

**5-Hydroxymethylcytosine Primes Neuronal Genes for Activation During Zebrafish
Retinal Progenitor Cell Differentiation**

by

Jennifer A. Spengler

Bachelors of Science, Millersville University of Pennsylvania , 2015

Submitted to the Graduate Faculty of the
School of Medicine in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This thesis was presented

by

Jennifer A. Spengler

It was defended on

December 16, 2020

and approved by

Dennis Kostka, Associate Professor, Department of Developmental Biology

Arjumand Ghazi, Associate Professor, Department of Pediatrics

Thesis Advisor: Jeffrey Gross, Professor, Department of Ophthalmology

Copyright © by Jennifer A. Spengler

2020

5-Hydroxymethylcytosine Primes Neuronal Genes for Activation During Zebrafish Retinal Progenitor Cell Differentiation

Jennifer A. Spengler, MS

University of Pittsburgh, 2020

The epigenetic mark 5-hydroxymethylcytosine (5hmC) is generated on DNA by the oxidation of 5-methylcytosine (5mC) by members of the ten-eleven-translocase (tet) enzyme family. 5hmC is detected at high levels within differentiated neurons and previous results from our lab have shown that activity of tet2 and tet3 enzymes is necessary for proper retinal differentiation in zebrafish (Seritrakul et al., 2017). However, the mechanism(s) by which tet activity and 5hmC regulate gene expression and influence differentiation in retinal progenitor cells (RPCs) is unknown. This study aimed to determine how 5mC and 5hmC modulate gene expression in early and late RPCs, and if 5mC and 5hmC relate to gene expression after RPCs differentiate to retinal ganglion cells (RGCs) in the developing zebrafish. 5hmC and 5mC levels in zebrafish RPCs were determined by oxidative-bisulfite and bisulfite sequencing. Gene expression levels were determined with RNA sequencing on RPCs and RGCs. Analyses were performed on RPCs at two developmental time points: 22 hours post-fertilization (HPF), when retinal progenitor cells are proliferative and not yet committed to retinal neuron fates, and 27 HPF, when retinal ganglion cell differentiation is underway. RGCs were also collected as they differentiate at 27 HPF to inform gene expression as a result of the epigenetic changes in progenitors. The resulting epigenetic and gene expression data from RPCs were integrated to identify how 5mC and 5hmC were distributed in the progenitor. The analysis performed here revealed that both 5mC and 5hmC were distributed to repress gene expression in retinal progenitor cells. Regions that gained 5hmC between early and late RPCs were enriched for neurogenic genes but did not correlate to gene expression in the RPCs. However, the increasing 5hmC was localized to neurogenic genes that were upregulated in RGCs. The data presented here indicate that the 5hmC generated

between early and late RPCs primes genes for activation after the onset of differentiation.

Table of Contents

1.0 Introduction	1
2.0 Materials and Methods.....	9
2.1 Zebrafish lines, maintenance, and husbandry.....	9
2.2 Single Cell dissociation for FACS.....	10
2.3 Sorting and Nucleic acid extraction	10
2.4 RNAseq library Preparation and sequencing	11
2.5 oxBS and BS Library preparation	11
2.6 Interrogation of internal cutting control.....	12
2.7 Initial QC and Processing the BS and oxBS data.....	13
2.8 RNAseq Analysis and Identification of Differentially expressed Genes ..	15
2.9 Generation of Deciles, Heatmaps and Boxplots.....	16
2.10 Differentially gene expression dependence on DMR and DHMRs	16
3.0 Results	17
3.1 Collection of RPCs and RGCs	17
3.2 Oxidative Bisulfite and Bisulfite sequencing and quality control assessment	19
3.3 5mC and 5hmC are distributed throughout genic regions	26
3.4 Changes in 5mC and 5hmC accumulation in neural development genes occur as RPCs mature	30
3.5 Few expression changes are present between RPCs at 22 and 27 HPF, but many exist between RPCs and RGCs.....	32
3.6 5mC and 5hmC changes with expression in Genic Regions	35
3.7 DEGs in RPCs are not dependent on 5mC and 5hmC changes	37
3.8 5hmC primes genes for activation after RPC restrictions	38
4.0 Discussion	40
5.0 Figures	50
6.0 Tables.....	68

Bibliography 82

List of Tables

Table 1. Coverage of the Genome.....	68
Table 2. Whole Genome Alignment to Reference.....	68
Table 3. Whole Genome C:T Conversion	69
Table 4. CpG Sites Filtered out from Samples.....	70
Table 5. Localization of DMRs.....	71
Table 6. Localization of DHMRs	72
Table 7. Differentially Expressed Genes between early and late RPCS	73
Table 8. Number of Differentially expressed genes with DHMR or DMR.....	73
Table 9. Contingency Table for Genes and DMR.....	74
Table 10. Contingency Table for Genes and DHMR	75
Table 11. Fisher's Exact Test Results on RPCs and DMRs and DHMRs	75
Table 12. Genes Upregulated in RGC with Increasing 5hmC	76
Table 13. Contingency Table for Genes DEG between RPC and RGC and DHMR	80
Table 14. Changes in expression in RGCs was dependent on changing 5hmC....	80
Table 15. Contingency Table for Genes Upregulated between RGC and RPC and increasing DHMR.....	81
Table 16. Changes in expression in RGCs was dependent on changing 5hmC..	81

List of Figures

Figure 1 Collection of RPCs.	50
Figure 2. 5mC and 5hmC distribution in the zebrafish genome was enriched in CpG Residues	51
Figure 3. 5mC and 5hmC content in CGI.....	52
Figure 4. 5mC and 5hmC were present in different genic regions.....	53
Figure 5. 5mC was not uniformly distributed around genic region.	54
Figure 6. 5hmC was patterned around genic regions.....	56
Figure 7. Genes containing increasing and increasing DHMRs and decreasing DMRs are Enriched for Neural Genes.....	59
Figure 8. Differentially expressed genes between early and late RPCs indicate few differences between cells.	61
Figure 9. Specified RGCs upregulate many genes involved in neurogenesis compared to RPCs.	63
Figure 10. 5mC and 5hmC change gene expression.....	64
Figure 11. 5mC changes in all genic regions with expression.....	65
Figure 12. 5hmC changes in most genic regions with expression	66
Figure 13. Genes upregulated in RGC fated progenitor contain increasing 5hmC and were enriched in neural genes	67

Preface

List of abbreviations used throughout this work

Abbreviation	
ESC	Embryonic stem cell
5hmC	5-hydroxymethylstyosien
5mC	5-methylcytosine
BS	Bisulfite
CGI	Cpg island
DEG	Differentially expressed genes
DHMR	Differentially hydroxymethylated region
DMR	Differentially methylated regions
ESC	Embryonic stem cell
FACS	Fluorescence activated cell sorting
GCL	Ganglion cell layer
HPF	Hours post fertilization
INL	Inner nuclear layer
IPL	Inner plexiform layer
mESC	Mouse embryonic stem cells
NPC	Neural progenitor cell
ONL	Outer nuclear layer
OPL	Outer plexiform layer
oxBS	Oxidative bisulfite
RGC	Retinal ganglion cell
RPC	Retinal progenitor cell

1.0 Introduction

The vertebrate eye is the component of the central nervous system (CNS) that takes light and converts it to a signal perceived by the brain as vision. At the back of the eye lies the thin neural tissue, the retina, responsible for converting light to an interpretable biological signal and carrying it to the brain. The vertebrate retina is made up of six neural- and one glial- cell types divided into five layers: three cellular layers separated by two plexiform layers, where neurons synapse. The outermost cellular layer, the outer nuclear layer (ONL), is made up of rod- and cone- photoreceptors that undergo depolarization after light exposure. The depolarization triggers a signaling cascade causing neurotransmitter release at the next layer, the inner plexiform layer (IPL) (Hoon et al., 2014). At the IPL, the synapses occur between the photoreceptor cells of the inner nuclear layer (INL) made up of bipolar, horizontal and amacrine neurons. The cells of the INL interpret the intensity of the light received by the photoreceptors and carry it to the inner plexiform layer (IPL) (Hoon et al., 2014). Synapses in the INL occur between the bipolar horizontal and amacrine cells of the INL and transmit signals to the ganglion cell layer (GCL), containing primarily ganglion cells and some amacrine cells. The retinal ganglion cells (RGC) of the GCL have long axons that bundle together to form the optic nerve, which connects the retina to the brain for interpretation as vision. Stratified within all the retinal layers are the Müller glia, which provides structural and trophic support to the neurons of the retina (Vecchio 2010). Each of these cell types must work together to create vision.

For the retina and eye to form properly, many steps of tissue specification and morphogenesis must occur. The future eye is specified from the forebrain by expression of the retinal homeodomain transcription factor, *Rx*, cells expressing this gene will undergo a series of evaginations and invaginations to generate the optic cup (Fuhrmann, 2010). Once the structure of the eye forms, the presumptive retina is established and can be identified by *pax6* and *vsx2* expression (Heavner and Pevny, 2012). The presumptive retina is made up of retinal progenitor cells (RPC), which are a multipotent population of cells able to give rise to all retinal cell types. Early in development, the RPC is multipotent and can differentiate into all the retinal cell types in a stereotypical order (Wetts et al., 1989). In zebrafish, the multipotent RPC can be identified by *vsx2* expression, which remains exclusive to the RPCs until 80 HPF when it also labels a subset of bipolar cells (Vitorino et al., 2009). However as retinal development progresses, some of the RPC population activates other genes that identify restricted RPC population (Cepko et al., 1996). The restricted RPC population expresses *atona1/7* which identifies RPCs that will generate RGCs, the first differentiated neurons (Masai et al., 2000). During retinal development, several restricted RPC populations arise, each forming a different subset of retinal neurons. This appears to be a conserved process in vertebrate development (Livesey and Cepko, 2001). RPCs collected after the differentiation of horizontal cells and cones are no longer competent to all cell fates, differentiating into a subset of the amacrine, Muller glia, rods and cones, (Turner and Cepko, 1987). When younger rat RPCs from the neuroepithelium are cultured *in vitro*, they differentiate into a set number of cell types that do not change in different culture conditions (Cayouette et al., 2003). Restriction of the early and late

RPC population in rats occurs and neither change the cell types they differentiate into. Co-culture of early rat RPCs with late postnatal RPCs found that although early RPCs will differentiate into different proportions of cell types, they are restricted to the same cell types that would differentiate *in vivo* (Belliveau and Cepko, 1999). RPCs collected from different time points have been shown to have different differentiation capabilities. This led to the theory of RPC competence, hypothesizing that multipotent RPCs go through different competence states where each state forms a specific subset of retinal neurons. It is theorized that external signaling helps direct the competent cells but that they are limited to a subset of fates by cell-intrinsic properties (Cepko, 2014). Transcription factors have been proposed to be the intrinsic element to competency, activating the gene networks necessary to form cell types (Stenkamp, 2007). However there is increasing evidence that DNA methylation and its newly identified derivative restricts cell fates and could be the intrinsic mechanism behind RPC competency (Kim and Costello, 2017; T. Li et al., 2015; Tahiliani et al., 2009).

DNA methylation is the process by which a methyl group is added to the 5' carbon of a cytosine residue by DNA methyltransferases (DNMTs) - either DNMT3A or DNMT3B - to create 5-methylcytosine (5mC). These 5mC marks are maintained during DNA replication by another DNMT, DNMT1, and are thus heritable modifications (Lyko, 2018). DNA methylation has been shown to restrict lineages keeping a progenitor cell limited to a set number of cell fates, with embryonic stem cells (ESC) without DNMT failing to undergo lineage restriction (Hemberger et al., 2009). Although 5mC can be maintained throughout cell division, it can be oxidized by members of the ten-eleven-

translocase (TET) family of enzymes, which generate 5-hydroxymethylcytosine (5hmC) (Tahiliani et al., 2009). There are three TET family members, TET1, TET2, and TET3, capable of generating 5hmC (Tahiliani et al., 2009). TETs can further oxidize 5hmC to 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5CaC), which can be recognized for base excision repair machinery or by thymidine glycosylases to restore an unmodified cytosine in the genome (He et al., 2011; Ito et al., 2010). The conversion of 5mC to 5hmC to an unmodified cytosine is referred to as active demethylation (Tahiliani et al., 2009). Importantly, due to the low affinity of TETs for 5hmC (He et al., 2011), 5hmC (like 5mC) also exists as a stable epigenetic mark within the genome (Bachman et al., 2014). 5mC and 5hmC are generated in all vertebrate tissues, but 5hmC is particularly enriched in neural tissue, whereas 5mC is more uniform (Globisch et al., 2010). At present, little is known about the function of 5hmC enrichment in the retina, but studies in the brain indicate a critical role for 5hmC during neuronal differentiation (Szulwach et al., 2011b).

In the developing mouse brain, 5hmC levels have been shown to increase over developmental time between perinatal and adult brains, implicating a role for 5hmC as neurons are undergoing differentiation *in vivo* (Song et al., 2011; Szulwach et al., 2011b). These findings have been corroborated by *in vitro* studies showing 5hmC levels also increase as mouse embryonic stem cells (mESC) undergo differentiation to a neural progenitor cell (NPC) (T. Li et al., 2015). mESCs are pluripotent and thus able to give rise to all cell types (Waisman et al., 2019). NPCs on the other hand are multipotent like RPCs, and able to differentiate into a select set of neural cell types

(Doe, 2008). This rise in 5hmC levels as cells differentiates indicates a potential role for 5hmC in restricting cell "potency." Tet3 was shown to be responsible for the increase of 5hmC during the commitment from mESC to NPC, as knockout of Tet3 lead to a decrease in 5hmC levels in both mESCs and NPCs (D. Li et al., 2015; T. Li et al., 2015). Interestingly, the mESCs could differentiate into NPCs with Tet3 knockout, but NPC differentiation to neurons was decreased compared to wild-type controls. The NPC proliferation increased after *Tet3* knockdown and expression of pluripotency genes, *Oct4*, *Nanog*, and *Sox2* was re-activated, making the NPC more stem cell-like (D. Li et al., 2015; T. Li et al., 2015; Santiago et al., 2020a, 2020b). The Tet3 KO NPCs were not fully stem-cell-like, still retaining the NPC marker *Nestin*, yet could not differentiate into as many neurons as NPCs with Tet3 (T. Li et al., 2015). These results indicate there is a requirement for Tet3 and 5hmC in restricting pluripotency and promoting neural differentiation. Thus, Tet proteins and 5hmC appear to be important features of neural development, leading to the repression of pluripotency genes in NPCs. However, impact on gene expression as a result of 5hmC depends greatly on its genic context and is likely tissue dependent.

5hmC can only appear in the genome where 5mC was present, so 5hmC localization is similar to 5mC, but with different impact on expression. 5hmC at the promoter and TSS, has been found to accumulate in both low- and highly expressed genes (Szulwach et al., 2011a), suggesting it can be both an activating and repressive mark. Furthermore, 5hmC in the promoter has been shown to have a negative correlation with expression in cardiac progenitors (Greco et al., 2016) and no correlation

with expression in neural tissues (Mellén et al., 2012; Tan et al., 2013). Without there being a consensus, it appears the role of 5hmC in promoters and TSS may be tissue dependent. In the gene body, 5mC has been positively correlated with gene expression (Anastasiadi et al., 2018), and in many cases, 5hmC has also been positively correlated with gene expression (Pastor et al., 2013). Conversely, there are some instances in neural development where an increase of 5hmC was found in the gene body during the transition from mESC to NPC but did not correlate with gene expression (Williams et al., 2011). Similarly, an increase in 5hmC within gene bodies without a change in expression correlation was also seen during differentiation of NPCs to neurons (C. Zhang et al., 2016). Further, it was found that during brain development, 5hmC increases in genes that will be developmentally activated but does not correlate well with expression (Szulwach et al., 2011a). 5hmC was found to accumulate on genes that have not yet been activated but will be expressed at later developmental time points (Szulwach et al., 2011b). These findings collectively indicate that 5hmC may not be an activating mark, but instead functions to prime genes for activity at later stages (Szulwach et al., 2011a). Thus, it appears that 5hmC, unlike 5mC, is not always clearly linked with gene expression, however some trends do exist.

5hmC has been shown to be a critical component of brain development and NPC differentiation, with RPCs functioning similar to NPCs, it is not surprising that there is some evidence of 5hmC having a role in retinal development and RPC differentiation. The first report of 5hmC in eye development came from studies in *Xenopus*, where *tet3* was knocked down by morpholino. The resulting morphants had an eyeless phenotype,

and genes necessary for eye development were downregulated (Xu et al., 2012). Although showing *tet3* and 5hmC are required for eye development, without an eye, this study could not inform on the role of 5hmC in RPCs. Evidence from our lab in zebrafish, strongly indicates 5hmC has a role in RPC restriction and downstream differentiation. Because zebrafish mutants can be generated easily and knockdown of a gene by morpholino can lead to off-target effects, such as non-specific cell death, zebrafish mutants were generated used to study tet proteins (Gerety and Wilkinson, 2011). Zebrafish truncation mutants created using targeted genome editing, led to catalytically inactive *tet2* and *tet3* (Seritrakul and Gross, 2017). Only *tet2*^{-/-};*tet3*^{-/-} double mutant fish showed an overt eye phenotype, an abnormally small eye, and had a reduction in 5hmC compared to wild type. In section, the phenotype's severity was apparent within the retina, having the appearance of a less mature retina with the laminated retinal structure failing to form. Furthermore, few cells were able to differentiate in *tet2*^{-/-};*tet3*^{-/-} double mutants and the retinal cell were proliferative longer into development. In addition, cells within the mutant retina underwent proliferation longer into development than a wild-type embryo suggesting cells were more progenitor-like. Initial cell specification of all retinal cell types occurred, but few retinal neurons fully differentiated, as seen with mouse NPCs (T. Li et al., 2015). Interestingly when mutant cells were transplanted to a wild-type embryo, the mutant cells did differentiate into all cell types (Seritrakul and Gross, 2017). Showing that the early RPCs are able to generate all cell fates, if the environment is permissive, but did not show if 5hmC is necessary in restricted progenitors to form specific cell types. The gene expression within the retina was also disrupted, indicating a need for 5hmC in the

genome to regulate gene expression. This work in the lab identified the 5hmC is required for proper retinal differentiation.

These studies show that 5hmC is necessary for proper retinal development and differentiation of all cell fates, but neither of the studies determined where in the genome 5hmC is present during specification or how it changes. The data discussed in this Masters thesis aim to determine and analyze the 5mC and 5hmC patterns of early and late RPC before restriction occurs to identify if changes in the epigenome are necessary to influence gene expression.

In this study the 5mC and 5hmC distribution between early and late RPCs was identified and showed that 5hmC mimics the distribution of 5mC in the genome. In addition, both 5mC and 5hmC were found to decrease as gene expression increases. This together with the similarity of 5mC and 5hmCs distribution indicates that 5hmC represses gene expression in RPCs. Specific small regions in the genome were observed to increase in 5hmC between early and late RPCs and these are enriched in genes related to neurogenesis. Combining the RGC expression data with the methylation analyses shows that regions increasing 5hmC content were present within the neurogenic genes that were upregulated once RGCs are differentiated. The gain of 5hmC in upregulated neurogenic genes indicates that 5hmC is generated in RPCs to prime neurogenic genes for activation after differentiation occurs.

2.0 Materials and Methods

2.1 Zebrafish lines, maintenance, and husbandry

Zebrafish adults were maintained at 28.2 °C on a 14:10 hour light: dark cycle. The transgenic zebrafish line Tg(*vsx2*: GFP) or Tg(*atoh7*:GFP) (Masai 2003) was outcrossed to wild-type TU (ZIRC cat# ZL57) for use in experiments. Crosses were set up 48 hours before the cell sorting time and separated by barriers. Barriers were removed 26 hours before the sort time, and fish were bred for 30 minutes before embryos were collected. After collection, embryos were split into groups of 50, non-viable embryos were removed at 8 HPF, and embryos were grown in Danieau's embryo media (58mM NaCl, 0.7mM KCl, 0.4 mM MgSO₄, 0.6mM Ca(NO₃)₂, 5mM Hepes pH 7.6) at 28.2 C until 22 HPF or 27 HPF. When the embryos reached the desired developmental time point, they were taken out of the incubator and screened for viability and GFP expression. Embryos were treated with 25 Units of Pronase (Sigma Aldrich # 11459643001) for 5 minutes to remove the chorion from the embryos, then rinsed three times with Danieau's embryo media then twice with cold PBS pH 7.4 (Fisher Scientific # 10010023). Embryos were treated with tricaine and kept on ice during dissection. Eyes were dissected using flame sharpened tungsten wire (0.025mm diameter) and, once isolated, transferred to an Eppendorf tube and kept on ice until cell dissociation.

2.2 Single Cell dissociation for FACS

After eyes were isolated, samples were centrifuged for 5 minutes at 4500 RPM at 4C. The supernatant was removed, and eyes were resuspended in 1 ml of 0.25% trypsin (Trypsin-EDTA(0.5%) Fisher Scientific # 15400054) for 5 minutes at room temperature. During incubation, eyes were resuspended with a pipette and sheared with a 25G needle. After the 5 minutes, the cell suspension was put through a 70 um filter and washed with a mix of 40mM CaCl₂ with 5% FBS to inactivate the trypsin. The suspension was then centrifuged again for 5 minutes at 4500 RPM at 4 C. The supernatant was removed, the pellet resuspended in cold PBS, and centrifuged. This process was repeated for a total of three washes. After the final wash, the cells were resuspended in 5% FBS and brought on ice to the Rangos Flow Cytometry Core at the John G. Rangos Sr. Research Center at the University of Pittsburgh.

2.3 Sorting and Nucleic acid extraction

Cells were sorted on a BD FACS Aria II, with an 85 µm nozzle and a sheath pressure of 40. Prior to sorting, cells were treated with propidium iodide (PI) to stain dead cells. Gates were initially set using a GFP- population stained with PI, an unstained GFP- population, and a PI stained GFP- population to collect the brightest individual live GFP positive cells. The same gates were used for each experiment. After sorting parameters were set, 1,000 cells were sorted into a 96-well plate containing Clontech lysis buffer. The lysate was then used to make cDNA and a

subsequent RNA sequencing library. The remainder of cells were sorted into an Eppendorf tube with PBS and 5% FBS. Sorted cells were centrifuged, and the supernatant was removed. DNA extraction was performed using the Zymo Quick-DNA/RNA Microprep Plus Kit according to kit instructions for cell pellets. DNA quantification was carried out using Qubit.

