KRUPPLE LIKE FACTOR 4 (KLF4) BINDING TO METHYLATED DNA UNCOVERS NOVEL

REGULATORY MECHANISM IN GLIOBLASTOMA MIGRATION

By

Olutobi Z. Oyinlade

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the

degree of Doctor of Philosophy

Baltimore, Maryland

September 2017

ABSTRACT

Protein microarray technology provides a versatile platform for characterization of hundreds to thousands of proteins in a parallel and high-throughput manner. Over the last decade, applications of functional protein microarrays in particular have flourished in studying protein function at a systems level and has led to the construction of networks and pathways describing these functions. Relevant areas of research include the detection of various binding properties of proteins, the study of enzyme-substrate relationships, the analysis of host-microbe interactions, and profiling antibody specificity. In addition, discovery of novel biomarkers in autoimmune diseases and cancers is emerging as a major clinical application of functional protein microarrays. The Zhu lab used a protein microarray-based approach to systematically survey the entire human transcription factor (TF) family and found numerous purified TFs exhibit specific binding activity to methylated and unmethylated DNA motifs of distinct sequences. DNA methylation, especially CpG methylation at promoter regions, has been generally considered as a potent epigenetic modification that prohibits (TF) recruitment, resulting in transcription suppression. To elucidate the underlying mechanism, we focused on Kruppel-like factor 4 (KLF4), and decoupled its mCpG- and CpG-binding activities via site-directed mutagenesis. This study suggested that mCpG-dependent TF binding activity is a widespread phenomenon and provides a new framework to understand the role and mechanism of TFs in epigenetic regulation of gene transcription

Further studies from the Xia Lab showed that KLF4 promotes cell adhesion, migration, and morphological changes, all of which are abolished by a single mutation in R458 to an alanine. Surprisingly, 116 genes are directly activated via mCpG-dependent KLF4 binding activity. These

studies demonstrate a new paradigm of DNA methylation-mediated gene activation and chromatin remodeling, and provides a general framework to dissect the biological functions of DNA methylation readers and effectors.

My thesis project focused on characterizing the KLF4 methylation dependent migration targets identified in previous studies. UDP-Glucose 6-dehydrogenase (UGDH) was identified as one of the downstream targets of KLF4-mCpG binding activity. This work shows that KLF4 upregulates UGDH expression in a mCpG-dependent manner, and UGDH is required for KLF4 induced cell migration in vitro. UGDH produces UDP-α-D-glucuronic acid, the precursors for glycosaminoglycans (GAGs) and proteoglycans (PGs) of the extracellular matrix. Elevated GAG formation has been implicated in a variety of human diseases, including glioblastoma (GBM). UGDH knockdown decreases glycosaminoglycan (GAG) abundance in GBM cells, as well as cell proliferation and migration in vitro. In intracranial xenografts, reduced UGDH inhibits tumor growth and decreases expression of extracellular matrix, e.g. tenascin C, brevican. These studies demonstrate a novel DNA methylation-dependent UGDH upregulation by KLF4. Developing UGDH antagonists to decrease the synthesis of extracellular matrix components will be a useful strategy for GBM therapy. Further studies of KLF4 mCpG interaction revealed a novel KLF4 binding to enhancer regions to regulate transcription such as the BLK gene. Through Chromosome Conformation Capture (3C) analysis, we found *BLK* was activated by KLF4 binding to enhancers in a methylation dependent manner. In addition, we found that in genome scale, KLF4 binds to methylated enhancer regions and activates gene transcription.

Ph.D. DISSERTATION REFEREES FOR OLUTOBI OYINLADE

Heng Zhu, PhD., Professor, Pharmacology and Molecular Sciences (Thesis Advisor)

Shuli Xia, PhD., Associate Professor, Department of Neurology (Thesis Advisor)

Carolyn Machamer, Ph.D., Professor, Department of Cell Biology (Thesis reader)

ACKNOWLEDGEMENTS

I would like to acknowledge and thank all the people who have helped me accomplish my goals as a scientist, and who have contributed to my life successes. First and foremost, I would like to thank my thesis advisors, Dr. Heng Zhu and Dr. Shuli Xia, for their careful guidance, patience, and unwavering support throughout my journey. Heng has always given me the freedom to pursue my own scientific avenues. I appreciate his trust in me which has inspired the confidence I have come to feel towards the end of my PhD candidacy. I am so thankful for giving me the opportunity to pursue my scientific interest in the Xia Lab. I will like to thank Shuli for accepting me into her lab in the middle of my graduate career. Shuli has always been supportive of me and I have learned so much from her. I'm grateful to both my mentors as they have been very instrumental in my development as a scientist.

I want to thank Dr. John Laterra for taking the time to direct the development and evolution of my project. I appreciate all your insightful contributions, valuable ideas, and suggestions over the years. I would also like to thank all of the members of the Laterra Lab for creating a uniquely collaborative and inspiring atmosphere to train in as a graduate student. I always look forward to presenting in our lab meetings as all your suggestions and insights really helped shape my project. In particular, I would like to thank Dr. Hernando Lopez. Hernando has been a model scientist and an excellent teacher. I must also thank the members of the Zhu lab, both past and present as they helped forged my development from the naïve first year graduate student I was.

I extend my deepest thanks to my family and friends for their unending love and support. I thank my mom Mrs. Olufunmi Ajala for being the paragon of excellence. My mother's death after ten years with Multiple sclerosis sparked a fresh moxie in me. She was and still is the force behind my passion for research. Witnessing my mother's resilience and optimism taught me that I should never give up and

TABLE OF CONTENTS

VIII

TABLE OF FIGURES

LIST OF TABLES

CHAPTER 1

APPLICATIONS IN HIGH-CONTENT FUNCTIONAL PROTEIN

MICROARRAYS

1.1 Introduction

The protein microarray has revolutionized high-throughput biology by allowing researchers to simultaneously interrogate biochemical function of hundreds of thousands of proteins arrayed at high density on a solid surface. In addition to being a highly parallel platform, protein microarrays are capable of detecting interactions and modifications to femtomole quantities of protein, reduce experiment-to-experiment variability, and require minimal sample input, making this technique a powerful tool appropriate for clinical and basic research. Protein microarrays can be categorized into three major groups: analytical, functional, and reverse phase. Functional protein microarrays are constructed from purified proteins or protein fragments encoded by an organism for biochemical assays to characterize binding interactions, modification, or enzymatic activity of proteins printed on the array [1]. Analytical microarrays utilize affinity reagents, most often antibodies, to detect and/or quantify a large number of proteins presented in a complex biological sample [2]. Conversely, reverse phase protein microarrays expand the applications of classical protein microarrays by spotting many complex mixtures, such as cell/tissue lysates or serum samples, directly on the surface of the slide and probes with specific binding characteristics can be used to identify components within each mixture [3].

Applications with functional protein microarrays in particular have dramatically expanded over the past decade. In this review, we will highlight the design and methodology of high-throughput screens with functional protein microarrays and summarize applications and recent achievements within the field (Outlined in Figure 1-1).

1.2 Design and Fabrication of Functional Protein Microarrays

When designing experiments with functional protein microarrays for high-throughput screens, it is essential that individually purified proteins of full-length are used, and that they remain in their native conformation once printed on the array surface. In addition, choosing a proper detection method and surface chemistry is crucial to the success of an assay because both can affect the outcome of the results. After each of these factors has been considered, functional protein microarrays can be fabricated for a variety of applications, which will be discussed in the next section [4].

Protein Production

Generating proteins necessary for high-content functional protein arrays presents a significant technical barrier and often calls for the parallel purification of hundreds to thousands of proteins. This is further complicated by the need for relatively large amounts of pure proteins with the correct post-translational modifications and conformation to produce active proteins. Recent endeavors in both industry and academia have applied a variety of ingenuities in vitro and in vivo expression/purification systems enabling the construction of various proteome libraries as summarized in Table 1-1.

Table 1: High-Content Functional Protein Microarrays

Name of Array	No. of proteins	Proteome Coverage	Company/Lab Group	References
HuProt (Human)	19,394	~80%	CDI Laboratories, Zhu Lab	[1-3]
ProtoArray (Human)	>9,000	~45%	Life Technologies	[4-6]
PrESTs (Human protein epitope signature tag)	>10,000 protein fragments	N/A	Uhlen Lab	[7-9]
Arabidopsis	~17,400	50%	Arabidopsis Biological Resource Center, Dinesh-Kumar/Snyder Labs	[10-12]
S. cerevisiae	5,800	~85%	Zhu/Snyder Labs	[13, 14]
Nucleic Acid Programable Protein Array (NAPPA)	~2,500	~15%	Labaer Lab	[15-17]
M. tuberculosis	4,262	>98%	BC-Bio, Tao Lab	[18]
E. coli (K12)	4,256	>98%	Zhu/Chen Labs	[19]
Herpesviridae (HSV-1, EBV, KSHV, HCMV)	350	N/A	Zhu/Hayward Labs	[20]
Pathogens Antigen (Various Species)	200 - 4,000	N/A	Antigen Discovery Inc., Felgner Lab	[21-24]
Influenza Array	127	N/A	Carter Lab	[25]

Surface Chemistry

Once a library has been constructed, proteins can be immobilized onto a functionalized surface or slide. These surfaces can consist of reactive groups that form covalent linkages with spotted proteins, a 3-dimensional matrix restricting protein movement, or affinity reagents that form a strong interaction with some component of the printed proteins (Table 2) [4]. Maintaining protein integrity and activity is essential; as a result, the benefits and shortcomings of each surface chemistry should be considered before selecting the method most appropriate for the desired application.

Table 2: Surface Chemistry for Protein Immobilization

Surface Type	Examples	Advantage	Disadvantage
Absorption (Non-covalent)	Nitrocellulose, polystyrene, polyvinylideen fluoride nitrocellulose (PVDF)	3D surface, high binding capacity	High background, random orientation, non-specific binding
Diffusion (Non-covalent)	Agarose gel, polyacrylamide	3D surface, high binding capacity	Random orientation, weak protein immobilization
Covalent	Aldehyde, Epoxy, NHS, carboxylic ester, etc.	Permanent attachment, low background	Random orientation, low binding capacity
Affinity Capture	Streptavidin, Ni^{+2} NTA, Glutathione, Sortase (Gly_3)	Uniform orientation, strong/specific attachment	Must have domains or tags (e.g., biotin)
Metal	Gold, Silver, Steel	Compatible with SPR and MS/MS	Expensive, non-transparent

Detection Methods

The use of radioisotope- or fluorescently-labeled probes is the most common means of detecting biological interactions and reactions (Figure 2). Such label-dependent detection is straight-forward and simple to apply to most assays; however, these end-point measurements are not amendable to determination of kinetic parameters and requires labeling a probe-of-interest/cofactor, or use of affinity reagents to detect signal. To minimize perturbation to biological samples, a number of label-free detection platforms have been developed or adapted to the protein microarray format, including mass spectrometry, surface plasmon resonance (SPR), and oblique incidence reflectivity difference (OIRD) [5,6,7,8,9]. Photonic crystal (PC)-based detection is another emerging sensitive label-free detection method that functions by measuring small changes in refractive index of pores embedded on the surface of a silicon wafer to monitor binding interactions [10-12]. An additional benefit of OIRD, PC, and SPR is that binding interactions can be monitored in real-time allowing for the association and dissociation rate constants, k_{on} and k_{off}, to be calculated. These rate constants give a more complete picture of binding interactions than the dissociation equilibrium constant, K_d, as multiple values of k_{on} and k_{off} can give rise to the same K_d.
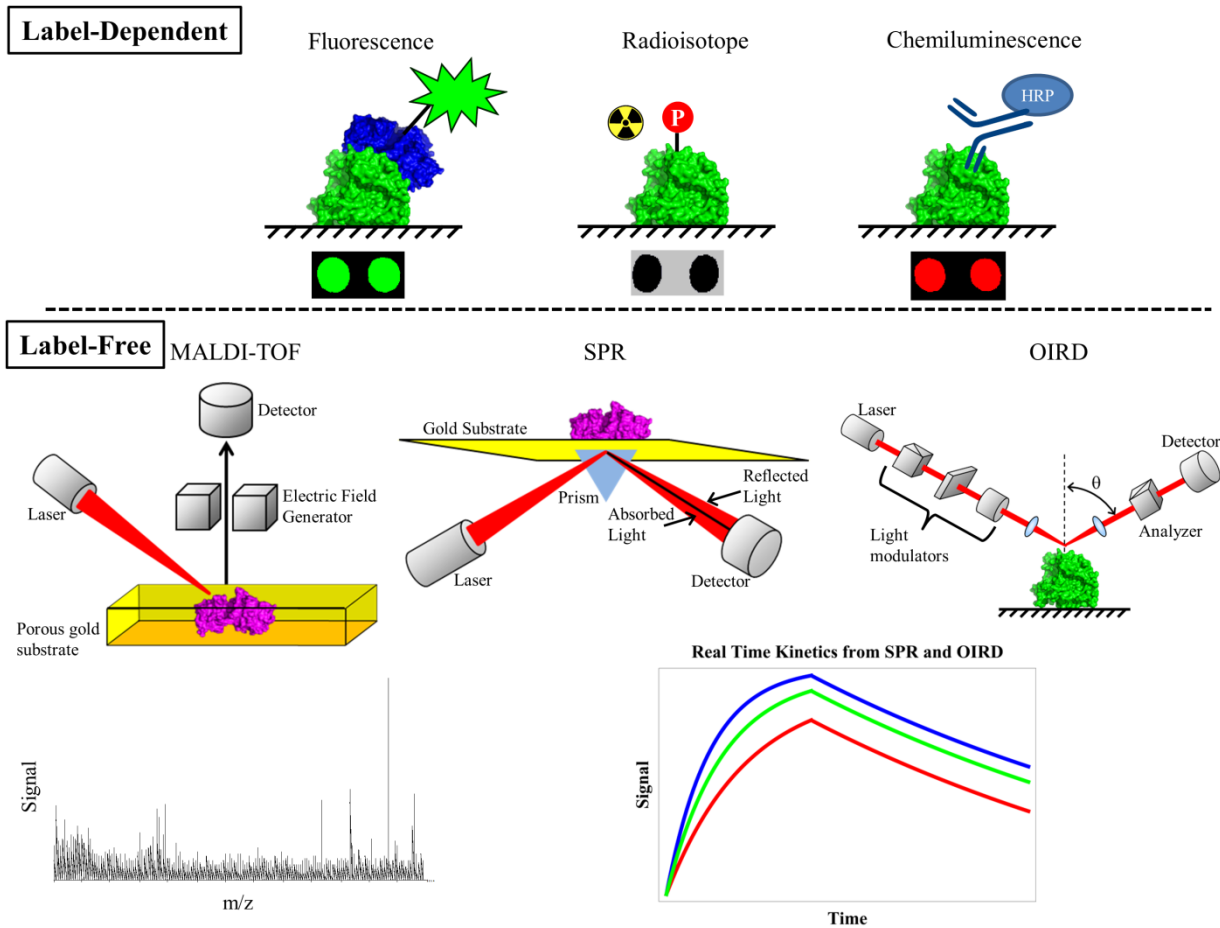
Figure 1-1: Detection Methods for Protein Microarrays

Label-dependent detection of protein interactions and modification is the most common (Top panel)

[26]. Label-free detection methods do not require that proteins be modified; consequently, protein will

have native conformation and is amendable to real-time kinetic monitoring (Bottom panel) [27-31].

Results respresentative of experiments are displayed below each detection method.

1-3. Applications in Basic Research

Array-based approaches have been powerful tools for protein expression profiling, biomarker discovery, post-translational modification study, and the examination of interactions with protein, DNA, RNA, peptide and small molecules. A typical large-scale project usually begins with a biological question, which dictates the design of the biochemical assays performed on a functional protein microarray (Figure 1-1, left panel). The power of the protein microarray is derived from the near-comprehensive, unbiased screening of a proteome. The resulting datasets are then processed with sophisticated bioinformatics tools to reveal otherwise hidden biological functions, based on which a new hypothesis may be generated. Finally, orthogonal in vitro and in vivo assays are performed to test the hypothesis.

Using Protein Microarrays to Characterize Protein-Nucleic Acid Interactions

Characterization of protein-nucleic acid interactions is essential for understanding regulation of genes and protein expression within the cell and is the focus of international collaborations such as the Encyclopedia of DNA Elements (ENCODE) [13]. Protein microarrays can play a vital role in the elucidating the molecule mechanisms responsible for the sophisticated regulation within cells [14]. For example, in one of our recent projects we questioned whether a large number of human transcription factors (TFs) existed that preferentially recognize methylated CpG (mCpG)-containing DNA motifs [15]. We designed a competition-based assay to identify preferential methylated DNA-binding by probing 154 methylated DNA motifs to our human TF protein microarrays. Bioinformatic analysis revealed that 47 TFs and co-factors could preferentially bind to methylated DNA motifs; 17 of them recognized both methylated and unmethylated motifs of very different sequences. This led us to hypothesize that some human TFs could recognize methylate promoters to activate transcription in cells. Further validation demonstrated that one of the 17 TFs, KLF4, binds to methylated promoters in human embryonic stem cells and such mCpG-dependent binding activity could activate downstream

8

gene transcription. Functional protein microarrays can be employed to identify novel protein interactions with coding- and non-coding RNA representing additional control and regulation within the cell [16].

Post-Translational Modifications

Protein posttranslational modifications (PTMs) play an important role in regulating protein activity, stability, and/or subcellular localization. Among the hundreds of PTMs identified so far, enzyme-dependent, reversible protein phosphorylation, ubiquitylation, SUMOylation, acetylation, and glycosylation are the best studied. The importance of creating systematic network maps of enzyme-substrate relationships is emphasized by the layered nature of protein regulation in biological machinery; the activity of modifying enzyme are frequently themselves controlled by PTMs. By performing phosphorylation reactions on human protein microarrays and integrating the existing phosphorylome datasets, Newman et al. were able to construct an activity-based human phosphorylation network that connects a kinase not only to its substrate(s), but also to a particular phosphorylated residue [17-20]. Such large-scale analysis of enzyme-substrate relationships is being expanded to study ubiquitylation, acetylation, and glycosylation and will be increasingly valuable as protein-localization data becomes available and further modifying enzymes are profiled [20-24].
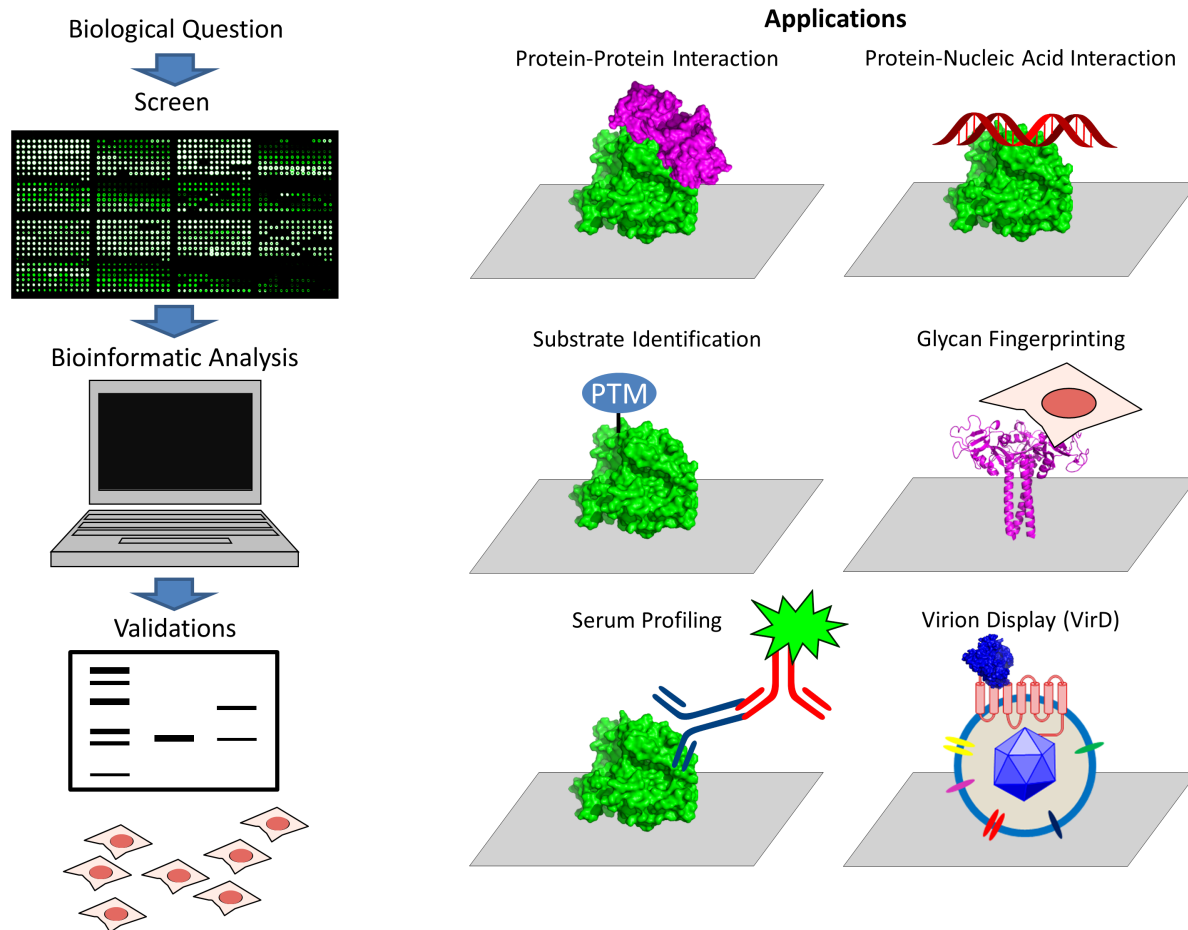
Figure 1-2 Applications of Protein Microarrays

Outline of scientific approach using functional protein microarrays. A protein microarray screen can be designed

to near-comprehensively interogate proteomes in an unbiased manner to answer biological questions. Results

from initial screen are processed using statistical analysis and orthogonal validations (Left panel). Screens on the

protein microarray can identify protein interactions, modifications, and proteins of interest for biomarker

discovery (Right panel).

1-4. Applications in Clinical Research

Biomarker Discovery

Biomarkers are detectable signatures that can serve as diagnostic tools in the clinical setting, and can facilitate vaccine and drug development in basic research. For instance, upon infection by a deleterious pathogen the immune system will respond and develop antibodies against a foreign agent. These antibodies become the hallmark of infection and allow for detection of the history of exposure to a pathogen. Researchers can view proteins on a functional protein microarray as antigens to accelerate biomarker discovery by identifying antibodies that statistically correlate with the disease of interest. Zhu and colleagues reported the first protein microarray-assisted serological study to identify Severe acute respiratory syndrome (SARS) infected patients in China and Canada [25]. Following this study, many new pathogen protein/antigen microarrays have been established, including malaria (Plasmodium), HIV-1, influenza, herpesviruses, and tuberculosis [26-34].

Biomarker identification is not limited to pathogens; diagnosis of many human diseases, especially autoimmune diseases and cancers, is challenging and often relies on pathological tests [35,36]. Hudson et al. used a functional protein microarray consisting of 5,005 human proteins to screen patient autoantibodies and discovered antigens useful for ovarian cancer diagnosis [37]. In 2009, Song et al. developed a cost-effective, streamlined two-stage screening approach for biomarker identification in autoimmune hepatitis (AIH) [4,38]. In the first stage, a human protein microarray with ~5,000 proteins was probed with a small cohort of patient sera to identify 11 candidate autoantigens. In the second stage, these candidates were printed in 12 duplicate blocks on the microarray surface to allow for groups of 12 patient serum samples from a larger cohort to be simultaneously profiled with a single array. Three of the candidates were confirmed as AIH biomarkers by bioinformatic analysis and validated with further experiments. These studies demonstrate the utility and versatility of functional

11

protein microarrays as a diagnostic and basic research tool in clinical applications and will continue to improve as the technology matures [39,40].

1-5. Conclusions

Functional protein microarrays, especially those of high-content, and their applications have evolved significantly since the first proteome microarray was introduced in 2001 by Zhu et. al [1]. These successes capitalize on femtomolar sensitivity for profiling full proteomes in high-throughput yet straightforward assays. We have described their utility for a myriad of applications that have resulted in impactful scientific findings, ranging from discovery of mCpG-dependent DNA-binding activity among human transcription factors to construction of kinase-substrate relationship networks, and biomarker discovery. While protein microarrays leverage the advantage of uniform protein sampling and high sensitivity, for proteomics, their impact is limited by the extent of coverage for different species. To this end, many research groups and companies have begun expanding the availability of high-content proteome arrays (Table 1). A remarkable advance was put forth by the Zhu laboratory with the construction of the first human proteome microarray containing over 17,000 full length proteins representing ~75% of the human proteome and is the largest available to date [15,39]. From a technological point of view, more reliable and higher density protein arrays are becoming more readily available (Table 1)

Despite these milestones, significant improvement can still be made. Transmembrane proteins, representing ~26% of the human proteome, are the preferred targets of nearly 70% of all drugs [41]. However, these proteins, especially those containing multipass transmembrane domains, are notoriously difficult to study because they must be embedded in a membrane to maintain their native conformation and many require proper post-translational modifications (PTMs) to be functional [42]. To overcome these hurdles, we recently developed Virion Display (VirD) technology through which

human transmembrane proteins could be assayed in their native conformation in microarray format. The unique advantages of this approach is that a human transmembrane protein is embedded in the viral envelop of Herpes Simplex Virus 1 (HSV-1) in a correct orientation with proper PTMs [42]. The VirD Array is expected to be readily transformed to a high-content platform that can display virtually all of the human membrane proteins close to their native conformation on a single glass slide.

