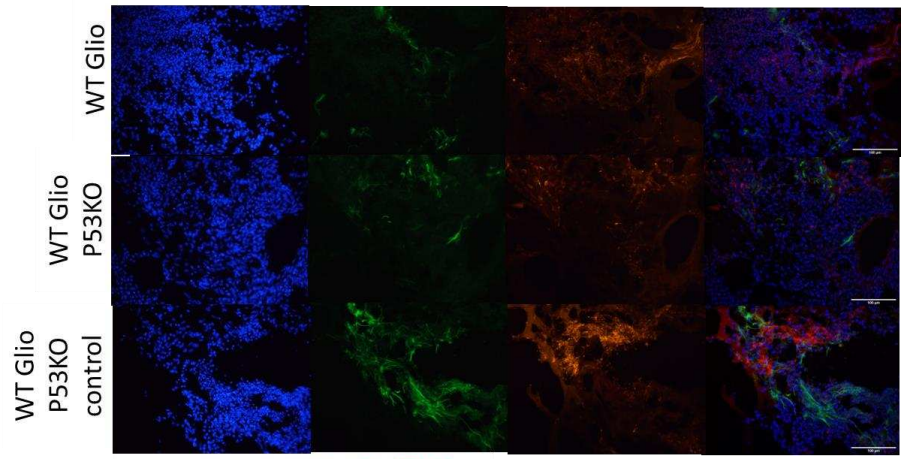
Supplementary Figure 18. Myc electroporated TP53KO Brain organoids immunohistochemistry assay of myc- electroporated BORGs

Immunofluorescence staining, with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for 1 month old Brain organoids of WT Glio, WT Glio P53KO and WT Glio P53KO control cell lines

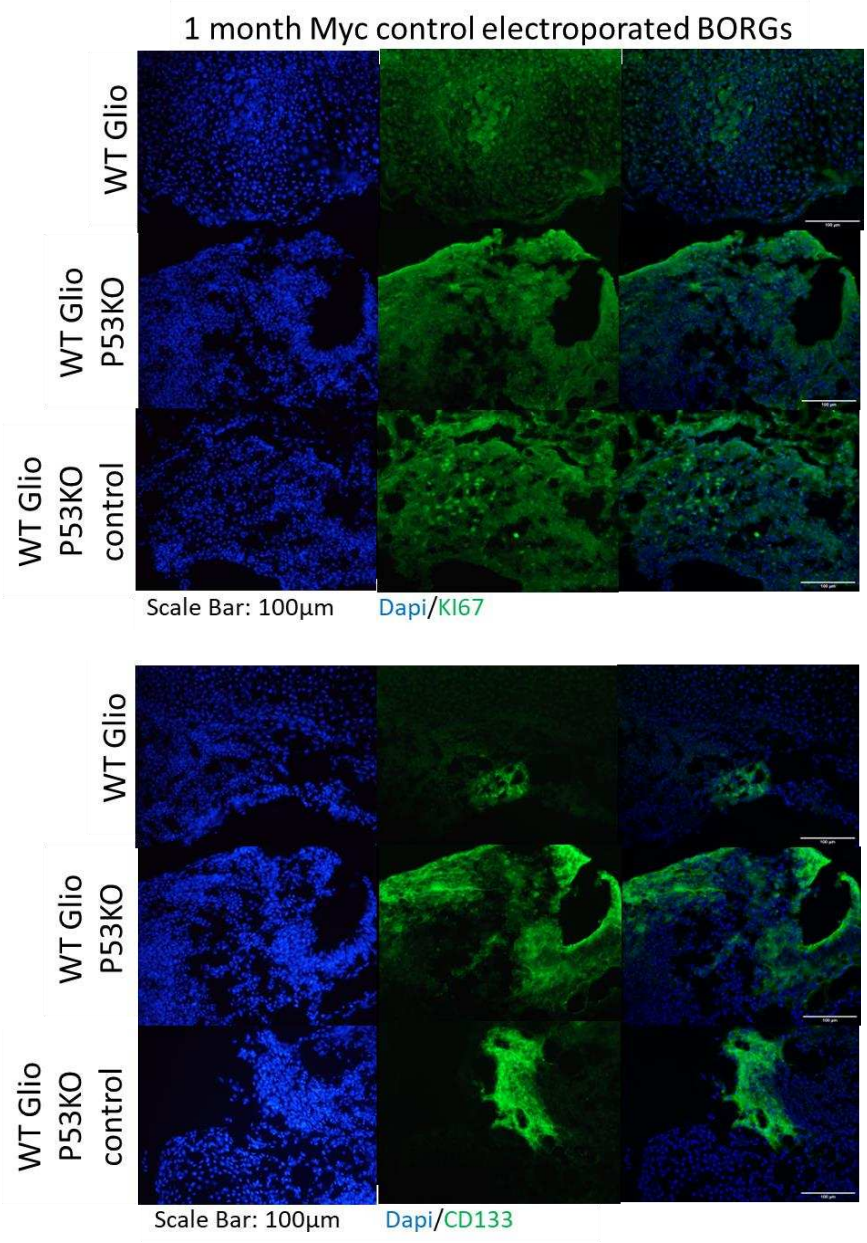
1 month Myc control electroporated BORGs



Scale Bar: 100µm Dapi/Tubb3/sox2



Scale Bar: 100µm Dapi/Gad67/Gfap



Supplementary Figure 19. Myc electroporated TP53KO Brain organoids immunohistochemistry assay of myc control- electroporated BORGs

Immunofluorescence staining, with antibodies for Sox2 (Neural stem cells),Tubb3 (immature neurons), Gfap (astrocytes), Gad67 (posymitotic gabaergic neurons), KI67 (proliferation), CD133 (Cancer stem cell) for 1 month old Brain organoids of WT Glio, WT Glio P53KO and WT Glio P53KO control cell lines

Synthèse en Français

Le système nerveux peut être classé en plusieurs parties : le système nerveux central (SNC) comprenant le cerveau et la moelle épinière, et le système nerveux périphérique (SNP) comprenant les nerfs, les ganglions et les plexus qui émergent du crâne et de la moelle épinière.

La première étape donnant naissance au système nerveux est le développement de la séquence primitive [1] au jour embryonnaire 6.5 (E6.5). À E8.0, il y a un approfondissement continu du sillon neural et le pli neural s'est délimité en l'ectoderme de la surface externe et le neuroectoderme interne [2,3]. E9.0, c'est lorsque le neuropore caudal (une ouverture qui existe entre 2 points de fermeture s'étendant l'un vers l'autre le long du tube neural) se ferme complètement. Vers E10, les parois télencéphaliques s'épaississent et se différencient. Le neuroectoderme a également commencé à évoluer vers la zone ventriculaire interne (VZ), une zone intermédiaire (couche du manteau) et une zone marginale externe. E12 à E15 sont marqués par des processus radicaux. Le plexus choroïde est visible et les nerfs olfactifs sont détectables lorsqu'ils se déplacent vers le cortex olfactif, et par la suite les lobes olfactifs deviendront plus distincts. Entre E16 et E18, le cortex cérébral est devenu bien structuré formant les 6 couches - la zone marginale, la plaque corticale, la sous-plaque corticale, la zone intermédiaire, la zone sous-ventriculaire (SVZ) et la VZ - et cette différenciation se poursuit jusqu'à quelques semaines après naissance avant de se stabiliser.

La thèse est divisée en deux sections. La première section se concentre sur la neurogenèse pilotée par les rétinoïdes. La deuxième section abordera la possibilité de générer des organoïdes cérébraux de souris récapitulant le glioblastome pédiatrique.

Section 1- Neurogenèse pilotée par les rétinoïdes

Au cours du développement neuronal, l'embryon reçoit une multitude de signaux provenant de diverses sources, et la combinaison de ces signaux d'une manière spatialement et temporellement distincte est à l'origine de la formation de régions distinctes du système nerveux. Les sources de ces signaux peuvent être largement classées en régulation intrinsèque et signalisation extrinsèque, y compris les molécules diffusibles, les interactions cellule à

cellule, les interactions cellule à matrice extracellulaire, les vaisseaux sanguins et le liquide céphalo-rachidien.

L'acide rétinoïque est un signal extrinsèque qui induit la différenciation cellulaire et l'apoptose des cellules. L'acide rétinoïque (RA) est un produit métabolique de la vitamine A, non produit de manière inhérente par le corps des mammifères, et donc ingéré à partir de sources externes sous forme de caroténoïdes et d'esters de rétinyle provenant respectivement de plantes et d'animaux. Le rétinol est converti en rétinaldéhyde via les alcool déshydrogénases (ADH) et les rétinol déshydrogénases (RDH), puis en acide rétinoïque via les rétinaldéhyde déshydrogénases (RALDH) [150]. L'acide rétinoïque se lie aux hétérodimères du récepteur de l'acide rétinoïque/récepteur X du rétinoïde (RAR/RXR), qui sont généralement liés aux éléments de réponse à l'acide rétinoïque (RARE) ou aux éléments de réponse X rétinoïque (RXRE) dans les régions promotrices ou amplificatrices des gènes cibles [151]. Les récepteurs RAR et RXR existent chacun sous forme d'isotypes, RARα/β/γ et RXRα/β/γ, codés par des gènes différents, et chacun d'eux possède ses propres isoformes, issues d'un épissage alternatif et/ou d'une utilisation alternative de promoteurs. Alors que les récepteurs RXR ne sont capables de se lier qu'à l'acide rétinoïque 9-cis (9-cis RA), les récepteurs RAR peuvent se lier avec une haute affinité à l'AtRA et au 9-cis RA.

Chaque récepteur isotype peut être activé avec des ligands spécifiques. Pour les récepteurs RAR, les agonistes pan-RAR tels que l'acide rétinoïque, le luffariellolide [25], sont largement étudiés alors qu'il n'existe aucun ligand naturel connu qui soit spécifique ou sélectif d'un isotype. Dans un tel cas, plusieurs ligands synthétiques ont été conçus comme Am 580, Am 80, BMS753, AGN 193835 et AGN 193836, qui sont spécifiques de RARα. Le BMS641 et les rétinoïdes dihydronaphtalènes 45 à 47 à substitution 2-thiényle sont des ligands modifiés qui sont spécifiés par RARβ, tandis que le BMS961 et les rétinoïdes liés à l'α-hydroxyacétamide 54 à 57 ont été décrits comme puissants pour l'activité RARγ, et le TTNPB comme un agoniste pan-RAR [23,26].

La capacité de chacun des agonistes synthétiques du RAR à réguler une réponse transcriptionnelle spécifique a été décrite très tôt par l'équipe de Pierre Chambon [27,29,152], et plus récemment, ces réponses spécifiques ont été démêlées au niveau des différents programmes de gènes contrôlés [30,32,153].

La répartition des RAR dans les embryons diffère d'une région à l'autre. Notamment, bien que la prégastrulation chez les modèles de souris et de rats ne révèle pas de niveaux détectables d'expression des gènes RAR/RXR, lors de la gastrulation (E7.5), l'expression de RARα et RARγ est omniprésente et diffuse, et l'expression de RARβ a été trouvée principalement dans les régions latérales du corps embryon [154]. À E7.5, il n'y a aucune signalisation détectable ou très faible des récepteurs RAR. De E8.5 à E13.5, on observe une présence accrue du RARα dans la région neuroectodermique. RARβ est fortement exprimé dans le mésencéphale à un stade précoce à E8.5, spécifique au tube neural. RARγ, en revanche, occupe l'extrémité opposée, se trouvant principalement dans la strie primitive en régression [152], et presque complètement absent du tissu mésodermique. Par E13.5, RARα est principalement exprimé dans le corps calleux, le corps striatum. À aucun moment RARγ n'est détecté dans le SNC en développement.

La fonctionnalité de l'acide rétinoïque fonctionne également en synchronisation avec des signaux complémentaires tels que le facteur de croissance des fibroblastes (FGF-8), le sonic hedgehog (SHH) et les protéines morphogénétiques osseuses (BMP). L'acide rétinoïque joue un rôle principalement dans la croissance et la différenciation des structures postérieures, tandis que dans la partie antérieure de l'embryon, son activité est régulée par les enzymes CYP26A1 et CYP26C1 dégradant la RA, favorisant la configuration antéro-postérieure [155–157]. Dans l'axe dorsoventral du tube neural en développement, l'acide rétinoïque est produit au niveau des sites somites, ainsi que SHH et BMP exprimés ventralement et dorsalement, respectivement. L'expression du FGF-8 est détectable à l'extrémité postérieure du tube neural en extension, et ces gradients de facteurs prédisent le sort ultime des neurones spécialisés qui émergent, tels que les interneurones, les neurones sensoriels et les motoneurones [158–160].

En plus de l'action directe de l'acide rétinoïque dans la structuration régionale du neurectoderme, plusieurs études *in vivo* ont discuté de l'importance de l'acide rétinoïque pour la génération des motoneurones et des progéniteurs ventraux [158,161,162], sous-ensembles de GABAergiques ou les neurones dopaminergiques [163–165], ainsi que la différenciation neuronale terminale dans les zones ventriculaires et sous-ventriculaires [166]. Dans les années 1990, Shanthini Sockanathan a prouvé l'effet des rétinoïdes synthétisés par les neurones qui influencent la capacité de différenciation des neurones LMC latéraux, mais aussi la quantité, l'identité du sous-type et le moment de la maturation au niveau des membres. Dans un autre

exemple, l'expression du facteur de transcription 2 des oligodendrocytes (Olig2), pilotée par l'acide rétinoïque et l'action de SHH, marque l'identité des progéniteurs des motoneurones et est étroitement liée à l'expression des régulateurs du domaine homéobox (HD). L'expression d'Olig2 est régulée négativement par l'expression de l'homéobox 2 NK2 (Nkx2.2) (pilotée par SHH), tandis qu'Olig2 régule négativement l'expression du facteur HD Pax6 [167]. De plus, il a été découvert que l'acide rétinoïque a un effet dose-dépendant sur l'identité des neurones p3 V3 différenciés par rapport aux neurones sérotoninergiques [165,168]. Les progéniteurs p3 constituent le domaine progéniteur le plus ventral présent dans la moelle épinière et le cerveau postérieur, les deux étant des populations équivalentes. Ils donnent naissance aux interneurones glutamatergiques V3 dans la moelle épinière et aux neurones sérotoninergiques (5-HT) dans le cerveau postérieur.

