Title

Dissecting Complex Neurological Processes with Next-generation Sequencing and Other Whole-genome Approaches

Author

Cheng, Yin

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Dissecting Complex Neurological Processes with Next-generation Sequencing and Other Whole-genome Approaches

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular & Medical Pharmacology

by

Yin Cheng

2014

ABSTRACT OF THE DISSERTATION

Dissecting Complex Neurological Processes with Next-generation Sequencing and Other Whole-genome Approaches

by

Yin Cheng

Doctor of Philosophy in Molecular & Medical Pharmacology

University of California, Los Angeles, 2014

Professor Yi Sun, Chair

Genome-wide approaches have been successfully applied to obtain a precise and comprehensive picture of the biological and pathological processes underlying neural development and neurological diseases. Currently two main approaches are used to generate large-scale data in a rapid and inexpensive manner, namely the microarrays and the next-generation sequencing (NGS) technologies. The work presented here focuses on elucidating the complex transcriptional network and epigenetic regulation during neural development and neurological disease using whole-genome approaches.

In Chapter 2 of this dissertation, I showed a link between the regulation of DNA methylation and astrocyte differentiation in embryonic neural progenitor cells (NPCs). DNA methylation is one of the essential epigenetic mechanisms involved in regulating gene expression and it is highly dynamic during the development as well as across

different cell types. The precise regulation of DNA methylation is crucial for normal development of central nerve system (CNS). Our lab has previously demonstrated that the *de novo* methyltransferases Dnmt3a is required for neurogenesis in postnatal neural stem cells. Here I showed that the expression of methylcytosine dioxygenase Tet2 is essential for astrocyte differentiation in the NPCs. By analyzing and comparing the gene expression profiles and genome-wide DNA methylation/hydroxymethylation pattern during the differentiation of NPCs, I found that Tet2 preferentially targets the proximal promoter of astrocytic genes. Tet2 mediated DNA demethylation at the promoter sites is essential for the suppression of the astrocytic genes. I also showed that the basic-Helix-Loop-Helix (bHLH) transcription factor Olig2 directly binds to the promoter of Tet2, and Olig2 represses the differentiation towards astrocyte lineage through transcriptional repression of Tet2.

In Chapter 3 of this dissertation, I described the application of Beadarray technology and bioinformatics analysis in characterizing the temporal changes in global gene expression in a spinal cord injury (SCI) mouse model. Using data-driven network based transcription analysis (Weighted Gene Co-expression Network Analysis, WGCNA) coupled with knowledge-driven Gene Ontology (GO) analysis, we can accurately and comprehensively capture the molecular events occur at different stages after SCI. In this study, I showed an example of how global gene profiling can be translated to identify clusters of genes as indicators of functional recovery and genes of interest as potential therapeutic targets.

The dissertation of Yin Cheng is approved.

Samson A Chow

Jing Huang

Bennett G Novitch

Yi Sun, Committee Chair

University of California, Los Angeles

2014

TABLE OF CONTENT

Abstract	ii
List of Figures	Vii
List of Tables	X
Acknowledgements	Xi
Vita	Xii
Chapter 1: Introduction	1
Figures	11
References	20
Chapter 2: Hydroxylation of 5-methylcytosine by Tet2 promotes astrocyte differentiation in embryonic cortical NPCs	24
Abstract	25
Introduction	26
Materials and Methods	28
Results	35
Discussion	43
Figures	46
References	65
Chapter 3: Digitizing molecular events after spinal cord injury	70
Abstract	71

Introduction 72

Materials and Methods 77

Results 80

Discussion 85

Figures 89

References 106

Chapter 4: Summary 110

LISTS OF FIGURES

Chapter 1

Figure 1.1	Cortical progenitor cells follow an intrinsic developmental sequence both *in vivo* and *in vitro*	11
Figure 1.2	Dynamic modification of cytosine by TET and TDG	13
Figure 1.3	Summary of Illumina's Solexa Sequencing Technology	14
Figure 1.4	Overview of the ChIP–seq experiment	17
Figure 1.5	Overview of the WGCNA methodology	19

Chapter 2

Figure 2.1	ChIP-on-chip overview	46
Figure 2.2	Olig2 overexpression represses astrocyte differentiation *in vitro*	47
Figure 2.3	Olig2 binding sites in E11 cortical NPCs and the GO enrichment for its targets	48
Figure 2.4	ChIP-chip and ChIP-qPCR validation of Olig2 targets in E11 cortical NPC	49
Figure 2.5	Expression level and PolII enrichment of Olig2 targets	50
Figure 2.6	Olig2 represses astrocyte differentiation through its down stream targets	51
Figure 2.7	mRNA level of Tet and Gadd45 family members during the spontaneous differentiation of E11 NPCs	52
Figure 2.8	Immunohistochemistry and quantification of Map2+ neurons in differentiating control and Tet2 overexpressing NPCs	53
Figure 2.9	Immunohistochemistry and quantification of CNPase+ oligodendrocytes in differentiating control and Tet2 overexpressing NPCs	54
Figure 2.10	Immunohistochemistry and quantification of Gfap+ astrocytes in differentiating control and Tet2 overexpressing NPCs	55

Figure 2.11 Exogenous Tet2 induces Gfap expression 56

Figure 2.12 Global differential expression induced by Tet2 overexpressed 57

Figure 2.13 Tet2 preferentially affects the expression of astrocytic genes 59

Figure 2.14 Expression changes of the lineage genes 61

Figure 2.15 DNA methylation and hydroxymethylation surrounding transcription start sites (TSS) in differentiating NPCs 62

Figure 2.16 Exogenous Tet2 induces demethylation at Gfap promoter 63

Figure 2.17 Summary of Tet2 and Olig2's roles in astrocytic differentiation 64

Chapter 3

Figure 3.1 Temporal changes in global gene expression changes after SCI 89

Figure 3.2 Top GO pathways enriched in the differentially expressed genes 90

Figure 3.3 WGCNA coexpression network 91

Figure 3.4 Ordered eigengenes expression profiles of each module 92

Figure 3.5 Correlation heatmap of between module eigengenes and sample groups 93

Figure 3.6 Eigengene expression changes in injury only group 94

Figure 3.7 Average expression level of module genes with PBS and 2MeSADP treatment 95

Figure 3.8 The top GO pathways enriched in each module 96

Figure 3.9 Average expression level of neuronal genes 99

Figure 3.10 Average expression level of oligodendrocyte lineage genes 100

Figure 3.11 Average expression level of astrocytes lineage genes 101

Figure 3.12 Average expression level of genes involved in immune response and inflammation 102

Figure 3.13 Average expression level of genes involved in cell cycle 103

Figure 3.14 Average expression level of P2ry1 downstream members 104

List of Tables

Chapter 2

Table 2.1 Component of Lysis buffer 1,2 and 3 65

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

I would like to express my deepest gratitude to my mentor, Dr. Yi Sun, for her excellent guidance, caring, patience, and providing me with an excellent atmosphere for research, not to mention her unsurpassed knowledge of neuroscience and epigenetic regulation. I would like to thank my committee members, Dr. Samson A Chow, Dr. Jing Huang, and Dr. Bennett Novitch, whose good advice and support have been invaluable on both an academic and a personal level.

I am also thankful to many current and former laboratory members, Dr. Hao Wu, Dr. Jun Koyama, Dr. Rosemarie Tsoa, Dr. Kent Ho, Gun Woo Byeon, Dr. Bruno Blanchi, Dr. Volkan Coskun, Weihong Ge, Dr. Quan Lin, Quan Li and Xiaoying Chen. I would like to especially thank Jun Koyama, Rosemarie Tsoa, Kent Ho, and Gun Woo Byeon for their help and contribute to the work in Chapter 2. I would like to thank the Sofroniew laboratory, especially Dr. Michael Sofroniew, Dr. Yan Ao, Dr. Joshua E. Burda and Dr. Ziru Zhao for their input on the work in Chapter 3. I would also like to thank Dr. Steve Horvath and Dr. Matteo Pellegrini for their help with data analysis.

In addition, I would like to acknowledge the financial and academic support of UCLA and the Chinese Scholarship Council.

Last but not least, I would like to thank my parents Jiaan Cheng and Xiaoyun Chen, my husband Jixuan Jiang for their unconditional love and support.

VITA

University of Leeds, School of Bioinformatics Leeds, UK

- M.S, Bioinformatics and Computational Biology (Joint program with Zhejiang University), Sep 2006 – Sep 2007

Zhejiang University (ZJU), College of Life Sciences Hangzhou, CN

- B.S., Bioinformatics, Sep 2003 - Sep 2007

PUBLICATIONS

- **Yin Cheng**, Rosemarie Tsoa, Volkan Coskun V, Jun Kohyama, Sun, Y.E Tet2 mediated epigenetic changes regulate astrocyte differentiation during embryonic stage (Manuscript under preparation)

- **Yin Cheng**, Ziru Zhao, Weihong Ge, Joshua Burda, Michael Sofroniew, Yi Eve Sun Genome-wide gene expression profiling in a spinal cord injury model (Manuscript under preparation)

- Brian Wilburn, Dobrila D. Rudnicki, Jing Zhao, Tara Murphy Weitz, **Yin Cheng**, Xiaofeng Gu, Erin Greiner, Chang Sin Park, Nan Wang, Bryce L. Sopher, Albert R. La Spada, Alex Osmand, Russell L. Margolis, Yi E. Sun, and X. William Yang (2011) A Novel Antisense CAG Repeat Transcript at JPH3 Locus Mediating Expanded Polyglutamine Protein Toxicity in Huntington's Disease-Like 2 (HDL2) Mice, Neuron 70(3), 427-40

- **Yin Cheng**, Biaoyang Lin et al., (2010) Identification of Novel SNPs by Next-Generation Sequencing of the Genomic Region Containing the APC Gene in Colorectal Cancer Patients in China, OMICS 14(3), 315-325

- Biaoyang Lin, Jun Wang, **Yin Cheng** (2008) Recent Patents and Advances in the Next-Generation Sequencing Technologies. Recent Patents on Biomedical Engineering Vol.1 60-67

Chapter 1

Introduction

Regulation of neural development in the embryonic stage

Neural progenitor cells (NPCs) (Fan et al.) are the self-renewing, multipotent stem cells that possess both the ability to proliferate and self-renew and to differentiate into three main types of cells in the central nervous system (CNS), namely neurons, astrocytes and oligodendrocytes. The development of the CNS is tightly regulated both temporarily and spatially(Wen et al., 2009, Juliandi et al., 2010). Both *in vivo* and *in vitro,* NPCs first differentiate into neurons then glial cells(Feng et al., Juliandi et al.). In the mouse cortex, neurons are mainly produced between embryonic day E10.5 to E16, astrocytes are main produced between E14 to neonatal and oligodendrocytes are mainly generated postnatally. Mouse cortical NPCs dissected from different embryonic stages also show different differentiation potentials *in vitro*. NPCs dissected from early embryonic stages (E10.2-E14) mainly give rise to neurons, even cultured in the presence of glial induction factors (i.e. BMP, LIF). During *in vitro* culturing, NPCs gradually acquire the competence for gliogenesis and lose their neurogenic potential (Fig 1.1), indicating the existence of an intrinsic switch mechanism from neurogenic to gliogenic (Sauvageot and Stiles, 2002, Fan et al., 2005). While extrinsic environmental signals and the specific transcription factor and co-factor networks are important in regulating cell fate (Wen et al., 2009, Juliandi et al., 2010, Feng et al., 2007), epigenetic modifications, such as histone modification, DNA methylation, chromatin remodeling and non-coding RNAs are also crucial in mediating the proper regulation of stage specific gene expression (Moore et al., 2013, Takizawa et al., 2001).

DNA methylation as one of the major epigenetic regulation mechanisms has been postulated to regulate the cell fate of NPCs(Moore et al., 2013). DNA methylation is

catalyzed by DNA methyltransferases (Dnmt). There are three members in the Dnmt family, the *de novo* DNA methyltransferases (Dnmt 3a, Dnmt 3b) and the maintenance DNA methyltransferases (Dnmt1). DNA methylation can interfere with the binding of transcription factors and recruit methyl-binding proteins (MBD), which then form complexes with histone deacetylase (HDAC) to initialize the transformation from open chromatin to heterochromatin state. Although DNA methylation was previously considered to be solely repressive, our lab has demonstrated that DNA methylation at the proximal promoters lead to silencing of the downstream genes, while DNA methylation at non-proximal promoter sites, such as gene bodies and distal promoters can act to facilitate the expression of the downstream genes and stabilize their transcription (Wu et al., 2010).

DNA methylation was once considered to be an irreversible modification, which can only be removed passively through cell division. Recent studies have suggested otherwise. The identification of Gadd45 family and Tet family indicates that DNA methylation is highly dynamic and the regulation of DNA methylation plays important role during the development, maturation and maintenance of the CNS (Barreto et al., 2007, Sultan and Sweatt, 2013, Ito et al., 2010, Ito et al., 2011, Kohli and Zhang, 2013). There are three members in the Tet family, Tet1, Tet2 and Tet3. During the active demethylation process, the methylated cytosine (5mC) is first converted to 5-hydroxymethylcytosine (5hmC, which is also known as the 'sixth base' while 5mC being the 'fifth base') by Tet proteins. 5hmC can be further oxidized into 5-formylcytosine (5fmC) and 5-carboxylcytosine (5caC) by Tet proteins. The thymine DNA glycosylase (TDG) mediates conversion of 5hmC/5fmC/5caC to demethylated cytosine (5C) through base

excision (He et al., 2011, Wu and Zhang, 2011).

Basic Helix-Loop-Helix transcription factors and neural development

Multi-potent NPCs can give rise to neurons, astrocytes and oligodendrocytes through the transmission via intermediate or lineage-committed progenitor cell populations. The differentiation to these three cell types is tightly regulated in a spatial and temporal manner. Several classes of transcription factors have been shown to regulate the differentiation of NPCs and to determine the cell fate. Recent evidences indicate that combinations of transcription factors and their co-factors can establish the molecular codes that determine when and what sub-types of neurons or glial cells will be generated and where the cells will migrate to. For instance, expression of Pax6 in progenitor cells induces the formation of neurons, whereas the loss of Pax6 leads to precocious formation of oligodendrocytes and astrocytes. Expression of Olig2 promotes neurogenesis or oligodendrogenesis depending on its phosphorylation status. Induction of Sox9 and Nfia promotes astroglial or oligodendroglial fate and is essential for switching the neurogenesis phase to gliogenesis phase.

DNA methylation and neural development

While proper regulation of the intracellular signaling and transcription factor pathways is essential for the transition between neurogenesis and gliogenesis, the dynamic changes in DNA methylation is also crucial for proper CNS development (Gabel and Greenberg, 2013). Teter et al. and Takizawa et al have shown that the methylation at the promoter of GFAP, a canonical astrocyte marker, is anti-correlated with the expression of GFAP as

well as the initiation of astrogliogenesis (Takizawa et al., 2001, Teter et al., 1994, Condorelli et al., 1994). CpGs at the Stat3 binding site of Gfap promoter is initially methylated, which blocks the binding to Stat3 and represses the expression of Gfap. With the switch from neurogenesis to astrogliogenesis, the Stat3 binding site becomes largely demethylated, which leads to the activation of Gfap expression.

Our lab has demonstrated that DNA methylation as a key mechanism in maintaining the neurogenisis and preventing premature astrogliogenesis in early NPCs. DNA methylation not only inhibit the activation of gliogenesis through hypomethylation at the promoter of astroglial genes (such as Gfap, S100b, Stat1) (Teter et al., 1994), more importantly, it also represses the activity of JAK-STAT pathway. We have shown an enhanced JAK-STAT signaling as well as early astrogliogenesis in NPCs lacking the maintenance methyltransferase Dnmt1 (Fan et al., 2005). We have also demonstrated the crucial function of the *de novo* DNA methyltransferase, Dnmt3a, in postnatal neurogenesis.

On the other hand, the Tet family members, which are involved in the process of active DNA demethylation, have been implicated in embryonic development and neural differentiation (Shen and Zhang, 2012). However, the underlying machinery of how DNA methylation and demethylation act in coordinate to regulate the fate of NPCs remains elucidated. Here, I look at how the DNA demethylation is regulated during the differentiation of NPCs and how Tet2 can push NPCs towards astrocytic lineage through epigenetic regulation.

Molecular mechanisms underlying acute and sub-acute phase of spinal cord injury

Spinal cord injury (SCI) is partial or complete damage to the spinal cord, which causes permanent changes in strength, sensation and other body functions below the site of the injury. Besides the immediate axonal damage, the presence of inhibitory extracellular matrix molecules and the lack of positive environmental stimuli (i.e. growth factors) can also be detrimental to axonal regeneration and functional recovery in the long term. Emerging evidence suggest that the cross talk between the immune cells and neurons and glial cells is crucial for homeostasis of the functional recovery after injury. Unraveling the detailed molecular events and identifying the key members and pathways would shed light on both understanding the lesion mechanism and developing therapeutic strategies. Due to the technical limitations, previous studies on SCI normally only focus on a small group of genes at certain time point after the lesion, which failed to draw a comprehensive picture of the complex mechanism and molecular networks. Moreover, target specific strategies also give us a biased view on these events. Therefore, we used the global transcriptome analysis spanning both acute and prolonged phase post lesion to elucidate the large amount of molecular and cellular reactions involved.

Sequencing and microarray for genome-wide analysis

Microarray is a solid chip with a collection of microscopic DNA probes attached onto the surface. Microarray technology has been developed for more than two decades and is widely used in biological researches (Maskos and Southern, 1992). DNA microarrays can be used to detect DNA (hybridization between fragmented genomic DNA and the probes) as well as RNA (most commonly the hybridization between cDNA reverse transcribed from mRNA and the probes). The application of microarray ranging from comparative genomic profiling, copy number variation (CNV) detection, single nucleotide polymorphism (SNP) detection, gene expression profiling and chromatin immunoprecipitation on Chip. The core principle behind microarrays is hybridization between the probes and the input DNA fragments. After washing away the non-specific binding and weak binding, the fluorescent label conjugated to the input fragment will generate a signal, whose intensity depends on the amount of input fragments binding to the probes that are presented on the chip. The readout of microarray is the relative signal intensity between two input samples for two-channel microarrays or the raw intensity of each probe for single-channel microarrays. While the traditional microarray features a solid-phase chip, the Illumina BeadChip array techonology deploys silica beads covered with hundreds of thousands of copies of every specific oligonucleotide.

Microarray first made the genome-wide analysis possible and has produced much important information. However, with recent advancements and a radical decline in sequencing costs, most researchers turn to next generation sequencing (NGS) technology. With the advent of capillary electrophoresis-based (CE) Sanger sequencing, scientists gained the ability to elucidate genetic information from any given biological system

(Sanger, 1981). In principle, the concept behind NGS technology is similar to CE—the bases of a small fragment of DNA are sequentially identified from signals emitted as each nucleic acid is re-synthesized from a template strand. NGS achieves high-throughput capacity by performing millions of reactions in a massively parallel fashion, rather than analyzing a single DNA fragments at a time. This advance enables rapid profiling of large stretches of sequences, for example the entire genome. Figure 1.3 presents a brief overview of Solexa's sequencing-by-synthesis strategy. The sample prep methods used differ slightly from that used in ABI's SOLiD system, but the basic strategies are the same. Sample is fragmented into a population of small segments that can be uniformly and accurately sequenced in millions of parallel reactions (Morozova and Marra, 2008, Puritz and Toonen, 2013, Wan et al., 2013). The newly identified strings of bases, called the reads, are then mapped to a known reference genome as a scaffold (resequencing), or assembled in the absence of a reference genome (*de novo* sequencing). The latest instruments (i.e. HiSeq 2500 from Illumina, Ion Torrent from Life Technologies) are capable of producing hundreds GB of data in a single run (Morozova and Marra, 2008). With the development of microarray technology and the availability of NGS, we now can profile the whole transcriptome as well as the genome-wide localization of a particular protein or epigenetic modification (Bernstein et al., 2004, Carey et al., 2009, Milne et al., 2009, Morozova and Marra, 2008).

WGCNA (weighted gene co-expression network analysis)

Microarray and RNA sequencing provide the approaches for genome-wide expression analysis and important insights into transcriptome profiling. While increasing our

understanding of biological processes, how to dissect and understand the complexity of the transcription network at a system-level remains a big challenge.

Gene co-expression analysis identifies differentially expressed and regulated genes that can be highly correlated to the phenotypes and traits under study. Standard differential expression analyses treat each gene as an individual entity, while Weighted Gene Co-expression Network Analysis (WGCNA) (Zhang and Horvath, 2005, Langfelder and Horvath, 2008) organizes genes into relatively few modules (groups of densely interconnected genes), which can reduce high dimensional data acquired from microarray and sequencing experiments and also better describe the functioning of the engine instead of enumerating individual nuts and bolts.

Briefly, each module is grouped base on the pairwise comparison between genes. By adopting the concept of "soft thresholding", the weighted gene coexpression network construction emphasizes higher correlations at the expense of lower correlations. Eigengene is the best representation of the gene expression in a particular module and it is extremely useful in calculating the correlation between modules and the correlation of each individual module to traits and genetic variables. Certain genes in the coexpression modules tend to have higher network connectivity, and they are thus called the intramodular 'hub' genes. Hub genes are can be considered the most central genes inside one module and are more likely to have biological and functional significances (Fuller et al., 2007, Langfelder and Horvath, 2008).

Unbiased, data driven analysis, such as WGCNA coupled with knowledge based analysis (i.e. pathways and gene ontology analysis) allows us to identify pathways (modules) that are highly correlated with the biological event under study and their key drivers (e.g., hub

genes) that are present in a given data.

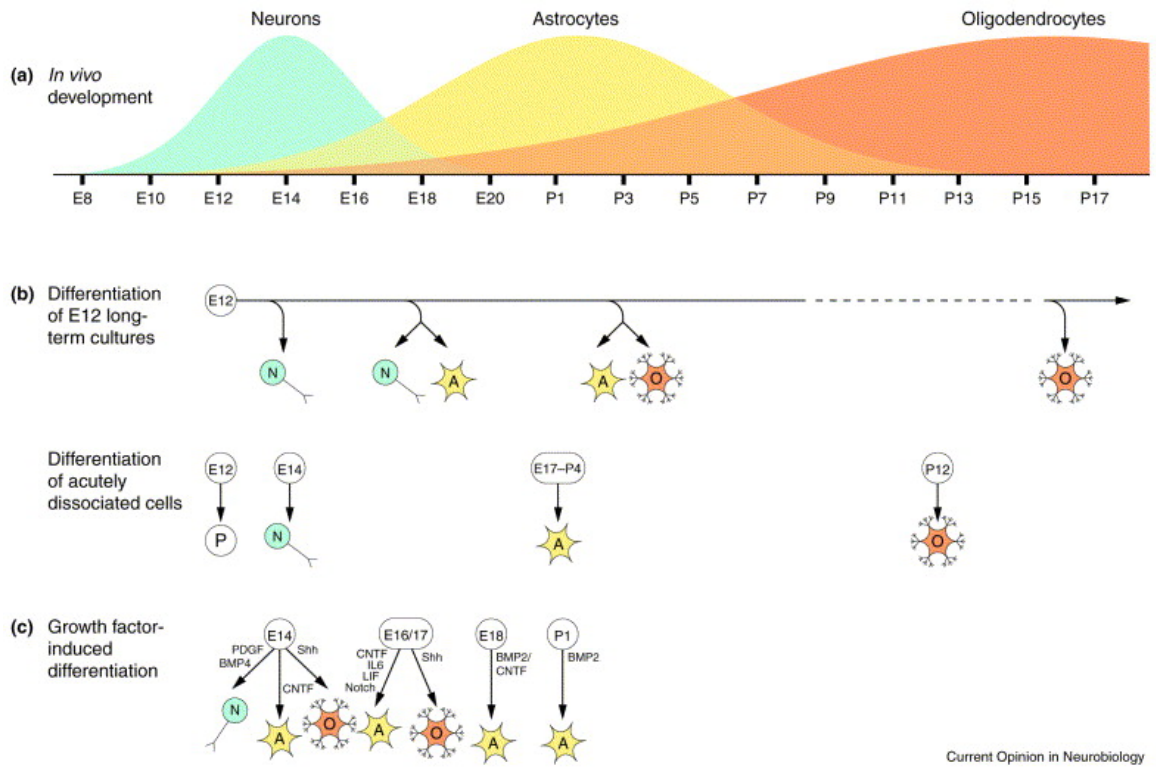


Figure 1.1 Cortical progenitor cells follow an intrinsic developmental sequence both *in vivo* and *in vitro* (Sauvageot and Stiles, 2002). (a) The generation of the neurons, astrocytes and oligodendrocytes in the prenatal and neonatal cortical brain region follows a temporally distinct yet overlapping pattern. (b) *In vitro* cultures recapitulate the differentiation pattern seen *in vivo*, suggesting an intrinsically programmed developmental pattern. Cells dissected at different developmental periods differentiate into progenies that reflect the developmental activity occurring at time of dissection, *i.e.* NPCs isolated at E12 will sequentially give rise to neurons, then astrocytes, and finally oligodendrocytes. (c) Cortical NPCs are fate-restricted and respond differently to extracellular cues (*i.e.* stimulation of growth factors) over time. Cells isolated at E14 showed tri-lineage potential, whereas NPCs isolated in a later stage preferentially adopt glial fates. Cytokines and Notch proteins promote astrocytic fate determination; PDGF

drives neuronal differentiation; Shh promotes oligodendrocyte; BMPs promote either neuronal or astrocytic fates depending on the age of the cell. A, astrocyte; N, neuron; O, oligodendrocyte; P, progenitor cell.