2.4 RNAseq library Preparation and sequencing

RNA was extracted from 1000 sorted cells and cDNA generated using Smart-Seq v4 Ultra-low input RNA kit (Takara Biosciences cat # 634894). RNAseq libraries were prepared using Nextera XT library kit from Illumina and according to manufacture instructions. Samples were run on a NextSeq500 to generate 75bp paired end reads, 40-50 million reads per samples were generated.

2.5 oxBS and BS Library preparation

DNA samples from the same time point were pooled to reach the recommended concentration of 200 ug needed for library preparation. The samples could be split for paired BS and oxBS reactions and accommodate the high DNA degradation during the BS reaction(Wreczycka et al., 2017). DNA was fragmented using Covaris ME220 using a duration of 260s, peak power of 75, duty of 20, 220 cycles per burst, and average power of 15 and carried out at 9°C. Fragment size distribution and concentration was

confirmed to be 200 bp on a Tapestation DS100 high sensitivity tape. The concentration of fragmented DNA was used to calculate a 5% by mass quantity of cutting control from a TrueMethyl Whole Genome Cambridge epigenomics kit (V3.1, out of production) for interrogation prior to sequencing, as recommended in personal communication with technical support from Tecan. Library preparation was completed using Tecan's TrueMethyl oxBS-Seq module (Tecan cat #0414-32, #0541-32) according to manufacturer instructions. In brief, fragmented sample was purified, and end repair was performed. Samples were split for a Bisulfite and Oxidative Bisulfite reaction in parallel, then sequencing adapters and indices were added. Samples were then purified and treated with either the Oxidant solution or mock with ultrapure water for 10 minutes at 40 C. Both samples were bisulfite converted, using the bisulfite reagent included in the kit, for 60 minutes. According to manufacturer parameters, libraries were purified and amplified, but with cycles repeated 15 or 20 times for BS and oxBS, respectively.

2.6 Interrogation of internal cutting control

PCR primers specific to the cutting control were used to amplify the control sequences. The protocol used was initial denaturation for 5 minutes at 95°C; 40 cycles of denaturation at 95°C for 30, annealing for 30 s at 60°C, extension for 15 s at 72°C followed by a final extension for 5 minutes at 72°C. Products were purified using Qiagen Qiaquick PCR purification kit (Qiagen cat # 28106), and a restriction fragment length polymorphism was performed for 18 hours using Taqql and was run on a gel.

This confirmed the conversion was successful, and the samples were submitted for sequencing. A shallow sequencing experiment of 0.5X coverage per sample was done on an Illumina MiSeq micro300 chip to confirm the library's viability. Then the sample underwent sequencing on an Illumina NovaSeq S4 chip generating 150bp paired end reads to generate data of at least 30X coverage. Between 600-700 million reads were generated per sample.

2.7 Initial QC and Processing the BS and oxBS data

Sequences were aligned using the BWA-meth aligner (Pederson 2005) to the zebrafish genome(GRCz11), percent alignment can be observed in **Table 1**. Sample coverage was calculated using Picard SamSort and were a minimum of 30X each(**Table 2**). MethPipe was used to process the aligned BS and oxBS samples in parallel, which were sorted, duplicates removed, and the rate of methylation was calculated for each cytosine residue in the genome. Then the frequency of methylation from the BS and oxBS sample were run through mlml, within MethPipe, to calculate the 5hmC within each set of BS and oxBS samples and adjust the methylation level (Song et al., 2013).

The 5mC, 5hmC, and C levels per base were loaded into R, where sites covered by fewer than 10 reads were filtered out, according to ENCODE standards(Davis 2018). Remaining bases were used to calculate the average the β values for 5mC and

5hmC with the genome and were plotted using the R package ggplot2 (Wickham, 2016). Average 5mC or 5hmC levels per replicate at each time point loaded into GraphPad Prism 9.0 for Windows, GraphPad Software, San Diego, California USA, where replicates were pooled to perform a two-tailed t-test between timepoints. CpG islands (CGI) were downloaded from the UCSC table browser (Karolchik et al., 2004). The CGI shores were annotated as 2000 bp upstream and downstream of the islands, and CGI shelves were annotated from the end of each shore and extended 2000 bp as recommended by previous methods (Bibikova et al., 2011).

To prevent a base from being counted more than once if it fell within multiple genic regions from different transcripts, a hierarchy was used to annotate the regions. The hierarchy used was promoters, 5'UTRs, exons, and lastly 3'UTRs. A base that could be annotated to a promoter transcript and another a 5'UTR, the base would be only considered a promoter. Any gaps that fell within a gene between exons was considered an intron. Anything that did not fall within a gene was considered intergenic. Genes or gene bodies were kept and analyzed separately. Each of the regions created by the annotation were used throughout the remainder of the analyses. They also served as a target to create the distribution heatmaps around genic regions with the package enrichedheatmap (Gu et al., 2018).

The average β value for 5mC and 5hmC at each CpG was calculated within each genic region for each gene. For every gene, an average 5mC and 5hmC for each corresponding promoter, 5'UTR, exons, 3'UTR, and introns was calculated. These

data were compiled into two SummarizedExperiment objects for use in analysis (Morgan et al., 2017).

DMR and DHMRs were identified using MethPipe, using the radmeth regression program, to find the probability of either a difference in 5mC or 5hmC between timepoints. Then regions of significance were identified and filtered using a p-value of 0.01. Once identified, DMR and DHMRs were filtered by a change of 25 % or more, or -25% or less, as recommended in other methods (Akalin 2012). The remaining DMR and DHMRs were annotated by overlaps with genic regions, using the package GenomicRanges. To calculate the frequency of 5mC or 5hmC change, the sum of bases contained within the DMR or DHMRs were calculated and divided by the sum of all the widths of the genic regions. The genes that contained a change in 5hmC or 5mC were run through GO-term analysis using the package clusterProfiler. Venn Diagrams were generated by inputting the gene names onto the web page Venny (Oliveros 2007-2015).

2.8 RNAseq Analysis and Identification of Differentially expressed Genes

RNAseq data were processed using a standard workflow using Rsubreads, for alignment and limma and edgeR to find gene counts and differentially expressed genes (Liao et al., 2019; Ritchie et al., 2015; Robinson et al., 2010). Samples were normalized for library size, and a linear model was used to identify differentially expressed genes

using a log₂- fold change of 1.5 and a false discovery rate of 0.1. Differentially expressed genes were run through GO-term and KEGG analysis using clusterProfiler (Yu et al., 2012).

2.9 Generation of Deciles, Heatmaps and Boxplots

The logTPM was calculated for all genes detected. Genes were ranked from increasing to decreasing logTPM values. The data set was split into 10 groups, using the package dplyr, creating 10 deciles. The geneIDs making up each decile were used to create deciles of the averaged 5mC and 5hmC data across genic regions.

5mC and 5hmC from each genic region was matched to expression deciles by matching the geneIDs. All genes in a genic region were averaged within each decile to create one average for 5mC or 5hmC per genic region per decile. The generated average per decile was plotted using the package complex heatmap (Gu et al., 2016).

2.10 Differential gene expression dependence on DMR and DHMRs

The differentially expressed genes were split based on whether they contained a region of changing 5mC or 5hmC. This data was tabulated to create a contingency table for use in Fisher's exact test, implemented in R, to test for dependence

3.0 Results

3.1 Collection of RPCs and RGCs

There is a large body of evidence showing that 5mC and 5hmC accumulation in genes and their associated regulatory regions is critical to neural differentiation during neural development (T. Li et al., 2015; Szulwach et al., 2011b; J. Zhang et al., 2016). Work from our laboratory has shown that in *tet2^{-/-};tet3^{-/-}* mutants, 5hmC is significantly reduced and differentiation of retinal cell types is impaired (Seritrakul and Gross, 2017). It is not known where in the genome of RPCs 5hmC or 5mC is distributed during differentiation, and therefore the goal of this research was to determine this and enable a more thorough understanding of how DNA hydroxymethylation influences retinal development.

To determine 5mC and 5hmC distributions throughout early retinal development, RPCs were collected using the transgenic zebrafish *Tg(vsx2:GFP)*. *vsx2* labels multipotent RPCs throughout retinal differentiation, starting at 15 HPF until 80 HPF when it labels a subset of bipolar neurons (Vitorino et al., 2009). RPCs were collected at 22 HPF to represent a naïve progenitor state, before lineage restriction but after the presumptive retina has been established. RPCs were also collected at 27 HPF, a time at which a subset of RPCs were expected to be restricted to RGC (and possibly other) cell fates. A differentiated cell population was also collected by isolating GFP+ cells

from Tg(*atoh7:GFP*), which labels cells differentiating RGCs and also those fated to become cones (Vitorino et al., 2009).

Both early and late RPCs were isolated by first selecting GFP⁺ embryos containing the Tg(*vsx2:GFP*)(**Figure 1A**). Heads were dissected to enrich the sample for GFP⁺ cells and then dissociated into a cell suspension to use for FACS. Before FACS the cell suspension was stained with propidium iodide (PI) to identify dead cells. Single, live GFP⁺ cells were collected by gating to select for cells without PI, and high GFP fluorescence (**Figure 1B-D**). One thousand cells were sorted for bulk RNAseq experiments, and the remainder of the cells were collected for DNA extraction follow by oxidative bisulfite (oxBS) and bisulfite (BS) library preparation. Each sorting session contained a minimum of 100 embryos, and between 60,000 to 180,000 cells were collected per session. DNA yield from sorted RPCs ranged from 11.5 to 26.8 ng/ul. Samples were pooled to reach the minimum recommended concentration of 200 ng for oxBS and BS library preparation. Isolated DNA was used to generate paired oxBS and BS sequencing libraries using Tecan's TrueMethyl library kit with oxBS module. BS and oxBS must be used together to distinguish 5mC, 5hmC and unmodified C because neither method alone can. BS identifies an unmodified C from a modified one, losing the distinction between 5hmC and 5mC. On the other hand, oxBS can definitively distinguish 5mC bases, but not 5hmC and unmodified C in the final result. By comparing the frequency that each aligned C was methylated or not in the BS compared to oxBS data, it is possible to determine how frequently each C in the sample was a 5mC, 5hmC or C (Booth et al., 2013).

The same process was repeated to collect RGCs using *Tg(ato7:GFP)*. Because *atona17* is expressed in specified RGCs and brain, individual eyes were dissected to eliminate unwanted brain cells before dissociation and FACS (Saul et al., 2008). The same gating strategies to collect RPCs were used to collect RGCs. Each FACS experiment for RGC isolation contained a minimum of 100 embryos, but the GFP⁺ fraction was less than 4% of cells in the suspension. With this low frequency of cells, it was possible to collect 1000 for RNAseq experiments, but the remainder of cells sorted for DNA extraction yielded at most 6,000 cells and < 1 ng/ul of DNA. Because of the high input requirement of oxBS and BS, it was not technically possible to collect enough RGCs at 27 HPF to reach input quantity requirements without exceeding input volume constraints. Therefore, only RNAseq was performed on 27 HPF RGCs.

3.2 Oxidative Bisulfite and Bisulfite sequencing and quality control assessment

oxBS and BS libraries for early and late RPCs were generated and they were sequenced and analyzed for 5mC and 5hmC content. Once the oxBS and BS sequencing data were received, sample quality for oxBS and BS libraries were assessed. First, sequencing depths were analyzed, and the results showed greater than 30-times coverage (30X) per sample, exceeding the recommended ENCODE standards (**Table 1**. Davis et al., 2018). Each sample was aligned to the zebrafish genome assembly GRCz11 using BWA-meth (Pedersen et al., 2014) and had >99% alignment (**Table 2**). Aligned oxBS and BS samples were processed through the

MethPipe pipeline to calculate the ratio of 5mC, 5hmC, and C at every C residue in the genome (Song et al., 2013). A computational conversion test was performed on the data to determine the efficiency of the bisulfite conversion, using bsrate within the MethPipe software (Song et al., 2013). Conversion efficiency was determined by comparing CpG sites, which are highly methylated and should not undergo bisulfite conversion, to all non-CpG Cs in the genome which are typically unmethylated and should not undergo conversion (Feng et al., 2010; Guo et al., 2014; Song et al., 2013; Ziller et al., 2011). Results from this test range from 0-100%, with 100% representing complete bisulfite conversion; samples used for this study ranged from 96.1-97.8% conversion (**Table 3**). Although the ENCODE conversion standard is 98%, a cutoff of 95% has been used in other studies and can yield meaningful results (Roeh et al., 2018). Importantly, this QC measure was developed for mammals that typically have no non-CpG methylation, whereas zebrafish do have some non-CpG methylation (Feng et al., 2010; Hernando-Herraez et al., 2015). This is likely contributing to the lower conversion rate. With this in mind, all samples were processed for further analysis. The ratio of 5mC, 5hmC, and C was calculated for every C residue in the genome using the MethPipe pipeline (Song et al., 2013). This resulted in a β value for every C in the genome for 5mC, 5hmC, and C. The β value represents the frequency each cytosine state was detected in the reads covering each C. For example, if a C in the genome has a β value of 0.6 for 5mC, 0.1 for 5hmC and 0.4 for C, it indicates that in 60% of the reads mapping to that site 5mC was detected, 10% of reads mapping to that site were 5hmC, and 40% of reads were C. The total number of reads covering each site determines how accurate the β value is (Wreczycka et al., 2017). Therefore, to

confidently call 5mC, 5hmC, and C, residues covered by fewer than 10 reads were removed, as recommended by ENCODE standards (Davis et al., 2018). For most samples, this filtered out between 40 – 70% of reads; however, the RPC_27_2 oxBS sample filtered out ~96% of the reads (**Table 4**).

Next, whole-genome content of both 5mC and 5hmC was assessed by averaging the β values of 5mC or 5hmC within the genome of each sample. 5mC content in all C of the zebrafish genome was on average around 8% (**Figure 2A**). Because this method analyzes all Cs, many of which may not be methylated, it was likely why there was lower than the reported 80% genomic methylation in this analysis (Feng et al., 2010; Guo et al., 2014; Song et al., 2013; Ziller et al., 2011). With 5mC occurring more frequently in CpG sites, 5mC content in only CpG sites was next examined, and resulting in genomic 5mC content of 78% (**Figure 2A**), close to the previously reported 80% 5mC (Feng et al., 2010). The slightly lower 5mC averages in these data were likely because BS-seq was used in Feng et al., 2010, which does not distinguish between 5mC and 5hmC, unlike the methods used here which can distinguish 5mC (Booth et al., 2013). To directly compare the 5mC in RPCs to the previous zebrafish study, a BS only experiment was mimicked by subtracting the ratio of unmodified C from one. This was done to directly compare the 5mC frequency in these data to the previous report on zebrafish 5mC (Feng et al., 2010). The data generated in this study had a higher 5mC than previously reported at 83% (**Figure 2B**). This difference, while modest, could be because the data reported by Feng et al, was performed on whole embryos, and there can be tissue variability in 5mC content (Globisch et al., 2010;

Nestor et al., 2012). Therefore, it was not unreasonable for isolated cells, like RPCs here, to have slightly higher 5mC content than that in an entire embryo. Whether or not 5mC increased between 22 and 27 HPF, replicates were pooled and a t-test performed, revealing no significant difference between early and late RPCs (**Figure 2C**).

Next, levels of 5hmC were assessed using similar techniques as above. 5hmC was more frequent in CpG sites than in all C, 4.54-5.48% and 1.3-2%, respectively (**Figure 2D**). The value in all Cs was higher than expected because 5hmC, although highest in neuronal tissues, has only been reported as 0.3%-0.7% of all C residues (Globisch et al., 2010), compared to 1.3-2.11% here (**Figure 2D**). However, when an individual cell type, Purkinje neurons, was investigated by thin-layer chromatography, 5hmC was 40% as frequent as 5mC (Kriaucionis and Heintz, 2009; Kriaucionis, n.d.). Although this was not shown on a per-base basis like the data here, it does provide evidence for cell-type-specific enrichment of 5hmC and suggests up to 40 % of the total 5mC observed in zebrafish RPCs could occur. There appeared to be an increase in 5hmC between 22 and 27 HPF RPC (**Figure 2D**), to assess if there was a change in 5mC or 5hmC between early and late RPCs, replicates were combined for each mark, and a t-test was performed on the pooled data. A significant increase was observed in 5hmC between 22 and 27 HPF (**Figure 2E**). Given that 5mC and 5hmC were enriched in CpG sites, (**Figure 2A, 2D**), the remainder of analyses were performed on CpG sites only.

5mC and 5hmC are biologically dependent on one another, with 5hmC only proven to occur at sites that were previously 5mC (Tahiliani et al., 2009). However, 5hmC has been seen to accumulate at different parts of the genome than 5mC during neural differentiation (Szulwach et al., 2011b; Wu and Zhang, 2011). There could be different 5mC and 5hmC distribution between early and late RPCs. To determine if 5mC and 5hmC were enriched in the same or different regions of the genome, 5mC and 5hmC distributions in genic regions were analyzed. First, CpG islands (CGI) were investigated because CGI are present within many promoters and are reported as a region of low 5mC (Deaton and Bird, 2011). Flanking the CGI is a CGI shore, a 2kb region that has conserved methylation and outside the shore is the CGI Shelves extending to 2kb off the shore (Qu et al., 2012). CGI have less 5mC than their surrounding Shores and Shelves (**Figure 3A**), however, the sample RPC_27_2 had a noticeably higher 5mC level than other samples. 5hmC was lower in CGIs when compared to Shores and Shelves as well (**Figure 3B**) as has been reported (Scourzic et al., 2015).

5mC and 5hmC can also be enriched in different parts of a gene (Szulwach et al., 2011a; Tan et al., 2013), functionally defined here as 5'UTRs, exons, introns and the 3'UTR. Genes are separated by intergenic regions. It has been shown that 5mC or 5hmC have differing impacts on gene expression depending on in which genic region they are found (Williams et al., 2011; Wu and Zhang, 2011). Thus, 5mC and 5hmC accumulation were next analyzed separately in each of these genic components (**Figure 4A**). Within genic regions, promoters have higher 5mC than genes and 3'UTRs

(**Figure 4B**), which was unexpected because promoters are usually reported as hypomethylated (Klose and Bird, 2006). Because promoter 5mC starts high in the promoter and decreases at the TSS, by looking at 5mC over the whole promoter region, the 5mC content was artificially increased (Lee et al., 2015). In the 5'UTR, 5mC and 5hmC were lowest as compared to other genic regions. For 5mC this was expected because low content in the 5'UTR has been reported (Klose and Bird, 2006) (**Figure 4B**). 5mC content was high in exons, introns, and intergenic regions. Genes showed less 5mC likely because the 5'UTR was within the gene body and was driving the decrease. When the average 5mC for the gene was calculated the low values of the 5'UTR were included, bringing down the total gene body 5mC average(**Figure 4A, 4B**).

Compared to other regions in the genome, 5hmC content was moderate in promoters (**Figure 4C**). The enrichment of 5hmC in promoters has been reported as high or low, but always within the context of expression, like 5mC (Tan et al., 2013). Without considering expression levels in the promoter, it was difficult to compare 5hmC content to previous work. Here 5hmC was moderately enriched compared to other regions, but like 5mC, this could also be due to averaging across the entire region or dependent on gene expression, a topic that will be investigated later in this thesis. There are confounding reports on 5'UTR 5hmC content, low 5hmC has been reported at the 5'UTR, as compared to the gene body, 3'UTR and intergenic regions (Madrid et al., 2018), but high 5hmC enrichment has also been reported (Wang et al., 2020). This contradicting result was from a cancer model study, where 5mC and 5hmC patterns are often disrupted, so may have no bearing on the embryonic state (Pfeifer et al., 2014).

5hmC content within the 3'UTR was similar to genes in zebrafish RPCs which was unlike other reports showing lower 5hmC in the 3'UTR compared to genes (Madrid et al., 2018). Within all the exons, introns, 3'UTRs, and intergenic regions the 5hmC content was relatively consistent. Genes had lower 5hmC, but likely also because the gene annotation includes the 5'UTR, which has the lowest 5hmC.

To summarize analyses thus far, revealed an increase in 5hmC between early and late RPCs, but not a significant change in 5mC. When the 5mC content in genic regions was investigated, 5'UTRs, and genes have lower methylation than promoters, exons, introns, 3'UTRs, and intergenic regions. 5hmC content was also low in the 5'UTR and was present at moderate levels in genes, promoters, 5'UTRs, exons, introns, 3'UTRs, and intergenic regions. This analysis brought up several questions. The first being does 5mC or 5hmC change within each of these genic regions. The second was where in the genome does 5hmC increase and the third does the 5mC and 5hmC content change with changing gene expression. Finally, within each of these analyses RPC_27_2, had a higher 5mC content and 5hmC content than the other replicates. There was no clear biological reason for this sample having higher 5mC and 5hmC content. This result was likely due to technical error with only 5% of all CpG represented in the oxBS sample of RPC_27_2. For these reasons and to avoid confounding further analyses, RPC_27_2 was excluded from further studies.

3.3 5mC and 5hmC were distributed throughout genic regions

5mC is known to be distributed through genic regions in a non-uniform pattern due to its inhibitory effect on the binding of many transcription factors and other DNA binding proteins (Klose and Bird, 2006). In addition methyl-binding proteins recognize 5mC and repress transcription (Mahé et al., 2018). In promoters of expressed genes, 5mC starts out higher but decreases as it reaches the TSS; 5mC at the TSS prevents the binding of transcriptional machinery and impairs gene expression (Lee et al., 2015; Mahé et al., 2018). Similarly, 5mC enrichment is a requirement for proper exon recognition during alternative splicing, with MeCP2 binding 5mC at alternatively spliced exons to slow down Pol II elongation and facilitate exon inclusion (Maunakea et al., 2013). Given these enrichment patterns, 5mC and 5hmC distributions within all genic regions were investigated to see how each were distributed, which might suggest potential regulatory roles in modulating gene expression. 5mC and 5hmC distribution were investigated by centering the 5mC or 5hmC data on each genic region of interest. The frequency of 5mC and 5hmC within each region was calculated over 200 bp bins and plotted for all data using the R package enrichedheatmap (Gu et al., 2018).