Des efforts consciencieux ont été déployés pour caractériser l'effet de la RA indépendamment et en combinaison avec d'autres facteurs de croissance et morphogènes, en imitant l'embryogenèse *in vivo*, *in vitro*. L'acide rétinoïque libéré par les somites et le gradient SHH produit par la plaque de plancher et la notocorde [63–65] fournissent respectivement un gradient rostral-caudal et ventral-dorsal, influençant l'émergence des domaines interneurones progéniteurs ventraux (p0-p3) et un domaine de motoneurone progéniteur (pMN) disposé dans l'axe ventral-dorsal ; qui finissent par devenir les classes d'interneurones ventraux et les motoneurones [63,66,67]. La RA induit un effet caudalisant (particulièrement visible dans les corps embryoïdes des cellules progénitrices neurales), et SHH crée une spécialisation des progéniteurs des motoneurones [69].

Ces mêmes études se sont étendues à l'utilisation de cellules souches pluripotentes humaines (hPSC) comme modèle de recherche, ouvrant la voie à des connaissances plus pertinentes et transférables sur le plan clinique sur le devenir des cellules neuronales. Dans le cas de la génération de motoneurones, le traitement par RA et SHH est jugé nécessaire pour les cellules neuroectodermiques Sox1+ dérivées de cellules souches embryonnaires humaines (CSEh), de préférence à faibles concentrations, avec un traitement à l'acide rétinoïque survenant précocement [72]. Ceci permet d'éviter la spécificité régionale que les cellules acquièrent une fois qu'elles deviennent Pax6+/Sox1+.

La différenciation neuronale à l'aide de l'acide rétinoïque continue d'être testée pour déterminer les meilleures conditions possibles pour générer des types de cellules neuronales.

Des durées d'exposition plus longues à la RA, la culture de cellules sous forme de corps embryonnaires, en particulier lors de leur exposition à l'acide rétinoïque, la culture de cellules à des densités cellulaires plus élevées, sont des facteurs qui augmentent le rendement neuronal [74]. Des densités cellulaires plus élevées favorisent la diaphonie cellulaire et ont un effet sur la régulation positive des facteurs de neurogenèse tels que Sox2, Neurod1, Pax6, influençant finalement la formation neuronale Tuj1+.

Le lien entre RAR/RXR et la spécialisation cellulaire a déjà été exploré dans des articles examinant la capacité de différenciation de la RA dans les cultures de cellules P19 et F9 [30,79,152,153]. La compréhension du rôle de RAR/RXR dans les études de spécialisation cellulaire reposait sur l'utilisation d'agonistes synthétiques, d'autant plus que RAR ne possède aucun agoniste naturel connu spécifique à un sous-type. Dans les années 1990, Taneja, Roy et Chambon ont exploré les fonctionnalités de l'activité RAR/RXR dans les cellules de carcinome embryonnaire [27,152]. À de très faibles concentrations, les agonistes synthétiques de chaque RAR ne sont pas capables d'induire une expression génique efficace. La présence de ligands spécifiques de RAR à des concentrations appropriées était suffisante pour induire la différenciation dans les cellules EC en activant l'hétérodimère RAR/RXR. En outre, ils discutent des redondances fonctionnelles des sous-types de récepteurs RAR dans l'induction de certains gènes sensibles à la RA dans la différenciation des cellules EC P19 et F9. Cette redondance dépend du type de cellule et du contexte du promoteur. Par exemple, RARγ est capable d'induire la différenciation des cellules F9, tandis que les cellules P19 peuvent être induites à le faire par RARα ou RARγ [152]. Un RAR particulier peut avoir la capacité d'induire une réponse génétique spécifique, mais cela dépend également de l'activité des autres sous-types de récepteurs et de sa « dominance » lorsque les 3 récepteurs sont actifs. Ceci est visible dans les études knock-out où l'un ou l'autre des récepteurs est capable de reprendre dans une certaine mesure les profils d'expression génique [27,80]. Comparé à l'induction par ATRA dans les cellules F9, BMS961 (agoniste RARγ) est capable de récupérer 62% des gènes cibles RXRα – RARγ, contrairement à BMS753 (agoniste RARα) et BMS641 (agoniste RARβ) induisant 40 et 10% du réseau. respectivement. Bien que RARα puisse médier l'activité de RARγ au cours de la réponse génique médiée par la RA, les niveaux de récupération peuvent être influencés principalement en raison du fait qu'il existe une abondance de RARγ dans les cellules F9, généralement par rapport aux autres sous-types.

Néanmoins, de multiples études sont en mesure de confirmer la redondance fonctionnelle, ainsi que l'influence spécifique du destin cellulaire de ces récepteurs. Plus précisément, dans les cellules F9, ATRA et BMS961 sont capables d'induire une différenciation cellulaire vers la lignée endodermique, alors que seuls ATRA et BMS753 produisent un devenir cellulaire neuronal dans les cellules P19 EC [153]. Bien que chaque agoniste du récepteur puisse induire une transition de destin cellulaire différente dans différentes lignées EC, ils activent tous deux également un noyau commun de programmes pouvant avoir des modèles d'expression temporelle différents, constituant plus de 60% des gènes exprimés différentiellement induits par la RA. Chaque voie de transition du destin cellulaire a révélé des TF spécifiques à l'engagement et spécifiques au programme commun. La différenciation induite par la RA dans les cellules souches embryonnaires de souris reconstitue une réponse combinée de ce qui est observé dans la différenciation P19 et F9. La différenciation n'est pas homogène et indique la pluripotentité et le non-engagement des cellules ES par rapport aux cellules EC. 65 et 75% des gènes régulés négativement et positivement dans les cellules P19 sont récupérés dans la différenciation ES, ce qui représente environ 30 % pour les gènes régulés positivement dans les cellules F9 [153]. Ceci fournit un outil utile dans la détermination d'une réponse spécifique au sous-type RAR à étudier dans des cellules pluripotentes non engagées. En 2017, Podleśny-Drabiniok a démontré que les cellules traitées au CD666 (agoniste RARγ) réussissaient le mieux à générer des neurones GABAergiques, c'est-à-dire, 77 %, et BMS641 (agoniste RARβ), le moins efficace, c'est-à-dire 28%.

Bien que des travaux approfondis aient été réalisés dans l'étude de la différenciation EC en relation avec l'activation des isotypes RAR/RXR, il manque un chaînon dans nos connaissances sur ces récepteurs dans la différenciation des cellules ES. Le lien entre les récepteurs RAR spécifiques et leur influence ultime sur le destin cellulaire reste encore à décoder, comme l'illustre Podleśny-Drabiniok et ses collègues [79].

L'acide rétinoïque est très clairement démontré comme étant un élément crucial pour le développement des organoïdes cérébraux [83,84]. Aux premiers stades du développement, il incite les cellules à s'engager dans la voie neuroectodermique, que ce soit pour la génération d'organoïdes cérébraux entiers sans modèle ou spécifiquement pour les organoïdes cérébraux à motifs cérébraux ou dorsaux. L'état actuel de la recherche évolue rapidement pour inclure l'analyse de cellules uniques [84,85,169], la transcriptomique spatiale et le traçage de lignées

[87–89], et nous sommes déjà témoins de la possibilité de développer des structures de type embryonnaire à partir de cellules souches (stembryos) [90,91]. Ainsi, la nécessité d'étudier l'influence de chaque RAR sur le développement du cerveau et le devenir des cellules neurales reste très forte, ce qui pourrait fournir un moyen de thérapie ciblée dans un avenir proche.

Résultats

Nous abordons le lien entre les récepteurs RAR spécifiques et leur influence sur le devenir cellulaire dans les cellules souches neuronales différenciées dans notre article récent [32]. Afin de déchiffrer le rôle de chaque récepteur individuel, un test de différenciation neuronale 2D a été réalisé sur des cellules P19 EC en utilisant des agonistes synthétiques pour chaque isotype de récepteur (individuellement et/ou en combinaisons) pendant 10 jours. La découverte majeure est que l'activation synergique de RARβ et RARγ conduit à des niveaux de récupération plus élevés des marqueurs associés aux cellules neuronales, et que les programmes génétiques activés par ces récepteurs sont capables de mieux récupérer la signature neuronale lorsqu'ils sont activés dans les cellules RARα KO. Ceci suggère un effet inhibiteur de RARα dans des conditions de type sauvage qui pourraient ne plus supprimer l'activité de RARβ et de RARγ. La comparaison des programmes activés par RARα et RARβ+ RARγ révèle un sous-ensemble de programmes de 235 gènes appelés « programmes inhibés par le RARα non ligandé ». Ces programmes sont constitués de facteurs de transcription qui peuvent être classés en fonction de leur influence sur le réseau de régulation en aval.

Étant donné que les expériences ont été réalisées à l'aide de données transcriptomiques globales, nous avons effectué une analyse transcriptomique unicellulaire sur des MESC neuronalement différenciées. Les cellules ont été différenciées pendant une période de 4, 8 et 16 jours. La bibliothèque a été préparée à partir des cellules à l'aide du kit 10X Genomics Chromium Single cell 3' Reagents v3 et les données ont été normalisées, filtrées et alignées. Nous sommes capables de visualiser 17 clusters différents. Dans la condition Bêta + Gamma du Jour 4, le cluster 12 est fortement exprimé. Au jour 16, les cellules traitées par l'agoniste Alpha présentent une forte expression pour les groupes 2, 5, 6, 8, 13, 17, tandis que les cellules traitées par Beta + Gamma présentent une forte expression pour les groupes 3 et 11. Au jour 4 et au jour 16, nous voir la présence de lignées d'endoderme et de mésoderme dans la population cellulaire, ce qui implique que nous avons de multiples voies de différenciation dans ces cellules, sans rapport avec le ligand auquel elles ont été exposées. Ceci est similaire

à ce qui est observé dans Mendoza-Parra et al, 2016, où la différenciation des MESC induite par la RA a abouti à des programmes de différenciation endodermique et ectodermique [153]. Au jour 16, le cluster 8 profile fortement les marqueurs neuronaux et le cluster 13 pour les astrocytes tandis que les OPC sont plus présents dans les clusters 5, 3, 11. Nous trouvons également les clusters 8, 13 et 3 mis en évidence pour l'expression principalement des neurones GABAergiques et glutamatergiques. Une autre observation est qu'au jour 16, dans les deux traitements, nous pouvons voir l'apparition robuste de nouveaux clusters tels que le cluster 10 qui semble avoir une signature endodermique, éventuellement capable d'exprimer également des marqueurs de l'endoderme mature, le cluster 8 avec une signature neuronale et plus spécifiquement les neurones GABAergiques et glutamatergiques, et le cluster 13 qui présente un profil mixte montrant des caractéristiques de cellules souches, de progéniteurs et de neurones glutamatergiques suggérant un état de transition intermédiaire pour ce cluster. Dans l'analyse préliminaire de l'ontologie des gènes, nous constatons que le groupe 3 n'a pas de type cellulaire distinct mais montre un possible phénotype progéniteur/interneurone apporté par les gènes LDHA, HMGB2, PCLAF, DUT, mais également une lignée mésodermique issue des gènes d'expression des cellules sanguines - RAN, NPM1. Le cluster 11 contenant FAT3, DACH1, NPAS3, HES5, CCND2 suggère une identité de cellules neurales immatures telles que des cellules gliales radiales, des interneurones, des cellules précurseurs et même des neurones matures (cellules de Cajal Retzius).

Les cellules cultivées dans des structures 2D ou 3D interagissent différemment. De nombreuses études knock-out RAR corroborent les découvertes sur la nécessité de certains récepteurs au cours du développement embryonnaire [158,161–166], et c'est ce qui a motivé l'allongement de la durée de nos études sur la différenciation neuronale et l'introduction de la complexité d'une structure tridimensionnelle en plus de les cultures monocouches. Pour réaliser la complexité de la différenciation neuronale régulée par les sous-types de récepteurs rétinoïdes dans un format 3D, nous avons utilisé la technologie des organoïdes cérébraux qui est durable sur une période plus longue (mois) et avons suivi leur développement en relation avec les ligands synthétiques des rétinoïdes (ATRA, BMS753 et BMS641+961). Nous sommes en mesure de constater que le traitement de ligands de récepteurs individuels est toujours capable de produire des BORG qui se développent dans la voie de différenciation neuroectodermique. Cependant, nous ne pouvons pas observer de différences frappantes

entre les traitements des ligands. Cela peut éventuellement être attribué à la redondance des récepteurs rétinoïdes en cours de développement, et une comparaison plus approfondie nécessiterait une approche telle que le séquençage de l'ARN unicellulaire et la transcriptomique spatiale pour élucider l'acquisition du destin cellulaire attribuée à l'activation de l'isotype du récepteur dans des cultures à long terme.

