Deciphering the mechanism of G9a spreading genome-wide

Dmytro Yevstafiev

This thesis is submitted as a partial fulfillment of the M.Sc. program in Cellular and Molecular Medicine.

Submitted September 29th, 2014

Department of Cellular and Molecular Medicine

Faculty of Medicine

University of Ottawa

Ottawa, Ontario, Canada

Abstract

The cell differentiation process is associated with activation and repression of different genes, whereby the formation of heterochromatin is mediated by spreading of repressor proteins along large chromatin domains. Some of these proteins are methyltransferases, including GLP and G9a that are implicated in the addition of mono- and dimethyl groups to lysine 9 at Histone 3. Despite extensive research the exact mechanism of binding and spreading of G9a and GLP is unclear. To better understand the molecular mechanisms through which G9a and GLP bind to chromatin we tested the *in vivo* binding of a mutant G9a that is unable to bind to H3K9^{me2} histone marks via its Ankyrin domain. Murine erythroleukemia (MEL) cell line with expression of mutant G9a was generated using recombinant DNA technologies; G9a binding targets genome-wide were detected by the analysis of ChIP-sequencing data. We validated ChIP-sequencing data providing a reliable tool to visualize G9a targets in MEL cells. We also found that G9a Ankyrin mutant bound to all tested regions suggesting that the Ankyrin domain is not the only factor that contributes to the binding of G9a on chromatin *in vivo.*

Table of Contents

List of Figures

List of abbreviations

β	Beta, β-globin adult gene
β maj	Beta major, mouse β-globin adult gene
A#	Alanine, location
ADP	Adenosine Diphosphate
Ac	Acetyl group
Ank	Ankyrin domain
ATP	Adenosine Triphosphate
bp	Base pair
CDS	Coding sequence
ChIP	Chromatin immunoprecipitation
ddH2O	Double-distilled water
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DTT	Dithiothreitol

E_{50}	50% of maximal effect
E_{max}	Maximal effect
FBS	Fetal bovine serum
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H4	Histone 4
HAT	Histone acetyltransferase
HDAC	Histone Deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HKDM	Histone lysine demethylase
HKMT	Histone lysine methyltransferase
HP1	Heterochromatin protein 1
KD	Knockdown
K#	Lysine, location

K_d	Dissociation constant
K_i	Affinity constant
KMTs	Lysine methyltransferases
LB	Luria-Bertani broth
LCR	Locus control region
Me1-3	Methyl group (mono-, di-, tri-)
MCS	Multiple cloning site
MEL	Murine erythroleukemia cell line
MEL-TR	MEL cells expressing the tetracycline repressor
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
mRNA	Messenger ribonucleic acid
N-liq	Liquid nitrogen
PBS	Phosphate buffered saline
PIC	Protease Inhibitor Cocktail
PTMs	Post-translational modifications
R#	Arginine, location

RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
RT-qPCR	Real time quantitative polymerase chain reaction
SDS-Page	Sodium Dodecyl Sulfate Polyacrylamide
SET	Set domain
shRNA	Small hairpin RNA
TetO2	Tetracycline operator 2
TFIIH	Transcription factor II H
UCSC	University of California, Santa Cruz
W#	Tryptophan, location
W844A	Substitution of 844 Tryptophan to Alanine (G9a)

Acknowledgments

I would like to thank Dr. Chandra-Prakash Chaturvedi for his time, guidance and support during my first steps in science. He was my first mentor and I enjoyed the time I spent with him in discussions and study.

Also I would like to acknowledge the importance of support from my labmates. These amazing people were always a beacon of hope and example for me. They showed me this whole new world of science and taught me how to succeed here.

I thank my Thesis Advisory Committee members Dr. Michael Rudnicki and Dr. William Stanford for their guidance and faith in me and my potential to develop my skills.

I also thank my supervisor Dr. Marjorie Brand for giving me a chance to prove myself in her laboratory.

Special thanks to my parents who made it possible for me to come to Canada and realize my dream to study and contribute to the science.

I hope my scientific path will continue and I will be able to achieve my goal to become a successful scientist.

Introduction

The structure and functions of any cell are directly dependent on non-covalent interaction kinetics of a wide variety of intra- and extracellular molecules. The properties of such interactions determine the organization of functional complexes and structural compartments of the cell. In other words, physical interaction of specific macromolecules determines their biological functions. For example, the structure of cytoskeleton depends on the association of dozens of different protein types (Wickstead and Gull 2011); cell migration is caused by formation and breakdown of certain bonds between different adhesive molecules on the cells' surface (Raghavan, Desai et al. 2010); cell's response to the variety of external stimuli triggers an activation cascade of messenger molecules (Bongrand and Malissen 1998), which will bind to corresponding receptors throughout the cell, changing the gene expression rates to adapt to the new conditions.

In biological studies a molecule, ion or protein, which is able to form an intermolecular complex, bind to proteins or DNA double helix, is called a ligand (Teif and Rippe 2010). Usually, ligand binding is based on non-covalent attractive intermolecular forces such as dipole-dipole, ion-dipole and Van der Waals interactions. For example, heme groups, which are ligands embedded in hemoglobin and myoglobin, are able to bind molecular

oxygen using ion-induced dipole forces. Associations involving non-covalent interactions are reversible. Besides the direct intermolecular forces, the binding in solutions is also driven indirectly by solvent properties (Baron, Setny et al. 2010) and macromolecular crowding.

The interactions of ligands with their receptors are characterized by the binding affinity. Efficacy of such binding is based on the affinity value, which is, in turn, dependent on strength of intermolecular forces. The higher affinity, the longer receptor will maintain the ligand at its binding site. The characteristics such as ligand affinity (K_i) and dissociation constant (K_d) are essential to its efficacy in term of binding to the substrate. The ligands with different characteristics but for the same substrate will compete with each other. The example is a carbon monoxide poisoning resulting from competition of carbon monoxide with oxygen in binding to hemoglobin (Prockop and Chichkova 2007).

In biological systems the prolonged binding of a ligand to the corresponding receptor protein has a physiological importance and usually involves conformational change of the target receptor. Agonist ligands for the specific receptor are able to alter its function and trigger a physiological response. The type of signaling cascade and optimal concentration needed to activate it characterize the agonist ligands. Ligands which bind the receptor

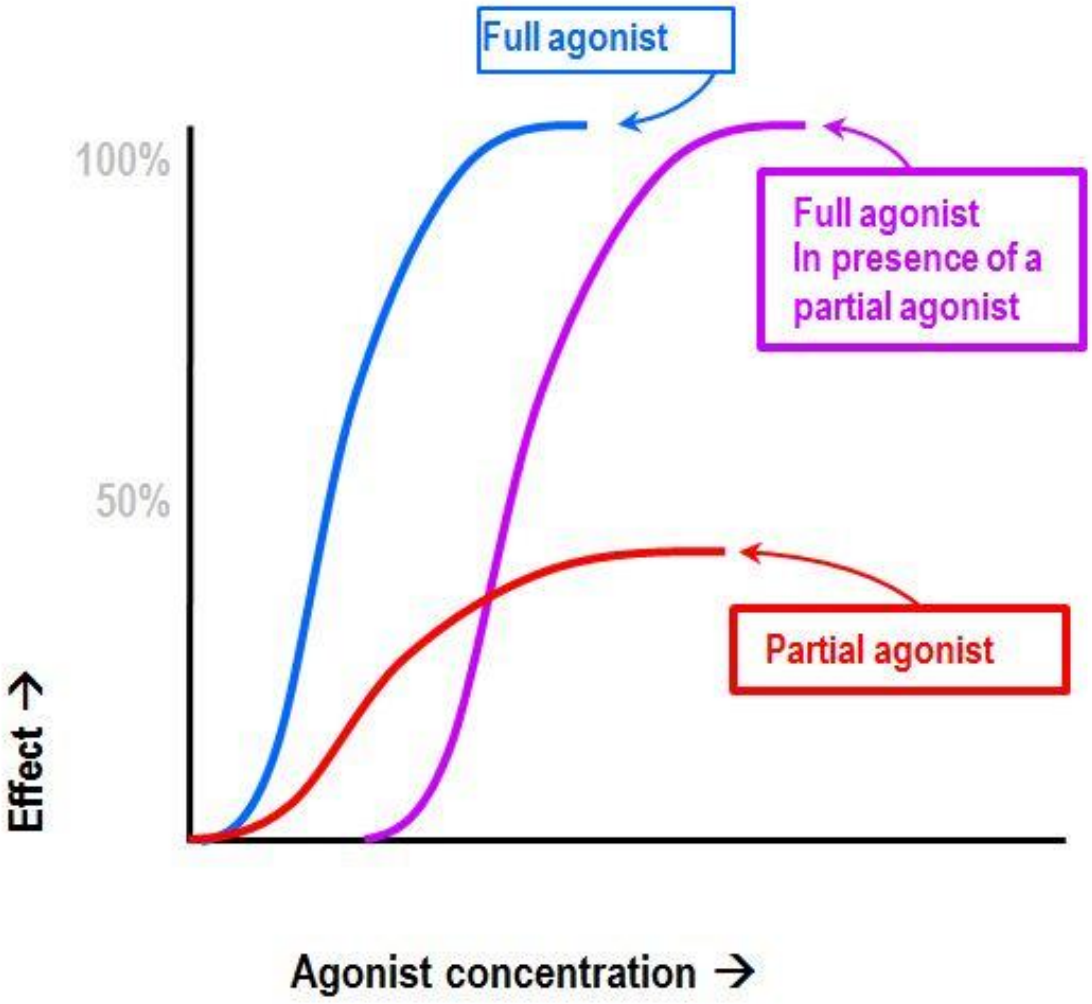


Figure 1. Dose-response curve for full and partial agonists in terms of stimulating a receptor.