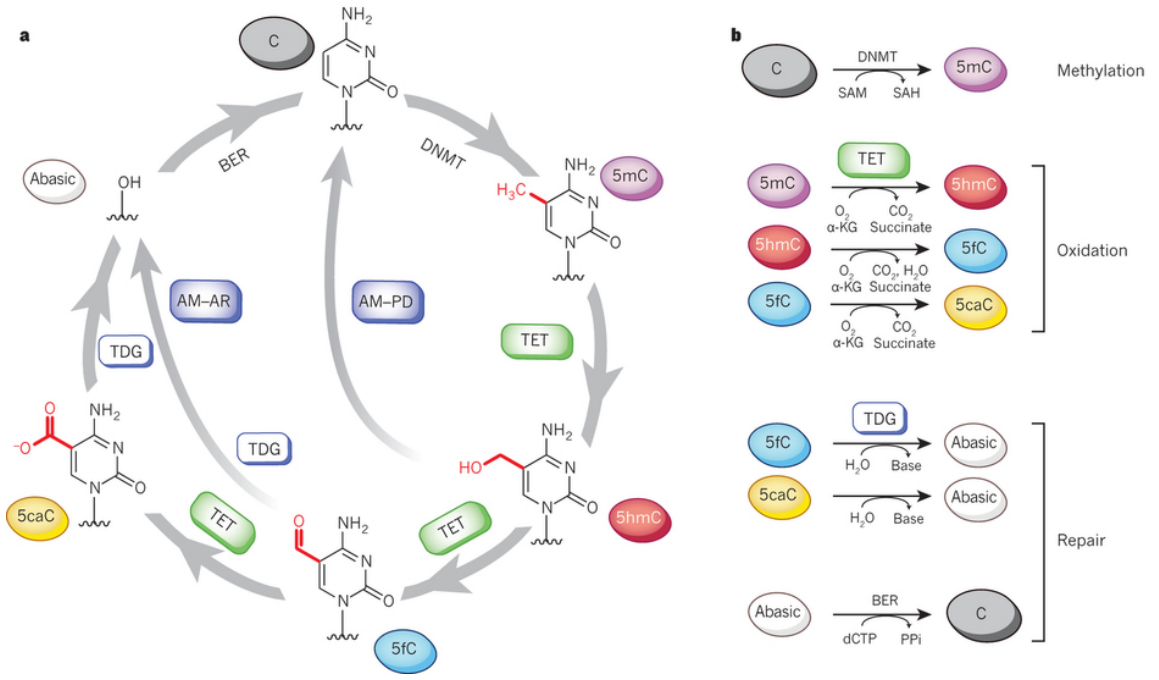
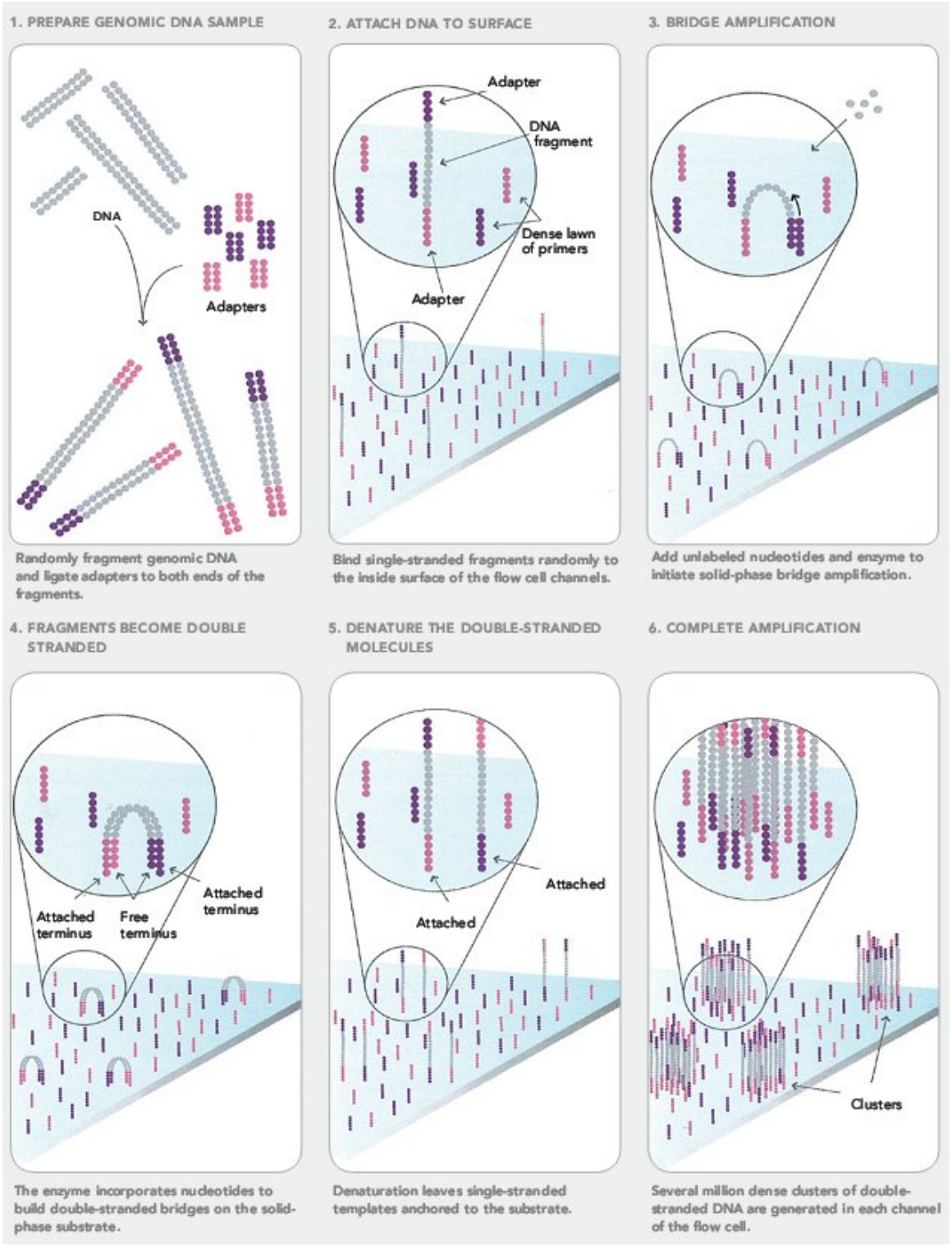


Figure 1.2 Dynamic modification of cytosine by TET and TDG (Kohli and Zhang, 2013). (a) Biochemical pathway of cytosine modification. Cytosine (C) can be methylated to 5mC by DNMTs, which can then be oxidized to 5hmC, 5fC and 5caC by TET. 5hmC can be active modified (AM) and passive diluted (PD) in a replication-dependent manner to regenerate unmodified C; while 5fmC and 5caC can be AM and active restorated (AR) by TDG through an abasic site as part of the base excision repair (Ito et al.) process that regenerates unmodified C. (b) The individual reactions in the pathway. The BER pathway involves excision of the abasic site, replacement of the nucleotide using unmodified deoxycytidine triphosphate (dCTP) by a DNA polymerase (generating pyrophosphate, PPi) and ligation to repair the nick. α-KG, α-ketoglutarate; SAM, *S*-adenosylmethionine; SAH, *S*-adenosylhomocysteine.


1. PREPARE GENOMIC DNA SAMPLE
DNA
Adapters
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.
2. ATTACH DNA TO SURFACE
Adapter
DNA fragment
Dense lawn of primers
Adapter
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.
3. BRIDGE AMPLIFICATION
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
4. FRAGMENTS BECOME DOUBLE STRANDED
Attached terminus
Free terminus
Attached terminus
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.
5. DENATURE THE DOUBLE-STRANDED MOLECULES
Attached
Attached
Denaturation leaves single-stranded templates anchored to the substrate.
6. COMPLETE AMPLIFICATION
Clusters
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

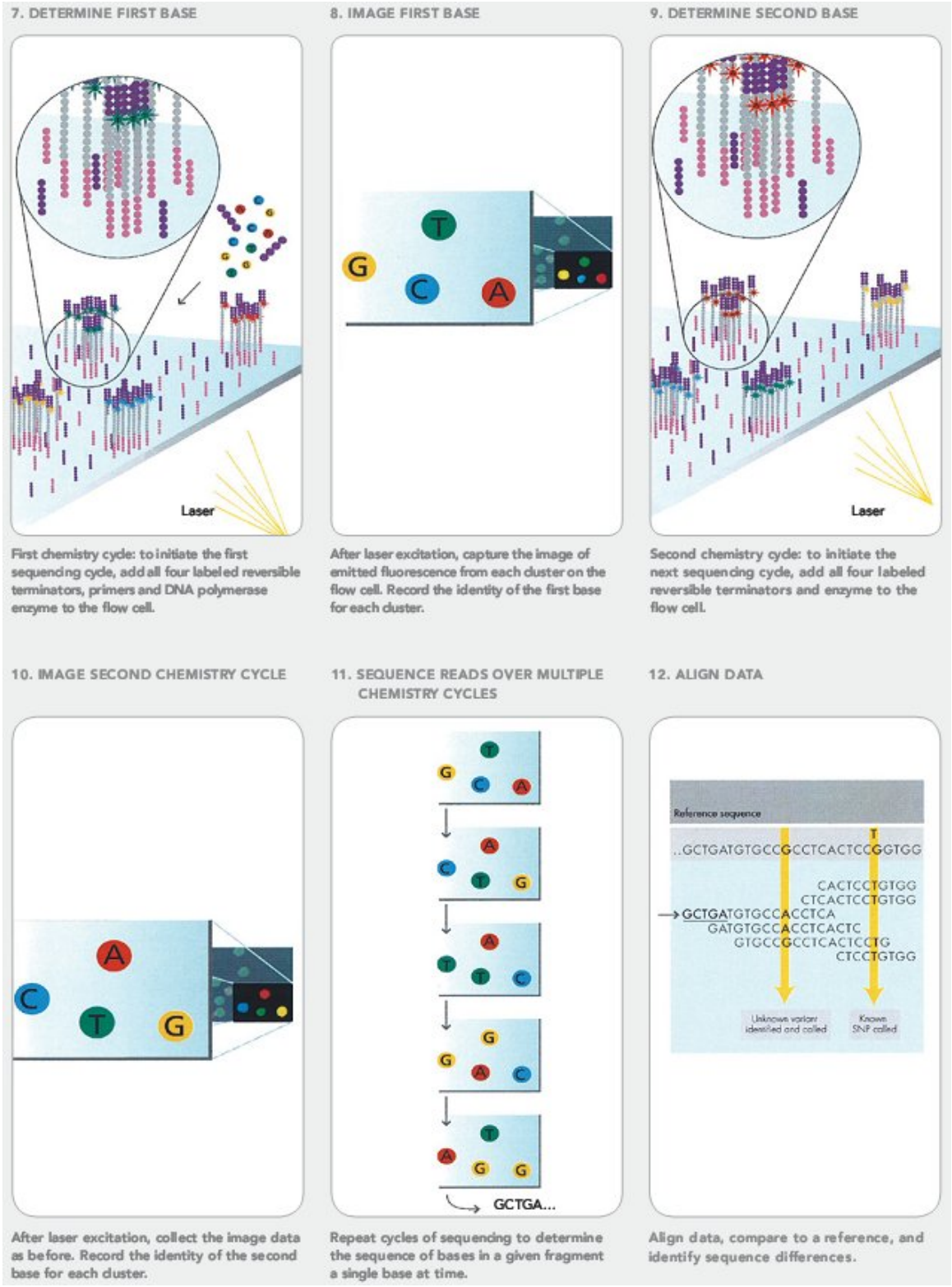
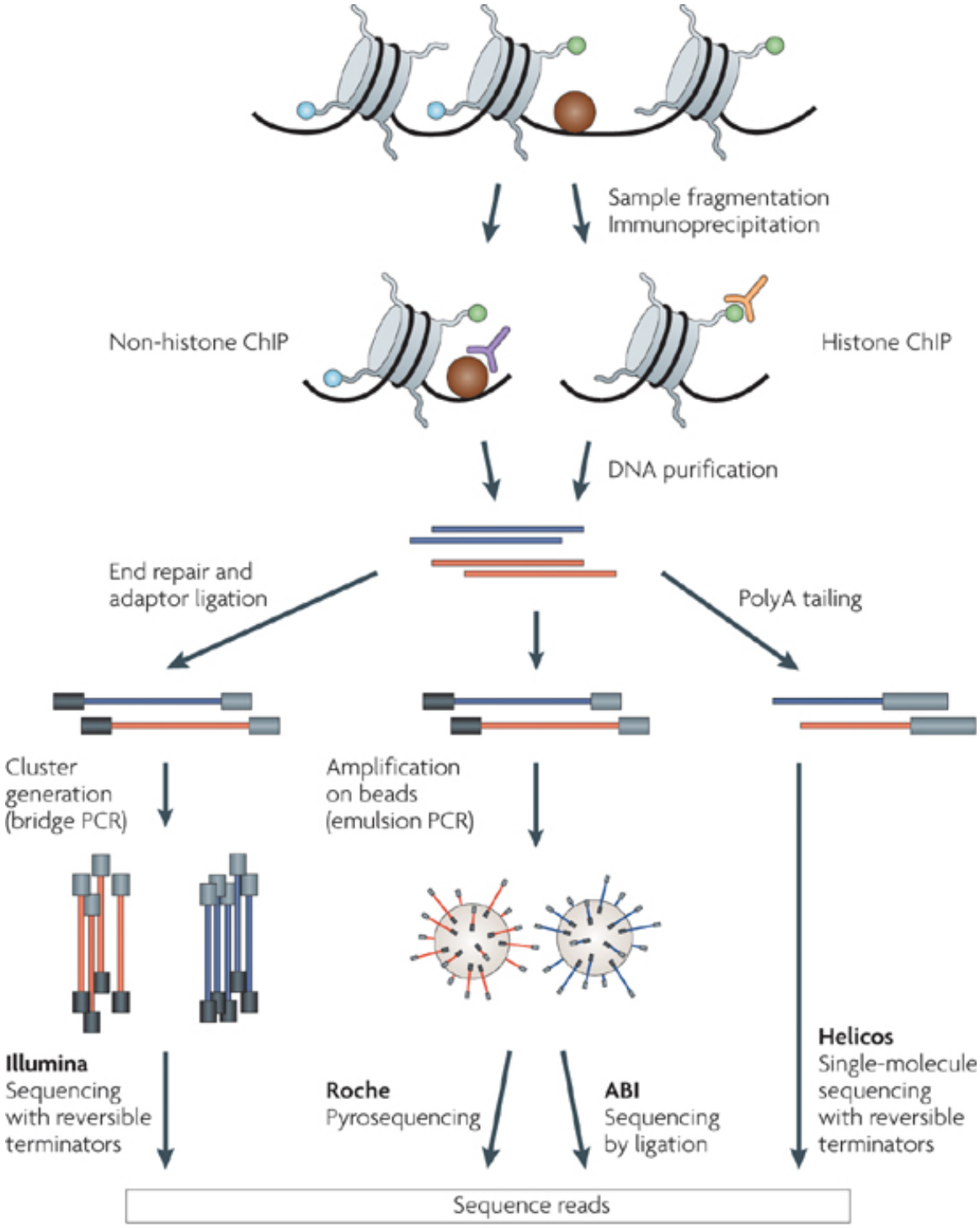


Figure 1.3 Summary of Illumina's Solexa Sequencing Technology (Figure adopted from the Illumina website). 1. The DNA sample (cDNA reverse transcribed from mRNA in

case of RNA-seq) is sheared to appropriate size, end-repaired and ligated to the adapters or indexes plus the adapters in case of multiplexing. Ligated fragments of appropriate sizes (range of 250-400bp) are isolated and amplified using limited cycles of PCR. 2. The flow cell surface is coated with single stranded oligonucleotides that complimentary to the adapters ligated during the sample preparation stage. Single-stranded fragments from the library are bound to the surface of the flow cell. 3-6. Attached fragments are amplified through a "bridged amplification", where repeated denaturation and extension results in localized amplification of single molecules in millions of unique locations across the flow cell surface. 7. A flow cell containing millions of unique clusters is n loaded into the sequencer for automated cycles of extension and imaging. Sequencing cycles consist the incorporation of a single fluorescent nucleotide, followed by high resolution imaging of the entire flow cell. 8. A simplified illustration of the image data collected after one of the sequencing cycles. The fluorescent emission identifies which of the four bases was incorporated at that position. 9-11. This cycle is repeated, incorporating one base at a time, generating a series of images each representing a single base extension at a specific cluster. 12. Base calls are derived with an algorithm that identifies the emission color over time.



Nature Reviews | Genetics

Figure 1.4 Overview of a ChIP–seq experiment (Park, 2009). Specific DNA sites in direct physical interaction with transcription factors and other proteins can be isolated by chromatin immunoprecipitation (ChIP). ChIP produces a library of target DNA sites bound to a protein of interest *in vivo*. Massively parallel sequence analyses are used to analyze the interaction pattern of the protein, such as transcription factors, polymerases and transcriptional machinery, structural proteins, binding or the pattern of any epigenetic modifications.

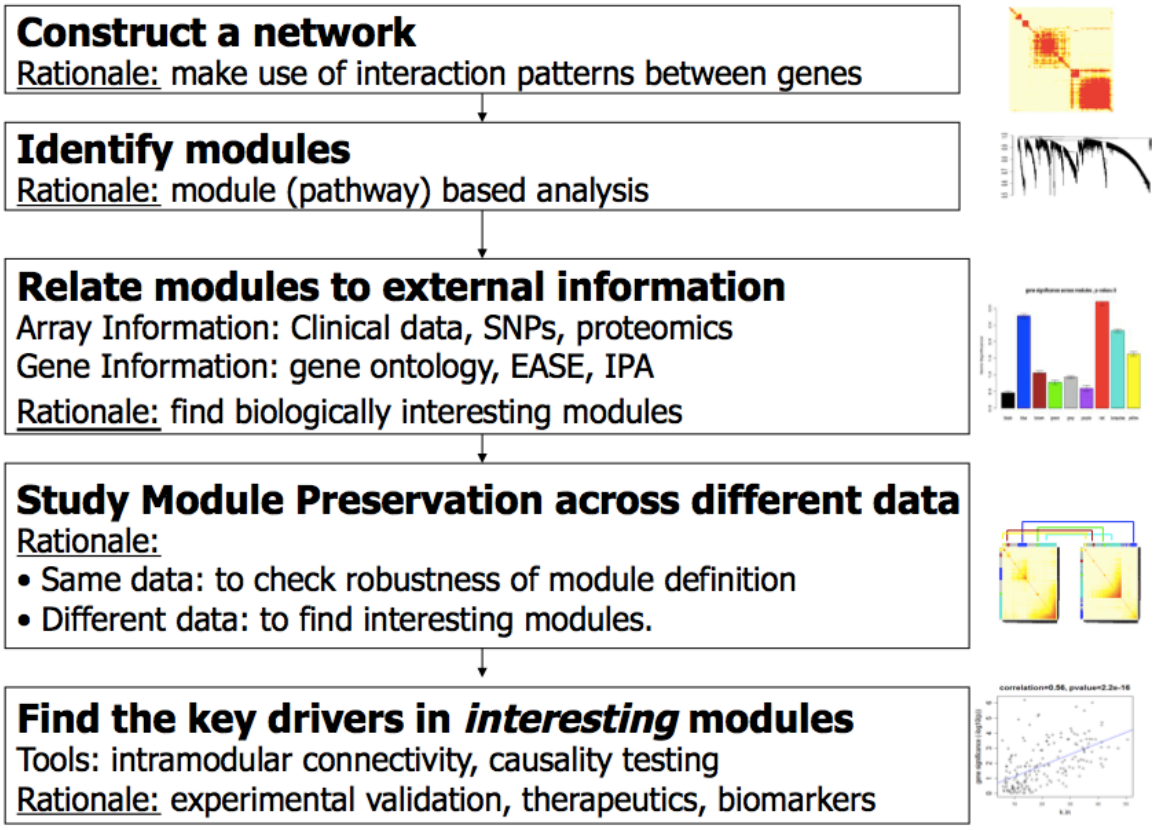


Figure 1.5 Overview of WGCNA methodologies. This flowchart presents a brief overview of the main steps of Weighted Gene Co-expression Network Analysis (modified based on Fig.1(Fuller et al., 2007)).

References

BARRETO, G., SCHAFER, A., MARHOLD, J., STACH, D., SWAMINATHAN, S. K., HANDA, V., DODERLEIN, G., MALTRY, N., WU, W., LYKO, F. & NIEHRS, C. 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature,* 445**,** 671-5.

BERNSTEIN, B. E., HUMPHREY, E. L., LIU, C. L. & SCHREIBER, S. L. 2004. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol,* 376**,** 349-60.

CAREY, M. F., PETERSON, C. L. & SMALE, S. T. 2009. Chromatin immunoprecipitation (ChIP). *Cold Spring Harb Protoc,* 2009**,** pdb prot5279.

CONDORELLI, D. F., NICOLETTI, V. G., BARRESI, V., CARUSO, A., CONTICELLO, S., DE VELLIS, J. & GIUFFRIDA STELLA, A. M. 1994. Tissue-specific DNA methylation patterns of the rat glial fibrillary acidic protein gene. *J Neurosci Res,* 39**,** 694-707.

FAN, G., MARTINOWICH, K., CHIN, M. H., HE, F., FOUSE, S. D., HUTNICK, L., HATTORI, D., GE, W., SHEN, Y., WU, H., TEN HOEVE, J., SHUAI, K. & SUN, Y. E. 2005. DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development,* 132**,** 3345-56.

FENG, J., FOUSE, S. & FAN, G. 2007. Epigenetic regulation of neural gene expression and neuronal function. *Pediatr Res,* 61**,** 58R-63R.

FULLER, T. F., GHAZALPOUR, A., ATEN, J. E., DRAKE, T. A., LUSIS, A. J. & HORVATH, S. 2007. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome,* 18**,** 463-72.

GABEL, H. W. & GREENBERG, M. E. 2013. Genetics. The maturing brain methylome.

Science, 341**,** 626-7.

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**,** 1303-7.

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**,** 1129-33.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

JULIANDI, B., ABEMATSU, M. & NAKASHIMA, K. 2010. Epigenetic regulation in neural stem cell differentiation. *Dev Growth Differ,* 52**,** 493-504.

KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472-9.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

MASKOS, U. & SOUTHERN, E. M. 1992. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. *Nucleic Acids Res,* 20**,** 1679-84.

MILNE, T. A., ZHAO, K. & HESS, J. L. 2009. Chromatin immunoprecipitation (ChIP) for analysis of histone modifications and chromatin-associated proteins. *Methods Mol Biol,* 538**,** 409-23.

MOORE, L. D., LE, T. & FAN, G. 2013. DNA methylation and its basic function. *Neuropsychopharmacology,* 38**,** 23-38.

MOROZOVA, O. & MARRA, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-64.

PARK, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet,* 10**,** 669-80.

PURITZ, J. B. & TOONEN, R. J. 2013. Next-generation sequencing for high-throughput molecular ecology: a step-by-step protocol for targeted multilocus genotyping by pyrosequencing. *Methods Mol Biol,* 1006**,** 89-99.

SANGER, F. 1981. Determination of nucleotide sequences in DNA. *Science,* 214**,** 1205-10.

SAUVAGEOT, C. M. & STILES, C. D. 2002. Molecular mechanisms controlling cortical gliogenesis. *Curr Opin Neurobiol,* 12**,** 244-9.

SHEN, L. & ZHANG, Y. 2012. Enzymatic analysis of Tet proteins: key enzymes in the metabolism of DNA methylation. *Methods Enzymol,* 512**,** 93-105.

SULTAN, F. A. & SWEATT, J. D. 2013. The role of the gadd45 family in the nervous system: a focus on neurodevelopment, neuronal injury, and cognitive neuroepigenetics. *Adv Exp Med Biol,* 793**,** 81-119.

TAKIZAWA, T., NAKASHIMA, K., NAMIHIRA, M., OCHIAI, W., UEMURA, A., YANAGISAWA, M., FUJITA, N., NAKAO, M. & TAGA, T. 2001. DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. *Dev Cell,* 1**,** 749-58.

TETER, B., OSTERBURG, H. H., ANDERSON, C. P. & FINCH, C. E. 1994. Methylation of

the rat glial fibrillary acidic protein gene shows tissue-specific domains. *J Neurosci Res,* 39**,** 680-93.

WAN, M., FARUQ, J., ROSENBERG, J. N., XIA, J., OYLER, G. A. & BETENBAUGH, M. J. 2013. Achieving high throughput sequencing of a cDNA library utilizing an alternative protocol for the bench top next-generation sequencing system. *J Microbiol Methods,* 92**,** 122-6.

WEN, S., LI, H. & LIU, J. 2009. Epigenetic background of neuronal fate determination. *Prog Neurobiol,* 87**,** 98-117.

WU, H., COSKUN, V., TAO, J., XIE, W., GE, W., YOSHIKAWA, K., LI, E., ZHANG, Y. & SUN, Y. E. 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science,* 329**,** 444-8.

WU, H. & ZHANG, Y. 2011. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev,* 25**,** 2436-52.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol,* 4**,** Article17.

Chapter 2

Hydroxylation of 5-methylcytosine by Tet2 promotes astrocyte differentiation in embryonic cortical NPCs

Abstract

In mammalian central nervous system, the differentiations of three major neural cell types (neuron, astrocyte, and oligodendrocyte) are tightly regulated in a temporal manner(Feng et al.). Increasing evidences have pointed out the important role of epigenetic regulation in regards of cell lineage switch during the CNS development. Recent studies have shown that once thought irreversible DNA methylation is in fact a dynamic epigenetic modification, and can be removed through active demethylation mechanisms mediated by either the Tet/Tdg pathway or the Gadd45 family proteins. Our results demonstrated that promoter DNA demethylation by Tet2 is crucial for the activation of astrocytic genes and differentiation towards astrocyte lineage. The bHLH factor Olig2 directly targets and represses the expression of Tet2, and then represses astrogliogenesis.

Introduction

DNA methylation is a major epigenetic factor that has been postulated to regulate cell fate. Our lab has demonstrated that astroglial marker (such as Gfap) and the astrogliogenic members in JAK-STAT pathway (such as STAT1) are repressed through DNA methylation by Dnmt1 in early NPCs, which can prevent premature astrogliogenesis. Selective demethylation at CpG sites surrounding Stat3-binding element (-1557 to -1280 of Gfap promoter) was observed in E11.5 NPCs over *in vitro* culturing (Fig 2.1) (Fan et al., 2005). We have also shown the crucial function of the *de novo* DNA methyltransferase Dnmt3a in neurogenesis in postnatal NPCs (Wu et al., 2010). Nakashima et al analyzed DNA methylation changes in mouse NPCs between the mid (E11.5) and late (E14.5) stages of development and found that many astrocytic genes, including Gfap are becoming demethylated in late-stage NPCs, which then enables the cells to become competent to express these genes.

On the other hand, the Tet family members, which are involved in the process of active DNA demethylation, have been implicated in embryonic development and neural differentiation. However, the underlying machinery of how DNA methylation and demethylation act in coordinate to regulate the fate of NPCs remains elucidated. It has been demonstrated that Olig2 can repress the astroglial differentiation by suppress the activation of astroglial genes. Our preliminary analysis suggested that Olig2 represses the expression of Gfap in a dosage dependent manner, however, Olig2 does not bind to Gfap promoter site directly. Therefore, we hypothesize that Olig2 can repress the astrogliogenesis indirectly through epigenetic regulatory elements.