5mC was low in the CGI and increases within the surrounding CGI Shore, then levels off in the CGI Shelves (**Figure 5A**), as previously reported in the zebrafish embryo (Feng et al., 2010; Lee et al., 2015). Assays on promoters in embryonic zebrafish show 5mC was higher at the start of the promoter and decreases at the TSS, as previously reported (Feng et al., 2010; Lee et al., 2015). RPCs at 22 and 27 HPF

exhibit this same pattern showing higher 5mC content upstream of the promoter and decreasing near the end at the TSS (**Figure 5B**). In 5'UTRs, 5mC was less enriched compared to upstream and downstream sequences (**Figure 5C**). This agrees with other studies in the zebrafish where 5mC is low proximal to the TSS (Feng et al., 2010; McGaughey et al., 2014). Exons had high 5mC when compared to upstream and downstream regions, which include the 5'UTR, 3'UTR and introns (**Figure 5D**). High 5mC in exons has been shown to influence exon inclusion during splicing, where the more 5mC an exon contains, the more likely it is that it will be included in the final transcript (Maunakea et al., 2013; McGaughey et al., 2014). Introns have high and consistent 5mC enrichment, with sharp peaks upstream and downstream of the intron (**Figure 5E**). This pattern has not been reported in zebrafish around introns, but similar patterns have been shown to influence splicing in human cell lines (Maunakea et al., 2013). At the end of the gene, the 3'UTR decreases 5mC from the start to end of the region (**Figure 5F**). 5mC accumulates within the gene body in mouse embryonic stem cells (Anastasiadi et al., 2018; Shi et al., 2017; Tan et al., 2013). In zebrafish RPCs, 5mC was high upstream to the start of the gene but as it approaches the gene start, it drops abruptly and reaches its lowest point immediately before the gene starts, (**Figures 5G**). Entering the gene body, 5mC levels increase sharply and remain high before decreasing after the transcription termination site (TTS) (**Figure 5G**), matching previous reports (Feng et al., 2010; Lee et al., 2015). This result shows that 5mC accumulates in genes compared to upstream and downstream regions. However, none of the genic regions showed an obvious difference between early and late RPCs, changes between timepoints will be investigated and addressed later.

5hmC can be distributed differently throughout the genome based on the potency of the cell type (Tan et al., 2013). To see if there was different 5hmC distribution between early and late RPCs, 5hmC distribution around genic regions was investigated. CGIs showed low 5hmC, but levels increased upstream in the surrounding shores and shelves (**Figure 6A**). Similar 5hmC distributions have been observed in human and monkey brains (Madrid et al., 2018). 5hmC distribution in promoters was reminiscent of 5mC content, with a higher 5hmC content upstream in the promoter, which decreased as it reached the TSS (**Figure 6B**). With 5hmC exhibiting the same pattern as 5mC it suggests 5hmC may negatively impact gene expression, a model that will be investigated in a later section. This is the first report of 5hmC content in zebrafish 5'UTR, and 5hmC content was lower within the 5'UTR when compared to surrounding sequences (**Figure 6C**). 5hmC has been reported in human and monkey brains where there appears to be a slight increase within the 5' UTR (Chopra et al., 2014; Madrid et al., 2018). However, these data may not be directly comparable because it was collected from mature neural tissues, and 5hmC has been shown to increase in aged tissues relative to developing ones (Song et al., 2011). The methods used were also vastly different, identifying 5hmC on an array, which loses the per-base resolution provided by the oxBS and BS (Chopra et al., 2014). Whereas 5mC increased in exons, the opposite was observed for 5hmC, which was lower in exons when compared to surrounding sequences (**Figure 6D**). This relative reduction has not been reported in zebrafish or other systems; in fact, the opposite has been observed, with 5hmC higher in exons than surrounding sequences, which has also been implicated in exon inclusion

(Gao et al., 2019). A decrease in 5hmC has been observed in the first exon in the human and monkey brain, although the gene body containing the remainder of exons still had high 5hmC (Madrid et al., 2018). These discrepancies could be due to 5hmC accumulating throughout differentiation, and so at this early stage any enrichment in exons has not yet occurred (Khare et al., 2012). Introns have consistent and high 5hmC in the immediate upstream and downstream regions surrounding the intron (**Figure 6E**). This pattern of 5hmC decreasing at the exon-intron boundary has also been observed in the mouse brain and is suggested to help distinguish the exon-intron boundary (Khare et al., 2012). This further suggests that 5mC and 5hmC patterning around the exons and introns influences exon recognition. Although an interesting finding, how 5mC and 5hmC may influence splicing will not be investigated in this study. In the 3'UTR, 5hmC increases (**Figure 6F**), an observation that has not been reported before within the 3'UTR, but has been reported to increase downstream of the 3'UTR (Anastasiadi et al., 2018). The function, if any, of this increase in 5hmC is unknown. 5hmC shows accumulation in the gene body (**Figure 6G**), when compared to upstream and downstream sequences, an observation that is consistent with reports showing 5hmC accumulation can lead to gene expression (T. Li et al., 2015; Szulwach et al., 2011b; J. Zhang et al., 2016).

Taken together, these analyses show that 5mC and 5hmC distribution in the genome have particular patterns, informing potential functions. Given the known repressive role of 5mC at promoters and 5'UTRs suggests that 5hmC may also have a repressive role when near the TSS. An accumulation of 5hmC in gene bodies was also

observed, suggesting that like in NPCs (Tan et al., 2013), RPC's gene body content of 5hmC could influence gene expression. Distribution of 5mC and 5hmC within exons and introns at both 22 and 27 HPF implicates them in exon inclusion, consistent with known roles (Khare et al., 2012).

3.4 Changes in 5mC and 5hmC accumulation in neural development genes occur as RPCs mature

The previous analysis showed 5mC and 5hmC were distributed throughout genic regions but did not indicate any difference between early and late RPCs. To determine whether 5mC and 5hmC change between 22 and 27 HPF RPCs, a beta-binomial regression functionality of the MethPipe pipeline was used (Song et al., 2013). This method was performed on 5mC and 5hmC separately to compare time points to find bases significantly changing in 5mC or 5hmC between timepoints (Song et al., 2013). From these analyses, changes in 5mC were termed differentially methylated regions (DMR) and those changing in 5hmC were termed differentially hydroxymethylated regions (DHMR).

DHMR and DMRs were split based on whether they increased or decreased (**Table 5** and **Table 6**.); there were far more regions with decreasing 5hmC overall, than has been previously reported (Tan et al., 2013). There were fewer DMRs compared to DHMRs, which was unexpected because 5mC and 5hmC are biologically dependent

(Tahiliani et al., 2009). 5hmC cannot increase without 5mC decreasing so it was expected to see a similar order of change in DMR and DHMRs. This clearly exhibits a limitation with current methods for analyzing 5mC and 5hmC data. To calculate changing 5mC and 5hmC, each dataset had to be separately analyzed, so potential correlations between changing 5hmC and 5mC were lost. Meaning that a small but significant change in 5hmC between time points which, although present in the 5mC data, was not statistically significant in the 5mC data, so not reported as a DMR (Kochmanski et al., 2019). With no well-established method to overcome this limitation, the regions of changing 5mC and 5hmC were considered independent of each other for the next set of analyses.

To test if the genes associated with changing in 5mC and 5hmC were involved in RPC functions such as proliferation, or RGC formation, such as neural differentiation, GO-term analysis was performed. Specifically, GO-term analysis for biological processes was performed using the clusterProfiler package in the R programming language (Yu et al., 2012). Gene names for the regions changing 5mC and 5hmC were analyzed, resulting in enrichment terms for increasing and decreasing DHMRs, and decreasing DMRs. Those associated with increasing DMRs did not reach the enrichment cutoffs of each GO category containing more than one gene. Genes decreasing in 5mC, increasing in 5hmC, and decreasing in 5hmC were all enriched in biological processes relevant to retinal ganglion cell development, containing terms for axonal development and synaptic signaling (**Figure 7A-C.**). To determine whether the same genes made up these classes and contain both DHMRs and DMRs, the names

were extracted and compared. 15.3% of genes overlap between the DHMRs and DMRs in several combinations (**Figure 7D**). This showed one gene could contain both changing 5mC and 5hmC in combination. Investigating the gene track of one of these genes, *dscama*, shows that DMR and DHMRs can occur in the same region and overlap (**Figure 7E**). Co-occurrence of decreasing DMR and increasing DHMR likely represents active demethylation.

Between 22 and 27 HPF in RPCs many changes in 5mC and 5hmC occur. The changing 5hmC along with decreasing 5mC suggests either an active demethylation process or an accumulation of 5hmC. With DMRs and DHMRs enriched in genes related to axonal development, it suggests those genes will be activated regardless of whether active demethylation of 5hmC accumulation was occurring, and if these were causal or not will need to be tested. To determine if genes changing in 5mC and 5hmC were changing expression, expression data from RPCs during maturation and after differentiation into RGC was investigated

3.5 Few expression changes were present between RPCs at 22 and 27 HPF, but many exist between RPCs and RGCs

To determine whether there were gene expression changes between 22 and 27 HPF RPCs, cells were collected for RNAseq, as described above. RPCs expressing Tg(*vsx2:GFP*) were collected by FACS, RNA was isolated and used to prepare

sequencing libraries. Raw data were aligned to the zebrafish transcriptome GRCz11, and the number of times each gene was detected was calculated. Differentially expressed genes (DEG) between 22 HPF and 27 HPF RPCs were identified using functionality from the limma and edgeR package for R (Ritchie et al., 2015; Robinson et al., 2010). A log₂-fold change of 1.5 and a false discovery rate of 0.02 were used to identify 124 differentially expressed genes, between 22 HPF and 27 HPF (**Figure 8A**, **Table 8**). Of these, 111 were downregulated at 27 HPF, and 13 upregulated in 27 HPF when compared to 22 HPF. To identify the biological processes the DEG between early and late RPCs were involved in, GO-term analysis was performed on the upregulated and downregulated DEGs using clusterProfiler (Yu et al., 2012). Genes downregulated in 27 HPF RPCs compared to 22 HPF RPCs were enriched for developmental processes such as cell adhesion and different types of morphogenesis (**Figure 8B**). Whereas genes upregulated in late RPCs were enriched for categories related to chromatin structure (**Figure 8C**). Neither early or late RPCs had DEG enriched for neurogenesis indicating these cell types were similar in gene expression. Combined the differentially expressed genes and their biological functions, indicate there were few differences in gene expression between early and late RPCs.

To identify genes expressed during the early stages of RGC development, RNAseq was performed on RGCs, isolated by FACS based on Tg(*ato7:GFP*) expression, and DEGs between RPCs and RGCs at 27 HPF were identified using the same methods to compare RPCs. In total, 3789 genes were upregulated and 2602 downregulated in RGCs compared to RPCs at 27 HPF. The top 100 DEGs were plotted

in a heatmap, using the package pheatmap in R (**Figure 9A**) (Kolde, 2012). To see if these genes were involved in processes relevant to RGC differentiation, GO term-analysis was used as previously described. As expected, isolated RGCs upregulate genes involved in ganglion cell growth processes such as “cell adhesion,” “synaptic signaling,” and “cell projection” (**Figure 9B**), much like the DHMRs and DMRs (**Figure 7**). Genes downregulated between RPCs and RGCs were all involved in DNA replication, cell cycle progression, and other terms related to cell division, likely because these cells will be undergoing their final mitosis before differentiation (**Figure 9C**) (Hu and Easter, 1999).

RNAseq analysis revealed that few genes change in expression between early and late RPCs, but many genes were differentially expressed after RGC specification. Although there were few DEG between 22 and 27 HPF RPCs, many changes occurred in 5mC and 5hmC, as evident in the DMRs and DHMRs. Whether or not 5mC and 5hmC change with expression in RPCs as they mature was unclear. If they change with expression, 5mC and 5hmC could change to regulated gene expression globally or have a specific role in regulating the differentially expressed genes. These possibilities were investigated in the next two sections of this thesis.

3.6 5mC and 5hmC changes with expression in Genic Regions

First, to test if there was any connection between the gene expression in RPCs and 5hmC and 5mC, expression levels for all genes detected in the RPCs were analyzed. This was accomplished first by ranking the logTPM values of the RNAseq and splitting it into 10, creating deciles, with decile 10 representing the highest 10% and decile 1 representing the lowest 10% of expressed genes. Genes from the epigenomic data were matched with their counterpart in expression deciles for the promoter, 5'UTR, exons, introns, 3'UTR, and gene body. Once the 5mC and 5hmC data were assigned a decile, the average 5mC or 5hmC content for each decile and region was taken and plotted in a heatmap using ComplexHeatmap in the R programming environment (Gu et al., 2016). This analysis shows that decreases in 5mC in the 5'UTR and promoters occurs as expression increases, but 5mC does not change in genes, exons, introns, and the 3'UTR at either time point (**Figure 10A**). Similar to 5mC, 5hmC decreases in promoters and the 5'UTR as expression increases (**Figure 10B**).

To confirm the changes seen in **Figure 10A**, individual bases making up the deciles in each genic region were plotted into a boxplot and a two-tailed Wilcoxon test performed. At 22 and 27 HPF, both the promoters and 5'UTRs significantly decrease in 5mC as expression increases between deciles 1 and 10 (**Figure 11A, 11B, 11H, 11I**) as seen in the heatmap (**Figure 10A**). Although not clear from the heatmap, a statistically significant change was observed in all regions at both 22 and 27 HPF. Since the two-tailed Wilcoxon test does not indicate the direction of the change, a Wilcoxon one-tailed greater than and less than test was both performed. This showed

the 3'UTR increased in 5mC with expression. 5mC content in genes, exons, and introns all decrease as expression increases (**Figure 11G, 11O**).

Although undetectable within the heatmaps, 5hmC content within introns and genes decreases as expression increases (**Figure 12D, 12F, 12G, 12K, 12M, 12L**). To confirm the direction observed change, a Wilcoxon two-tailed test was also performed, showing a significant change only in the 5'UTRs (**Figure 12B, 12I**). In the 3'UTRs, there was a significant increase in 5hmC as expression increases (**Figure 12E, 12G, 12L, 12O**). There was not an increase in 5hmC within the promoters or genes in early or late RPCs. Given that 5hmC is known to be enriched in genes and promoters with high expression, it was expected to see an increase in 5hmC as gene expression increased in RPCs (D. Li et al., 2015; T. Li et al., 2015; Szulwach et al., 2011a). However other studies have been contradictory with 5hmC being present in promoters of activated or repressed genes (D. Li et al., 2015; T. Li et al., 2015; Szulwach et al., 2011a), indicating tissue dependence of 5hmC distribution. Therefore, in the context of RPCs, these results show that 5hmC in promoters was higher in lowly expressed genes, as in decile 1.

Taken together, these analyses demonstrated that 5mC and 5hmC change globally with expression in RPCs, but 5hmC does not increase with expression except for the 3'UTR. Specifically, as expression increases in RPCs, 5mC decreases in 5' UTR, promoters, genes, exons, and introns (**Figure 10A, 11A, 11H, 11B, 11I**). The 3'UTR increases 5mC content with expression. 5hmC, on the other hand, does

decrease within the 5'UTR, but not in the promoters. The remaining genic regions, the exons, intron, and genes also decrease in 5hmC as expression increases on a global scale. These experiments show that 5mC and 5hmC change with the expression globally but does not answer whether 5mC and 5hmC changes observed in the DMR and DHMRs were correlated to genes differentially expressed between 22 and 27 HPF RPCs.

3.7 DEGs in RPCs were associated with changes in 5mC and 5hmC

With both 5mC and 5hmC increasing and decreasing, and genes differentially expressed between 22 HPF and 27 HPF RPCs, these changes may work together to drive gene expression. To determine if epigenetic and gene expression changes were dependent in RPCs, it was first asked if any DEG contained either a DMR or DHMR. Many DEGs contain a DHMR, specifically in genes, but few regions contained a DMR (**Table 8**). Next, to test if changes in gene expression were independent of changes in 5mC or 5hmC, a Fisher's exact test was performed. Each gene was assessed for whether it was differentially expressed between RPCs, and if it contained either a DMR or DHMR, within one of the genic regions.

A contingency table for each genic region with either 5hmC or 5mC and DEG was generated; examples of a contingency table are in **Table 9** and **Table 10**. A Fisher's exact test was run on each of these tables and showed expression was only

dependent on DMRs located in exons and introns, but the odds ratio being less than one indicates changing 5mC in exons or introns have a negative association with expression (**Table 11**). DEGs between RPCs at 27 and 22 HPF were not dependent on DHMRs or DMRs. Without evidence to support DEG dependence on DMRs and DHMRs, the change between early and late RPCS may serve a different function. Several other studies have shown that increasing 5hmC can prepare genes for activation and expression at later stages (Tan et al., 2013; Wu and Zhang, 2011; Xu et al., 2011).

3.8 5hmC primes genes for activation after RPC restrictions

To determine whether changing 5mC and 5hmC between early and late RPCs were related to DEG after RPC restriction to RGC fate, DEGs between RGCs and RPCs were compared to the genes containing DHMRs. Increasing DHMRs were tested specifically because they could indicate either active demethylation or accumulation of 5hmC in genes, decreasing DHMRs and DMRs only inform active demethylation. Of the genes increasing in 5hmC and upregulated in RGCs, 512 genes overlapped (**Figure 13A, Table 12**). To identify the biological processes in which these genes were involved, GO-Term analysis was performed on overlapping genes as previously described. Genes that overlap were all involved in axon development and synaptic signaling, functions that is critical for RGCs, suggesting 5hmC may be priming genes for activation (**Figure 13B**) (Szulwach et al., 2011; Xu et al., 2011). Genes increasing in 5hmC were expressed in RGCs and involved in axonal development, but a dependence

between 5hmC gain and expression was not determined. To test for dependence, a contingency table of change in gene expression and change in 5hmC was created (**Table 13**), and a Fisher's exact test was performed. This test indicates a dependence between changing 5hmC in the gene body (as well as exons, introns and the 3'UTR contained within) and changing expression in RGCs (**Table 14**). Next genes that were specifically upregulated in RGCs compared to RPCs were tested for dependence with increasing 5hmC between early and late RPCs, by creating a contingency table (**Table 15**). A Fisher's exact test was performed resulting in a p-value that rejected the null hypothesis, suggesting there was an association between increasing expression in RGCs and increasing 5hmC (**Table 15, and Table 16**). The odds ratio was greater than one, indicating that there was a positive association between the two. Together these data suggest that gain in 5hmC was a feature of genes that will be activated after RPCs were restricted to ganglion fates.

In summary 5mC and 5hmC were distributed through genic regions, which in some cases appear to be specific to zebrafish RPCs, and these data suggest that in each genic region, 5mC and 5hmC may have a functional role. Between early and late RPCs there were many changes occurring in 5mC and 5hmC content, but few transcriptional changes, and these were not dependent on the 5mC and 5hmC content. Although 5mC and 5hmC generally decrease as expression increases, they do not inform the genes that were currently being differentially expressed. Instead, the regions changing 5hmC and 5mC content were predictive of what genes will be expressed after RPC restriction to ganglion cell fates.

4.0 Discussion

RPCs are a multipotent cell population that gives rise to all retinal cell types. As retinogenesis progresses, RPCs from different timepoints generate subsets of retinal neurons, but how this is regulated is unknown. Our lab has recently show that in *tet2^{-/-};tet3^{-/-}* zebrafish mutants, which cannot convert 5mC to 5hmC, have impaired terminal retinal neural differentiation. In *tet2^{-/-};tet3^{-/-}* zebrafish mutants all the retinal cell types are specified but terminal differentiation does not occur and gene expression is disrupted, particularly in genes related to terminal differentiation, suggesting 5hmC contributes to gene regulation (Seritrakul and Gross, 2017). Previous work in our lab has shown that the conversion of 5mC to 5hmC is an important feature of retinal neurogenesis but did not identify where in the genome 5hmC is distributed and how it is related to gene expression. Throughout differentiation, 5hmC has been found to have a complex tissue-dependent relationship with gene expression. During NPCs differentiation, enrichment of 5hmC in promoters and genes can show positive, negative, or no correlation to gene expression (T. Li et al., 2015; Szulwach et al., 2011a; J. Zhang et al., 2016). In cases of no correlation to gene expression, 5hmC can correlate better with genes that will be expressed later in development, suggesting 5hmC primes genes for activation (Szulwach et al., 2011a). A gene is considered primed when a repressive modification such as 5mC is removed, but gene expression is still low. Priming is thought as a component of lineage restriction, occurring in uncommitted cells, to prepare genes for expression once the cell has committed to a lineage (Bonifer and Cockerill, 2017).

Although how 5hmC regulates genes expression is still under investigation, 5hmC has been shown to be critical for terminal differentiation of neurons (T. Li et al., 2015; Santiago et al., 2020b). Within the retina, 5hmC is also needed during terminal differentiation, with *tet2^{-/-};tet3^{-/-}* zebrafish mutants lacking 5hmC compared to wild-type and failing terminal neurogenesis (Seritrakul and Gross, 2017). How the lack of 5hmC prevented terminal differentiation in the retina was unclear, but where in the genome 5hmC is distributed on genes expressing during RPC differentiation could shed some light on this. In this thesis, the distribution of 5mC and 5hmC was investigated to determine where it was patterned during RPC differentiation and its relationship to gene expression during retinogenesis.