Full agonist is able to completely stimulate its receptor, while partial agonist only partially. Competition of full and partial agonists for the receptor leads to the increased values of E_{max} and EC_{50}.

but do not activate the receptor response are called antagonists. The higher affinity implies the relatively lower concentration of ligand is enough to occupy the binding sites on receptors and trigger signaling cascade. Full agonists are able to completely stimulate the receptor, while partial agonists only partially. Combination of full and partial agonists for the same receptor will cause partial agonist to act as a competitive antagonist. This will require higher concentration of a full agonist to reach E_{max} and EC_{50}, values representing concentration of a compound whereby maximal or 50% of maximal effect is observed (Figure 1).

As the other biological processes, the epigenetic regulation of gene transcription is based on ligand/receptor interactions. Molecular basis of epigenetic regulation is a control of gene expression without altering the sequence of DNA (Jaenisch and Bird 2003). This control is achieved through the attraction, dissociation or modification of transcription factors and chromatin proteins associated with DNA. The compaction of DNA into chromatin is a key element for the regulation of eukaryotic gene transcription, replication and repair of DNA (Teif and Bohinc 2011).

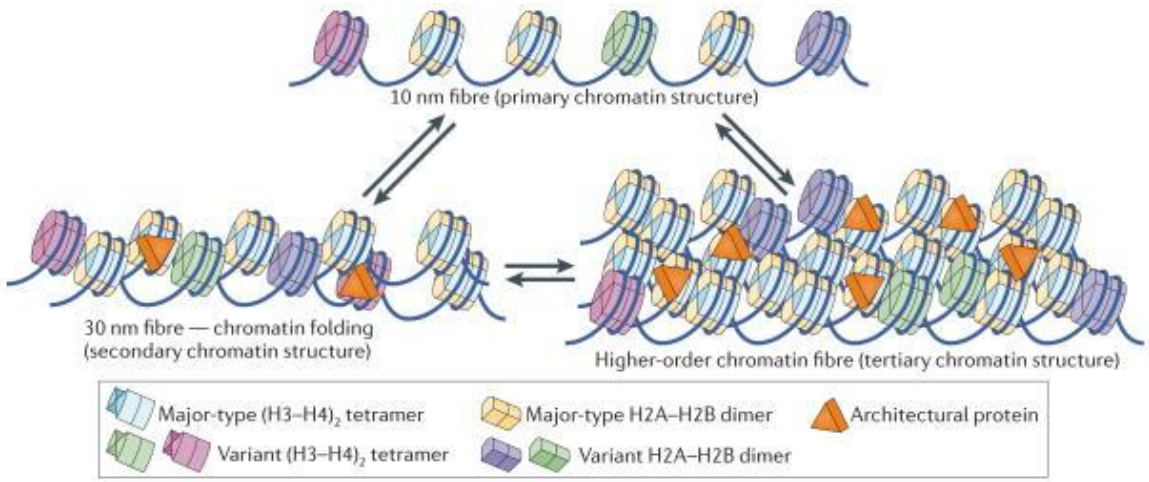


Figure 2. Levels and variations of chromatin organization.

The figure depicts the ability of nucleosomes to incorporate different histone variants and form primary, secondary and tertiary structures. Primary structure called "beads-on-a-string" comprises separate nucleosomes connected with the linker DNA. Histone 1 proteins connect and stabilize a couple of primary structure chains to from a defined chromatin fiber. A large-scale tertiary structures represent the highest level of chromatin compaction and arise through a variety of intermolecular interactions (Luger, Dechassa et al. 2012).

Histones

The basic structural proteins involved in DNA packaging and ordering are histones. They are present in eukaryotic cell nuclei and form a tight complex with genomic DNA termed nucleosome. The nucleosome comprises a histone octamer formed by H3, H4, H2A and H2B subunits and ~145-147 bp of DNA wrapped around each core. Nucleosomes are connected together with linker DNA and form a so-called "beads-on-a-string" primary structure of chromatin. These linear arrays fold into more complicated secondary structures with the help of linker histones H1 and other architectural proteins to form a defined chromatin fiber. Further compaction into large-scale tertiary structures is achieved via a variety of intermolecular interactions. Nucleosomes vary in their histone composition: incorporation of different histones variants and post-translational modifications (PTMs) of histone tails affect the stability and dynamics of high-order compaction of chromatin (Luger, Dechassa et al. 2012).

In addition to the variety of nucleosome components, there are other factors affecting chromatin structure. For example, nucleosome-binding proteins which recognize specific histone tail modifications, ATP-dependent chromatin remodelers and other architectural proteins (Vincent, Kwong et al. 2008). Although the complexity of chromatin structure prevents the access of most transcriptional factors to the DNA template, it is very dynamic. The

changes may be applied at the different levels of chromatin. The small scale but the most frequent change is the establishment of an active promoter prior to gene transcription. The larger scale modification is formation of special chromatin structure such as chromosome centromeres and telomeres. This dynamics allows cells to express genes precisely in temporal and spatial scale. Importantly, such precision defines which genes will be expressed in the specific cell types, while others will be silenced.

Histone modifications

Post-translational covalent modifications (PTMs) of the histone tails (unfolded N- and C-terminal polypeptide chains) play a key role in regulating the access to DNA for the vide variety of nuclear proteins related to transcription of genes, DNA replication and repair. The first histone modifying enzymes that were involved in the regulation of transcription were histone acetyltransferases (HATs) (Brownell, Zhou et al. 1996) and histone deacetylases (HDAC) (Taunton, Hassig et al. 1996) of the lysine residues. Since then many more modifications were identified and described, including lysine methylation, SUMOylation, ubiquitination, crotonylation, butyrylation and propionylation (Musselman, Lalonde et al. 2012). Besides the lysine it appeared that other histone residues may be modified: arginine (methylation, citrullination and ADP-ribosylation), serine and threonine (phosphorylation and glycosylation of both) (Musselman, Lalonde et al.

2012). With the expanding knowledge of the role of different PTMs and enzymes responsible for their writing, erasing and reading, histone code hypothesis arose. This hypothesis states that the combinatorial action of several modifications on one or more histone tails specify unique downstream function (Strahl and Allis 2000). Basically, histone tails act as the antennas stabilizing (or prevent binding of) nuclear proteins on the specific location of nucleosome. Their structure and the total charge define the affinity value for these proteins implying the length of residency and the amount of enzymatic activity. For example, the transfer of acetyl functional groups on histone tails neutralizes the positive charge on histones which decreases their interaction with negatively charged phosphate groups of DNA. As a result, condensed chromatin gains more relaxed conformation which is more accessible to the transcriptional machinery and chromatin-modifying enzymes. The further sequential modification results in a unique platform for specialized enzyme complexes. The process of histone mark establishment is very dynamic and based on the competition between writers that introduce the marks and erasers that remove them.