The capabilities of microarray technology are further expanding with the development of label-free detection methods that are capable of monitoring the kinetics of biomolecular interactions in real time. By observing molecular reactions in real time, OIRD, SPR, and PC are capable of determining the kinetic parameters k_{on} and k_{off} [5,7,10]. In 2015 Liu et. al demonstrated the ability of the OIRD to simultaneously determine these parameters for multiple monoclonal antibodies [7].

Lastly, there is an increasing trend in the use of protein microarray technology for biomarker discovery. Human diseases leave a characteristic fingerprint on the proteome through alterations to expression and PTMs. Experiments performed by Woodard et al. have demonstrated that functional protein microarrays can identify differentially phosphorylated proteins using xenograft tumor and cell lysates [43]. Consequently, protein microarray arrays are poised to promote discovery of novel biomarkers and improve clinical diagnostics by providing access to the human proteome on a single chip.

References

1. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, et al.: Global analysis of protein activities using proteome chips. Science 2001, 293:2101-2105.

2. Haab BB, Dunham MJ, Brown PO: Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2001

3. Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, Emmert-Buck MR, Roth MJ, Petricoin IE, Liotta LA: Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 2001, 20:1981-1989.

4. Sutandy FX, Qian J, Chen CS, Zhu H: Overview of protein microarrays. Curr Protoc Protein Sci 2013, Chapter 27:Unit 27 21.

5. Liang W, Wang S, Festa F, Wiktor P, Wang W, Magee M, LaBaer J, Tao N: Measurement of small molecule binding kinetics on a protein microarray by plasmonic-based electrochemical impedance imaging. Anal Chem 2014, 86:9860-9865.

6. Landry JP, Sun YS, Guo XW, Zhu XD: Protein reactions with surface-bound molecular targets detected by oblique-incidence reflectivity difference microscopes. Appl Opt 2008, 47:3275-3288.

7. Liu S, Zhang H, Dai J, Hu S, Pino I, Eichinger DJ, Lyu H, Zhu H: Characterization of monoclonal antibody's binding kinetics using oblique-incidence reflectivity difference approach. MAbs 2015, 7:110-119.

8. Evans-Nguyen KM, Tao SC, Zhu H, Cotter RJ: Protein arrays on patterned porous gold substrates interrogated with mass spectrometry: detection of peptides in plasma. Anal Chem 2008, 80:1448-1458.

9. Zhang X, Zhu S, Deng C, Zhang X: An aptamer based on-plate microarray for high-throughput insulin detection by MALDI-TOF MS. Chem Commun (Camb) 2012, 48:2689-2691.

10. Shtenberg G, Massad-Ivanir N, Moscovitz O, Engin S, Sharon M, Fruk L, Segal E: Picking up the pieces: a generic porous Si biosensor for probing the proteolytic products of enzymes. Anal Chem 2013, 85:1951-1956.

11. Lee MR, Fauchet PM: Two-dimensional silicon photonic crystal based biosensing platform for protein detection. Opt Express 2007, 15:4530-4535.

12. Zhang JT, Cai Z, Kwak DH, Liu X, Asher SA: Two-dimensional photonic crystal sensors for visual detection of lectin concanavalin A. Anal Chem 2014, 86:9036-9041.

13. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J et al.: Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A 2014, 111:6131-6138.

14. Xie Z, Hu S, Blackshaw S, Zhu H, Qian J: hPDI: a database of experimental human protein-DNA interactions. Bioinformatics 2010, 26:287-289.

15. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C et al.: DNA methylation presents distinct binding sites for human transcription factors. Elife 2013, 2:e00726.

16. Siprashvili Z, Webster DE, Kretz M, Johnston D, Rinn JL, Chang HY, Khavari PA: Identification of proteins binding coding and non-coding human RNAs using protein microarrays. BMC Genomics 2012, 13:633.

17. Newman RH, Zhang J, Zhu H: Toward a systems-level view of dynamic phosphorylation networks. Front Genet 2014, 5:263.

18. Hu J, Rho HS, Newman RH, Zhang J, Zhu H, Qian J: PhosphoNetworks: a database for human phosphorylation networks. Bioinformatics 2014, 30:141-142.

19. Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, Tao SC, Qian J, Zhao Y, Boeke JD, et al.: Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. Cell 2009, 136:1073-1084.

20. Newman RH, Hu J, Rho HS, Xie Z, Woodard C, Neiswinger J, Cooper C, Shirley M, Clark HM, Hu S, et al.: Construction of human activity-based phosphorylation networks. Mol Syst Biol 2013, 9:655.

21. Li Y, Tao SC, Bova GS, Liu AY, Chan DW, Zhu H, Zhang H: Detection and verification of glycosylation patterns of glycoproteins from clinical specimens using lectin microarrays and lectin-based immunosorbent assays. Anal Chem 2011, 83:8509-8516.

22. Dias WB, Cheung WD, Hart GW: O-GlcNAcylation of kinases. Biochem Biophys Res Commun 2012, 422:224-228.

23. Lu JY, Lin YY, Qian J, Tao SC, Zhu J, Pickart C, Zhu H: Functional dissection of a HECT ubiquitin E3 ligase. Mol Cell Proteomics 2008, 7:35-45.

24. Li R, Wang L, Liao G, Guzzo CM, Matunis MJ, Zhu H, Hayward SD: SUMO binding by the Epstein-Barr virus protein kinase BGLF4 is crucial for BGLF4 function. J Virol 2012, 86:5412-5421.

25. Zhu J, Liao G, Shan L, Zhang J, Chen MR, Hayward GS, Hayward SD, Desai P, Zhu H: Protein array identification of substrates of the Epstein-Barr virus protein kinase BGLF4. J Virol 2009, 83:5219-5231.

26. Desbien AL, Van Hoeven N, Reed SJ, Casey AC, Laurance JD, Baldwin SL, Duthie MS, Reed SG, Carter D: Development of a high density hemagglutinin protein microarray to determine the breadth of influenza antibody responses. Biotechniques 2013, 54:345-348.

27. Deng J, Bi L, Zhou L, Guo SJ, Fleming J, Jiang HW, Zhou Y, Gu J, Zhong Q, Wang ZX, et al.: Mycobacterium tuberculosis proteome microarray for global studies of protein function and immunogenicity. Cell Rep 2014, 9:2317-2329.

28. Liang L, Juarez S, Nga TV, Dunstan S, Nakajima-Sasaki R, Davies DH, McSorley S, Baker S, Felgner PL: Immune profiling with a Salmonella Typhi antigen microarray identifies new diagnostic biomarkers of human typhoid. Sci Rep 2013, 3:1043.

29. Kalantari-Dehaghi M, Chun S, Chentoufi AA, Pablo J, Liang L, Dasgupta G, Molina DM, Jasinskas A, Nakajima-Sasaki R, Felgner J, et al.: Discovery of potential diagnostic and vaccine antigens in herpes simplex virus 1 and 2 by proteome-wide antibody profiling. J Virol 2012, 86:4328-4339.

30. Liang L, Felgner PL: A systems biology approach for diagnostic and vaccine antigen discovery in tropical infectious diseases. Curr Opin Infect Dis 2015, 28:438-445.

31. Gaze S, Driguez P, Pearson MS, Mendes T, Doolan DL, Trieu A, McManus DP, Gobert GN, Periago MV, Correa Oliveira R, et al.: An immunomics approach to schistosome antigen discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas. PLoS Pathog 2014, 10:e1004033.

32. Dotsey EY, Gorlani A, Ingale S, Achenbach CJ, Forthal DN, Felgner PL, Gach JS: A High Throughput Protein Microarray Approach to Classify HIV Monoclonal Antibodies and Variant Antigens. PLoS One 2015, 10:e0125581.

33. Reusken C, Mou H, Godeke GJ, van der Hoek L, Meyer B, Muller MA, Haagmans B, de Sousa R, Schuurman N, Dittmer U, et al.: Specific serology for emerging human coronaviruses by protein microarray. Euro Surveill 2013, 18:20441.

34. Helb DA, Tetteh KK, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H, Ssewanyana I, Kamya MR, Beeson JG, et al.: Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. Proc Natl Acad Sci USA 2015, 112:E4438-4447.

35. Santacatterina F, Chamorro M, de Arenas CN, Navarro C, Martin MA, Cuezva JM, Sanchez-Arago M: Quantitative analysis of proteins of metabolism by reverse phase protein microarrays identifies potential biomarkers of rare neuromuscular diseases. J Transl Med 2015, 13:65.

36. Hu B, Niu X, Cheng L, Yang LN, Li Q, Wang Y, Tao SC, Zhou SM: Discovering cancer biomarkers from clinical samples by protein microarrays. Proteomics Clin Appl 2015, 9:98-110.

37. Hudson ME, Pozdnyakova I, Haines K, Mor G, Snyder M: Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proc Natl Acad Sci USA 2007, 104:17494-17499.

38. Song Q, Liu G, Hu S, Zhang Y, Tao Y, Han Y, Zeng H, Huang W, Li F, Chen P, et al.: Novel autoimmune hepatitis-specific autoantigens identified using protein microarray technology. J Proteome Res 2010, 9:30-39.

39. Hu CJ, Song G, Huang W, Liu GZ, Deng CW, Zeng HP, Wang L, Zhang FC, Zhang X, Jeong JS, et al.: Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. Mol Cell Proteomics 2012, 11:669-680.

40. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, et al.: Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 2012, 148:1293-1307.

41. Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L: Prediction of the human membrane proteome. Proteomics 2010, 10:1141-1149.

42. Hu S, Feng Y, Henson B, Wang B, Huang X, Li M, Desai P, Zhu H: VirD: a virion display array for profiling functional membrane proteins. Anal Chem 2013, 85:8046-8054.

43. Woodard CL, Goodwin CR, Wan J, Xia S, Newman R, Hu J, Zhang J, Hayward SD, Qian J, Laterra J, et al.: Profiling the dynamics of a human phosphorylome reveals new components in HGF/c-Met signaling. PLoS One 2013, 8:e72671.

44. Tarrant MK, Rho HS, Xie Z, Jiang YL, Gross C, Culhane JC, Yan G, Qian J, Ichikawa Y, Matsuoka T, et al.: Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. Nat Chem Biol 2012, 8:262-269.

45. Jeong JS, Jiang L, Albino E, Marrero J, Rho HS, Hu J, Hu S, Vera C, Bayron-Poueymiroy D, Rivera-Pacheco ZA, et al.: Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. Mol Cell Proteomics 2012, 11:O111 016253.

46. Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA: Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics 2012, 11:M111 011593.

47. Uhlen M, Bjorling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al.: A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics 2005, 4:1920-1932.

48. Sjoberg R, Sundberg M, Gundberg A, Sivertsson A, Schwenk JM, Uhlen M, Nilsson P: Validation of affinity reagents using antigen microarrays. N Biotechnol 2012, 29:555-563.

49. Buus S, Rockberg J, Forsstrom B, Nilsson P, Uhlen M, Schafer-Nielsen C: High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. Mol Cell Proteomics 2012, 11:1790-1800.

50. Popescu SC, Popescu GV, Bachan S, Zhang Z, Seay M, Gerstein M, Snyder M, Dinesh-Kumar SP: Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. Proc Natl Acad Sci U S A 2007, 104:4730-4735.

51. Manohar M, Tian M, Moreau M, Park SW, Choi HW, Fei Z, Friso G, Asif M, Manosalva P, von Dahl CC, et al.: Identification of multiple salicylic acid-binding proteins using two high throughput screens. Front Plant Sci 2014, 5:777.

52. Lee HY, Bowen CH, Popescu GV, Kang HG, Kato N, Ma S, Dinesh-Kumar S, Snyder M, Popescu SC: Arabidopsis RTNLB1 and RTNLB2 Reticulon-like proteins regulate intracellular trafficking and activity of the FLS2 immune receptor. Plant Cell 2011, 23:3374-3391.

53. Takulapalli BR, Qiu J, Magee DM, Kahn P, Brunner A, Barker K, Means S, Miersch S, Bian X, Mendoza A, et al.: High density diffusion-free nanowell arrays. J Proteome Res 2012, 11:4382-4391.

54. Miersch S, Bian X, Wallstrom G, Sibani S, Logvinenko T, Wasserfall CH, Schatz D, Atkinson M, Qiu J, LaBaer J: Serological autoantibody profiling of type 1 diabetes by protein arrays. J Proteomics 2013, 94:486-496.

55. Yu X, Wallstrom G, Magee DM, Qiu J, Mendoza DE, Wang J, Bian X, Graves M, LaBaer J: Quantifying antibody binding on protein microarrays using microarray nonlinear calibration. Biotechniques 2013, 54:257-264.

56. Chen CS, Korobkova E, Chen H, Zhu J, Jian X, Tao SC, He C, Zhu H: A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. Nat Methods 2008, 5:69-74.

CHAPTER 2

NOVEL REGULATORY ROLE OF DNA METHYLATION IN GENE ACTIVATION

2-1. DNA Methylation

DNA methylation at the 5 position of the cytosine base (5mC) is the primary epigenetic modification on the mammalian genome. Methyl groups are covalently added at the 5 carbon of the cytosine ring, resulting in 5-methylcytosine (5mC). 4% of mammalian DNA is composed of 5-methylcytosine at CpG dinucleotides (CpGs) [1]. DNA methylation plays important roles in key biological processes, including genomic imprinting, X-chromosome inactivation, suppression of transposable elements, and carcinogenesis [2, 3]. As genome wide methylation profiles associated with specific diseases are increasingly becoming available, interpreting these often descriptive and correlative methylome datasets remains a big challenge.

In higher eukaryotes, methylation of CpG sites, especially at promoter regions, is generally considered as the hallmark of gene silencing [4]. The molecular consequence of CpG methylation is generally believed to disrupt TF–DNA interactions either directly [5], or indirectly by recruiting sequence-independent methylated DNA-binding proteins that occupy the methylated promoters and compete for the TF binding sites [6] **(Figure 2-1).**
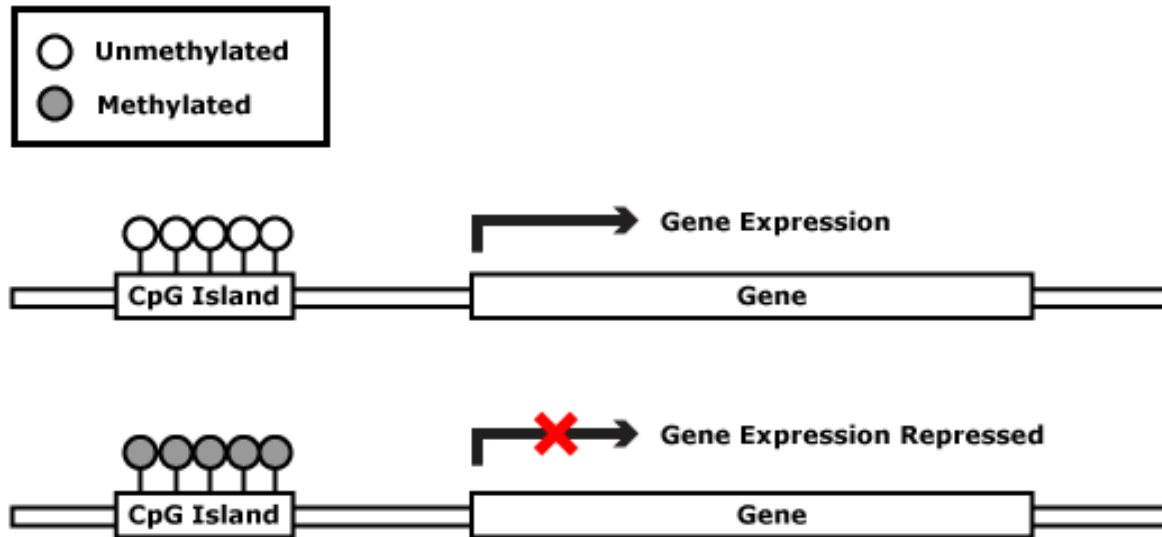
Figure 2-1. Traditional View of DNA methylation effect on gene expression
Many genes in the human genome have upstream CG-rich regions called CpG islands. DNA methylation of a

gene's CpG island represses gene expression. Different cell types have different methylation patterns, which

contributes to the differences in gene expression in different cell types. (Adapted from USCF SOM

missinglink.edu).

So far, only MeCP2, MBD1, MBD2, and a few zinc finger proteins have been identified as bonafide

methylated DNA-binding proteins [7]. It is unclear whether the methylated DNA binding activity is

widespread among different TF subfamilies. Furthermore, the transcriptional regulatory activity of

these methylation-dependent TF–DNA interactions has not been explored. Finally, the structural basis

of these methylation-dependent TF–DNA interactions remains elusive.

Transcription is regulated by the interaction of DNA cis-regulatory sequences and trans-acting factors.

High throughput approaches have recently been developed to study novel protein-DNA interactions

across genomes. This has created new frontiers as well as challenges in understanding the control of

gene expression. The Zhu lab has previously developed a large-scale, unbiased method to characterize

Protein-DNA Interactions (PDIs). Using a protein microarray containing the human repertoire of transcription factors (TFs), which includes 1321 TFs, and 210 co-factor proteins, over 40 TFs and 6 TF co-factors were found to bind preferentially to mCpG-carrying DNA motifs (Hu et al, 2013) [7]. Each TF chip was probed with one of 154 methylated DNA motifs in the presence of 10-fold excess of the unlabeled counterpart. Among the 47 DNA binding proteins (DBPs) that showed binding activity to mCpG, 15% demonstrate broad binding activity to over 50% of all methylated motifs. One transcription factor that showed methylation specific binding is KLF4 [7].

2-2. Krupple Like Factors (KLFs)

The krüppel-like factors (KLFs) compose a family of transcription factors that bind to GC rich regions of DNA to activate or repress transcription. The first human KLF was identified by Page et al. in 1987. The first mammalian klf gene was cloned soon after in 1993 [8]. Since then, 16 other mammalian KLFs have been identified and are designated KLF1-KLF17 [9]. KLFs have three highly conserved consecutive Cys2-His2 zinc fingers located near the carboxyl terminus of the protein that enables them to interact with GC-rich sequences in target gene promoters and enhancers **(Figure 2-2)** [10]. However, they have unique amino-terminal sequences that permit the transcription factors to interact with specific binding partners. Through transcriptional regulation of a variety of genes, KLFs regulate numerous cellular processes including cell growth, proliferation and apoptosis [11]. The functions of the 17 KLF family members are in some cases overlapping and in others widely divergent. As transcriptional regulators, KLFs control essential cellular processes such as proliferation, differentiation and migration [11]. More recently, KLFs have also been implicated in the establishment and maintenance of pluripotency [12, 13]. Along with these roles for KLFs in normal cells and tissues, important tumor suppressive and oncogenic functions have been defined for KLFs in cancer [11].

In general, members of the KLF family can be grouped on the basis of structural or functional relationships. Moreover, some members of the family are ubiquitously expressed, whereas others are tissue restricted, leading to the possibility of both exclusive and redundant functions for each KLF. Deletion of KLF2, KLF5 or KLF6, for example, is lethal in mice, which is indicative of non-redundant functions during development, whereas several other KLF-knockout mice are viable, which suggests functional compensation by other factors [14].
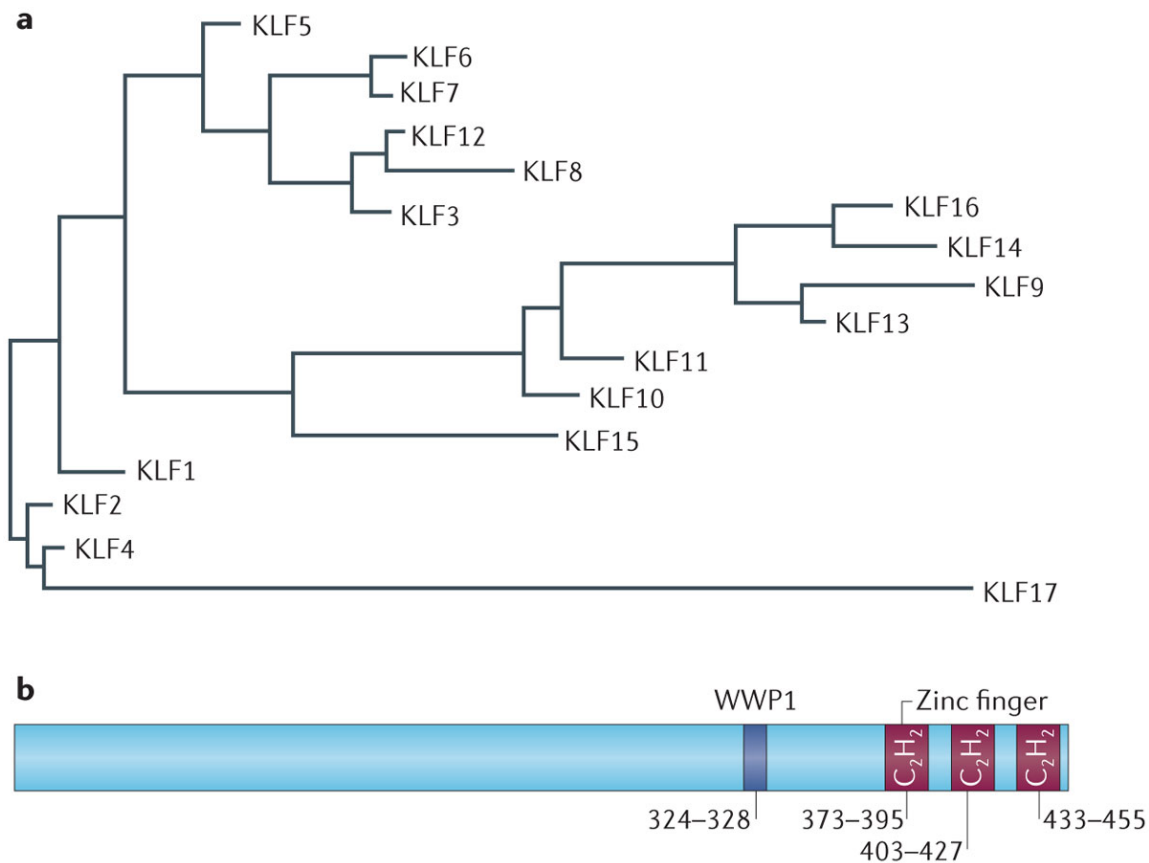
Figure 2-2. Phylogenetic tree of the human Kruppel-like factors (KLFs)

(A).Phylogenetic tree of the human Krüppel-like factors (KLFs). Currently, there are 17 known KLFs in

mammals. (B). The defining feature of the KLF family is the carboxy-terminal DNA-binding domain, which

consists of three zinc fingers that bind GC-rich DNA sequences. The rest of the amino acid sequence is

divergent, and each KLF typically contains at least one transactivation or transrepression domain. KLF5, for

example, as shown in the figure, consists of 457 amino acids with three zinc fingers at the C terminus and a

transactivation domain surrounding a proline-rich motif (amino acids 324–328), which binds the E3 ubiquitin

ligase WWP1.