The data generated and analyzed in this thesis indicate that 5hmC negatively regulates gene expression in early and late RPCs. When 5mC and 5hmC distribution were plotted within promoters, 5'UTR, CGI, and genes, 5mC and 5hmC showed nearly identical distribution patterns. 5mC has been shown to repress gene activity in these regions either by preventing transcription factor binding or recruiting other repressive factors. Therefore within these regions 5mC must be low for expression to occur (Klose and Bird, 2006). With 5hmC showing the same distribution as the 5mC pattern, it suggests that 5hmC was also repressive and must be removed for gene expression to occur. In addition, if 5hmC was not repressive and its increases leads to genes expression in RPCs, it would have been expected to see some accumulation in or around promoters, as observed by Szulwach et al. (Szulwach et al., 2011a). If the

distribution of the highest expressed genes were plotted it may be possible to observe an accumulation of 5hmC in the promoters, but it was unlikely that an accumulation would be found because 5hmC decreases as expression increases (**Figure 10**). 5mC and 5hmC content within introns and exons has not been correlated to increased expression; however, the distribution of these within the RPCs suggests a role in alternative splicing.

The sharp spikes of 5mC flanking introns, high 5mC in exons and the drop in 5hmC content, in both early and late RPCs suggest that 5mC and 5hmC may be associated with alternative splicing (**Figure 5D, 5E, 6D, 6E**). Included exons have higher 5mC than excluded exons, and high 5hmC can denote the intron-exon boundary (Khare et al., 2012; Maunakea et al., 2013). RPCs exhibit a similar distribution of 5mC and 5hmC at the intron-exon boundary suggesting 5mC and 5hmC may define which exons are included in genes expressed in early and late RPCs. In addition, many DMRs and DHMRS were identified in introns and exons but whether they are located at the boundary was not assayed. Whether 5mC and 5hmC influence splicing could be determined by reanalyzing the data on a transcript basis and may identify more transcriptional differences between early and late RPCs.

Gene based analysis was performed because changes between expressed genes is easier to interpret biologically than transcript based analysis, particularly when the significance of different isoforms is unknown (Soneson et al., 2015). Between early and late RPCs, there were only 124 differentially expressed genes. These were not

enriched in a particular GO categories that could be reasonably explained based on a reported difference between early and late RPCs. This indicates early and late RPCs are not significantly different based on gene expression. However, the genes that were downregulated between 22 and 27 HPF were enriched in different morphogenic categories, which could be due to the eye finishing its invagination and evagination by 27 HPF (Schmitt and Dowling, 1994). In addition, the enrichment for chromatin remodeling in upregulated genes could be explained by the chromatin changing confirmation to allow for neurogenic gene activation after differentiation. Another explanation for these differences could be the from subpopulations of RPCs identified in single-cell RNAseq of RPCs, at 24 HPF based on the expression of marker genes (Xu et al., 2020). This shows that there are some transcriptional differences in RPCs at a single-cell level, which likely explains some of the differentially expressed genes identified. Integrating the differentially expressed genes to the 5mC and 5hmC data and performing a Fisher's exact test, did not indicate a relationship between the DEG and change in 5mC or 5hmC. No association between 5mC or 5hmC was observed, making it unlikely 5mC and 5hmC directly regulate the DEG, although there was an overall repressive trend of 5mC and 5hmC with global gene expression.

Integrating the expression data with 5mC and 5hmC content, supported a negative regulatory role for both 5mC and 5hmC in relation to gene expression (**Figure 10**). In genes exhibiting the highest expression, there was less 5mC in all genic regions investigated except for the 3'UTR, which showed increased 5mC content. How increasing 5mC in the 3'UTR could lead to an increase in gene expression was unclear,

but a positive correlation between 5mC and gene expression has been reported in both cancers and T-cells (McGuire et al., 2019). Within gene bodies it was unexpected to see the repressive trend with gene expression but could be due to the 5'UTR, which decreases in 5mC content, being included within the gene body designation. Except for the 3'UTR, 5mC decreases with gene expression and was consistent with its established role as a negative gene regulator.

When plotted based on expression, there was also a decrease in 5hmC in the 5'UTR, introns, and genes as expression increases (**Figure 10B**, **Figure 11**), indicating 5hmC was likely repressive in these genic contexts. Genes decreasing in 5hmC content as expression increased was unexpected but further supports 5hmC as a repressive mark in RPCs. 5hmC also increased with expression in the 3'UTR at both time points and in exons at 27 HPF (**Figure 10B**, **Figure 12**). This increase in 5hmC in the 3'UTR of more highly expressed genes has not been reported before but could be related to the higher 5mC in highly expressed genes. The 5mC increase in exons could be related to changing distribution during splicing, but this was not confirmed. Based on these data it appears that 5hmC negatively regulates gene expression within most of the genic regions. However, looking at regions changing in 5mC and 5hmC between time points, there was enrichment for neurogenic genes (**Figure 7A-C**). Surprisingly there were fewer DMRs detected than DHMRs (**Table 5 and Table 6**) even though there is a biological dependence between 5mC and 5hmC (Tahiliani et al., 2009). This is likely because changes between 5mC and 5hmC in early and late RPCs were detected separately. 5mC content in zebrafish RPCs can range from 16% in the 5'UTR

to 80% in introns (**Figure 4B**), indicating that large changes can occur in 5mC content. On the other hand 5hmC ranges from 2.5% in the 5'UTR to 5.5% in introns (**Figure 4C.**), indicating smaller changes likely occur in 5hmC. By identifying DMRs and DHMRs separately, the larger, more extreme changes in 5mC are called as DMR, whereas the comparatively smaller changes in 5hmC are identified as DHMRs. If DMR and DHMR could be identified together, it would be expected that more small changes in 5mC would also be identified, but is a limitation to current analysis (Kochmanski et al., 2019). Combining the 5hmC changes with gene expression from RGCs, suggests that increasing 5hmC primes genes for activation after differentiation occurs.

Comparing the expression data from RGCs to the RPCs at 27 HPF, many genes were upregulated and involved in neurogenesis (**Figure 9B**), and contained an increasing 5hmC (**Figure 13A, 13B**). This suggests that 5hmC gain within these genes was co-occurs with gene activation in differentiated cells. A similar trend was observed in mouse cerebellum with 5hmC accumulating on genes that will be activated in later stages of development but did not correlate to genes expressed before differentiation (Szulwach et al., 2011a). It has been suggested that 5hmC serves to prime genes for activation, keeping them repressed but ready for activation after differentiation.

Not only were neurogenic genes show to increase in 5hmC content, but by comparing the DEG between RPC and RGCs to increasing DHMRs, it was also found that they are statistically dependent on each other by a Fisher's exact test (**Table 14**). The number of genes containing both increasing 5hmC and changing gene expression

are statistically unlikely to occur by chance, suggesting a biological significance to this finding. Thus, it seems likely 5hmC increases on genes that will be active once differentiation occurs, but whether the 5hmC was maintained after differentiation was unclear from these data. Whether 5hmC was generated by active demethylation or accumulated and maintained cannot be supported without a 5mC and 5hmC dataset for RGCs. In either case, by converting the 5mC to 5hmC these genes are being poised for activation.

5hmC may serve to prime genes themselves for activation but may also prime enhancers. Although not further investigated here, many DHMRs were within intergenic regions, some of which likely represent enhancers. Enhancers serve as a long-distance on/off switches for gene expression during development and help mediate precise spatial-temporal gene regulation (Furlong and Levine, 2018). 5hmC has been found enriched with histone modifications that together label active and poised enhancers (Verma et al., 2018). Therefore, it was possible that with 5hmC gain in the intergenic regions, 5hmC was generated on enhancers priming them as well. Enhancers likely play a role during RPC differentiation and 5hmC priming them should be investigated but requires additional datasets to definitively identify enhancers.

The data generated here support 5hmC as a repressive mark that was priming neurogenic genes for activation once differentiation begins. A similar phenomenon occurs in NPCs during differentiation to neurons with 5hmC found on the promoters in neurogenesis genes. When Tet3 is knocked down by shRNA in mESC, these same

genes were hypermethylated and neural differentiation was impaired (Santiago et al., 2020b). Interestingly with Tet3 knockdown, proliferation increases and pluripotency genes, *Oct4* and *Nanog* were derepressed, making the NPC more ESC like. In *tet2*^{-/-}; *tet3*^{-/-} zebrafish mutants, RPCs also had impaired neural differentiation and increased proliferation. This suggested 5hmC is needed to pattern the genome for neural differentiation and absence of 5hmC makes cells more stem-cell like. Impaired neural differentiation was observed in the *tet2*^{-/-}; *tet3*^{-/-} zebrafish but lacked information on 5hmC distribution (Seritrakul and Gross, 2017). The genome-wide 5hmC content generated here shows 5hmC was gained in neurogenesis genes. Thus, it can be speculated that in *tet2*^{-/-}; *tet3*^{-/-} zebrafish RPCs 5hmC was not generated in genes which remain fully repressed by 5mC and inactive during differentiation. However, when *tet2*^{-/-}; *tet3*^{-/-} zebrafish blastomeres were transplanted to wild-type embryos, all retinal cell types differentiated. It can be argued that without 5hmC cells are in a more ESC-like state, like Tet3 deficient NPCs, and competent to all cell types if the proper extracellular environment was present to direct the cells. However, there are alternative possibilities for the role of 5hmC during retinal neuron differentiation.

In addition to the differentiation defects in the *tet2*^{-/-}; *tet3*^{-/-} zebrafish, Wnt and Notch signaling were overactive (Seritrakul and Gross, 2017). This indicates that *tet2* and *tet3* modulate the activity of Wnt and Notch signaling in the embryo, but whether signaling is disrupted directly from lack of 5hmC or another role of *tet2* and *tet3* independent of 5hmC generation is unclear. The extracellular signaling defect in the *tet2*^{-/-}; *tet3*^{-/-} mutants is not conducive to retinal neuron differentiation, with wild-type cells

transplanted to the *tet2*^{-/-};*tet3*^{-/-} zebrafish failing to differentiate, suggesting that the lack of 5hmC is not the only factor limiting neurogenesis in the *tet2*^{-/-};*tet3*^{-/-} mutants (Serittrakul and Gross, 2017). However, the 5hmC pattern in transplanted cells was not investigated. It is possible that the signaling defects prevented normal 5hmC patterning from occurring in the transplanted cells, limiting differentiation or the defect could be independent of 5hmC. This could be tested by inhibiting Notch and Wnt in wild type zebrafish, collecting RPCs and determining if 5hmC is still generated within the developing retina. Further supporting the complexity of the differentiation defect, RGCs can be rescued in *tet2*^{-/-};*tet3*^{-/-} mutants when Wnt and Notch pathways are inhibited but not terminally differentiated photoreceptors. Thus 5hmC may be more critical to differentiation of certain retinal neurons. This could be investigated with targeted *tet2* and *tet3* knockout in specified RPCs for different cell types.

From the 5mC and 5hmC data generated here and the *tet2*^{-/-};*tet3*^{-/-} mutant studies, it is clear that 5hmC is a necessary component of retinal cell differentiation. However, whether or not 5hmC itself serves a function or is simply generated as a component of active demethylation is unclear. Tet proteins are a critical component of active demethylation, without functional tets in the *tet2*^{-/-};*tet3*^{-/-} mutants, active demethylation does not occur and 5hmC cannot accumulate on genes. Whether 5hmC accumulates on genes to drive gene activation or is generated only as a result of active demethylation during differentiation is unclear and these possibilities cannot be separated without 5mC and 5hmC data from differentiated cells. By analyzing a differentiated RGC population these possibilities could be better distinguished.

In conclusion 5hmC was generated on neurogenic genes between early and late RPCs and appears to prime them for activation in differentiated neurons. Between early and late RPCs, 5hmC increases but was distributed like repressive 5mC throughout the genome suggesting 5hmC negatively regulated gene expression in RPCs. In addition, both 5mC and 5hmC decreases on a global scale as gene expression increases. Specific regions can be identified with increasing 5hmC between timepoints and were enriched in neurogenic genes many of which were upregulated in RPCs. Indeed, the number of genes that were gaining 5hmC and expressed in RGCs was significant indicating that 5hmC must be accumulated on these genes for activation. The data presented here support 5hmC as a repressive mark that primes neurogenic genes for activity during retinal differentiation.

5.0 Figures

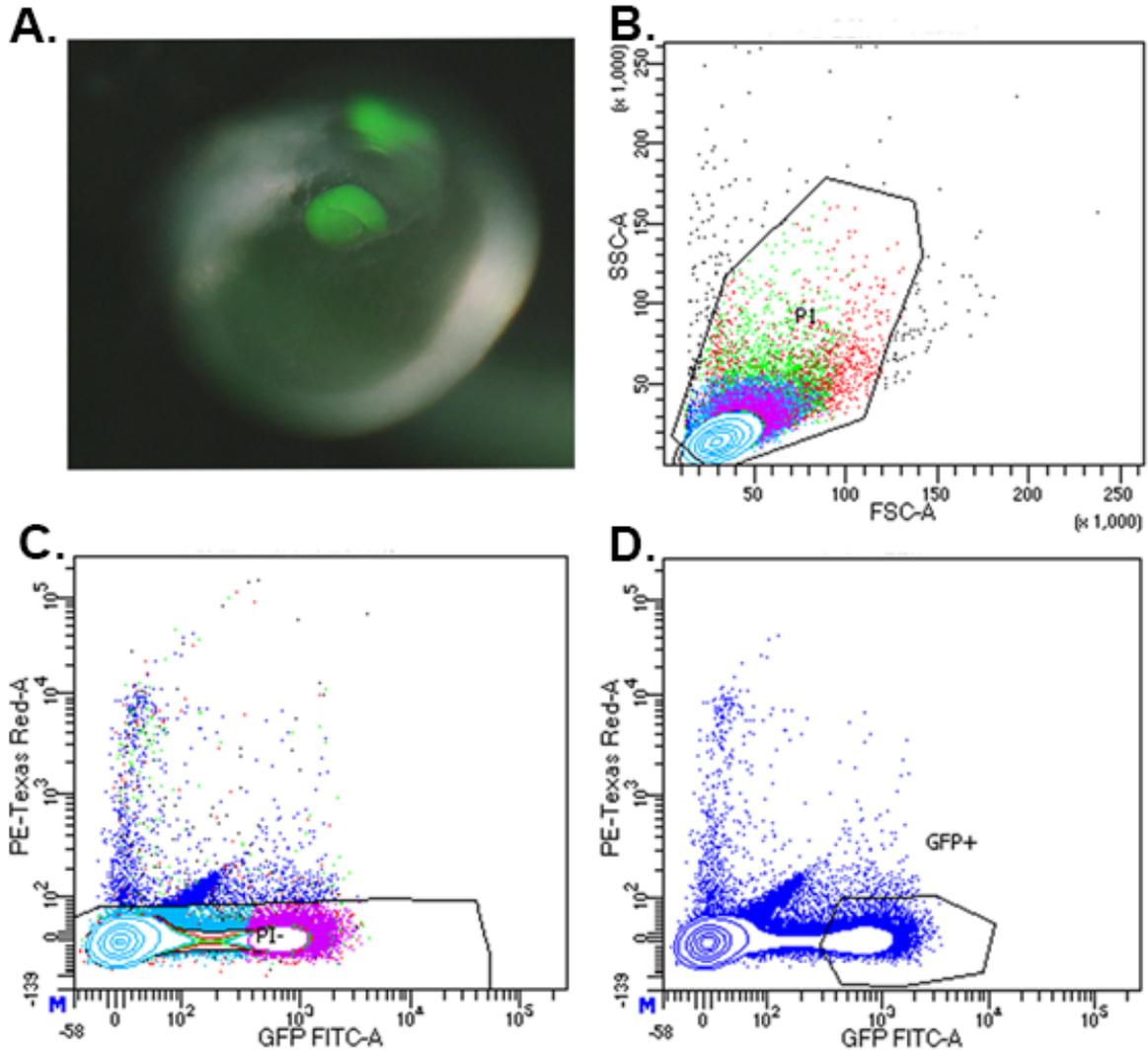


Figure 1 Collection of RPCs. A. Representative image of a *Tg(vsx2:GFP)* embryo used to collect RPCs at 22 HPF. B. Forward-Scatter Side-Scatter gating to select single cells at 22 HPF. C. Gating for selection against dead cells. D. Selection of GFP+ cells for collection.

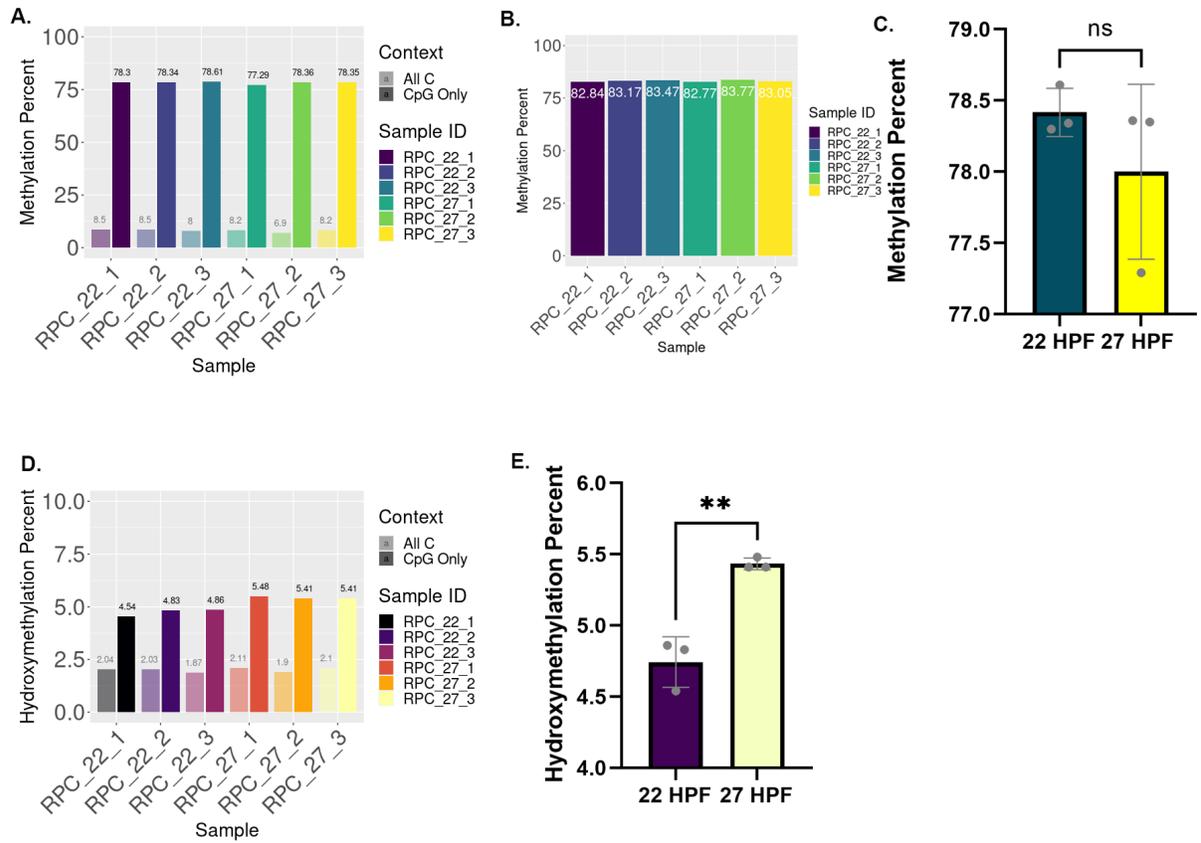


Figure 2. 5mC and 5hmC distribution in the zebrafish genome was enriched in CpG Residues **A.** 5mC in all C residues compared to CpG sites. **B.** Mock WGBS data 5mC level when 5mC and 5hmC are not separated. **C.** Change in 5mC between 22 and 27 HPF was not significant **D.** 5hmC in all C residues compared to only CpG residues. **E.** Combined replicates of 5hmC between 22 and 27 HPF testing for change. (* P-value = 0.0027)

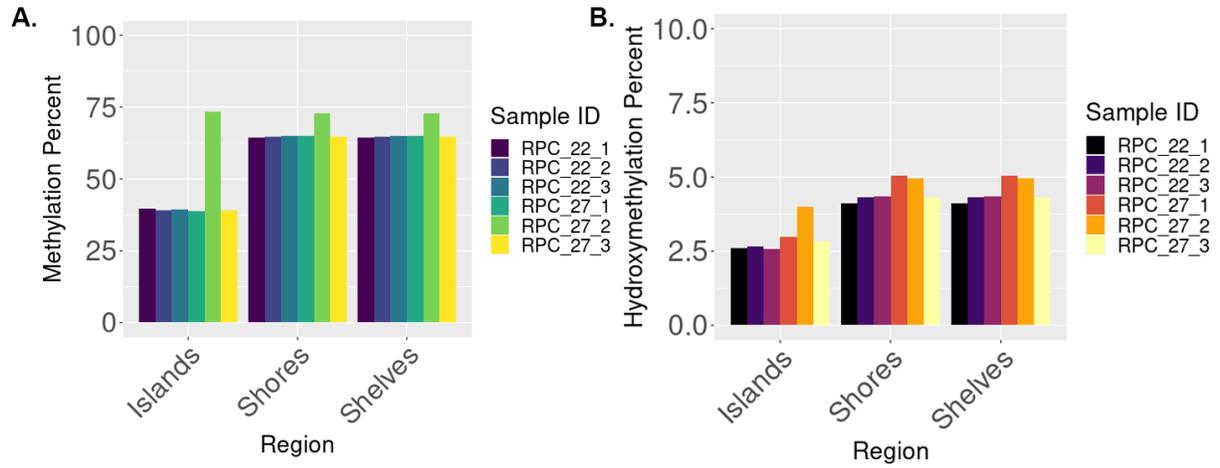


Figure 3. 5mC and 5hmC content in CGI A. 5mC in CGI, CGI Shores, CGI Shelves.
B. 5hmC frequency in CGI, shores, and shelves.