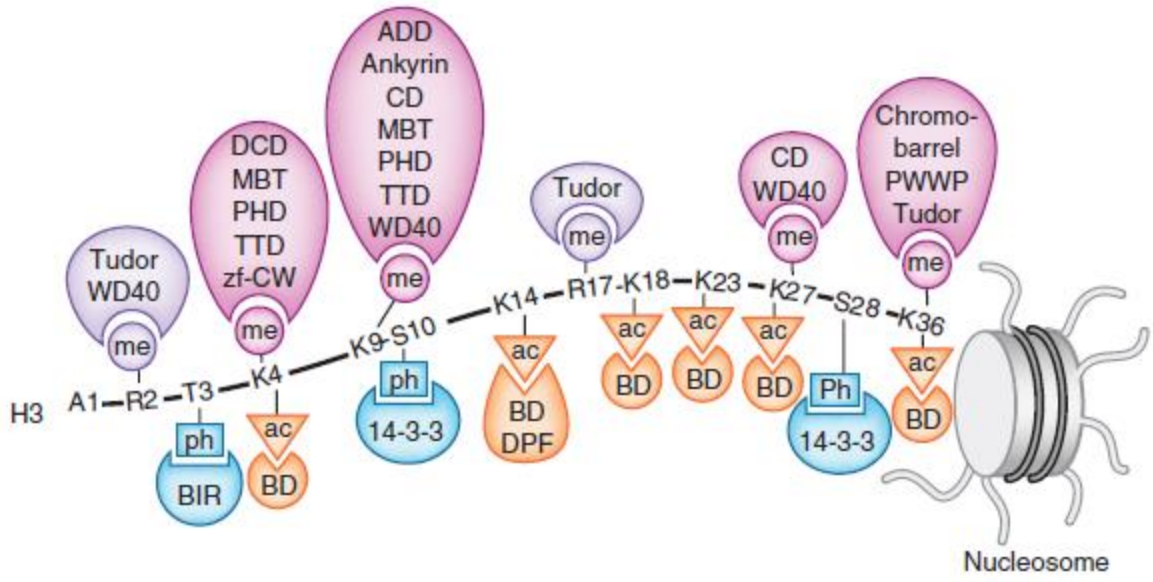


Figure 3. The examples of H3 residues modification and the readers (domains) able to recognize and bind them.

(Musselman, Lalonde et al. 2012).

Acronym definitions:

Modifications: me – methylation; ac – acetylation; ph – phosphorylation. Binding domains: WD40 - beta-transducin repeat domain; DCD - double chromodomain; MBT - malignant brain tumor repeat domain; PHD - plant homeodomain; TTD - tandem Tudor domain; zf-CW – CW-type Zinc Finger; ADD - ATRX, DNMT3, DNMT3L domain; CD – chromodomain; PWWP - Pro-Trp-Trp-Pro domain; BIR - chromosomal passenger complex subunit Survivin domain; BD – bromodomain; DPF - double PHD finger.

Lysine methylation

Methylation is one of the most versatile modifications among other PTMs. Lysine and arginine residues can be mono-, di- and trimethylated. Although addition of methyl groups does not change the charge of modified residue, it does alter the size of the residue and its hydrophobic potential. The common methylation sites of lysine on N-terminal tail of H3 are K4, K9, K26, K27, K36 and C-terminal K79. Arginine residues, in turn, can be mono- and dimethylated symmetrically or asymmetrically (Rme2s or Rme2a) at H3R2, H3R8, H3R17, H3R26, H4R3, H2AR11 and H2AR29 sites. The number of identified readers able to recognize and bind methyl lysine marks is increasing. This group includes ADD (ATRX-DNMT3-DNMT3L), ankyrin, bromo-adjacent homology (BAH), chromo-barrel, chromodomain, double chromodomain (DCD), MBT (malignant brain tumor), PHD (plant homeodomain), PWWP (Pro-Trp-Trp-Pro), tandem Tudor domain (TTD), Tudor, WD40 and the zinc finger CW (zf-CW) (Musselman, Lalonde et al. 2012) (Figure 3). The binding of methyllysine with these readers is achieved through the interaction with an aromatic cage formed by 2 to 4 residues. Surrounding the extended chain of methylated lysine, the aromatic residues stabilizing the complex via cation-π, hydrophobic and Van der Waals interactions.

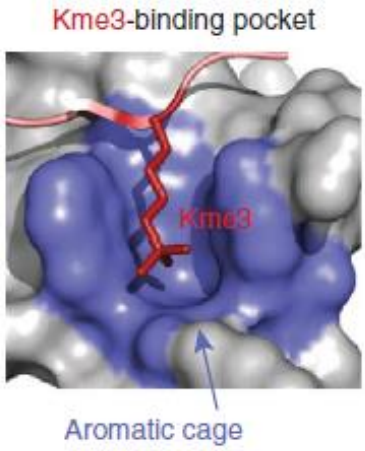


Figure 4. The example of the Kme3 binding pocket.

The figure shows the accommodation of trimethyllysine by an aromatic cage of the reader (Musselman, Lalonde et al. 2012).

The size of the pocket determines its ability to accommodate mono-, di- or trimethyllysines. The ability of different methyllysine readers to stabilize on the histone tail may be impaired by surrounding residues. On the other hand, some of them are more specific while recognizing surrounding modified or unmodified residues.

The chromodomains usually prefer trimethylated lysines, such as ones of Polycomb and HP1 proteins which recognize H3K27me3 and H3K9me3 modifications (Lachner, O'Carroll et al. 2001, Blus, Wiggins et al. 2011). Methylation of H3K9 and H2K27 is mostly related to gene silencing and formation of facultative (transcriptionally silent genes) and constitutive heterochromatin (repetitive DNA elements) (Peters, Kubicek et al. 2003). Different marks have been shown to be associated with different states of chromatin. For example, repressed regions of euchromatin are enriched in H3K27me3 and H3K9me2 marks, while H3K9me3 and H3K27me1 — pericentromeric regions (Rice, Briggs et al. 2003). The actual formation and propagation of

heterochromatin is carried out by recruitment of HP1 proteins to H3K9me3 marks.

G9a and GLP methyltransferases

Histone methylation had been extensively studied since the first discovery of the evolutionary conserved SET domain (Rea, Eisenhaber et al. 2000) responsible for methyltransferase activity. To date, most of the identified histone lysine methyltransferases (HKMTs) contain a SET domain. Acting together with histone lysine demethylases (HKDMs) HKMTs maintain dynamic pattern of histone tail methylation.

G9a/EHMT2 and its homologue GLP/EHMT1 are members of the Suv39h subgroup of SET domain-containing molecules (Shinkai and Tachibana 2011). The primary targets of their catalytic activity (mono- and dimethylation of lysine residues) are H3K9 and H3K27 (Tachibana, Sugimoto et al. 2002, Chaturvedi, Hosey et al. 2009). Besides the histone methylation G9a and GLP are capable to methylate other non-histone proteins, including self-methylation. GLP and G9a possess similar domain composition (Figure 5) and substrate specifications that could make them interchangeable in their functionality. The knockdown studies of either G9a or GLP demonstrated a severe decrease in the level of H3K9me1-2 (Tachibana, Ueda et al. 2005),

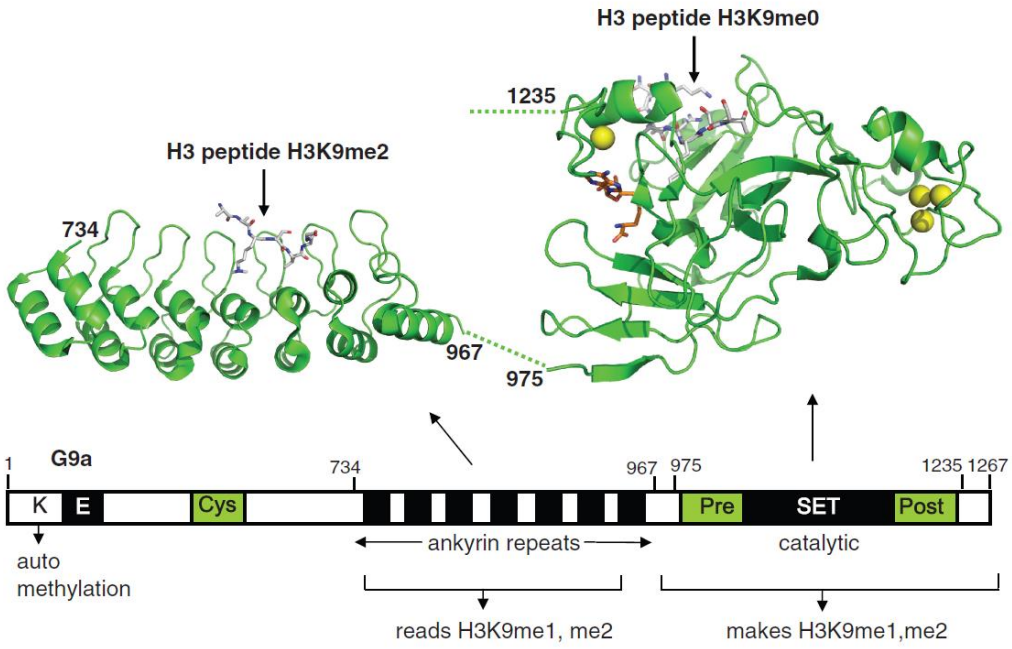


Figure 5. G9a/GLP methyltranferase domain organization.

G9a and GLP consists of a catalytic SET domain (addition of mono- and dimethyl groups to H3K9 and H3K27, also responsible for heteromeric complex formation), the product-binding Ankyrin repeat domain and N-terminal auto-methylation site resembling H3K9. (Collins and Cheng 2010).