Materials and Methods

Cell culture

Primary neural progenitor cells (NPCs) are prepared from time-pregnant CD1 mice. Telencephalon was dissected from E11 mice was first coarsely dissociated by mechanical force then treated with Papain (Worthington) for 5 minutes at 37°C with constant shaking. 3×10^6 dissociated cells were then plated onto a poly-ornithine (Sigma, 15 μg/ml in H2O) and fibronectin (Sigma, 2 μg/ml in PBS) coated 10 cm dishes in serum-free medium containing DMEM/F12 (Invitrogen), 1% B27 (Invitrogen), and penicillin-streptomycin (50 μg/ml and 50 U/ml, respectively). Cells were fed with basic fibroblast growth factor (bFGF, PeproTech) at a final concentration of 10 ng/ml on a daily base. NPCs were passaged with enzymatic dissociation using StemPro Accutase (Life Technologies) upon reaching confluency, and re-plated on PO/FN coated plate at a density of $1 – 2 \times 10^6$ cells per 10cm dish. Cultured cells displayed a bipolar shape and over 95% of the cells were Olig2 and Sox2 positive. For NPC differentiation, cells were cultured in medium containing DMEM/F12 (Invitrogen), 1% B27 (Invitrogen), and penicillin-streptomycin (50 μg/ml and 50 U/ml, respectively) without bFGF.

Immunocytochemistry

NPCs were dissociated, plated on coverslips pre-coated with poly-ornithine (Sigma, 15 μg/ml in H2O) and fibronectin (Sigma, 2 μg/ml in PBS) and cultured till confluency or certain days into differentiation.

Cells were fixed with 4% formalin/PBS solution at room temperature for 10 min and

washed with PBS for three times. Fixed cells were permeabilized using 0.4% Trixon-X (Sigma), incubated with primary antibodies at 4°C overnight, followed by secondary antibodies at room temperature for 60 min. Cells were washed with PBS for 3 times after both primary and secondary antibodies. Cells were stained with NPC markers Olig2 (Millipore; 1:2000), Sox2 (Santa Cruz; 1:200), Nestin (Sigma; 1:1000) neuronal marker Tuj1 (Covance; 1:1000), oligodendrocyte marker CNPase (Millipore; 1:1000) or astrocyte marker Gfap (Sigma; 1:1000). Hoechst staining was used to label the nuclei. Images were captured using Olympus fluorescence microscope and processed using Imaris and Adobe Photoshop CS5 software.

Quantitative RT-PCR

Total RNA was extracted using Trizol (TRIzol, Invitrogen) following manufacturer's instructions. Genomic DNA contamination was removed using Turbo DNase (Ambion), and cDNA was synthesized using SuperScript® III First-Strand Synthesis System (Invitrogen) following the manufacturer's instruction. Quantitative RT-PCR was performed on a StepOnePlus Real-Time PCR System (Life Technologies) using Fast SYBR Green Master Mix (Applied Biosystems). Melting curves were analyzed to confirm a single species of each PCR product. Gapdh cDNA was used as an internal control to quantify the relative expression of each cDNA.

Western Blotting

Tissue or cultured cells were homogenized in lysis buffer containing 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P-40, 10% glycerol, and protease

inhibitor cocktail (Roche). The protein concentration was measured using Pierce BCA Protein Assay Kit (Thermo Scientific) following the manufacturer's instruction. The lysates (10 - 25 μg protein per lane) were separated by SDS-PAGE gel (6% - 10%) and transferred to nitrocellulose membrane (Bio-Rad) for immunoblotting. The primary antibodies used were Olig2 (Millipore; 1:2000), Tet2 (Santa Cruz, 1:200), Gfap (Sigma; 1:1000), Flag (Sigma, 1:2500) and β-actin (Sigma, 1:3000). β-actin was used as an internal control.

Chromatin immunoprecipitation

ChIP-on-chip were carried out using Agilent Mammalian ChIP-on-chip protocol (Fig 2.1). In brief, 5×10^7 to 1×10^8 mouse NPCs were dissociated and suspended in 10ml PBS at room temperature. Cells were then chemically crosslinked by adding 1ml 11% formaldehyde (Sigma) solution containing 50mM Hepes-KOH, pH7.5, 100mM NaCl, 1mM EDTA, pH 8.0 and 0.5mM EGTA, pH 8.0 to every 10ml of cell suspension and rotating at room temperature for 10 minutes. 0.5ml of 2.5M glycine (Sigma) was added to neutralize the formaldehyde. Cells were collected by centrifugation at 1,350 ×g for 5 minutes at 4°C and washed with cold PBS twice. Pellets were either flash frozen using liquid nitrogen and stored at -80°C or directly proceeded to the next step.

Crosslinked pellets were then lysed using subsequent treatment of lysis buffer 1, 2, and 3 (Table 2.1) supplemented with protease inhibitor cocktail (Roche). Nuclei pellets were collected by centrifugation and suspended in 0.3ml lysis buffer 3 with 1% Triton-X (Sigma) and sonicated using Bioruptor (Diagenode). Samples are sheared for 15 rounds of sonication cycles (30 sec ON/30 sec OFF) at high power setting with the Bioruptor

combined with the Bioruptor water cooler (Diagenode), resulting fragments of 400 - 600 base pairs (bp) in length.

100 µg of sheered chromatin were mixed with 30 µl of Dynabeads (Dynabeads® M-280 Sheep Anti-Mouse IgG or Dynabeads M-280 Sheep anti-Rabbit IgG, Invitrogen) pre-incubated with 2 – 10 µg of antibodies (mouse/rabbit IgG, Tet2, Olig2) overnight at 4°C on a rotating platform. The beads were washed three times with RIPA buffer the next day. Chromatin was eluted from beads by incubation in elution buffer containing 50mM Tris-HCl, pH8.0, 10mM EDTA, pH 8.0 and 1% SDS at 65°C for 15 minutes with brief mixing on vortex. Eluted chromatin and WCE (whole cell extract) were then reverse-crosslinked at 65°C for 6 hr to overnight. IP and WCE were treated with RNase and protease K and purified using phenol-chloroform extraction. Purified DNA was used in chip, sequencing and qPCR in order to identify Tet2 and Olig2 binding sites.

MeDIP and hMeDIP

MeDIP and hMeDIP were performed as previously described (Blecher-Gonen et al., 2013). For methylated and hydroxymethylated DNA immunoprecipitation (IP), genomic DNA was extracted using phenol-chloroform. 1µg of genomic DNA was used per IP with similar procedure as ChIP described previously.

Purified genomic DNA was fragmented by Covaris (Covaris) and mixed with 1µg of 5mC (eurogentec) and 5hmC (Active motif) antibody conjugated with Dynabeads® M-280 Sheep Anti-Rabbit IgG (Invitrogen) respectively. DNA fragments pulled down from MeDIP and hMeDIP, as well as genomic DNA (input) were end-repaired by T4 DNA polymerase and phosphorylated. A single ‘A’ base was added to the 3′ end with

Klenow. Adaptors with indexes were ligated to the fragments with multiplexing sample preparation kit (Illumina). Ligation products between 300 and 500 bp were purified using AMPure beads (NEB) and amplified by PCR. Libraries were quantified with PicoGreen and QC with Bioanalyzer then analysed by Illumina Hiseq2000 platform.

ChIP targets validation

Site-specific primers were designed for Olig2 and Tet2 binding sites and methylation/hydroxymethylation sites identified from ChIP-chip, ChIP-seq or adopted from previously published studies. Quantitative PCR was performed on a StepOnePlus Real-Time PCR System (Life Technologies) using Fast SYBR Green Master Mix (Applied Biosystems). Fold enrichment were calculated by IP over IgG or IP over WCE.

Whole-genome expression analysis

mRNA library was prepared using NEBNext Utra mRNA Library Prep Kit for Illumina (NEB) following manufacturer's protocol. Briefly, total RNA was extracted following manufacturer's instructions (TRIzol, Invitrogen). mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and fragmented using Covaris (Covaris). cDNA was synthesized using random priming, followed by end repair and 5' phosphorylation. dA-tailing was added and adaptors with indexes were ligated. Ligation product was amplified using PCR, products between 300 and 500 bp were purified using AMPure beads (NEB) and amplified by PCR. Libraries were quantified with PicoGreen and QC with Bioanalyzer before analyzed by Illumina Hiseq2000

platform.

Data analysis

Data QC, trimming and mapping

FASTQ files were generated from HiSeq2000 with at approximately 25 million reads (50bp) per lane. Reads were first de-multiplexed according to the corresponding indexes. Quality control was performed using FastQC (The Picard BAM/SAM Libraries), the indexes and base sequences with a quality score below 20 are removed. Trimmed reads were aligned to mouse reference genome (MM10) using TopHat and Bowtie2 (Asmann et al., 2011, Roberts et al., 2011a, Roberts et al., 2011b) with default settings (allow maximum 2 mismatches).

RNA-seq analysis

Reads were mapped to the mouse genome (mm10) using Bowtie and TopHat using default parameters. BAM files generated from mapping were then submitted to Cuffdiff for differential gene expression detection. Results were visualized using UCSC Genome Browser and CummeRbund (Kent et al., 2002).

ChIP and MeDIP/hMeDIP data analysis

Reads were mapped to the mouse genome (mm10) using Botwie2. Binding peaks were identified using MACS (Feng et al., 2011) with default settings and visualized using UCSC Genome Browser (Kent et al., 2002). MeDIP and hMeDIP data were first analyzed using MEDIPS with the parameters suggested in the manual (Lienhard et al.,

2014). The average signal enrichment of methylation/hydroxymethylation at TSS was plotted using SitePro from the CEAS (Cis-regulatory Element Annotation System) with a profiling resolution of 100nt and spanning 1500bp of the TSS region (Shin et al., 2009, Liu et al., 2011).

Functional Annotation and Statistical Significance of Gene Lists

Functional annotation was performed with the online tool, DAVID (http://niaid.abcc.ncifcrf.gov/) (Dennis et al., 2003)

Results

Olig2 acts to repress NPCs differentiation towards astrocyte lineage *in vitro*

It has been previously reported that Olig2 is crucial for the generation of motor neurons and oligodendrocytes in the CNS and the phosphorylation status of Olig2 is critical in regulating the proliferation of neural progenitors and the cell fate (Li et al., 2011). Besides its function in the motor neuron and oligodendrocyte lineage, Olig2 also acts to repress the astrocyte lineage in embryonic mouse cortex. Forced expression of Olig2 in NPCs lead to increase the percentage of oligodendrocyte and decrease in the precentage of astrocytes both *in vitro* and *in vivo* (Setoguchi and Kondo, 2004, Zhou et al., 2001). Therefore, we first tested if this phenomenon can be recapitulated in our *in vitro* culturing system.

NPCs derived from telencephalon of E11 mouse embryos were cultured *in vitro* with 10ng/ml bFGF for 1 to 2 passages (P1 to P2). NPCs cultured *in vitro* express progenitor markers Nestin and Sox2, and were positive for Olig2 as well (data not shown here). To study the relationship between Olig2 and astrocyte differentiation, NPCs were infected with FG12 (ctrl) or FUIGW-hOlig2-HA (HA tagged human Olig2 cDNA) lenti-virus at least 6 days before differentiation. Infection rate was calculated base on number of GFP positive cells over total number of cells (GFP+/total) and the overall infection rate of 85% to 95% was achieved. Spontaneous differentiation was induced by withdraw bFGF when cells reached approximately 70% confluency. NPCs were harvested at different time points up to 30 days after bFGF withdraw for immunohistochemistry, total RNA was also extracted for qPCR and RNA-seq.

NPCs did not express astrocyte marker Gfap when they were maintain in a progenitor cell stage (cultured with bFGF). Upon differentiation, we observed less Gfap+ cells in Olig2 overexpressing cells comparing with the control cells. We also noticed that Olig2 and Gfap were mutually exclusive (Fig. 2.2a, 2.2b). Not surprisingly, overexpressing Olig2 also repressed the mRNA level of Gfap, qPCR results showed a reduced Gfap mRNA level in cells with forced expression of exogenous Olig2 (Fig 2.2c).

Genome-wide binding profile of Olig2

Olig2 contains one basic helix-loop-helix (bHLH) domain and binds to a DNA consensus sequence (CANNTG) known as the E box (Tabu et al., 2006, Yu et al., 2013). ChIP-chip experiment using NimbleGen mouse promoter array was performed to profile the binding sites of Olig2 in E11 cortical NPCs.

5140 binding sites were identified using MACS with FDR<0.05 through out the mouse genome, among which 2978 were located in the proximal promoter region (-3kb to +1kb from TSS). Gene ontology (GO) analysis using DAVID showed that majority of the Olig2 targets are associated with neuron differentiation, axongenesis, cell proliferation and neuron projection (Fig 2.3), which suggested Olig2 is heavily involved in the development of central nervous system.

While Olig2 repressed the expression of classic astrocytic genes, such as Gfap and S100b, Olig2 did not bind their promoters. Comparing with direct targets of Olig2 (*i.e.* Tet2, Nfia), the fold enrichment of Olig2 is much lower at the promoter of Gfap and S100b (Fig 2.4a). ChIP-qPCR validation showed the same results as ChiP-chip experiment (Fig 2.4b).

To further investigate the expression regulation by Olig2, we looked at both the mRNA level and the binding of PolII (RNA polymerase II) as a direct indication of active transcription at the TSS (transcription starting site) of both direct and indirect Olig2 targets up to 10 days into spontaneous differentiation. Upon spontaneously differentiation, Olig2 expression was immediately dropped to 15 - 20% of its original level. Both Tet2 and Gfap were expressed at an extremely low level in undifferentiated NPCs. Tet2 expression increased within 2 days of differentiation and maintained at a relative high level before it started to decrease at day 6; while Gfap expression continuously increased (Fig 2.5a-c). Enrichment of PolII at the Tet2 TSS also showed a transient increase and then decreased, while PolII binding showed a continuous increase at the Gfap TSS. On the other hand, Olig2 binding was only seen at Tet2 promoter, but not the Gfap promoter in undifferentiated NPCs. The decrease of Olig2 ‘s enrichment at Tet2 promoter was correlated with the expression decrease of Olig2.

Our observation that Olig2 represses astrocytic genes without direct association of Olig2 to their promoters suggested Olig2 is repressing the expression and may be also the astrocyte differentiation indirectly. Our ChIP-chip results indicate that Olig2 is enriched at the promoter of multiple genes that are known for driving astrogliogenesis, such as Nfia, Notch1, Id2 and Id4. Therefore, we hypothesized that rather than directly repress the expression of astrocytic genes, Olig2’s repression on astrocyte differentiation through direct repressing of astrocytic differentiation factors Nfia, Notch1, ID2 and ID4 and possibly Gadd45a, Tet2 (Fig 2.6).

Tet2 overexpression induces astrocyte differentiation

Target-specific DNA methylation and demethylation are crucial for the switch from neurogenesis to astrogliogenesis in NPCs, we hypothesize that Tet2 and Gadd45a, both being the direct targets of Olig2 and implicated in DNA demethylation, act as the intermediates in Olig2's repression of astrocyte differentiation.

Both Tet2 and Gadd45a were identified as the direct binding target of Olig2; their expression was repressed with Olig2 overexpression. However, when we looked at the expression change upon spontaneous differentiation, Tet2 and Gadd45a showed different patterns. Tet2 expression was immediately induced after bFGF withdraw, while Gadd45a expression was barely detectable even after 10 days of differentiation. Therefore, we hypothesized that Olig2 can repress the initiation of astrocyte differentiation through epigenetic modification mediated by Tet2. On the other hand, Gadd45s might be involved in the maintenance and maturation of astrocytes.

To investigate Tet2's function in NPC differentiation, we compared the differentiation potential of E11 NPCs infected with control (pMXs) and Tet2 overexpressing (pMYs-Tet2CD, Tet2 catalytic domain) virus. Exogenous Tet2 expression gave rise to a reduced percentage of Map2 positive cells comparing with control at day3 but not at day7 of differentiation, suggesting that Tet2 may repress or delay the differentiation towards neuronal lineage (Fig 2.8). Exogenous Tet2 did not seem to affect the percentage of CNPase positive cells (Fig 2.9). However, the overall percentage of CNPase positive cells was extremely low. Gliogenesis in mice occurs at neonatal stage, it is possible that NPCs derived from E11 cortical region do not have the full potential of generating oligodendrocytes. On the other hand, we used a simple culturing condition, which may not favor the oligodendrocyte lineage; it is difficult to say that Tet2 does not

affect oligodendrocyte differentiation. Tet2 overexpression promoted NPCs differentiation towards astrocyte lineage. NPCs with Tet2 overexpression showed a three-fold increases in the percentage of Gfap positive cells comparing to control NPCs (Fig 2.10a,b). As expected, Olig2 overexpression largely decreased the percentage of Gfap positive cells (Fig 2.10a,b).

Exogenous Tet2 induces Gfap expression

We have shown that Tet2 overexpression promoted astrocyte differentiation. Then we look at the mRNA and protein level of the astrocyte marker Gfap with Tet2 or Olig2 overexpression, Tet2 shRNA knock-down and double overexpression of Tet2 with Olig2. NPCs were infected with the corresponding virus and differentiated for 7 days with bFGF withdrawal. As expected, Tet2 overexpression significantly promoted the expression of Gfap and Tet2 knockdown lead to the opposite trend (Fig 2.11a,b). Although NPCs with Olig2 overexpression did not show an increase in the number of CNPase positive cells, Olig2 did promote the expression of CNPase (Fig 2.11a). Both Olig2 overexpression and Tet2 knock-down showed a decreased level of Gfap expression (Fig 2.11a,b). Tet2 overexpression was able to partially rescue the repression of Olig2 (Fig 2.11a).

Genome-wide expression changed induced by Tet2

E11 cortical NPCs infected with control (pMXs) or Tet2 overexpressing (pMYs-Tet2CD, Tet2 coding region) virus were cultured with bFGF (Day 0) and differentiated for 3 days (Day 3) or 7 days (Day7). Total RNA was extracted and sequenced on the HiSeq 2000

using single-end (1×50bp) multiplex sequencing. Total number of raw reads generated was 331 million. After de-multiplexing, there were 18.57 million reads in each sample on average, and 89.7% of the de-multiplexed reads can be uniquely mapped to the mouse Refseq database (MM10).

RNA-seq analysis identified more than 1000 significantly differential expressed gene in pairwise comparison. A general clustering based on Pearson correlation coefficients showed that the overall expression were extremely similar between the control and Tet2 overexpressing NPCs at undifferentiated stage (Day 0), there were merely 6 genes that were significantly differential expressed between the two conditions (Fig 2.12a,b, Fig 2.13c). Control and Tet2 overexpressing NPCs showed very similar expression regulation during the differentiation. 878 and 1136 genes that were significantly up-regulated or down-regulated comparing during differentiation in control and Tet2 overexpressing NPCs respectively and 731 of them overlap (Fig 2.12c). This indicates that Tet2 does not affect the expression profile of the progenitors and Tet2 does not alter the general process of differentiation.

We then look at the difference between control and Tet2 overexpressing NPCs at each time point. The number of differential expressed genes was relatively small, there are 6, 39 and 70 genes showed a significant difference at day 0, differentiation day 3 and day 7 respectively. Interestingly, all of them were up-regulated by Tet2 (Fig 2.13. a, c). This matches Tet2's function in DNA demethylation, as DNA methylation is largely considered to be a repressive marker and removal of DNA methylation can potentially release the repression. When we look into the function of the genes that are up-regulated by Tet2, we observed a bias over the astrocyte lineage. Majority of the

genes that were up-regulated are astrocyte specific genes, there were some oligodendrocyte specific genes and very few neuronal genes (Fig 2.13b). Expression of the lineage specific genes also showed the same trend. Astrocytic genes were up-regulated in Tet2 overexpressing NPCs (Fig 2.14), indicating that Tet2 preferentially up-regulate the astrocytic genes.

Epigenetic landscape during the differentiation of NPCs

To investigate the changes in DNA methylation during the differentiation of NPCs, we performed genome-wide DNA methylation and hydroxymethylation profile of the undifferentiated NPCs (day 0) and NPCs after 3 days into differentiation (Day 3) using MeDIP-seq and hMeDIP-seq. Sequencing results from fragmented DNA (input) at day 0 and day 3 were used as background respectively. Sequencing were performed using Illumina Hiseq2000, a total number of 269 million reads were obtained, which gave us over 30 million mappable reads per sample.

We focused on the distribution of DNA modification in promoter regions and plotted the average enrichment of 5mC and 5hmC surrounding the TSS (+-1.5kb). There as no obvious change on the average enrichment of 5mC or 5hmC surrounding the TSS between day 0 and day 3 (Fig 2.15). We then looked at the genes that were induced by Tet2 over-expression (Fig 2.15b, d) and the binding targets of Tet2 (Fig15a, c) (Chen et al., 2013). We observed a decrease of 5mC level at -1.5kb to -1kb from the TSS of both the Tet2 binding targets and genes that were induced by Tet2 over-expression; and we also observed an increase in the 5hmC level at those loci, suggesting Tet2's role in inducing gene expression via target specific DNA demethylation.

Tet2 is involved in the demethylation of Gfap promoter

Since Tet2 is implicated in DNA demethylation, we then ask whether Tet2 is directly involved in regulation of DNA methylation. We used Gfap, whose expression was significantly unregulated by Tet2 overexpression as an example, and looked at the methylation level at three different loci of Gfap TSS and promoter. Primary cultured astrocytes were used as control. We saw a significant decrease in DNA methylation at both Gfap proximal promoter and TSS in Tet2 overexpressing NPCs even at the progenitor stage. After 5 days of differentiation, the methylation level in Tet2 overexpressing cells was not significantly different from the primary astrocyte. We did not see the demethylation in Olig2 overexpressing cells nor Tet2 knockdown cells (Fig 2.16), indicating the Tet2 is crucial for target specific demethylation at the promoter of astrocytic genes during NPC differentiation. Although the DNA methylation level was decreased in undifferentiated NPCs, we did not see an increase in its mRNA level, indicating another layer of regulation besides DNA methylation.

Discussion

In the mammalian central nervous system (CNS), the differentiation of the three major neural cell types (neuron, astrocyte, and oligodendrocyte) is tightly regulated in a spatial and temporal manner. As one of the basic-helix-hoop-helix (bHLH) transcription factors, Olig2 plays a central role in guiding oligodendrocytes and motor neuron development and also shows inhibitory effects on astrocytic differentiation, yet the mechanism remains elusive. Spontaneous differentiation of *in vitro* cultured NPCs can give rise to neurons, oligodendrocytes as well as astrocytes, while forced expression of Olig2 leads to a decrease in astrocytes population as well as a decrease in the expression of astrocytic genes, *i.e.* Gfap, Aldh1l1, S100b. To investigate the mechanism underlying Olig2-dependant repression on astrocytic lineage, we mapped genome-wide Olig2 occupancy in embryonic NPCs by ChIP-chip. I identified 5140 binding sites with high confidence (FDR<0.05), among which 2978 were located in the proximal promoter region (-3kb to +1kb from TSS). Mapping Olig2 binding sites to the regions flanking TSS revealed that Olig2 does not physical associated to the promoter of the astrocytic genes, despite the fact the expressional repression. On the other hand, Olig2 binds to Id2/4, Notch1 and Nfia, which are known to play a crucial role in the onset of astroglial development.

It has been reported that the Gfap promoter, especially the Stat3 binding region is heavily methylated in early embryonic stages and becomes demethylated in late-stage NPCs prior to the onset of astrocyte differentiation (Fan et al., 2005, Takizawa et al., 2001, Wu et al., 2010, Sultan and Sweatt, 2013, Condorelli et al., 1994). Both Gadd45a and Tet2 have

been implicated in DNA demethylation, and I hypothesize that Olig2 can indirectly regulate the DNA methylation level at the promoter of astrocytic genes by repressing genes related to DNA demethylation. To test this hypothesis, I first looked at the expression change of Gadd45 and Tet family members during spontaneous differentiation. The expression of Tet family members picked up within the first 2 days upon bFGF withdraw, and started to drop after day 6; while Gadd45a expression did not start to increase till the later stage (Day 7 to Day 15). These data indicate that instead of triggering the initiation of astrocyte differentiation, Gadd45a is more likely to be involved in maintenance or maturation of astrocytes. On the other hand, Tet2 is highly likely to be involved in the active demethylation of astrocytic genes and the initialization of astrocyte differentiation.

Since the discovery of demethylation catalyzed by Tet family proteins, there have been extensive studies on their roles in expression regulation, embryonic development and stem cell differentiation (Williams et al., 2012, Shen and Zhang, 2012, Shen et al., 2013, Kohli and Zhang, 2013, He et al., 2011, Hamby et al., 2008). qPCR and western blot showed that Tet2 is expressed at an extremely low level in E11 cortical NPCs, and its expression level rapidly, yet transiently increased upon differentiation *in vitro*. Overexpression of Tet2 promoted the differentiation towards astrocyte lineage, delayed the differentiation towards neuronal lineage and had little affect on the differentiation towards oligodendrocyte lineage. Depletion of Tet2 using shRNA knockdown partially blocked the differentiation towards astrocyte lineage.

DNA methylation at the promoter sites is commonly considered as a repressive marker, and it has been proposed to regulate the switch from neurogenesis to astrogliogenesis

during the development of the brain (Condorelli et al., 1994, Takizawa et al., 2001, Fan et al., 2005). Tet proteins can oxidize 5mC (5-methylcytosine) into 5hmC (5-hydroxymethylcytosine), which can then be removed passively by cell division or actively through Tdg/base excision pathway (Iyer et al., 2009, Tahiliani et al., 2009, Hu et al., 2013). 5hmC is abundant in in brain, and a high ratio of 5hmC over 5mC is associated with active transcription (Ito et al., 2011). Tet2 mediated converting from 5meC to 5hmC may be associated with gene activation (Wrighton, 2013).

Genome-wide expression analysis using RNA-seq revealed that Tet2 overexpression did not affect the process of NPC proliferation or differentiation. All of the genes that showed differential expressed in Tet2 overexpression cells were up-regulated, which is not surprising considering Tet2's function in DNA demethylation. Over 67% of the genes that were up-regulated in Tet2 overexpressing NPCs belong to astrocyte specific, while less then 14% were oligodendrocyte or neuron specific, indicating Tet2 preferentially promote the expression of astrocytic genes.

The MeDIP and hMeDIP experiments showed both an increased in hydroxymethylation level and a decrease in the DNA methylation level at the proximal promoter (-1.5kb to -1kb from TSS) of Tet2 targets during the differentiation of NPCs. On the other hand, there was no obvious change at global DNA hydroxymethylation and DNA methylation, suggesting Tet2 is actively and specifically targeting the astrocytic genes.