Section 2- Développement d'organoïdes cérébraux de souris récapitulant le glioblastome pédiatrique muté H3.3

Dans la section précédente, nous nous concentrons sur les processus normaux de développement cérébral et sur l'étude du destin cellulaire piloté par les récepteurs de l'acide rétinoïde. Mais l'étude des processus de développement modifiés est tout aussi intéressante. Et cela peut être dû à des maladies dégénératives ou à la survenue d'un cancer. En ce qui concerne le développement cérébral, nous nous sommes intéressés à étudier les caractéristiques du glioblastome, plus spécifiquement du glioblastome pédiatrique. Actuellement, les moyens de traitement pour ces patients atteints de cancer sont très limités. De plus, les modèles de recherche actuels s'appuient sur des échantillons dérivés de patients et des lignées d'iPSC humaines portant les mutations spécifiques du glioblastome pour étudier cette maladie. Dans cette section, je discuterai de l'état actuel de la recherche sur la modélisation du glioblastome et proposerai l'utilisation d'un modèle organoïde de cerveau de souris pour étudier les tumeurs spécifiques à une mutation.

Les gliomes sont des tumeurs cérébrales provenant de cellules gliales cancéreuses. Le glioblastome provient des astrocytes, le principal composant des cellules gliales du SNC. Bien qu'elles présentent des marges distinctives lors de l'imagerie, elles sont infiltrantes de manière diffuse [94]. Les tumeurs du SNC sont la deuxième tumeur la plus fréquente chez les enfants après la leucémie et les tumeurs solides les plus courantes avec une incidence de 30 par million [95]. Selon les définitions de l'âge et les grades définis par l'OMS pour le glioblastome pédiatrique (pGBM), l'incidence peut varier de 3 à 15 % [96]. La période médiane de survie varie de 13 à 73 mois avec un taux de survie à 5 ans inférieur à 20 %. À ce jour, aucune référence en matière de traitement n'a été établie. Le protocole le plus efficace et le plus largement suivi

est la résection chirurgicale de la tumeur en association avec un traitement oral au témozolomide [94]. Les pGBM présentent des variations de signature par rapport au glioblastome adulte en termes de mutations et de pronostics correspondants, ce qui peut modifier considérablement les modalités et l'efficacité du traitement. Les tumeurs du SNC les plus courantes chez l'adulte sont les HGG diffuses, alors que seulement 10 % des tumeurs cérébrales pédiatriques peuvent être classées dans cette classification. Par conséquent, beaucoup moins de marqueurs moléculaires ont été établis. Les cas de glioblastome chez les enfants sont caractérisés par l'absence de mutations IDH, l'absence de co-délétion 1p/19q [93] et moins de mutations d'EGFR et/ou PTEN. En revanche, il existe une forte corrélation avec les altérations génétiques p53 [94], ATRX et PDGFRA [97].

Typiquement, les gliomes survenant chez les enfants peuvent être séparés en 2 catégories : les gliomes de bas grade qui sont généralement mutants dans les gènes BRAF, affectant la signalisation MAPK en aval, et les gliomes de haut grade qui présentent généralement des mutations dans le locus H3F3A [93], connu comme variante 3 de l'Histone 3. Dans les pLGG (gliomes pédiatriques de bas grade), les anomalies les plus fréquemment détectées se concentrent sur BRAF (48%), le faux-sens FGFR1 (17,6%), NF1 (8,8%) et TP53 (5,6%) [97]. Alternativement, TP53 (49 %), H3F3A (37,6 %), ATRX (24,2 %), NF1 (22,2 %) et PDGFRA (21,7 %) représentent les mutations les plus fréquentes du pHGG (gliomes pédiatriques de haut grade). L'histone 3.3 (H3.3) est essentielle à la transcription [100]. Plusieurs sites de cette protéine sont modifiés post-traductionnellement, ce qui entraîne des effets en aval sur la différenciation, le maintien de l'état des cellules souches, etc. En raison de leur profil évolutif hautement conservé, les maladies impliquant des histones mutées sont rares. H3.3 est une variante exprimée de manière constitutive et présente de manière diffuse. La production de H3.3 est codée par les gènes H3F3A et H3F3B, et le variant diffère des autres variants H3 par plusieurs acides aminés qui régulent l'affinité des protéines chaperons et l'expression de la chromatine [101]. Les deux mutations les plus courantes liées au locus H3F3A se trouvent sur la queue amino-terminale de l'histone, H3K27M et H3G34R/V où la lysine en 27ème position est convertie en méthionine et la glycine en position 34 est convertie respectivement en arginine/valine. Les rôles de ces acides aminés dans la régulation des événements de transition cellulaire ont été décrits récemment, mais de manière peu approfondie. Comme le rapportent plusieurs études, 30 % des patients pGBM expriment les

mutations H3.3, soit K27M, soit G34R/V. Des mutations TP53 sont survenues dans 86 % des échantillons présentant des mutations H3F3A et/ou ATRX [102]. Korshunov et al, 2015, ont présenté plus de 50 % des cas de pGBM profilés comme présentant des mutations H3F3A (43 % pour K27 et 15 % pour G34) [103].

H3K27 est normalement une cible d'EZH2, devenant di- ou tri-méthylé. Le H3K27 méthylé, H3K27me3, agit comme un répresseur transcriptionnel, inhibant l'association et la liaison des complexes de transcription. Puisque la plupart des variantes H3.3 occupent des positions du génome impliquées dans le maintien d'états indifférenciés, le H3K27M muté peut conduire à des profils oncogènes. On suppose que la mutation K27M soit provoque un obstacle stérique à la liaison de EZH2, soit entre très probablement en compétition pour la liaison avec EZH2, séquestrant et inactivant ainsi le complexe PRC2 de manière permanente et provoquant une réduction globale de la méthylation de K27 [106,107]. Cela entraîne une augmentation de l'expression des gènes au niveau des sites d'expression régulés par H3K27 et au niveau des promoteurs bivalents (H3K27 et H3K4). Bien que K27M ait été déduit comme étant une mutation motrice du pGBM, ce n'est probablement pas le seul événement mutagène capable de propulser la formation de tumeurs [106]. Il a été fréquemment démontré que les mutations H3F3A sont capables d'induire une hyperprolifération cellulaire mais pour donner naissance à des tumeurs voraces, d'autres mutations motrices semblent être nécessaires comme les mutations TP53 [99]. Cela corrobore les études sur la souris, dans lesquelles la mutation H3K27M, couplée à la perte de P53 dans les progéniteurs de Nestin, était capable de générer des amas ectopiques proliférants mais pas de gliomes à grande échelle [108].

La glycine en position 34, bien qu'elle ne subisse pas de modifications post-traductionnelles, se trouve à proximité immédiate de K36, dont le modèle de méthylation affecte la transcription. La triméthylation H3K36 joue un rôle actif dans la réparation de l'ADN. L'acide aminé triméthylé interagit avec le domaine PWWP de MutSα, une protéine impliquée dans la réparation des mésappariements des dommages à l'ADN. Le recrutement de MutSα sur le site du dommage initie un processus de réparation des mésappariements [110]. H3G34R/V/D convertit le résidu glycine en un résidu de chaîne latérale volumineux qui inhibe stériquement les interactions protéiques en position K36. Ces interactions impliquent la SETD2, une triméthyl transférase spécifique de H3K36, qui reconnaît le motif G33-34. De même, la

liaison de la NSD1/2 méthyltransférase au site est également inhibée [107,110]. De plus, H3K36me3 ne peut pas interagir avec MutSα.

Au cours de la dernière décennie, les organoïdes ont été largement utilisés dans la recherche pour modéliser les processus de développement normaux, mais également pour étudier les états cellulaires modifiés, comme on le voit dans le cancer, les troubles du développement, les maladies, etc. Les organoïdes sont des cultures tissulaires tridimensionnelles et auto-organisées qui reconstituent soit un organe entier ou une partie de celui-ci. Ils sont dérivés de cellules souches, différant essentiellement des cultures sphéroïdes qui sont des cultures de tissus flottants et non définis composés de tissus cancéreux. Ils peuvent récapituler de manière plus réaliste l'environnement *in vivo*, ainsi que le processus chronologique de développement et de maturation de l'organe. Des organoïdes ont maintenant été générés pour une multitude d'organes, notamment le cerveau, l'intestin, l'estomac, les reins, les poumons, le cœur, etc.

Même avec le succès relatif des organoïdes dans la recherche fondamentale, de nombreux défis subsistent lors de la culture et de la maintenance de ces systèmes, le premier étant le manque de reproductibilité [104,114]. Étant donné que les organoïdes se développent généralement dans des gouttelettes de Matrigel en suspension sans support externe ni moules, ils donnent naissance à des structures uniques dans chaque culture individuelle. De plus, même si les apports nutritionnels et moyens, les températures et les procédures de culture restent standardisés, il existe toujours des variations entre les lots, les compositions cellulaires et l'architecture. De plus, il existe une complication selon laquelle toutes les cellules induites à se différencier ne se déplaceront pas vers l'engagement neuroectodermique [115]. La deuxième limitation rencontrée est l'absence de système vasculaire. Une fois que les organoïdes atteignent quelques millimètres de diamètre, ils ne peuvent plus transporter efficacement l'oxygène par diffusion vers les noyaux les plus internes. Cela conduit à des noyaux nécrotiques et sous-développés. Cela peut expliquer en partie pourquoi la plupart des modèles organoïdes ne survivent pas au-delà de 9 à 12 mois de culture et ne dépassent pas 5 à 7 millimètres de diamètre. Ce problème a été partiellement contourné grâce à l'utilisation de bioréacteurs à rotation, des solutions plus efficaces sont toujours à l'étude et ont également montré des résultats fructueux, comme l'utilisation de constructions à fibres creuses [116], des cocultures avec des cellules endothéliales [117] et la modification génétique de gènes tels que

ETV2 [118]. Troisièmement, la maturation complète des tissus. La majeure partie du développement précoce et de la différenciation de l'organoïde est bien formée mais ne parvient pas à se poursuivre de la même manière au fil du temps ; certaines structures présentant des architectures distinctes tandis que d'autres parties restent indifférenciées. Comme le montrent les cultures 3D établies par Lancaster et Knoblich [82], les organoïdes n'atteignent pas leur pleine maturité et continuent de conserver une population de cellules souches/cellules souches neuronales.

Le glioblastome est une maladie largement étudiée, récapitulée *in vitro* dans les systèmes organoïdes cérébraux à l'aide de diverses méthodologies. Ceci est mis en évidence dans l'étude de Bian et al, 2018, dans laquelle ils ont développé un organoïde cérébral néoplasique (neoCOR), dans lequel des constructions CRISPR/Cas9 et/ou transposon sont utilisées pour introduire des oncogènes ou perturber les gènes suppresseurs de tumeurs dans le hiPSC (souche pluripotente induite par l'homme). cellules) dérivés d'organoïdes cérébraux [119]. Les organoïdes cérébraux surexprimés par Myc ont tendance à présenter une surabondance de Sox2 et de CD99, tandis que les organoïdes du groupe Glioblastoma ont une plus grande importance de l'expression de S100β et de GFAP (astrocytes). De la même manière, CRISPR/Cas9 a été utilisé pour introduire une mutation HRasG12V dans un locus TP53, dans des organoïdes cérébraux dérivés de CSEh (cellules souches embryonnaires humaines) âgés de 4 mois [170]. Étonnamment, dans les 4 mois suivant l'électroporation, l'organoïde a montré une présence accrue de la construction HRASG12V tdTomato-positive dans jusqu'à 86 % de l'organoïde. Une population cellulaire importante souvent trouvée chez les patients atteints de glioblastome est constituée de GSC (cellules souches de gliome). Comme le montrent plusieurs études [121–123], les GSC restent souvent inaccessibles à la résection tumorale et développent une résistance à la chimiothérapie actuellement approuvée. Afin de pouvoir les étudier *in vitro*, le modèle GLICO a été développé. Il implique des organoïdes cérébraux dérivés de CSEh ou de CSPi qui sont co-cultivés avec des GSC dérivées de patients et surveillés de près pour suivre l'invasion et la prolifération [124]. En une semaine, des GSC GFP+ peuvent être observées présentant un bord infiltrant diffus, rappelant la morphologie tumorale *in vivo*, et plus de 20 % de ces tumeurs présentent une coloration KI67+ (marqueur de prolifération). Des efforts ont également été déployés pour utiliser directement des échantillons cliniques de glioblastome pour générer du tissu organoïde dans un milieu

entièrement défini et sans sérum [125]. Ces GBO (organoïdes de Glioblastoma) sont découpés en morceaux de 1 mm et directement placés dans un agitateur orbital, fournissant des organoïdes entièrement formés en moins de 2 semaines. Cela contraste fortement avec d'autres protocoles dérivés de l'iPSC, qui nécessitent un minimum d'un mois pour commencer à montrer des marqueurs de maturité cellulaire. Puisque ces échantillons proviennent directement des patients, certaines structures telles que les vaisseaux sanguins et le système vasculaire CD31+ subsistent, récapitulant presque à l'identique le microenvironnement tumoral chez un patient.