H3K27me1-2 (Chaturvedi, Hosey et al. 2009) and heterochromatin protein 1 (HP1) (Collins and Cheng 2010) binding on euchromatic regions. However, the double knockout of G9a and GLP did not further decrease the level of methylation, implying their inability to compensate for each other's function. Furthermore, studies on methyltransferase-deficient mutants demonstrated that the catalytic activity of GLP (but not that of G9a) is dispensable for methyltransferase function *in vivo* (Tachibana, Matsumura et al. 2008). It was also shown that G9a and GLP preferentially form a heteromeric complex via their SET domains in different human and mouse cells. To date studies indicated that G9a and GLP are unable to function independently and exist predominantly as a stoichiometric complex in the cells, which is essential for methylation *in vivo* (Tachibana, Ueda et al. 2005, Tachibana, Matsumura et al. 2008). Furthermore, the knockout or knockdown of GLP leads to the concomitant reduction of G9a protein levels in the cell (Ueda, Tachibana et al. 2006). Together, these data imply a structural role of the cross-talk between these methyltransferases.

Generally, H3K9me2 marks appear on the extended islands of modified nucleosomes, sometimes covering mega-base long chromatin domains which are poised for heterochromatinization – recruitment of HP1 proteins to the

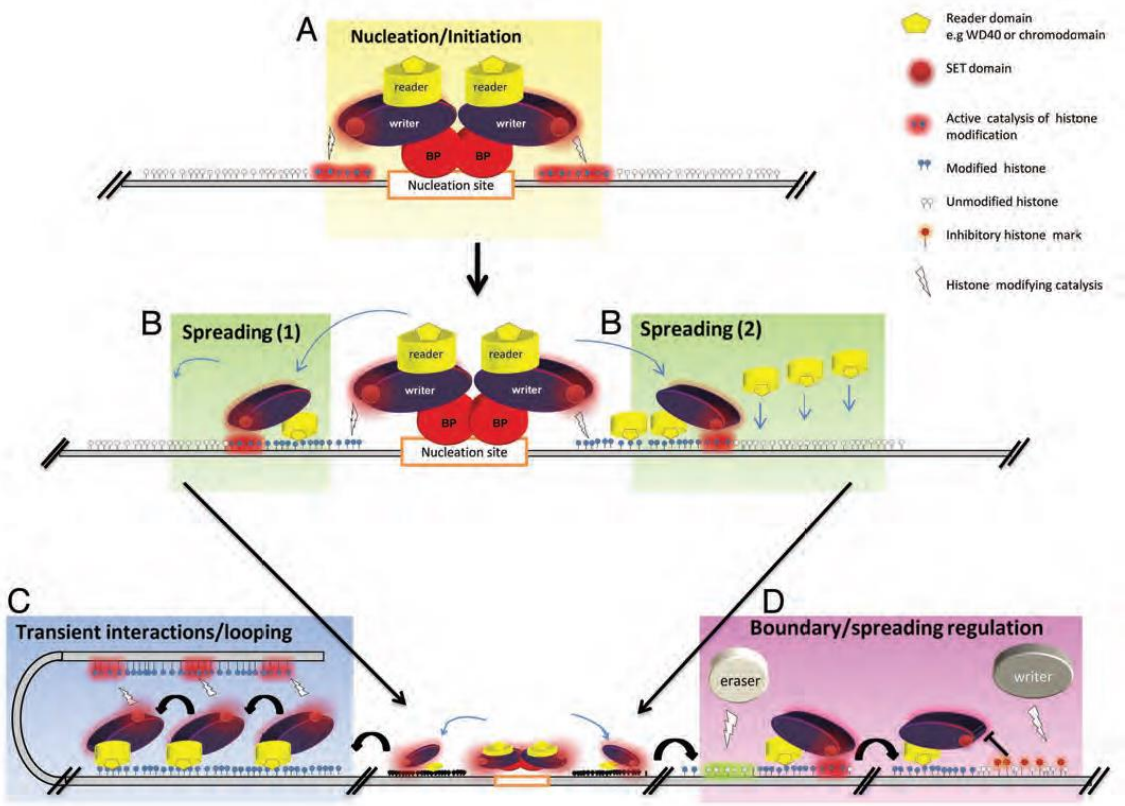


Figure 6. Model for Spreading:

"A. A complex is nucleated at a particular chromatin region by mechanisms such as DNA sequence recognition, non-coding RNA (ncRNA) transcripts or by cooperation of protein complexes bound on a number of DNA elements. This complex or a subsequent complex it attracts will contain a histone modifying complex containing both 'reader' and 'writer' domains. B. The recruited complex modifies nearby nucleosomes. These modified nucleosomes represent binding sites for a 'reader' subset of the complex resulting in either (1) the direct

movement of the 'writers' with the 'readers' to the adjacent nucleosomes or (2) the extraction of the 'writer' portion from that recognition complex to be picked up by nearby chromatin bound 'reader' factors. Either general mechanism results in cycles of binding of the complex, in modification of adjacent nucleosomes and in the progressive spread of the histone modifying complex and the resultant chromatin state. C. In addition to linear spreading of a complex along a chromatin domain, HMT complexes could also transiently associate with and modify other chromatin domains in close physical proximity (looping). D. Every step in the process of spreading may be regulated by other factors encountered by the complex in transit. For example, this could include 'eraser' complexes that remove the histone mark, which is the 'reader' recognition motif, and thereby limit the performance of the spreading complex or act as a 'boundary' factor. Other factors could prevent further spread by inducing an inhibitory chromatin state, such as catalyzing histone modifications that hinder the 'writer' activity. Factors may also locally limit the histone modifying activity while not preventing the spread of the complex (not shown). BP: Binding protein (with specificity for DNA/RNA)" (Hosey, Chaturvedi et al. 2010).

euchromatin regions. Given that the chromodomain of HP1 proteins recognizes H3K9me2 and H3K9me3 marks and G9a self-methylation is also involved in HP1 recruitment (Sampath, Marazzi et al. 2007), the spreading mechanism of G9a/GLP complex was proposed (Bannister, Zegerman et al. 2001). The example of writer-reader spreading mechanism is given from the Hosey, Chaturvedi et al. review (Figure 6).

Importantly, the studies on Ankyrin repeat domain of GLP/G9a (Collins, Northrop et al. 2008) demonstrated its ability to recognize and bind H3K9me2 marks *in vitro* – the product of their SET domain. Three tryptophan's indole groups of Ankyrin domain are forming a pocket that is able to accommodate mono- and dimethyl lysine groups of histone 3. Several mutations substituting tryptophan to alanine destabilize the cavity and prevent interaction of Ankyrin domain with $K9me_{1-2}$ *in vitro* (Figure 7).

Project rationale and hypothesis

Taking together, the propagation of $H3K9me^2$ marks and subsequent heterochromatinization is mediated by the G9a/GLP methyltransferases. However, the mechanism by which G9a/GLP spreads across the chromatin domains *in vivo* remains elusive. Given the ability of the Ankyrin domain to recognize and bind to $H3K9me_{1-2}$ marks, we hypothesize that the interaction

of G9a/GLP Ankyrin domain with H3K9me^2 mark is important for its binding on chromatin *in vivo*.

In order to test this hypothesis we set up the following specific aims: 1) *assess the binding of GLP and G9a genome-wide*; 2) *determine the extent to which Ankyrin domain interaction with H3K9me^2 participates in G9a/GLP binding*.

To approach to the first aim we set up a goal: **design and generate G9a and GLP Ankyrin mutant cell lines.** This goal includes a plan to modify original Murine Erythroleukemia cells such that we could induce the expression of only mutant form of G9a (and separately GLP) which cannot bind to H3K9$me2$ marks.

By testing our hypothesis we expect that introduction of mutation in the Ankyrin domain of GLP and/or G9a will limit or completely prevent its *in vivo* binding on the anticipated spreading regions of G9a/GLP complexes via interaction with H3K9$me1\text{-}2$ marks. By spreading regions we imply the regions of chromatin where G9a/GLP complexes bind without direct recruitment by the DNA-binding activators and remain there performing catalytic activity.

In this project we were using Murine Erythroleukemia cell line.

Murine Erythroleukemia cell line and β-globin locus

Murine Erythroleukemia (MEL) cell line was designed as a model system to study terminal differentiation of erythroid cells. It was established by transformation with Friend virus arresting cells at the stage of erythroblasts (Friend, Scher et al. 1971). Friend spleen focus-forming virus (SFFV) is a replication-deficient retrovirus that can be used to infect susceptible strains of mice and induce erythroleukemia. It carries an *env* gene encoding an envelope glycoprotein gp55 responsible for its high pathogenicity (Yoshimura, D'Andrea et al. 1990). The transformation takes two stages. In the first stage, glycoprotein gp55 interacts with the erythropoietin receptor and activates tyrosine kinase sf-Stk, leading to activation of signal transducing molecules and development of Epo-independent erythroid hyperplasia (Nishigaki, Hanson et al. 2006). This stage is characterized by splenomegaly and polycythemia (Nishigaki, Thompson et al. 2001). On the second stage, the virus integrates into the *Sfpi-1* locus leading to inappropriate expression of its product – PU.1 myeloid transcription factor (Paul, Schuetze et al. 1991). This results in a differentiation block and outgrowth of erythroleukemia cells (MEL).