2-3. Krupple Like Factor 4 (KLF4)

KLF4 was initially identified as a growth arrest-associated gene in the intestinal epithelium that

suppresses DNA synthesis when expressed ectopically. KLF4 is a versatile transcription factor that

regulates numerous cellular processes [15]. KLF4-mediated gene transactivation or trans-repression is

regulated on multiple levels by modulating KLF4's status through phosphorylation, acetylation,

methylation, and ubiquitination in a context-dependent manner [14]. KLF4 is one of four transcription

factors rigorously described by Yamanaka that, when ectopically expressed, reprogram somatic cells to

pluripotent stem cells [16, 17]. Depletion of KLF4 has also been shown to result in embryonic stem

cell differentiation [12]. KLF4 can function as both an oncogene and a tumor suppressor by regulating

cell cycle promoting genes and genes involved in apoptosis depending on the cellular context. The zinc

finger TF KLF4 is a cancer driver gene shown to be upregulated in glioblatomas [18, 19].
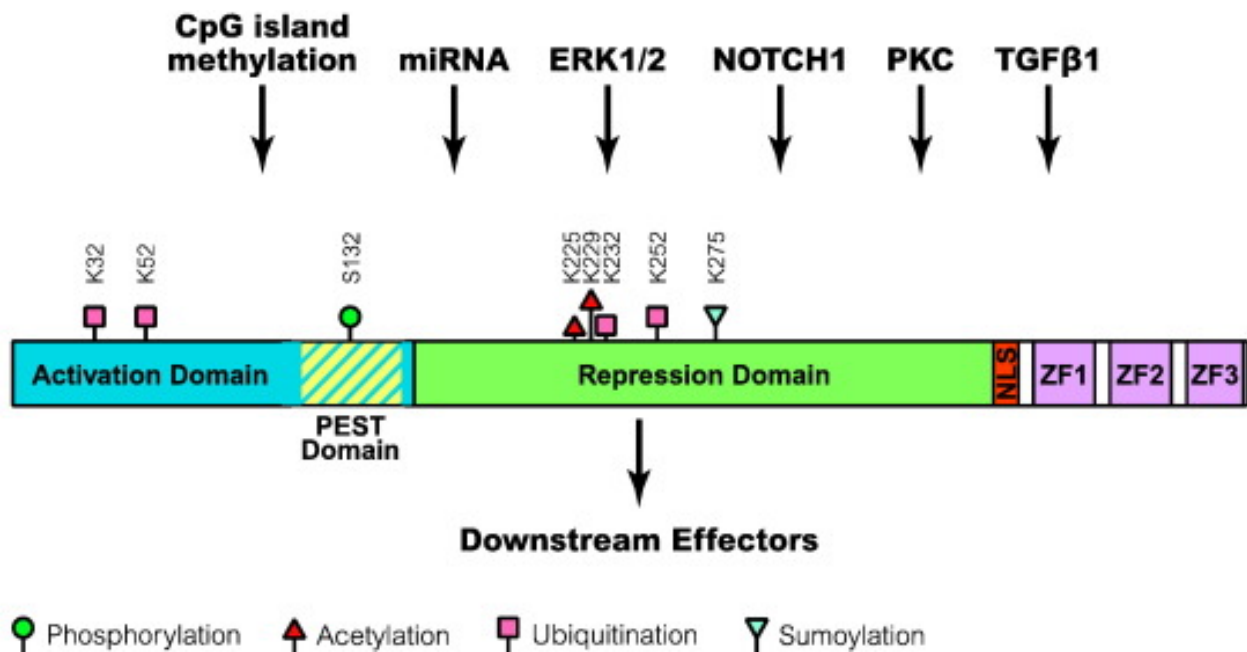
Figure 2-3. Schematic of KLF4 gene

KLF4 belongs to the family of SP/KLF factors that are characterized by three zinc finger motifs within their carboxyl terminal sequences. Within its amino terminus, KLF4 possess a transactivation domain (TAD) and adjacent to it, a repression domain, together of which determine the specificity of KLF4's transcriptional regulating activity by interacting with other factors and modulating DNA binding efficiency. Two nuclear localization signals (NLS) have been identified in mouse KLF4. The first is directly adjacent to the most amino terminal zinc finger motif and the second spans the first and half of the second zinc finger domain. Mouse KLF4 contains 483 amino acids with a predicted molecular weight of 53 kDa, and is 91% identical to human KLF4. Several splice variants of the human KLF4 gene have been identified in normal and cancer cells. (Adapted from: Amr M. Ghaleb, Vincent W. Yang. Krüppel-like factor 4 (KLF4): What we currently know)

Mutational and structural analysis from Hu et al 2013 [7] identified R458 as the critical residue for

DNA methylation binding and recognition. Mutating R458 to alanine abolished KLF4's ability to bind

to methylated DNA **(Figure 2-5).**
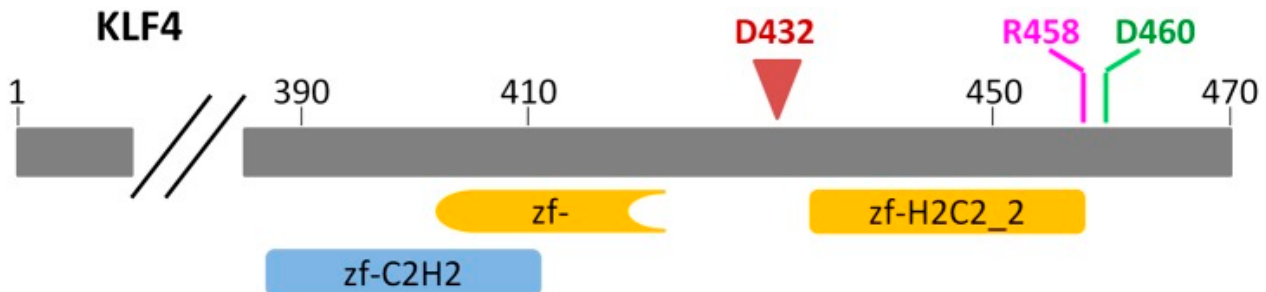
Figure 2-4. KLF4's mCpG-dependent binding activity is decoupled from its binding activity to unmethylated motifs
 KLF4 encodes three zinc finger DNA-binding domains at its C-terminus. Residues R458 and D460, which were

predicted to interact with the 5-methyl group in the cytosine, are located in the zf-H2C2 domain. (Adapted from

Hu et al 2013; DNA methylation presents distinct binding sites for human transcription factors [7])

References

1. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.

2. Feinberg, A.P., *Phenotypic plasticity and the epigenetics of human disease.*

3. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.*

4. Baylin, S.B., *DNA methylation and gene silencing in cancer.* Nat Clin Pract Oncol, 2005. **2 Suppl 1**: p. S4-11.

5. Nan, X., S. Cross, and A. Bird, *Gene silencing by methyl-CpG-binding proteins.* Novartis Found Symp, 1998. **214**: p. 6-16; discussion 16-21, 46-50.

6. Boyes, J. and A. Bird, *DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein.* Cell, 1991. **64**(6): p. 1123-34.

7. Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* eLife, 2013. **2**: p. e00726.

8. Miller, I.J. and J.J. Bieker, *A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins.* Mol Cell Biol, 1993. **13**(5): p. 2776-86.

9. Suske, G., E. Bruford, and S. Philipsen, *Mammalian SP/KLF transcription factors: bring in the family.* Genomics, 2005. **85**(5): p. 551-6.

10. Pearson, R., et al., *Kruppel-like transcription factors: a functional family.* Int J Biochem Cell Biol, 2008. **40**(10): p. 1996-2001.

11. McConnell, B.B. and V.W. Yang, *Mammalian Kruppel-like factors in health and diseases.* Physiol Rev, 2010. **90**(4): p. 1337-81.

12. Jiang, J., et al., *A core Klf circuitry regulates self-renewal of embryonic stem cells.* Nat Cell Biol, 2008. **10**(3): p. 353-360.

13. Yamanaka, S., *Strategies and new developments in the generation of patient-specific pluripotent stem cells.* Cell Stem Cell, 2007. **1**(1): p. 39-49.

14. Tetreault, M.-P., Y. Yang, and J.P. Katz, *Kruppel-like factors in cancer.* Nat Rev Cancer, 2013. **13**(10): p. 701-713.

15. Ghaleb, A.M. and V.W. Yang, *Krüppel-like factor 4 (KLF4): What we currently know.* Gene, 2017. **611**: p. 27-37.

16. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* (0092-8674 (Print)).

17. Wernig, M., et al., *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.* (1476-4687 (Electronic)).

18. Elsir, T., et al., *A study of embryonic stem cell-related proteins in human astrocytomas: Identification of Nanog as a predictor of survival.* International Journal of Cancer, 2014. **134**(5): p. 1123-1131.

19. Holmberg, J., et al., *Activation of Neural and Pluripotent Stem Cell Signatures Correlates with Increased Malignancy in Human Glioma.* Plos One, 2011. **6**(3).

CHAPTER 3

TARGETING UDP-α-D-GLUCOSE 6-DEHYDROGENASE INHIBITS GLIOBLASTOMA GROWTH

3-1. Introduction

KLF4 plays multiple roles in normal physiology and disease. It is one of the Yamanaka factors that induce pluripotency in somatic cells [1]. KLF4 also functions as a cancer driver gene [2], and is involved in cancer stem cell maintenance [3-5]. KLF4 has also been shown to be upregulated in high-grade brain tumors [6, 7], such as glioblastoma (GBM) [8, 9]. In addition to driving tumor malignancy, KLF4 can act as a tumor suppressor in distinct cellular contexts [10-12].

Studies by Wan et al, dissected the biological function of KLF4 binding to methylated DNA in malignant brain tumor cells by taking advantage of the R458A mutant lacking the ability to bind to methylated DNA. We showed that KLF4-mCpG interaction promotes brain tumor cell migration via the transactivation of genes involved in cell motility pathways [13].

Glioblastoma (GBM, Grade IV glioma) is one of the most devastating forms of cancer and characterized by highly proliferative tumor growth and intensive tumor cell infiltration into normal brain tissues [14, 15]. An increased understanding of the molecular mechanisms underlying the aggressive behavior of tumor cells and the microenvironment in which they invade could provide insights into novel treatment strategies for this deadly disease. The extracellular matrix (ECM) is one of the critical components of the tumor microenvironment and provides essential biochemical and mechanical cues that direct cell growth, survival, migration and differentiation [16, 17]. Cell adhesion to the ECM permits growth factor-dependent activation of oncogenic signals, which promotes cell cycle progression and cell proliferation, while also functioning as either a barrier or a movement track to inhibit or promote cell migration [18]. The ECM is mainly composed of fibrous proteins (e.g., collagen) and gel-like substance, such as glycosaminoglycans (GAGs), which are long polysaccharide chains with negative charges that attract water and soluble molecules including growth factors [19].

31

GAGs are synthesized by an enzyme called UDP-glucose 6-dehydrogenase (UGDH). In our prior work, we found that krᴜ̈ppel-like factor 4 (KLF4) binds to methylated CpGs (mCpG) in *cis*-regulatory elements, and activates gene expression, including *UGDH*, in GBM cells to exert phenotypic changes such as increased cell migration [13]. DNA methylation at *cis*-regulatory regions, mostly occurring at the CpG dinucleotide sites, is linked to gene repression and more recently gene activation [13, 20]. Given the importance of extracellular matrix function, specifically GAGs, in GBM; and the fact that UGDH is implicated as a rate-limiting and essential step in GAG monosaccharide synthesis [21-23], the mCpG-dependent activation of *UGDH* prompts us to investigate the biological function of UGDH in GBM.

GAG formation is part of glucose metabolism: glucose is converted to glucose-1-phosphate then to UDP-glucose (UDP-Glu), an active form of glucose, which is further converted to UDP-glucuronic acid (UDP-GlcA). UDP-GlcA is the indispensable precursor for the synthesis of GAGs. The enzyme UDP-glucose 6-dehydrogenase (UGDH) catalyzes the biosynthetic oxidation of UDP-glucose to UDP-glucuronic acid [21, 23], which are the building blocks of GAGs including hyaluronic acid and proteoglycans such as brevican, versican, aggregan etc. GAG synthesis pathways and key players are shown in **Figure 3-1.**
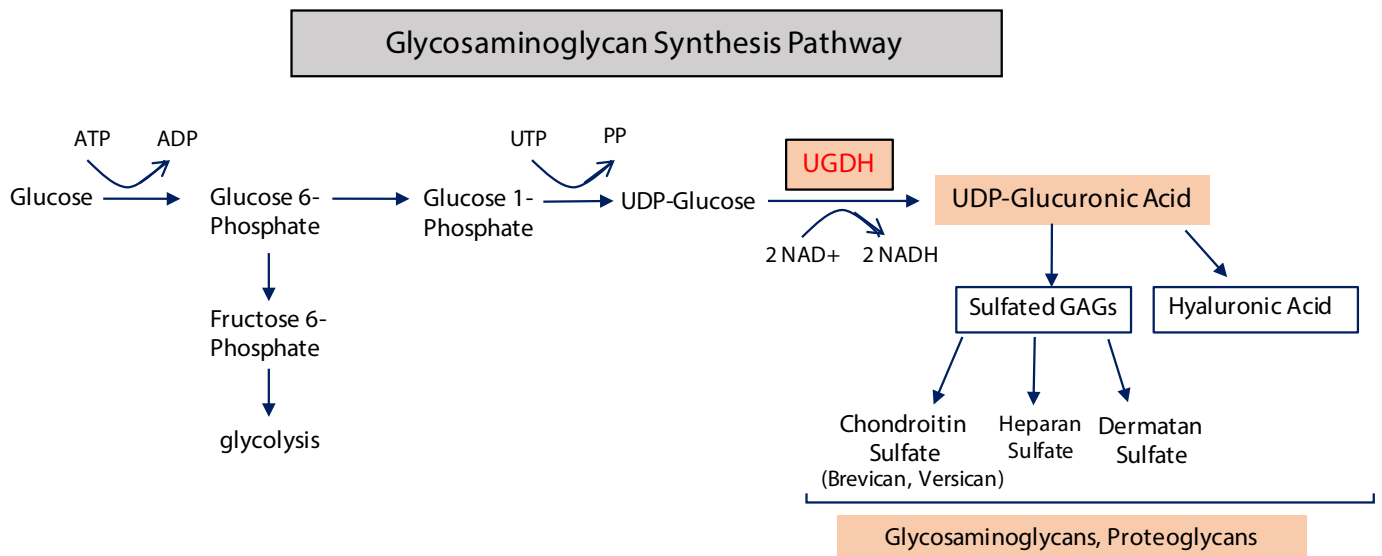
Figure 3-1. Schematic Illustration of GAG synthesis pathway, different GAGs and UGDH function in GAG Synthesis

GAGs reside in the extracellular space providing structural support for cells, as well as promoting cell adhesion, motility, angiogenesis and wound healing [24, 25]. Elevated GAG formation is implicated in a variety of human diseases, including the progression of epithelium tumors, breast cancers and brain tumors [19, 26]. Although GAGs are shown to be implicated in tumor progression, decrease in the synthesis of GAG precursor UDP-glucuronic acid in GBM biology has not been investigated. In this current work, we investigated the methylation-dependent regulation of UGDH, as well as the biological function of UGDH in GBM cells. These findings identify UGDH as a potential therapeutic target for GBMs.

3-2. KLF4-mCpG interactions activate genes involved in migration

In our previous studies, we firmly established that KLF4 could activate transcription of >100 genes via a non-canonical binding activity to methylated CpG (mCpG) to promote GBM cell migration [20]. We observed that one third of such genes are involved in cytoskeletal organization, extracellular matrix formation and cell migration **(Figure 3-2).** To further interrogate the biological function of KLF4-mCpG binding activity, two previously characterized tet-on inducible human glioblastoma (GBM) U87 cell lines were used: one expressing KLF4 wild type (KLF4 WT) and the other KLF4 site-specific mutant (KLF4 R458A), which lacks the mCpG-dependent binding activity but retains KLF4 canonical binding activity [13]. Real-time PCR (RT-PCR) and ChIP-PCR were utilized to examine the expression of selected KLF4 downstream target genes in GBM cells. Genes activated by KLF4 WT but not by KLF4 R458A would be recognized as a putative KLF4-mCpG target.

PANTHER Protein Class
Total # Genes: 59   Total # protein class hits: 63
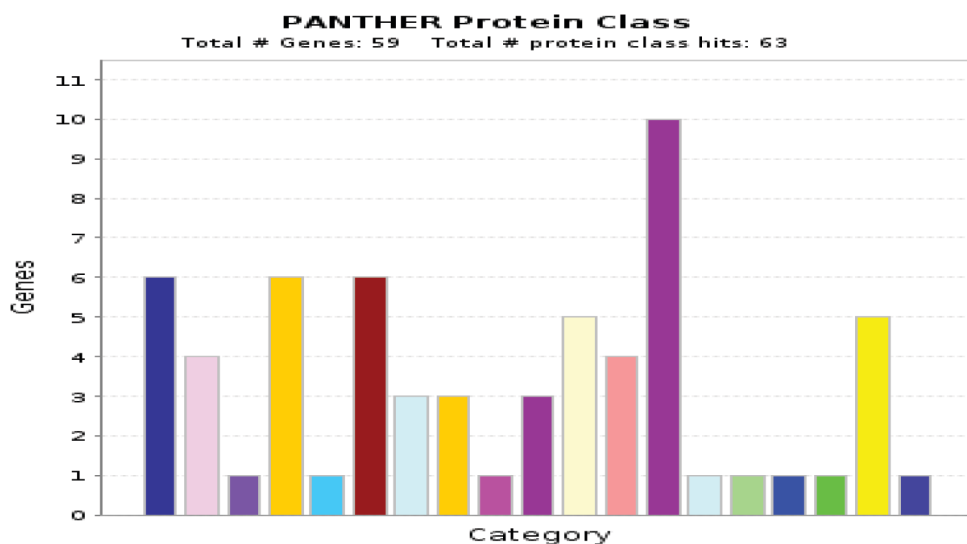
■ calcium-binding protein (PC00060)
■ cell adhesion molecule (PC00069)
■ chaperone (PC00072)
■ cytoskeletal protein (PC00085)
■ defense/immunity protein (PC00090)
■ enzyme modulator (PC00095)
■ extracellular matrix protein (PC00102)
■ hydrolase (PC00121)
■ membrane traffic protein (PC00150)
■ nucleic acid binding (PC00171)
■ oxidoreductase (PC00176)
■ receptor (PC00197)
■ signaling molecule (PC00207)
■ storage protein (PC00210)
■ structural protein (PC00211)
■ transcription factor (PC00218)
■ transfer/carrier protein (PC00219)
■ transferase (PC00220)
■ transporter (PC00227)

Figure 3-2. Pathway analysis by Panther of KLF4-mCpG direct targets
Pathways analysis in GBM cells showed that many genes are involved in cytoskeletal organization, extracellular

matrix formation, cell adhesion and migration

Twenty genes from our RNA-sequencing analysis [13] were selected for further studies by RT-PCR based on biological function, antibody availability, and association with GBM malignancy. Gene expression was quantified in tet-on inducible U87 KLF4 WT cells +/- Dox, and tet-on inducible U87 KLF4 R458A cells +/- Dox (**Figure 3-3A**). Our results demonstrated that KLF4 WT and KLF4 R458A differently induced gene expression. Specifically, eleven of the twenty genes, such as neuronal guanine nucleotide exchange factor (*NGEF),* pleckstrin homology like domain family B member 2 (*PHLDB2),* RAB guanine nucleotide exchange factor (GEF) 1 (*RABGEF1),* and UDP-α-D-glucose 6-dehydrogenase (*UGDH),* were highly induced by KLF4 WT, but not by KLF4 R458A at the mRNA level (**Figure 3-3A**), indicating that activation of these genes was dependent on mCpG-dependent binding of KLF4. To further confirm that the gene activation occurred via KLF4 binding to mCpGs in *cis*-regulatory regions, we performed bisulfite sequencing to examine the methylation status of the KLF4-binding regions of these gene based on our previous ChIP-seq data [13]. We found that nine of the eleven KLF4 WT-upregulated genes were associated with highly methylated regions, and some examples are shown in **Figure 3-3B**.

We next performed ChIP-PCR to examine binding activity of KLF4 WT and KLF4 R458A to the *cis*-regulatory regions of a subset of the nine genes, including *PHLDB2, RABGEF1, NGEF* and *UGDH*. An anti-KLF4 antibody that recognizes both KLF4 WT and KLF4 R458A was used to ChIP the chromatins in KLF4 WT and KLF4 R458A expressing cells 48-hrs post Dox induction [13]. The resulting ChIP-PCR data demonstrated that KLF4 WT preferentially bound to the *cis*-regulatory regions of these upregulated genes, whereas KLF4 R458A showed much weaker or no binding signals to the same region (**Figure 3-3C**). In addition, KLF4 WT increased protein expression of PHLDB2, RABGEF1, and UGDH in comparison to KLF4 R458A (**Figure 3-3D**), corroborating that these genes could be functional KLF4-mCp direct targets.
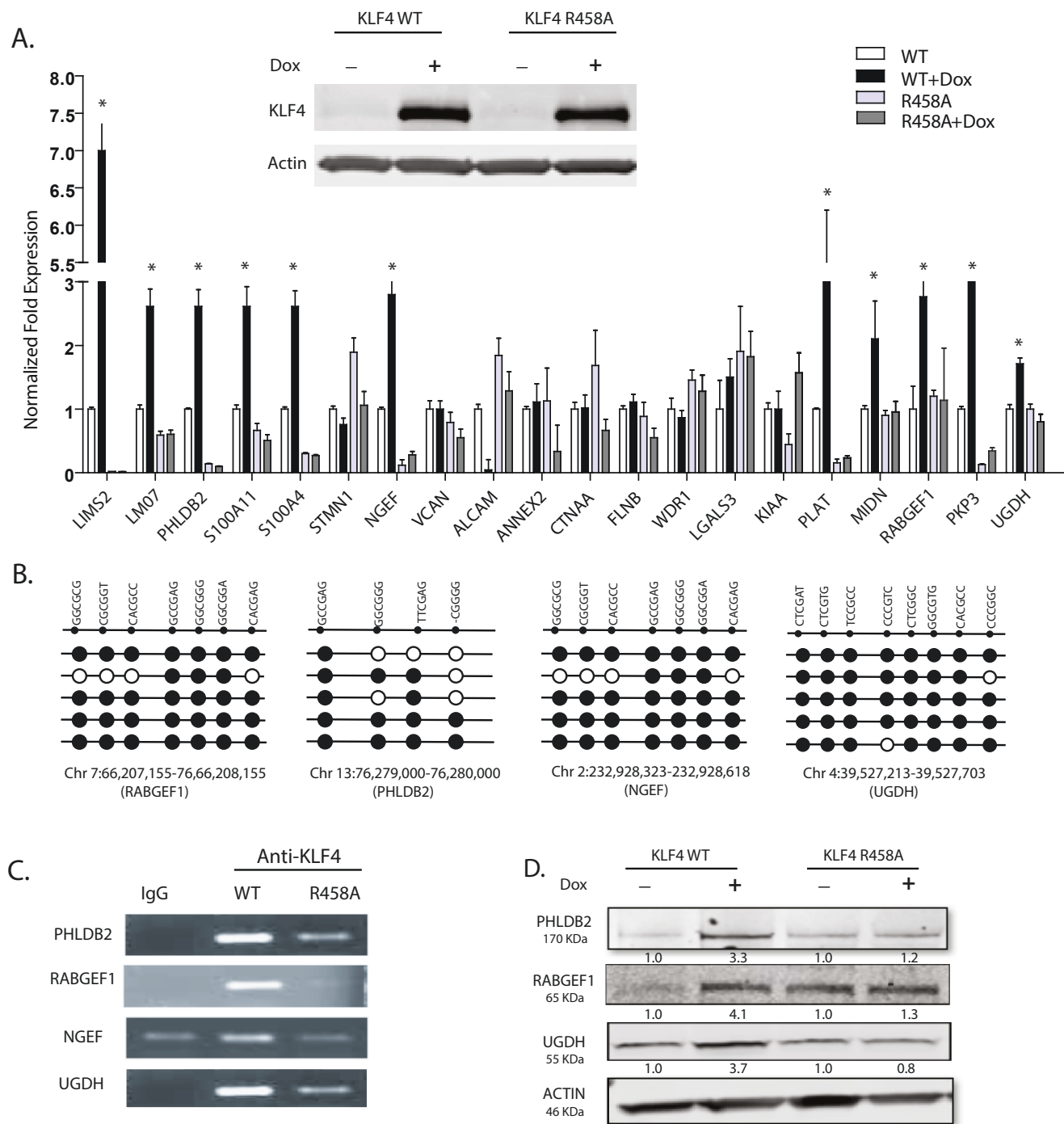
Figure 3-3. KLF4-mCpG binding activity activates genes involved in GBM cytoskeletal organization and migration

(A) Upper panel: Western blot analysis showed KLF4 WT or KLF4 R458A expression in tet-on stable U87 GBM cell lines upon doxycyline treatment (1 ug/ml, 48 hrs). Lower panel: Twenty putative KLF4-mCpG gene targets involved in cell migration pathway were picked from our previous RNA-seq studies. Real time-PCR (RT-PCR) revealed 11 of the 20 genes were significantly upregulated by KLF4 WT only, with no change in KLF4 R458A expressing cells (+Dox, 48 hr), confirming a mCpG-dependent gene activation mechanism). (B) Sanger bisulfite sequencing indicated DNA methylation in tested cis-regulatory regions of putative KLF4-mCpG targets. Examples of four genes showed highly methylated KLF4 binding regions of these genes. Each row represents one sequenced clone; each column represents one CpG site; filled circles stand for methylation. (C) Confirmation that KLF4 WT but not KLF4 R458R preferentially bound to the methylated cis-regulatory regions of selected genes. A KLF4 antibody was used to precipitate cross-linked genomic DNA from U87 cells expressing KLF4 WT or KLF4 R458A. Rabbit IgG was used to control for non-specific binding. De-crosslinked DNA samples were served as the input for ChIP-PCR. KLF4 binding to the selected regions was enriched in KLF4 WT expressing cells. (D) Western blot analysis indicating increased protein expression of the selected targets by KLF4 WT, but not by KLF4 R458A.