Dans des cas très limités, des organoïdes cérébraux de souris ont été développés avec succès. Mais les avantages du développement d'un tel modèle clinique résident dans les cycles de division plus courts, conduisant à un délai plus court pour atteindre la maturité, utile pour l'optimisation des protocoles avec des coûts de production inférieurs [126]. Il existe également une pénurie de données expérimentales *in vivo* sur des souris déjà facilement disponibles qui peuvent être examinées de manière croisée pour comparer la physiopathologie, les marqueurs de développement, les tests de toxicité des médicaments, etc. Des protocoles antérieurs ont été développés pour utiliser des cellules souches embryonnaires de souris auto-agrégées (MESC) pour se développer en structures qui ressemblent aux structures corticales et à la cupule optique [127,128]. Plus récemment, des organoïdes cérébraux de souris ressemblant à des structures cérébrales avec un phénotype de cerveau antérieur dorsal ont été développés [129]. En l'espace de 5 à 6 semaines, ils ont produit des organoïdes corticaux matures, hautement reproductibles et robustes. Ainsi, il est possible de modéliser spécifiquement le glioblastome dans les organoïdes cérébraux de souris. Il offre des avantages tels que des coûts de réactifs moins élevés, des délais éventuellement plus rapides et un moyen de vérifier les protocoles avant d'investir dans des organoïdes cérébraux dérivés d'iPSC humaines pour des études de maladies, comme souligné précédemment.

Les recherches actuelles se concentrent sur l'utilisation d'échantillons provenant de patients et de lignées de cellules souches humaines pour étudier les maladies. Cependant, il existe un avantage à développer des modèles de souris *in vitro* capables de récapituler avec succès de telles maladies. Ici, nous travaillerons avec des lignées cellulaires mutantes H3.3 – H3.3K27M, H3.3G34R – pour développer des organoïdes cérébraux de souris qui présentent des caractéristiques tumorigènes telles que la prolifération et la prolifération.

Résultats

Le premier lot d'organoïdes cérébraux a été développé à l'aide du protocole décrit dans Lancaster et al, 2013, pour les cellules souches pluripotentes induites par l'homme (hiPSC) avec une période initiale d'engagement neuronal de 10 jours et induction, avant que les EB (corps embryoïdes) générés ne soient immergés dans des gouttelettes de Matrigel et transférés dans un milieu de maturation. Leur croissance sur une période de 3 mois ne présente globalement pas de différences significatives. À 2 mois, il existe toujours une niche de cellules souches neurales (Sox2) et des taches de prolifération (KI67). Nous avons observé un regain de tige au bout de 3 mois qui était plus robuste dans les organoïdes G34R, par rapport aux 2 autres lignées cellulaires. On remarque une bonne population d'astrocytes et, étonnamment, une population de microglies également. Puisque ces cellules émergent de la lignée mésodermique, elles prouvent que le protocole n'est pas pénétrant à 100%. Il existe probablement des niches de cellules dans les stades initiaux qui ne sont pas exposées en quantités suffisantes aux ligands initiaux et qui s'échappent vers le destin mésodermique/endodermique. Il existe également une expression confirmée de marqueurs neuronaux matures/postmitotiques. Bien qu'aucune des lignées cellulaires ne présente une différence comparable dans l'expression de KI67, une expression plus élevée est observée pour CD133 dans les organoïdes K27M et G34R. Certains modèles observés dans l'organoïde dérivé de lignées cellulaires mutantes H3.3 suggèrent un profil oncogène, mais il n'y a pas d'expression manifeste d'une prolifération tumorale, d'une prolifération excessive et d'une expansion caractéristiques du glioblastome. Nous avons donc utilisé le protocole de Lancaster et al, 2013, pour produire un nouveau lot d'organoïdes avec des délais d'induction neuroectodermique plus courts.

Ce nouveau lot d'organoïdes cérébraux a été développé avec une période initiale d'engagement et d'induction neuronale de 6 jours, avant que les EB générés ne soient immergés dans des gouttelettes de Matrigel et transférés dans un milieu de maturation. Comme visible au jour 10, il existe une abondance de plis neuroectodermiques rappelant le développement cortical du cerveau, absents dans le lot précédent d'organoïdes. Leur croissance sur une période de 3 mois a été suivie mais n'a pas montré de différences de taille entre les 2 lignées cellulaires - WT Glio et G34R. Il existe une expression dispersée de KI67 dans tous les organoïdes, et le CD133 a également tendance à être localisé par plaques dans les

deux organoïdes. L'analyse des profils d'expression des gènes révèle une première baisse des marqueurs de pluripotence qui réapparaît à partir d'un mois. Tous les marqueurs neuronaux matures sont fortement exprimés, indiquant un engagement neuroectodermique robuste. Pour élucider une vue approfondie des profils d'expression de nos organoïdes, nous avons effectué une transcriptomique globale comparant les organoïdes cérébraux de souris WT Glio et G34R sur 2 mois. À partir de l'analyse des gènes communs régulés positivement dans les deux lignées cellulaires, nous avons pu confirmer la différenciation neuronale des organoïdes, montrant les signatures du striatum dorsal, de l'amygdale, du bulbe olfactif et du cortex cérébral ainsi que la présence d'astrocytes, de cellules gliales et de neurones. En analysant spécifiquement les gènes régulés positivement dans les organoïdes G34R par rapport aux organoïdes WT Glio, nous trouvons une importance pour la voie P53. Nous savons, grâce aux études de représentation clinique, que les patients atteints de glioblastome présentant un phénotype mutant H3.3 présentent généralement également fréquemment une mutation dans le TP53 [98,100,103,109,110]. De plus, avec les organoïdes générés jusqu'à présent, nous n'avons pas pu remarquer une expansion excessive des tissus qui indiquerait une infiltration et une prolifération tumorale. Cela peut être dû au fait que les mutations H3.3, tout en étant décrites comme des mutations motrices, nécessitent des profils mutationnels supplémentaires afin d'exprimer une morphologie tumorale. Dans cet esprit, nous avons procédé à la création d'un modèle d'inactivation TP53 dans la lignée cellulaire G34R, afin de débloquer un éventuel point restrictif de régulation du cycle cellulaire qui pourrait inhiber les premiers stades de la tumorigenèse.

La lignée cellulaire G34R a été sélectionnée pour être transfectée avec un plasmide P53 KO afin de comparer l'effet tumorigène d'un G34R P53KO par rapport aux organoïdes cérébraux dérivés de G34R non transfectés. La lignée cellulaire G34R ainsi que sa lignée cellulaire isogénique P53KO et la lignée cellulaire transfectée par le plasmide témoin ont été générées pendant une période de 3 mois. Nous commençons à remarquer une différence de taille à partir du 30e jour de culture, les BORG G34R P53KO semblant généralement nettement plus gros que les organoïdes témoins. Au bout de 3 mois de croissance, les BORG P53KO sont 0,5mm plus grands que les BORG témoins. Alors que l'expression globale de Gfap pour les lignées cellulaires transfectées (contrôle P53KO et P53KO) est inférieure à celle de la lignée cellulaire non transfectée, l'expression d'Olig2 est visiblement inférieure pour les organoïdes

G34R P53KO. La comparaison de l'expression de KI67 et CD133 sur les 3 lignées cellulaires ne montre aucune différence, et il existe une surexpression transitoire de Pdgfra observée dans les organoïdes G34R et G34R P53KO à 2 mois, qui est réprimée plus tard.

Outre les aberrations TP53 chez les patients atteints de glioblastome, il existe également d'autres profils mutationnels observés chez les patients atteints de glioblastome, par exemple l'amplification de Myc. Les lignées cellulaires WT Glio et G34R ont toutes deux été transfectées avec une surexpression de Myc ou un plasmide témoin. Les organoïdes cérébraux transfectés par Myc et par contrôle Myc atteignent un diamètre moyen maximum de 2 à 2,5 mm sur la période de 3 mois et sans tissu surexpansé apparent. Nous avons observé une répression plus faible des facteurs de pluripotence dans les organoïdes transfectés par Myc (WT Glio et G34R) et de plus une forte expression du marqueur de cellules souches Oct4 dans les organoïdes G34R Myc, tout au long de sa durée de vie. Les cellules transfectées par Myc qui ont été différenciées en organoïdes cérébraux ne présentent pas de phénotype tumoral prononcé. Il existe des différences dans les niveaux d'expression génique lorsque l'on compare les G34R Myc BORG aux autres conditions, ce qui indique un phénotype tumoral précoce qui ne peut probablement pas surmonter les points normaux de régulation de la croissance cellulaire.

Afin d'imiter le phénotype de l'électroporation de Myc dans les organoïdes, nous avons suivi le protocole de Bian et al, 2018 [119] et électroporé les organoïdes cérébraux à un stade précoce de développement, au lieu d'électroporer les cellules MESC). Les organoïdes ont un profil de croissance comparable sur 3 mois, ne présentant aucune différence. Nous pouvons confirmer l'électroporation des organoïdes avec le plasmide Myc, mais ne sommes pas en mesure de constater une invasion tumorigène. L'utilisation de profils d'expression génique et d'immunocoloration confirme une différenciation neuronale et la présence de niches de cellules souches neurales et de neurones postmitotiques ainsi que la présence de cellules positives à prolifération (KI67).

Ainsi, pour pousser l'invasion tumorale, nous avons combiné les cellules P53KO avec le profil de surexpression de Myc et observé leur croissance. Nous avons étudié 3 lignées cellulaires différentes – WT Glio, WT Glio P53KO et WT Glio P53KO contrôle – qui avaient été électroporées avec une surexpression de Myc ou des plasmides de contrôle Myc au jour 6 de la différenciation organoïde et les avons comparées à des organoïdes cérébraux non électroporés (un total de 9 conditions). Nous trouvons une différence significative dans les

tailles observées à 1 mois de croissance et un graphique statistique de leurs diamètres moyens révèle que les tailles des BORG WT Glio P53KO sont significativement plus grandes que celles de leurs homologues témoins WT Glio et WT Glio P53KO, lorsqu'on les compare aux mêmes Profil d'électroporation Myc. Cela valide en partie le fait que les lignées cellulaires P53KO évoluent vers un état hyperprolifératif qui peut être attribué à leur plus grande taille. De plus, nous sommes en mesure de déduire des données d'expression génique qu'il existe une réémergence de facteurs de pluripotence spécifiques (Nanog, Oct4, Klf4), une neurogenèse affectée (Nes, Tuj), une surexpression de marqueurs OPC (Olig2) et un maintien/régulation positive dans les BORG P53KO sur la période de 1 mois. Les données dérivées de cette expérience démontrent la capacité de la lignée cellulaire P53KO à créer spécifiquement un profil tumorigène.

Pour conclure, j'ai généré un protocole de travail de BORG de signature P53KO qui peuvent démontrer une prolifération significative en 1 mois de culture. A partir de là, il serait bénéfique de refaire notre travail sur des lignées cellulaires mutées H3.3 confirmées, afin de comparer avec leurs lignées témoins isogéniques. Plus précisément, en utilisant P53KO dans des BORG mutants H3.3 confirmés, nous devrions être en mesure de générer un BORG de souris H3.3 tumorigène, qui peut être analysé plus en détail à une résolution cellulaire unique et spatialement en utilisant des outils déjà établis dans l'équipe tels que MULTILAYER [144].

Bibliography

1. Goumans, M.J.; Mummery, C. Functional Analysis of the TGFbeta Receptor/Smad Pathway through Gene Ablation in Mice. *The International Journal of Developmental Biology* **2000**, *44*, 253–265, doi:10.1387/ijdb.10853822.
2. Chen, V.S.; Morrison, J.P.; Southwell, M.F.; Foley, J.F.; Bolon, B.; Elmore, S.A. Histology Atlas of the Developing Prenatal and Postnatal Mouse Central Nervous System, with Emphasis on Prenatal Days E7.5 to E18.5. *Toxicol Pathol* **2017**, *45*, 705–744, doi:10.1177/0192623317728134.
3. Kaufman, M.H.; Bard, J.B.L. *The Anatomical Basis of Mouse Development*; Gulf Professional Publishing, 1999; ISBN 978-0-12-402060-3.
4. Kazanis, I.; Lathia, J.; Moss, L.; ffrench-Constant, C. The Neural Stem Cell Microenvironment. In *StemBook*; Harvard Stem Cell Institute: Cambridge (MA), 2008.
5. Bond, A.M.; Bhalala, O.G.; Kessler, J.A. The Dynamic Role of Bone Morphogenetic Proteins in Neural Stem Cell Fate and Maturation. *Dev Neurobiol* **2012**, *72*, 1068–1084, doi:10.1002/dneu.22022.
6. MacDonald, B.T.; Tamai, K.; He, X. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Dev Cell* **2009**, *17*, 9–26, doi:10.1016/j.devcel.2009.06.016.
7. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell* **2008**, *134*, 921–931, doi:10.1016/j.cell.2008.09.002.
8. Semba, R.D. On the "discovery" of Vitamin A. *Ann Nutr Metab* **2012**, *61*, 192–198, doi:10.1159/000343124.
9. Hopkins, F.G. Feeding Experiments Illustrating the Importance of Accessory Factors in Normal Dietaries. *J Physiol* **1912**, *44*, 425–460, doi:10.1113/jphysiol.1912.sp001524.
10. Wolbach, S.B.; Howe, P.R. TISSUE CHANGES FOLLOWING DEPRIVATION OF FAT-SOLUBLE A VITAMIN. *J Exp Med* **1925**, *42*, 753–777, doi:10.1084/jem.42.6.753.
11. Giguère, V.; Evans, R.M. Chronicle of a Discovery: The Retinoic Acid Receptor. *J Mol Endocrinol* **2022**, *69*, T1–T11, doi:10.1530/JME-22-0117.
12. Blomhoff, R.; Blomhoff, H.K. Overview of Retinoid Metabolism and Function. *J Neurobiol* **2006**, *66*, 606–630, doi:10.1002/neu.20242.
13. Kawaguchi, R.; Yu, J.; Honda, J.; Hu, J.; Whitelegge, J.; Ping, P.; Wiita, P.; Bok, D.; Sun, H. A Membrane Receptor for Retinol Binding Protein Mediates Cellular Uptake of Vitamin A. *Science* **2007**, *315*, 820–825, doi:10.1126/science.1136244.
14. Ang, H.L.; Deltour, L.; Hayamizu, T.F.; Zgombić-Knight, M.; Duester, G. Retinoic Acid Synthesis in Mouse Embryos during Gastrulation and Craniofacial Development Linked to Class IV Alcohol Dehydrogenase Gene Expression. *J Biol Chem* **1996**, *271*, 9526–9534, doi:10.1074/jbc.271.16.9526.
15. Deltour, L.; Foglio, M.H.; Duester, G. Metabolic Deficiencies in Alcohol Dehydrogenase Adh1, Adh3, and Adh4 Null Mutant Mice. Overlapping Roles of Adh1 and Adh4 in Ethanol Clearance and Metabolism of Retinol to Retinoic Acid. *J Biol Chem* **1999**, *274*, 16796–16801, doi:10.1074/jbc.274.24.16796.
16. Molotkov, A.; Fan, X.; Duester, G. Excessive Vitamin A Toxicity in Mice Genetically Deficient in Either Alcohol Dehydrogenase Adh1 or Adh3. *Eur J Biochem* **2002**, *269*, 2607–2612, doi:10.1046/j.1432-1033.2002.02935.x.
17. Parker, R.O.; Crouch, R.K. Retinol Dehydrogenases (RDHs) in the Visual Cycle. *Exp Eye Res* **2010**, *91*, 788–792, doi:10.1016/j.exer.2010.08.013.