MEL cells can be induced to undergo differentiation towards mature erythroid cells by treatment with dimethyl sulfoxide (DMSO) (Marks and Rifkind 1978). Differentiation is accompanied by the biological and

morphological changes of the cells. Among these changes is alteration of transcription activity of a variety of genes. That makes this cell line valuable to study the mechanisms of epigenetic regulation and erythroid differentiation in temporal scale of induced differentiation.

After 48 hours of induction with DMSO transcription of the adult β-globin gene (Bmaj) is enhanced significantly (Ganguly and Skoultchi 1985) accompanied with the production of hemoglobin within the cells. This gene is located on β-like globin gene cluster, and it is the dominant β-like globin gene expressed in adult mouse erythroid cells. Transcription of β-like globin genes is regulated by an enhancer region called locus control region (LCR) and located several kilobases upstream of the cluster (Moon and Ley 1990).

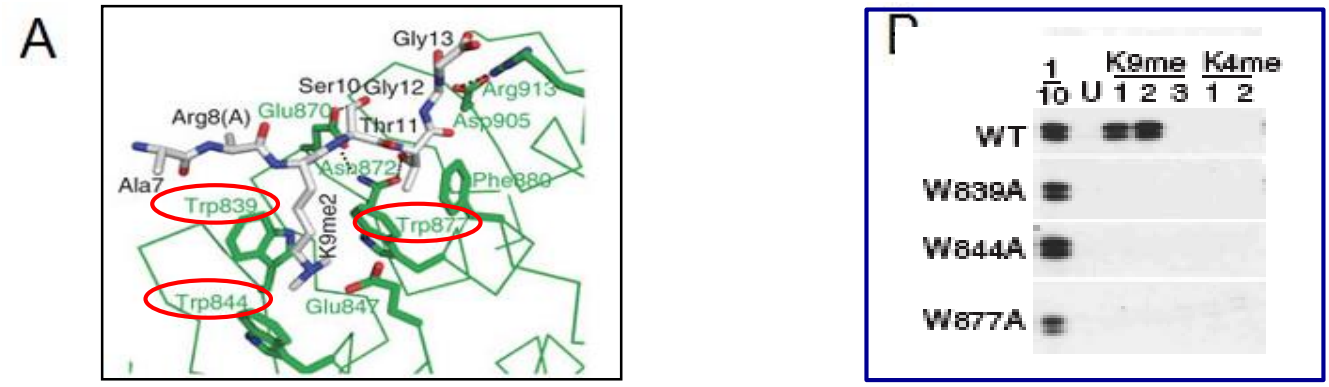


Figure 7. G9a Ankyrin domain interaction with K9me2 peptide.

A. The binding of H3K9me2 in the cavity formed by three tryptophan residues: W839, W844 and W877 (circled in red). B. Substitution of any of the tryptophan residues with alanine results in inability to accommodate mono- and dimethyllysines *in vitro* (Collins, Northrop et al. 2008).

Materials and methods

Cell culture and growth conditions

In this project we are using Murine Erythroleukemia (MEL) cells. These cells are arrested at the level of proerythroblasts and can be induced to differentiate towards the erythroid lineage by incubation with Dimethyl Sulfoxide (DMSO) (Antoniou 1991). Because of the important role of G9a in erythropoiesis (reported previously by our laboratory (Chaturvedi, Hosey et al. 2009)), MEL cells make a good model to study in the context of our hypothesis. Differentiated MEL cells do not become erythrocytes and stay at the level of normoblasts which still retain a cell nucleus. Production of hemoglobin in MEL cells starts at 48 hours of incubation (2% v/v DMSO) and increases to plateau level at 72 hours. For this project MEL cells were differentiated for 72 hours while they are not yet small and remain nucleated, which facilitates chromatin studies.

All parental and clonal MEL cells were growing at 37 °C, 5% CO_2 in RPMI-1640 (HyClone®) medium with addition of 10% tetracycline-free fetal bovine serum (FBS) and 1% penicillin/streptomycin (Gibco®). Induction of differentiation was performed using 2% v/v DMSO. Addition of selection antibiotics to clonal cells is as follows: MEL-TR clone expressing tetracycline repressor – 20 μg/ml of Blasticidine (Wisent Inc.);

GLP and G9a stable knockdown clones – 877 μg/mL G418 Geneticin® (Wisent Inc.); V5-G9a expression clone – 800 μg/mL Hygromycin. To induce expression of shRNA and V5-G9a we added 5 μg/mL Doxycycline to the medium for 72 hours.

Generation of inducible knockdown (GLP)

We used expression of small hairpin RNA (shRNA) as the RNA interference agent. I designed three different shRNA sequences (see Appendix) complementary to one GLP coding sequence (shRNA3) and two GLP 3' untranslated regions (shRNA1 and shRNA2). shRNAs cross-reacted with all GLP isoform transcripts excluding G9a and other proteins. Sequences were ordered from Invitrogen™. For annealing, sense and antisense oligonucleotides were mixed with addition of Medium Restriction Buffer (MRB) and then boiled in a beaker for 10 minutes. I ligated annealed shRNA into pGJ10 plasmid (contains Tetracycline-ON promoter, ampicillin and G418 resistance genes) using 3:1 insert-plasmid ratio, Ligation Buffer and T4 DNA Ligase (Invitrogen™). Ligation products were transfected to TOP10 Z-Competent bacterial cells grown on LB agar plates with addition of 100 μg/mL ampicillin. I used plasmid miniprep followed by restriction digestion to find positive insertion of shRNA and electroporation to transfect newly amplified and purified plasmids into MEL-TR cells. Cells stayed in selection medium (877 μg/mL G418) for 10 days (medium refreshed each 3

days). Induction of shRNA expression was achieved by addition of 5 μg/mL Doxycycline in medium for 72 hours. Next, 20 million cells from each shRNA KD batch cell line were harvested for subsequent nuclear extraction and Western Blot assay. Single-cell colony screening was performed in order to distinguish different clones in batch culture and find the best knockdown condition. For this I made serial dilution of cells from batch culture in 96-well plates. Monitoring each day the cell growth in the wells I was able to locate single-cell colonies and exclusively harvest them for subsequent nuclear extraction and Western Blot assay.

Nuclear Extraction

20 million cells were harvested and sedimented by centrifugation at 1500 RPM for 10 min. Cell pellet was washed with 4 °C 1x PBS twice followed by single wash with hypotonic buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1x proteinase inhibitor cocktail (PIC)). Cells were kept in 200 μL Buffer A for 10 min on ice to let them swell. Cells were homogenized with 1 mL syringe with 27 gauge ½ inch needle. To remove cytoplasmic fraction I used brief spin at 14 000 RPM for 8 sec. Nuclei then were disrupted with addition of Buffer C (20 mM HEPES pH 7.9, 1.5 mM $MgCI_2$, 0.6 mM KCl, 25%(v/v) Glycerol, 0.2 mM EDTA, 0.5 mM DTT, 1x PIC) and incubation on ice for 30 minutes with accidental vortexing. Then nuclear fraction was equilibrated with Buffer D- (20 mM

HEPES pH 8, 5 mM $MgCI_2$, 0.6 mM KCl, 20% (v/v) Glycerol, 0.3 mM DTT, 1x PIC) and sedimented at 11 000 RPM for 15 minutes to remove nuclear debris. To determine protein concentration nuclear extract was subjected to Bradford assay and then stored at -80°C.

Western Blot

For loading I used 20 μg of proteins per each SDS-PAGE gel well. Proteins mixed with the loading dye were denatured at 95°C for 5 minutes before loading on 8% gel. As a protein standard ladder I used Precision-Plus mix (Bio-Rad Laboratories Inc.). Proteins were resolved in the gel at 30 mA constant current and then transferred on nitrocellulose membrane using wet method at 150V for 1 hour 10 minutes. Following transfer membrane blots were blocked in 2% skim milk powder suspended in PBS. Incubation with primary antibodies (1:1000 in 2% milk) was performed overnight at 4°C. Washing with PBS included 5 washes for 5 min. each after overnight incubation. Incubation with secondary anti-rabbit and anti-mouse antibodies (conjugated to horseradish peroxidase) was performed for 1 hour followed by 3 washes for 5 min with PBS. To develop films I incubated blots with peroxide solution for 1 minute.