I observed a decreased DNA methylation at Gfap promoter and TSS sites with Tet2 overexpressin, which correlated with the expression up-regulation. The process of demethylation and also expression change was abolished when I knockdown Tet2. These data established a direct link between Tet2 and expression regulation.

In summary, I have demonstrated that the activation of Tet2 dependent demethylation at the differentiation onset is necessary to initiate and promote astrocyte lineage progression. Genome-wide transcriptome and methylation/hydroxymethylation enrichment profiling revealed that Tet2 mediated active DNA demethylation and expression activation is specific to the promoter of astrocytic genes. Genome-wide ChIP-chip revealed that besides a subset of astrocytic regulatory genes, including Id2, Id4, Nfia and Notch, Olig2 also binds and represses Tet2. Together, the results demonstrate that the proper DNA demethylation by Tet2 is critical to initiate and establish the transcriptional program that promotes astrocyte differentiation. Olig2 not only represses the expression of astrocytic regulatory factors to block premature astrocytic differentiation, it also directly represses the methylcytosine dioxygenase Tet2 to maintain a repressive chromatin states of astrocytic genes (Fig 2.17).

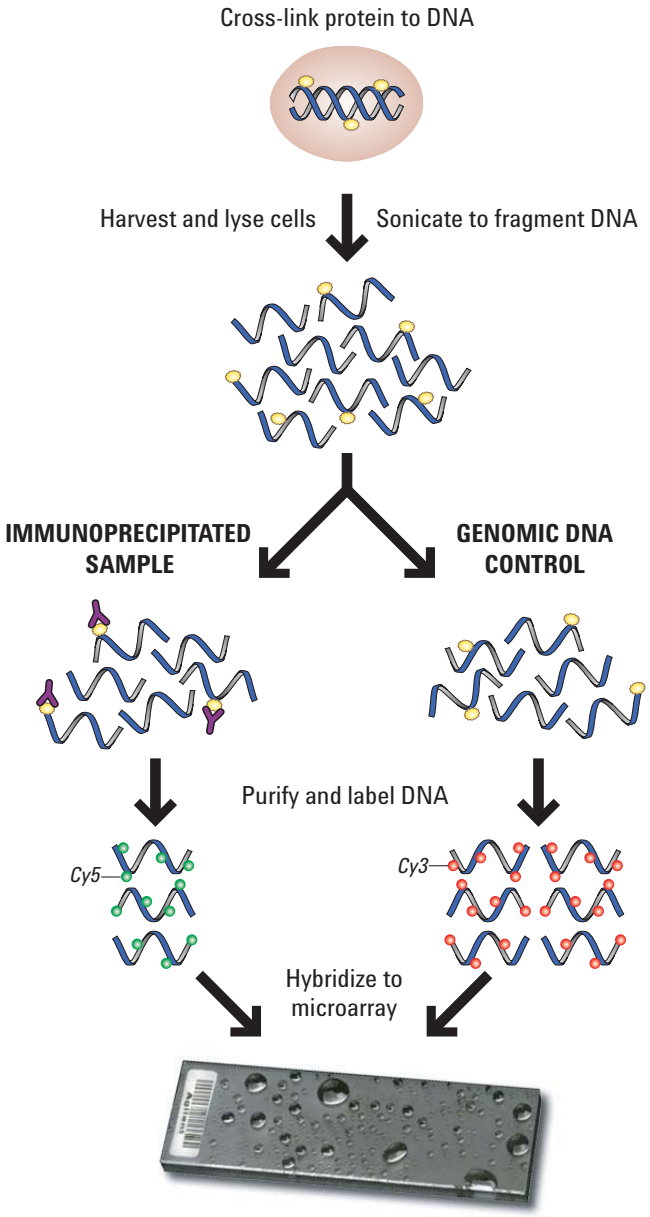


Figure 2.1 ChIP-on-chip overview. (Adopted from Agilent Mammalian ChIP-on-chip protocol)

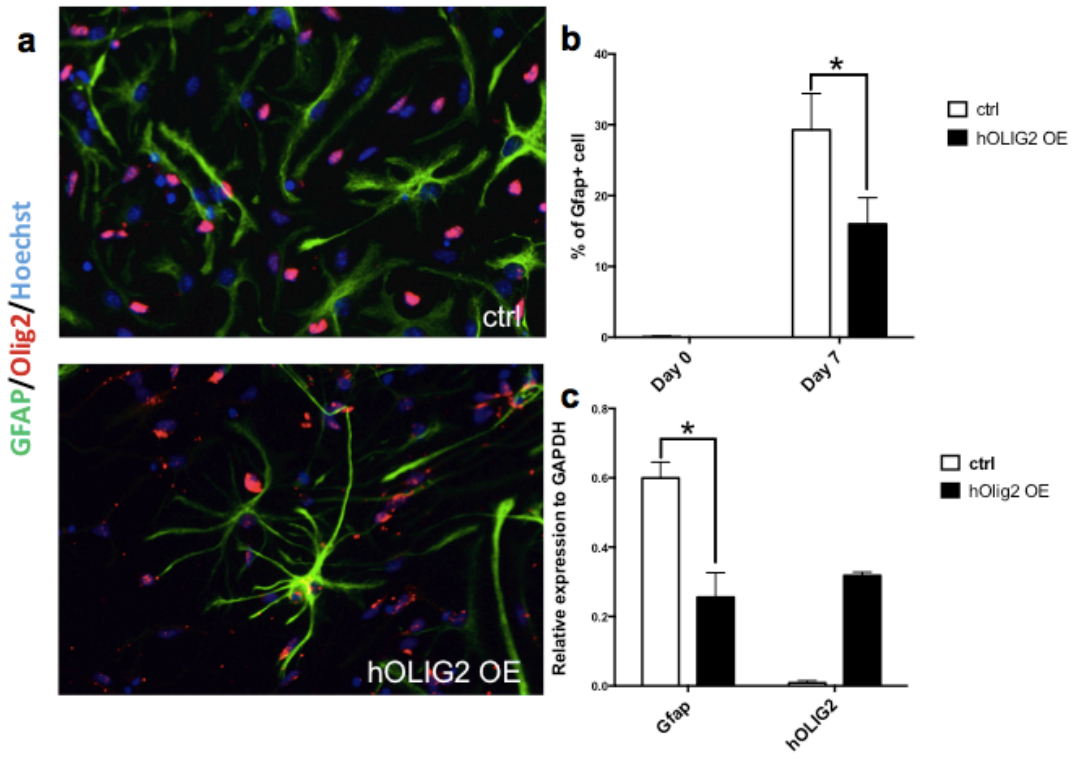


Figure 2.2 Olig2 overexpression represses astrocytic differentiation *in vitro*.

(a) NPCs infected with FG12 (ctrl) or FUIGW-hOlig2 (hOlig2) were spontaneously differentiated by withdraw bFGF for 7 days. Cells were labeled with Olig2 (red), Gfap (green) and Hoechst (blue). (b) Percentage of Gfap+ cells at day 0 (undifferentiated NPCs) and day7 in both control and Olig2 overexpressing system are shown in bar graph (p value < 0.05 by t test). (c) The mRNA level of Gfap and exogenous Olig2 are shown as the relative expression to GAPDH (p value < 0.05 by t test).

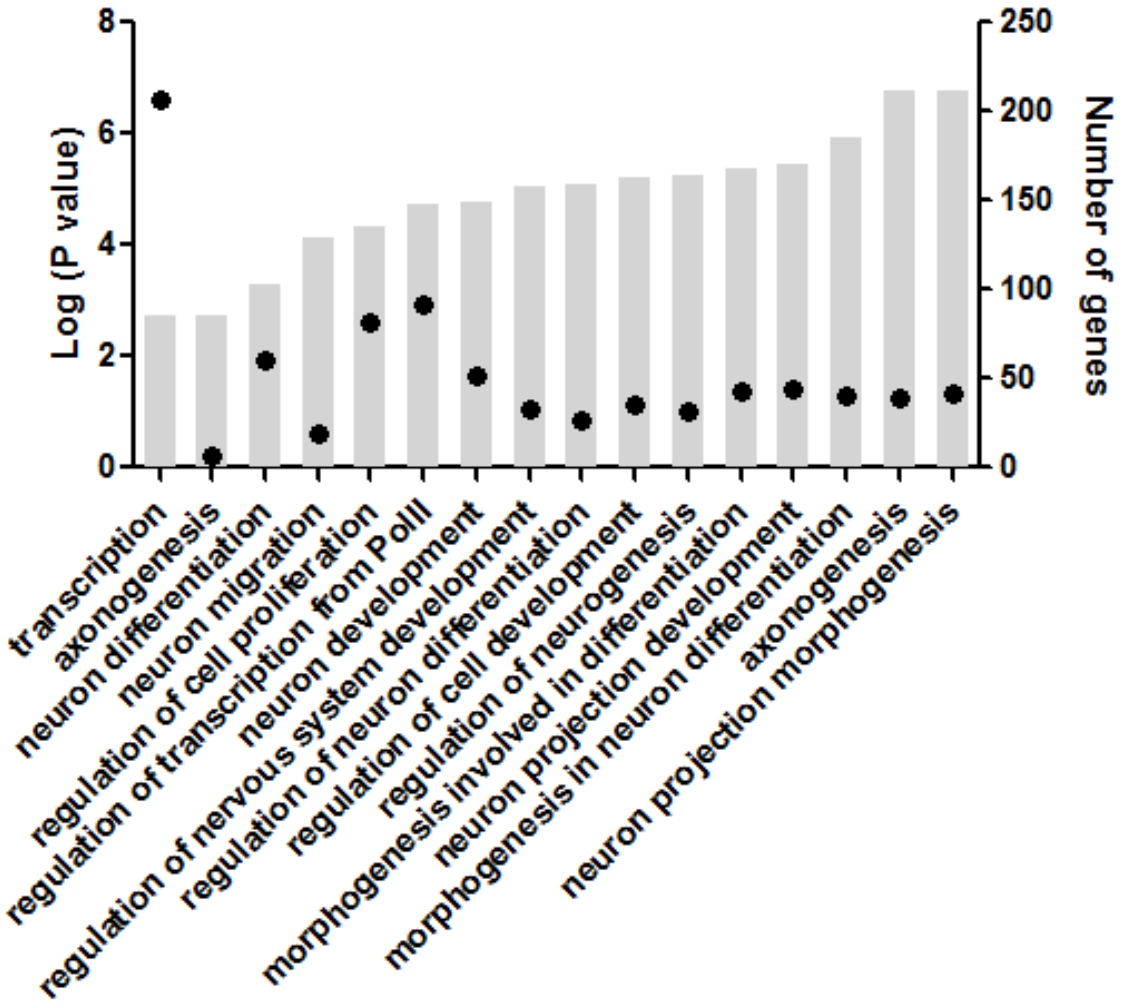


Figure 2.3 Olig2 binding sites in E11 cortical NPCs and the GO enrichment for its targets.

Left y axis (grey bar): enrichment of corresponding GO term (- log ratio of the p value)

Right y axis (black dot): number of the genes in corresponding GO category

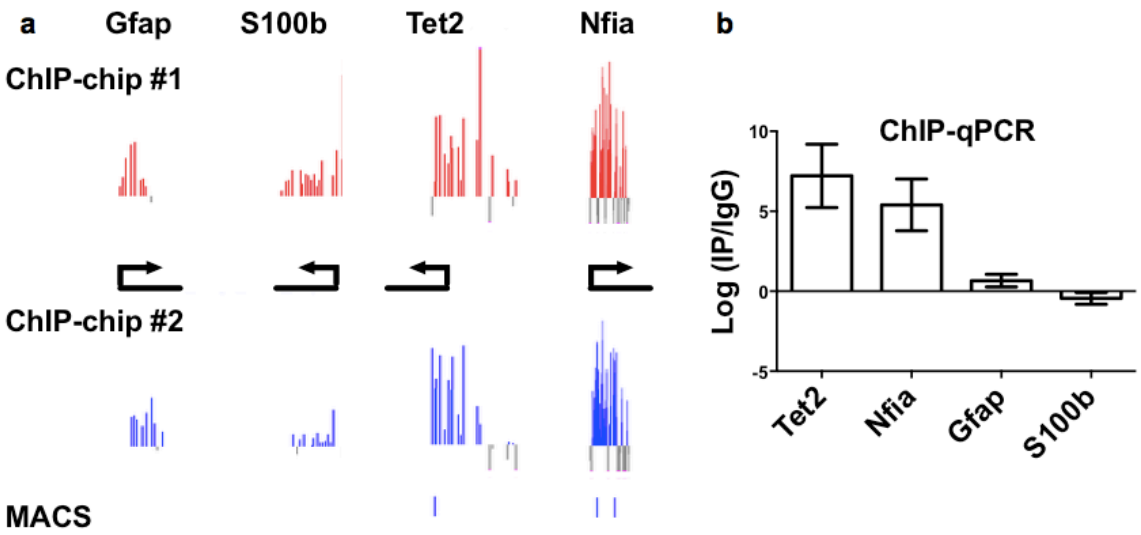


Figure 2.4 ChIP-chip (a) and ChIP-qPCR validation (b) of Olig2 targets (Tet2, Nfia) and non-Olig2 targets (Gfap, S100b) in E11 cortical NPC. Chip-chip replicates are shown in red and blue bars, the height of the bar plot indicates the enrichment of Olig2 at the corresponding locus, black arrow indicates the direction of transcription, blue bars at the bottom of the graphs indicates the true binding peaks identified by MACS.

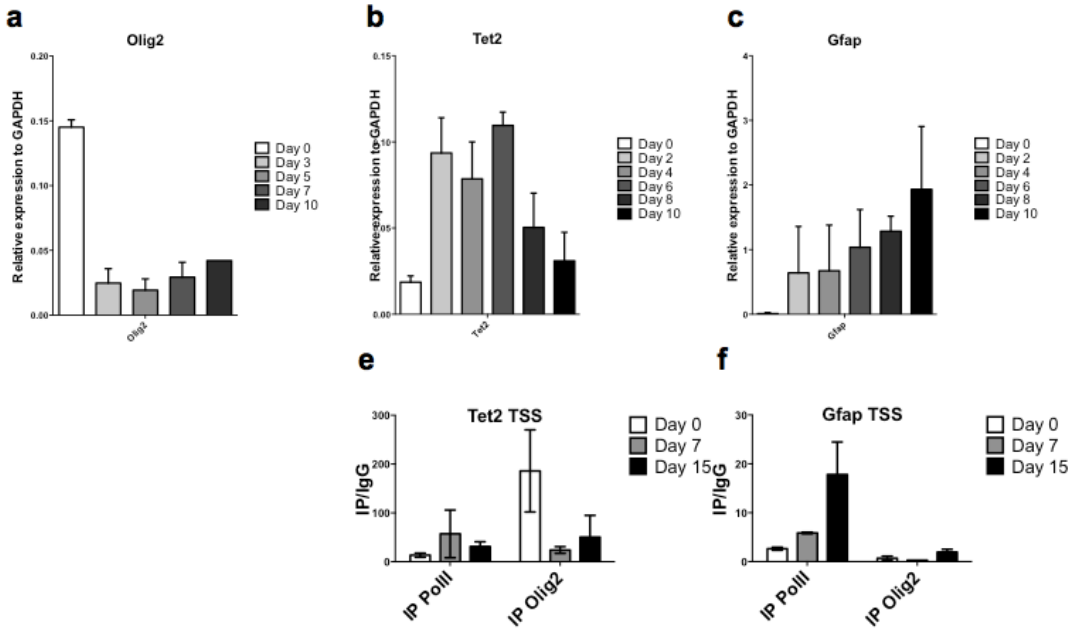


Figure 2.5 Expression level (a-c) and PolII enrichment (e-f) of Olig2 targets in E11 cortical NPC and NPCs differentiated for up to 10 days.

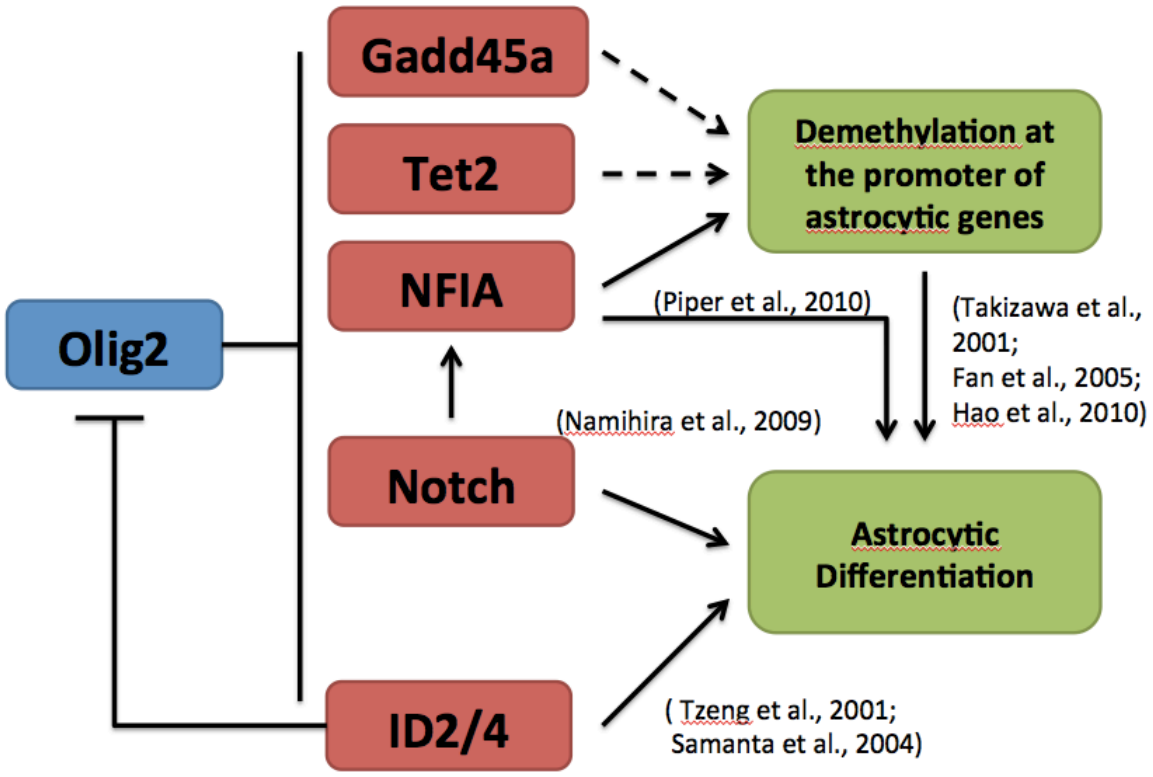


Figure 2.6 Olig2 represses astrocyte differentiation through its down stream targets. Olig2 does not directly repress astrocytic genes. ChIP-chip result indicates that Olig2 is enriched at the promoter of multiple genes that are known for regulating astrogliogenesis, such as Nfia, Notch1, ID2 and ID4 (as shown with solid arrows). Given that DNA methylation is crucial for the switch from neurogenesis to astrogliogenesis in NPCs, we hypothesize that Tet2 and Gadd45a, both being the direct targets of Olig2 and implicated in DNA demethylation, also act as the intermediates in Olig2's repression of astrocyte differentiation.

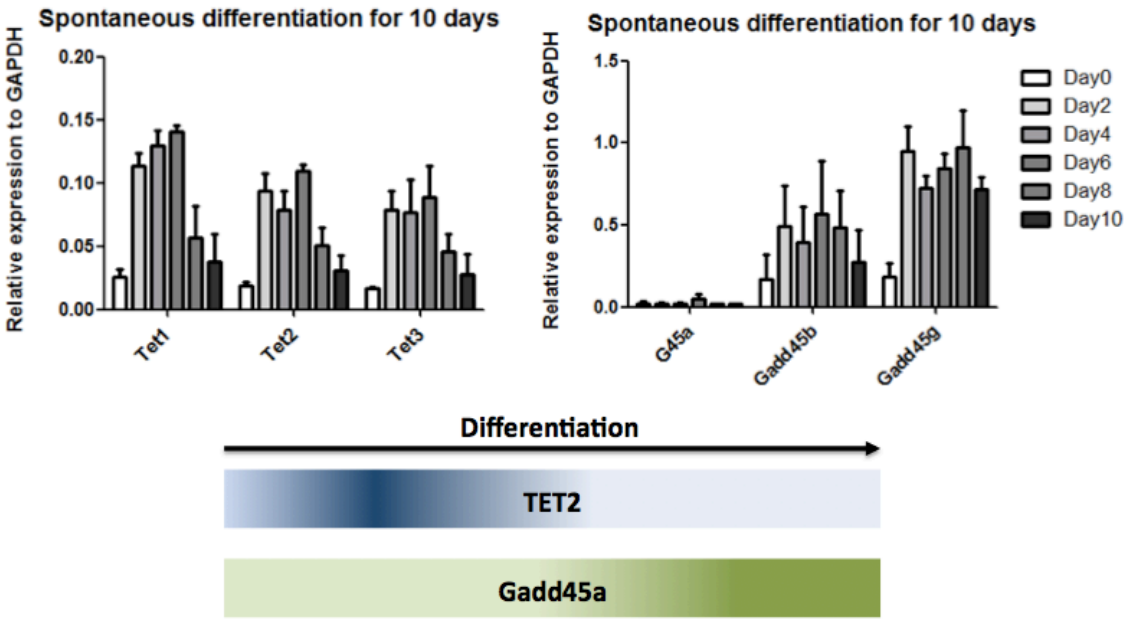


Figure 2.7 mRNA level of Tet and Gadd45 family members during the spontaneous differentiation of E11 NPCs.

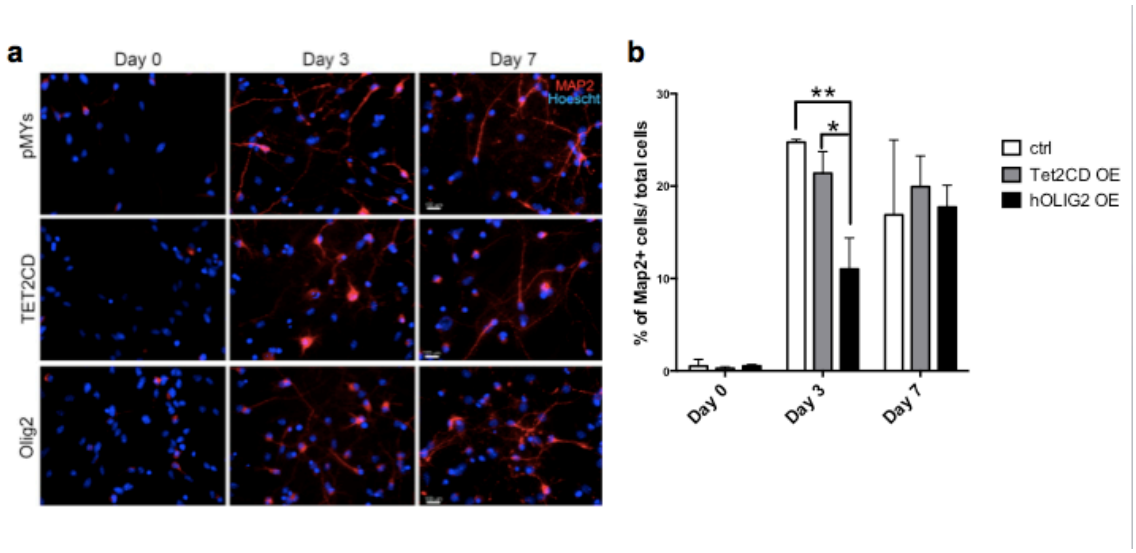


Figure 2.8 Immunohistochemistry and quantification of Map2+ neurons in differentiating control and Tet2 overexpressing NPCs. (a) E11 cortical NPCs were infected with control, Olig2 or Tet2 overexpression virus and cultured with bFGF or withdraw bFGF for up to 7 days. Cells were labeled with neuronal marker Map2 (red) and Hoechst (blue). (b) Percentage of Map2+ cells at day 0 (undifferentiated NPCs) day3 and day7 in control, Tet2 and Olig2 overexpressing cells are shown in bar graph (p value < 0.05 by t test).

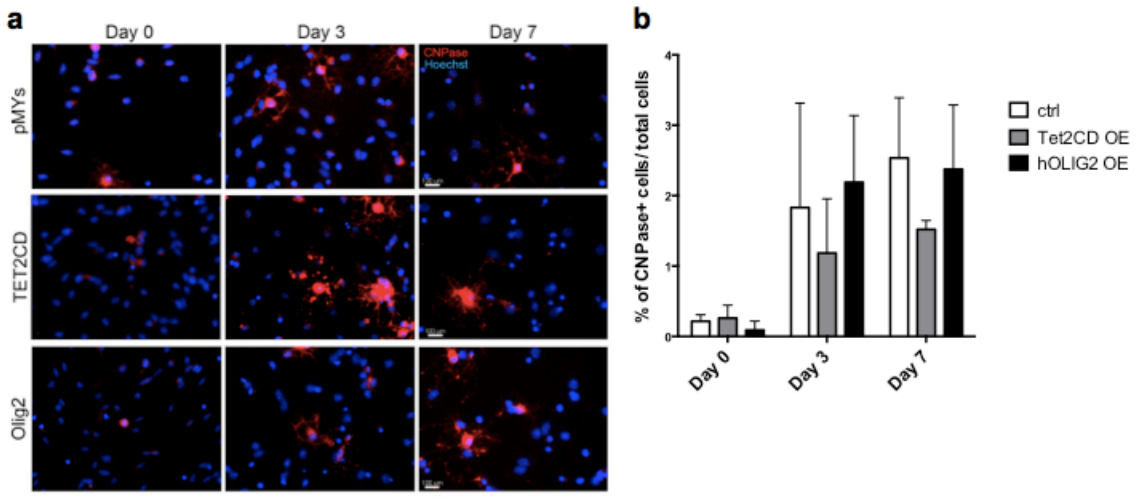


Figure 2.9 Immunohistochemistry and quantification of CNPase+ oligodendrocytes in differentiating control and Tet2 overexpressing NPCs. (a) E11 cortical NPCs were infected with control, Olig2 or Tet2 overexpression virus and cultured with bFGF or withdraw bFGF for up to 7 days. Cells were labeled with oligodendrocyte marker CNPase (red) and Hoechst (blue). (b) Percentage of CNPase+ cells at day 0 (undifferentiated NPCs) day3 and day7 in control, Tet2 and Olig2 overexpressing cells are shown in bar graph (no significant differences between groups).