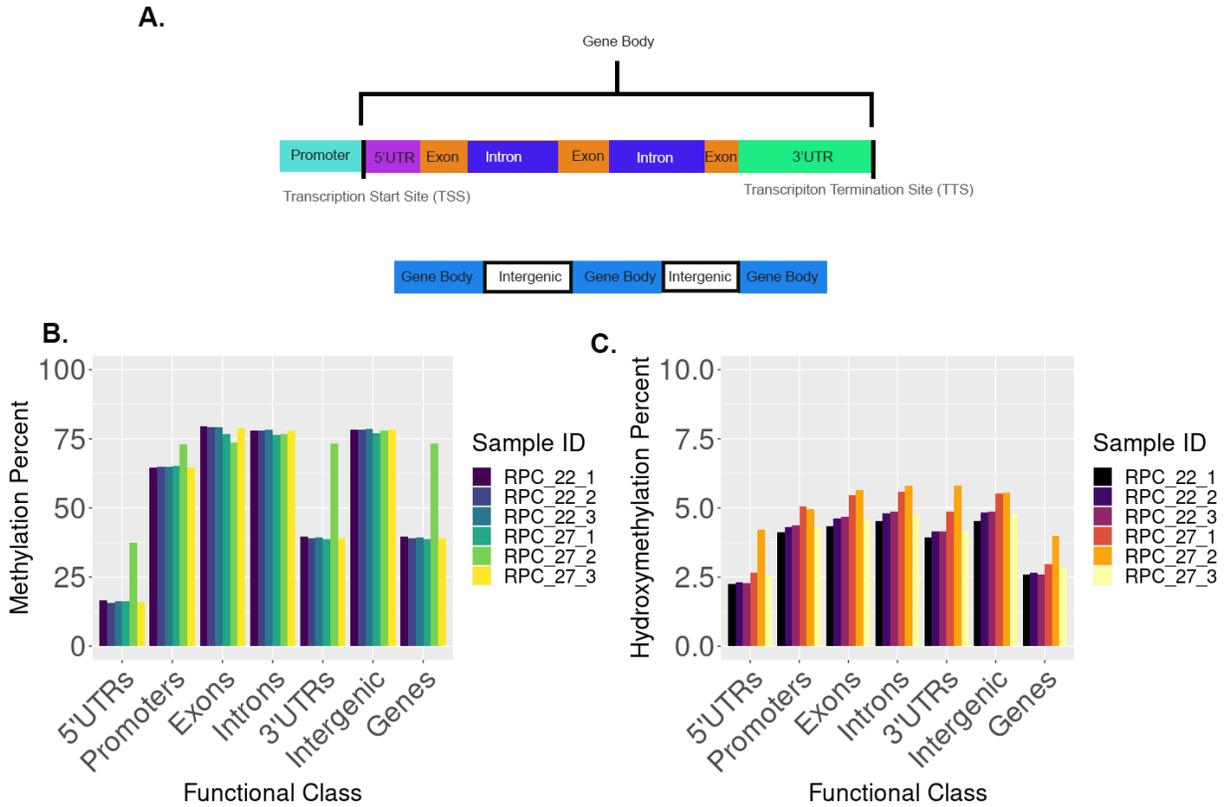


Figure 4. 5mC and 5hmC were present in different genic regions **A.** Schematic of genic regions analyzed for 5mC and 5hmC content. **B.** 5mC content in genic regions showing high 5mC in promoters, exons, introns, and intergenic regions. 5'UTRs, 3'UTRs, and genes have lower 5mC. RPC_27_2 shows inconsistent 5mC content compared to other samples. **C.** 5hmC content in genic regions. Promoters and genes have lower 5hmC than remaining genic regions. Samples RPC_27_1 and 27_2 have higher 5hmC content than other samples

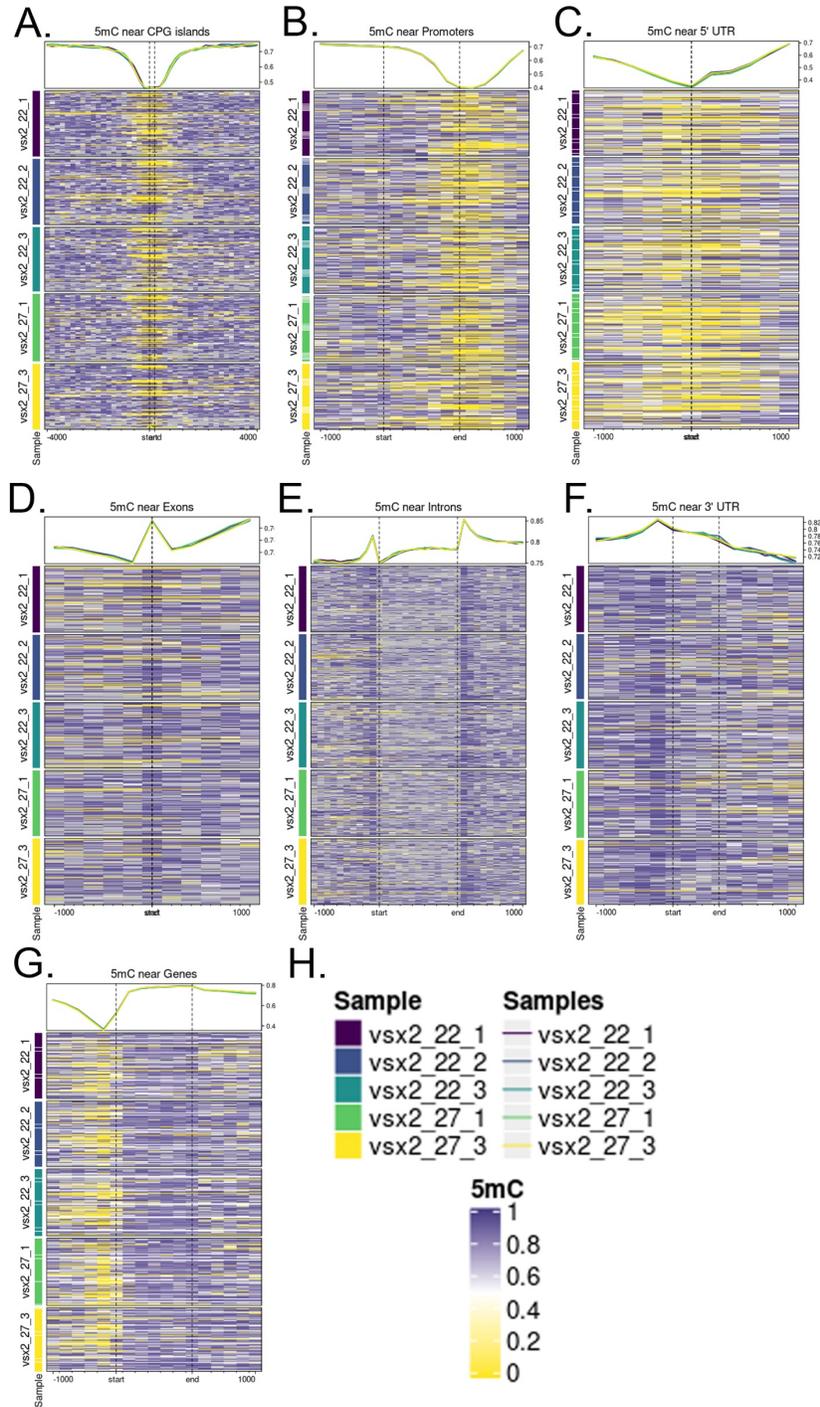


Figure 5. 5mC was not uniformly distributed around genic region. A. 5mC in CGI was low and increases in surround CGI Shores and CGI Shelves. **B.** In the promoter, 5mC starts high and decreases as it reaches the TSS. **C.** The 5'UTR has low 5mC and was higher upstream and downstream. **D.** Exons have high 5mC compared to

surrounding sequences. **E.** Regions of high 5mC flank introns. **F.** 5mC decreases throughout the 3'UTR. **G.** Gene body 5mC starts low at the TSS then increases and stays high throughout the gene body. **H.** Key for heatmaps and line plots, indicating samples and dynamic range.

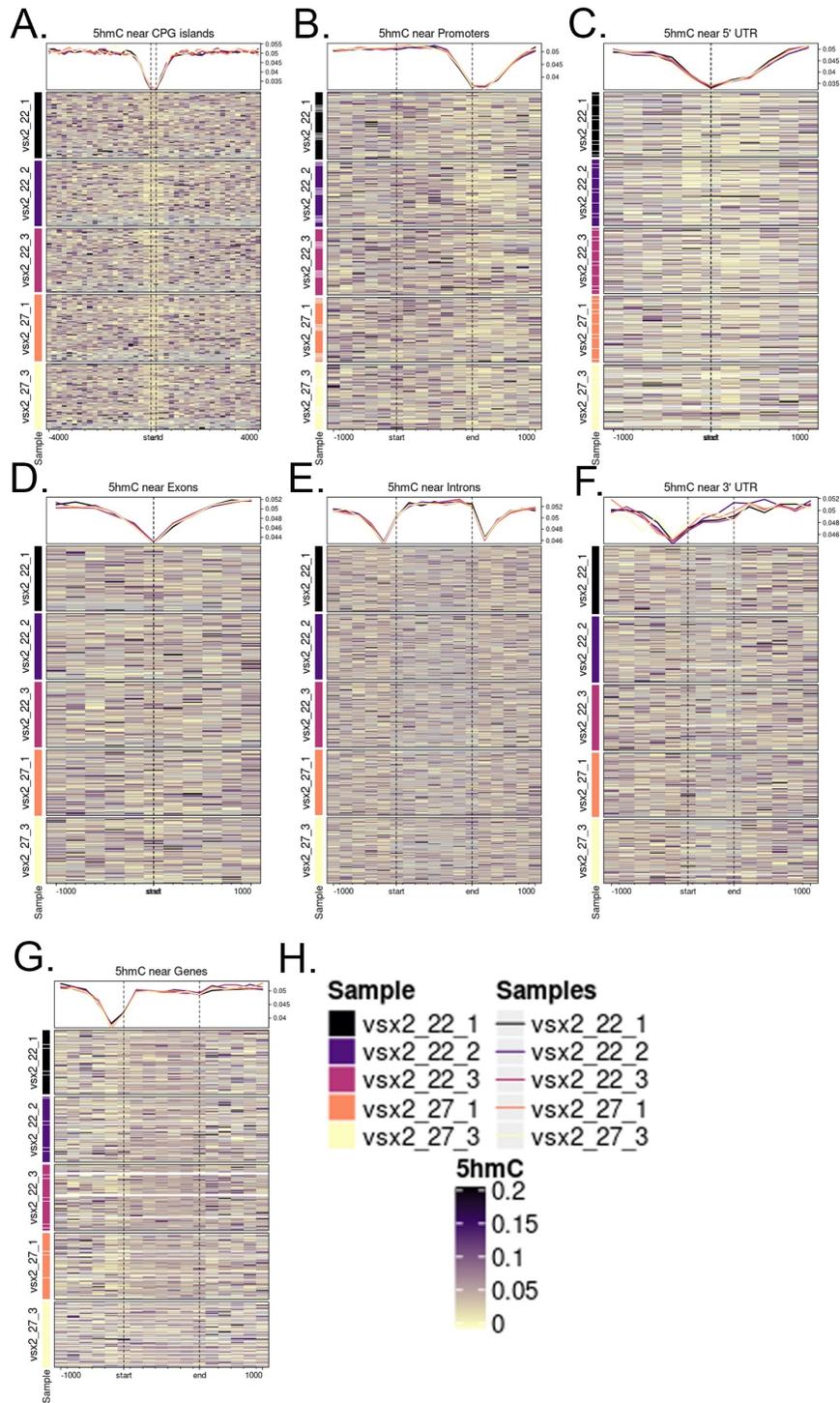
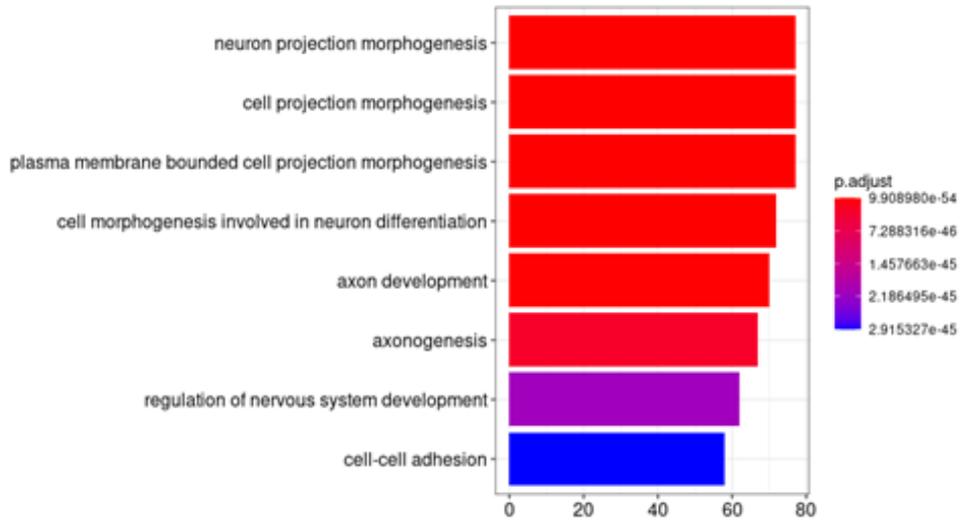


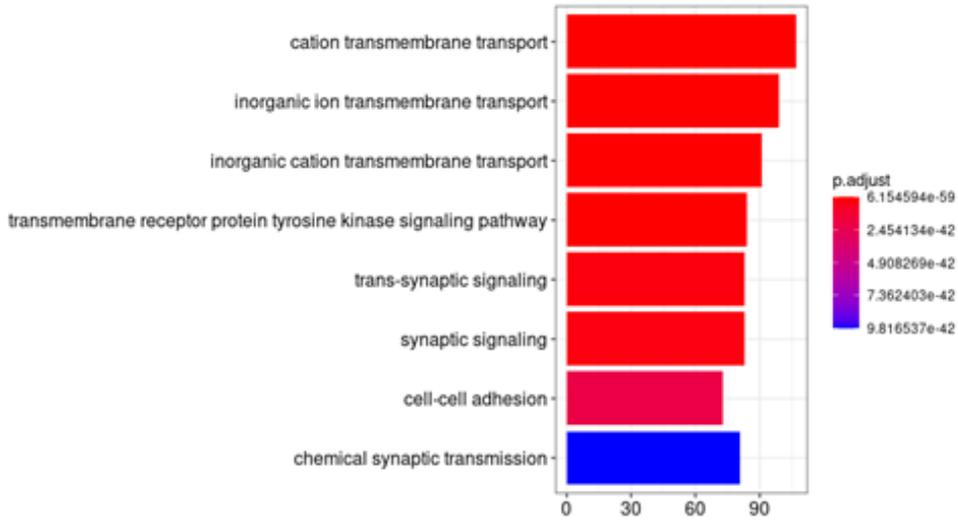
Figure 6. 5hmC was patterned around genic regions. A. 5hmC in CPG was low and increases in Surround CGI Shores and CGI Shelves. **B.** In promoters, 5hmC starts high and decreases around the TSS. **C.** 5hmC was low in 5'UTR compared to surrounding regions. **D.** Exons sharply decrease in 5hmC compared to surrounding sequences. **E.**

Introns have high 5hmC compared to upstream and downstream sequences. **F.** 3'UTRs increase in 5hmC. **G.** Genes have low 5hmC around the TSS, but high 5hmC throughout the gene. **H.** Heatmap key indicating sample colors and dynamic range of the heatmap.

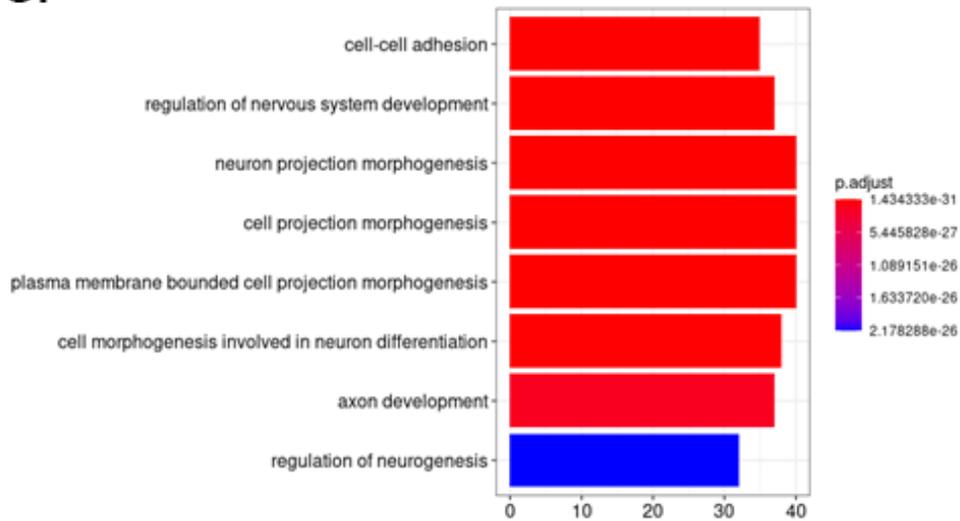
A.



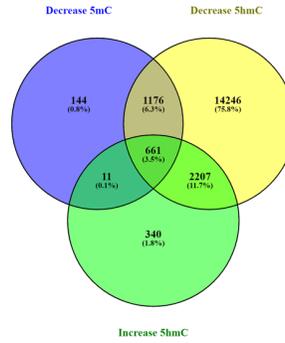
B.



C.



D.



E.

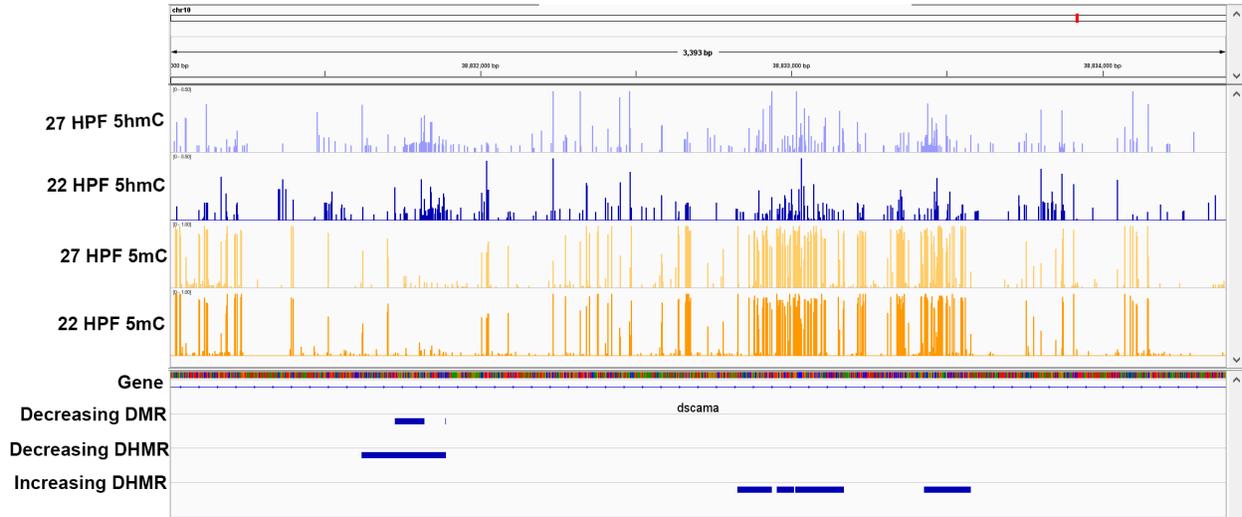
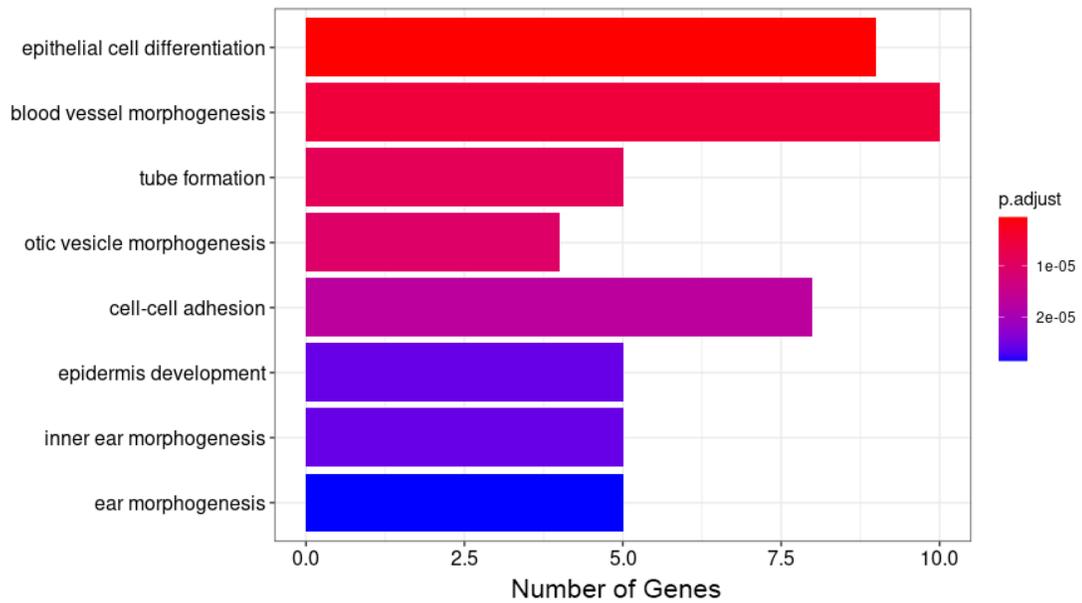


Figure 7. Genes containing increasing and increasing DHMRs and decreasing DMRs are Enriched for Neural Genes. A. GO Term of genes containing decreasing 5mC, number of genes indicated in the x-axis and p values by color. **B.** GO Terms of genes containing increasing 5hmC. **C.** GO terms of genes containing decreasing in 5hmC. **D.** Representative track of a gene containing Increasing and decreasing DHMRs and decreasing DMRs. **E.** Venn diagram of the number of genes containing increases 5hmC, decrease 5hmC and decreasing 5mC

B.



C.