3-3. UGDH correlates with KLF4 expression in GBMs and is regulated by KLF4 via a DNA methylation-dependent mechanism

We decided to focus on understanding the biological function of UGDH in GBM as it is an important enzyme for the synthesis of extracellular matrix components. We first queried the TCGA database to assess gene expression in tumor and non-tumor specimens, and to determine whether selected genes correlate with KLF4 expression and patient survival in GBM. *UGDH* was found to be upregulated in GBM samples when compared with normal brains (**Figure 3-4A**). *UGDH* expression also moderately correlated with KLF4 expression in GBM samples (R= 0.32, *P*<0.001) (**Figure 3-4B**). Furthermore, we found that high expression of *UGDH* is associated with poor survival in GBM patients (**Figure 3-4C)**

With evidence supporting *UGDH* was upregulated by KLF4 WT only, but not the KLF4 R458A, we determined whether *UGDH* expression is activated by KLF4 via a methylation-dependent mechanism. We tested the effect of 5-aza-2′-deoxycytidine (5-Aza), a potent inhibitor of DNA methyltransferase, on KLF4 binding to the *UGDH cis*-regulatory region, and on *UGDH* expression. Cells treated with 5-Aza (1µmol/L) for 2 weeks had 80% fewer methylated sites in *UGDH cis*-regulatory regions (**Figure 3-4D**). Consistent with the reduction of mCpG sites, we observed a complete loss of KLF4 WT binding to the *UGDH cis*-regulatory region in the presence of 5-Aza (**Figure 3-4E**). Furthermore, 5-Aza abrogated KLF4 WT-induced *UGDH* upregulation by nearly 60% at both the mRNA and protein levels (**Figure 3-4F, G**). These results demonstrate that KLF4 activates *UGDH* transcription in a DNA methylation-dependent manner.
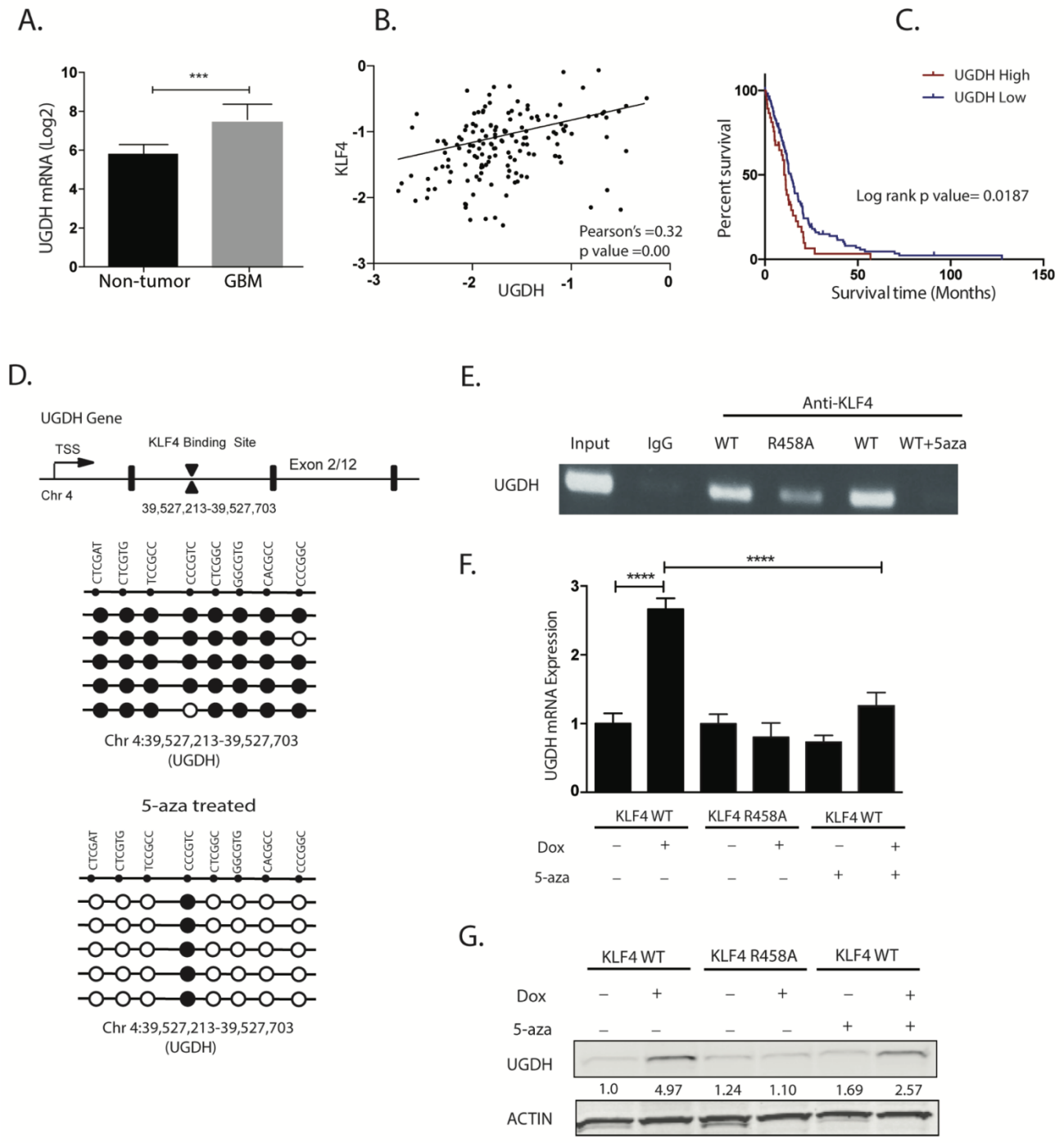
Figure 3-4. UGDH correlates with KLF4 expression in GBM samples and is regulated via a mCpG dependent mechanism by KLF4

(A) UGDH gene expression intensity (mean) from REMBRANDT database. UGDH expression is significantly up-regulated in GBM samples when compared with non-neoplastic brain samples (P<0.001). (B) Pearson's correlation plots of UGDH and KLF4 expression in glioma patients of mesenchymal subtype. Plots show positive correlation in the expression of UGDH and KLF4 (R=0.32, P < 0.001). (C) Kaplan-Meier Survival Plots for glioma patients based on differential gene expression of UGDH (from REMBRANT database). P<0.05, UGDH high n=38, median survival =10.4 months; UGDH low n=118, median survival =13.9 months). The probability of survival is significantly lower in samples with high UGDH gene expression compared to samples with low expression in GBM mesenchymal subtype (P < 0.001). (D) Upper panel: schematic of UGDH gene structure showing KLF4 binding site on the first intron. Lower panels: sanger bisulfite sequencing of DNA methylation on KLF4 binding site of UGDH gene before and after treatment with DNA methyltransferase inhibitor 5-Aza. Cells pretreated with 5-aza showed 80% reduction in DNA methylation. (E) ChIP-PCR showed KLF4 WT preferentially bound to the methylated cis-regulatory region of UGDH; 5-aza treatment decreased KLF4 WT binding to UGDH. (F) RT-PCR showed UGDH mRNA was significantly induced by KLF4 WT only, which was abrogated by 5-aza. (G) Western blot showed induction of UGDH by KLF4 WT only. Consistent with ChIP-PCR and RT-PCR, 5-aza treatment partially blocked UGDH induction by KLF4 WT, in keeping with a methylation-dependent mechanism.

3-4. *UGDH* knockdown decreases GAG abundance and cell migration

Given the importance of extracellular matrix on GBM progression and the relatively unknown functional significance of UGDH in GBM, we focused our efforts on determining the biological function of UGDH in GBM cells. We hypothesized that knocking down UGDH would inhibit tumor cell migration and proliferation, because UGDH catalyzes the reaction that generates the key precursor for glycosaminoglycans, building blocks for extracellular matrix components [23]. Two GBM model systems, U87 cells and GBM neurosphere cells HSR-GBM1A (GBM1A), were used to conduct UGDH loss-of-function studies. Each cell line was transfected with lentivirus containing either one of

the two distinct *UGDH* shRNAs (sh#1, sh#2). Both shRNAs induced a significant reduction (80%-85%) in UGDH protein in U87 (**Figure 3-5A**, left panel) and GBM1A cells (**Figure 3-5A,** right panel). We tested the level of the end products of UGDH, GAGs, by utilizing a polysaccharide binding dye 1,9 dimethyl-methylene blue (DMMB). There was a substantial reduction in the abundance of cell-associated GAGs in UGDH knockdown cells when compared to cells transfected with control shRNAs (e.g., 48% by sh#1, *P*<0.001; and 79% by sh#2 in U87 cells, *P*<0.001) (**Figure 3-5B;** left panel). Similar effects were observed in GBM 1A cells (**Figure 3-5B;** right panel). To examine the effect of UGDH knockdown on GBM cell migration, transwell and wound healing scratch migration assays were performed in 0.1% serum medium to minimize cell proliferation. Transwell assays 24 hrs after cell plating revealed that UGDH silencing significantly decreased the migratory ability of GBM cells (by ~50% in U87, *P*<0.01, and 30% in GBM1A cells, *P*<0.05) (**Figure 3-5C,D)**. UGDH knockdown also decreased cell motility in wound healing assays by 25% (*P*<0.05) when compared to cells transfected with control shRNAs after 24 hrs of scratch, and by 28% after 48 hrs (*P*<0.05) (**Figure 3-5E,F)**. Supplementing the culture medium with one of the GAGs, hyaluronic acid (HA, 100 ng/ml) rescued migration inhibition by *UGDH* shRNAs (**Figure 3-5G**), supporting the hypothesis that the effect of silencing UGDH on cell migration response is GAG-dependent. To substantiate the notion that the decreased cell migration in UGDH knockdown cells was not due to reduced cell growth under the given conditions and time points, we quantified the total cell number at 48 hrs under 0.1% FBS, and did not find any significant difference in cell growth (115,000 ± 3000 vs 113,000 ±2000, **Figure 3-6A,B).**

A.

U87

GBM 1A

UGDH shRNA
Control   #1   #2

UGDH
1.0   0.08   0.07

ACTIN

UGDH shRNA
Control   #1   #2

UGDH
1.0   0.1   0.1

ACTIN

B.

%sGAG

Control   UGDH sh#1   UGDH sh#2

**
***

%sGAG

Control   UGDH sh#1   UGDH sh#2

**
***

C.

# of Migrated Cells

Control   UGDH sh#1   UGDH sh#2

***
**

# of Migrated Cells

Control   UGDH sh#1   UGDH sh#2

*
*

D.

Control   UGDH sh#1   UGDH sh#2
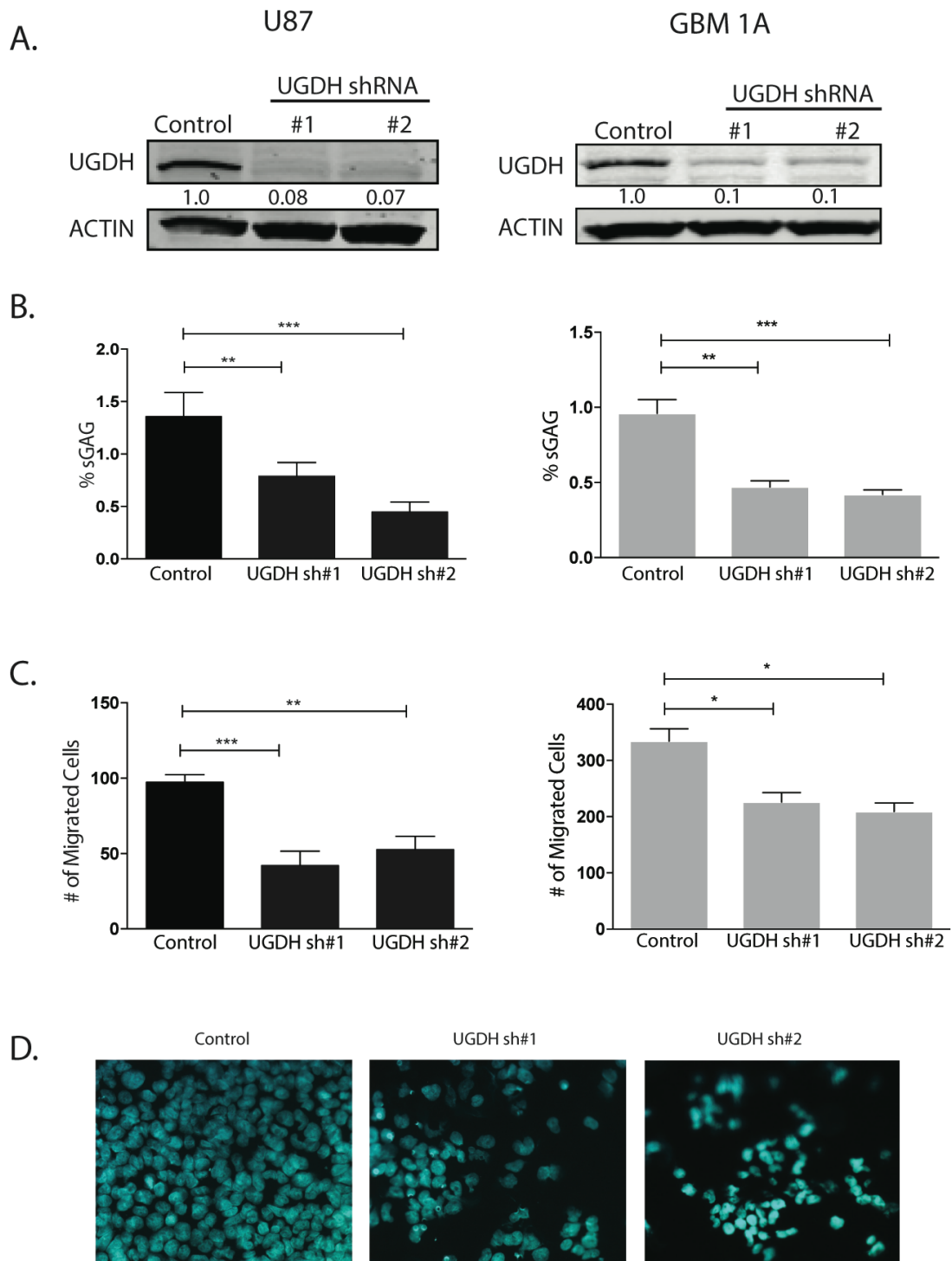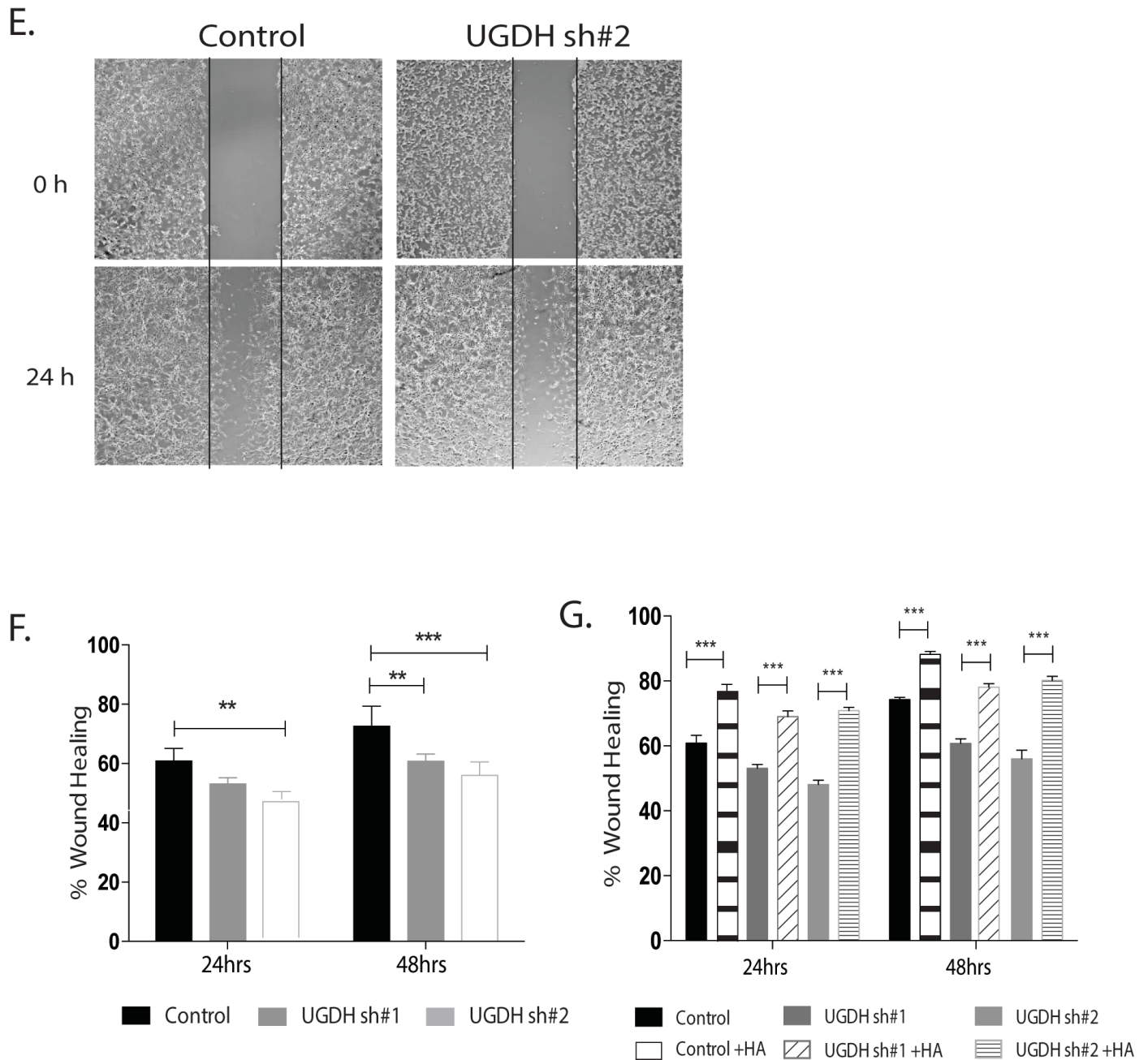
Figure 3-5. UGDH knockdown decreases GAG abundance and cell migration
(A) U87 and HSR-GBM1A (GBM1A) cells were infected with lentivirus coding for control shRNA, UGDH

shRNA #1 or UGDH shRNA #2. UGDH knockdown was quantified by immunoblot analysis. Both UGDH

shRNAs decreased UGDH protein level by more than 80%. (B) Sulfated GAG (sGAG) concentration was

quantified by DMMB assay. UGDH knockdown showed significant reduction in GAG concentration in both

U87 (Left panel) and GBM1A cells (right panel). (C) Transwell migration assays showed significant decrease in

migration of UGDH knockdown U87 cells (left panel) and GBM1A cells (right panel). (D). Microphotographs

of transwell analysis of U87 cell migratory ability of control and UGDH silenced cells. After 24 hrs, cells

migrated on the other side on the membrane were stained with DAPI and counted 5 fields/well. (E) Wound

healing scratch assays showed UGDH knockdown decreased migration in U87 cells. (F) Quantification of

wound healing scratch assays showed UGDH knockdown decreased migration in U87 cells. (G). Exogenous HA

rescued cell migration in scratch assays, in both U87 control and UGDH shRNA transfected cells.

A.

B.
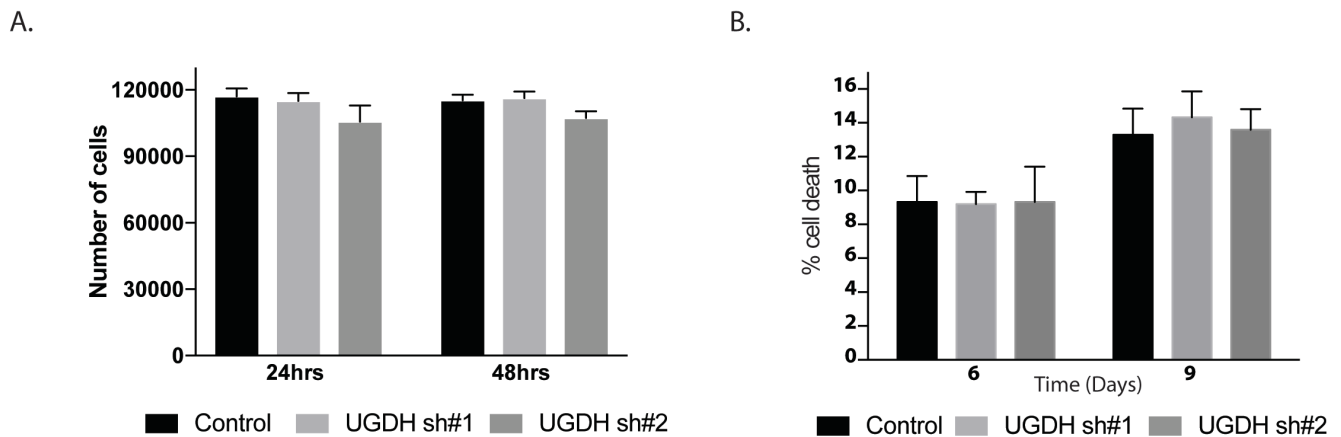
Figure 3-6. Cell proliferation assays for scratch assays
 (A) Cell proliferation assays under the conditions and time course of scratch assays. Cells grown in 0.1% FCS

for 48 hrs were counted manually. No difference in cell number was recorded, confirming an anti-migratory role

of UGDH knockdown under the similar conditions. (B) Trypan blue staining showed that in cultures grown at 6

and 9 days after plating, both control cells and UGDH knockdown cells had comparable percentage of cell

death, suggesting that the reduced cell growth after long time in culture was not due to cell death.

3-5. UGDH knockdown decreases GBM cell proliferation and clonogenicity

Studies show that, in addition to a prominent role in migration and metastasis, GAGs can influence

signal transduction, proliferation and differentiation [22]. We hypothesized that reduction of UGDH

could also influence cell proliferation. Cell counting demonstrated that UGDH knockdown inhibited cell proliferation by 45% and 20-50% after 9 days in U87 and GBM1A cultures, respectively (**Figure 3-7A,** *P*<0.001). Trypan blue staining demonstrated the reduced cell growth was not due to cell death as both control cells and UGDH knockdown cells had comparable percentage of non-viable cells. Next, we examined the effects of UGDH knockdown on clonogenicity of U87 cells and the capacity for neurosphere formation in GBM1A cells. We found that compared to controls, reduction of UGDH dramatically impaired the ability of U87 cells to form colonies in soft agar (by 50-60%, *P*<0.001) (**Figure 3-7B;** left panel). We also observed a ~70% reduction in the neurosphere forming capacity of GBM neurosphere cells as compared with the controls (*P*<0.001) **Figure 3-7B;** right panel). Next, we analyzed cell cycle progression in U87 control and UGDH knockdown cells. Cell cycle was synchronized by incubating cells in 0.1% serum for 48 hrs, followed by stimulation with 10% serum for up to 32 hrs. Cell cycle progression was analyzed at 0, 8, 16, 24 and 32 hr post serum addition. At 0 h, both control and UGDH knockdown cells had ~80% of cells in G1/G0 phase. After replenishing serum for 32 hrs, 83% of UGDH knock down cells remained in G1/G0, in comparison to only 60% in control cells (*P*<0.001, **Figure 3-7C**). Thus, UGDH knockdown displayed a delay in G1/G0 to S phase transition. Analysis of cell cycle regulators by Western blot revealed a 65%-80% reduction in cyclin E and a 40-50% decrease in cyclin D1 in UGDH knockdown cells (**Figure 3-7D**), providing further evidence that UGDH loss-of-function promoted a delay in cell cycle progression, which led to decreased cell proliferation in GBM cells.
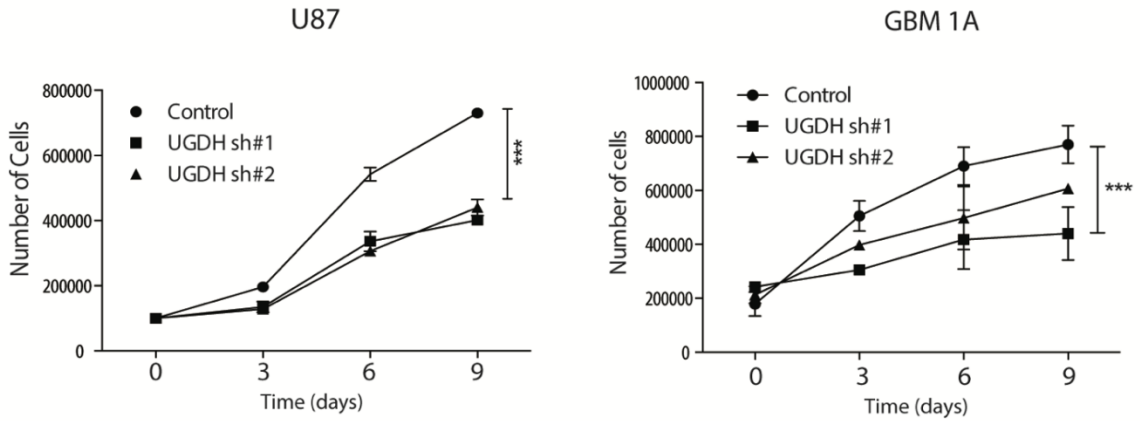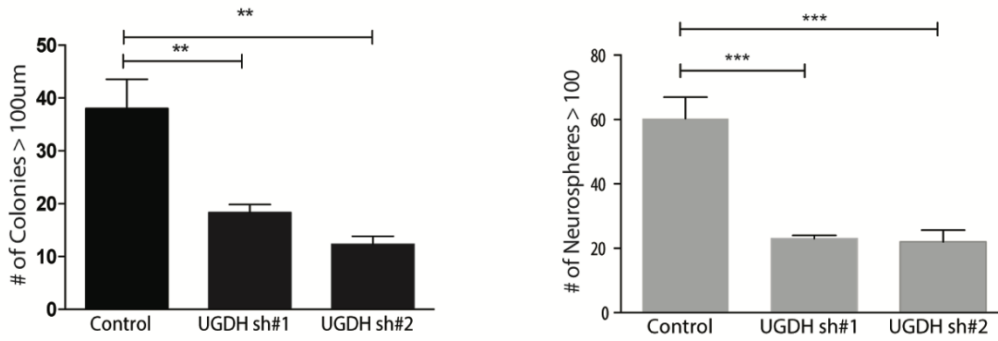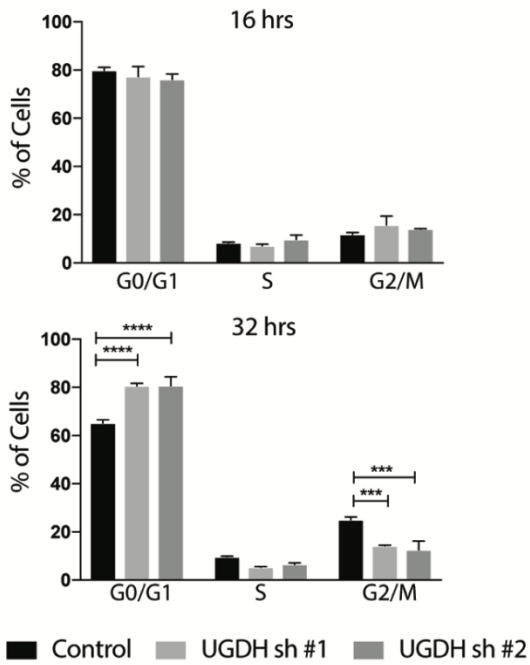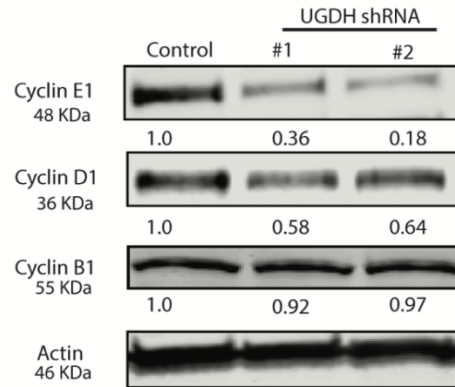
Figure 3-7. UGDH knockdown decreases GBM cell proliferation and clonogenicity
(A) Cell growth curve up to 9 days after plating. Trypsinized cells were stained with Trypan blue viable cells were counted on the days indicated. (B) Colony formation assays showing significant decrease in anchorage independent clongenicity in UGDH knockdown U87 cells (right panel). For GBM neurosphere cells, equal numbers of viable GBM1A cells were plated and cultured for 14 days to allow neurosphere formation. Neurospheres (>100 μm diameter) were counted with MCID software. UGDH silencing inhibited neurosphere formation. (C) Cell cycle was synchronized. There was a delayed progression to S phase in U87 UGDH knockdown cells compared to controls after 32 hrs of serum replenishing. (D) UGDH silencing decreased cyclin D1 and E protein levels in GBM cells. (*: $P < 0.05$; ** $P<0.01$; *** $P<0.001$).