For this project I used the following antibodies:

GLP (Perseus Proteomics, PP-B0422–00)

G9a (Perseus Proteomics, PP-A8620A-00)

TFIIH p89 (S-19) (Santa Cruz biotechnology, sc-293)

V5-tag (Abcam, ab15828)

Cloning

Coding sequences (CDS) of GLP and G9a were obtained from Dr. Y. Shinkai from the University of Kyoto, Japan. Substitution W844A for G9a and W880A for GLP was performed using site-directed mutagenesis protocol. As an expression vector we used pcDNA5/TO (Invitrogen™) (Figure 13) with Tetracycline-ON promoter system, ampicillin and hygromycin resistance genes and added V5-Flag-Biotin tag at the C-terminus of the insert location. To ligate GLP wild-type and mutant coding sequence into the pcDNA5/TO, first, I designed primers specific (see Appendix) to the C- and N-termanal ends with addition of restriction sequences for BglII and SalI (respectively) and amplified coding sequences using these primers. I ran amplification products on the 1% agarose gel and then cut out the frame with the DNA of interest and purified it with gel extraction kit (Qiagen). Then I performed double digestion of the vector using BamHI and XhoI restriction enzymes (their target sequences are complementary to BglII and SalI after digestion). After PCR purification procedure (Qiagen kit) I used T4 Ligase (Invitrogen™) to ligate

pcDNA5/TO with wild-type and mutant GLP CDS. Following ligation I transformed TOP10 Z-competent bacteria cells by incubation of 100 uL of cells with 5 uL of ligation product for 30 min on ice and accidental tapping to mix the cells. Bacteria cells were then inoculated onto the agar plates (100ug/mL ampicillin) and incubated at 37°C overnight. Separate colonies were then expanded in 2 mL LB medium with addition of 100ug/mL ampicillin and used in miniprep assay to extract amplified plasmid. These plasmids underwent series of digestions to determine the colonies with positive ligation. Selected colonies were expanded further in 150mL flasks and underwent maxiprep assay to extract plasmids for subsequent use in transfection.

Electroporation

To transfect plasmids to the cells we used electroporation – a method designed to increase the electrical conductivity and permeability of the cell membrane by applying an electrical pulse. For each transfection I used 16ug of plasmid and $15*10^6$ cells. Cells were sedimented for 5 min at 1000 rpm. Meanwhile 16ug of plasmid was added to the bottom of 4mm electroporation cuvette (BioRad ™). Cell pellet was dissolved in 400 uL of medium and then transferred to the cuvette. After gentle mixing of the cuvette content it was subjected to electrical pulse with the following conditions: Voltage 260V; Capacitance 950 mF. Then 800 uL more of

medium was added to the cuvette and then everything was transferred to the 25cm^2 flask (Corning®) and filled up to 10 mL.

Chromatin Immunoprecipitation

Chromatin was prepared from MEL cells. 50 million cells were harvested and sedimented in the centrifuge at 1500 RPM for 10 min (RT). The pellet was dissolved in 39 mL of 37°C RPMI-1640 + 10% FBS followed by addition of 1.081 mL of Formaldehyde (1% total) for crosslinking. Solution was mixed and rotated for 25 minutes at RT. To stop the reaction 2mL of Glycine 2.5M (0,125M final) was added. Then cells were washed with cold PBS and underwent N-liq freezing. Stored at -80°C.

Antibody binding to the magnetic beads (dynabeads M280) was performed in parallel with chromatin extraction. 20uL of beads were washed 2x with 1mL of IP100 Buffer. 5ug of antibody was mixed with beads and filled up to 100ul with IP100 buffer. Solution was rotated for 2 hours minimum at RT.

Chromatin extraction was performed on ice. Crosslinked cells were dissolved in 1mL of Swelling Buffer, vortexed and incubated on ice for 20 minutes. Then cells were homogenized in 2mL Dounce homogenizer (Pestle B) by 40 strokes followed by centrifugation at 2000 RPM for 5 min at 4°C. The pellet was dissolved in 300 uL of Sonication Buffer 0.7% SDS in

siliconized tubes (~0.6% SDS final; ~350 volume final). Solution was sonicated with Bioruptor with the following conditions: water 4°C; High Power setting; 35 cycles of 30" ON and 30" OFF corresponding to 17.5 min cumulative sonication time; total time 35 min. To lower the SDS concentration solution was diluted in 1750 uL of Sonication Buffer no-SDS (~0.1% SDS final in 2.1 mL). Solution then was centrifuged at 14 000 RPM, 4°C for 15 minutes to remove cell debris. 50 uL of chromatin was taken for input sample. For blocking chromatin was supplemented with 2 uL of sonicated λ DNA (0.5ug/uL) and 20 uL of Ovalbumin (1ug/uL final). 20 uL of washed beads were used to pre-clear the chromatin for 1 hour at 4°C rotation. For the IP 1 mL of chromatin was mixed with the beads conjugated with antibodies and rotated overnight at 4°C.

Washing of beads included the following steps: 2x with 1 mL Sonication buffer 0.1% SDS, 2x with Wash Buffer A, 2x with Wash buffer B and 2x with TE buffer. For elution beads were dissolved in 200 uL of Elution buffer and heated while mixing at 65°C for 10 minutes. This step was performed twice. The final elution sample (400uL) was supplemented with 16 uL of 5M NaCl (0.2M final) and 4 uL of RNAse A (1ug/uL) (10ug/mL final). Input sample was diluted in 350 uL of Elution buffer and

treated the same way as the elution sample (including the above step). Reverse crosslink was performed in 65°C water bath overnight.

To purify the DNA the samples were supplemented with 4 uL of 0.5M EDTA and 1uL of proteinase K (10ug/uL) and incubated at 42°C for 2 hours. To extract DNA we used phenol/chloroform method. To each sample 400 uL of phenol was added. The mixture was vortexed and centrifuged at 8000 RPM for 10 minutes (RT). The upper layer was transferred to the fresh tube and treated with chloroform: 400 uL of chloroform, vortexes and centrifuged at 8000 for 10 minutes (RT). The upper layer was also transferred to the fresh tube and supplemented with 5 uL of Glycogen (20ug/uL), 40 uL of 3M Sodium Acetate (ph 5.5) and 0.8 mL of 100% Ethanol (-20°C) for DNA precipitation. The mixture was kept at -20°C overnight.

To collect precipitated DNA the mixture was centrifuged at 14 000 RPM for 30 minutes (4°C) and washed with 0.5 mL of 70% Ethanol. The pellet then was dried and dissolved in 100 uL of 10mM TRIS (ph 7.4). DNA was used for subsequent real time qPCR analysis.

See Appendix for buffer references and compositions.

ChIP-Seq – Identification of binding sites by high throughput DNA sequencing

Single-end sequencing was performed at the McGill University and Génome Québec Innovation Centre (Montréal, QC) on an Illumina HiSeq 2000 following the standard procedures. DNA for the sequencing was obtained through the chromatin immunoprecipitation (ChIP) using V5-tag antibody (Abcam®) from Dox-induced undifferentiated and differentiated wild-type MEL cells expressing V5-G9a. Bioinformatics analysis was performed by Alphonse Chu. 50 bp reads were aligned to the mm9 reference genome from UCSC genome browser with BWA version 0.6.2 PMID: 19451168 using default options. Reads mapping to multiple locations on the genome and repeated sequences were discarded for downstream analysis. To prevent high noise level in sequencing coverage within detection of diffuse signals of G9a we used SICER version 1.1 PMID: 19505939 with a window size of 200 bp and gap size of 600 bp. This method was designed as a clustering approach to detect enriched domains of chromatin modifications from ChIP-seq data (Zang, Schones et al. 2009). To remove background peaks we used corresponding mock ChIP (non-induced MEL cells without expression of V5-G9a) and then compared the ChIP-Seq with or without Dox induction to detect differential binding.

Results

The goal of this project is to design and generate G9a and GLP Ankyrin mutant cell lines and test the ability of G9a/GLP complexes to bind spreading regions of chromatin *in vivo*. To achieve this goal we have designed a research plan (Figure 8). We are using Murine Erythroleukemia cells as a model system to elucidate the mechanism of G9a/GLP spreading. These cells are expressing G9a and GLP proteins which have a role in regulating beta-globin locus expression (Chaturvedi, Hosey et al. 2009, Chaturvedi, Somasundaram et al. 2012). The fact that G9a is bound over the whole locus (Chaturvedi, Hosey et al. 2009) without direct recruitment by activators inspired us to investigate this process.