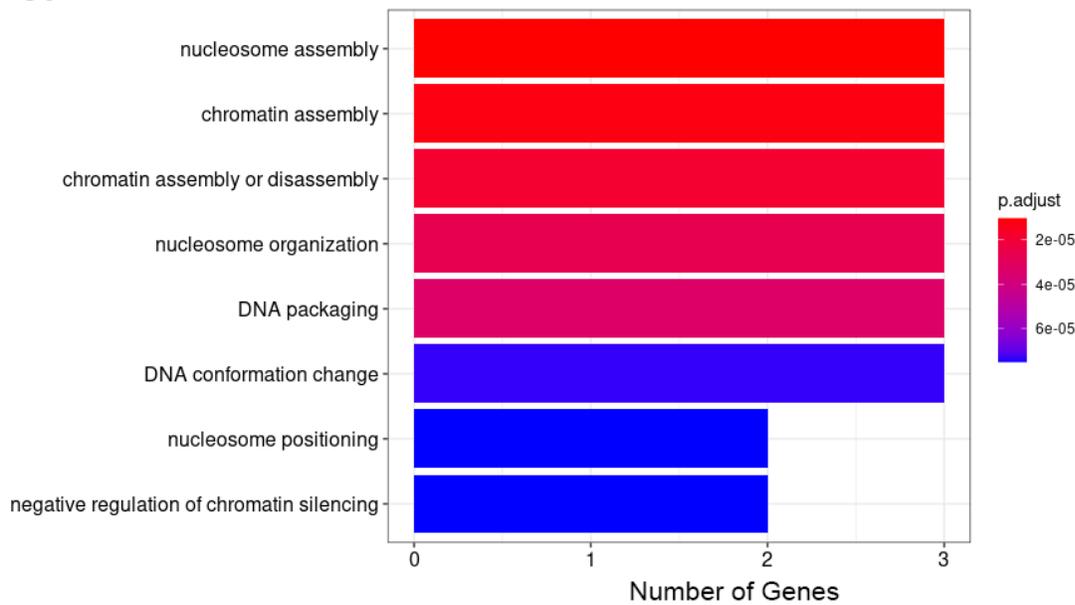
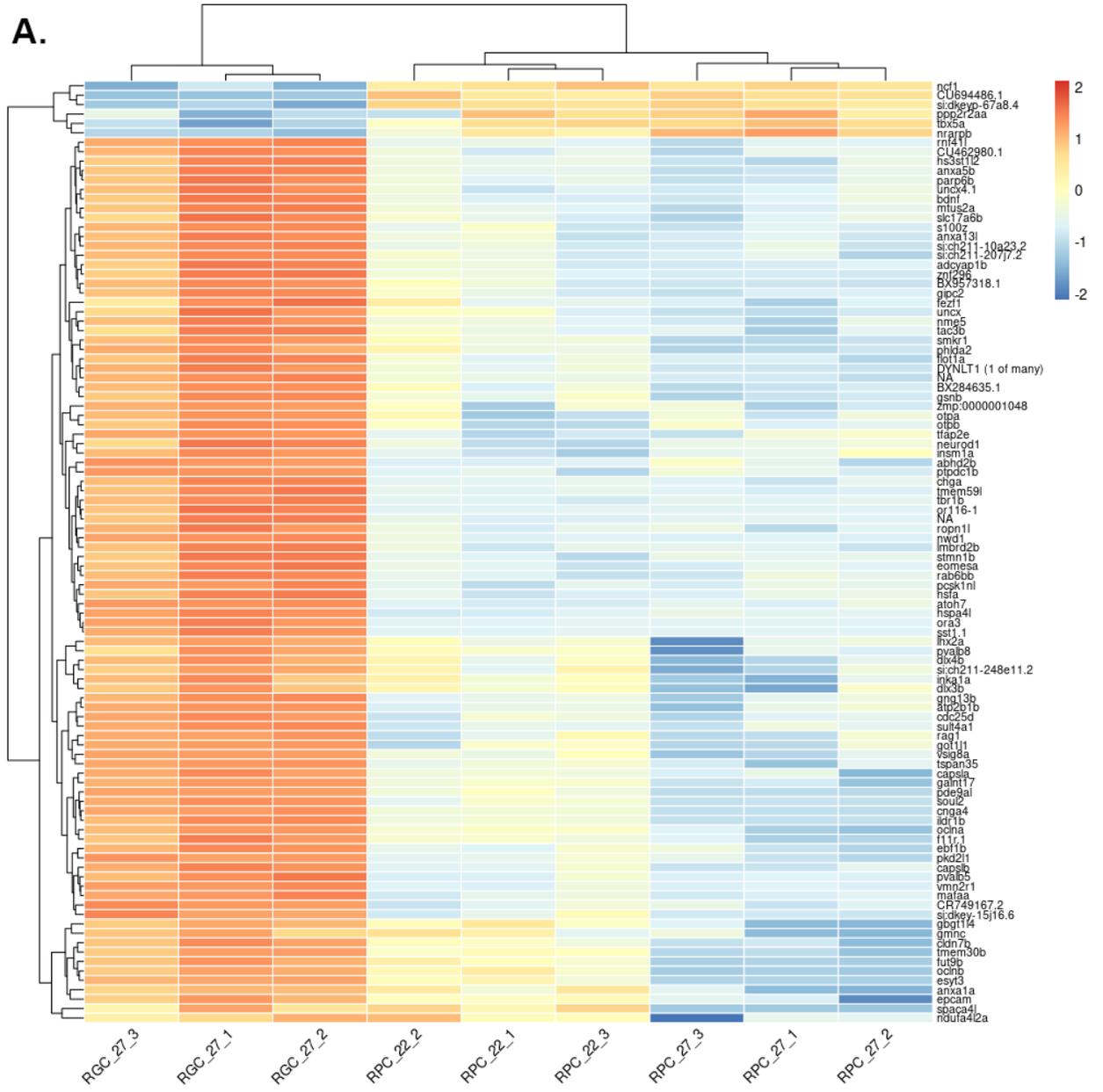


Figure 8. Differentially expressed genes between early and late RPCs indicate few differences between cells. A. heatmap of top 100 differentially expressed genes between 22 and 27 HPF RPCs. **B.** Go-terms of genes increases between 22 and 27 HPF. **C.** Go terms of genes decreasing in expression between early and late RPCs.

A.



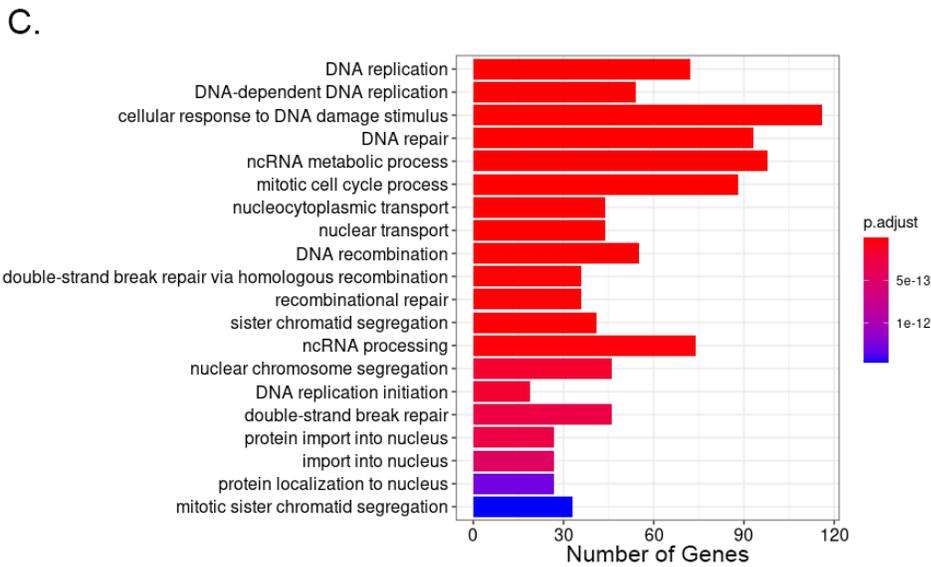
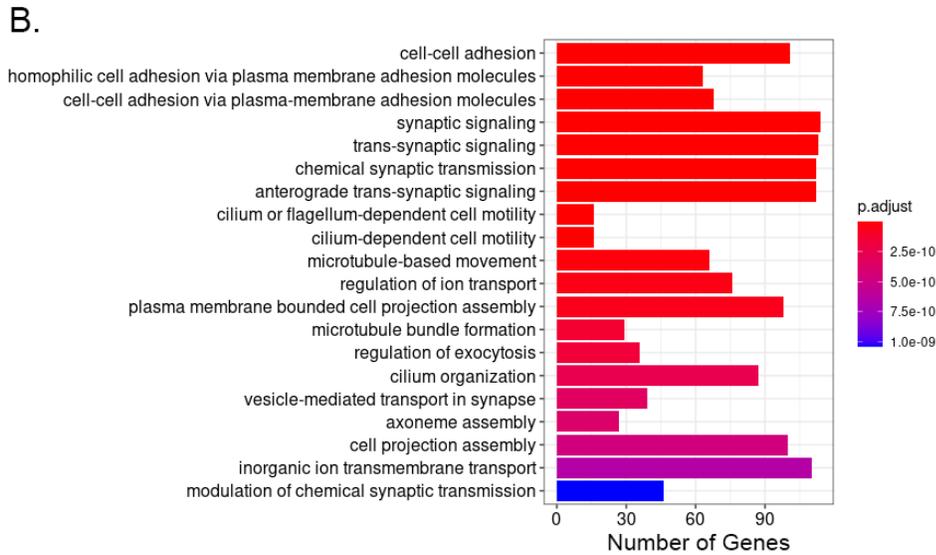


Figure 9. Specified RGCs upregulate many genes involved in neurogenesis compared to RPCs. A. Heatmap of the top 100 differentially expressed genes between 27 HPF RPCs and specified RGCs at 27 HPF. **B.** Go Terms of all upregulated genes are involved in neurogenesis. **C.** GO terms of genes downregulated are in DNA replication.

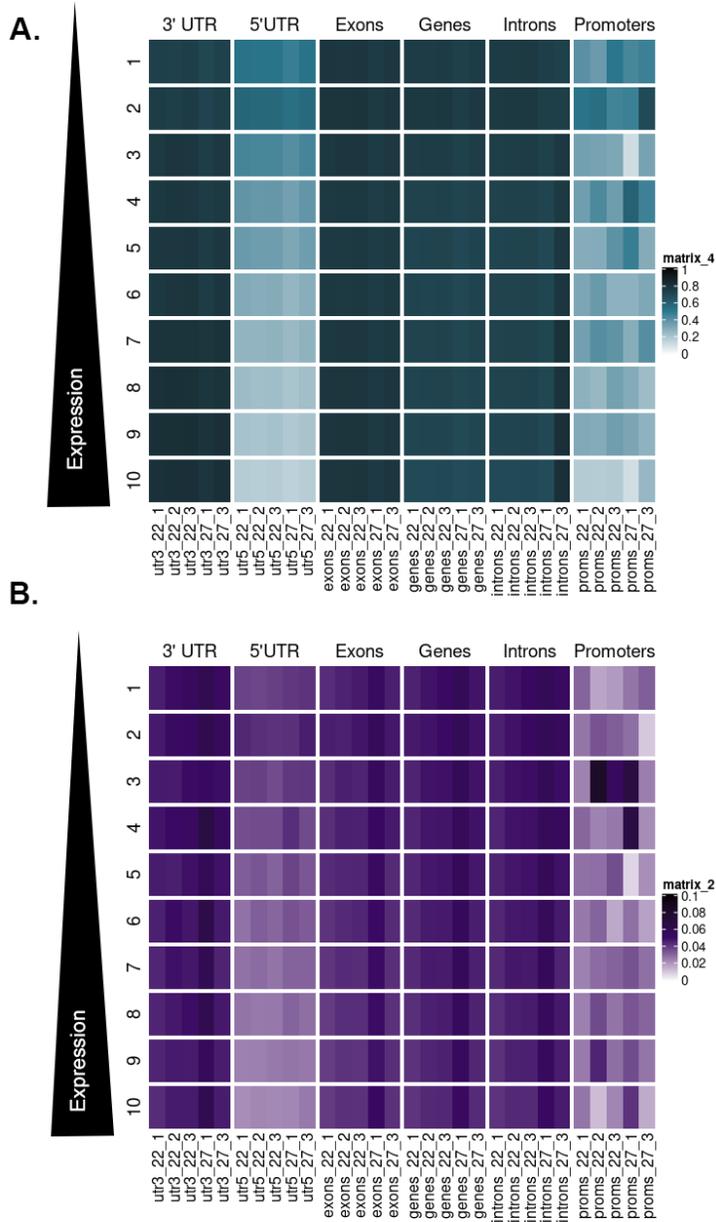


Figure 10. 5mC and 5hmC change gene expression. A. 5mC changes with increasing gene in promoters and 5'UTR. **B.** 5hmC content changes with changing gene expression in the promoters and 5'UTR.

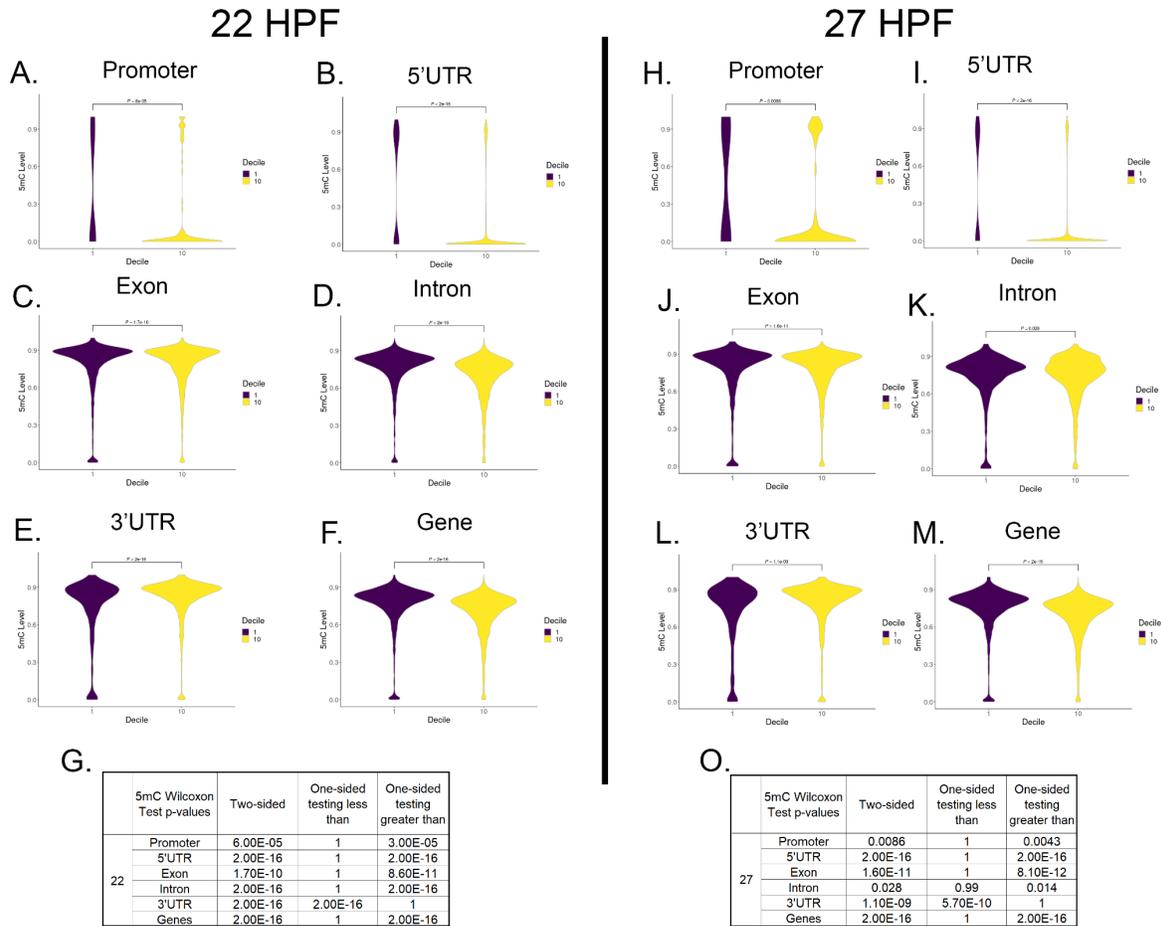


Figure 11. 5mC changes in all genic regions with expression 5mC changes in all genic regions with expression 5mC levels at bases in deciles 1 and 10 within the genic region at 22 (**A-G**) and 27 HPF (**H-O**), plotted with p-values for two-tailed Wilcoxon results. **G.** P values for Wilcoxon test, two-sided, one-sided less than and one-sided greater than for 22 HPF. **O.** P values for Wilcoxon test, two-sided, one-sided less than and one-sided greater than for 27 HPF

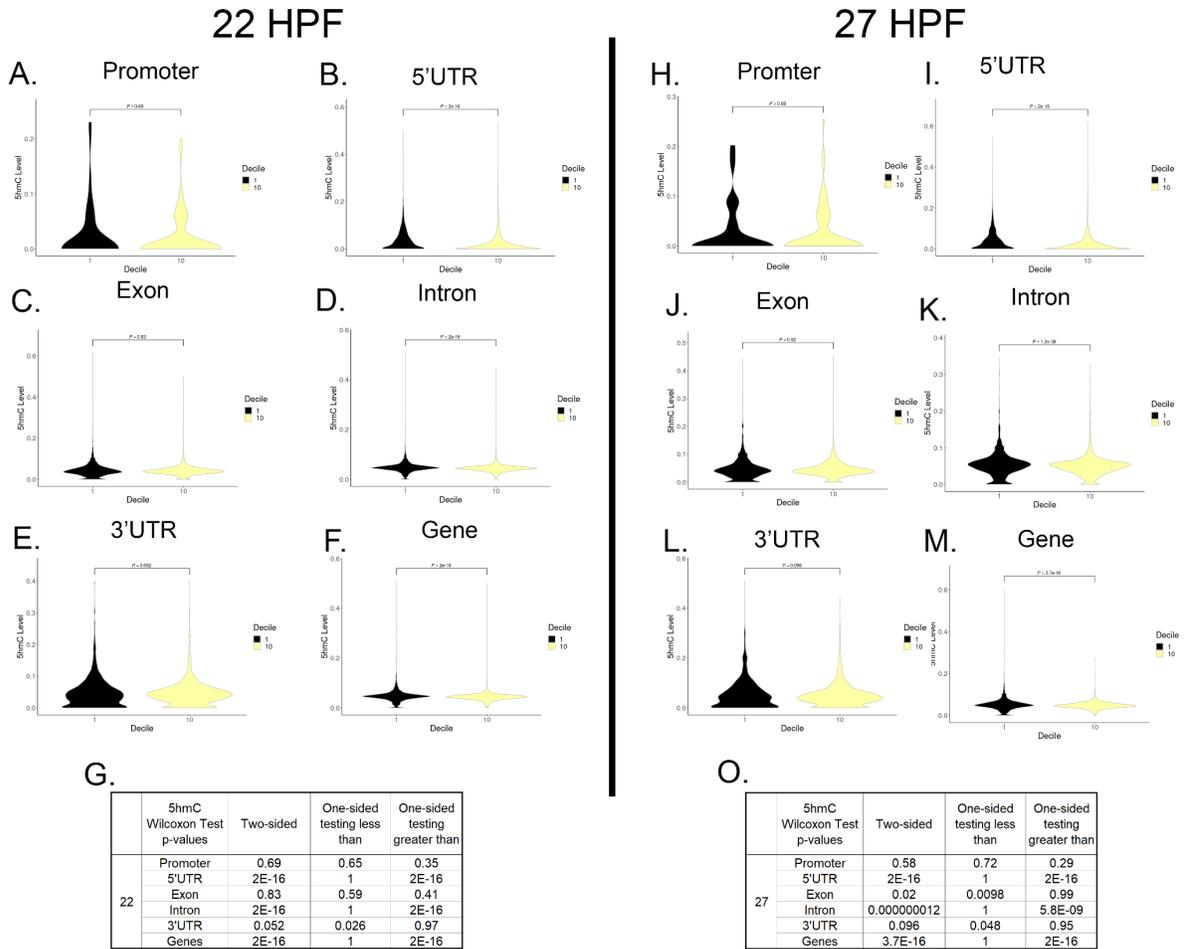


Figure 12. 5hmC changes in most genic regions with expression Comparison between deciles 1 and 10 in 5hmC content in genic regions at 22 (**A-G**) and 27 HPF (**H-O**), p-values for two-tailed Wilcoxon tests plotted. **G.** P values for Wilcoxon test, two-sided, one-sided less than and one-sided greater than for 22 HPF. **O.** P values for Wilcoxon test, two-sided, one-sided less than and one-sided greater than for 27 HPF.

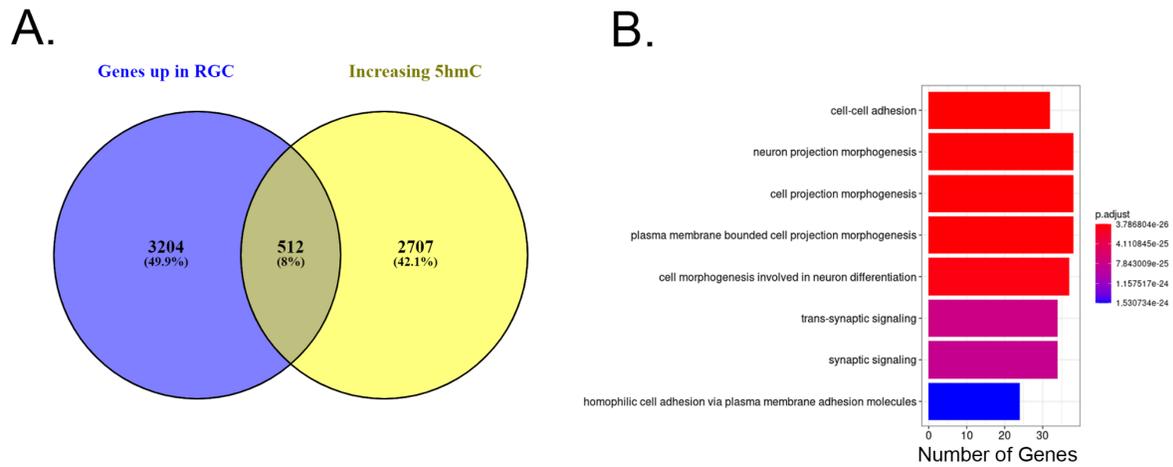


Figure 13. Genes upregulated in RGC fated progenitor contain increasing 5hmC and were enriched in neural genes **A.** Venn diagram of genes upregulated in specified RGCs compared to genes increases in 5hmC between early and late RPCs. **B.** GO-terms of genes increasing in expression and 5hmC are enriched in neural functions

6.0 Tables

Table 1. Coverage of the Genome

Table 1. Coverage of the Genome			
Timepoint	Replicate	Bisulfite Sample	Oxidative Bisulfite Sample
22 HPF	1	40.44	48.33
	2	33.42	37.85
	3	42.77	38.06
27 HPF	1	40.4	48.91
	2	32.26	38.75
	3	58.4	36.16

The sequencing coverage per sample calculated by SamSort in Picard. The number indicates the number of times the whole genome should be represented in the data.