3-6. UGDH knockdown reduces growth of GBM xenografts

The results that UGDH knockdown dramatically decreased GBM cell growth and migration *in vitro* prompted us to hypothesize that dramatic reduction of GAGs by UGDH knockdown would alter tumor growth *in vivo*. To examine the effect of UGDH silencing on orthotropic GBM growth, we employed GBM1A cells that form infiltrative orthotropic xenografts in immunodeficient (SCID) mice to examine the effect of UGDH silencing *in vivo*. GBM1A cells stably expressing either control shRNA or *UGDH* shRNA#1 were implanted into the brains of SCID mice and sacrificed 50 days after implantation according to our previous studies using this model [27]. Coronal brain sections were stained for hematoxylin and eosin. Tumor size measurement demonstrated a ~65% reduction in tumor growth in mice implanted with *UGDH* shRNA transfected cells (2.0 mm^3) compared to control shRNA transfected cells (6.0 mm^3) (**Figure 3-8A,** $P<0.001$). The anti-tumor effects of UGDH inhibition could be explained in part by a ~25% inhibition of the tumor cell proliferation index measured by anti-Ki67 immunohistochemical staining (**Figure 3-8B,** $P<0.001$), and ~25% inhibition of the vascular density index measured by anti-laminin staining, respectively (**Figure 3-8C,** $P<0.001$).

3-7. UGDH knockdown decreases expression of extracellular matrix proteins in GBM xenografts

Given the evidence that UGDH knockdown inhibits GAG expression *in vitro*, we investigated the extracellular matrix components in UGDH knockdown xenografts, hypothesizing that GAG expression inhibited by UGDH knockdown would alter extracellular matrix formation and therefore impact tumor growth. Expression of brevican, a proteoglycan downstream of UGDH pathway and tenascin C, a glycoprotein that binds and connects proteoglycans in the ECM, was quantified by immunohistochemical staining of intracranial xenograft sections from control and UGDH knockdown tumor groups. Brevican and tenascin C staining was reduced by 40-50% in tumors harboring *UGDH* shRNA compared to control tumors as evidenced by Image J software quantification of staining intensity (**Figure 3-8D, E,** *P*<0.001). *In vitro* Western blot analysis of U87 and GBM1A cells further confirmed reduced brevican (50%) and tenascin C (~65%) expression in response to UGDH silencing (**Figure 3-8F,** *P*<0.001). These results demonstrate that UGDH knockdown decreases the expression of GAGs and other extracellular matrix components *in vitro* and *in vivo*.
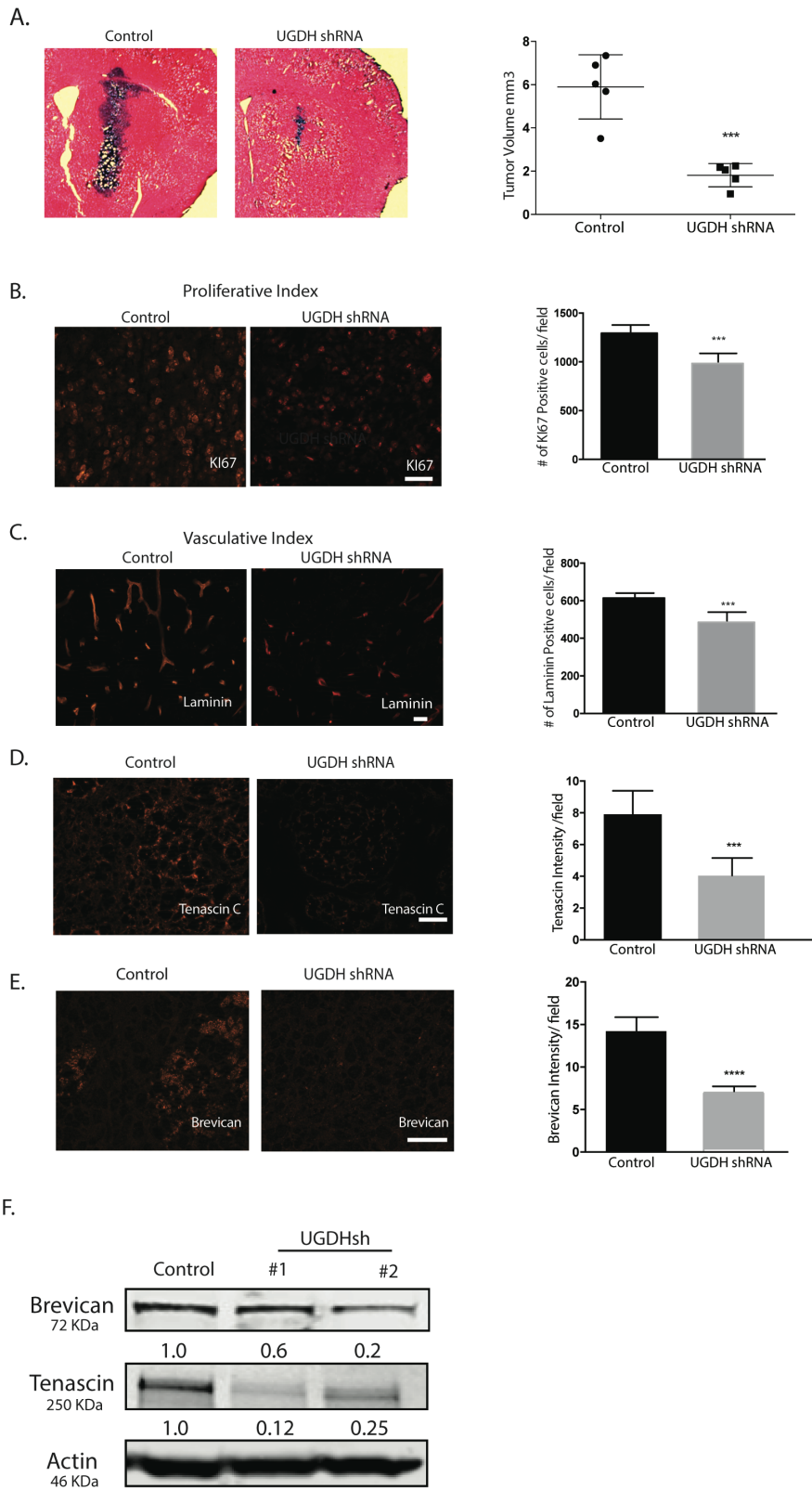
Figure 3-8. UGDH knockdown inhibits GBM growth *in vivo*

(A) Control shRNA or UGDH shRNA transfected GBM1A cells (100,000) were implanted by stereotactic injection to caudate/putamen of severe combined immunodeficiency mice (SCID). Animals were sacrificed 50 days after implantation. Hematoxylin and eosin-stained coronal brain sections (20 μm) obtained from animals showed dramatically decreased tumor size in UGDH knockdown groups (left panels, bar = 100 mm). Right panel: quantification of xenograft tumor volume shows that silencing UGDH repressed xenograft growth by more than 65% (6.4 in control vs 2.3 in UGDH sh#1, $P < 0.001$). B. UGDH knockdown significantly inhibited tumor cell proliferation by 31% as evidenced by Ki67 staining. C. UGDH loss-of-function inhibited blood vessel density by 25% as evidenced by laminin staining. (D, E) Xenografts with UGDH knockdown decreased the abundance of key extracellular matrix components tenascin C and brevican in GBM1A xenografts. (F) Western blots showing decreased brevican and tenascin C protein in U87 UGDH knockdown cells in vitro.

3-8. UGDH is required for induction of GBM cell migration by KLF4-mCpG interactions

Finally, we investigated the involvement of UGDH in cellular phenotype changes induced by KLF4-mCpG interactions. In tet-on U87 and GBM1A KLF4 WT cells, we knocked down *UGDH* expression with shRNAs to generate stable cells lines. Immunoblot analysis showed an ~80% to ~90% inhibition of UGDH in both U87 KLF4 WT and GBM1A KLF4 WT model systems (**Figure 3-9A**). KLF4 expression increased UGDH protein expression in the control shRNA transfected U87 and GBM1A cells as we have shown before, but failed to do so in the UGDH knockdown cell models (**Figure 3-9A**). A concomitant ~40-60% increase in GAGs in both U87 and GBM 1A cells was also observed after KLF4 WT expression was induced. However, UGDH knockdown abolished the increased GAG levels induced by KLF4 (**Figure 3-9B**). Next, we examined the effects of UGDH knockdown on KLF4-mCpG mediated GBM cell migration. KLF4 WT expression increased cell migration in transwell and wound healing assays, while UGDH knockdown reversed these changes induced by

KLF4-mCpG interactions (**Figure 3-9C, D**). These results support that UGDH is required for cell migration mediated via KLF4 binding to methylated CpGs.
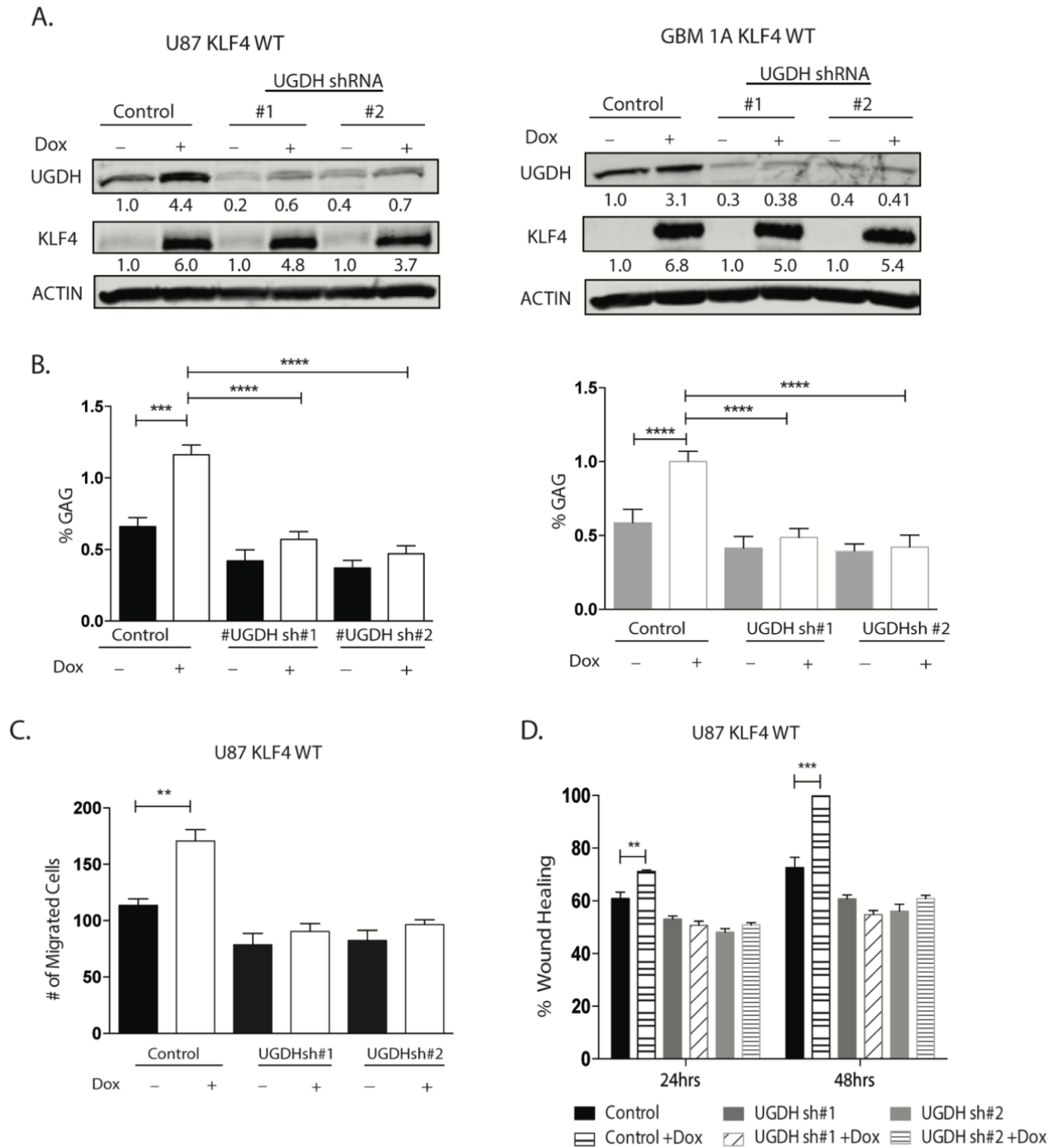
Figure 3-9. UGDH is required for KLF4-mCpG dependent increase in GBM cell migration

(A) U87 KLF4 WT and GBM1A KLF4 WT cells were infected with lentivirus coding for control shRNA, UGDH shRNA #1 or UGDH shRNA #2. Cells were treated with Dox for 48 hrs before each analysis. Western blot showed UGDH knockdown significantly reverses KLF4 WT dependent induction of UGDH in both U87 and GBM1A cells. (B) Sulfated GAG (sGAG) concentration in U87 and GBM1A cells expressing KLF4 WT was determined by DMMB assay. UGDH knockdown significantly decreased the KLF4 dependent increase in GAG concentration (C) UGDH knockdown significantly reversed the KLF4-dependent increase in cell migration in transwell assays. Cell migration was evaluated 24 hrs later by counting DAPI-stained cells. Five fields per well were counted. (D) U87 KLF4 WT cells harboring UGDH shRNA were treated with Dox for 5 days till confluence. A scratch was made and cells were maintained in 0.1% FCS medium overnight. Microphotographs were taken 0 hr and 24 hrs after scratching. Bar = 100 um. UGDH knockdown inhibited the increased ability of KLF4 WT cells to migrate towards scratched area.

3-9. Discussion

Elevated glycosaminoglycan (GAG), major components in tumor microenvironment, has been shown to regulate multiple oncogenic pathways including tumor growth, invasion and migration [17, 19, 25]. Consistent with these findings, several studies have demonstrated that inhibiting GAG synthesis diminishes tumor growth and metastasis [24, 25, 28]. In our prior study, we found that krÜppel-like factor 4 (KLF4) promotes GBM cell migration by binding to methylated DNA (mCpG) and activating gene expression. In this study, we investigated a subset of KLF4-mCpG direct targets and focused on UDP-glucose dehydrogenase (UGDH), given its importance as an enzyme involved in the synthesis of the precursors for GAGs. Although UGDH is implicated as a rate-limiting and essential step in GAG monosaccharide synthesis [21-23] the biological function of this enzyme in GBMs has not been explored. We demonstrate that UGDH is regulated via a methylation-dependent pathway; UGDH regulates cell migration and proliferation in vitro; silencing UGDH decreases levels of GAGs and

some key components of the extracellular matrix in vitro and in vivo, and results in inhibition of GBM growth. A model of UGDH's regulation and implication in GBM biology is proposed in **Figure 3-10**
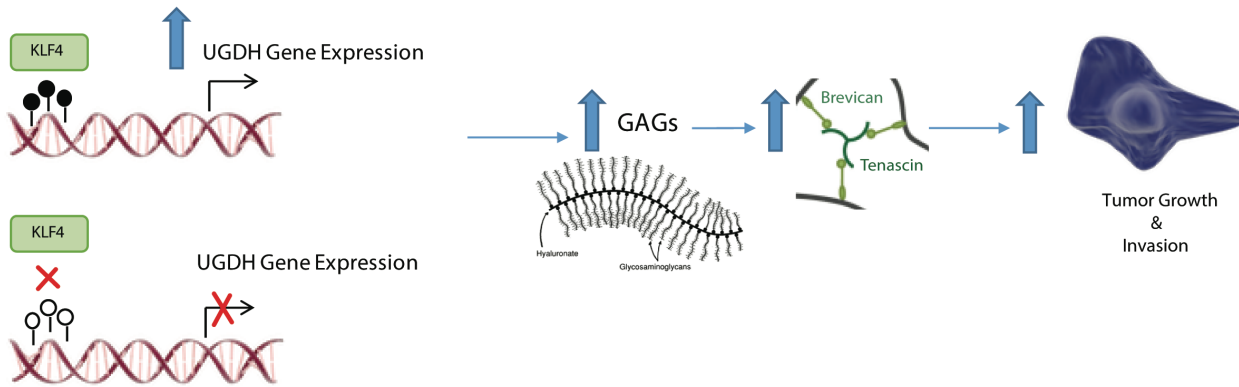
 Figure 3-10. Working model of KLF4 mediated increase in UGDH leading to increased GBM tumor migration and invasion

Expression of proteoglycan brevican and glycoprotein tenascin C, both of which are overexpressed in primary brain tumors as well as in experimental models of glioma, have been implicated in GBM progression [28-30]. In our study, cell migration was rescued by exogenous HA in UGDH knockdown cells, supporting the notion that the biological function of UGDH occurs via GAG production. Although it is not surprising that silencing UGDH expression leads to decreased tumor progression, our work provides a direct link between GAGs and other extracellular matrix proteins including tenascin C, a glycoprotein with a short carbohydrate chain. Tenascin C is shown to be elevated in the extracellular matrix of malignant brain tumor models and mediates tumor progression [30]. In our GBM models, we found that a reduction of GAGs facilitates decreased expression of tenascin C in vitro and in vivo, possibly via a post-translational mechanism. It has been reported that in human brains, the most abundant gel-like, long charge GAG, hyaluronic acid, serves as the backbone for the connection of other proteoglycans such as the brevican, which are further linked by tenascin C. It is conceivable that a decrease in glycosaminoglycan expression would disrupt the formation of these

54

complex networks, resulting in a decreased deposit of tenascin C. The exact mechanism through which silencing UGDH induces a decrease in tenascin C requires further detailed studies.

Whereas silencing UGDH dramatically decreased tumor cell migration and proliferation in vitro, our in vivo UGDH knockdown xenografts showed a dramatic decrease in tumor growth, mainly via decreased tumor cell proliferation. Unlike our previously published studies on tenascin C [30] in which the tumor/normal brain interface demonstrated well-demarcated borders in tenascin C knockdown orthotopic xenografts, our UGDH knockdown models did not support a role for UGDH knockdown in tumor cell infiltration in vivo. This could be due to a moderate decrease in extracellular matrix components, such as brevican and tenascin C, in UGDH knockdown xenografts, as opposed to a complete elimination of these components in TNC knock down xenografts [30].

DNA methylation, mainly at the C5 of the CpG dinucleotides, is present throughout the genome and more than 70% of CpGs are methylated in GBM [31, 32]. In fact, DNA hypermethylation is among the most commonly investigated epigenetic alterations in GBMs [32]. An increased understanding of how abnormal DNA methylation patterns contribute to transcription factor binding ultimately leads to downstream gene transcription is important for developing effective epigenetic-related therapies against GBM tumor progression. In this work, the anti-proliferative and pro-migratory role of UGDH are shown in vitro; a DNA methylation-dependent mechanism for UGDH regulation is identified, and modulation of GBM cell migration via KLF4-mCpG interactions is demonstrated. While little is known about the epigenetic regulation of UGDH expression, this study shows for the first time that KLF4 up-regulates UGDH expression via a methylation-dependent manner and also increases GAGs in GBM cells. Since our previous studies indicated that KLF4-mCpG binding activity promotes

migration but not proliferation in GBM cells, we focused our efforts on studying the effect of UGDH knockdown on KLF4-induced GAG production and cell migration in GBM cells.

Finally, considering the importance of the enzyme UGDH in GAG synthesis and the critical role of GAG in tumor growth, we provide evidence supporting the notion that UGDH could be a potential therapeutic target to treat GBM malignancy. Future studies with small molecular inhibitors specific for UDGH would test if UGDH could serve as a therapeutic target alone or in combination with other therapies for GBMs.

References

1. Nandan, M.O. and V.W. Yang, *The role of Kruppel-like factors in the reprogramming of somatic cells to induced pluripotent stem cells.* Histol Histopathol, 2009. **24**(10): p. 1343-55.

2. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.

3. Leng, Z.W., et al., *Kruppel-Like Factor 4 Acts as an Oncogene in Colon Cancer Stem Cell-Enriched Spheroid Cells.* Plos One, 2013. **8**(2).

4. Yu, F., et al., *Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion.* Oncogene, 2011. **30**(18): p. 2161-2172.

5. Zhu, L.X., et al., *The PGI-KLF4 pathway regulates self-renewal of glioma stem cells residing in the mesenchymal niches in human gliomas.* Neoplasma, 2014.

6. Elsir, T., et al., *A study of embryonic stem cell-related proteins in human astrocytomas: Identification of Nanog as a predictor of survival.* International Journal of Cancer, 2014. **134**(5): p. 1123-1131.

7. Holmberg, J., et al., *Activation of Neural and Pluripotent Stem Cell Signatures Correlates with Increased Malignancy in Human Glioma.* Plos One, 2011. **6**(3).

8. Carlsson, S.K., S.P. Brothers, and C. Wahlestedt, *Emerging treatment strategies for glioblastoma multiforme.* EMBO Mol Med, 2014. **6**(11): p. 1359-1370.

9. Quick, A., et al., *Current therapeutic paradigms in glioblastoma.* Rev Recent Clin Trials, 2010. **5**(1): p. 14-27.

10. Evans, P.M. and C. Liu, *Roles of Krupel-like factor 4 in normal homeostasis, cancer and stem cells.* Acta Biochim Biophys Sin (Shanghai), 2008. **40**(7): p. 554-64.

11.   Rowland, B.D., R. Bernards, and D.S. Peeper, *The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene.* Nat Cell Biol, 2005. **7**(11): p. 1074-82.

12.   Tetreault, M.P., Y. Yang, and J.P. Katz, *Kruppel-like factors in cancer.* Nat Rev Cancer, 2013. **13**(10): p. 701-13.

13.   Wan, J., et al., *Methylated cis-regulatory elements mediate KLF4-denpendent gene transactivation and cell migration.* eLife, 2017. **6**: p. e20068.

14.   Adamson, C., et al., *Glioblastoma multiforme: a review of where we have been and where we are going.* Expert Opin Investig Drugs, 2009. **18**(8): p. 1061-83.

15.   Batash, R., et al., *Glioblastoma Multiforme, Diagnosis and Treatment; Recent Literature Review.* Curr Med Chem, 2017.

16.   Marzenna Wiranowska and Mumtaz V. Rojiani, D.A.G., *Extracellular Matrix Microenvironment in Glioma Progression, Glioma - Exploring Its Biology and Practical Relevance.* InTech, 2011.

17.   Venning, F.A., L. Wullkopf, and J.T. Erler, *Targeting ECM Disrupts Cancer Progression.* Front Oncol, 2015. **5**: p. 224.

18.   Lu, P., V.M. Weaver, and Z. Werb, *The extracellular matrix: a dynamic niche in cancer progression.* J Cell Biol, 2012. **196**(4): p. 395-406.

19.   Afratis, N., et al., *Glycosaminoglycans: key players in cancer cell biology and treatment.* FEBS J, 2012. **279**(7): p. 1177-97.

20.   Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* eLife, 2013. **2**: p. e00726.

21.   Egger, S., et al., *UDP-glucose dehydrogenase: structure and function of a potential drug target.* Biochem Soc Trans, 2010. **38**(5): p. 1378-85.

22.   Clarkin, C.E., et al., *Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and release, control sulfated GAG synthesis, and promote chondrogenesis.* J Cell Physiol, 2011. **226**(3): p. 749-61.