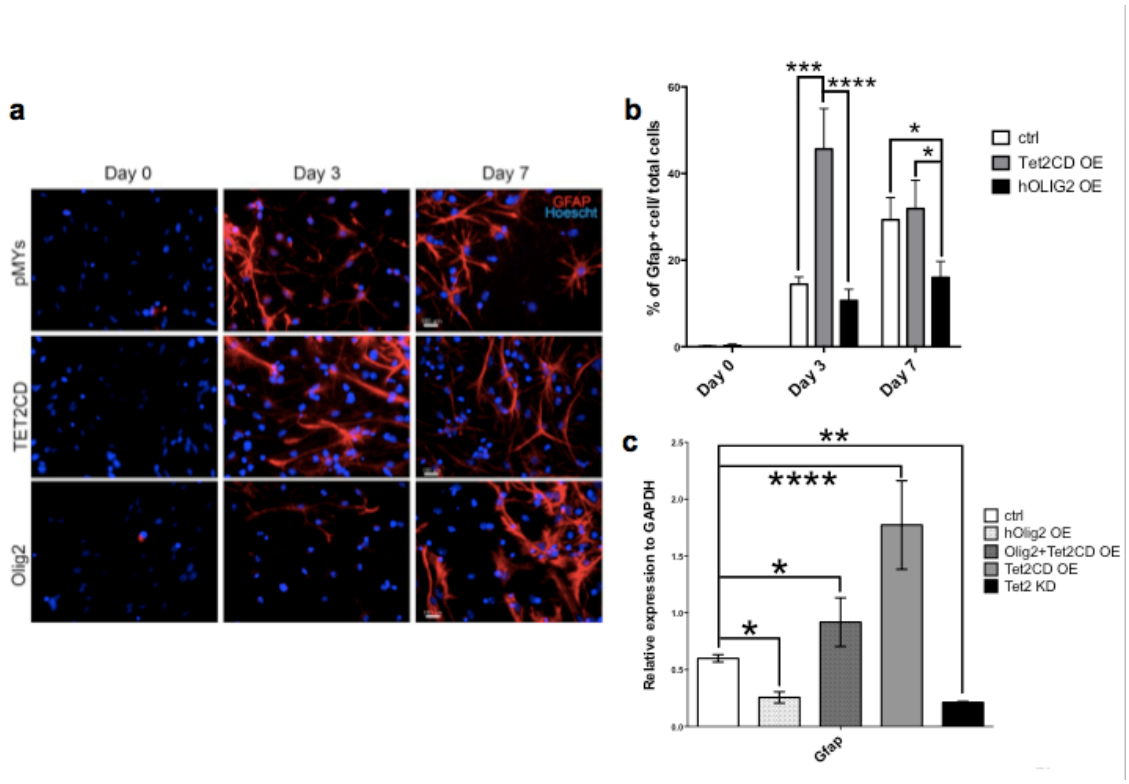


Figure 2.10 Immunohistochemistry and quantification of Gfap+ astrocytes in differentiating control and Tet2 overexpressing NPCs. (a) E11 cortical NPCs were infected with control, Olig2 or Tet2 overexpression virus and cultured with bFGF or withdraw bFGF for up to 7 days. Cells were labeled with astrocyte marker Gfap (red) and Hoechst (blue). (b) Percentage of Gfap+ cells at day 0 (undifferentiated NPCs) day3 and day7 in control, Tet2 and Olig2 overexpressing cells are shown in bar graph (p value < 0.05 by t test). (c) mRNA level of Gfap at day7 of differentiation.

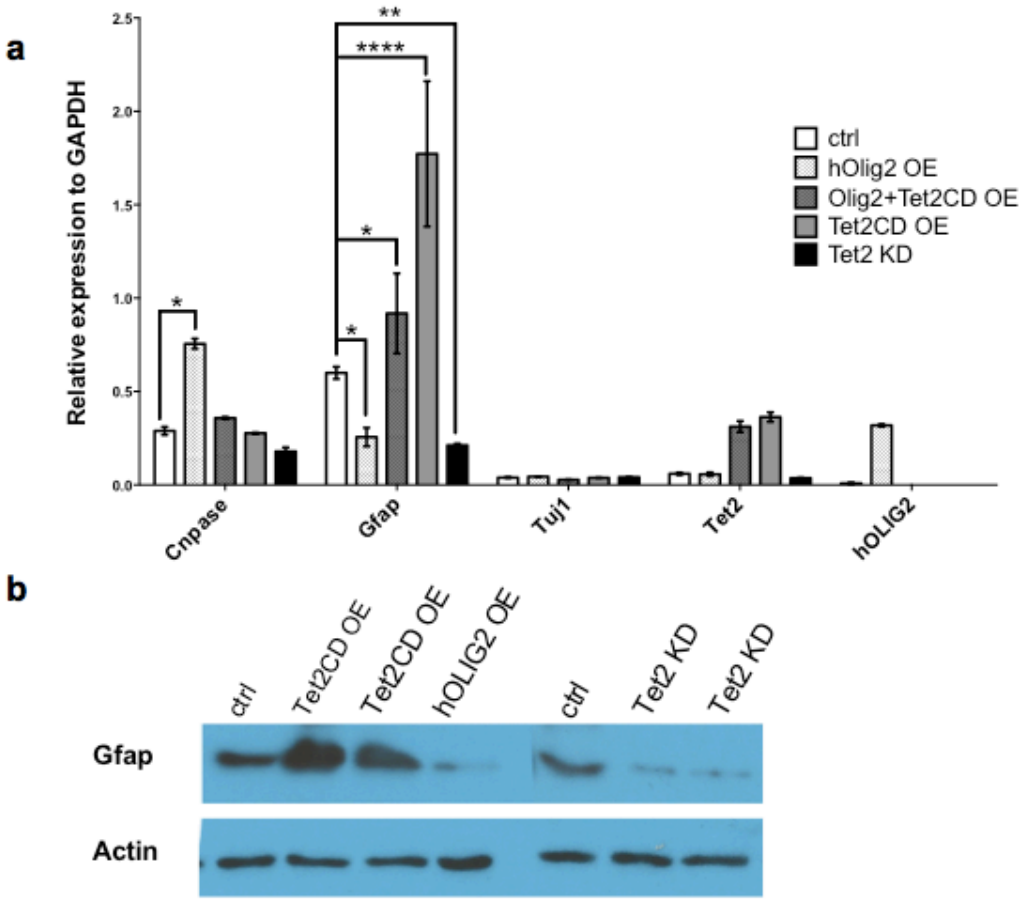


Figure 2.11 Exogenous Tet2 induces Gfap expression. NPCs infected with control (ctrl), Tet2 overexpression (Tet2CD), Olig2 overexpression (hOlig2) and shTet2 (Tet2 KD) virus were differentiated for 7 days. (a) mRNA level of lineage markers CNPase, Gfap and Tuj1 were measured using qPCR. (b) The protein level of Gfap was measured using western blot with Actin as internal control.

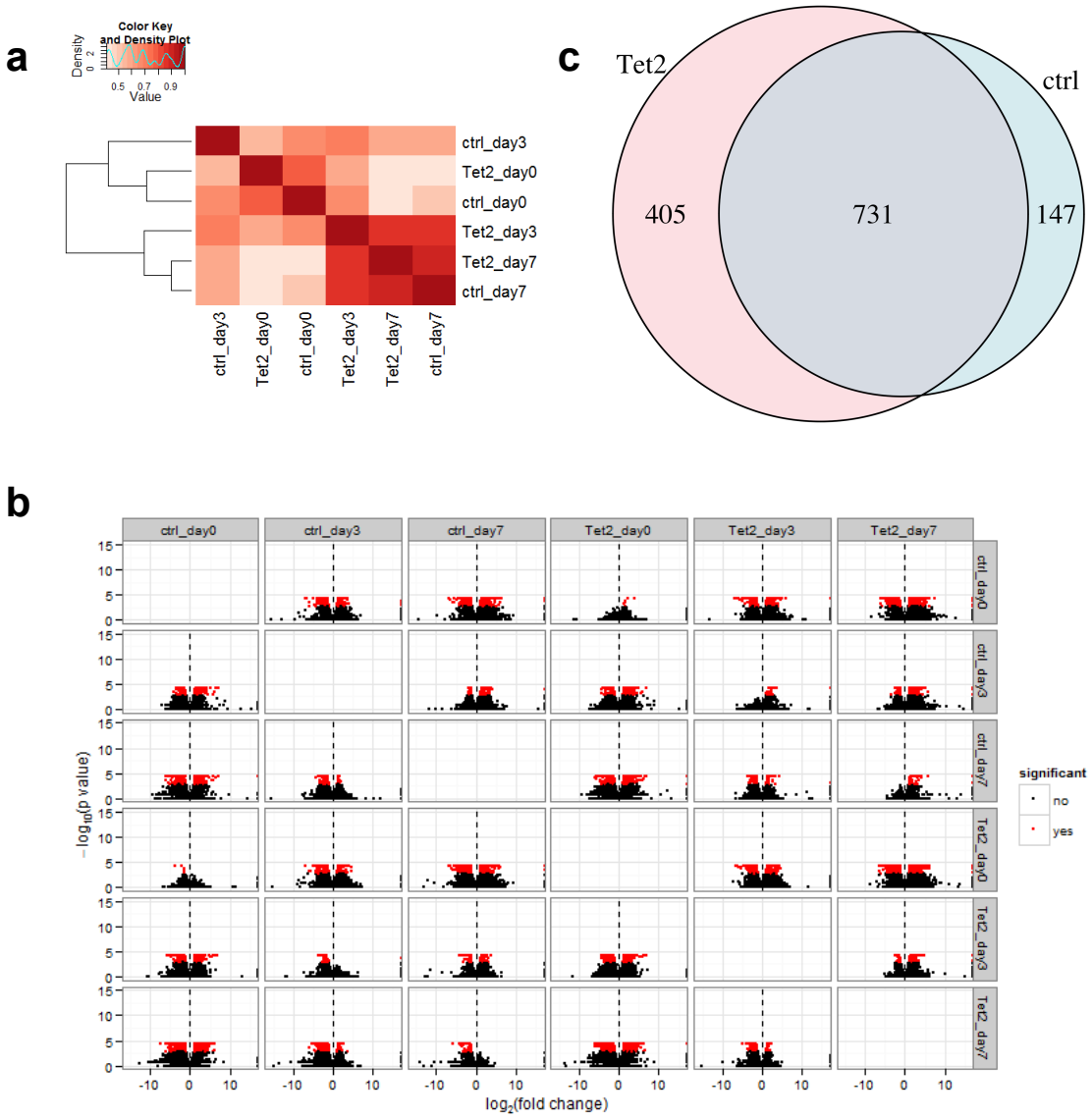
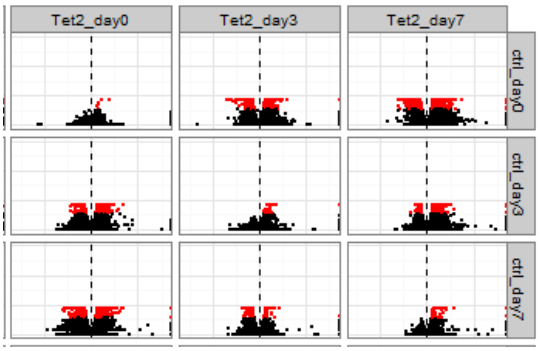


Figure 2.12 Global differential expression induced by Tet2 overexpressed. (a) A pairwise comparison with Pearson correlation coefficients. Red indicates a high similarity (cor = 1) and white indicates a lower similarity (cor = 0) between the two samples under comparison. (b) CummeRbund volcano plot reveals genes that differ significantly between each pair. Red indicates significant differential expression. Dots to the right of the vertical dash line indicate an increase in expression, while dots to the left of the vertical dash line indicates a decrease in expression. (c) Vann diagram

showing the number of overlapped genes between significantly differential expressed genes in control (blue) and Tet2 (pink) overexpressing NPCs.

a



b

Day 3 (Top 25)		Day 7 (Top 25)	
Gene ID	log2 FC	Gene ID	log2 FC
Agt	3.79901	Gpr37l1	5.7388
Gpr37l1	3.64966	Tet2	4.07033
Cryab	3.12793	Chac1	3.99901
Tet2	3.0721	Ucma	3.71914
Rmrp	2.64429	Ifitm3	3.50677
Ifi27l1	2.5737	Cyp2d22	3.48614
Lcat	2.41896	Gm11127	3.28306
Pygm	2.36465	Gjb6	3.2761
Neat1	2.36372	Lgals3bp	3.20692
Gfap	2.30717	Pdk4	3.15844
C4b	2.28082	S100b	3.0615
Hspb8	2.26514	Apoc1	2.9829
Grin2c	2.25981	Agt	2.97708
Kcne1l	2.23716	Eif4ebp1	2.79891
Rpph1	2.09518	Neat1	2.71426
Gm19757	2.06061	Aldh1a1	2.69098
S100a1	1.99596	Ddr2	2.68668
S100b	1.68227	Chchd10	2.55175
Tmem38b	1.66793	1700019G17Rik	2.52865
Mdk	1.57217	Gas6	2.51019
Clu	1.56723	Lcat	2.4992
Aox1	1.55649	Prr5	2.47986
Fxyd1	1.55013	Rmrp	2.44829
Mthfd2	1.52895	Atf5	2.43996
Foxj1	1.50088	Grin2c	2.43498

(Astrocyte/Oligodendrocyte/Neuron)

c

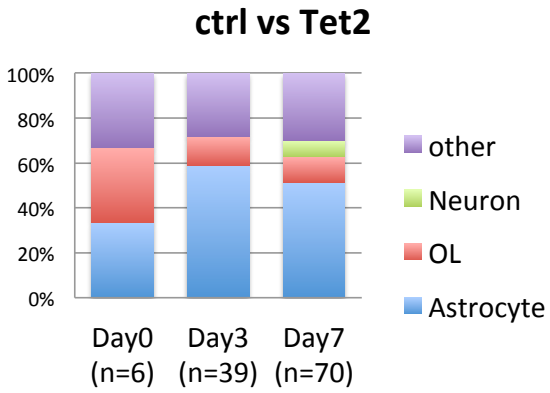


Figure 2.13 Tet2 preferentially affects the expression of astrocytic genes. (a) CummeRbund volcano plot reveals genes that differ significantly between each pairs of conditions with red indicate significant differential expression. Dots to the right of the vertical dash line indicate an increase in expression, while dots to the left of the vertical dash line indicates a decrease in expression. (b) Top 25 up-regulated genes in Tet2 overexpression NPCs after day 3 and day 7 into differentiation. Green indicates astrocytic genes; blue indicates oligodendrocyte genes and red indicates neuronal genes. (c) Lineage composition of the genes up-regulated in Tet2 overexpression NPCs at day 0,

differentiation day 3 and day 7.

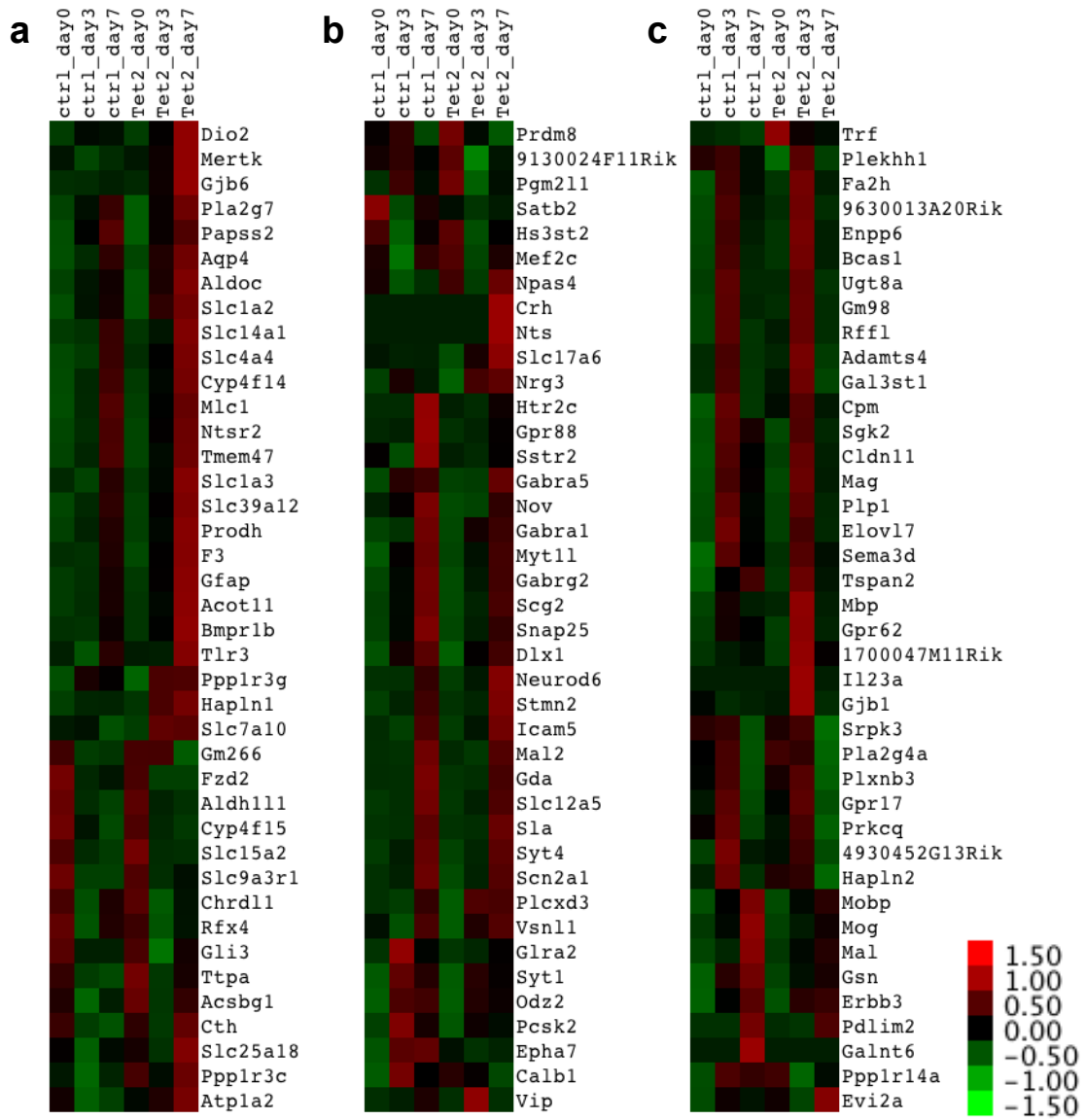


Figure 2.14 Expression changes of the lineage genes (from left to right: astrocyte, neuron, oligodendrocyte).

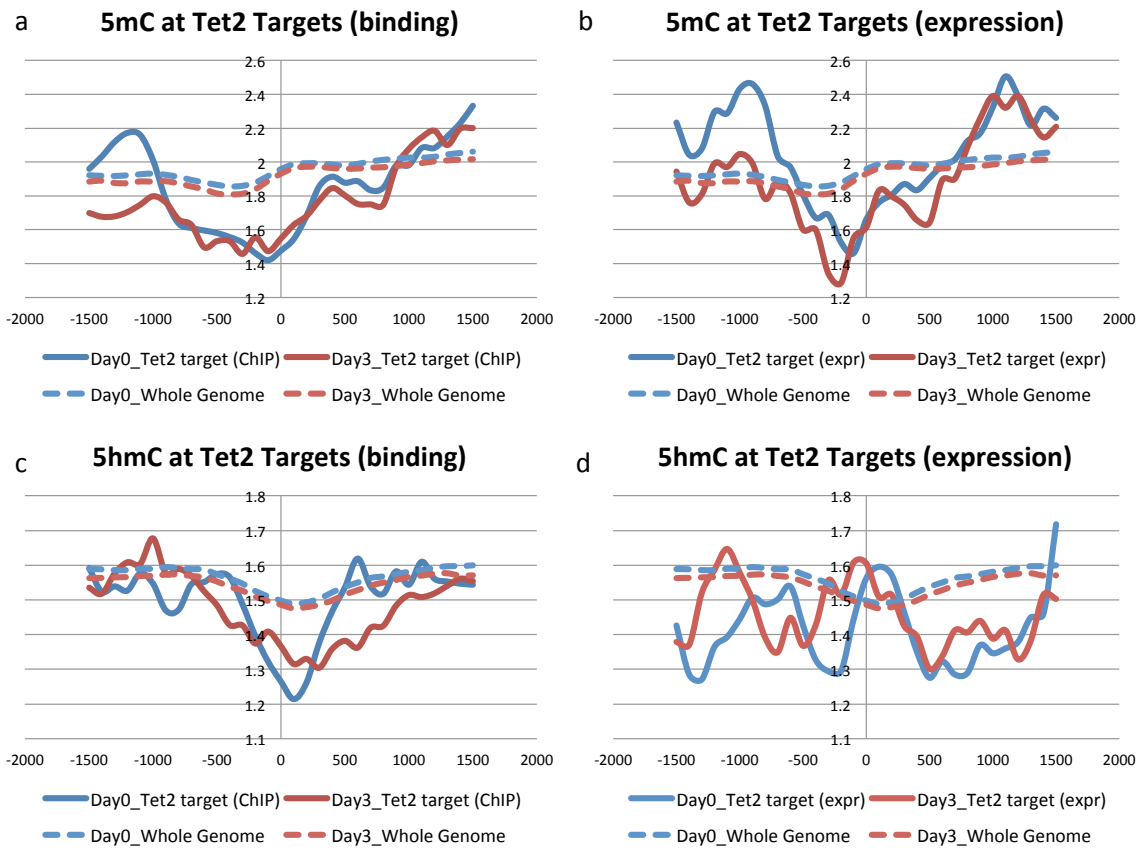


Figure 2.15 DNA methylation and hydroxymethylation surrounding transcription start sites (TSS) in undifferentiated NPCs (blue) and differentiated NPCs at day 3 (red). Average enrichment of DNA methylation (a,b) and hydroxymethylation (c,d) surrounding the TSS was calculated based on the enrichment of MeDIP over input using the SitePro from CEAS. Solid lines represent the TSS of genes that are Tet2 binding targets(Chen et al., 2013) (a, c) or genes whose expression was induced by Tet2-overepsression during NPC differentiation (b, d).

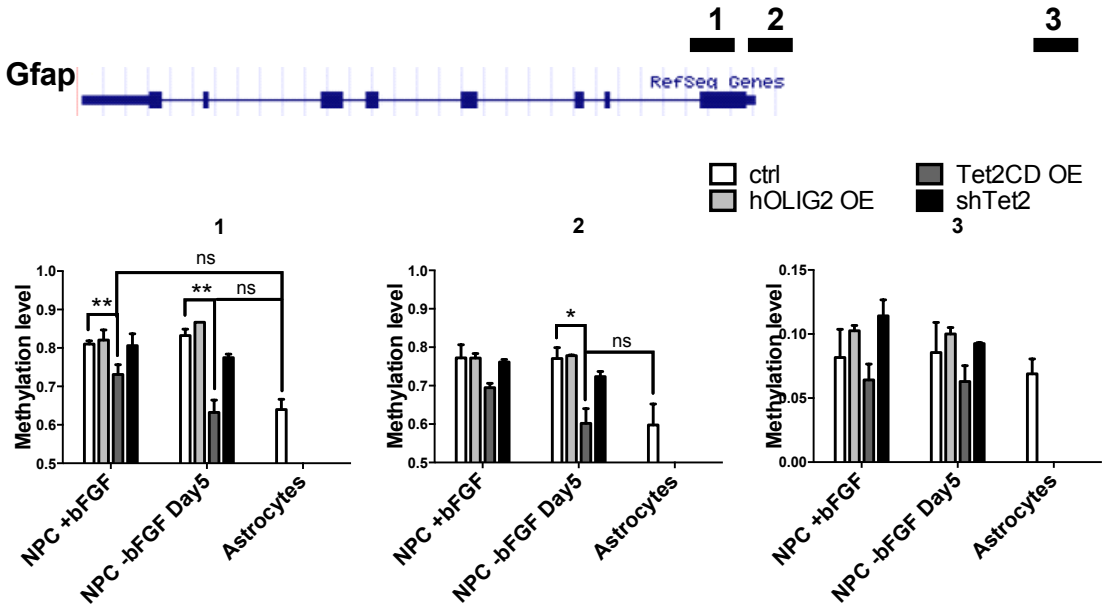


Figure 2.16 Exogenous Tet2 induces demethylation at Gfap promoter. Methylation level at the TSS (1,2) and proximal promoter (3) of Gfap in undifferentiated and differentiated NPCs with control (ctrl), Tet2 overexpression (Tet2CD), Olig2 overexpression (hOlig2) and shTet2 (Tet2 KD) and primary astrocytes.