18. Niederreither, K.; McCaffery, P.; Dräger, U.C.; Chambon, P.; Dollé, P. Restricted Expression and Retinoic Acid-Induced Downregulation of the Retinaldehyde Dehydrogenase Type 2 (RALDH-2) Gene during Mouse Development. *Mech Dev* **1997**, *62*, 67–78, doi:10.1016/s0925-4773(96)00653-3.
19. Rochette-Egly, C. Retinoic Acid Signaling and Mouse Embryonic Stem Cell Differentiation: Cross Talk between Genomic and Non-Genomic Effects of RA. *Biochim Biophys Acta* **2015**, *1851*, 66–75, doi:10.1016/j.bbalip.2014.04.003.
20. Kam, R.K.T.; Deng, Y.; Chen, Y.; Zhao, H. Retinoic Acid Synthesis and Functions in Early Embryonic Development. *Cell Biosci* **2012**, *2*, 11, doi:10.1186/2045-3701-2-11.
21. Balmer, J.E.; Blomhoff, R. Gene Expression Regulation by Retinoic Acid. *J Lipid Res* **2002**, *43*, 1773–1808, doi:10.1194/jlr.r100015-jlr200.
22. Brtko, J.; Dvorak, Z. Natural and Synthetic Retinoid X Receptor Ligands and Their Role in Selected Nuclear Receptor Action. *Biochimie* **2020**, *179*, 157–168, doi:10.1016/j.biochi.2020.09.027.
23. Thacher, S.; Vasudevan, J.; Chandraratna, R. Therapeutic Applications for Ligands of Retinoid Receptors. *CPD* **2000**, *6*, 25–58, doi:10.2174/1381612003401415.
24. Almeida, N.R.; Conda-Sheridan, M. A Review of the Molecular Design and Biological Activities of RXR Agonists. *Med Res Rev* **2019**, *39*, 1372-1397, doi:10.1002/med.21578.
25. Wang, S.; Wang, Z.; Lin, S.; Zheng, W.; Wang, R.; Jin, S.; Chen, J.; Jin, L.; Li, Y. Revealing a Natural Marine Product as a Novel Agonist for Retinoic Acid Receptors with a Unique Binding Mode and Inhibitory Effects on Cancer Cells. *Biochem J* **2012**, *446*, 79–87, doi:10.1042/BJ20120726.
26. *The Biochemistry of Retinoic Acid Receptors I: Structure, Activation, and Function at the Molecular Level; Asson-Batres*; M.A., R.-E., C., Eds.; Subcellular Biochemistry; Springer: Netherlands, 2014; Vol. 70; ISBN 978-94-017-9049-9.
27. Roy, B.; Taneja, R.; Chambon, P. Synergistic Activation of Retinoic Acid (RA)-Responsive Genes and Induction of Embryonal Carcinoma Cell Differentiation by an RA Receptor α (RARα)-, RARβ-, or RARγ-Selective Ligand in Combination with a Retinoid X Receptor-Specific Ligand. *Molecular and Cellular Biology* **1995**, *15*, 6481–6487, doi:10.1128/MCB.15.12.6481.
28. Taneja, R.; Roy, B.; Plassat, J.L.; Zusi, C.F.; Ostrowski, J.; Reczek, P.R.; Chambon, P. Cell-Type and Promoter-Context Dependent Retinoic Acid Receptor (RAR) Redundancies for RAR Beta 2 and Hoxa-1 Activation in F9 and P19 Cells Can Be Artefactually Generated by Gene Knockouts. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6197–6202.
29. Chiba, H.; Clifford, J.; Metzger, D.; Chambon, P. Specific and Redundant Functions of Retinoid X Receptor/Retinoic Acid Receptor Heterodimers in Differentiation, Proliferation, and Apoptosis of F9 Embryonal Carcinoma Cells. *J. Cell Biol* **1997**, *139*, 735-747, doi:10.1083/jcb.139.3.735.
30. Mendoza-Parra, M.A.; Walia, M.; Sankar, M.; Gronemeyer, H. Dissecting the Retinoid-Induced Differentiation of F9 Embryonal Stem Cells by Integrative Genomics. *Molecular Systems Biology* **2011**, *7*, 538, doi:10.1038/msb.2011.73.
31. Mendoza-Parra, M.-A.; Malysheva, V.; Mohamed Saleem, M.A.; Lieb, M.; Godel, A.; Gronemeyer, H. Reconstructed Cell Fate-Regulatory Programs in Stem Cells Reveal Hierarchies and Key Factors of Neurogenesis. *Genome Res.* **2016**, *26*, 1505–1519, doi:10.1101/gr.208926.116.
32. Koshy, A.; Mathieux, E.; Stüder, F.; Bramoulle, A.; Lieb, M.; Colombo, B.M.; Gronemeyer, H.; Mendoza-Parra, M.A. Synergistic Activation of RARβ and RARγ Nuclear Receptors

Restores Cell Specialization during Stem Cell Differentiation by Hijacking RARα-Controlled Programs. *Life Sci. Alliance* **2023**, *6, e202201627*, doi:10.26508/lsa.202201627.
33. Dollé, P.; Izpisúa-Belmonte, J.C.; Falkenstein, H.; Renucci, A.; Duboule, D. Coordinate Expression of the Murine Hox-5 Complex Homoeobox-Containing Genes during Limb Pattern Formation. *Nature* **1989**, *342*, 767–772, doi:10.1038/342767a0.
34. Ruberte, E.; Dolle, P.; Krust, A.; Zelent, A.; Morriss-Kay, G.; Chambon, P. Specific Spatial and Temporal Distribution of Retinoic Acid Receptor Gamma Transcripts during Mouse Embryogenesis. *Development* **1990**, *108*, 213–222, doi:10.1242/dev.108.2.213.
35. Ruberte, E.; Dolle, P.; Chambon, P.; Morriss-Kay, G. Retinoic Acid Receptors and Cellular Retinoid Binding Proteins. II. Their Differential Pattern of Transcription during Early Morphogenesis in Mouse Embryos. *Development* **1991**, *111*, 45–60, doi:10.1242/dev.111.1.45.
36. Hale, L.A.; Tallafuss, A.; Yan, Y.-L.; Dudley, L.; Eisen, J.S.; Postlethwait, J.H. Characterization of the Retinoic Acid Receptor Genes Raraa, Rarab and Rarg during Zebrafish Development. *Gene Expr Patterns* **2006**, *6*, 546–555, doi:10.1016/j.modgep.2005.10.007.
37. Tallafuss, A.; Hale, L.A.; Yan, Y.-L.; Dudley, L.; Eisen, J.S.; Postlethwait, J.H. Characterization of Retinoid-X Receptor Genes Rxra, Rxrba, Rxrbb and Rxrg during Zebrafish Development. *Gene Expr Patterns* **2006**, *6*, 556–565, doi:10.1016/j.modgep.2005.10.005.
38. Waxman, J.S.; Yelon, D. Comparison of the Expression Patterns of Newly Identified Zebrafish Retinoic Acid and Retinoid X Receptors. *Dev Dyn* **2007**, *236*, 587–595, doi:10.1002/dvdy.21049.
39. Blumberg, B.; Mangelsdorf, D.J.; Dyck, J.A.; Bittner, D.A.; Evans, R.M.; De Robertis, E.M. Multiple Retinoid-Responsive Receptors in a Single Cell: Families of Retinoid "X" Receptors and Retinoic Acid Receptors in the Xenopus Egg. *Proc Natl Acad Sci U S A* **1992**, *89*, 2321–2325, doi:10.1073/pnas.89.6.2321.
40. Ellinger-Ziegelbauer, H.; Dreyer, C. A Retinoic Acid Receptor Expressed in the Early Development of Xenopus Laevis. *Genes Dev* **1991**, *5*, 94–104, doi:10.1101/gad.5.1.94.
41. Koide, T.; Downes, M.; Chandraratna, R.A.; Blumberg, B.; Umesono, K. Active Repression of RAR Signaling Is Required for Head Formation. *Genes Dev* **2001**, *15*, 2111–2121, doi:10.1101/gad.908801.
42. Shiotsugu, J.; Katsuyama, Y.; Arima, K.; Baxter, A.; Koide, T.; Song, J.; Chandraratna, R.A.S.; Blumberg, B. Multiple Points of Interaction between Retinoic Acid and FGF Signaling during Embryonic Axis Formation. *Development* **2004**, *131*, 2653–2667, doi:10.1242/dev.01129.
43. Biga, L.M.; Bronson, S.; Dawson, S.; Harwell, A.; Hopkins, R.; Kaufmann, J.; LeMaster, M.; Matern, P.; Morrison-Graham, K.; Oja, K.; et al. *Anatomy & Physiology*; OpenStax/Oregon State University, 2019;
44. Niederreither, K.; Subbarayan, V.; Dollé, P.; Chambon, P. Embryonic Retinoic Acid Synthesis Is Essential for Early Mouse Post-Implantation Development. *Nat Genet* **1999**, *21*, 444–448, doi:10.1038/7788.
45. Uehara, M.; Yashiro, K.; Mamiya, S.; Nishino, J.; Chambon, P.; Dolle, P.; Sakai, Y. CYP26A1 and CYP26C1 Cooperatively Regulate Anterior-Posterior Patterning of the Developing Brain and the Production of Migratory Cranial Neural Crest Cells in the Mouse. *Dev Biol* **2007**, *302*, 399–411, doi:10.1016/j.ydbio.2006.09.045.

46. Hernandez, R.E.; Putzke, A.P.; Myers, J.P.; Margaretha, L.; Moens, C.B. Cyp26 Enzymes Generate the Retinoic Acid Response Pattern Necessary for Hindbrain Development. *Development* **2007**, *134*, 177–187, doi:10.1242/dev.02706.
47. Ribes, V.; Otto, D.M.E.; Dickmann, L.; Schmidt, K.; Schuhbaur, B.; Henderson, C.; Blomhoff, R.; Wolf, C.R.; Tickle, C.; Dollé, P. Rescue of Cytochrome P450 Oxidoreductase (Por) Mouse Mutants Reveals Functions in Vasculogenesis, Brain and Limb Patterning Linked to Retinoic Acid Homeostasis. *Dev Biol* **2007**, *303*, 66–81, doi:10.1016/j.ydbio.2006.10.032.
48. Novitch, B.G.; Wichterle, H.; Jessell, T.M.; Sockanathan, S. A Requirement for Retinoic Acid-Mediated Transcriptional Activation in Ventral Neural Patterning and Motor Neuron Specification. *Neuron* **2003**, *40*, 81–95, doi:10.1016/j.neuron.2003.08.006.
49. Wilson, L.; Maden, M. The Mechanisms of Dorsoventral Patterning in the Vertebrate Neural Tube. *Dev Biol* **2005**, *282*, 1–13, doi:10.1016/j.ydbio.2005.02.027.
50. Diez del Corral, R.; Storey, K.G. Opposing FGF and Retinoid Pathways: A Signalling Switch That Controls Differentiation and Patterning Onset in the Extending Vertebrate Body Axis. *Bioessays* **2004**, *26*, 857–869, doi:10.1002/bies.20080.
51. Sockanathan, S.; Jessell, T.M. Motor Neuron-Derived Retinoid Signaling Specifies the Subtype Identity of Spinal Motor Neurons. *Cell* **1998**, *94*, 503–514, doi:10.1016/s0092-8674(00)81591-3.
52. Sockanathan, S.; Perlmann, T.; Jessell, T.M. Retinoid Receptor Signaling in Postmitotic Motor Neurons Regulates Rostrocaudal Positional Identity and Axonal Projection Pattern. *Neuron* **2003**, *40*, 97–111, doi:10.1016/s0896-6273(03)00532-4.
53. Holzschuh, J.; Barrallo-Gimeno, A.; Ettl, A.-K.; Durr, K.; Knapik, E.W.; Driever, W. Noradrenergic Neurons in the Zebrafish Hindbrain Are Induced by Retinoic Acid and Require Tfap2a for Expression of the Neurotransmitter Phenotype. *Development* **2003**, *130*, 5741–5754, doi:10.1242/dev.00816.
54. Chatzi, C.; Brade, T.; Duester, G. Retinoic Acid Functions as a Key GABAergic Differentiation Signal in the Basal Ganglia. *PLoS Biol* **2011**, *9*, e1000609, doi:10.1371/journal.pbio.1000609.
55. Jacob, J.; Kong, J.; Moore, S.; Milton, C.; Sasai, N.; Gonzalez-Quevedo, R.; Terriente, J.; Imayoshi, I.; Kageyama, R.; Wilkinson, D.G.; et al. Retinoid Acid Specifies Neuronal Identity through Graded Expression of Ascl1. *Curr Biol* **2013**, *23*, 412–418, doi:10.1016/j.cub.2013.01.046.
56. Rataj-Baniowska, M.; Niewiadomska-Cimicka, A.; Paschaki, M.; Szyszka-Niagolov, M.; Carramolino, L.; Torres, M.; Dollé, P.; Krężel, W. Retinoic Acid Receptor β Controls Development of Striatonigral Projection Neurons through FGF-Dependent and Meis1-Dependent Mechanisms. *J Neurosci* **2015**, *35*, 14467–14475, doi:10.1523/JNEUROSCI.1278-15.2015.
57. Lobjois, V.; Bel-Vialar, S.; Trousse, F.; Pituello, F. Forcing Neural Progenitor Cells to Cycle Is Insufficient to Alter Cell-Fate Decision and Timing of Neuronal Differentiation in the Spinal Cord. *Neural Development* **2008**, *3*, 4, doi:10.1186/1749-8104-3-4.
58. Dasen, J.S.; De Camilli, A.; Wang, B.; Tucker, P.W.; Jessell, T.M. Hox Repertoires for Motor Neuron Diversity and Connectivity Gated by a Single Accessory Factor, FoxP1. *Cell* **2008**, *134*, 304–316, doi:10.1016/j.cell.2008.06.019.
59. Sagner, A.; Gaber, Z.B.; Delile, J.; Kong, J.H.; Rousso, D.L.; Pearson, C.A.; Weicksel, S.E.; Melchionda, M.; Mousavy Gharavy, S.N.; Briscoe, J.; et al. Olig2 and Hes Regulatory