23.   Wen, Y., et al., *UDP-glucose dehydrogenase modulates proteoglycan synthesis in articular chondrocytes: its possible involvement and regulation in osteoarthritis.* Arthritis Research & Therapy, 2014. **16**(6): p. 484.

24.   Toole, B.P., *Hyaluronan: from extracellular glue to pericellular cue.* Nat Rev Cancer, 2004. **4**(7): p. 528-39.

25.   Park, J.B., H.-J. Kwak, and S.-H. Lee, *Role of hyaluronan in glioma invasion.* Cell Adhesion & Migration, 2008. **2**(3): p. 202-207.

26.   Lu, P., et al., *Extracellular Matrix Degradation and Remodeling in Development and Disease.* Cold Spring Harbor perspectives in biology, 2011. **3**(12): p. 10.1101/cshperspect.a005058 a005058.

27.   Rath, P., et al., *In Vivo c-Met Pathway Inhibition Depletes Human Glioma Xenografts of Tumor-Propagating Stem-Like Cells.* Transl Oncol, 2013. **6**(2): p. 104-11.

28.   Wade, A., et al., *Proteoglycans and their roles in brain cancer.* The FEBS journal, 2013. **280**(10): p. 2399-2417.

29.   Nutt, C.L., et al., *Brain enriched hyaluronan binding (BEHAB)/brevican increases aggressiveness of CNS-1 gliomas in Lewis rats.* Cancer Res, 2001. **61**(19): p. 7056-9.

30.   Xia, S., et al., *Tumor microenvironment tenascin-C promotes glioblastoma invasion and negatively regulates tumor proliferation.* Neuro Oncol, 2016. **18**(4): p. 507-17.

31.   Craig, J.M. and W.A. Bickmore, *The distribution of CpG islands in mammalian chromosomes.* Nat Genet, 1994. **7**(3): p. 376-382.

32.   Medvedeva, Y.A., et al., *Effects of cytosine methylation on transcription factor binding sites.* BMC Genomics, 2014. **15**(1): p. 119.

CHAPTER 4

CHROMOSOME CONFORMATION CAPTURE (3C) ANALYSIS REVEALS KLF4 BINDS TO METHYLATED CPG AT ENHANCER REGIONS TO ACTIVATE GENE EXPRESSION

4-1. Introduction

Using zinc finger krueppel-like factor 4 (KLF4) as the first candidate, we investigated the function of TF-mCpG binding in biological relevant systems, e.g. human GBM cells, and obtained strong evidence that KLF4 binds to mCpG at cis-element regions and activate GBM adhesion and migration. Employing KLF4 site-specific mutant (KLF4R458A) that lacks KLF4 binding ability to mCpG but retains its binding to canonical, non-methylated CpGs, we discovered that this novel KLF4-mCpG interaction could activate gene expression [1]. A follow-up study aimed to determine the cellular function of KLF4-mCpG interactions in GBM cells [2]. Our studies revealed that KLF4 mediates brain tumor cell migration by binding to mCpGs in the cis-regulatory elements and activating cell motility gene expression, e.g. RHOC, RAC1 etc [2]. All these in-depth studies of KLF4-mCpG interactions focused on KLF4 binding to mCpG at promoter regions. The function of KLF4-mCpG interactions at enhancer regions requires more investigation.

We also identified many other direct targets of KLF4-mCpG binding, including the Src family member of B Lymphocyte Kinase (BLK). The Src family of tyrosine kinases consists of non-receptor tyrosine kinases including LYN, LCK, HCK and BLK. Src kinases function in the signal transduction of multiple cellular processes including migration, adhesion, invasion, angiogenesis, proliferation and differentiation. BLK is primarily expressed by B lineage cells but it is also expressed in non B-lineage cell types such as pancreatic ß cells and human thymocytes [3]. BLK is activated upon B cell receptor (BCR) stimulation and phosphorylates others in the BCR signaling pathway. Besides a role of BLK in B cell receptor signaling and B cell development, the cellular functions of BLK in brain tumors are poorly defined [4].

In the current study, we focused our studies on BLK regulation and biological function. Through 3C analysis, we found BLK was activated by KLF4 binding to enhancers in a methylation dependent

manner. In addition, we found that in genome scale, KLF4 binds to methylated enhancer regions and activates gene transcription.

4-2. Identification of proteins regulated by KLF4-mCpG binding

In our previous studies by Hu et al [1], KLF4 point mutation R458A was generated to study KLF4 binding to methylated CpG (mCpG). Compared to KLF4 wild type (KLF4 WT), KLF4 R458 loses KLF4 binding ability to mCpG but retains its binding to non-mCpG. The lack of binding to mCpG by KLF4 R458A mutant was further confirmed by our recent publication using whole genome bisulfite sequencing and chromatin immunoprecipitation (ChIP)-sequencing in our previously engineered human U87 GBM cells expressing tet-on KLF4 WT or KLF4 R458A [2]. In these models, cells without doxycycline treatment served as a negative control; upon doxycycline (Dox) treatment, KLF4 WT and KLF4 R458A were induced by ~20 fold, similar to that during cancer cell reprogramming when challenged by growth factors [5] (**Fig. 4-1A**). Therefore, the distinct phenotypes and downstream targets driven by KLF4 WT and KLF4 R458A have been considered as biological function of KLF4-mCpG interactions. It was critical to detect our KLF4-mCpG targets at the protein level. Therefore, we performed proteomics to identify protein expression changes driven by KLF4 WT and KLF4 R458A.

The protein expression changes mediated by KLF4 WT and KLF4 R458A mutant were measured under four conditions: U87 KLF4 WT +/- Dox and U87 KLF4 R458R +/- Dox after cells being treated with Dox for 48 h. In all, we detected 6875 proteins (or 7008 isoforms) in our proteomics analysis. We compared protein expression changes in U87 KLF4 WT cells after Dox treatment with that before Dox treatment, and found 73 proteins were up-regulated and 37 down-regulated in KLF4 WT expressing cells (**Fig. 4-1B**). In contrast, no differentially expressed proteins were found upon KLF4R458A induction when we used the same cut-off as the KLF4 WT cells (data not shown). Expression levels of

a given protein significantly altered (up-regulated or down-regulated) by KLF4 WT, but not by KLF4R458A would be the putative KLF4-mCpG direct targets. With these criteria, our large-scale proteomics data analysis generated a total of 110 proteins differentially affected by KLF4 WT and KLF4 R458A mutant. To further determine if these proteins were direct targets of KLF4 binding activities, we examined the binding of KLF4 to the *cis*-regulatory regions of these proteins. In total, 39 proteins were up-regulated by with KLF4 binding to mCpGs at either upstream/5'UTR/exon or predicted enhancer location in U87 cells (**Fig. 4-1C**). For example, our data indicated that the protein expression level of small GTPase RHOC was increased by KLF4 WT but not KLF4 R458A, consistent with our previous RNA-seq, RT-PCR and Western blot analysis [2], indicating that KLF4-mCpG interactions drove the transcriptional and translational activation of RHOC. Another validated KLF4-mCpG target, UGDH was also found only up-regulated by KLF4 WT in our proteomics studies.
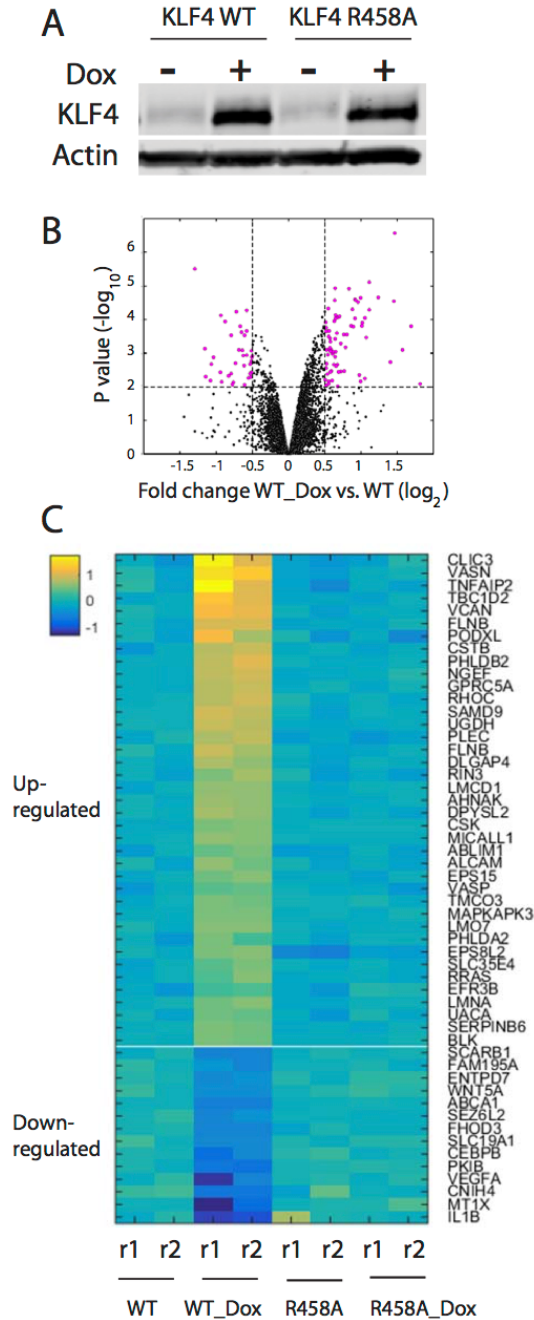
Figure 4-1. Protein Expression changes induced by KLF4 WT and KLF4 R458A in U87 cells
A. U87 KLF4 WT and U87 KLF4 R458R cells were treated with Dox (1 ug/ml) for 48 h, Western blot analysis

verified the induction of KLF4 WT and KLF4 R458A proteins. B. Graph of differentially expressed proteins

regulated by KLF4 WT and KLF4 R458A. Protein samples were collected from untreated and Dox treated U87

KLF4 WT and KLF4 R458A cells, and subjected to mass spectrometry for protein analysis. In total, expression

of 65 proteins was altered by KLF4 WT, but not by KLF4R458A. Approximately 40 proteins were up-regulated,

and ~25 down-regulated by KLF4 WT only. C. Heatmap of proteomics data with our previously published RNA-seq data

4-3. BLK was activated by KLF4 in a mCpG-dependent manner

In our previous studies, we used chromatin immunoprecipitation (ChIP) sequencing to search for KLF4-mCpG binding sites, and we performed RNA-sequencing to identify gene expression changes driven by KLF4-mCpG interactions. By combining KLF4-mCpG binding sites and KLF4-mCpG-regulated genes, a list of ~130 genes were generated as direct gene targets of KLF4-mCpG [2]. From our large scale RNA-sequencing (published in [2]) and proteomics analysis, we noticed one of the putative KLF4-mCpG targets, BLK, was up-regulated by KLF4 WT at both the mRNA (**Fig. 4-2A**) and protein level (**Fig. 4-1C**). We further validated the expression of BLK, by quantitative real time PCR (RT-PCR) and Western blot analysis. RT-PCR indicated that *BLK* mRNA level was increased ~2.2 fold upon KLF4 WT expression for 48 h, but not changed by KLF4 R458A (**Fig. 4-2B,** $P<0.05$). Western blot analysis confirmed that BLK protein was significantly increased by ~3.9 fold by KLF4 WT only, but not by KLF4 R458A (**Fig. 4-2C,** $P<0.001$). The differential expression of BLK in KLF4 WT and KLF4 R458A cells suggested that BLK was activated by KLF4 binding to methylated CpGs in cis-regulatory elements. To find out the binding sites of KLF4, we queried our KLF4 ChIP-sequencing data for *BLK* in U87 KLF4 WT and KLF4 R458A expressing cells. We found two possible binding sites for KLF4 WT on the entire genome of the *BLK* gene, one at the 5'-UTR site, and the other at the 3'-UTR site (**Fig. 4-2D**)**.** The 5'-UTR KLF4 binding site did not contain CpG sites, but the putative KLF4 binding site of *BLK* gene at 3'-UTR contained multiple mCpG sites, therefore we focused our following studies on the 3'UTR site. We queried KLF4 ChIP-seq data and found that this site was only bound by KLF4 WT, but not by KLF4 R458A (**Fig. 4-2E),** suggesting that mCpG mediated KLF4 binding at this site.

Further investigation of the enhancer mark H3K27ac ChIP-seq data at the same site revealed that there was an enrichment of H3K27ac binding in U87 KLF4 WT expressing cells, but not in control or KLF4 R458A expressing cells (**Fig. 4-2F**).
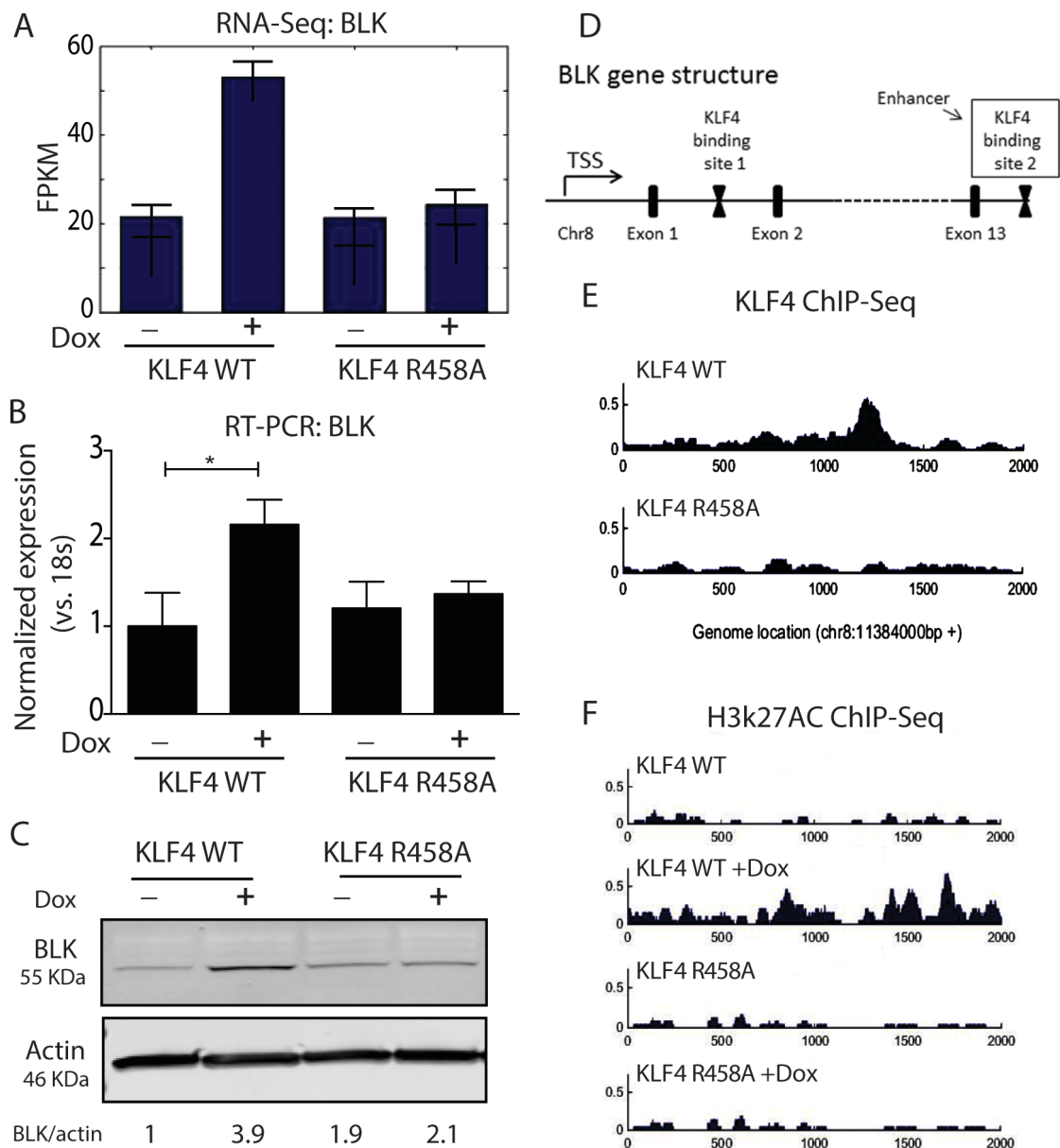
Figure 4-2. BLK was only bound and activated by KLF4 WT, not KLF4 R458A.

A. RNA-sequencing data showing that *BLK* gene was only up-regulated by KLF4 WT, but not KLF4 R458A. B. Real time-PCR verified that BLK mRNA level was increased by ~2.6 fold in KLF4 WT expressing cells only. C. Western blot analysis showed that BLK was up-regulated ~3 fold by KLF4 WT, but not changed by KLF4 R458A. D. KLF4 ChIP-sequencing data of *BLK* in U87 cells. There are two possible binding sites for KLF4 at the entire genomic structure of *BLK*, one is that the 5'-UTR site, the other one is at the 3'-UTR site.

To determine if KLF4 indeed bound to the 3'-UTR site of *BLK* gene via mCpG-dependent mechanism, we performed ChIP-PCR. An antibody against KLF4 was used to precipitate genomic DNA from KLF4 WT and KLF4 R458A expressing cells (+Dox 48h). This antibody recognized N-terminal of KLF4 and has been previously shown to equally precipitate KLF4 WT and KLF4 R458A as the mutated site is at the C-terminal of KLF4 [2]. We found that binding fragment at the 3'-UTR of *BLK* gene was only enriched in KLF4 WT expressing cells, but not in KLF4 R458A expressing cells, indicating that the mCpG-dependent binding activity of KLF4 was involved in *BLK* transcription (**Fig. 4-3A**). To confirm the methylation status of the *BLK* binding regions, bisulfite sequencing was performed. There was a 100% methylation level at 3 CpG sites of the putative BLK binding site at 3'-UTR (**Fig. 4-3B**).

To confirm that methylation at the 3'-UTR is responsible for KLF4 binding and gene activation in vivo, we treated the cells with the DNA methyltransferase inhibitor 5-aza, which has been previously proven by us to significantly decrease genome wide methylation and reverse gene expression in U87 cells [2]. 5-Aza completely erased the methyl group at the CpG sites we tested (**Fig. 4-3C**). ChIP-PCR also indicated that 5-aza completely blocked KLF4 WT binding to the 3'-UTR of the *BLK* gene (**Fig. 4-3D**). RT-PCR revealed that 5-aza reversed *BLK* induction in KLF4 WT expressing cells (**Fig. 4-3E**). All these suggest that the methylated sites at the 3'-UTR of the *BLK* gene bound to KLF4 and is responsible for up-regulated *BLK* expression.
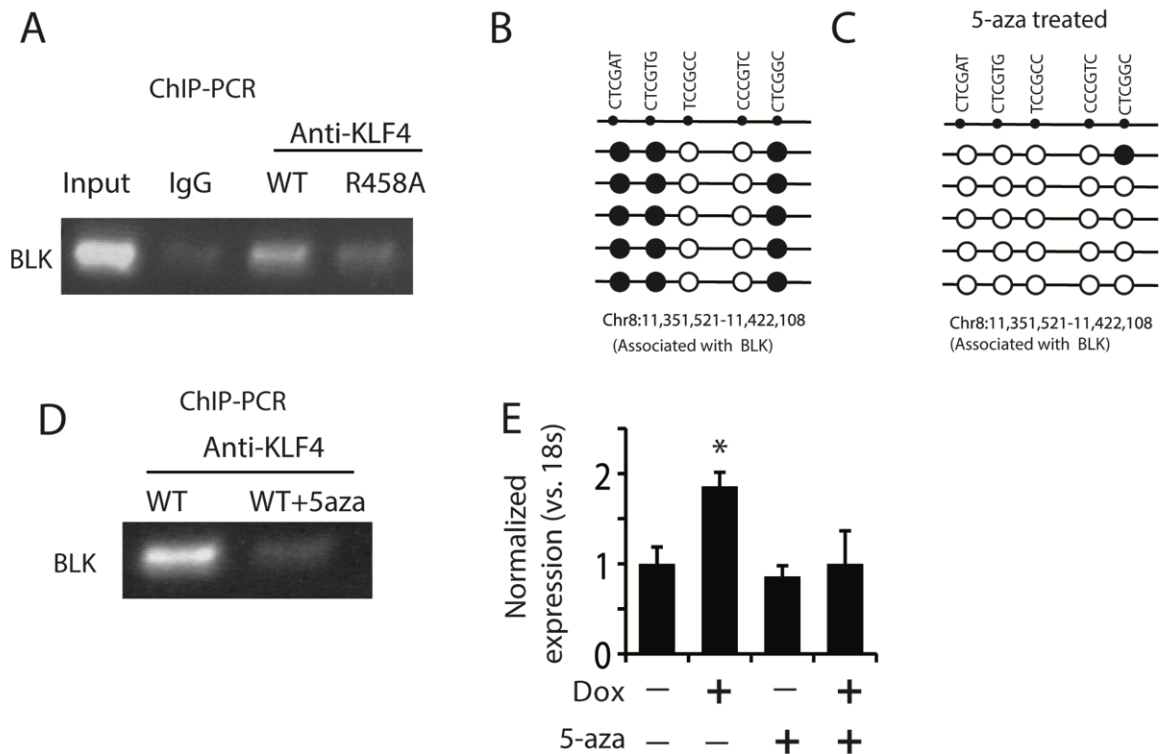
Figure 4-3. KLF4 bound to the enhancer region of BLK in a methylation dependent manner to activate gene expression

A. Genomic DNA from U87 KLF4 WT and KLF4 R458A cells immunoprecipitated by a KLF4 antibody and served as templates for PCR. The 3'UTR binding site of BLK gene amplified and was only enriched in KLF4 WT expressing cells, not in KLF4 R458A expressing cells. Input genomic DNA served as a positive control and isogenic IgG negative control. B. Bisulfite sequencing analysis of the KLF4 binding site at 3'-UTR of BLK gene indicated that 3of 5 CpG sites were 100% methylated in U87 cells. C. U87 KLF4 WT cells were treated with DNA methyltransferase inhibitor 5-Aza. 5-Aza completely erased the methyl group at the CpG sites of the 3'UTR region of BLK gene. D. ChIP-PCR of the same 3'-UTR region in U87 KLF4 WT cells after 5-Aza treatment. 5-Aza abolished KLF4 WT binding to this fragment. E. RT-PCR indicated that 5-Aza reversed *BLK* induction in KLF4 WT expressing cells.

69

4-4. 3C showed that KLF4 binds to the 3-D structure of BLK genome

The putative KLF4 binding site of BLK gene at 3'-UTR contained multiple mCpG sites and more likely serves as an enhancer to activate gene expression. We speculate that KLF4 binds to the *BLK* promoter region at 5'-UTR and the enhancer at 3'-UTR through a chromatin loop to activate gene expression (**Fig. 4-4A**). chromosome conformation capture (3C) technique was performed to investigate whether KLF4 binding to methylated CpG at enhancer regions could activate gene expression. Experiments were performed under four conditions, U87 KLF4 WT +/- Dox and U87 KLF4 R458A +/- Dox. Genomic DNA from cells was cross-linked and fragmented by digesting with two enzymes: Bgl I and Hind III. The digested DNA fragments were ligated and served as templates for PCR. Two primers were designed to amplify the adjacent sites of the looped structure, the left primer spanned the end of 5'-UTR KLF4 binding site, and the right primer spanned the beginning of 3'-UTR KLF4 binding site. A fragment of 298 bp PCR product was predicted if our hypothesized *BLK* activation model in **Fig. 4A** was correct. Indeed, PCR products that linked the 5'-UTR and 3'-UTR binding fragments with the corrected size (~298 bp) were detected in U87 KLF4 WT cells before Dox treatment, and increased after KLF4 WT induction. In contrast, there was no enrichment of this putative linked fragment in U87 KLF4 R458A cells (**Fig. 4-4B).** Since the same amount of ligated DNA from each condition was used as the PCR template, we were able to semi-quantify a ~9.2 fold increase in PCR product of U87 KLF4 WT expressing cells, when compared to cells before Dox treatment (**Fig. 4-4C,** *P*<0.001). Sanger sequencing showed that the PCR fragment sequence was a 100% match to the putative loop linked sequence as **AATA**AGCTTGC (**Fig. 4-4D, E** showing the linked sequence). All these data indicates that *BLK* was activated by KLF4 binding to the enhancer at 3'-UTR and forming a 3D structure with the promoter region at 5'-UTR.

We also found that KLF4 R458A did not increase the PCR production of the ligated fragment from the putative loop structure, indicating that methylation dependent KLF4 binding to the 3'-UTR of BLK genome structure is responsible for the loop formation. To corroborate this, the KLF4 WT cells were treated with 5-Aza before adding Dox, and we failed to amplifying the loop fragment in U87 KLF4 WT expressing cell under this condition (**Fig. 4-4F**), further supporting our hypothesis that KLF4 binds to mCpGs at the 3'- UTR enhancer region to activate gene expression.