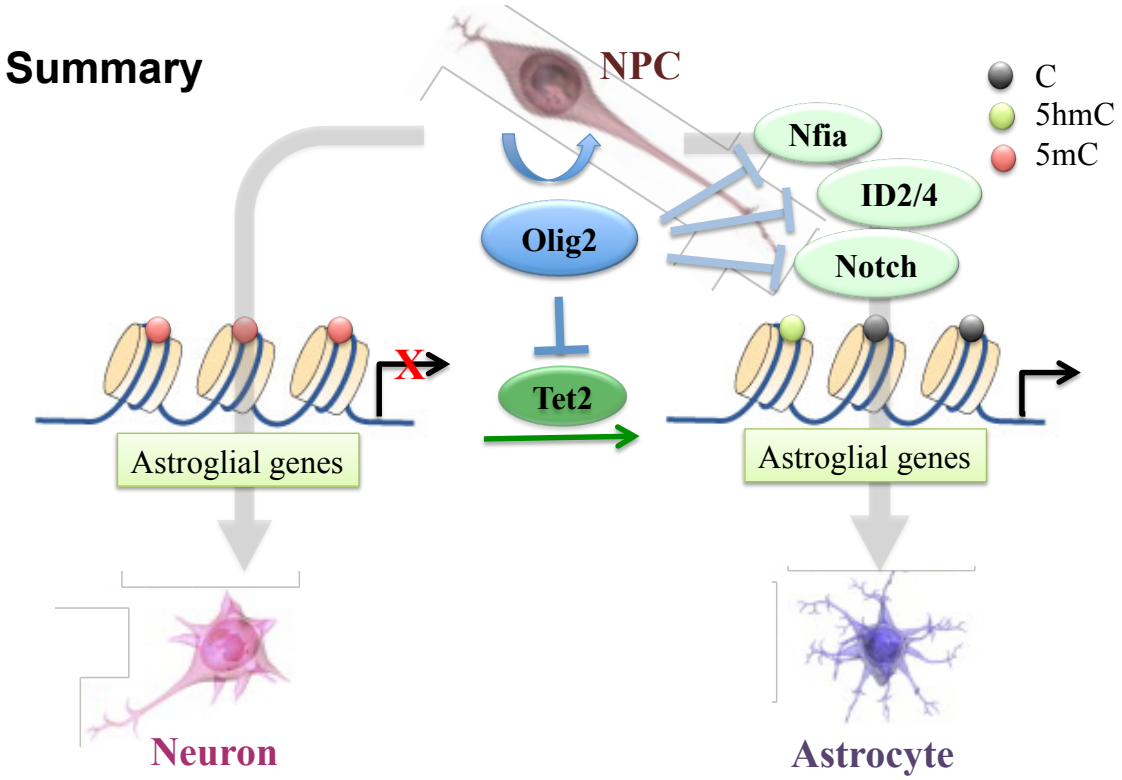


Figure 2.17 Summary of Tet2 and Olig2's roles in astrocytic differentiation. DNA demethylation by Tet2 is critical to initiate and establish the transcriptional program that promotes astrocyte differentiation. Olig2 not only represses the expression of astrocytic regulatory factors to block premature astroglial differentiation, it also directly represses the methylcytosine dioxygenase Tet2 to maintain a repressive chromatin states of astroglial genes

Table 4 Lysis Buffer 1 (LB1)

Stock	For 100 mL	Final Concentration
1M Hepes-KOH, pH 7.5	5.0 mL	50 mM
5M NaCl	2.8 mL	140 mM
0.5M EDTA	0.2 mL	1 mM
50% glycerol	20.0 mL	10%
10% Igepal	5.0 mL	0.5%
10% Triton X-100	2.5 mL	0.25%
ddH_2O	64.5 mL	

Table 5 Lysis Buffer 2 (LB2)

Stock	For 100 mL	Final Concentration
1M Tris-HCl, pH 8.0	1.0 mL	10 mM
5M NaCl	4.0 mL	200 mM
0.5M EDTA, pH 8.0	0.2 mL	1 mM
0.5M EGTA, pH 8.0	0.1 mL	0.5 mM
ddH_2O	94.7 mL	

Table 6 Lysis Buffer 3 (LB3)

Stock	For 100 mL	Final Concentration
1M Tris-HCl, pH 8.0	1.0 mL	10 mM
5M NaCl	2.0 mL	100 mM
0.5M EDTA, pH 8.0	0.2 mL	1 mM
0.5M EGTA, pH 8.0	0.1 mL	0.5 mM
10% Na-Deoxycholate	1.0 mL	0.1%
20% N-lauroylsarcosine	2.5 mL	0.5%
ddH_2O	93.2 mL	

Table 2.1 Component of Lysis buffer 1,2 and 3 (Agilent Mammalian ChIP-on-chip protocol)

References

ASMANN, Y. W., HOSSAIN, A., NECELA, B. M., MIDDHA, S., KALARI, K. R., SUN, Z., CHAI, H. S., WILLIAMSON, D. W., RADISKY, D., SCHROTH, G. P., KOCHER, J. P., PEREZ, E. A. & THOMPSON, E. A. 2011. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res,* 39**,** e100.

BLECHER-GONEN, R., BARNETT-ITZHAKI, Z., JAITIN, D., AMANN-ZALCENSTEIN, D., LARA-ASTIASO, D. & AMIT, I. 2013. High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. *Nat Protoc,* 8**,** 539-54.

CHEN, Q., CHEN, Y., BIAN, C., FUJIKI, R. & YU, X. 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature,* 493**,** 561-4.

CONDORELLI, D. F., NICOLETTI, V. G., BARRESI, V., CARUSO, A., CONTICELLO, S., DE VELLIS, J. & GIUFFRIDA STELLA, A. M. 1994. Tissue-specific DNA methylation patterns of the rat glial fibrillary acidic protein gene. *J Neurosci Res,* 39**,** 694-707.

DENNIS, G., JR., SHERMAN, B. T., HOSACK, D. A., YANG, J., GAO, W., LANE, H. C. & LEMPICKI, R. A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol,* 4**,** P3.

FAN, G., MARTINOWICH, K., CHIN, M. H., HE, F., FOUSE, S. D., HUTNICK, L., HATTORI, D., GE, W., SHEN, Y., WU, H., TEN HOEVE, J., SHUAI, K. & SUN, Y. E. 2005. DNA methylation controls the timing of astrogliogenesis through regulation of

JAK-STAT signaling. *Development,* 132**,** 3345-56.

FENG, J., FOUSE, S. & FAN, G. 2007. Epigenetic regulation of neural gene expression and neuronal function. *Pediatr Res,* 61**,** 58R-63R.

FENG, J., LIU, T. & ZHANG, Y. 2011. Using MACS to identify peaks from ChIP-Seq data. *Curr Protoc Bioinformatics,* Chapter 2**,** Unit 2 14.

HAMBY, M. E., COSKUN, V. & SUN, Y. E. 2008. Transcriptional regulation of neuronal differentiation: the epigenetic layer of complexity. *Biochim Biophys Acta,* 1779**,** 432-7.

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**,** 1303-7.

HU, L., LI, Z., CHENG, J., RAO, Q., GONG, W., LIU, M., SHI, Y. G., ZHU, J., WANG, P. & XU, Y. 2013. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell,* 155**,** 1545-55.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

IYER, L. M., TAHILIANI, M., RAO, A. & ARAVIND, L. 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle,* 8**,** 1698-710.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC.

Genome Res, 12**,** 996-1006.

KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472-9.

LI, H., DE FARIA, J. P., ANDREW, P., NITARSKA, J. & RICHARDSON, W. D. 2011. Phosphorylation regulates OLIG2 cofactor choice and the motor neuron-oligodendrocyte fate switch. *Neuron,* 69**,** 918-29.

LIENHARD, M., GRIMM, C., MORKEL, M., HERWIG, R. & CHAVEZ, L. 2014. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics,* 30**,** 284-6.

LIU, T., ORTIZ, J. A., TAING, L., MEYER, C. A., LEE, B., ZHANG, Y., SHIN, H., WONG, S. S., MA, J., LEI, Y., PAPE, U. J., POIDINGER, M., CHEN, Y., YEUNG, K., BROWN, M., TURPAZ, Y. & LIU, X. S. 2011. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol,* 12**,** R83.

ROBERTS, A., PIMENTEL, H., TRAPNELL, C. & PACHTER, L. 2011a. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics,* 27**,** 2325-9.

ROBERTS, A., TRAPNELL, C., DONAGHEY, J., RINN, J. L. & PACHTER, L. 2011b. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol,* 12**,** R22.

SETOGUCHI, T. & KONDO, T. 2004. Nuclear export of OLIG2 in neural stem cells is essential for ciliary neurotrophic factor-induced astrocyte differentiation. *J Cell Biol,* 166**,** 963-8.

SHEN, L., WU, H., DIEP, D., YAMAGUCHI, S., D'ALESSIO, A. C., FUNG, H. L., ZHANG, K. & ZHANG, Y. 2013. Genome-wide analysis reveals TET- and

TDG-dependent 5-methylcytosine oxidation dynamics. *Cell,* 153**,** 692-706.

SHEN, L. & ZHANG, Y. 2012. Enzymatic analysis of Tet proteins: key enzymes in the metabolism of DNA methylation. *Methods Enzymol,* 512**,** 93-105.

SHIN, H., LIU, T., MANRAI, A. K. & LIU, X. S. 2009. CEAS: cis-regulatory element annotation system. *Bioinformatics,* 25**,** 2605-6.

SULTAN, F. A. & SWEATT, J. D. 2013. The role of the gadd45 family in the nervous system: a focus on neurodevelopment, neuronal injury, and cognitive neuroepigenetics. *Adv Exp Med Biol,* 793**,** 81-119.

TABU, K., OHNISHI, A., SUNDEN, Y., SUZUKI, T., TSUDA, M., TANAKA, S., SAKAI, T., NAGASHIMA, K. & SAWA, H. 2006. A novel function of OLIG2 to suppress human glial tumor cell growth via p27Kip1 transactivation. *J Cell Sci,* 119**,** 1433-41.

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**,** 930-5.

TAKIZAWA, T., NAKASHIMA, K., NAMIHIRA, M., OCHIAI, W., UEMURA, A., YANAGISAWA, M., FUJITA, N., NAKAO, M. & TAGA, T. 2001. DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. *Dev Cell,* 1**,** 749-58.

WILLIAMS, K., CHRISTENSEN, J. & HELIN, K. 2012. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep,* 13**,** 28-35.

WRIGHTON, K. H. 2013. Chromatin: TET2 keeps histones sweet. *Nat Rev Mol Cell Biol,*

14**,** 64-5.

WU, H., COSKUN, V., TAO, J., XIE, W., GE, W., YOSHIKAWA, K., LI, E., ZHANG, Y. & SUN, Y. E. 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science,* 329**,** 444-8.

YU, Y., CHEN, Y., KIM, B., WANG, H., ZHAO, C., HE, X., LIU, L., LIU, W., WU, L. M., MAO, M., CHAN, J. R., WU, J. & LU, Q. R. 2013. Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell,* 152**,** 248-61.

ZHOU, Q., CHOI, G. & ANDERSON, D. J. 2001. The bHLH transcription factor Olig2 promotes oligodendrocyte differentiation in collaboration with Nkx2.2. *Neuron,* 31**,** 791-807.

Chapter 3

Digitizing molecular events after spinal cord injury

Abstract

Spinal cord injury (SCI) induces a progressive pathophysiology affecting cell survival and neurological integrity via complex and evolving molecular cascades whose interrelationships are not fully understood. Injury in spinal cord tissue leads to a rapid destruction of cells at the site of injury, an intense inflammatory response, secondary necrotic and apoptotic cell death, and reparative responses. These responses to injury can be mediated and reflected by changes in mRNA concentrations, regardless of whether these changes are due to regulated gene expression or altered cellular composition. Here, I used genome-wide transcritptome analysis approaches coupling data-based WGCNA (weighted gene co-expression network analysis) and knowledge-based network analysis to constructing gene regulatory networks following acute spinal cord injury (SCI).

Introduction

Research on the lesion and repair mechanism is one of the hot areas of spinal cord injury. The primary and secondary injury processes take place sequentially after spinal cord injury, resulting in varying degrees of neuronal and glial cell necrosis, apoptosis, and axonal and demyelinating fracture. Immediate after the lesion, massive cell death takes place surrounding the injury site, which is coupled with necrosis and demyelination. In the early stage, the vasculature destruction of the spinal cord injury, loss of neurons and myelin in both the gray matter and the white matter. Over time, the occurrence of secondary injury, such as inflammation infiltration, glial fibrillary barrier formation, etc., affecting the axon regeneration, formation of synaptic connection, remyelination and functional recovery. Cytological and histological studies have found that secondary injury after spinal cord injury, certain pathological processes, including the micro-environment of the local inflammatory response, glial scar formation, cell edema, microenvironment and changes in pH, and the reconstruction of neural circuits are in very close relationship. Moreover, intervention of certain pathological processes at different stage of injury appeared to generate different outcomes.

Histopathological study found that inflammatory monocytes are the first to participate in secondary injury by releasing proteases that damage endothelial cells and lead to increased permeability of blood vessels to expand, causing wilder distribution of local inflammatory factors, which cause further damage and also inhibit the proliferation of both neurons and myelinating oligodendrocytes. Previous studies focusing on molecular mechanisms of the inflammatory response reviewed many inflammatory

factors that participated in the process of nerve regeneration and repair, i.e. integrins, cytokines (IL-6, LIF), chemokine (CXCL12), nuclear receptors (RXRs) and galactic (Gelatins) (Ulloa and Messmer, 2006). However, due to the complex nature of the inflammatory response, it is not clear what kind of factors and which signaling pathways are the main contributors of the secondary injury. Therefore, the investigation of the molecular mechanisms of inflammation, the types of cells that are involved in inflammatory event and their behavior (i.e. pattern of expression, secreted cytokine, inflammatory cell surface molecules) is important for the understanding spinal cord injury (Kernie et al., 2001, Leal-Filho, 2011).

Spinal cord injury glial scarring is also common pathological process of pathophysiological responses. Histopathological study showed that during the early spinal cord injury, microglia (microglia) and peripheral mononuclear cells, including macrophages, neutrophils are recruited to the lesion area, causing inflammatory reaction. The activated inflammatory cells release cytokines stimulate proliferation of astrocyte or differentiation of neural precursor cells surrounding the injury site, creating the glial scar (Leal-Filho, 2011, Kernie et al., 2001). Although scar tissue can block the formation of synaptic connection and neuron regeneration, it has also been shown that scar tissue formed during the early stage of injury can also protect the neuronal cells from the exposure to the toxic environment and inflammatory cytokines.

Functional recovery and treatment after spinal cord injury

The first step to carry out the treatment of spinal cord injury is to understand what exactly happened at the damaged site, especially changes in cellular and molecular level. Most

of the current studies focus on histopathology, animal behavior or only a handful of genes, but rather lack a comprehensive understanding of the entire event and the underlying complex networks. We used the genome-wide gene expression analysis combined with data-based Weighted Co-expression Network Analysis (WGCNA) and knowledge-based network analysis to better understand the pathological events during and after spinal cord injury and to identify potential targets for therapeutic purposes (Fuller et al., 2007).

The regeneration of neurons and reconstruction of synaptic connection are crucial for the behavioral recovery after spinal cord injury. Inflammatory cytokines infiltration, glial scar formation, cell edema, and pH instability after spinal cord injury are common pathological processes that hinder cellular and behavior recovery. Therefore it is difficult to rely solely on endogenous neurogenesis to restore damaged neural circuits. The goal of treating spinal cord injury is to improve motor function and sensation in patients with lesion. Multiple attempts have been made to treat spinal cord injury with small molecule compounds. BDNF has been shown to promote the regeneration of sensory neurons after the lesion, yet it is difficult for BDNF to cross the blood-brain barrier and therefore problematic for clinical administration (Song et al., 2008, Ochs et al., 2000). Approaches of intervention have gone beyond the traditional neurotropic factors. Drugs that focusing on cleaning up neural toxins, promoting the survival and regeneration of oligodendrocytes and modulating inflammation response etc., have shown certain degree of functional improvement in rodent models (Due et al., 2014, Song et al., 2008). And here we tested the protective effect of 2-methyl-thio-ADP (2MeSADP) after lesion. 2MeSADP is a specific agonist of the P2Y1 purinergic receptor, which belongs to a family of purinergic G protein-coupled receptors that can be

stimulated by nucleotides such as ATP, 2MeSADP, UTP, UDP and UDP-glucose (Bourdon et al., 2006, Domercq et al., 2006). P2Y1 is presented in a wide range of tissues, including platelets, megakaryoblastic cells and glia cells. It has been shown that the stimulating P2Y1 receptor using 2MeSADP can enhance astrocyte mitochondrial metabolism, significantly reduce cytotoxic edema and reactive gliosis in brain injury and partially reverse the damage to neurons (Talley Watts et al., 2013, Zheng et al., 2013). Our results suggested that 2MeSADP treated animals showed a higher neuronal survival rate, less myelination loss and inflammation response than the PBS treated animals after injury.

Materials and Methods

Animals and Surgical procedures

Male C57BL/6 mice were housed in a 12 h light/dark cycle in a specific pathogen-free facility with controlled temperature and humidity and were allowed free access to food and water. All experiments were conducted according to protocols approved by the Animal Research Committee of the Office for Protection of Research Subjects at University of California Los Angeles.

All surgical procedures were performed under general anesthesia with isoflurane in oxygen-enriched air using an operating microscope (Zeiss), and rodent stereotaxic apparatus (David Kopf). Laminectomy of a single vertebra was performed and moderately severe crush SCIs were made at the level of T9/T10 using No. 5 Dumont forceps (Fine Science Tools) ground down to a tip width of 0.5 mm and modified with a 0.25 mm spacer to compress the cord laterally from both sides for 10 s as described previously (Faulkner et al., 2004, Herrmann et al., 2008, Wanner et al., 2013). Animals were randomly assigned to PBS and 2MeSADP treatment group.

Drug Administration

P2Y1-specific agonists 2MeSADP (0.25 mg/kg, 100 μM) or same volumn of PBS was administrated by tail-vein injection for 4 times (6hr, 12hr, 24 hr and 72 hr) after lesion.

Total RNA extraction and microarray

Spinal cord tissue (5mm spanning both sides of the lesion area) was isolated from uninjured animals and injured animals 3, 7 and 14 days post lesion. Samples were temporarily stored in Trizol (Invitrogen) at -80 °C. To extract RNA, spinal cord tissue was homozygous, total RNA was extracted following manufacturer's instructions (TRIzol, Invitrogen). Genomic DNA contamination was removed using Turbo DNase (Ambion). RNA quality was checked using Bioanalyzer (Agilent), samples with RNA integrity number (RIN) over 8 were used for expression profiling.

Whole-genome expression profiling was performed on MouseRef-8 v2.0 Expression BeadChips (Illumina) at SCGC Gene Expression Core, UCLA. Briefly, cRNA was amplified and labeled using the Ambion TotalPrep kit (Illumina). 1st and 2nd strand cDNA was synthesized using the Ambion kit and purified using a robotic assisted magnetic capture step. Biotinylated cRNA is produced from the cDNA template in a reverse transcription reaction. Amplified and labeled cRNA was then hybridized overnight at 58° C to the BeadChip, followed by washing, blocking, staining and drying using Little Dipper processor (SciGene). BeadChips were scanned on Beadarray Reader (Illumina). Un-normalized expression data was extracted and compiled using BeadStudio software (Illumina).

Data analysis

Data processing was done using Limma package from R Bioconductor (Smyth, 2005). Raw data was first normalized against background. Probes were filtered based on detection values generated by BeadStudio, probes with a detection p-value <0.05 in at least one sample were kept. ComBat was used to remove batch effect between the

arrays (Storey). Probes with a FDR<0.05 in pairwise comparison were considered as significantly differential expressed.

WGCNA was used to identify co-expression modules (Langfelder and Horvath, 2008). In short, a "signed" weight adjacency matrix (A) was constructed using soft-threshold (β=12). The topological overlap dissimilarity (dissTOM, dissTOM = 1 - TOM) was used as input of hierarchical clustering to define modules. Modules with a high correlation coeffiency and similar eigengene were merged.

Gene ontology (GO) analysis of genes in each module was done using DAVID (http://david.abcc.ncifcrf.gov) (Huang da et al., 2009).

Results

In this study, I looked at the expressional changes following the progression of SCI in a comprehensive fashion using Illumina beadarray technique. I evaluated both earlier (within 48 hrs) and later time points (up to 10 days post lesion) to better define the temporal pattern of changes. Instead of focusing on certain population, I used the mRNA extracted directly from whole tissue. Although the signal might be diluted by the heterogeneity of sample and thus decreasing the resolution of the expression profiling, we gained a more comprehensive understanding of the molecular events and networks. Later on, we compared injured animals between PBS and 2MeSADP treated groups, which showed the protective effect of 2MeSADP at a molecular level and also showed that expression analysis as a more sensitive approach in detecting subtle changes comparing with traditional assays, such as behavior analysis.

In order to interpret the expression data, I employed both hierarchical clustering and WGCNA, both methods allowed us to cluster the genes sharing a high correlation coefficient in expression regulation. I also showed that genes bellowed to the same clusters or modules are more likely to share a specific function.

Preliminary expression analysis revealed molecular events after lesion

In this preliminary study, I first looked at the expression from lesion site at 2 days, 5 days and 10 days post lesion (dpi). I used an un-supervised k-mean clustering method to group the probes that showed a significant expression change in at least one of the

pairwise comparison. Probes were grouped into 6 clusters as shown in Fig 3.1. GO-analysis using DAVID was then performed to identify the underlying molecular events of each cluster (Fig 3.2).

Not surprisingly, genes related to synaptic transmission, neuronal differentiation, regulation of neurotransmitter levels etc., showed an immediate decrease (within 2 dpi) and a gradual increase starting from 5 dpi. Inflammatory, oxidative stress and immune response related genes were induced by the lesion and showed an increase at 2 dpi or5 dpi. Besides that, genes related to cell cycle also peaked at the same time window, indicating a local proliferation and the proliferating cells were largely immune cells and reactive astrocytes. Genes associated with angiogenesis and myelination was not previously reported in SCI, and we also saw the increase of these genes at 5 dpi, strongly suggesting a repair mechanism.

Co-expression modules identified from PBS and 2MeSADP treated animals

Using the Illumina beadarray, I looked at the transcriptome of 1cm spinal cord tissue (5mm up- and dowm-stream of lesion site) in 9 uninjured and 21 injuried animals with either PBS or 2MeSADP injection at 3, 7 or 14 dpi. Genes that are highly correlated at their expression level are biologically interesting as they might be regulated by common mechanisms and participate in shared biological processes. i grouped transcripts with correlated expression levels into gene coexpression modules using WGCNA, which maximizes the network's scale-free properties. WGCNA identified 18 modules. Figure 3.3 showed the clustering dendrogram and the scaled average signal intensity of probes in each of the module and the correlation coefficient was calculated from the

module eigengenes (i.e., first principal component of the expression values across subjects).

The expression level of genes in module lightgreen, turquoise, red and yellow showed decrease after lesion; module black, purple, greenyellow and pink showed a immediate decrease after lesion and a slow recovery between 7 dpi and 14 dpi; expression level of the genes in module grey60, lightcyan, green, cyan and salmon showed a continuous increase after lesion; genes in module midnightblue, blue and brown increased rapidly after injury and decreases in later time point; while genes in module tan and magenta changed constantly (Figure 4.3). Using Eigengene-trait correlations, 11 out of 18 modules (lightcyan, midnightblue, blue, brown, turquoise, red, yellow, black, purple, greenyellow and salmon) were identified to be significantly correlated (p value<0.05) (Figure 3.5).

I then tested each module for enrichment of specific biological functions or pathways using GO analysis. Gene ontology indicated these modules were associated with Apoptosis (midnightblue, blue and salmon modules), cell cycle (brown and green module), protein localization (magenta modules), mRNA processing (greenyellow module), immune response (lightcyan, blue and salmon modules), oxidative reaction (red module), and cell adhesion (grey60 module), angiogenesis (midnightblue, brown and tan module), neuron/glia differentiation (turquoise, purple and pink module), synaptic transmission (yellow and black module), muscle development (lightgreen and cyan module) (Figure 3.6). The expression changes of genes in those modules showed consistent trend with the preliminary study.

2MeSADP treatment showed neural protective effect in SCI animals in latter time points

I then look at the expression change induced by 2MeSADP treatment. Figure 3.7 shows the average expression level of genes in each module; samples were grouped based on treatment groups.

In general, the global expression pattern of the 2MeSADP treated animals was closer to the uninjured condition than control group at 14 dpi. For instance expression of genes in module black and yellow, which were enriched for synaptic transmission (p value = 1.86 E-3 and 2.11 E-3) showed a continuous decrease after lesion in control group. While 2MeSADP treatment seem to lead to a more dramatic decrease at 3 dpi, the expression of genes in these two modules also showed more increase from 3 dpi to 14 dpi, and the average expression level was much closer to the uninjured samples. Genes in the pink module, which was enriched for neuron projection and development (p value = 3.79 E-2, 4.07 E-2) showed a smaller decrease comparing to the control group after injury and the average expression was also closer to the uninjured samples at dpi14 (Figure 3.7, 3.8), suggesting a neural protective effect of 2MeSADP.

Besides that, we noticed that 2MeSADP treatment caused different affect on modules bellowe to the same GO term. Both tan and brown modules were enriched for angiogenesis and blood vessel development and 2MeSADP treatment does not seem to cause any changes to the genes in brown module. However, genes in tan module remained the same level as uninjured group till 3 dpi, indicating the dys-regulation of genes in tan module by the lesion was delayed by 2MeSADP treatment. But the underlying mechanism is still unknown, we speculate that extending 2MeSADP injection

might prolong the flat phase. Similar observation was made when looking into modules associated with cell cycle and immune response. 2MeSADP treatment did not seem to affect genes in brown and blue modules, while it induced the expression of genes in green and lightcyan module.

Discussion

Spinal cord injury induces a series of complex cellular responses that involve multiple cell types at different stages post lesion. Using the high-throughput microarray technique, I examined the changes of 18,779 probes covering 13,638 genes induced by the lesion at multiple time point. We found a time-dependent increase in the expression of the genes related to cell cycle, oxidative stress, inflammation, immune response and angiogenesis; and decrease in the expression of the genes related to neuron, glia, synaptic transmission and cell adhesion.

Cell cycle and apoptosis

WGCNA and GO analysis identified 2 modules related to cell cycle (brown and green module) and 3 modules related to apoptosis (midnightblue, blue and salmon module). Genes in both GO clusters showed an increased expression immediately after lesion, midnightblue and brown module showed a decrease starting at 3dpi, while the other 3 modules showed a continuous increase till 14dpi. We saw a very subtle decrease in the expression of apoptosis related genes in 2MeSADP treated group comparing with control (Fig 3.8). Caspase-3 (Casp3) is a protein associated with apoptosis, its expression increased after the lesion, suggesting a continuous wave of cell death. Its expression was decreased at both 3dpi and 14pdi with 2MeSADP treatment comparing with control group, indicating 2MeSADP treatment can reduce cell death not only during the treatment (0-3dpi), its effect lasted till 14dpi. Cd40, which is a membrane protein of the

tumor necrosis factor receptor family that has been reported to contribute to tissue necrosis during early phase after lesion, showed a decreased expression in 2MeSADP treated groups. On the other hand, cell cycle related gene Pcna showed an increased expression at both 3dpi and 14pdi with 2MeSADP treatment comparing with control group, indicating 2MeSADP treatment can promote cell proliferation (Fig 3.13).

Immune response and inflammation related genes

Three modules (lightcyan, blue, salmon module) were associated with immune response, defense response, antigen processing and presentation (lightcyan and blue module); myeloid cell differentiation, lymphocyte differentiation and immune system development (salmon module). 2MeSADP treatment led to a slight decrease in the expression of salmon and blue module and a significant decrease in the expression of lightcyan module at 14 dpi, suggesting 2MeSADP treatment's role in modulating immune response (Fig 3.8). Genes involved in antigen presentation, B/T cell activation, such as H2-Eb1, Cd47, Cd3d (T cell marker), Cd22 (B cell activation marker) etc., showed decreased expression in 2MeSADP treated group at 3dpi, 7 dpi and 14dpi (Fig3.12).