Table 2. Whole Genome Alignment to Reference

Table 2. Whole Genome Alignment to Reference			
Timepoint	Replicate	Bisulfite Sample	Oxidative Bisulfite Sample
22 HPF	1	99.93	99.92
	2	99.92	99.94
	3	99.93	99.94
27 HPF	1	99.93	99.93
	2	99.93	99.94
	3	99.91	99.93

Generated oxBS and BS samples were aligned to the zebrafish reference genome assembly GRCz11. The percent of reads aligned to the reference are reported here.

Table 3. Whole Genome C:T Conversion

Table 3. Whole Genome C:T Conversion			
Timepoint	Replicate	Bisulfite Sample	Oxidative Bisulfite Sample
22 HPF	1	97.5	97.5
	2	97.6	97.7
	3	97.8	97.7
27 HPF	1	97.5	96.9
	2	97.8	96.1
	3	97.3	97.4

Unmethylated non-CpG residues were compared to CpG residues that are methylated.

The frequency of C:T in the non-CpG residues were compared to C frequency in CpG sites to estimate conversion efficiency from 100-0% conversion, with 100% representing complete conversion.

Table 4. CpG Sites Filtered out from Samples

Table 4: CpG Sites Filtered out from Samples						
Cell Type	Timepoint (HPF)	Replicate	Treatment	Total Number of CpG sites	CpGs covered by 10 or more reads	Percent of Reads Kept after Filtering
RPC	22	1	BS	59432803	36096163	60.7
RPC	22	1	OxBS	59432803	35730475	60.1
RPC	22	2	BS	59432803	29012963	48.8
RPC	22	2	OxBS	59432803	27682146	46.6
RPC	22	3	BS	59432803	36405287	61.3
RPC	22	3	OxBS	59432803	23320843	39.2
RPC	27	1	BS	59432803	33753825	56.8
RPC	27	1	OxBS	59432803	15900820	26.8
RPC	27	2	BS	59432803	17764022	29.9
RPC	27	2	OxBS	59432803	2749391	4.6
RPC	27	3	BS	59432803	43300641	72.9
RPC	27	3	OxBS	59432803	22035590	37.1

CpG sites in oxBS and BS samples from both time points were subjected to a filter of 10 or more reads. The total number of CpG sites in the zebrafish genome was reported along with the number of CpG sites covered by 10 or more reads; these were compared to determine the percent of reads kept after the filter was applied.

Table 5. Localization of DMRs

Table 5. Localization of DMRs						
Region	#DMRs Increasing 5mC	#DMRs Decreasing 5mC	#Genes Increasing 5mC in	#Genes Decreasing 5mC in	Percent of Bases Increasing in 5mC	Percent of Bases Decreasing in 5mC
Promoters	0	94	0	13	0.0000	0.0528
5'UTR	0	24	0	20	0.0000	0.3959
Exons	4	437	0	271	0.0000	0.2071
Introns	16	1425	14	736	0.0000	0.0782
3'UTR	1	28	0	30	0.0000	0.2016
Genes	24	2359	19	1244	0.0004	0.0760
Intergenic	40	3850	NA	NA	0.0003	0.0552

The number of DMRs reported by MethPipe and the region they occur in. The first two columns are the raw number of DMRs within each genic region. The number of genes that containing DMRs were reported. The percent of bases changing in 5mC content were calculated by dividing the number of bases changing in each genic region by the genic region's total length.

Table 6. Localization of DHMRs

Table 6. Localization of DHMRs						
Region	#DHMRs Increasing 5hmC	#DHMRs Decreasing 5hmC	#Genes with Increasing 5hmC in	#Genes with Decreasing 5hmC in	Percent of Bases Increasing in 5hmC	Percent of Bases Decreasing in 5hmC
Promoters	155	3290	2	526	0.182	1.997
5'UTR	13	740	12	714	0.011	8.889
Exons	411	14117	401	9488	0.034	5.561
Introns	2603	54373	1935	11002	0.019	3.149
3'UTR	55	1460	55	1209	0.012	3.854
Genes	4354	90397	3248	18695	0.018	2.937
Intergenic	7558	150024	NA	NA	0.014	2.193

The number of DHMRs reported by MethPipe and the region they occur in. The first two columns are the raw number of DHMRs contained in genic regions. The number of genes that contain the DHMRs were reported in the middle columns, split by whether they are increasing or decreasing. This indicates in each gene there can be multiple DHMRs. The percent of bases changing in 5hmC content were calculated by dividing the number of bases changing in each genic region by the genic region's total length.

Table 7. Differentially Expressed Genes between early and late RPCS

Table 7 . Differentially Expressed Genes between early and late RPCS				
<i>akap12b</i>	<i>cldnf</i>	<i>etv2</i>	<i>krt8</i>	<i>quo</i>
<i>angpt2b</i>	<i>cldnh</i>	<i>f7i</i>	<i>krt97</i>	<i>rpe65b</i>
<i>anxa1a</i>	<i>cnga4</i>	<i>fabp7a</i>	<i>lhx2a</i>	<i>rpz5</i>
<i>anxa2a</i>	<i>cnmd</i>	<i>fkbp10b</i>	<i>lipia</i>	<i>s100t</i>
<i>armc3</i>	<i>COLEC10</i>	<i>fkbp7</i>	<i>lmo2</i>	<i>s100v1</i>
<i>atg9a</i>	<i>CR354540.2</i>	<i>flt4</i>	<i>mapk12a</i>	<i>s1pr5a</i>
<i>b3gnt3.3</i>	<i>CR457445.1</i>	<i>foxn4</i>	<i>mgaa</i>	<i>si:ch211-113a14.18</i>
<i>bmpcr</i>	<i>ctsk</i>	<i>fut9b</i>	<i>mtus1a</i>	<i>si:ch211-113a14.24</i>
<i>CABZ01086029.1</i>	<i>cyp2aa6</i>	<i>galnt17</i>	<i>mybpc2a</i>	<i>si:ch211-243g18.2</i>
<i>camk2d1</i>	<i>daw1</i>	<i>gbgt114</i>	<i>myo15aa</i>	<i>si:ch211-98n17.5</i>
<i>capn8</i>	<i>dkk1b</i>	<i>gmnc</i>	NA	<i>si:ch73-160p18.3</i>
<i>ccdc125</i>	<i>dlx3b</i>	<i>gpa33a</i>	<i>ndufa4l2a</i>	<i>si:ch73-334d15.1</i>
<i>ccl25b</i>	<i>dlx4b</i>	<i>has2</i>	<i>neurod4</i>	<i>si:ch73-56d11.5</i>
<i>ccn2a</i>	<i>ednrab</i>	<i>hes2.2</i>	<i>nipal4</i>	<i>sox10</i>
<i>cdcp1a</i>	<i>EIF3BB</i>	<i>inka1a</i>	<i>npas4l</i>	<i>sox11b</i>
<i>cdh1</i>	<i>epcam</i>	<i>jakmip1</i>	<i>oclna</i>	<i>spaca4l</i>
<i>cdh7b</i>	<i>esama</i>	<i>klf3</i>	<i>oclnb</i>	<i>spint1b</i>
<i>cenpf</i>	<i>esamb</i>	<i>kmt2a</i>	<i>ovol1b</i>	<i>stard15</i>
<i>cldn7b</i>	<i>esrp2</i>	<i>krt18a.1</i>	<i>parvab</i>	<i>tal1</i>
<i>cldnb</i>	<i>esyt3</i>	<i>krt222</i>	<i>pcdh20</i>	<i>tmem238a</i>
<i>pde9al</i>	<i>si:dkey-222f8.3</i>	<i>zgc:174938</i>	<i>slc25a24</i>	<i>tmem30b</i>
<i>pdgfra</i>	<i>si:dkey-262k9.2</i>	<i>zgc:194551</i>	<i>smkr1</i>	<i>txnipb</i>
<i>phlda2</i>	<i>si:dkey-52l18.4</i>	<i>zgc:195001</i>	<i>soul2</i>	<i>vox</i>
<i>pkp3a</i>	<i>si:dkey-74k8.3</i>	<i>zgc:85932</i>	<i>plk2b</i>	<i>vsig8a</i>
<i>pvalb8</i>	<i>zgc:153405</i>	<i>ppp1r3b</i>	<i>zbtb18</i>	

All genes differentially expressed between 22 and 27 HPF RPCs are listed.

Table 8. Number of Differentially expressed genes with DHMR or DMR

Table 8.Number of Differentially expressed genes with DHMR or DMR		
	DEG with DHMR	DEG with DMR
Promoter	9	0
5'UTR	9	0
Exons	68	3
Introns	53	6
3' UTR	12	0
Genes	96	11

Differentially expressed genes in RPCs were compared to the DHMRs and DMRs. This identified that some DEG did contain DHMRs and DMRs.

Table 9. Contingency Table for Genes and DMR

Table 9. Contingency Table for Genes and DMR			
		Change in Gene expression?	
		No	Yes
Change in 5mC?	No	30492	121
	Yes	1904	3

Whether or not a gene changes in expression were compared to where it changes in 5mC to create a contingency table. This format was repeated for each genic region containing a DMR.

Table 10. Contingency Table for Genes and DHMR

Table 10. Contingency Table for Genes and DHMR			
		Change in Gene expression?	
		No	Yes
Change in 5hmC?	No	12517	48
	Yes	19879	76

For each gene in the genome, whether it changes gene expression in RPCs was compared to its change to 5hmC. Similar tables were created for each genic region.

Table 11. Fisher's Exact Test Results on RPCs and DMRs and DHMRs

Table 11. Fisher's Exact Test Results on RPCs and DMRs and DHMRs				
	5mC		5hmC	
	p-value	Odds-ratio	p-value	Odds-ratio
Promoter	1	0	0.18	0.37
5'UTR	1	0	0.26	0.52
Exon	0.03	0	0.93	1
Intron	0.02	0	0.89	0.97
3'UTR	1	0	0.14	0.51
Gene	0.12	0.39	1	0.99

A Fisher's exact test for dependence between DEG of RPCs to DMR or DHMRs was performed. This table reports the p-values for each of the tests performed along with the odds ratio. This showed significance only for DMRs within exons and introns.

Table 12. Genes Upregulated in RGC with Increasing 5hmC

Table 12. Genes Upregulated in RGC with Increasing 5hmC					
<i>aamdc</i>	<i>col5a1</i>	<i>hs3st4</i>	<i>nsmfb</i>	<i>sema6e</i>	<i>zfhx3</i>
<i>aatkb</i>	<i>cpe</i>	<i>hs6st1b</i>	<i>ntm</i>	<i>sept5a</i>	<i>zfpm2a</i>
<i>abca12</i>	<i>cpne4a</i>	<i>hspg2</i>	<i>ntrk2a</i>	<i>sez6l2</i>	<i>zgc:100920</i>
<i>ablim3</i>	<i>cpne5a</i>	<i>igsf11</i>	<i>ntrk3a</i>	<i>sgsm2</i>	<i>zgc:153759</i>
<i>acsl1a</i>	<i>cpne8</i>	<i>igsf9ba</i>	<i>ntrk3b</i>	<i>sh3rf2</i>	<i>zgc:162324</i>
<i>actn2b</i>	<i>crebrf</i>	<i>inpp4aa</i>	<i>nyap2b</i>	<i>shank3a</i>	<i>zgc:162928</i>
<i>adam22</i>	<i>ctnna2</i>	<i>insyn1</i>	<i>olfm1b</i>	<i>shisa7a</i>	<i>zgc:77849</i>
<i>adarb1a</i>	<i>cygb2</i>	<i>iqca1</i>	<i>opcml</i>	<i>shisal1b</i>	<i>zgc:77880</i>
<i>adgrb1a</i>	<i>cyth1a</i>	<i>iqsec1b</i>	<i>oprl1</i>	<i>si:ch211-106a19.1</i>	<i>zgc:92140</i>
<i>adgrb3</i>	<i>dagla</i>	<i>iqsec2a</i>	<i>pacrg</i>	<i>si:ch211-12m10.1</i>	<i>zmp:0000000619</i>
<i>adgrl1a</i>	<i>dapk1</i>	<i>iqsec3a</i>	<i>pak6a</i>	<i>si:ch211-209l18.2</i>	<i>zfn296</i>
<i>adgrl3.1</i>	<i>dchs1b</i>	<i>itga3a</i>	<i>parp9</i>	<i>si:ch211-212o1.2</i>	<i>zfn804a</i>
<i>adipor1a</i>	<i>dclk1a</i>	<i>itga6b</i>	<i>pcdh11</i>	<i>si:ch211-233a24.2</i>	
<i>aebp1</i>	<i>dip2ca</i>	<i>jakmip2</i>	<i>pcdh1b</i>	<i>si:ch211-239f4.1</i>	
<i>aff2</i>	<i>dlg1l</i>	<i>jhy</i>	<i>pcdh1g11</i>	<i>si:ch211-242b18.1</i>	
<i>agbl4</i>	<i>dlg4b</i>	<i>kalrna</i>	<i>pcdh1g18</i>	<i>si:ch211-257p13.3</i>	
<i>ago3b</i>	<i>dmxl2</i>	<i>kalrnb</i>	<i>pcdh1g2</i>	<i>si:ch211-59d17.3</i>	
<i>agrnl</i>	<i>dnal1</i>	<i>katnal2</i>	<i>pcdh1g22</i>	<i>si:ch211-66k16.27</i>	
<i>ak5</i>	<i>dnmt3ab</i>	<i>kcnab1b</i>	<i>pcdh1g26</i>	<i>si:ch73-215d9.1</i>	
<i>ak8</i>	<i>doc2b</i>	<i>kcnc4</i>	<i>pcdh1g29</i>	<i>si:ch73-222h13.1</i>	
<i>akap6</i>	<i>dock3</i>	<i>kcnh7</i>	<i>pcdh1g3</i>	<i>si:ch73-60h1.1</i>	
<i>akt3b</i>	<i>DOCK4</i>	<i>kcnj3a</i>	<i>pcdh1g30</i>	<i>si:dkey-42p14.3</i>	
<i>alcamb</i>	<i>dok6</i>	<i>kcnk10b</i>	<i>pcdh1g33</i>	<i>si:dkey-97l20.6</i>	
<i>alkbh3</i>	<i>dpf1</i>	<i>kcnq3</i>	<i>pcdh1g9</i>	<i>slain1a</i>	
<i>ank3a</i>	<i>dscama</i>	<i>kctd1</i>	<i>pcdh1gb2</i>	<i>slc15a4</i>	

Table 12. Continued

<i>anks1b</i>	<i>dscamb</i>	<i>kctd7</i>	<i>pcdh1gb</i> 9	<i>slc4a10a</i>
<i>ano1</i>	<i>dsg2.1</i>	<i>kdm2bb</i>	<i>pcdh1gc5</i>	<i>slc4a10b</i>
<i>ano5a</i>	<i>DST</i>	<i>khdrbs2</i>	<i>pcdh1gc6</i>	<i>slc7a14a</i>
<i>anxa11b</i>	<i>dusp14</i>	<i>kif1ab</i>	<i>pcdh2g1</i> 7	<i>slc8a2b</i>
<i>apba1b</i>	<i>DYNC2H</i> 1	<i>kif26ba</i>	<i>pcdh9</i>	<i>slit3</i>
<i>apbb2b</i>	<i>edil3a</i>	<i>KIRREL3</i>	<i>pcloa</i>	<i>smap1</i>
<i>arf3b</i>	<i>eef1a1a</i>	<i>klf12a</i>	<i>pcsk2</i>	<i>smoc1</i>
<i>arhgdig</i>	<i>efna3a</i>	<i>klf7b</i>	<i>pdcd4b</i>	<i>smpd3</i>
<i>arhgef4</i>	<i>eipr1</i>	<i>klhl5</i>	<i>pde4ba</i>	<i>snap25b</i>
<i>arl15b</i>	<i>ek1</i>	<i>lgmn</i>	<i>pde5ab</i>	<i>snx25</i>
<i>arvcfa</i>	<i>elfn1b</i>	<i>lhfp16</i>	<i>pdlim5a</i>	<i>spns2</i>
<i>astn1</i>	<i>elmo2</i>	<i>lin7a</i>	<i>pe1i3</i>	<i>spon1a</i>
<i>atf6</i>	<i>emc4</i>	<i>lingo2a</i>	<i>pfkla</i>	<i>spsb4a</i>
<i>atp11a</i>	<i>EML6</i>	<i>lmbird1</i>	<i>pfkpb</i>	<i>sptan1</i>
<i>atp2b2</i>	<i>eno1a</i>	<i>lrp1aa</i>	<i>pgm2</i>	<i>srgap1b</i>
<i>atp2b3b</i>	<i>epha6</i>	<i>lrp2a</i>	<i>phkb</i>	<i>srrm4</i>
<i>atp6ap1b</i>	<i>ephb1</i>	<i>lrrc4ba</i>	<i>pias4b</i>	<i>ssbp3a</i>
<i>atp6v0a2a</i>	<i>erc1a</i>	<i>LRRC75A</i>	<i>picalma</i>	<i>st7</i>
<i>atp8b1</i>	<i>esrrb</i>	<i>lrrn3a</i>	<i>pip5k1ca</i>	<i>stau2</i>
<i>ATP9A</i>	<i>fam131ba</i>	<i>luzp2</i>	<i>pitpnab</i>	<i>stox2a</i>
<i>atrnl1a</i>	<i>fam131c</i>	<i>macrod2</i>	<i>pkib</i>	<i>strbp</i>
<i>auts2a</i>	<i>fam155a</i>	<i>map2</i>	<i>pkig</i>	<i>stxbp1a</i>
<i>b4galnt3b</i>	<i>fam184b</i>	<i>MAP3K13</i>	<i>pkp3b</i>	<i>stxbp5a</i>
<i>b4galt2</i>	<i>fam20b</i>	<i>MDFI</i>	<i>plch2a</i>	<i>susd4</i>
<i>b4galt3</i>	<i>faxcb</i>	<i>mdga1</i>	<i>plppr4a</i>	<i>syn1</i>
<i>baiap3</i>	<i>fbrsl1</i>	<i>mdga2a</i>	<i>ppargc1a</i>	<i>syt9b</i>
<i>bmpr1ba</i>	<i>fbxl6</i>	<i>megf11</i>	<i>ppm1e</i>	<i>TANC2</i>
<i>brsk2b</i>	<i>fbxo16</i>	<i>meis1a</i>	<i>ppp1r9bb</i>	<i>tbcelb</i>
<i>c1galt1a</i>	<i>fhdc3</i>	<i>mff</i>	<i>ppp2r5b</i>	<i>tbkbp1</i>
<i>ca10a</i>	<i>fmn2a</i>	<i>mphas1</i>	<i>ppp3cb</i>	<i>tbxas1</i>
<i>ca16b</i>	<i>fmnl2b</i>	<i>mgat4b</i>	<i>prickle2a</i>	<i>tead1b</i>
<i>cacna1bb</i>	<i>foxo1a</i>	<i>micu3a</i>	<i>prkag2a</i>	<i>TENM2</i>
<i>cacna1c</i>	<i>foxo6a</i>	<i>mink1</i>	<i>prkar1b</i>	<i>tenm4</i>
<i>cacna1g</i>	<i>foxp4</i>	<i>morn3</i>	<i>prkg1b</i>	<i>tex264a</i>