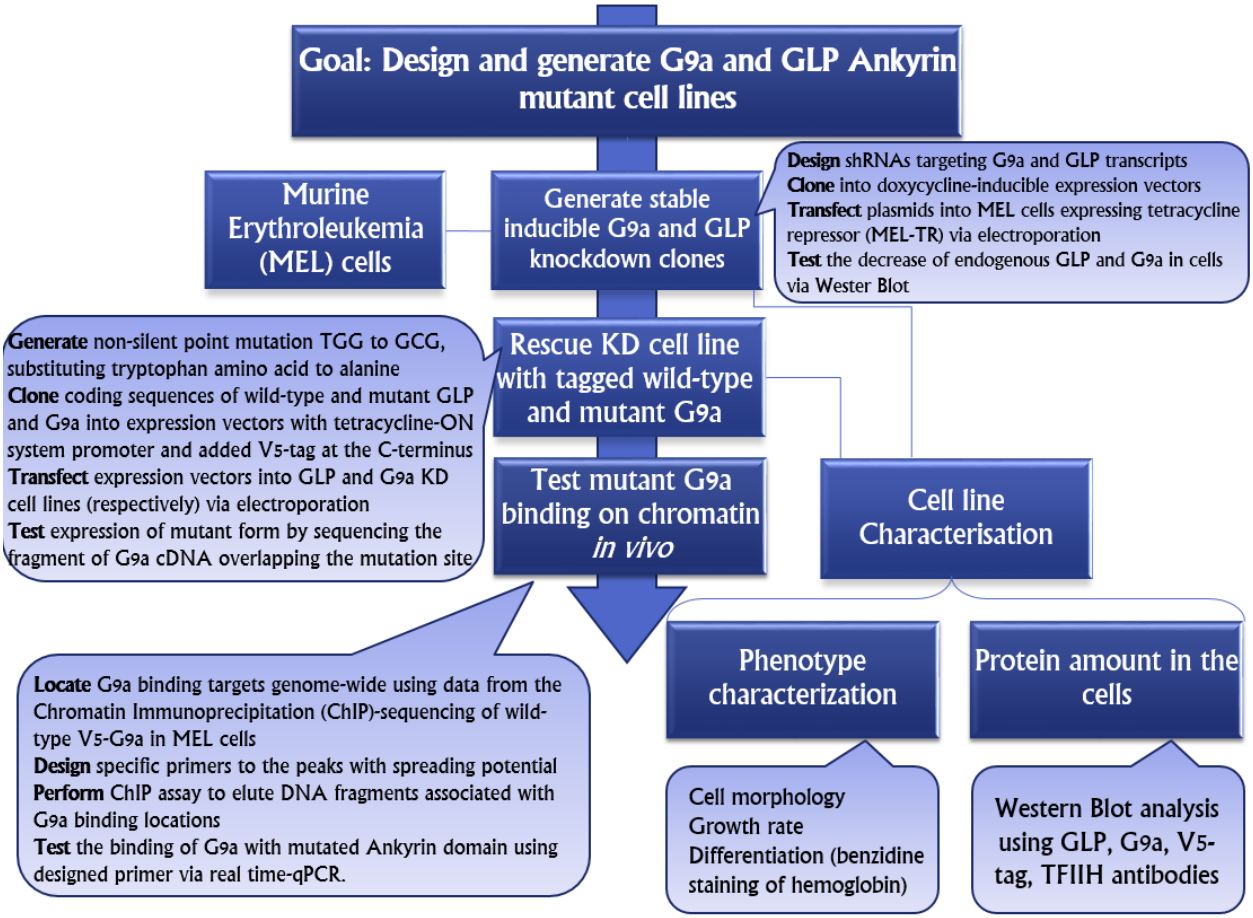


Figure 8. Research plan designed to achieve the goal: design and generate G9a and GLP Ankyrin mutant cell lines.

This schematics demonstrates a step-by-step plan to modify original MEL cells such that we could induce the expression of only mutant form of G9a (and separately GLP) which cannot bind to H3K9me2 marks. Each step is described in terms of methodology used to achieve it.

Designing and characterizing of cell lines

In order to modify MEL cell line such that we can induce the expression of only mutant form of GLP (and separately G9a) which would be unable to bind to its substrate H3K9^{me2} we first designed a method to decrease the levels of endogenous G9a and GLP. For GLP knockdown I designed three different shRNAs targeting all isoforms of GLP transcripts: one for coding sequence (CDS) and two for 3'-UTR regions (untranslated regions) (see Appendix). These sequences were tested for negative cross-reaction with other protein transcripts by *in silico* PCR. I cloned these shRNAs into PGJ10 doxycycline-inducible expression vectors containing the resistance gene to G418 antibiotic. I transfected the GLP shRNA-PGJ10 plasmids into the MEL cells with the expression of tetracycline repressor (MEL-TR) via electroporation. To select the cells with the plasmids, we treated the batch culture with G418 antibiotic (0.733 mg/mL) for 7 days. After the step of selection, I induced the knockdown of GLP by addition of doxycycline (Dox) (5ug/mL) to the medium for 72 hours; the cells were harvested for Western Blot analysis. The batch cultures for all three shRNAs did not have a visible knockdown of GLP (data not shown), probably due to heterogeneity of the transfection. In order to select the clones with the positive knockdown, I performed serial dilution of cells from each batch culture in the 96-well plates and separated single-cell

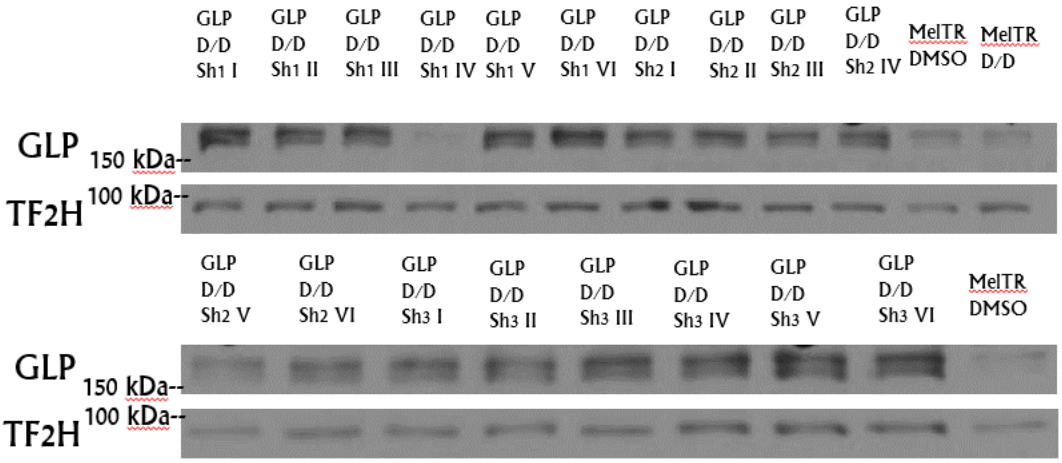


Figure 9. Western Blot for GLP KD single-cell colony screening.

The figure demonstrates the amount of GLP proteins in single-cell colonies (six for each shRNA) from the batch culture after 7 days of selection in G418 antibiotic (0.733 mg/mL) and 72 hours of induction with doxycycline using GLP and TF2H (internal control) antibodies. GLPsh1 IV is showing the best level of GLP knockdown.

Conditions: D/D – DMSO (2% v/v) + Doxycycline (5ug/mL).

colonies. After culturing of separate colonies (six colonies for each shRNA) I performed another Western Blot assay. The results for single-colony screening (Figure 9) demonstrated a significant decrease in signal for GLPsh1 IV clone (4^{th} lane, top panel). This data suggested that the knockdown level for this clone was the best among all tested. For this reason I chose this clone (further GLP KD) for future work. For the knockdown of G9a (G9a KD) I used the clone generated previously by Dr. Chandra-Prakash Chaturvedi using the same method.

In order to check the quality of the knockdown cell lines, I characterized both knockdown clones: counted cells during induction of Dox for 3 days, performed benzidine staining to stain hemoglobin after induction of differentiation and Western Blot to assess protein levels. The growth curve for GLP and G9a KD cell lines (Figure 11) demonstrated the rate of cell growth during incubation with DMSO and DMSO + Dox (D/D) for 72 hours. The data indicated no significant changes in GLP KD cell growth after transfection and GLP knockdown. However, G9a KD cell line did not reach the peak of $1.5*10^6$ cell/mL like MelTR cells before (DMSO) and after (D/D) induction of knockdown. This findings indicated a moderate deregulation of cellular growth due to transfection and G9a knockdown.

To further investigate the differentiation level of these cells I performed benzidine staining (Figure 11). This method is used to assess the overall production of hemoglobin by cells after induction of differentiation towards erythroid lineage. For the staining I used benzidine hydrochloride – compound used to stain hemoglobin within the cells. The results showed that around 90% cells from GLP KD cell line were stained before and after induction of GLP knockdown. The same trend was observed for G9a KD cell line before induction of knockdown (DMSO). However, induction of G9a knockdown (D/D) decreased the overall production of hemoglobin. The data supported the previous findings from the study of G9a KD (Chaturvedi, Hosey et al. 2009) in our lab: the knockdown of G9a decreases the expression of adult beta-globin genes and, thereby, overall hemoglobin production.