Dynamics during Motor Neuron Differentiation Revealed by Single Cell Transcriptomics. *PLoS Biol* **2018**, *16*, e2003127, doi:10.1371/journal.pbio.2003127.
60. Carcagno, A.L.; Di Bella, D.J.; Goulding, M.; Guillemot, F.; Lanuza, G.M. Neurogenin3 Restricts Serotonergic Neuron Differentiation to the Hindbrain. *J Neurosci* **2014**, *34*, 15223–15233, doi:10.1523/JNEUROSCI.3403-14.2014.
61. Casarosa, S.; Fode, C.; Guillemot, F. Mash1 Regulates Neurogenesis in the Ventral Telencephalon. *Development* **1999**, *126*, 525–534, doi:10.1242/dev.126.3.525.
62. Murakami, A.; Grinberg, D.; Thurlow, J.; Dickson, C. Identification of Positive and Negative Regulatory Elements Involved in the Retinoic Acid/cAMP Induction of Fgf-3 Transcription in F9 Cells. *Nucleic Acids Res* **1993**, *21*, 5351–5359.
63. Ericson, J.; Morton, S.; Kawakami, A.; Roelink, H.; Jessell, T.M. Two Critical Periods of Sonic Hedgehog Signaling Required for the Specification of Motor Neuron Identity. *Cell* **1996**, *87*, 661–673, doi:10.1016/S0092-8674(00)81386-0.
64. Ericson, J.; Briscoe, J.; Rashbass, P.; van Heyningen, V.; Jessell, T.M. Graded Sonic Hedgehog Signaling and the Specification of Cell Fate in the Ventral Neural Tube. *Cold Spring Harb Symp Quant Biol* **1997**, *62*, 451–466.
65. Okada, Y.; Shimazaki, T.; Sobue, G.; Okano, H. Retinoic-Acid-Concentration-Dependent Acquisition of Neural Cell Identity during in Vitro Differentiation of Mouse Embryonic Stem Cells. *Developmental Biology* **2004**, *275*, 124–142, doi:10.1016/j.ydbio.2004.07.038.
66. Jessell, T.M. Neuronal Specification in the Spinal Cord: Inductive Signals and Transcriptional Codes. *Nat Rev Genet* **2000**, *1*, 20–29, doi:10.1038/35049541.
67. Pierani, A.; Moran-Rivard, L.; Sunshine, M.J.; Littman, D.R.; Goulding, M.; Jessell, T.M. Control of Interneuron Fate in the Developing Spinal Cord by the Progenitor Homeodomain Protein Dbx1. *Neuron* **2001**, *29*, 367–384, doi:10.1016/S0896-6273(01)00212-4.
68. Brown, C.R.; Butts, J.C.; McCreedy, D.A.; Sakiyama-Elbert, S.E. Generation of V2a Interneurons from Mouse Embryonic Stem Cells. *Stem Cells Dev* **2014**, *23*, 1765–1776, doi:10.1089/scd.2013.0628.
69. Wichterle, H.; Lieberam, I.; Porter, J.A.; Jessell, T.M. Directed Differentiation of Embryonic Stem Cells into Motor Neurons. *Cell* **2002**, *110*, 385–397, doi:10.1016/S0092-8674(02)00835-8.
70. Salero, E.; Hatten, M.E. Differentiation of ES Cells into Cerebellar Neurons. *Proceedings of the National Academy of Sciences* **2007**, *104*, 2997–3002, doi:10.1073/pnas.0610879104.
71. Gaspard, N.; Bouschet, T.; Herpoel, A.; Naeije, G.; van den Ameele, J.; Vanderhaeghen, P. Generation of Cortical Neurons from Mouse Embryonic Stem Cells. *Nat Protoc* **2009**, *4*, 1454–1463, doi:10.1038/nprot.2009.157.
72. Li, X.-J.; Du, Z.-W.; Zarnowska, E.D.; Pankratz, M.; Hansen, L.O.; Pearce, R.A.; Zhang, S.-C. Specification of Motoneurons from Human Embryonic Stem Cells. *Nat Biotechnol* **2005**, *23*, 215–221, doi:10.1038/nbt1063.
73. Amoroso, M.W.; Croft, G.F.; Williams, D.J.; O'Keeffe, S.; Carrasco, M.A.; Davis, A.R.; Roybon, L.; Oakley, D.H.; Maniatis, T.; Henderson, C.E.; et al. Accelerated High-Yield Generation of Limb-Innervating Motor Neurons from Human Stem Cells. *J Neurosci* **2013**, *33*, 574–586, doi:10.1523/JNEUROSCI.0906-12.2013.
74. Tonge, P.D.; Andrews, P.W. Retinoic Acid Directs Neuronal Differentiation of Human Pluripotent Stem Cell Lines in a Non-Cell-Autonomous Manner. *Differentiation* **2010**, *80*, 20–30, doi:10.1016/j.diff.2010.04.001.

75. McBurney, M.W.; Jones-Villeneuve, E.M.; Edwards, M.K.; Anderson, P.J. Control of Muscle and Neuronal Differentiation in a Cultured Embryonal Carcinoma Cell Line. *Nature* **1982**, *299*, 165–167, doi:10.1038/299165a0.
76. Edwards, M.K.; McBurney, M.W. The Concentration of Retinoic Acid Determines the Differentiated Cell Types Formed by a Teratocarcinoma Cell Line. *Dev Biol* **1983**, *98*, 187–191, doi:10.1016/0012-1606(83)90348-2.
77. Smith, S.C.; Reuhl, K.R.; Craig, J.; McBurney, M.W. The Role of Aggregation in Embryonal Carcinoma Cell Differentiation. *J Cell Physiol* **1987**, *131*, 74–84, doi:10.1002/jcp.1041310112.
78. Monzo, H.J.; Park, T.I.H.; Montgomery, J.M.; Faull, R.L.M.; Dragunow, M.; Curtis, M.A. A Method for Generating High-Yield Enriched Neuronal Cultures from P19 Embryonal Carcinoma Cells. *J Neurosci Methods* **2012**, *204*, 87–103, doi:10.1016/j.jneumeth.2011.11.008.
79. Podleśny-Drabiniok, A.; Sobska, J.; de Lera, A.R.; Gołembiowska, K.; Kamińska, K.; Dollé, P.; Cebrat, M.; Krężel, W. Distinct Retinoic Acid Receptor (RAR) Isotypes Control Differentiation of Embryonal Carcinoma Cells to Dopaminergic or Striatopallidal Medium Spiny Neurons. *Sci Rep* **2017**, *7*, 13671, doi:10.1038/s41598-017-13826-x.
80. Taneja, R.; Bouillet, P.; Boylan, J.F.; Gaub, M.P.; Roy, B.; Gudas, L.J.; Chambon, P. Reexpression of Retinoic Acid Receptor (RAR) Gamma or Overexpression of RAR Alpha or RAR Beta in RAR Gamma-Null F9 Cells Reveals a Partial Functional Redundancy between the Three RAR Types. *Proc Natl Acad Sci U S A* **1995**, *92*, 7854–7858.
81. Brown, G. Retinoic Acid Receptor Regulation of Decision-Making for Cell Differentiation. *Front Cell Dev Biol* **2023**, *11*, 1182204, doi:10.3389/fcell.2023.1182204.
82. Lohnes, D.; Kastner, P.; Dierich, A.; Mark, M.; LeMeur, M.; Chambon, P. Function of Retinoic Acid Receptor γ in the Mouse. *Cell* **1993**, *73*, 643–658, doi:10.1016/0092-8674(93)90246-M.
83. Lancaster, M.A.; Renner, M.; Martin, C.-A.; Wenzel, D.; Bicknell, L.S.; Hurles, M.E.; Homfray, T.; Penninger, J.M.; Jackson, A.P.; Knoblich, J.A. Cerebral Organoids Model Human Brain Development and Microcephaly. *Nature* **2013**, *501*, 373–379, doi:10.1038/nature12517.
84. Sloan, S.A.; Andersen, J.; Paşca, A.M.; Birey, F.; Paşca, S.P. Generation and Assembly of Human Brain Region-Specific Three-Dimensional Cultures. *Nat Protoc* **2018**, *13*, 2062–2085, doi:10.1038/s41596-018-0032-7.
85. Kanton, S.; Boyle, M.J.; He, Z.; Santel, M.; Weigert, A.; Sanchís-Calleja, F.; Guijarro, P.; Sidow, L.; Fleck, J.S.; Han, D.; et al. Organoid Single-Cell Genomic Atlas Uncovers Human-Specific Features of Brain Development. *Nature* **2019**, *574*, 418–422, doi:10.1038/s41586-019-1654-9.
86. Li, C.; Fleck, J.S.; Martins-Costa, C.; Burkard, T.R.; Themann, J.; Stuempflen, M.; Peer, A.M.; Vertesy, Á.; Littleboy, J.B.; Esk, C.; et al. Single-Cell Brain Organoid Screening Identifies Developmental Defects in Autism. *Nature* **2023**, *621*, 373–380, doi:10.1038/s41586-023-06473-y.
87. Gabriel, E.; Albanna, W.; Pasquini, G.; Ramani, A.; Josipovic, N.; Mariappan, A.; Schinzel, F.; Karch, C.M.; Bao, G.; Gottardo, M.; et al. Human Brain Organoids Assemble Functionally Integrated Bilateral Optic Vesicles. *Cell Stem Cell* **2021**, *28*, 1740-1757.e8, doi:10.1016/j.stem.2021.07.010.
88. Lozachmeur, G.; Bramoulle, A.; Aubert, A.; Stüder, F.; Moehlin, J.; Madrange, L.; Yates, F.; Deslys, J.-P.; Mendoza-Parra, M.A. Three-Dimensional Molecular Cartography of Human