A. 1802bp core Promoter

Enhancer 3' UTR Exon

KLF4

-91bp

+1

1711bp

Promoter
Exon2: 33,510-33,910

Position: chr8: 11,351,521-11,422,108
Size: 70,588 Total Exon count: 13 Strand: +

B. PCR after 3C

KLF4 WT KLF4 R458A

Dox   −   +   −   +

298 bp

C.
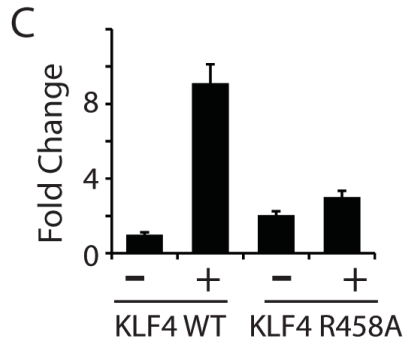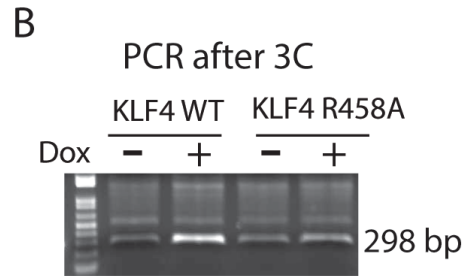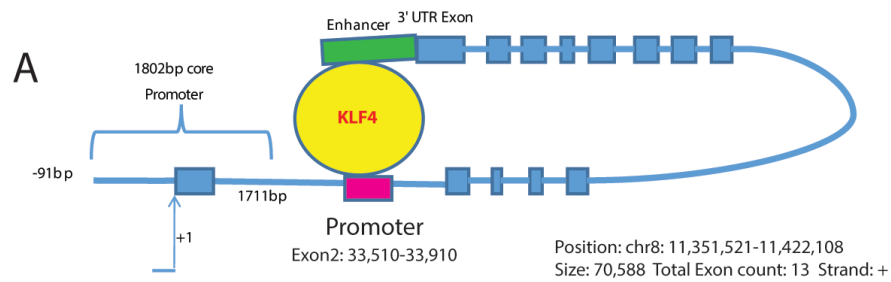
D.

AGGTGAATAACCCCAAAAATCAACCAAATATGGGGTCATCAGCAGAAGCTCAAGAACAAT

GGATGAGGCTGGAATCTAATAACAGGTGTGCTGTAGCTGTTTTGAAACATAATTTTTCTC

TCTCCAGTTTATTTTTACTAAAGACAAGACCAATTTATTTGCAAAATAAGCTTGCATGGC

TTTTCCCACAGGGGCCAAGTTCCCCATCAAGTGGACAGCCCCGGAAGCCATCCACTTCGG

GGTCTTCACCATCAAAGCAGACGTGTGGTCGTTTGGAGTCCT

E.

AATAAGCTT

F. KLF4 WT

298 bp

Dox   −   +   −   +
5-Aza   −   −   +   +
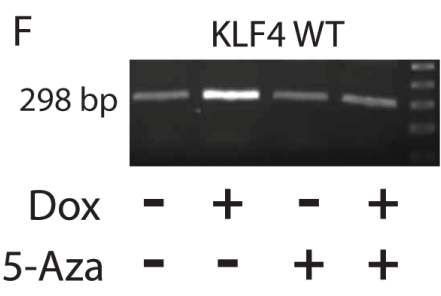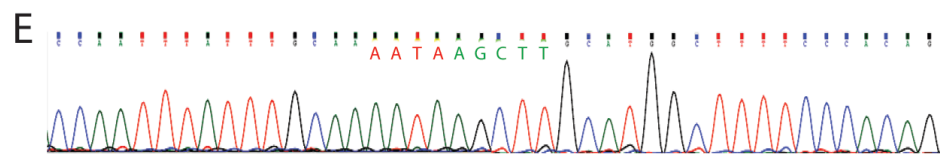
Figure 4-4. Chromosome conformation capture (3C) showed KLF4 bound to methylated enhancer to activate BLK expression.
A. Hypothesized 3D structure of *BLK* gene activation showing KLF4 binds to the 5'-UTR and 3'-UTR fragment

through a chromatin loop to activate gene expression. A fragment with the size of 298 bp B. 3C was performed

72

under four conditions, U87 KLF4 WT +/- Dox and U87 KLF4 R458R +/- Dox. Cross-linked genomic DNA was fragmented by digesting with Bgl I and Hind III followed by ligation of adjacent fragments. A primer set was designed to partially match the 5'-UTR and 3'-UTR KLF4 binding sites, respectively. same amount of template was used for PCR to amplify the putative linked site. Indeed we detected a PCR fragment with 298 bp that link the 5'-UTR and 3'-UTR binding fragment together in control cells, which was significantly increased after KLF4 WT expression (+Dox), but not changed in KLF4 R458A expressing cells. C. Quantification of the enrichment of PCR product in U87 KLF4 WT and KLF4 R458A cells indicated a 4.6 fold increase in the putative linked sites in KLF4 WT expressing cells. E. Sanger sequencing confirmed that the PCR fragment 100% matched our putative loop with the linked site sequencing as shown in B. D. 5-Aza treatment in U87 KLF4 WT expressing cells prevented loop formation and no enrichment of PCR product was detected.

4-5. BLK knockdown decreased cell migration

BLK is a non-receptor tyrosine kinase belonging to the SRC family kinases and its known to be functionally involved in B-cell receptor signaling and B-cell development. To determine the biological function of BLK in glioblastoma (GBM) cells, we studied the loss-of-function of BLK in U87 cells by using two BLK shRNAs to generate cell lines with BLK knock down. Two distinct constructs of BLK shRNA reduced BLK expression more than 80%in U87 GBM cells (**Fig. 4-5A**). Cell proliferation and cell migration were analyzed in these cells. We found that BLK knockdown did not change cell proliferation and clone formation ability (**Fig. 4-5B, C**). In contrast, BLK knockdown significantly increased cell adhesion (**Fig. 4-5D**, *P*<0.01), and reduced cell migration in transwell assays (**Fig. 4-5E**) and scratch wound healing assays (**Fig. 4-5F**). All these suggest that methylation on enhancer regions dictate transcription factor binding via 3D structure formation and promote cancer cell migration.
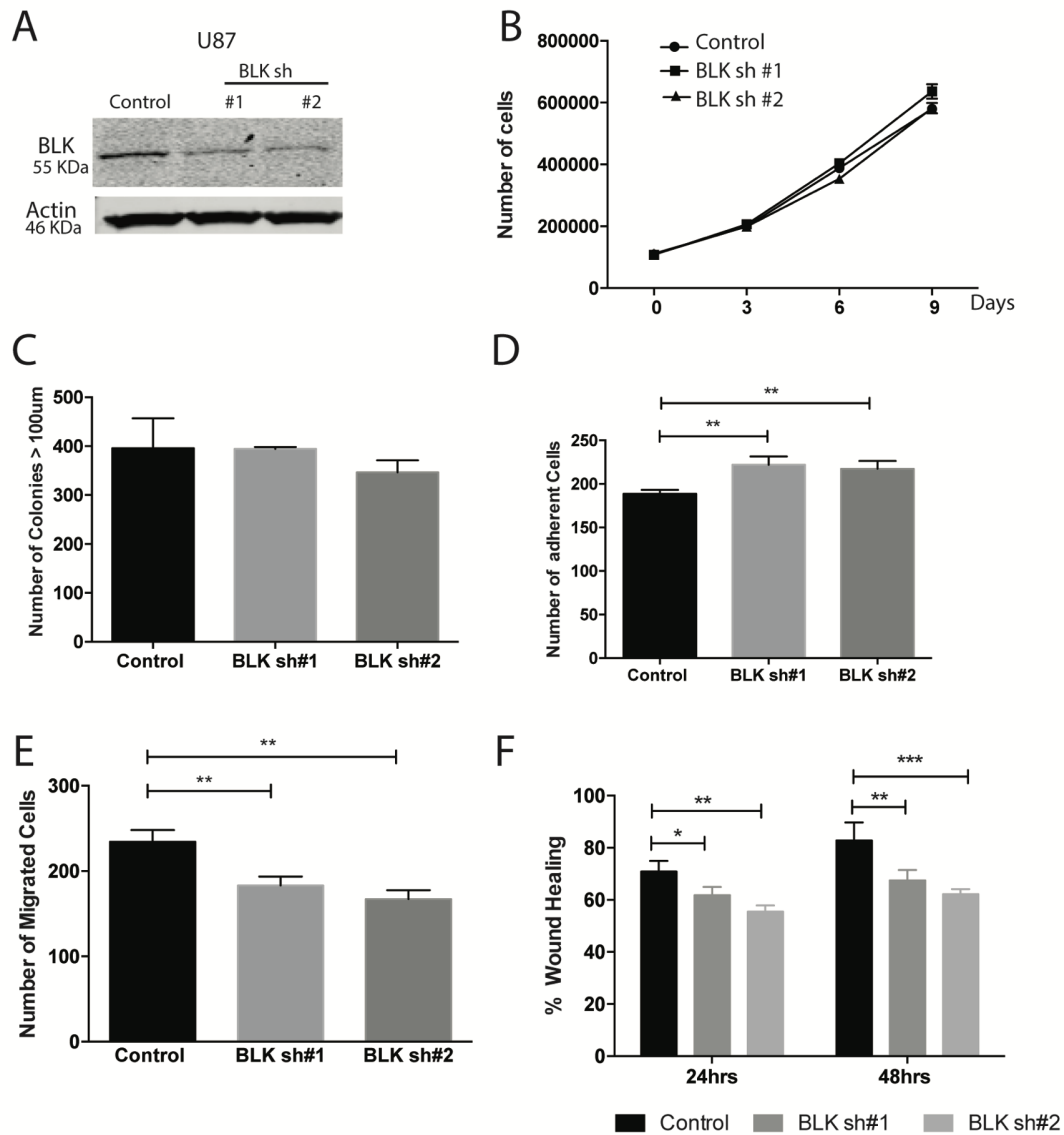
Figure 4-5. Biological function of BLK in GBM cells.
A. BLK was knockdown by ~80% in U87 cells with two distinct constructs of BLK shRNA. B. BLK knockdown did not change cell proliferation. C. BLK knockdown significantly reduced cell migration in transwell assays. D, E. BLK down-regulation partially blocked cell migration induced by KLF4. ***: $P<0.001$

4-6. Global analysis of KLF4 binding to methylated enhancer regions to activate gene expression

We believe that KLF4 binding to mCpGs on enhancer regions to activate gene expression applied to many more downstream targets of KLF4-mCpG interactions. We therefore performed more

sophisticated data analysis by integrating KLF4 WT ChIP sequencing, whole genome bisulfite sequencing and the enhancer marker H3K27ac ChIP sequencing data. In our previous studies, we identified a total of 3890 KLF4 WT binding fragments in U87 cells. Among them 1545 were highly methylated after overlapping with the whole genome bisulfite sequencing data. We used the histone mark H3K27ac as an enhancer mark, combined the H3K27ac ChIP-sequencing data with the highly methylated KLF4 WT binding fragments, and found 547 putative methylated enhancer regions in U87 cells (**Fig. 4-6A**). Motif analysis of the 547 fragments shown in Fig. 6B. The 6-mer CCCGCC has been previously identified as the motif of KLF4 binding to methylated DNA. In total, we found 149 genes that could be activated by KLF4 via its binding to methylated CpG at enhancer regions. Some examples of these genes, e.g. *UGDH, LMO7, GRIN1* and *HSP90AB1*, shown in **Fig. 4-6C, D.**

Besides *BLK*, 3C analysis was performed to validate whether other gene targets were activated by KLF4-mCpG interactions at enhancer regions. Mostly, we focused on those 149 putative gene targets that have been shown upregulated by KLF4 WT at protein level **(Fig. 4-1A).** Among the targets we tested, we found that LMO7 and UGDH were also regulated by KLF4 binding to methylated mCpGs at the enhancer regions via a 3D loop formation. We observed an enrichment of the predicated ligation fragments by binding to KLF4 to the LMO7 (**Fig. 4-6E**) or UGDH (**Fig. 4-6F**) genes in KLF4 WT expressing cells. In contrast, KLF4 R458A expression did not enrich the PCR products of predicted ligated fragments when compared to untreated cells. A negative control was seen in IDH1 gene, there are two possible binding sites for IDH1 and our 3C analysis failed to amplify the putative loop structure based on the two sites from our ChIP-seq data (data not shown). All these suggest that our 3C analysis of transcription factor-enhancer binding reveals a novel methylation-dependent mechanism in gene transcription.
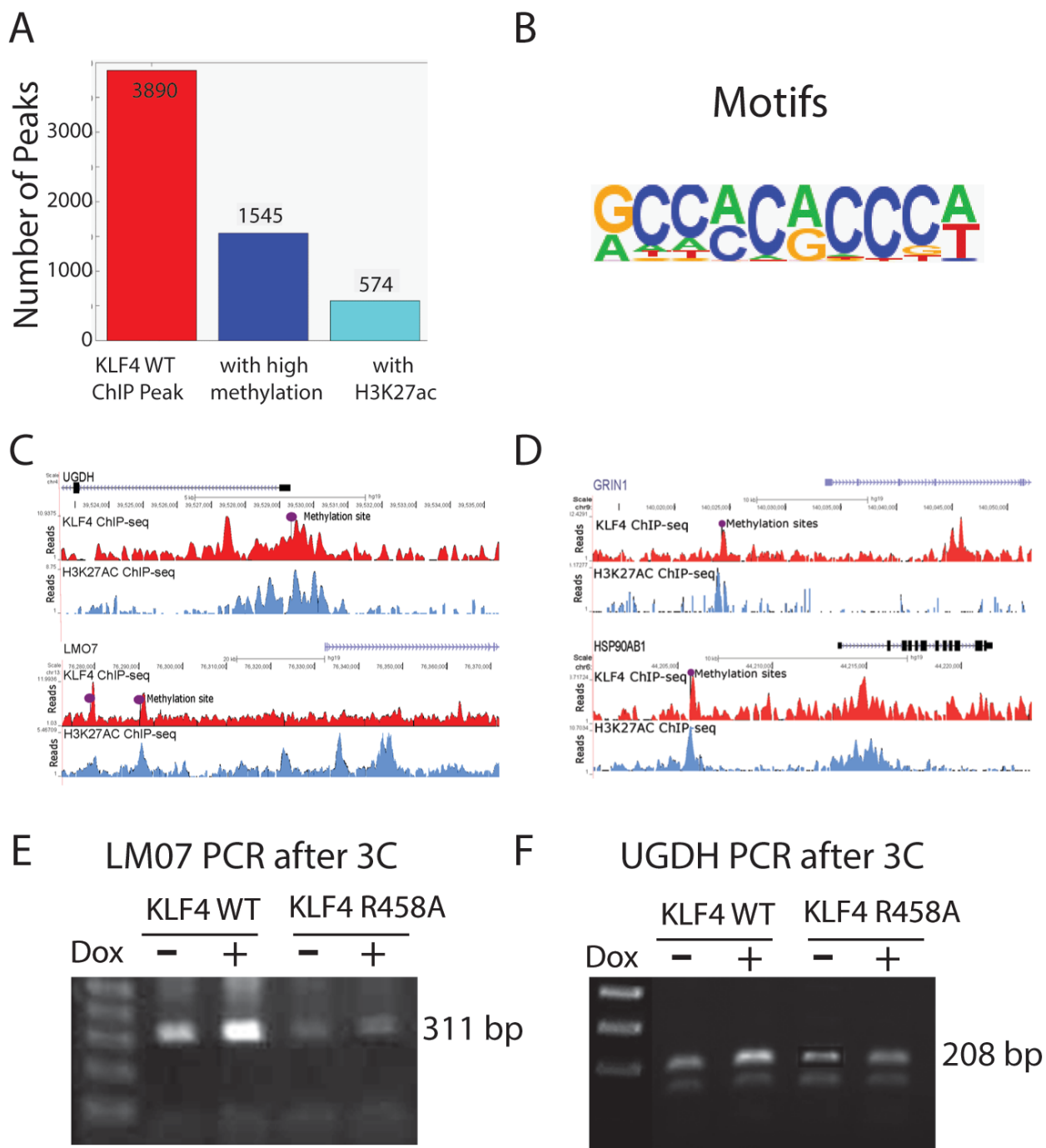
Figure 4-6. Global Analysis of KLF4 binding to mCpGs at enhancer regions to activate gene expression

A. Global analysis of KLF4 ChIP-sequencing, whole genome bisulfite sequencing and the enhancer

marker H3K27ac ChIP-sequencing data identified 547 putative methylated enhancer regions in U87

cells. **B.** Motif analysis of the KLF4 binding of the putative enhancer regions that was not within 10 kb of gene transcription regions (non-promoter regions)**.**

4-7. Discussion

By employing a KLF4 site-specific mutant that lacks KLF4 binding to mCpGs and chromosome conformation capture (3C), we found that mCpG-dependent KLF4 binding at enhancer regions promotes 3D structure of chromatin loop formation with promoters, thereby activating gene expression. Our 3C analysis also found that a couple of gene targets, including BLK, UGDH, LM07, were activated by KLF4 binding to mCpGs at enhancer regions, indicating that KLF4-mCpG interactions at enhancer regions presents an important gene activation mechanism, which has only been sparsely reported until now. Our work, with an emphasis on actual 3D structure formation in individual gene activation, support genome-wide studies showing that mCpG at enhancer regions positively correlates with gene activation. Furthermore, our functional study of BLK revealed that BLK is also involved in tumor cell migration. Thus, our study is in agreement with other works that challenges the traditional view that promoter CpG methylation only plays a passive role in transcriptional silencing, as these studies, including our own, establishes a new concept that DNA methylation plays an active role in transcriptional activation, and this case, at enhancer regions.

This study is a logical extension of our previous study utilizing unique KLF4 site-specific mutagenesis to dissect the biological function of mCpG dependent KLF4 function in GBM cells. There are several technical issues that warrant discussion. First, the expression level of KLF4 protein was within physiological range using our dox-inducible system. For example, we have reported that KLF4 is upregulated ~25-fold during GBM cell reprogramming induced by growth factors [5]. Second, in our previous studies our whole genome bisulfite sequencing in U87 cells has unambiguously proven that KLF4 R458A indeed is defective in binding to mCpG motifs in vivo, corroborating with our early

luciferase assays showing that KLF4 R458A only bound to non-mCpG motifs. Thus, we are confident that the difference we saw in our current 3C analysis between KLF4 WT and KLF4 R458A is attributed to KLF4-mCpG interactions. Third, we used ChIP-PCR to identify differential DNA sequences that complex with KLF4 WT or KLF4 R458A. We are confident that the differential binding of KLF4 WT and KLF4 R458A represented KLF4 binding via mCpG-mediated mechanism. The KLF4 antibody we used recognized the N-terminal of KLF4 and therefore did not interfere with KLF4 R458A. In our previous ChIP-PCR studies with a positive control, we have shown that this antibody was able to equally precipitate both KLF4 WT and KLF4 R458A.

References

1. Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* Elife, 2013. **2**: p. e00726.

2. Wan, J., et al., *Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration.* Elife, 2017. **6**.

3. Dymecki, S.M., J.E. Niederhuber, and S.V. Desiderio, *Specific expression of a tyrosine kinase gene, blk, in B lymphoid cells.* Science, 1990. **247**(4940): p. 332-6.

4. Wu, Y.Y., et al., *Concordance of increased B1 cell subset and lupus phenotypes in mice and humans is dependent on BLK expression levels.* J Immunol, 2015. **194**(12): p. 5692-702.

5. Li, Y.Q., et al., *c-Met signaling induces a reprogramming network and supports the glioblastoma stem-like phenotype.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(24): p. 9951-9956.

6. Galli, R., et al., *Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma.* Cancer Res, 2004. **64**(19): p. 7011-21.

7. Yu, F., et al., *Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion.* Oncogene, 2011. **30**(18): p. 2161-2172.

8. Tilghman, J., et al., *HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells.* Cancer Res, 2014. **74**(11): p. 3168-79.

9. Ying, M., et al., *Kruppel-like factor-9 (KLF9) inhibits glioblastoma stemness through global transcription repression and integrin alpha6 inhibition.* J Biol Chem, 2014. **289**(47): p. 32742-56.

10. Wu, Y., et al., *Regulation of glioblastoma multiforme stem-like cells by inhibitor of DNA binding proteins and oligodendroglial lineage-associated transcription factors.* Cancer Sci, 2012. **103**(6): p. 1028-37.

11. Rath, P., et al., *In Vivo c-Met Pathway Inhibition Depletes Human Glioma Xenografts of Tumor-Propagating Stem-Like Cells.* Transl Oncol, 2013. **6**(2): p. 104-11.

12. Xia, S., et al., *Tumor microenvironment tenascin-C promotes glioblastoma invasion and negatively regulates tumor proliferation.* Neuro Oncol, 2016. **18**(4): p. 507-17.

13. Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* eLife, 2013. **2**: p. e00726.

14. Wan, J., et al., *Methylated cis-regulatory elements mediate KLF4-denpendent gene transactivation and cell migration.* eLife, 2017. **6**: p. e20068.

15. Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.* Bioinformatics, 2011. **27**(11): p. 1571-2.

16. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.

17. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

18. Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.

19. Barbosa, I., et al., *Improved and simple micro assay for sulfated glycosaminoglycans quantification in biological extracts and its use in skin and muscle tissue studies.* Glycobiology, 2003. **13**(9): p. 647-653.

20. Wang, S.D., et al., *EphB2 receptor controls proliferation/migration dichotomy of glioblastoma by interacting with focal adhesion kinase.* Oncogene, 2012. **31**(50): p. 5132-43.

21. Reznik, T.E., et al., *Transcription-dependent epidermal growth factor receptor activation by hepatocyte growth factor.* Mol Cancer Res, 2008. **6**(1): p. 139-50.

Materials and Methods

Reagents and Cell Cultures

All reagents were purchased from Sigma-Aldrich unless otherwise stated. Doxycline (Dox) was diluted to a concentration of 1μg/ml in cell culture medium as a working concentration. The human glioblastoma (GBM) cell lines U87 were originally purchased from ATCC (Manassas, VA). GBM neurosphere culture (HSR-GBM1A) were originally established by Vescovi and colleagues [1] and further characterized by us [2-4]. Both cells lines are free from mycoplasma and authenticated with short tandem repeat (STR) profiling by Johns Hopkins Genetic Resources Core facility using Promega GenePrint 10 system (Madison, WI). U87 cells were cultured in Minimum Essential Media (MEM, Thermo Fisher Scientific, Grand Island, NY) supplemented with sodium pyruvate (1%), sodium bicarbonate (2%), non-essential amino acid (1%) and 10% fetal calf serum (FCS, Gemini Bio-products, West Sacramento, CA). HSR-GBM1A (GBM1A) cells contain CD133+ GBM stem-like cells and form infiltrative orthotropic xenografts that have been extensively characterized by others and our group [5, 6]. GBM1A neurospheres were cultured in DMEM/F12 medium supplemented with epidermal growth factor (EGF) (Peprotech, Rocky Hill, NJ) and fibroblast growth factor (FGF) (Peprotech). Cells were incubated in a humidified incubator containing 5% CO_2/95% air at 37°C, and passaged every 4-5 days.

Lentiviral Transduction

UGDH shRNA lentiviral particles were purchased from Dharmacon (Buckinghamshire, UK). *UGDH* sh#1 clone ID V2LHS-171838 and *UGDH* sh#2 clone ID V3LHS-412961. GBM cells were transfected with virus for 48 hrs prior to puromycin selection (1 μg/ml) as previously described [7].

Quantitative real-time PCR

Total RNA was extracted using RNeasy Mini Kit (Qiagen, Mansfield, MA). After reverse transcription using MuLV reverse transcriptase (Applied Biosystems, Calsbad, CA) and Oligo(dT) primer, quantitative real-time PCR (qRT-PCR) was performed using SYBR Green PCR Mix (Applied Biosystems) and IQ5 detection system (Bio-Rad, Hercules, CA). Primer sequences are listed below. Relative gene expression was normalized to 18S rRNA.

Table 3: RTPCR primers

Gene	Primer (left) 5'-3'	Primer (right) 5'-3'
LIMS2	ACTGCAGCCATGTGATTGAA	AGCTCCGACAGCTTCTTCAG
LM07	TTCCCCTTCAGCTTCACAGT	GCAACAGTGCTTTCGTATGG
PHLDB2	GATCGGAACAAGCGAACATT	GCGATGGGGCTACCATATAA
S100A11	ACAGAGACTGAGCGGTGCAT	CCATCACTGTTGGTGTCCAG
S100A4	GGGTGACAAGTTCAAGCTCA	CTTCCTGGGCTGCTTATCTG
STMN1	AAAGAGAACCGAGAGGCACA	TGGCCAGTACAGTCTTTGGA
NGEF	CATGCTAAAGGCGTCCTCTC	TGTCCAGGATGTTGAGGATG
VCAN	TTTGGAAAGATGAAACCTCGTT	AATTGTCCTTTGCTGATGAGG
ALCAM	CAGAACACGATGAGGCAGAC	TCCATATTACCGAGGTCCTTGT
ANNEX2	ATGACTCCATGAAGGGCAAG	GTGTCGGGCTTCAGTCATCT
CTNNA1	CACCATTGCAGACCATTGC	GCACCACAGCATTCATCAAG
FLNB	GACCTGCTATAGCGCCATTC	CTGGTTGCCTACATGCTTCA
WDR1	GCAAGGTGGTCACAGTGTTC	GGTGTGCATCTTGGATCTTG
LGALS3	GGAATGATGTTGCCTTCCAC	CTGCAACCTTGAAGTGGTCA
KIAA1199	TCAACTATGTGGCGACCATC	TGGCTTTGTGATCCTCAGTG

PLAT	AACAGTCACCGACAACATGC	GTGTACACACCCGGGACATC
MIDN	AGAAACGGCTCCGTAGAAAGG	GACTTGATGTCAGGGTGGACTTC
RABGEF1	CATAAACCGGCAACCAGCAT	GGCCTGTCTTGTGGAAGGTC
PKP3	ATTTTGACGGACTCCGAAAG	ACGTCACCTTCTCCTCCAGA
UGDH	CCAGCCTTTATCTTCGATGG	TCACAAATAAAAATGGCAATCTC

Immunoblot and Immunocytochemistry

Total cellular protein was extracted with radioimmunoprecipitation assay (RIPA) buffer containing protease and phosphatase inhibitors (Millipore, Billerica, MA) and sonicated for 15 seconds; the suspensions were centrifuged at 3,000 g for 10 minutes. SDS-PAGE was performed with 30-60 μg total proteins using 4% to 12% gradient Tris-glycine gels (LI-COR Biosciences, Lincoln, NE). Western blot analysis was performed using the Quantitative Western Blot System, with secondary antibodies labeled by IRDye infrared dyes (LI-COR Biosciences). Antibodies were purchased from: anti-KLF4 (Santa-Cruz, Dallas, TX); anti-RHOC (Cell signaling, Danvers, MA); anti-RAC1 (Cell signaling), anti-UGDH, anti- RABGEF1, anti-NGEF, anti-Brevican (Abcam, San Francisco, CA); anti- PHLDB2 (Bethyl Labs, Montgomery, TX); anti-tenascin C (Millipore), anti-cyclin B1, E1 and D1 (Cell signaling, Danvers, MA) and anti-β-actin.