Although immune response can help to eliminate pathogens, clear debris and aiding repair, inflammation reaction in the later stage of lesion can lead to neuronal cell death. 2MeSADP may act to protect neuronal cells by reducing immune response and inflammation.

Neurons and Glia cells

2MeSADP is an agonist with high specificity and affinity for the purinergic receptor P2ry1, possibly P2ry12 and P2ry13 as well. Activation of purinergic receptors in cultured astrocytes triggers glutamate release, which involves the metabotropic IP3-mediated intracellular Ca2+ signaling and the additional transductive events mediated by tumor necrosis factor-α (TNFα) and prostaglandins (PG) (Zheng et al., 2013, Domercq et al., 2006). Indeed, we observed a great induction of Ip3, Tnfα and prostaglandins in our 2MeSADP treated groups, while the increase of Ip3 and Tnf was most significant immediate after 2MeSADP injection at 3 dpi, the induction of prostaglandins lasted till 14 dpi (Fig 3.14).

Astrocytes are known to play critical roles in neuronal survival subsequent to CNS lesion. It has been shown that immediate administration of 2MeSADP after brain infarction can significantly reduce cell swelling, necrosis and brain infarcts (Zheng et al., 2013).

There was more neuronal cell loss in control group comparing with 2MeSADP treated groups. We saw an overall larger percentage of Nfm (neurofilament) positive cells at the lesion site at 14dpi and 28 dpi and less glial scar in 2MeSADP treated groups (data nor shown here). Our expression profiling also showed a similar trend. There were three modules (turquois, balck and yellow) that were primarily associated with neuronal cells, all of which showed a higher average expression at 14dpi with 2MeSADP treatment comparing to control (Fig 3.8). Expression level of neuronal markers and transcription factors that are crucial for neural differentiation, such as Sox2, NeuN, Dlx2, Dcx, and Nefh, were higher in 2MeSADP treated groups. The expression of some other neural differentiation associated genes, such as Neurog2 and Neurod6, were significantly higher

at 3dpi but dropped afterwards, suggesting 2MeSADP treatment might be protecting certain neuronal population over the others and a extended treatment might be beneficial (Fig 3.10). The expression of oligodendrocyte marker genes and myelin related genes, such as Mbp, Mog, Mag and Cnp, also showed a higher expression in 2MeSADP treated groups, while astrocytic genes only showed a subtle decrease in 2MeSADP treated groups comparing with control (Fig 3.11 and 3.12)

In summary, genome-wide expression profiling allows us to capture the dynamic response of a mix population during the acute and chronic stages SCI on a molecular level and it was sensitive enough to pick up subtle changes. 2MeSADP treatment was able to (i) modulate immune response and inflammation, (ii) promote neuron survival and regrowth and (iii) protect against myelin loss, (iv) promote angiogenesis. We were able to show the positive effect of 2MeSADP treatment using histochemistry approaches, but we did not see any improvement of motor activity using traditional behavior analysis. The strong agreement between expression profile and histochemistry study suggest that our systematic approach is useful at revealing the underlying molecular processes contributing to the SCI pathology and sensitive at capturing injury and recovery events.

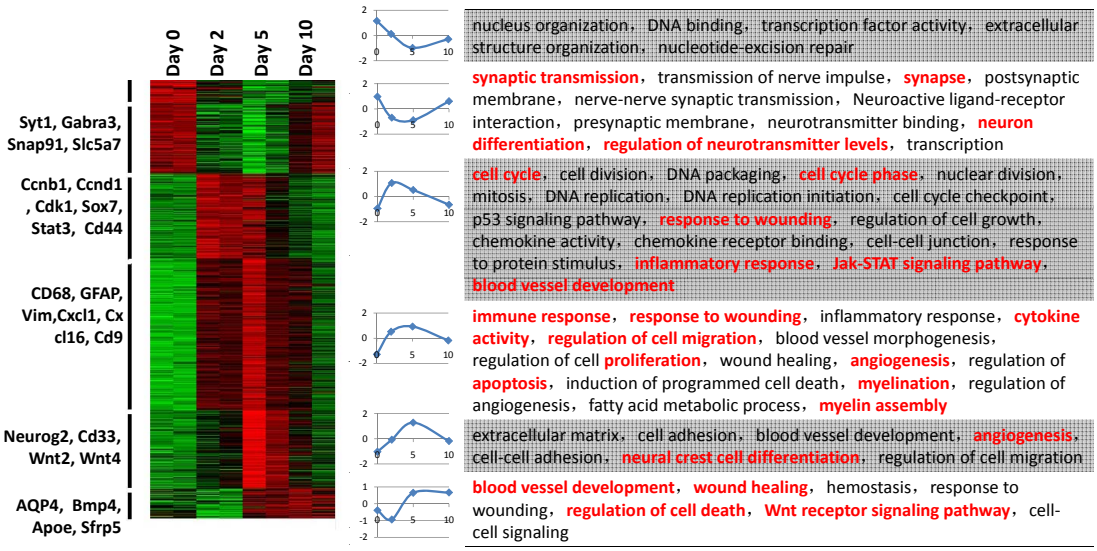


Figure 3.1 Temporal changes in global gene expression changes after SCI. Heatmap showed the expression changes up to 10 days after spinal cord lesion. Genes were grouped into 6 clusters using k-mean clustering method (left). Go terms enriched in each cluster were shown in the table (right). Go terms enriched with a higher significance (p-value) were highlighted with red.

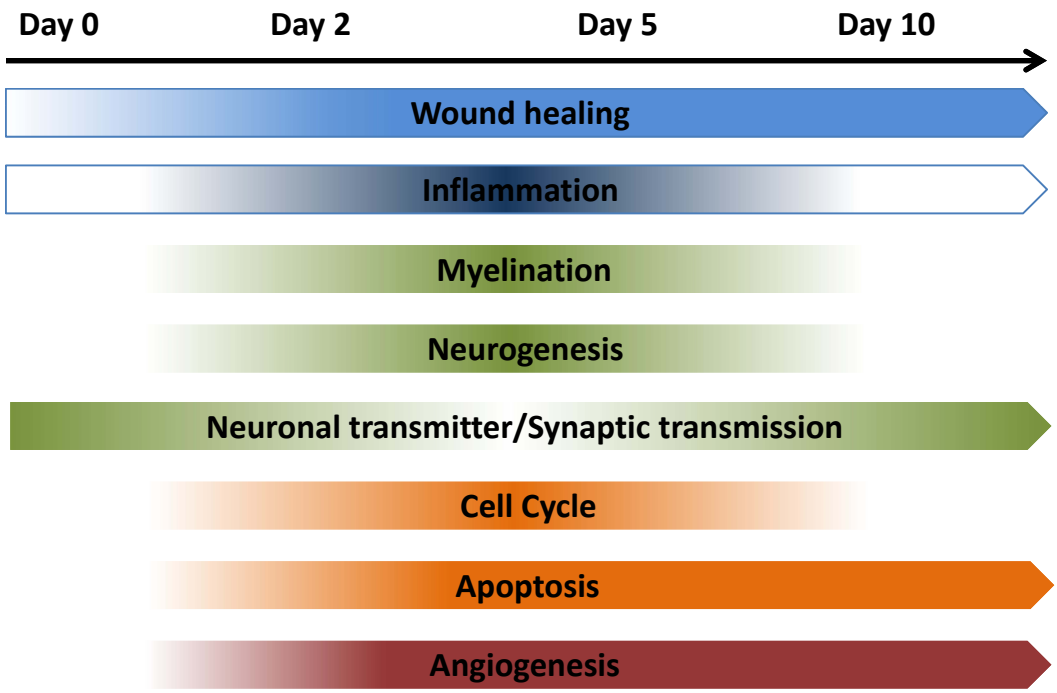


Figure 3.2 Top GO pathways enriched in the differentially expressed genes.

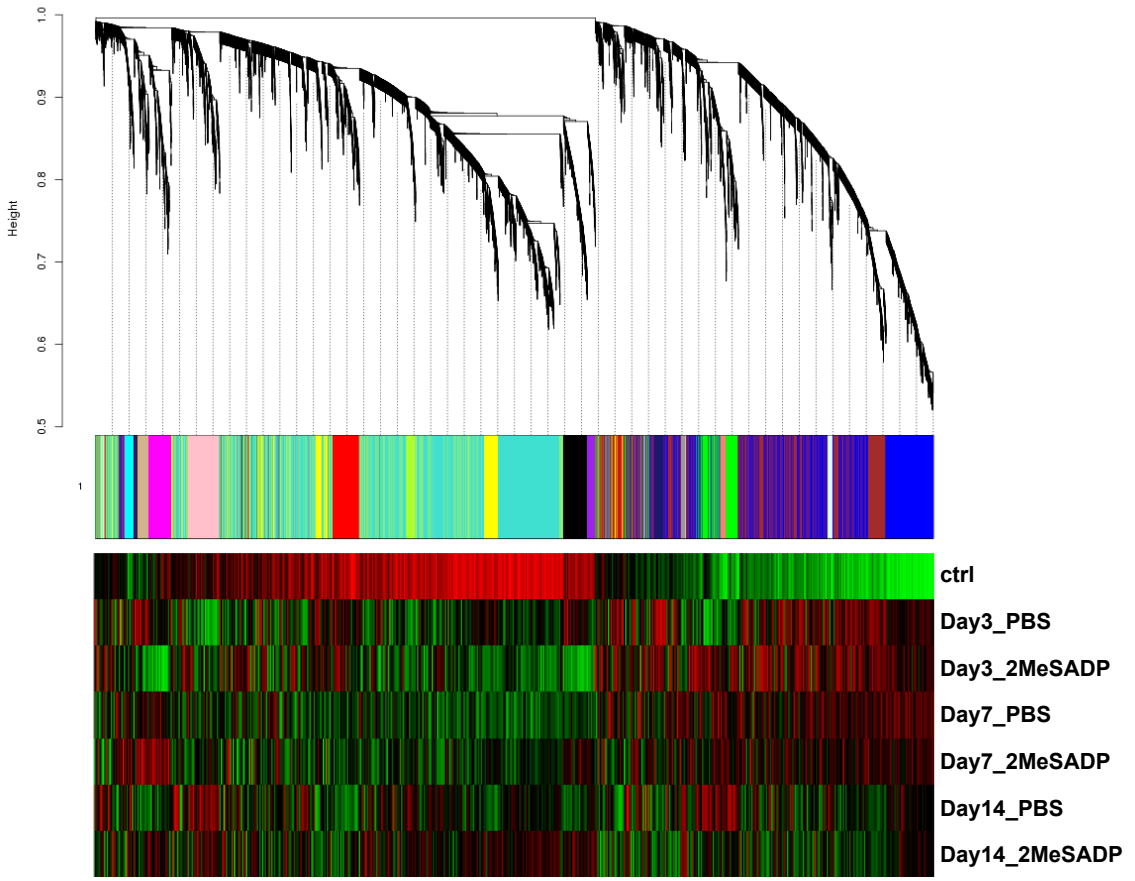


Figure 3.3 WGCNA coexpression network. Clustering dendrogram obtained by average linkage hierarchical clustering using significantly differential expressed probes, each line indicates one probe. The branches correspond to modules of highly interconnected groups of genes. The tips of the branches represent genes that are the least dissimilar and thus share the most similar network connections. The color row underneath the dendrogram shows the module assignment determined by the Dynamic Tree Cut with average expression of each probes. Heatmap at the bottom shows the scaled average expression of each probe in the corresponding group (red indicating higher expression and green indicating lower expression).

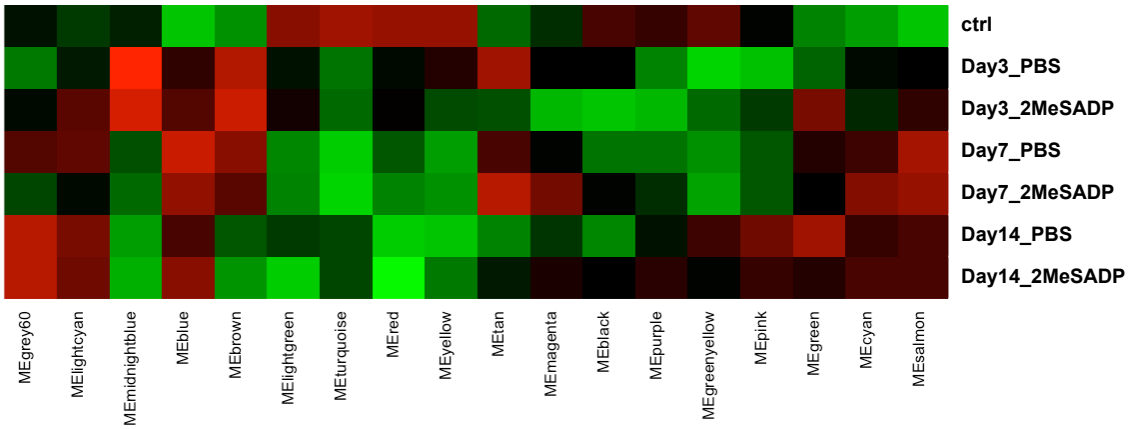


Figure 3.4 Ordered eigengenes expression profiles of each module. Heatmap shows the average eigengene expression in each module (red indicating higher expression and green indicating lower expression).

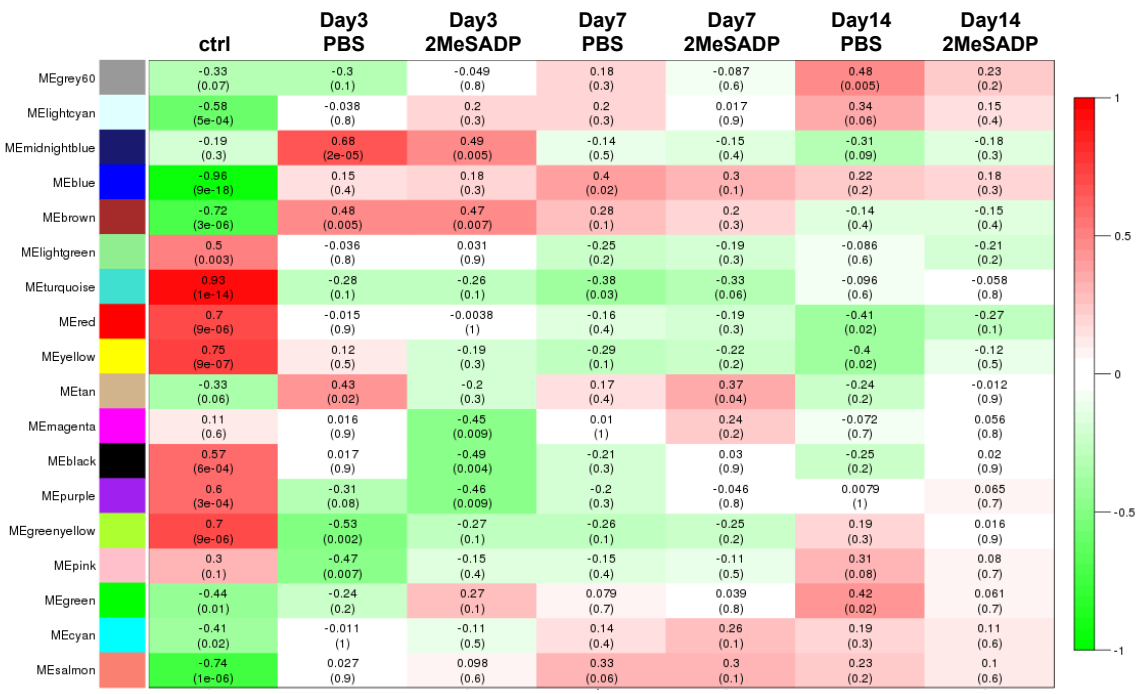


Figure 3.5 Correlation heatmap of between module eigengenes and sample groups. Each row in the table corresponds to a module, and each column to a sample. Numbers in the table reflect the correlations of the corresponding module eigengenes and sample, with the p-values printed below the correlations in parentheses. The table is color coded by correlation according to the color legend.

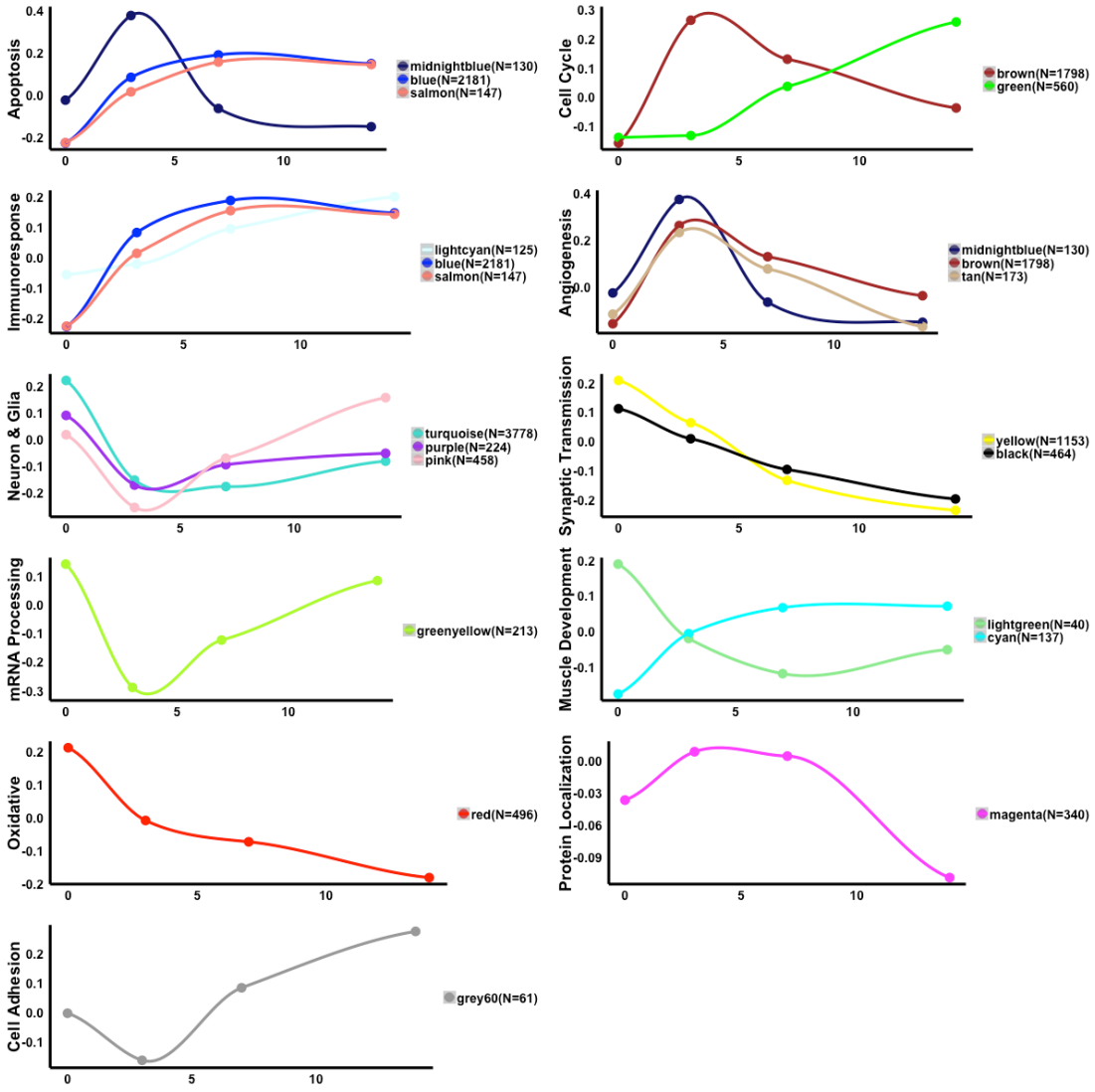


Figure 3.6 Eigengene expression changes in injury only group. Modules are grouped based on the GO enrichment for genes in each module.

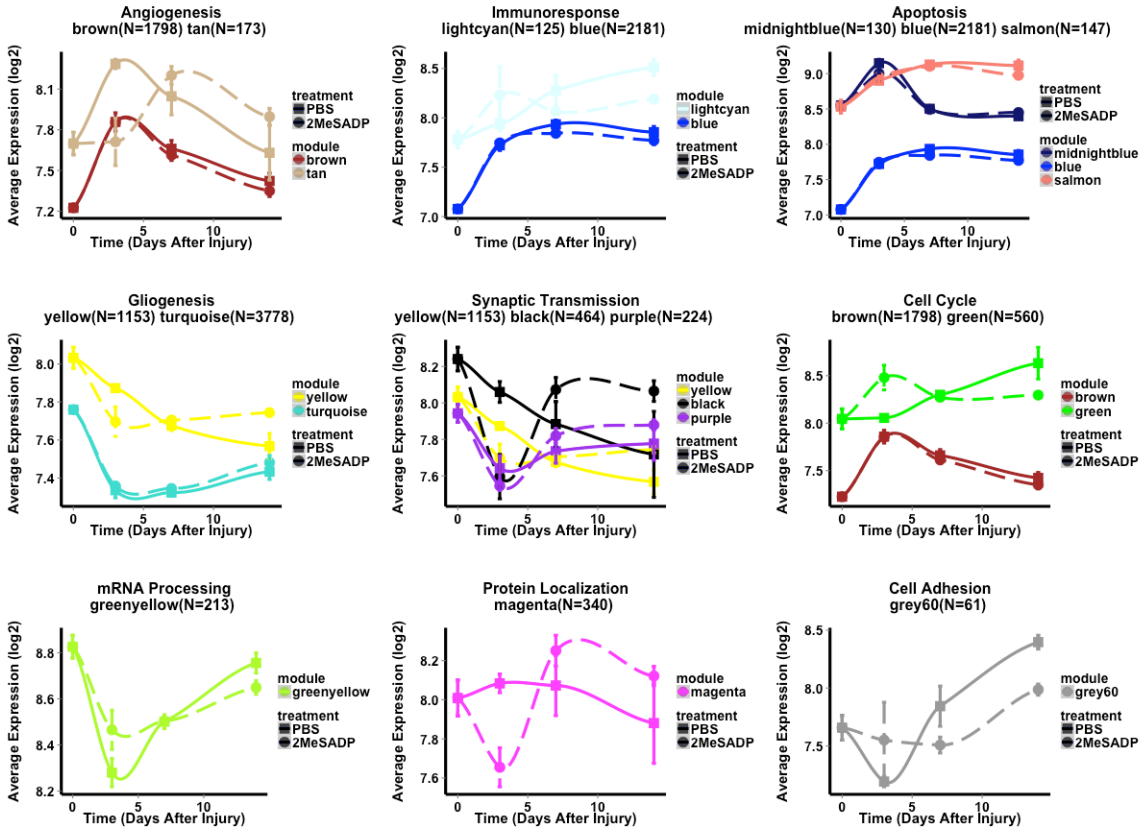
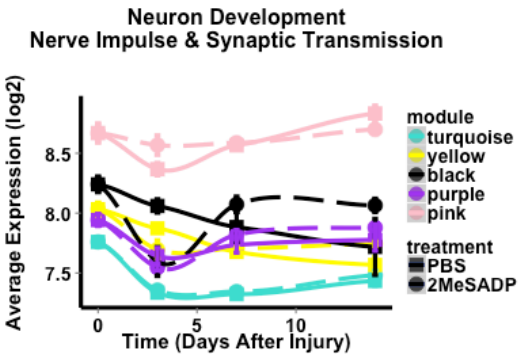
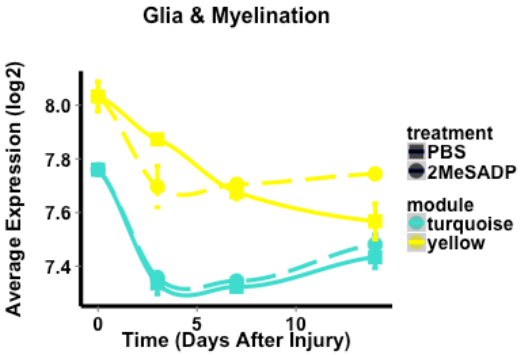


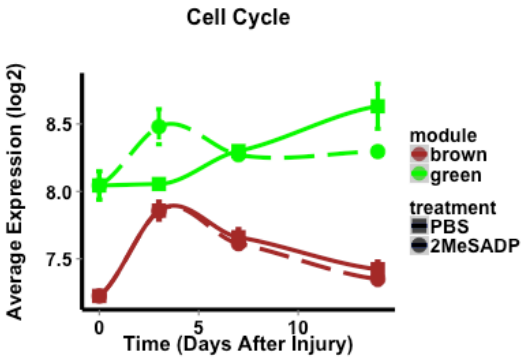
Figure 3.7 Average expression levels of module genes with PBS (solid line) and 2MeSADP treatment (dashed line). Modules are grouped based on the GO enrichment for genes in each module.