Cerebral Organoids Revealed by Double-Barcoded Spatial Transcriptomics. *Cell Reports Methods* **2023**, *3*, 100573, doi:10.1016/j.crmeth.2023.100573.
89. Uzquiano, A.; Kedaigle, A.J.; Pigoni, M.; Paulsen, B.; Adiconis, X.; Kim, K.; Faits, T.; Nagaraja, S.; Antón-Bolaños, N.; Gerhardinger, C.; et al. Proper Acquisition of Cell Class Identity in Organoids Allows Definition of Fate Specification Programs of the Human Cerebral Cortex. *Cell* **2022**, *185*, 3770-3788.e27, doi:10.1016/j.cell.2022.09.010.
90. He, Z.; Maynard, A.; Jain, A.; Gerber, T.; Petri, R.; Lin, H.-C.; Santel, M.; Ly, K.; Dupré, J.-S.; Sidow, L.; et al. Lineage Recording in Human Cerebral Organoids. *Nat Methods* **2022**, *19*, 90–99, doi:10.1038/s41592-021-01344-8.
91. Oldak, B.; Wildschutz, E.; Bondarenko, V.; Comar, M.-Y.; Zhao, C.; Aguilera-Castrejon, A.; Tarazi, S.; Viukov, S.; Pham, T.X.A.; Ashouokhi, S.; et al. Complete Human Day 14 Post-Implantation Embryo Models from Naïve ES Cells. *Nature* **2023**, 1–3, doi:10.1038/s41586-023-06604-5.
92. Weatherbee, B.A.T.; Gantner, C.W.; Iwamoto-Stohl, L.K.; Daza, R.M.; Hamazaki, N.; Shendure, J.; Zernicka-Goetz, M. Pluripotent Stem Cell-Derived Model of the Post-Implantation Human Embryo. *Nature* **2023**, 1–10, doi:10.1038/s41586-023-06368-y.
93. Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.K.; Pfister, S.M.; Reifenberger, G.; et al. The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro Oncol* **2021**, *23*, 1231–1251, doi:10.1093/neuonc/noab106.
94. Wirsching, H.-G.; Galanis, E.; Weller, M. Chapter 23 - Glioblastoma. In *Handbook of Clinical Neurology*; Berger, M.S., Weller, M., Eds.; Gliomas; Elsevier, 2016; Vol. 134, pp. 381–397.
95. Das, K.K.; Kumar, R. Pediatric Glioblastoma. In *Glioblastoma*; De Vleeschouwer, S., Ed.; Codon Publications: Brisbane (AU), 2017 ISBN 978-0-9944381-2-6.
96. Ostrom, Q.T.; Gittleman, H.; Fulop, J.; Liu, M.; Blanda, R.; Kromer, C.; Wolinsky, Y.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol* **2015**, *17*, iv1–iv62, doi:10.1093/neuonc/nov189.
97. Singla, A.K.; Madan, R.; Gupta, K.; Goyal, S.; Kumar, N.; Sahoo, S.K.; Uppal, D.K.; Ahuja, C.K. Clinical Behaviour and Outcome in Pediatric Glioblastoma: Current Scenario. *Radiat Oncol J* **2021**, *39*, 72–77, doi:10.3857/roj.2020.00591.
98. Johnson, A.; Severson, E.; Gay, L.; Vergilio, J.-A.; Elvin, J.; Suh, J.; Daniel, S.; Covert, M.; Frampton, G.M.; Hsu, S.; et al. Comprehensive Genomic Profiling of 282 Pediatric Low- and High-Grade Gliomas Reveals Genomic Drivers, Tumor Mutational Burden, and Hypermutation Signatures. *The Oncologist* **2017**, *22*, 1478–1490, doi:10.1634/theoncologist.2017-0242.
99. Gielen, G.H.; Gessi, M.; Hammes, J.; Kramm, C.M.; Waha, A.; Pietsch, T. H3F3A K27M Mutation in Pediatric CNS Tumors: A Marker for Diffuse High-Grade Astrocytomas. *American Journal of Clinical Pathology* **2013**, *139*, 345–349, doi:10.1309/AJCPABOHBC33FVMO.
100. Hoffman, M.; Gillmor, A.H.; Kunz, D.J.; Johnston, M.; Nikolic, A.; Narta, K.; Zarrei, M.; King, J.; Ellestad, K.; Dang, N.H.; et al. Intratumoral Genetic and Functional Heterogeneity in Pediatric Glioblastoma. *Cancer Res* **2019**, *79*, 2111–2123, doi:10.1158/0008-5472.CAN-18-3441.
101. Waterborg, J.H. Evolution of Histone H3: Emergence of Variants and Conservation of Post-Translational Modification Sites [1] This Article Is Part of Special Issue Entitled

Asilomar Chromatin and Has Undergone the Journal's Usual Peer Review Process. *Biochem. Cell Biol.* **2012**, *90*, 79–95, doi:10.1139/o11-036.

102. Lowe, B.R.; Yadav, R.K.; Henry, R.A.; Schreiner, P.; Matsuda, A.; Fernandez, A.G.; Finkelstein, D.; Campbell, M.; Kallappagoudar, S.; Jablonowski, C.M.; et al. Surprising Phenotypic Diversity of Cancer-Associated Mutations of Gly 34 in the Histone H3 Tail 2020, 2020.12.10.419184.
103. Schwartzentruber, J.; Korshunov, A.; Liu, X.-Y.; Jones, D.T.W.; Pfaff, E.; Jacob, K.; Sturm, D.; Fontebasso, A.M.; Quang, D.-A.K.; Tönjes, M.; et al. Driver Mutations in Histone H3.3 and Chromatin Remodelling Genes in Paediatric Glioblastoma. *Nature* **2012**, *482*, 226–231, doi:10.1038/nature10833.
104. Korshunov, A.; Ryzhova, M.; Hovestadt, V.; Bender, S.; Sturm, D.; Capper, D.; Meyer, J.; Schrimpf, D.; Kool, M.; Northcott, P.A.; et al. Integrated Analysis of Pediatric Glioblastoma Reveals a Subset of Biologically Favorable Tumors with Associated Molecular Prognostic Markers. *Acta Neuropathol* **2015**, *129*, 669–678, doi:10.1007/s00401-015-1405-4.
105. Pathak, P.; Jha, P.; Purkait, S.; Sharma, V.; Suri, V.; Sharma, M.C.; Faruq, M.; Suri, A.; Sarkar, C. Altered Global Histone-Trimethylation Code and H3F3A-ATRX Mutation in Pediatric GBM. *J Neurooncol* **2015**, *121*, 489–497, doi:10.1007/s11060-014-1675-z.
106. Mackay, A.; Burford, A.; Carvalho, D.; Izquierdo, E.; Fazal-Salom, J.; Taylor, K.R.; Bjerke, L.; Clarke, M.; Vinci, M.; Nandhabalan, M.; et al. Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **2017**, *32*, 520-537.e5, doi:10.1016/j.ccell.2017.08.017.
107. Harutyunyan, A.S.; Krug, B.; Chen, H.; Papillon-Cavanagh, S.; Zeinieh, M.; De Jay, N.; Deshmukh, S.; Chen, C.C.L.; Belle, J.; Mikael, L.G.; et al. H3K27M Induces Defective Chromatin Spread of PRC2-Mediated Repressive H3K27me2/Me3 and Is Essential for Glioma Tumorigenesis. *Nat Commun* **2019**, *10*, 1262, doi:10.1038/s41467-019-09140-x.
108. Shi, L.; Shi, J.; Shi, X.; Li, W.; Wen, H. Histone H3.3 G34 Mutations Alter Histone H3K36 and H3K27 Methylation In Cis. *Journal of Molecular Biology* **2018**, *430*, 1562–1565, doi:10.1016/j.jmb.2018.04.014.
109. Lewis, P.W.; Müller, M.M.; Koletsky, M.S.; Cordero, F.; Lin, S.; Banaszynski, L.A.; Garcia, B.A.; Muir, T.W.; Becher, O.J.; Allis, C.D. Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science* **2013**, *340*, 857–861, doi:10.1126/science.1232245.
110. Lulla, R.R.; Saratsis, A.M.; Hashizume, R. Mutations in Chromatin Machinery and Pediatric High-Grade Glioma. *Science Advances* **2016**, *2*, e1501354, doi:10.1126/sciadv.1501354.
111. Fang, H.-T.; EL Farran, C.A.; Xing, Q.R.; Zhang, L.-F.; Li, H.; Lim, B.; Loh, Y.-H. Global H3.3 Dynamic Deposition Defines Its Bimodal Role in Cell Fate Transition. *Nat Commun* **2018**, *9*, 1537, doi:10.1038/s41467-018-03904-7.
112. Sturm, D.; Witt, H.; Hovestadt, V.; Khuong-Quang, D.-A.; Jones, D.T.W.; Konermann, C.; Pfaff, E.; Tönjes, M.; Sill, M.; Bender, S.; et al. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* **2012**, *22*, 425–437, doi:10.1016/j.ccr.2012.08.024.
113. Korshunov, A.; Schrimpf, D.; Ryzhova, M.; Sturm, D.; Chavez, L.; Hovestadt, V.; Sharma, T.; Habel, A.; Burford, A.; Jones, C.; et al. H3-/IDH-Wild Type Pediatric Glioblastoma Is Comprised of Molecularly and Prognostically Distinct Subtypes with Associated Oncogenic Drivers. *Acta Neuropathol* **2017**, *134*, 507–516, doi:10.1007/s00401-017-1710-1.

114. Ashok, A.; Choudhury, D.; Fang, Y.; Hunziker, W. Towards Manufacturing of Human Organoids. *Biotechnology Advances* **2020**, *39*, 107460, doi:10.1016/j.biotechadv.2019.107460.
115. Chen, H.I.; Song, H.; Ming, G. Applications of Human Brain Organoids to Clinical Problems: Clinical Translation Of Brain Organoids. *Dev. Dyn.* **2019**, *248*, 53–64, doi:10.1002/dvdy.24662.
116. Quadrato, G.; Nguyen, T.; Macosko, E.Z.; Sherwood, J.L.; Min Yang, S.; Berger, D.R.; Maria, N.; Scholvin, J.; Goldman, M.; Kinney, J.P.; et al. Cell Diversity and Network Dynamics in Photosensitive Human Brain Organoids. *Nature* **2017**, *545*, 48–53, doi:10.1038/nature22047.
117. Zhu, Y.; Wang, L.; Yin, F.; Yu, Y.; Wang, Y.; Liu, H.; Wang, H.; Sun, N.; Liu, H.; Qin, J. A Hollow Fiber System for Simple Generation of Human Brain Organoids. *Integr. Biol.* **2017**, *9*, 774–781, doi:10.1039/C7IB00080D.
118. Wörsdörfer, P.; Dalda, N.; Kern, A.; Krüger, S.; Wagner, N.; Kwok, C.K.; Henke, E.; Ergün, S. Generation of Complex Human Organoid Models Including Vascular Networks by Incorporation of Mesodermal Progenitor Cells. *Sci Rep* **2019**, *9*, 15663, doi:10.1038/s41598-019-52204-7.
119. Cakir, B.; Xiang, Y.; Tanaka, Y.; Kural, M.H.; Parent, M.; Kang, Y.-J.; Chapeton, K.; Patterson, B.; Yuan, Y.; He, C.-S.; et al. Engineering of Human Brain Organoids with a Functional Vascular-like System. *Nat Methods* **2019**, *16*, 1169–1175, doi:10.1038/s41592-019-0586-5.
120. Bian, S.; Repic, M.; Guo, Z.; Kavirayani, A.; Burkard, T.; Bagley, J.A.; Krauditsch, C.; Knoblich, J.A. Genetically Engineered Cerebral Organoids Model Brain Tumour Formation. *Nat Methods* **2018**, *15*, 631–639, doi:10.1038/s41592-018-0070-7.
121. Ogawa, J.; Pao, G.M.; Shokhirev, M.N.; Verma, I.M. Glioblastoma Model Using Human Cerebral Organoids. *Cell Rep* **2018**, *23*, 1220–1229, doi:10.1016/j.celrep.2018.03.105.
122. Bao, S.; Wu, Q.; McLendon, R.E.; Hao, Y.; Shi, Q.; Hjelmeland, A.B.; Dewhirst, M.W.; Bigner, D.D.; Rich, J.N. Glioma Stem Cells Promote Radioresistance by Preferential Activation of the DNA Damage Response. *Nature* **2006**, *444*, 756–760, doi:10.1038/nature05236.
123. Auffinger, B.; Spencer, D.; Pytel, P.; Ahmed, A.U.; Lesniak, M.S. The Role of Glioma Stem Cells in Chemotherapy Resistance and Glioblastoma Multiforme Recurrence. *Expert Rev Neurother* **2015**, *15*, 741–752, doi:10.1586/14737175.2015.1051968.
124. Eckerdt, F.; Platanias, L.C. Emerging Role of Glioma Stem Cells in Mechanisms of Therapy Resistance. *Cancers* **2023**, *15*, 3458, doi:10.3390/cancers15133458.
125. Linkous, A.; Balamatsias, D.; Snuderl, M.; Edwards, L.; Miyaguchi, K.; Milner, T.; Reich, B.; Cohen-Gould, L.; Storaska, A.; Nakayama, Y.; et al. Modeling Patient-Derived Glioblastoma with Cerebral Organoids. *Cell Rep* **2019**, *26*, 3203-3211.e5, doi:10.1016/j.celrep.2019.02.063.
126. Jacob, F.; Salinas, R.D.; Zhang, D.Y.; Nguyen, P.T.T.; Schnoll, J.G.; Wong, S.Z.H.; Thokala, R.; Sheikh, S.; Saxena, D.; Prokop, S.; et al. A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-Tumoral Heterogeneity. *Cell* **2020**, *180*, 188-204.e22, doi:10.1016/j.cell.2019.11.036.
127. Marshall, J.J.; Mason, J.O. Mouse vs Man: Organoid Models of Brain Development & Disease. *Brain Research* **2019**, *1724*, 146427, doi:10.1016/j.brainres.2019.146427.
128. Eiraku, M.; Sasai, Y. Mouse Embryonic Stem Cell Culture for Generation of Three-Dimensional Retinal and Cortical Tissues. *Nat Protoc* **2012**, *7*, 69–79, doi:10.1038/nprot.2011.429.