For staining, GBM cells grown on chamber slides were fixed with 4% paraformaldehyde for 30 min at 4°C and permeabilized with PBS containing 0.1% Triton X-100 for 10 min. The cells were then incubated with primary antibodies at 4^0C overnight and then incubated with appropriate corresponding secondary antibodies conjugated with alexa fluorescent 488 or cy3 for 1hr at room temperature. Slides were mounted with vectashield antifade solution containing DAPI (Vector Laboratories, Burlingame, CA) and observed under fluorescent microscopy. Immunofluorescent images were taken and analyzed using Axiovision software (Zeiss, Germany).

Chromatin immunoprecipitation (ChIP)-PCR

A commercial ChIP-grade anti-KLF4 antibody (H180; Santa Cruz) recognizing the N-terminal region of KLF4 was used for ChIP (DNA-binding domains of KLF4 are located to the C-terminus). Tet-on KLF4 WT and R458A GBM cells were treated with Dox for 48 hr followed by ChIP using the anti-KLF4 antibody and Dynabeads Protein A/G (Thermo Fisher Scientific) according to a protocol described previously [8, 9]. Primers targeting KLF4 binding sites were identified from previous ChIP-Seq analysis [9]. Primers targeting promoter regions lacking KLF4 binding sites were used as a negative control.

Assessment of CpG methylation status by bisulfite sequencing

Sanger bisulfite sequencing was performed as previously described [8]. Purified genomic DNA from GBM cells were treated by EZ DNA Methylation-Direct Kit (Zymo Research, Irvine, CA). After bisulfite conversion, regions of interest were PCR-amplified using Taq polemerase. PCR products were cleaned up and cloned into a TA vector (Thermo Fisher Scientific). Individual clones were sequenced (Genewiz, Cambridge, MA) and aligned with the reference sequence. Primer sequences used are listed in table below

Table 4: Bisulfite Sequencing Primers

Gene	Primer (left) 5'-3'	Primer (right) 5'-3'
PHLDB2	TGTAATTTTAGTATTTTTGGGAGGT	TACAATCTTAACTCATTACAACCCC
RABGEF1	GTTTAGGTTTTTTTTGATGGTGATAG	CAAACTAATCTCAAACTCCTAACCTC
NGEF	CGTTAGGGTCGTCGTAGTT	CGCCTAAAAACTCCCGAAAT
UGDH	TTTTATTATGTTAGTTAGGATGGTTT	ATTTAATTTATTATAACTCTCCCAAAATAC

Genome-wide profiling of gene expression and KLF4 binding in GBM cells

RNA-seq: RNAs from KLF4 WT and R458A-expressing cells (0 hr and 48 hr +Dox) was subjected to Illumina HiSeq next generation sequencing following the standard amplification and library construction protocol provided by the Johns Hopkins Deep Sequencing and Microarray Core Facility. Sequencing was performed using 76-base single-end reads, with 23- to 33-million reads generated from each sample. We first used Tophat2 to map all reads to human genome (hg19) then employed Cufflink to summarize the gene/transcript expression based on mapped reads. An R package, DEGseq, was taken to identify DEGs for $p < 0.001$ between 0 hr and 48 hr in KLF4 WT and R458A cells, respectively.

Whole Genome Bisulfite Sequencing (WGBS) Analysis

We employed the software package bismark [10] to perform WGBS analysis. First, we built bismark reference human genome, then mapped sequence reads onto these specific references. Two files were generated afterwards. The text file includes the summary about total reads, mapping efficiency, total methylated C's in CpG/CHG/CHH context. The other file in the same format was used for next step to extract methylation. Finally, we used bismark2bedGraph followed by coverage2cytosine to achieve the methylated and unmethylated reads of all CpG sites. The β value was calculated for each CpG site as the ratio of number of methylated reads to sum over methylated and unmethylated.

Motifs Analysis

To identify methylated motifs enriched in KLF4 WT-specific peaks, we first used WGBS information to selected all 6-mers including mCpG (β > 0.6), then enumerated these 6-mers to compare their occurrence in 2733 KLF4 WT-specific peaks and all KLF4 binding peaks. The p-values were

calculated based on hypergeometric model to represent the significance of methylated 6-mers' frequencies in KLF4 WT-specific peaks compared to all, followed by multiple-test Bonferroni correction. The 6-mers with $p < 0.01$ were selected to construct the motif logo. The package, MEME (Multiple EM for Motif Elicitation) [11], was used to evaluate motifs significantly over-represented in KLF4 shared peaks, compared to all KLF4 binding peaks.

Gene Ontology Analysis

Gene Ontology analysis [12] was performed for the 116 differential expressed genes (DEGs) up-regulated by WT only, compared to that for total 12,824 genome-wide expressed genes (FPKM > 0.5). The statistical significance of the enrichment was evaluated by p-value based on hypergeometric distribution model. The p-values were then adjusted by multiple-test correction via false discovery rate (FDR). A cutoff of FDR < 0.05 was used to identify significantly enriched GO terms.

Chromatin immunoprecipitation (ChIP)-seq

 A commercial ChIP-grade anti-KLF4 antibody (H180; Santa Cruz) recognizing the N-terminal region of KLF4 (DNA-binding domains of KLF4 are located to the very C-terminus) was used for ChIP. Tet-on KLF4 WT and R458A cells were treated with Dox for 48 hr followed by ChIP using the anti-KLF4 antibody and Dynabeads Protein A/G (Thermo Fisher Scientific) according to a protocol described previously [13]. DNA library construction and sequencing was performed at Johns Hopkins Deep Sequencing and Microarray Core Facility. The antibodies used for ChIP experiments were as follows: anti-KLF4 (Santa Cruz, H-180, sc-20691); anti-H3K27ac (Abcam, ab4279) [14]; anti-H3K27me3 (Millipore, 07-449) [14] and anti-H3K9me3 (Abcam, ab8898) **[14].**

KLF4 ChIP-Seq data were mapped by Bowtie2, followed by MACS 1.4 being used to call peaks with cutoff of $p < 1E-5$. We first obtained binding peaks for KLF4 WT and R458A, respectively. The peaks

identified for both KLF4 WT and R458A at the same locus were referred as shared KLF4 binding peaks. Then we used KLF4 WT as foreground and R458A as background control to call peaks again. The new peaks were marked as KLF4 WT-specific ChIP peaks, only if they were not overlapped with shared ones which had been already identified. Same approach was used to obtain R458A specific binding peaks for which the foreground was KLF4 R458A ChIP-Seq data compared to the background of KLF4 WT ChIP-seq data.

We utilized MACS2 to recognize broad peaks of H3K27ac based on their ChIP-Seq data mapped by Bowtie2. The cutoff of broad peak call was q < 0.1. The same procedure as that for KLF4 ChIP-Seq was taken to distinguish H3K27ac peaks at 0 hr only, at 48 hr only, or shared at both 0 hr and 48 hr, for KLF4 WT and R458A, respectively.

Quantitative measurement of glycosaminoglycans (GAG)

The quantity of GAG was determined using the 1,9 dimethyl-methylene blue (DMMB) binding method with modifications as described by Barbosa et al [15]. Briefly, the cells were digested with 0.2% papain in reaction buffer consisting of 0.1 M NaH2PO4, 5 mM Na2-EDTA, and 5 mM cysteine–HCl pH 6.0, at 60 °C for 6 h. Total cell lysates were concentrated and adjusted to a concentration of 0.1 mg protein in 50 µl before mixing with 200 µl of DMMB solution (40 mM glycine, 40 mM NaCl, 9.5 mM HCl, and 0.0016% DMMB, pH 3.0) in each well of a 96-well microtiter plate. The absorbance was immediately recorded at 525 nm using a plate reader (Perkin Elmer, Waltham, MA). Serial dilutions of shark chondroitin-6-sulphate, ranging from 5 to 25 µg/ml, were used to establish a standard curve, and duplicate wells with 50 µl of papain digestion buffer were used as blanks. Sample concentrations were calculated using linear regression of the standard curves.

Transswell migration assay

Cell migration assays were performed using transwell chambers as we previously described [7, 16]. The upper chamber medium consisted of either neurosphere cell culture medium without EGF/FGF or U87 cell media without FBS, and the lower chamber medium consisted of DMEM with 10% FBS. After 4-24 hours, cells that had migrated through the filter were fixed with Diff-Quick kit (Thermo Fisher Scientific). Cells on the upper side of the transwells were gently wiped off with Q-tips. Cells migrating through the filter were stained with 4'-6-Diamidino-2-phenylindole (DAPI). Migration was quantified by counting cells on five randomly selected fields per transwell in at least three independent experiments [16].

Scratch Assay

GBM cells were grown under 10% FCS medium in 35 mm dishes until confluent. Several scratches were created using a 10 μl pipette tip through the confluent cells. Dishes were washed with PBS for 3 times and cells were grown in 0.1% FCS medium for 24-48 hr. Phase contrast pictures were taken at different time points. The width of the scratch was measured and areas of wound healing were measured and quantified using ImageJ.

Neurosphere formation assays

Viable cells (2×10^3/well or 2×10^4/well) were cultured in 48-well or 6-well plates, respectively. After 7-14 days, neurospheres were fixed in medium with 1% agarose, stained with 1% Wright stain solution and counted by computer-assisted morphometry (MCID software, Cambridge, UK) by measuring the number of neurospheres (>50 μm or >100 μm in diameter, as indicated) in three random fields per well.

Colony formation Assay

 Anchorage-independent tumor cell proliferation was assessed by colony formation in soft agar. UGDH knockdown U87 cells or control-transfected cells were plated 10,000 cells per well in 0.5% agarose medium

on top of regular medium containing 1% agarose and incubated for 2 weeks. The cells were stained blue with 1% Wright stain solution, and the number of colonies larger than 100µm in diameter were determined by computer-assisted image analysis MCID.

Cell Cycle Analysis

Cell cycle was analyzed by flow cytometry on a FACSCalibur (Becton-Dickinson, Mountain View, CA) [17]. After plating in normal medium overnight, U87 cells were synchronized by changing into 0.1% FCS medium for 48 hrs followed by stimulation with 10% FCS for the indicated time points. To harvest, cells were trypsinized and dissociated by pipetting, fixed with 75% ethanol at 4 °C for 30 min. Cells were then incubated with DNase-free RNase at 37 °C for 30 min followed by propidium iodide (100 ng/ml) for 1 h at 37 °C. The percentage of cells at each cell-cycle phase (G1/G0, S and G2/M) was analyzed using CellQuest software (Becton-Dickinson).

Intracranial tumor implantation

All animal protocols were approved by the Johns Hopkins School of Medicine Animal Care and Use Committee. SCID immunodeficient mice received 10,000 viable neurosphere cells in 2 µL PBS by stereotactic injection to the right caudate/putamen (AP = 0 mm, ML = −2.5 mm, DV = −3.0 mm). After 7 weeks, mice were sacrificed and perfused with 4% paraformaldehyde; the brains were removed for histological analysis. Tumor sizes were quantified by measuring maximum tumor volume on hematoxylin and eosin–stained brain coronal sections using computer-assisted morphometry (MCID software) and then applying the formula Volume = (square root of maximum cross-sectional area)^3 [6]. The primary antibodies used for immunofluorescent staining are the following: monoclonal anti-Ki67 (BD Biosciences, Franklin Lakes, NJ), anti-TNC (Millipore), anti-Brevican (Abcam), and anti-laminin (Millipore).

Chromosome Conformation Capture (3C)

U87MG cells were cultured in 15 cm plates to 80% confluence and were fixed with 1% formaldehyde. The nuclear fraction was isolated using a lysis buffer containing 10mM Tris HCl pH 8, 10 mM NaCl, 0.2% NP40 and a protease inhibitor cocktail (Roche). DNA samples were digested with 375 units of HindIII and BgI (NEB) and ligated using T4 DNA ligase (NEB). Samples were then digested with Proteinase K (Qiagen), phenol chloroform extracted and ethanol precipitated prior to PCR analysis. Unligated and ligated samples were analyzed on an agarose gel and a control PCR was performed to confirm successful ligation using control primers.

Statistical analysis

Statistical analysis was performed using Prizm software (GraphPad, La Jolla, CA, www.graphpad.com). Post hoc tests included the Students T-test and Tukey multiple comparison tests as appropriate. All *in vitro* experiments reported here represent at least three independent replications. All data are represented as mean value ± standard error of mean (S.E.); significance was set at $P < 0.05$.

References

1.   Galli, R., et al., *Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma.* Cancer Res, 2004. **64**(19): p. 7011-21.

2.   Yu, F., et al., *Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion.* Oncogene, 2011. **30**(18): p. 2161-2172.

3.   Tilghman, J., et al., *HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells.* Cancer Res, 2014. **74**(11): p. 3168-79.

4.   Ying, M., et al., *Kruppel-like factor-9 (KLF9) inhibits glioblastoma stemness through global transcription repression and integrin alpha6 inhibition.* J Biol Chem, 2014. **289**(47): p. 32742-56.

5.   Wu, Y., et al., *Regulation of glioblastoma multiforme stem-like cells by inhibitor of DNA binding proteins and oligodendroglial lineage-associated transcription factors.* Cancer Sci, 2012. **103**(6): p. 1028-37.

6.   Rath, P., et al., *In Vivo c-Met Pathway Inhibition Depletes Human Glioma Xenografts of Tumor-Propagating Stem-Like Cells.* Transl Oncol, 2013. **6**(2): p. 104-11.

7.   Xia, S., et al., *Tumor microenvironment tenascin-C promotes glioblastoma invasion and negatively regulates tumor proliferation.* Neuro Oncol, 2016. **18**(4): p. 507-17.

8.   Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* eLife, 2013. **2**: p. e00726.

9.   Wan, J., et al., *Methylated cis-regulatory elements mediate KLF4-denpendent gene transactivation and cell migration.* eLife, 2017. **6**: p. e20068.

10.   Krueger, F. and S.R. Andrews, *Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.* Bioinformatics, 2011. **27**(11): p. 1571-2.

11.   Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.

12.   Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

13.   Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* Elife, 2013. **2**: p. e00726.

14.   Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.

15.   Barbosa, I., et al., *Improved and simple micro assay for sulfated glycosaminoglycans quantification in biological extracts and its use in skin and muscle tissue studies.* Glycobiology, 2003. **13**(9): p. 647-653.

16.   Wang, S.D., et al., *EphB2 receptor controls proliferation/migration dichotomy of glioblastoma by interacting with focal adhesion kinase.* Oncogene, 2012. **31**(50): p. 5132-43.

17.   Reznik, T.E., et al., *Transcription-dependent epidermal growth factor receptor activation by hepatocyte growth factor.* Mol Cancer Res, 2008. **6**(1): p. 139-50.

Conclusion

The role of DNA methylation in gene expression has previously been described as a repressed mark by disrupting transcription factor (TF)-DNA interactions directly or indirectly via the recruitment of methyl-CpG-binding domains (MBD). This traditional view has been challenged by several discoveries. First, a handful of TFs, especially C2H2 zinc finger proteins, were anecdotally identified to specifically bind to methylated DNA sequences [1-4]. Second, earlier in 2013, Mann et al using methylated-CpG microarrays found that two out of 12 basic leucine zipper TFs bind to mCpG-containing DNAs [5]. Third, a mass spectrometry-based screening approach has revealed that some proteins specifically bind to methylated DNA [6]. Finally, we have employed a protein microarray-based approach and found that 42 human TFs, including KLF4, preferentially recognize mCpG-containing DNA motifs [7]. Moreover, specific mCpG-dependent TF-DNA interactions were found to transactivate downstream gene expression. All these findings point out that some TFs might serve as "epigenetic readers" and regulate gene expression and cell phenotypes in a methylation-dependent manner. All this suggests that current dogma regarding DNA methylation misses an important dimension of the epigenetic landscape by ignoring mCpG-dependent TF-DNA interactions. The biological function of such TF-mCpG binding is also largely unknown.

Using zinc finger krueppel-like factor 4 (KLF4) as the first candidate, we investigated the function of TF-mCpG binding in biological relevant systems, e.g. human GBM cells, and obtained strong evidence that KLF4 binds to mCpG at *cis*-element regions and activate GBM adhesion and migration. Employing KLF4 site-specific mutant (KLF4R458A) that lacks KLF4 binding ability to mCpG but retains its binding to canonical, non-methylated CpGs, we discovered that this novel KLF4-mCpG interaction could activate gene expression. We performed follow up studies aimed at determining the cellular function of KLF4-mCpG interactions in GBM cells. Our studies revealed that

KLF4 mediates brain tumor cell migration by binding to mCpGs in the *cis*-regulatory elements and activating cell motility gene expression, e.g. *RHOC, RAC1* etc.

The mechanism adopted by brain tumor migration/invasion is an intricate program that recalls what takes place in carcinoma cells during metastasis initiation. Yet, no cancer driver genes have been implicated in tumor metastasis [8], suggesting that other mechanisms, including epigenetic regulation, play a role in tumor metastasis. DNA methylation-mediated gene silencing has been shown to regulate tumor cell migration and invasion [9-11]. Clinical studies from human brain tumor samples indicate that IDH1 mutated gliomas are relatively more invasive than IDH1 wild-type gliomas [12, 13], supporting an active role for DNA methylation in GBM invasion. We identified and dissected the mCpG-dependent activation of cell motility genes, including *RHOC, RAC1 and UGDH,* in GBM cells. All these in-depth studies of KLF4-mCpG interactions focused on KLF4 binding to mCpG at promoter regions.

In summary, our studies promise to significantly advance our understanding of epigenetic mechanisms in tumor biology, which will greatly expand the epigenetic landscape in a novel direction. Determining DNA methylation-mediated gene transactivation mechanisms at promoter and enhancer regions and their impact on GBM migration/invasion would create a totally new dimension for interpreting current genome-wide methylation data in cancer. This work will provide a roadmap to studying this novel gene regulatory mechanism in other cancer subtypes, as well as characterizing other mCpG-binding TFs.

References

1.   Filion, G.J., et al., *A family of human zinc finger proteins that bind methylated DNA and repress transcription.* Mol Cell Biol, 2006. **26**(1): p. 169-81.

2.   Rishi, V., et al., *CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes.* Proc Natl Acad Sci U S A, 2010. **107**(47): p. 20311-6.

3.   Sasai, N., M. Nakao, and P.A. Defossez, *Sequence-specific recognition of methylated DNA by human zinc-finger proteins.* Nucleic Acids Res, 2010. **38**(15): p. 5015-22.

4.   Serra, R.W., et al., *A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype.* Elife, 2014. **3**.

5.   Mann, I.K., et al., *CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo.* Genome Res, 2013. **23**(6): p. 988-97.

6.   Spruijt, C.G., et al., *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.* Cell, 2013. **152**(5): p. 1146-59.

7.   Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* Elife, 2013. **2**: p. e00726.

8.   Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.

9.   Esteller, M., *Epigenetic gene silencing in cancer: the DNA hypermethylome.* Hum Mol Genet, 2007. **16 Spec No 1**: p. R50-9.

10.   Martinez, R. and M. Esteller, *The DNA methylome of glioblastoma multiforme.* Neurobiol Dis, 2010. **39**(1): p. 40-6.

11.   Waha, A., et al., *Epigenetic silencing of the protocadherin family member PCDH-gamma-A11 in astrocytomas.* Neoplasia, 2005. **7**(3): p. 193-9.

12.   Baldock, A.L., et al., *Invasion and proliferation kinetics in enhancing gliomas predict IDH1 mutation status.* Neuro Oncol, 2014. **16**(6): p. 779-86.

13.   Sabit, H., et al., *Characterizing invading glioma cells based on IDH1-R132H and Ki-67 immunofluorescence.* Brain Tumor Pathol, 2014.

CURRICULUM VITAE FOR Ph.D. CANDIDATES

The Johns Hopkins University School of Medicine

Olutobi Z. Oyinlade   September 8th 2017

Educational History

Ph.D expected   2017   Johns Hopkins School of Medicine
                                                                       Biochemistry, Cellular and Molecular
                                                                       Biology Program. Department of
                                                                       Pharmacology and Molecular Sciences

B.S.   2010   College of Saint Elizabeth
                                                                       Department of Chemistry

Other Professional Experience

Johns Hopkins School of Medicine Post Baccalaureate Research 2010-2012
Merck Research Laboratories Summer Intern 2009-2010

Scholarships and Fellowships

Scientist Mentoring and Diversity Program Scholar, Johnson & Johnson and Amgen   2016-2017
National Academies of Sciences Ford Foundation Pre-Doctoral Fellow   2013-2016
Johns Hopkins School of Medicine PREP Scholar   2010-2012
United Negro College Fund/Merck Science Initiative Fellowship   2009-2010
Minorities for Academic Careers Fellowship   2009-2010
New Jersey Commission on Higher Education Outstanding Academic Achievement Award   2010
Pi Mu Epsilon-National Mathematics Honor Society Outstanding Academic Achievement Award 2010
Kappa Gamma Pi - St. Catherine National Achievement Award   2009

Publications

Olutobi Z Oyinlade, Shuang Wei, Bachchu Lal John Laterra Heng Zhu, C. Rory Goodwin, Shuyan Wang, Ding Ma, Jun Wan, Shuli Xia. Targeting UDP-α-D-glucose 6-dehydrogenase inhibits glioblastoma growth. (Submitted to Oncogene July 2017)

Olutobi Oyinlade*, Shuang Wei*, Sheng Liu, Shuyan Wang, Ding Ma, Kai Hammers Jiang Qian, Heng Zhu, Jun Wan and Shuli Xia. Chromosome conformation capture (3C) analysis reveals KLF4 binds to methylated CpG at enhancer regions to activate gene expression. (In preparation)

Jun Wan, Yijing Su, Qifeng Song, Brain Tung, **Olutobi Oyinlade,** Sheng Liu, Mingyao Ying, Guo-li Ming, Hongjun Song, Jiang Qian, Heng Zhu, Shuli Xia, Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. Elife 2017

C. Rory Goodwin, Prakash Rath, **Olutobi Oyinlade**, Hernando Lopez, Salman Mughal, Shuli Xia, Yunqing Li, Harsharan Kaur, Xin Zhou, Sandra Ho, Alessandro Olivi, and Bachchu Lal. Crizotinib and Erlotinib inhibits growth of c-Met$^+$/EGFRvIII$^+$ primary human glioblastoma xenografts. (Submitted to Translational Oncology, February 2017)

Ijeoma Uzoma, Jianfei Hu, Eric Cox, Shuli Xia, **Olutobi Z Ajala** , Jianying Zhou, Hee-Sool Rho, Catherine Guzzo, Corry Paul, Rory Goodwin, Heng Zhu et al. Global Identification of SUMO Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration. (Submitted to Molecular Systems Biology, August 2016, Under Revision)

Olutobi Z Ajala, Cedric D Moore, Heng Zhu, Applications in high-content functional protein microarrays, Current Opinion in Chemical Biology, Volume 30, February 2016, Pages 21-27, ISSN 1367-5931

Prakash Rath, Bachchu Lal, **Olutobi Z Ajala,** Yunqing Li, Shuli Xia, Jin Kim, John Laterra, In Vivo c-Met Pathway Inhibition Depletes Human Glioma Xenografts of Tumor-Propagating Stem-Like Cells, Translational Oncology, Volume 6, Issue 2, April 2013, Pages 104-111, IN1, ISSN 1936-5233

Presentations

Olutobi Oyinlade et al. UGDH is Required for the KLF4 mCpG dependent increase in GBM cell migration. UMBC Biotech Symposium. Baltimore, MD. January 2017 (Awarded for best talk)

Olutobi Oyinlade et al. UGDH is Required for the KLF4 mCpG dependent increase in GBM cell migration. Johnson and Johnson Consumer. Spring House, PA. September 2016

Olutobi Ajala et al. Novel Role of Histone Deacetylase HDAC1 as a methyl binding protein. National Academies of Sciences Ford Foundation Conference. Washington, DC. September 2014

Service and Leadership

Founder, Women Empowered by Grace (WEG)	2014-Present
Mentor, New Jersey Educational Opportunity Program	2009-Present
President, American Chemical Society, College of St. Elizabeth	2009-2010
Student Academic Life Representative, CSE Student Government Association	2009-2010