Table 12. Continued

<i>cacna2d3</i>	<i>fryb</i>	<i>mpp2b</i>	<i>prlhr2a</i>	<i>tgm2b</i>
<i>cacnb2a</i>	<i>fstl5</i>	<i>mpp3b</i>	<i>prr13</i>	<i>tiparp</i>
<i>cadps2</i>	<i>gabbr1b</i>	<i>mroh1</i>	<i>ptmaa</i>	<i>tm2d1</i>
<i>camk1da</i>	<i>gabbr2</i>	<i>ms4a17a.11</i>	<i>ptprea</i>	<i>tmed7</i>
<i>camk2g1</i>	<i>gabra5</i>	<i>mtmr7a</i>	<i>ptprga</i>	<i>tmem150c</i>
<i>camsap3</i>	<i>gabbr3</i>	<i>mtmr7b</i>	<i>ptprsa</i>	<i>tmem63c</i>
<i>camta1a</i>	<i>galnt9</i>	<i>mtss1b</i>	<i>ptprua</i>	<i>tmtops2b</i>
<i>camta1b</i>	<i>gamt</i>	<i>mtus1a</i>	<i>ptprz1a</i>	<i>tnk2a</i>
<i>capn12</i>	<i>GARNL3</i>	<i>mtus1b</i>	<i>pvr12l</i>	<i>traf4b</i>
<i>capn2l</i>	<i>gchfr</i>	<i>myo10l1</i>	<i>rab11bb</i>	<i>trhde.2</i>
<i>cblb</i>	<i>gdf11</i>	<i>myo5aa</i>	<i>rab3ip</i>	<i>trip10a</i>
<i>cc2d1b</i>	<i>GGT7</i>	<i>naalad2</i>	<i>ralaa</i>	<i>trpc4b</i>
<i>ccdc136b</i>	<i>glcci1a</i>	<i>nalcn</i>	<i>ralgps1</i>	<i>tshz3b</i>
<i>ccdc65</i>	<i>glra4a</i>	<i>nbeaa</i>	<i>raph1b</i>	<i>tspan15</i>
<i>cd99</i>	<i>gnao1a</i>	<i>nbeal2</i>	<i>rasgrf2b</i>	<i>tspan2a</i>
<i>cdadc1</i>	<i>gpc6a</i>	<i>ncam2</i>	<i>raver2</i>	<i>tspan5b</i>
<i>cdc42ep4a</i>	<i>gphnb</i>	<i>nckap5l</i>	<i>rbfox3a</i>	<i>ttbk2a</i>
<i>CDCP1</i>	<i>gpr12</i>	<i>ncs1a</i>	<i>rca3</i>	<i>ttc26</i>
<i>cdh10a</i>	<i>gpr153</i>	<i>ndfip1</i>	<i>reep3b</i>	<i>TTC28</i>
<i>cdh13</i>	<i>gpr158a</i>	<i>ndrg4</i>	<i>rem2</i>	<i>ttc39c</i>
<i>cdh18a</i>	<i>gpr158b</i>	<i>ndst3</i>	<i>rgs11</i>	<i>ttl3</i>
<i>celf5a</i>	<i>gria1a</i>	<i>ndufa11</i>	<i>rhbdl3</i>	<i>ttl7</i>
<i>celsr3</i>	<i>gria2a</i>	<i>necab2</i>	<i>rims1a</i>	<i>ttyh1</i>
<i>cep170ab</i>	<i>gria3a</i>	<i>neo1b</i>	<i>rims1b</i>	<i>tub</i>
<i>cerkl</i>	<i>gria3b</i>	<i>neto1l</i>	<i>RIMS2</i>	<i>ubn1</i>
<i>cers6</i>	<i>gria4b</i>	<i>neto2b</i>	<i>rnf11b</i>	<i>ubr4</i>
<i>cfap300</i>	<i>grid1b</i>	<i>neurl1aa</i>	<i>rnf130</i>	<i>ugt8</i>
<i>cfap57</i>	<i>grid2</i>	<i>nexmifb</i>	<i>robo2</i>	<i>unc5a</i>
<i>chchd6a</i>	<i>GRIK2</i>	<i>nfatc1</i>	<i>robo3</i>	<i>unc5da</i>
<i>chm</i>	<i>grin2da</i>	<i>nid1b</i>	<i>rrm2b</i>	<i>unc5db</i>
<i>cib2</i>	<i>grm2b</i>	<i>nkain2</i>	<i>rtn4r</i>	<i>ush1ga</i>
<i>cica</i>	<i>GRM7</i>	<i>nlgn1</i>	<i>rufy2</i>	<i>vav2</i>
<i>clstn2</i>	<i>grm8b</i>	<i>nlgn2a</i>	<i>rundc3b</i>	<i>vcanb</i>
<i>clstn3</i>	<i>gucy1a1</i>	<i>nlgn2b</i>	<i>runx1</i>	<i>vldlr</i>
<i>clta</i>	<i>gucy1b1</i>	<i>nptx2a</i>	<i>runx1t1</i>	<i>wasf1</i>
<i>cnih2</i>	<i>hacd2</i>	<i>nrbp2b</i>	<i>satb1b</i>	<i>whrna</i>
<i>cnot2</i>	<i>hcn1</i>	<i>nrg1</i>	<i>sbf2</i>	<i>wnk1b</i>

Table 12. Continued

<i>cntn5</i>	<i>hecw1b</i>	<i>nrxn1a</i>	<i>sema3d</i>	<i>ywhag1</i>
<i>cntnap2a</i>	<i>hnrnpa3</i>	<i>nrxn2a</i>	<i>SEMA4F</i>	<i>zbbx</i>
<i>col11a1a</i>	<i>hnrnpd</i>	<i>nrxn3b</i>	<i>sema5a</i>	<i>zbtb46</i>

The 516 genes that increase expression in RGCs that have increasing 5hmC between early and late RPCs.

Table 13. Contingency Table for Genes DEG between RPC and RGC and DHMR

Table 13. Contingency Table for Genes DEG between RPC and RGC and DHMR			
		Change in Gene expression?	
		No	Yes
Change in 5hmC?	No	11135	1430
	Yes	14996	4549

The number of genes differentially expressed between late RPCs and RGCs was compared to changing 5hmC.

Table 14. Changes in expression in RGCs was dependent on changing 5hmC

Table 14. Changes in expression in RGCs was dependent on changing 5hmC				
	5mC		5hmC	
	p-value	Odds-ratio	p-value	Odds-ratio
Promoter	4.00E-01	1.4	0.26	0.92
5'UTR	0.58	1.25	0.37	1.06
Exon	1	0.99	0.047	1.03
Intron	9.94E-08	1.47	2.20E-16	1.96
3'UTR	0.081	0.55	0.029	1.13
Gene	2.20E-16	1.68	2.20E-16	2.57

The results of the Fisher's exact test are listed here, showing significance in both 5mC and 5hmC.

Table 15. Contingency Table for Genes Upregulated between RGC and RPC and increasing DHMR

Table 15. Contingency Table for Genes Upregulated between RGC and RPC and increasing DHMR			
		Increase in Gene expression?	
		No	Yes
Increase in 5hmC?	No	26034	3238
	Yes	2698	550

The number of genes upregulated in RGCs compared to RPCs was compared to the genes containing an increase in 5hmC between early and late RPCs.

Table 16. Changes in expression in RGCs was dependent on changing 5hmC.

Table 16.Changes in expression in RGCs was dependent on changing 5hmC		
	P-value	Odds ratio
Gene	2.20E-16	1.63

Fisher's exact test results from **Table 15** are reported here, including the p-value and odds ratio.

Bibliography

- Anastasiadi, D., Esteve-Codina, A., Piferrer, F., 2018. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11, 37. doi:10.1186/s13072-018-0205-1
- Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., Balasubramanian, S., 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055. doi:10.1038/nchem.2064
- Belliveau, M.J., Cepko, C.L., 1999. Extrinsic and intrinsic factors control the genesis of amacrine and cone cells in the rat retina. *Development* 126, 555–566.
- Bonifer, C., Cockerill, P.N., 2017. Chromatin priming of genes in development: Concepts, mechanisms and consequences. *Exp. Hematol.* 49, 1–8. doi:10.1016/j.exphem.2017.01.003
- Booth, M.J., Ost, T.W.B., Beraldi, D., Bell, N.M., Branco, M.R., Reik, W., Balasubramanian, S., 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* 8, 1841–1851. doi:10.1038/nprot.2013.115
- Cayouette, M., Barres, B.A., Raff, M., 2003. Importance of intrinsic mechanisms in cell fate decisions in the developing rat retina. *Neuron* 40, 897–904.
- Cepko, C.L., Austin, C.P., Yang, X., Alexiades, M., Ezzeddine, D., 1996. Cell fate determination in the vertebrate retina. *Proc. Natl. Acad. Sci. USA* 93, 589–595. doi:10.1073/pnas.93.2.589
- Chopra, P., Papale, L.A., White, A.T.J., Hatch, A., Brown, R.M., Garthwaite, M.A., Roseboom, P.H., Golos, T.G., Warren, S.T., Alisch, R.S., 2014. Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. *BMC Genomics* 15, 131. doi:10.1186/1471-2164-15-131
- Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., Oate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y., Cherry, J.M., 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46, D794–D801. doi:10.1093/nar/gkx1081
- Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022. doi:10.1101/gad.2037511

- Doe, C.Q., 2008. Neural stem cells: balancing self-renewal with differentiation. *Development* 135, 1575–1587. doi:10.1242/dev.014977
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M., Jacobsen, S.E., 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA* 107, 8689–8694. doi:10.1073/pnas.1002720107
- Fuhrmann, S., 2010. Eye morphogenesis and patterning of the optic vesicle. *Curr Top Dev Biol* 93, 61–84. doi:10.1016/B978-0-12-385044-7.00003-5
- Furlong, E.E.M., Levine, M., 2018. Developmental enhancers and chromosome topology. *Science* 361, 1341–1345. doi:10.1126/science.aau0320
- Gao, D., Pinello, N., Nguyen, T.V., Thoeng, A., Nagarajah, R., Holst, J., Rasko, J.E., Wong, J.J.-L., 2019. DNA methylation/hydroxymethylation regulate gene expression and alternative splicing during terminal granulopoiesis. *Epigenomics* 11, 95–109. doi:10.2217/epi-2018-0050
- Gerety, S.S., Wilkinson, D.G., 2011. Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. *Dev. Biol.* 350, 279–289. doi:10.1016/j.ydbio.2010.11.030
- Globisch, D., Münzel, M., Müller, M., Michalakakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., Carell, T., 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5, e15367. doi:10.1371/journal.pone.0015367
- Greco, C.M., Kunderfranco, P., Rubino, M., Larcher, V., Carullo, P., Anselmo, A., Kurz, K., Carell, T., Angius, A., Latronico, M.V.G., Papait, R., Condorelli, G., 2016. DNA hydroxymethylation controls cardiomyocyte gene expression in development and hypertrophy. *Nat. Commun.* 7, 12418. doi:10.1038/ncomms12418
- Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi:10.1093/bioinformatics/btw313
- Gu, Z., Eils, R., Schlesner, M., Ishaque, N., 2018. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics* 19, 234. doi:10.1186/s12864-018-4625-x
- Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G., Song, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* 17, 215–222. doi:10.1038/nn.3607

- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., Xu, G.-L., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi:10.1126/science.1210944
- Heavner, W., Pevny, L., 2012. Eye development and retinogenesis. *Cold Spring Harb. Perspect. Biol.* 4. doi:10.1101/cshperspect.a008391
- Hemberger, M., Dean, W., Reik, W., 2009. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington’s canal. *Nat. Rev. Mol. Cell Biol.* 10, 526–537. doi:10.1038/nrm2727
- Hernando-Herraez, I., Garcia-Perez, R., Sharp, A.J., Marques-Bonet, T., 2015. DNA Methylation: Insights into Human Evolution. *PLoS Genet.* 11, e1005661. doi:10.1371/journal.pgen.1005661
- Hoon, M., Okawa, H., Della Santina, L., Wong, R.O.L., 2014. Functional architecture of the retina: development and disease. *Prog Retin Eye Res* 42, 44–84. doi:10.1016/j.preteyeres.2014.06.003
- Ito, S., D’Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133. doi:10.1038/nature09303
- Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., Tice, R., Tochigi, M., Morera, S., Nazarians, A., Belsham, D., Wong, A.H.C., Blencowe, B.J., Wang, S.C., Kapranov, P., Kustra, R., Labrie, V., Klimasauskas, S., Petronis, A., 2012. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* 19, 1037–1043. doi:10.1038/nsmb.2372
- Kim, M., Costello, J., 2017. DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med* 49, e322. doi:10.1038/emm.2017.10
- Klose, R.J., Bird, A.P., 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* 31, 89–97. doi:10.1016/j.tibs.2005.12.008
- Kochmanski, J., Savonen, C., Bernstein, A.I., 2019. Reconciling Base-Pair Resolution 5-methylcytosine and 5-hydroxymethylcytosine Data in Neuroepigenetics. *BioRxiv.* doi:10.1101/630772
- Kriaucionis, S., n.d. The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and enriched in Purkinje neurons.
- Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi:10.1126/science.1169786

- Lee, H.J., Lowdon, R.F., Maricque, B., Zhang, B., Stevens, M., Li, D., Johnson, S.L., Wang, T., 2015. Developmental enhancers revealed by extensive DNA methylome maps of zebrafish early embryos. *Nat. Commun.* 6, 6315. doi:10.1038/ncomms7315
- Li, D., Guo, B., Wu, H., Tan, L., Lu, Q., 2015. TET family of dioxygenases: crucial roles and underlying mechanisms. *Cytogenet Genome Res* 146, 171–180. doi:10.1159/000438853
- Li, T., Yang, D., Li, J., Tang, Y., Yang, J., Le, W., 2015. Critical role of Tet3 in neural progenitor cell maintenance and terminal differentiation. *Mol. Neurobiol.* 51, 142–154. doi:10.1007/s12035-014-8734-5
- Liao, Y., Smyth, G.K., Shi, W., 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47, e47. doi:10.1093/nar/gkz114
- Livesey, F.J., Cepko, C.L., 2001. Vertebrate neural cell-fate determination: lessons from the retina. *Nat. Rev. Neurosci.* 2, 109–118. doi:10.1038/35053522
- Ludwig, A.K., Zhang, P., Cardoso, M.C., 2016. Modifiers and readers of DNA modifications and their impact on genome structure, expression, and stability in disease. *Front. Genet.* 7, 115. doi:10.3389/fgene.2016.00115
- Madrid, A., Chopra, P., Alisch, R.S., 2018. Species-Specific 5 mC and 5 hmC Genomic Landscapes Indicate Epigenetic Contribution to Human Brain Evolution. *Front. Mol. Neurosci.* 11, 39. doi:10.3389/fnmol.2018.00039
- Mahé, E.A., Madigou, T., Salbert, G., 2018. Reading cytosine modifications within chromatin. *Transcription* 9, 240–247. doi:10.1080/21541264.2017.1406435
- Masai, I., Stemple, D.L., Okamoto, H., Wilson, S.W., 2000. Midline signals regulate retinal neurogenesis in zebrafish. *Neuron* 27, 251–263. doi:10.1016/S0896-6273(00)00034-9
- Maunakea, A.K., Chepelev, I., Cui, K., Zhao, K., 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* 23, 1256–1269. doi:10.1038/cr.2013.110
- McGaughey, D.M., Abaan, H.O., Miller, R.M., Kropp, P.A., Brody, L.C., 2014. Genomics of CpG methylation in developing and developed zebrafish. *G3 (Bethesda)* 4, 861–869. doi:10.1534/g3.113.009514
- McGuire, M.H., Herbrich, S.M., Dasari, S.K., Wu, S.Y., Wang, Y., Rupaimoole, R., Lopez-Berestein, G., Baggerly, K.A., Sood, A.K., 2019. Pan-cancer genomic analysis links 3'UTR DNA methylation with increased gene expression in T cells. *EBioMedicine* 43, 127–137. doi:10.1016/j.ebiom.2019.04.045

- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., Heintz, N., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi:10.1016/j.cell.2012.11.022
- Morgan, M., Obenchain, V., Hester, J., Pagès, H., 2017. SummarizedExperiment: SummarizedExperiment container. R package version.
- Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., Meehan, R.R., 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22, 467–477. doi:10.1101/gr.126417.111
- Pastor, W.A., Aravind, L., Rao, A., 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 14, 341–356. doi:10.1038/nrm3589
- Pedersen, B.S., Eyring, K., De, S., Yang, I.V., 2014. Fast and accurate alignment of long bisulfite-seq reads. *arXiv preprint arXiv*
- Pfeifer, G.P., Xiong, W., Hahn, M.A., Jin, S.-G., 2014. The role of 5-hydroxymethylcytosine in human cancer. *Cell Tissue Res.* doi:10.1007/s00441-014-1896-7
- Qu, Y., Lennartsson, A., Gaidzik, V.I., Deneberg, S., Bengtzén, S., Arzenani, M.K., Bullinger, L., Döhner, K., Lehmann, S., 2012. Genome-Wide DNA Methylation Analysis Shows Enrichment of Differential Methylation in “Open Seas” and Enhancers and Reveals Hypomethylation in DNMT3A Mutated Cytogenetically Normal AML (CN-AML). *Blood* 120, 653–653. doi:10.1182/blood.V120.21.653.653
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Roeh, S., Wiechmann, T., Sauer, S., Ködel, M., Binder, E.B., Provençal, N., 2018. HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing. *Epigenetics Chromatin* 11, 39. doi:10.1186/s13072-018-0209-x
- Santiago, M., Antunes, C., Guedes, M., Iacovino, M., Kyba, M., Reik, W., Sousa, N., Pinto, L., Branco, M.R., Marques, C.J., 2020a. Correction to: Tet3 regulates cellular identity and DNA methylation in neural progenitor cells. *Cell Mol. Life Sci.* 77, 2885. doi:10.1007/s00018-019-03443-4

- Santiago, M., Antunes, C., Guedes, M., Iacovino, M., Kyba, M., Reik, W., Sousa, N., Pinto, L., Branco, M.R., Marques, C.J., 2020b. Tet3 regulates cellular identity and DNA methylation in neural progenitor cells. *Cell Mol. Life Sci.* 77, 2871–2883. doi:10.1007/s00018-019-03335-7
- Saul, S.M., Brzezinski, J.A., Altschuler, R.A., Shore, S.E., Rudolph, D.D., Kabara, L.L., Halsey, K.E., Hufnagel, R.B., Zhou, J., Dolan, D.F., Glaser, T., 2008. Math5 expression and function in the central auditory system. *Mol. Cell. Neurosci.* 37, 153–169. doi:10.1016/j.mcn.2007.09.006
- Scourzic, L., Mouly, E., Bernard, O.A., 2015. TET proteins and the control of cytosine demethylation in cancer. *Genome Med.* 7, 9. doi:10.1186/s13073-015-0134-6
- Seritrukul, P., Gross, J.M., 2017. Tet-mediated DNA hydroxymethylation regulates retinal neurogenesis by modulating cell-extrinsic signaling pathways. *PLoS Genet.* 13, e1006987. doi:10.1371/journal.pgen.1006987
- Shi, D.-Q., Ali, I., Tang, J., Yang, W.-C., 2017. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front. Genet.* 8, 100. doi:10.3389/fgene.2017.00100
- Soneson, C., Love, M.I., Robinson, M.D., 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. [version 2; peer review: 2 approved]. *F1000Res.* 4, 1521. doi:10.12688/f1000research.7563.2
- Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T.J., Zhang, B., Godley, L.A., Hicks, L.M., Lahn, B.T., Jin, P., He, C., 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72. doi:10.1038/nbt.1732
- Song, Q., Decato, B., Hong, E.E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., Smith, A.D., 2013. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* 8, e81148. doi:10.1371/journal.pone.0081148
- Stenkamp, D.L., 2007. Neurogenesis in the fish retina. *Int Rev Cytol* 259, 173–224. doi:10.1016/S0074-7696(06)59005-9
- Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., Yoon, Y.-S., Ren, B., He, C., Jin, P., 2011a. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7, e1002154. doi:10.1371/journal.pgen.1002154
- Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., Vasanthakumar, A., Godley, L.A., Chang, Q., Cheng, X., He, C., Jin, P., 2011b. 5-hmC-mediated epigenetic dynamics during postnatal

- neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616. doi:10.1038/nn.2959
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi:10.1126/science.1170116
- Tan, L., Xiong, L., Xu, W., Wu, F., Huang, N., Xu, Y., Kong, L., Zheng, L., Schwartz, L., Shi, Y., Shi, Y.G., 2013. Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Res.* 41, e84. doi:10.1093/nar/gkt091
- Turner, D.L., Cepko, C.L., 1987. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* 328, 131–136. doi:10.1038/328131a0
- Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., Papapetrou, E.P., He, C., Elemento, O., Huangfu, D., 2018. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* 50, 83–95. doi:10.1038/s41588-017-0002-y
- Vitorino, M., Jusuf, P.R., Maurus, D., Kimura, Y., Higashijima, S.-I., Harris, W.A., 2009. *Vsx2* in the zebrafish retina: restricted lineages through derepression. *Neural Dev.* 4, 14. doi:10.1186/1749-8104-4-14
- Waisman, A., Seivlever, F., Elías Costa, M., Cosentino, M.S., Miriuka, S.G., Ventura, A.C., Guberman, A.S., 2019. Cell cycle dynamics of mouse embryonic stem cells in the ground state and during transition to formative pluripotency. *Sci. Rep.* 9, 8051. doi:10.1038/s41598-019-44537-0
- Wang, Z., Du, M., Yuan, Q., Guo, Y., Hutchinson, J.N., Su, L., Zheng, Y., Wang, J., Mucci, L.A., Lin, X., Hou, L., Christiani, D.C., 2020. Epigenomic analysis of 5-hydroxymethylcytosine (5hmC) reveals novel DNA methylation markers for lung cancers. *Neoplasia* 22, 154–161. doi:10.1016/j.neo.2020.01.001
- Wetts, R., Serbedzija, G.N., Fraser, S.E., 1989. Cell lineage analysis reveals multipotent precursors in the ciliary margin of the frog retina. *Dev. Biol.* 136, 254–263. doi:10.1016/0012-1606(89)90146-2
- Williams, K., Christensen, J., Helin, K., 2011. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep.* 13, 28–35. doi:10.1038/embor.2011.233
- Wreczycka, K., Godschan, A., Yusuf, D., Grüning, B., Assenov, Y., Akalin, A., 2017. Strategies for analyzing bisulfite sequencing data. *J. Biotechnol.* 261, 105–115. doi:10.1016/j.jbiotec.2017.08.007

- Wu, H., Zhang, Y., 2011. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* 25, 2436–2452. doi:10.1101/gad.179184.111
- Xu, B., Tang, X., Jin, M., Zhang, H., Du, L., Yu, S., He, J., 2020. Unifying Developmental Programs for Embryonic and Post-Embryonic Neurogenesis in the Zebrafish Retina. *Development*. doi:10.1242/dev.185660
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., Shi, Y.G., 2011. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464. doi:10.1016/j.molcel.2011.04.005
- Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H.H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G.J.C., Xu, G., Kaiser, U.B., Liu, X.S., Sugrue, S.P., He, X., Min, J., Kato, Y., Shi, Y.G., 2012. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 151, 1200–1213. doi:10.1016/j.cell.2012.11.014
- Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi:10.1089/omi.2011.0118
- Zhang, C., Hoshida, Y., Sadler, K.C., 2016. Comparative epigenomic profiling of the DNA methylome in mouse and zebrafish uncovers high interspecies divergence. *Front. Genet.* 7, 110. doi:10.3389/fgene.2016.00110
- Zhang, J., Chen, S., Zhang, D., Shi, Z., Li, H., Zhao, T., Hu, B., Zhou, Q., Jiao, J., 2016. Tet3-Mediated DNA Demethylation Contributes to the Direct Conversion of Fibroblast to Functional Neuron. *Cell Rep.* 17, 2326–2339. doi:10.1016/j.celrep.2016.10.081
- Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., Gnirke, A., Meissner, A., 2011. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* 7, e1002389. doi:10.1371/journal.pgen.1002389