129. Eiraku, M.; Takata, N.; Ishibashi, H.; Kawada, M.; Sakakura, E.; Okuda, S.; Sekiguchi, K.; Adachi, T.; Sasai, Y. Self-Organizing Optic-Cup Morphogenesis in Three-Dimensional Culture. *Nature* **2011**, *472*, 51–56, doi:10.1038/nature09941.
130. Ciarpella, F.; Zamfir, R.G.; Campanelli, A.; Ren, E.; Pedrotti, G.; Bottani, E.; Borioli, A.; Caron, D.; Di Chio, M.; Dolci, S.; et al. Murine Cerebral Organoids Develop Network of Functional Neurons and Hippocampal Brain Region Identity. *iScience* **2021**, *24*, 103438, doi:10.1016/j.isci.2021.103438.
131. Million, K.; Tournier, F.; Houcine, O.; Ancian, P.; Reichert, U.; Marano, F. Effects of Retinoic Acid Receptor–Selective Agonists on Human Nasal Epithelial Cell Differentiation. *Am J Respir Cell Mol Biol* **2001**, *25*, 744–750, doi:10.1165/ajrcmb.25.6.4549.
132. Koshy, A.; Mathieux, E.; Stüder, F.; Bramoulle, A.; Lieb, M.; Colombo, B.M.; Gronemeyer, H.; Mendoza-Parra, M.A. Synergistic Activation of RARβ and RARγ Nuclear Receptors Restores Cell Specialization during Stem Cell Differentiation by Hijacking RARα-Controlled Programs. *Life Science Alliance* **2023**, *6*, doi:10.26508/lsa.202201627.
133. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool. *BMC Bioinformatics* **2013**, *14*, 128, doi:10.1186/1471-2105-14-128.
134. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update. *Nucleic Acids Res* **2016**, *44*, W90-97, doi:10.1093/nar/gkw377.
135. Xie, Z.; Bailey, A.; Kuleshov, M.V.; Clarke, D.J.B.; Evangelista, J.E.; Jenkins, S.L.; Lachmann, A.; Wojciechowicz, M.L.; Kropiwnicki, E.; Jagodnik, K.M.; et al. Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **2021**, *1*, e90, doi:10.1002/cpz1.90.
136. Mitsui, K.; Nakajima, D.; Ohara, O.; Nakayama, M. Mammalian Fat3: A Large Protein That Contains Multiple Cadherin and EGF-like Motifs. *Biochemical and Biophysical Research Communications* **2002**, *290*, 1260–1266, doi:10.1006/bbrc.2002.6338.
137. Okajima, T.; Gu, Y.; Teruya, R.; Yano, S.; Taketomi, T.; Sato, B.; Chiba, T.; Tsuruta, F. Atypical Cadherin FAT3 Is a Novel Mediator for Morphological Changes of Microglia. *eNeuro* **2020**, *7*, ENEURO.0056-20.2020, doi:10.1523/ENEURO.0056-20.2020.
138. Castiglioni, V.; Faedo, A.; Onorati, M.; Bocchi, V.D.; Li, Z.; Iennaco, R.; Vuono, R.; Bulfamante, G.P.; Muzio, L.; Martino, G.; et al. Dynamic and Cell-Specific DACH1 Expression in Human Neocortical and Striatal Development. *Cereb Cortex* **2019**, *29*, 2115–2124, doi:10.1093/cercor/bhy092.
139. Davis, R.J.; Shen, W.; Sandler, Y.I.; Amoui, M.; Purcell, P.; Maas, R.; Ou, C.-N.; Vogel, H.; Beaudet, A.L.; Mardon, G. Dach1 Mutant Mice Bear No Gross Abnormalities in Eye, Limb, and Brain Development and Exhibit Postnatal Lethality. *Mol Cell Biol* **2001**, *21*, 1484–1490, doi:10.1128/MCB.21.5.1484-1490.2001.
140. Liu, J.-W.; Li, H.; Zhang, Y. Npas3 Regulates Stemness Maintenance of Radial Glial Cells and Neuronal Migration in the Developing Mouse Cerebral Cortex. *Front Cell Neurosci* **2022**, *16*, 865681, doi:10.3389/fncel.2022.865681.
141. Li, Y.; Fan, T.; Li, X.; Liu, L.; Mao, F.; Li, Y.; Miao, Z.; Zeng, C.; Song, W.; Pan, J.; et al. Npas3 Deficiency Impairs Cortical Astrogenesis and Induces Autistic-like Behaviors. *Cell Reports* **2022**, *40*, 111289, doi:10.1016/j.celrep.2022.111289.
142. Bansod, S.; Kageyama, R.; Ohtsuka, T. Hes5 Regulates the Transition Timing of Neurogenesis and Gliogenesis in Mammalian Neocortical Development. *Development* **2017**, *144*, 3156–3167, doi:10.1242/dev.147256.

143. Pirozzi, F.; Lee, B.; Horsley, N.; Burkardt, D.D.; Dobyns, W.B.; Graham, J.M.; Dentici, M.L.; Cesario, C.; Schallner, J.; Porrmann, J.; et al. Proximal Variants in CCND2 Associated with Microcephaly, Short Stature, and Developmental Delay: A Case Series and Review of Inverse Brain Growth Phenotypes. *Am J Med Genet A* **2021**, *185*, 2719–2738, doi:10.1002/ajmg.a.62362.
144. Fiorenzano, A.; Sozzi, E.; Birtele, M.; Kajtez, J.; Giacomoni, J.; Nilsson, F.; Bruzelius, A.; Sharma, Y.; Zhang, Y.; Mattsson, B.; et al. Single-Cell Transcriptomics Captures Features of Human Midbrain Development and Dopamine Neuron Diversity in Brain Organoids. *Nat Commun* **2021**, *12*, 7302, doi:10.1038/s41467-021-27464-5.
145. Moehlin, J.; Mollet, B.; Colombo, B.M.; Mendoza-Parra, M.A. Inferring Biologically Relevant Molecular Tissue Substructures by Agglomerative Clustering of Digitized Spatial Transcriptomes with Multilayer. *Cell Systems* **2021**, *12*, 694-705.e3, doi:10.1016/j.cels.2021.04.008.
146. Bowling, S.; Sritharan, D.; Osorio, F.G.; Nguyen, M.; Cheung, P.; Rodriguez-Fraticelli, A.; Patel, S.; Yuan, W.-C.; Fujiwara, Y.; Li, B.E.; et al. An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **2020**, *181*, 1410-1422.e27, doi:10.1016/j.cell.2020.04.048.
147. Mátés, L.; Chuah, M.K.L.; Belay, E.; Jerchow, B.; Manoj, N.; Acosta-Sanchez, A.; Grzela, D.P.; Schmitt, A.; Becker, K.; Matrai, J.; et al. Molecular Evolution of a Novel Hyperactive Sleeping Beauty Transposase Enables Robust Stable Gene Transfer in Vertebrates. *Nat Genet* **2009**, *41*, 753–761, doi:10.1038/ng.343.
148. Matsuda, T.; Cepko, C.L. Electroporation and RNA Interference in the Rodent Retina in Vivo and in Vitro. *Proc Natl Acad Sci U S A* **2004**, *101*, 16–22, doi:10.1073/pnas.2235688100.
149. Wang, J.; Xie, G.; Singh, M.; Ghanbarian, A.T.; Raskó, T.; Szvetnik, A.; Cai, H.; Besser, D.; Prigione, A.; Fuchs, N.V.; et al. Primate-Specific Endogenous Retrovirus-Driven Transcription Defines Naive-like Stem Cells. *Nature* **2014**, *516*, 405–409, doi:10.1038/nature13804.
150. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell* **2008**, *134*, 921-931, doi:10.1016/j.cell.2008.09.002.
151. Kam, R.K.T.; Deng, Y.; Chen, Y.; Zhao, H. Retinoic Acid Synthesis and Functions in Early Embryonic Development. *Cell Biosci* **2012**, *2, 11*, doi:10.1186/2045-3701-2-11.
152. Taneja, R.; Roy, B.; Plassat, J.L.; Zusi, C.F.; Ostrowski, J.; Reczek, P.R.; Chambon, P. Cell-Type and Promoter-Context Dependent Retinoic Acid Receptor (RAR) Redundancies for RAR Beta 2 and Hoxa-1 Activation in F9 and P19 Cells Can Be Artefactually Generated by Gene Knockouts. *Proc Natl Acad Sci U S A* **1996**, *93*, 6197–6202.
153. Mendoza-Parra, M.-A.; Malysheva, V.; Saleem, M.A.M.; Lieb, M.; Godel, A.; Gronemeyer, H. Reconstructed Cell Fate–Regulatory Programs in Stem Cells Reveal Hierarchies and Key Factors of Neurogenesis. *Genome Res.* **2016**, *26*, 1505–1519, doi:10.1101/gr.208926.116.
154. Ruberte, E.; Dolle, P.; Chambon, P.; Morriss-Kay, G. Retinoic Acid Receptors and Cellular Retinoid Binding Proteins : II. Their Differential Pattern of Transcription during Early Morphogenesis in Mouse Embryos. *Development* **1991**, *111*, 45–60, doi:10.1242/dev.111.1.45.
155. Uehara, M.; Yashiro, K.; Mamiya, S.; Nishino, J.; Chambon, P.; Dolle, P.; Sakai, Y. CYP26A1 and CYP26C1 Cooperatively Regulate Anterior-Posterior Patterning of the Developing

Brain and the Production of Migratory Cranial Neural Crest Cells in the Mouse. *Dev Biol* **2007**, *302*, 399-411, doi:10.1016/j.ydbio.2006.09.045.
156. Hernandez, R.E.; Putzke, A.P.; Myers, J.P.; Margaretha, L.; Moens, C.B. Cyp26 Enzymes Generate the Retinoic Acid Response Pattern Necessary for Hindbrain Development. *Development* **2007**, *134*, 177-187, doi:10.1242/dev.02706.
157. Ribes, V.; Otto, D.M.E.; Dickmann, L.; Schmidt, K.; Schuhbaur, B.; Henderson, C.; Blomhoff, R.; Wolf, C.R.; Tickle, C.; Dollé, P. Rescue of Cytochrome P450 Oxidoreductase (Por. In *Mouse Mutants Reveals Functions in Vasculogenesis, Brain and Limb Patterning Linked to Retinoic Acid Homeostasis*; Dev Biol, 2007; Vol. 303, pp. 66-81,.
158. Novitch, B.G.; Wichterle, H.; Jessell, T.M.; Sockanathan, S. A Requirement for Retinoic Acid-Mediated Transcriptional Activation in Ventral Neural Patterning and Motor Neuron Specification. *Neuron* **2003**, *40*, 81–95, doi:10.1016/j.neuron.2003.08.006.
159. Wilson, L.; Maden, M. The Mechanisms of Dorsoventral Patterning in the Vertebrate Neural Tube. *Dev Biol* **2005**, *282*, 1-13, doi:10.1016/j.ydbio.2005.02.027.
160. Corral, R.; Storey, K.G. Opposing FGF and Retinoid Pathways: A Signalling Switch That Controls Differentiation and Patterning Onset in the Extending Vertebrate Body Axis. *Bioessays* **2004**, *26*, 857-869, doi:10.1002/bies.20080.
161. Sockanathan, S.; Jessell, T.M. Motor Neuron–Derived Retinoid Signaling Specifies the Subtype Identity of Spinal Motor Neurons. *Cell* **1998**, *94*, 503–514, doi:10.1016/S0092-8674(00)81591-3.
162. Sockanathan, S.; Perlmann, T.; Jessell, T.M. Retinoid Receptor Signaling in Postmitotic Motor Neurons Regulates Rostrocaudal Positional Identity and Axonal Projection Pattern. *Neuron* **2003**, *40*, 97–111, doi:10.1016/S0896-6273(03)00532-4.
163. Holzschuh, J.; Barrallo-Gimeno, A.; Ettl, A.-K.; Dürr, K.; Knapik, E.W.; Driever, W. Noradrenergic Neurons in the Zebrafish Hindbrain Are Induced by Retinoic Acid and Require *Tfap2a* for Expression of the Neurotransmitter Phenotype. *Development* **2003**, *130*, 5741–5754, doi:10.1242/dev.00816.
164. Chatzi, C.; Brade, T.; Duester, G. Retinoic Acid Functions as a Key GABAergic Differentiation Signal in the Basal Ganglia. *PLoS Biol* **2011**, *9*, e1000609, doi:10.1371/journal.pbio.1000609.
165. Jacob, J.; Kong, J.; Moore, S.; Milton, C.; Sasai, N.; Gonzalez-Quevedo, R.; Terriente, J.; Imayoshi, I.; Kageyama, R.; Wilkinson, D.G.; et al. Retinoid Acid Specifies Neuronal Identity through Graded Expression of Ascl1. *Curr Biol* **2013**, *23*, 412–418, doi:10.1016/j.cub.2013.01.046.
166. Rataj-Baniowska, M.; Niewiadomska-Cimicka, A.; Paschaki, M.; Szyszka-Niagolov, M.; Carramolino, L.; Torres, M.; Dollé, P.; Krężel, W. Retinoic Acid Receptor β Controls Development of Striatonigral Projection Neurons through FGF-Dependent and Meis1-Dependent Mechanisms. *J Neurosci* **2015**, *35*, 14467–14475, doi:10.1523/JNEUROSCI.1278-15.2015.
167. Sagner, A.; Gaber, Z.B.; Delile, J.; Kong, J.H.; Rousso, D.L.; Pearson, C.A.; Weicksel, S.E.; Melchionda, M.; Mousavy Gharavy, S.N.; Briscoe, J. Olig2 and Hes Regulatory Dynamics during Motor Neuron Differentiation Revealed by Single Cell Transcriptomics. *PLoS Biol* **2018**, *16, e2003127*, doi:10.1371/journal.pbio.2003127.
168. Carcagno, A.L.; Bella, D.J.; Goulding, M.; Guillemot, F.; Lanuza, G.M. Neurogenin3 Restricts Serotonergic Neuron Differentiation to the Hindbrain. *J Neurosci* **2014**, *34*, 15223-15233, doi:10.1523/JNEUROSCI.3403-14.2014.

169. Li, C.; Fleck, J.S.; Martins-Costa, C.; Burkard, T.R.; Themann, J.; Stuempflen, M.; Peer, A.M.; Vertesy, Á.; Littleboy, J.B.; Esk, C. Single-Cell Brain Organoid Screening Identifies Developmental Defects in Autism. *Nature* **2023**, *621*, 373-380, doi:10.1038/s41586-023-06473-y.
170. Bian, S.; Bian, S.; Knoblich, J. Genetic Engineering to Initiate Tumorigenesis in Cerebral Organoids. *Protocol Exchange* **2018**, doi:10.1038/protex.2018.071.