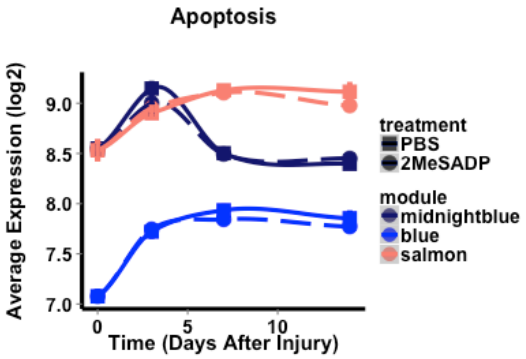
turquoise	
GO:0019226~transmission of nerve impulse	1.22E-05
GO:0046928~regulation of neurotransmitter secretion	2.05E-03
GO:0031175~neuron projection development	6.22E-03
yellow	
GO:0007268~synaptic transmission	9.66E-06
GO:0007274~neuromuscular synaptic transmission	2.11E-03
GO:0050804~regulation of synaptic transmission	7.23E-03
black	
GO:0007610~behavior	6.01E-04
GO:0007268~synaptic transmission	1.86E-03
GO:0007268~synaptic transmission	1.86E-03
purple	
GO:0016192~vesicle-mediated transport	4.96E-03
GO:0007268~synaptic transmission	1.33E-02
pink	
GO:0031175~neuron projection development	3.79E-02
GO:0048666~neuron development	4.07E-02



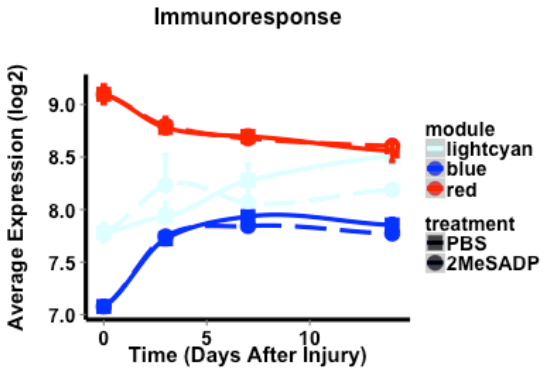
turquoise	
GO:0042552~myelination	3.86E-02
GO:0008366~axon ensheathment	5.23E-02
yellow	
GO:0014013~regulation of gliogenesis	3.61E-02
GO:0045685~regulation of glial cell differentiation	3.61E-02



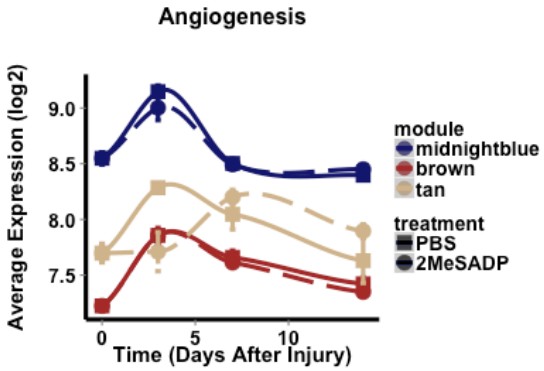
brown	
GO:0007049~cell cycle	1.52E-11
GO:0022403~cell cycle phase	3.84E-08
GO:0042127~regulation of cell proliferation	1.59E-06
green	
GO:0007049~cell cycle	5.17E-04
GO:0022402~cell cycle process	1.69E-03



midnightblue	
GO:0006915~apoptosis	3.27E-02
GO:0008219~cell death	4.84E-02
blue	
GO:0006915~apoptosis	1.13E-06
salmon	
GO:0006915~apoptosis	2.62E-02



lightcyan	
GO:0019886~antigen processing and presentation of exogenous peptide antigen via MHC class II	3.89E-03
GO:0006955~immune response	2.08E-02
blue	
GO:0006955~immune response	5.12E-26
GO:0048002~antigen processing and presentation of peptide antigen	4.33E-13
GO:0006952~defense response	3.64E-19
salmon	
GO:0030099~myeloid cell differentiation	2.86E-03
GO:0002520~immune system development	6.28E-03
GO:0030098~lymphocyte differentiation	7.50E-03



midnightblue	
GO:0048514~blood vessel morphogenesis	4.10E-02
GO:0001525~angiogenesis	5.67E-02
brown	
GO:0001568~blood vessel development	1.54E-08
GO:0001525~angiogenesis	1.84E-08
GO:0001763~morphogenesis of a branching structure	1.90E-05
GO:0048754~branching morphogenesis of a tube	3.28E-05
tan	
GO:0001568~blood vessel development	9.02E-03
GO:0001944~vasculature development	1.02E-02
GO:0035295~tube development	4.13E-02

Figure 3.8 The top GO pathways enriched in each module. Average expression levels of module genes with PBS (solid line) and 2MeSADP treatment (dashed line) are shown on the left with GO enrichment of each module on the right. Modules are grouped based on the GO enrichment for genes in each module.

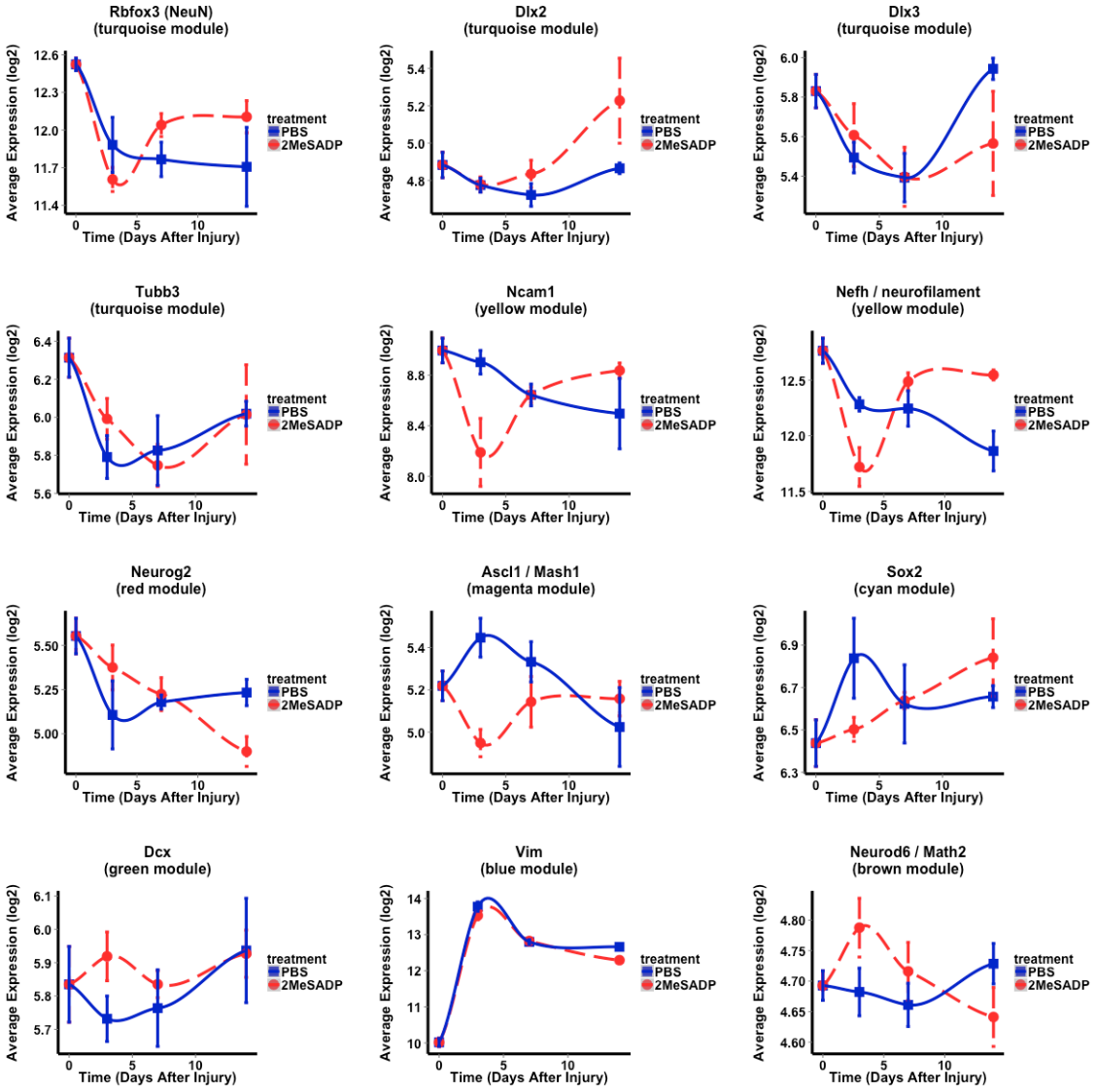


Figure 3.9 Average expression levels of neuronal genes

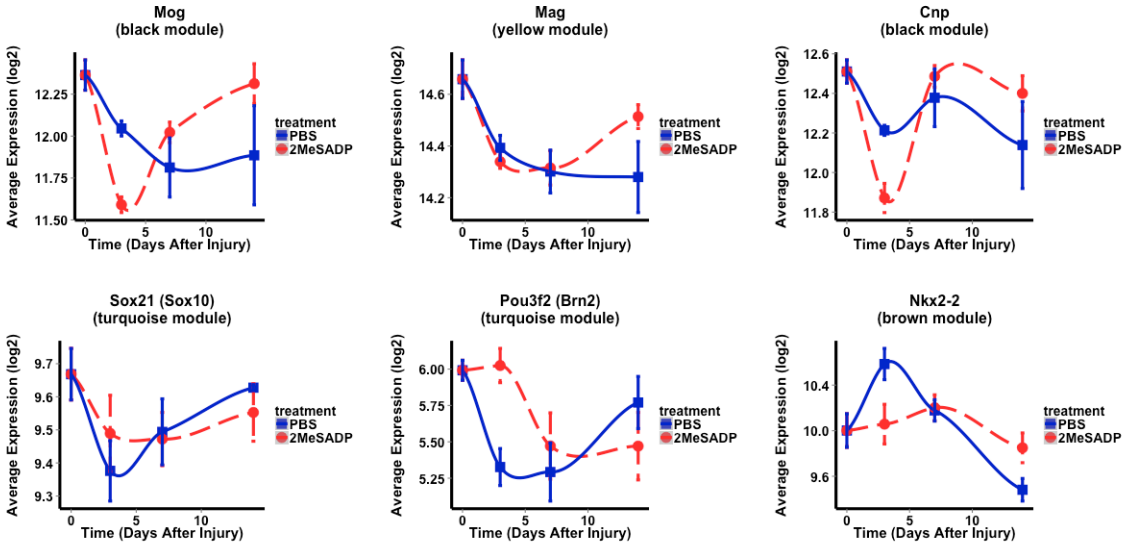


Figure 3.10 Average expression level of oligodendrocyte lineage genes

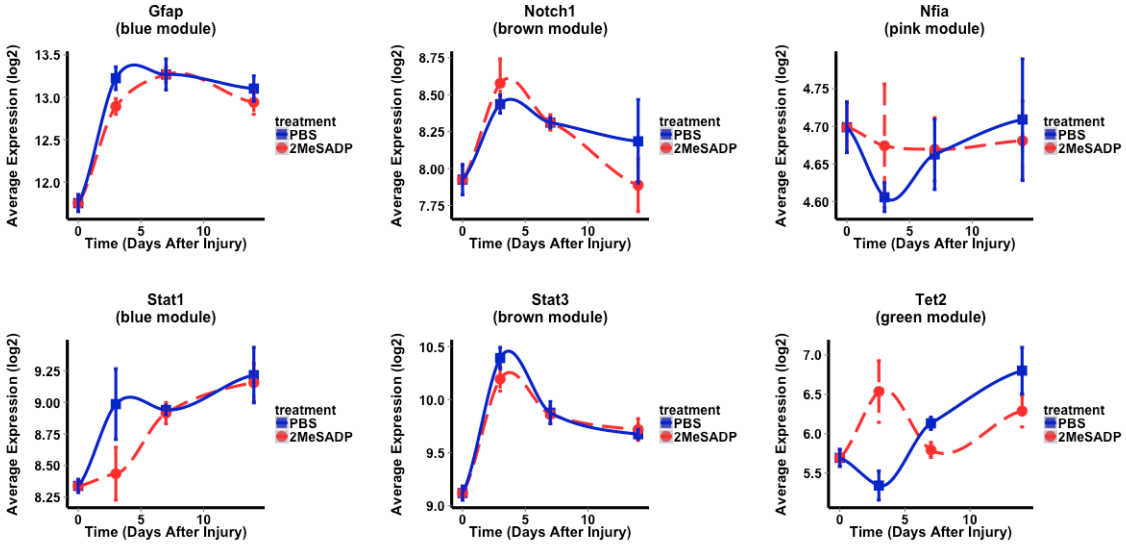


Figure 3.11 Average expression level of astrocytes lineage genes

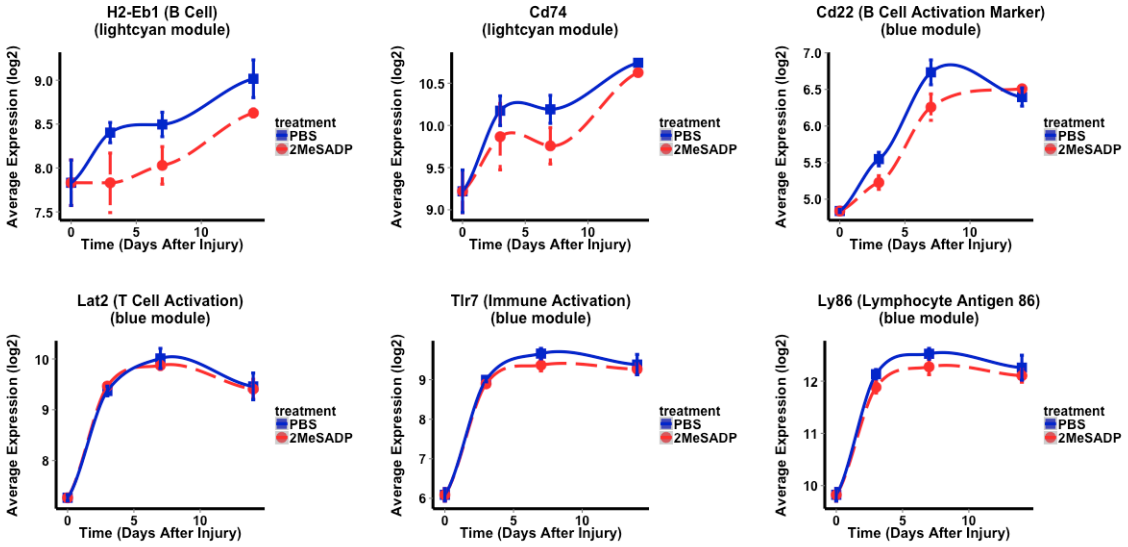


Figure 3.12 Average expression level of genes involved in immune response and inflammation

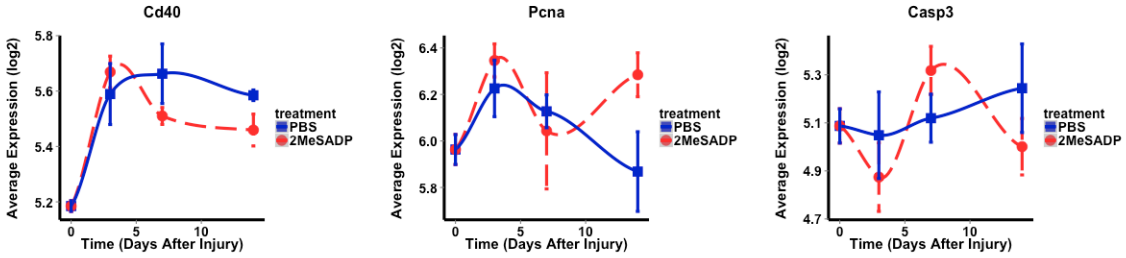


Figure 3.13 Average expression level of genes involved in cell cycle

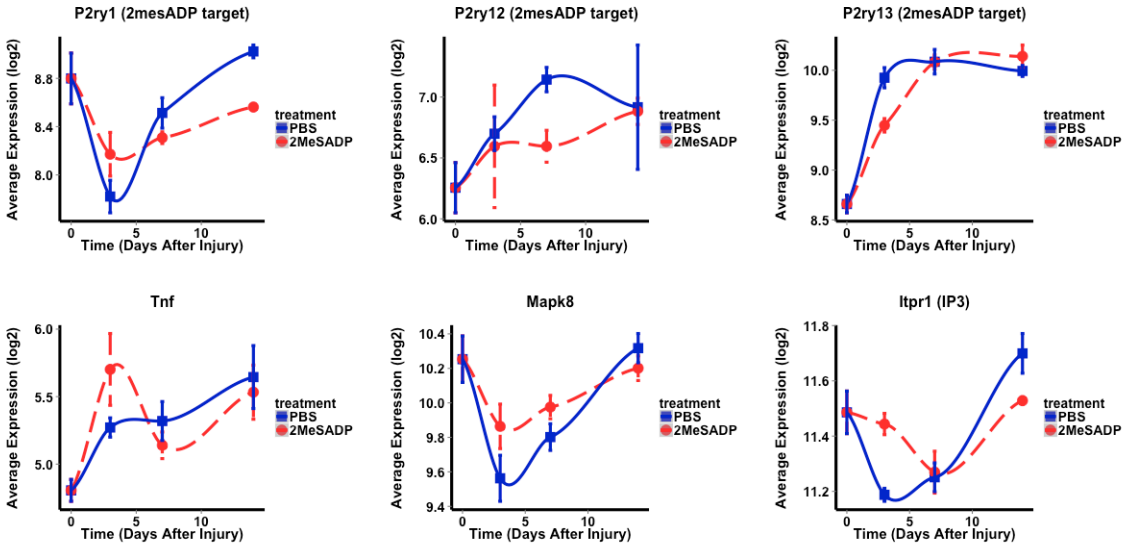


Figure 3.14 Average expression level of P2ry1 downstream members

References

BOURDON, D. M., MAHANTY, S. K., JACOBSON, K. A., BOYER, J. L. & HARDEN, T. K. 2006. (N)-methanocarba-2MeSADP (MRS2365) is a subtype-specific agonist that induces rapid desensitization of the P2Y1 receptor of human platelets. *J Thromb Haemost,* 4**,** 861-8.

CHU, X. P. & XIONG, Z. G. 2013. Acid-sensing ion channels in pathological conditions. *Adv Exp Med Biol,* 961**,** 419-31.

DOMERCQ, M., BRAMBILLA, L., PILATI, E., MARCHALAND, J., VOLTERRA, A. & BEZZI, P. 2006. P2Y1 receptor-evoked glutamate exocytosis from astrocytes: control by tumor necrosis factor-alpha and prostaglandins. *J Biol Chem,* 281**,** 30684-96.

DUE, M. R., PARK, J., ZHENG, L., WALLS, M., ALLETTE, Y. M., WHITE, F. A. & SHI, R. 2014. Acrolein involvement in sensory and behavioral hypersensitivity following spinal cord injury in the rat. *J Neurochem,* 128**,** 776-86.

FAN, G., MARTINOWICH, K., CHIN, M. H., HE, F., FOUSE, S. D., HUTNICK, L., HATTORI, D., GE, W., SHEN, Y., WU, H., TEN HOEVE, J., SHUAI, K. & SUN, Y. E. 2005. DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development,* 132**,** 3345-56.

FAULKNER, J. R., HERRMANN, J. E., WOO, M. J., TANSEY, K. E., DOAN, N. B. & SOFRONIEW, M. V. 2004. Reactive astrocytes protect tissue and preserve function after spinal cord injury. *J Neurosci,* 24**,** 2143-55.

FULLER, T. F., GHAZALPOUR, A., ATEN, J. E., DRAKE, T. A., LUSIS, A. J. & HORVATH, S. 2007. Weighted gene coexpression network analysis strategies applied to

mouse weight. *Mamm Genome,* 18**,** 463-72.

HERRMANN, J. E., IMURA, T., SONG, B., QI, J., AO, Y., NGUYEN, T. K., KORSAK, R. A., TAKEDA, K., AKIRA, S. & SOFRONIEW, M. V. 2008. STAT3 is a critical regulator of astrogliosis and scar formation after spinal cord injury. *J Neurosci,* 28**,** 7231-43.

HU, R., DUAN, B., WANG, D., YU, Y., LI, W., LUO, H., LU, P., LIN, J., ZHU, G., WAN, Q. & FENG, H. 2011. Role of acid-sensing ion channel 1a in the secondary damage of traumatic spinal cord injury. *Ann Surg,* 254**,** 353-62.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

KERNIE, S. G., ERWIN, T. M. & PARADA, L. F. 2001. Brain remodeling due to neuronal and astrocytic proliferation after controlled cortical injury in mice. *J Neurosci Res,* 66**,** 317-26.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LEAL-FILHO, M. B. 2011. Spinal cord injury: From inflammation to glial scar. *Surg Neurol Int,* 2**,** 112.

OCHS, G., PENN, R. D., YORK, M., GIESS, R., BECK, M., TONN, J., HAIGH, J., MALTA, E., TRAUB, M., SENDTNER, M. & TOYKA, K. V. 2000. A phase I/II trial of recombinant methionyl human brain derived neurotrophic factor administered by intrathecal infusion to patients with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord,* 1**,** 201-6.

SMYTH, G., R. GENTLEMAN, V. CAREY, S. DUDOIT, R. IRIZARRY, W. HUBER 2005. *Limma: linear models for microarray data*.

SONG, X. Y., LI, F., ZHANG, F. H., ZHONG, J. H. & ZHOU, X. F. 2008. Peripherally-derived BDNF promotes regeneration of ascending sensory neurons after spinal cord injury. *PLoS One,* 3**,** e1707.

STOREY, J. T. L. A. W. E. J. A. H. S. P. A. A. E. J. A. J. D. sva: Surrogate Variable Analysis.

TALLEY WATTS, L., SPRAGUE, S., ZHENG, W., GARLING, R. J., JIMENEZ, D., DIGICAYLIOGLU, M. & LECHLEITER, J. 2013. Purinergic 2Y1 receptor stimulation decreases cerebral edema and reactive gliosis in a traumatic brain injury model. *J Neurotrauma,* 30**,** 55-66.

ULLOA, L. & MESSMER, D. 2006. High-mobility group box 1 (HMGB1) protein: friend and foe. *Cytokine Growth Factor Rev,* 17**,** 189-201.

WANNER, I. B., ANDERSON, M. A., SONG, B., LEVINE, J., FERNANDEZ, A., GRAY-THOMPSON, Z., AO, Y. & SOFRONIEW, M. V. 2013. Glial scar borders are formed by newly proliferated, elongated astrocytes that interact to corral inflammatory and fibrotic cells via STAT3-dependent mechanisms after spinal cord injury. *J Neurosci,* 33**,** 12870-86.

WU, Z., HUANG, K., YU, J., LE, T., NAMIHIRA, M., LIU, Y., ZHANG, J., XUE, Z., CHENG, L. & FAN, G. 2012. Dnmt3a regulates both proliferation and differentiation of mouse neural stem cells. *J Neurosci Res,* 90**,** 1883-91.

XU, J. C., LYTLE, C., ZHU, T. T., PAYNE, J. A., BENZ, E., JR. & FORBUSH, B., 3RD 1994. Molecular cloning and functional expression of the bumetanide-sensitive

Na-K-Cl cotransporter. *Proc Natl Acad Sci U S A,* 91**,** 2201-5.

ZHENG, W., TALLEY WATTS, L., HOLSTEIN, D. M., WEWER, J. & LECHLEITER, J. D. 2013. P2Y1R-initiated, IP3R-dependent stimulation of astrocyte mitochondrial metabolism reduces and partially reverses ischemic neuronal damage in mouse. *J Cereb Blood Flow Metab,* 33**,** 600-11.

Chapter 4

Summary

Hydroxylation of 5-methylcytosine by Tet2 promotes astrocyte differentiation in embryonic cortical NPCs

The mammalian central nervous system (CNS) is established through a well-organized sequence of events during development. The differentiations of neurons and glial cells from neural projenitor cells (NPCs) are tightly regulated. NPCs are characterized by their self-renewing and multipotent capacities, and it has been shown that extrinsic signaling and intrinsic epigenetic regulation together orchestrate the temporal and spatial lineage differentiations in CNS. Both *in vivo* and *in vitro,* CNS progenitor cells are first differentiates into neurons then glial cells. During development and several passages, NPCs gradually acquire competence for gliogenesis, indicated the existence of a neurogenic to gliogenic switch mechanism.

The basic-helix-hoop-helix (bHLH) transcription factor Olig2 plays a central role in oligodendrocytes and motor neuron development and shows inhibitory effects on astrocytic differentiation, yet the mechanism remains elusive. It has been demonstrated that the presence of Olig2 protein in the nuclei can inhibit the activation of astrocytic genes, yet Olig2 is not directly associated with their promoter. In Charpter 2 of this dissertation, I focused on studying the molecular mechanisms underlying the cell fate specification between neuron and astrocyte by Olig2, namely how Olig2 regulates neural progenitor cells' commitment to neurons or astrocyte. I demonstrated that Olig2 inhibits astrogliagenesis through direct repression of Tet2, the latter plays critical role in NPC fate commitment by activating astrocyte differentiation programs through active and target specific DNA demethylation.

DNA methylation as one of the major epigenetic regulation mechanisms has been postulated to regulate neural differentiations. Our lab has demonstrated that astroglial marker (such as GFAP) and the astrogliogenic members in JAK-STAT pathway (such as STAT1) are repressed through DNA methylation in early NPCs as a mechanism to prevent premature astrogliogenesis. Our lab has previously shown the crucial function of the *de novo* DNA methyltransferase, Dnmt3a, in postnatal neurogenesis. On the other hand, Tet family members, which are involved in the process of active DNA demethylation, have been implicated in embryonic development and neural differentiation. However, the underlying machinery of how DNA methylation and demethylation act in coordinate to regulate the fate of NPCs remains elucidated. In this dissertation, I demonstrated that Tet2 is involved in the regulation of DNA methylation and the neuronal to astroglial lineage switch. My study suggested that Tet2 specifically promotes the expression of astroglial genes during the differentiation of NPCs. There was a significant decrease of DNA methylation and increase in DNA hydroxymethylation at the proximal promoter sites of the astroglial genes during NPCs differentiation, which correlates to the increase in Tet2 expression. Overexpression of Tet2 in NPCs led to active DNA demethylation at the Gfap promoter, which enabled active transcription once the NPCs were exposed to an extrinsic trigger for differentiation.

Digitize molecular events after spinal cord injury

Spinal cord injury (SCI) is characteristically accompanied by a period of secondary cellular degeneration that occurs in injured tissue over a course of hours and days after the initial insult, which often results in permanent neurologic deficit and functional loss.

The initial local mechanical tissue damage would first lead to cellular necrosis, and then induce series of pathophysiological and neuropathological events, including immune system response, reactive astrogliosis and scar formation. To better understand the pathological changes and underlying molecular mechanisms, I looked at the cellular composition and response using transcriptome profiling at different time points after lesion. Coupling data-based network analysis and knowledge-based gene ontology analysis, I constructed a network of how individual types of cells behave and interact with each other and how they may adapt to as well as modulate the local environment, which lead to deeper understanding the mechanisms of injury and leading to safe and effective therapies for SCI. I described the immediate loss of neuronal population and increase in immune response, coupled with a increased activity of cell cycle and DNA replication, RNA processing genes in the early stage after lesion (0 dpi to 3 dpi); I also observed the occurrence of angiogenesis, neurogenesis, synaptic connection and myelination in the late stage (7 dpi to 14 dpi), which indicate a spontaneous recovery. Although traditional behavior analysis did not show any sign of improvement with the treatment of 2MeSADP, there was an increased neuronal survival, synaptic connection and myelination in treated groups comparing with control groups in later stage (14 dpi - 28 dpi) shown by both histochemistry approaches and the transciptome analysis. WGCNA analysis has also identified the 2MeSADP target P2Y1 receptor as one of the hub genes, indicating the possibility that it is one of the key members in mediating the lesion and recovery events.

In summary, the strong agreement between expression profile and histochemistry study

suggests that our systematic approach is useful at revealing the underlying molecular processes contributing to the SCI pathology and sensitive at capturing subtle injury and recovery events. Our system-based analysis framework can be translated to identifying key determinants in the global gene networks as well as genes interested for functional tests and therapeutic targets in the future.