Finally, I tested the levels of GLP and G9a proteins in GLP KD cells before and after induction of knockdown. Western Blot (Figure 12A, left panel) indicated a significant decrease in GLP level after induction of shRNA expression (compare third and fourth lane). This result confirmed the data from the single-colony screening for the chosen GLP KD clone. Interestingly, I found that the level of G9a in cells dropped concomitantly with the knockdown of GLP (Figure 12B, left panel). G9a dependence on

GLP presence in the cell was previously reported in literature (Tachibana, Ueda et al. 2005, Shinkai and Tachibana 2011). It was speculated that GLP as a member of GLP/G9a heteromeric complex is essential to stabilize G9a and prevent its degradation. In order to test whether these findings are not the result of cross-reaction of shRNA with both GLP and G9a transcripts I performed real-time qPCR with the primers specific to GLP and G9a mRNA. The data demonstrated that the overall level of G9a transcripts (Figure 12B, right panel) within the cells did not change after induction of GLP knockdown, implying that there was no cross-reaction of shRNA with G9a transcripts. Also the level of GLP transcripts within the cells decreased accordingly to the GLP knockdown (Figure 12A, right panel). Western Blot for G9a KD cell line was performed previously by Dr. Chandra-Prakash Chaturvedi (Figure 12C). It demonstrated a decrease of G9a signal after addition of Doxycycline (Dox), implying an efficient G9a KD in the cells. Also the induction of G9a knockdown did not affect the amount of GLP in the cells. This data suggested that GLP is stable within the cells in the absence of G9a. Together, these results confirmed efficient knockdowns of GLP and G9a in generated cell lines. This allowed us to move on with our goal – rescue KD cell lines with Ankyrin mutants – GLP and G9a proteins that are unable to bind to their substrate.

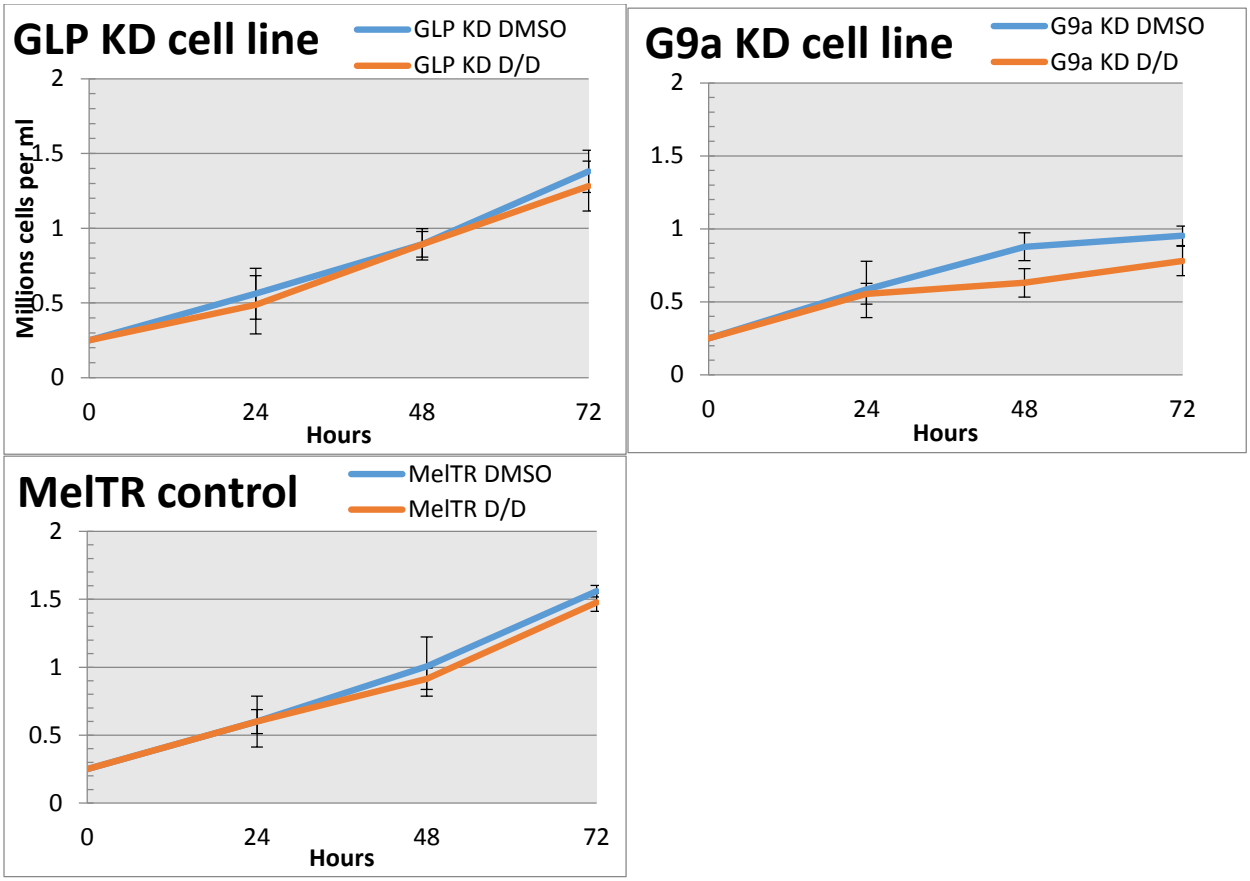


Figure 10. The growth curves for differentiated (DMSO) and with addition of knockdown inducer Doxycycline (D/D) GLP KD cell line, G9a KD cell line and Mel-TR as a control parental cells.

Starting at $0.25*10^6$ cells/mL GLP KD cell line reached a peak of around $1.3*10^6$ cell/mL in 72 hours. Similar result to the control Mel TR cells – $1.5*10^6$ cells/mL. Induction of knockdown didn't affect cell growth for GLP KD cell line. G9a KD cells reached the peak of 1 million cell/mL in differentiated condition and $0.8*10^6$ cell/mL with addition of Doxycycline.

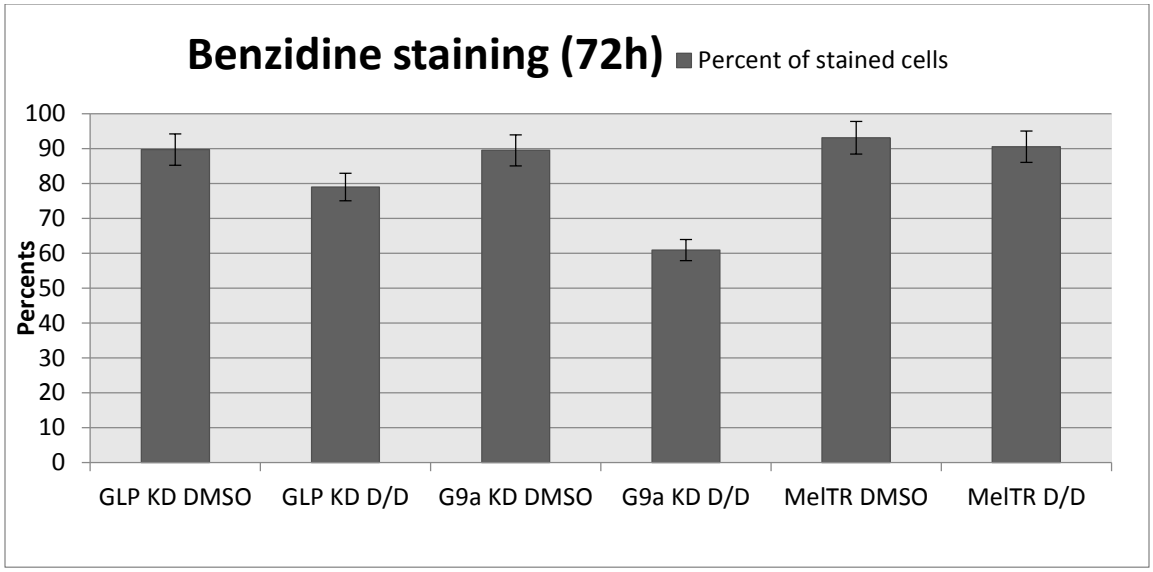


Figure 11. Benzidine staining for differentiated GLP and G9a KD cell lines and Mel TR as a control.

The graph represents an amount of cells stained by benzidine hydrochloride. GLP KD cell line had around 90% differentiation level (most of the cell were stained) similar to parental Mel-TR cells. G9a KD cell line had lesser stained cells in case of induction of G9a knockdown with Doxycycline (D/D). I reproduced the result previously obtained in our lab with G9a KD (Chaturvedi, Somasundaram et al. 2012).

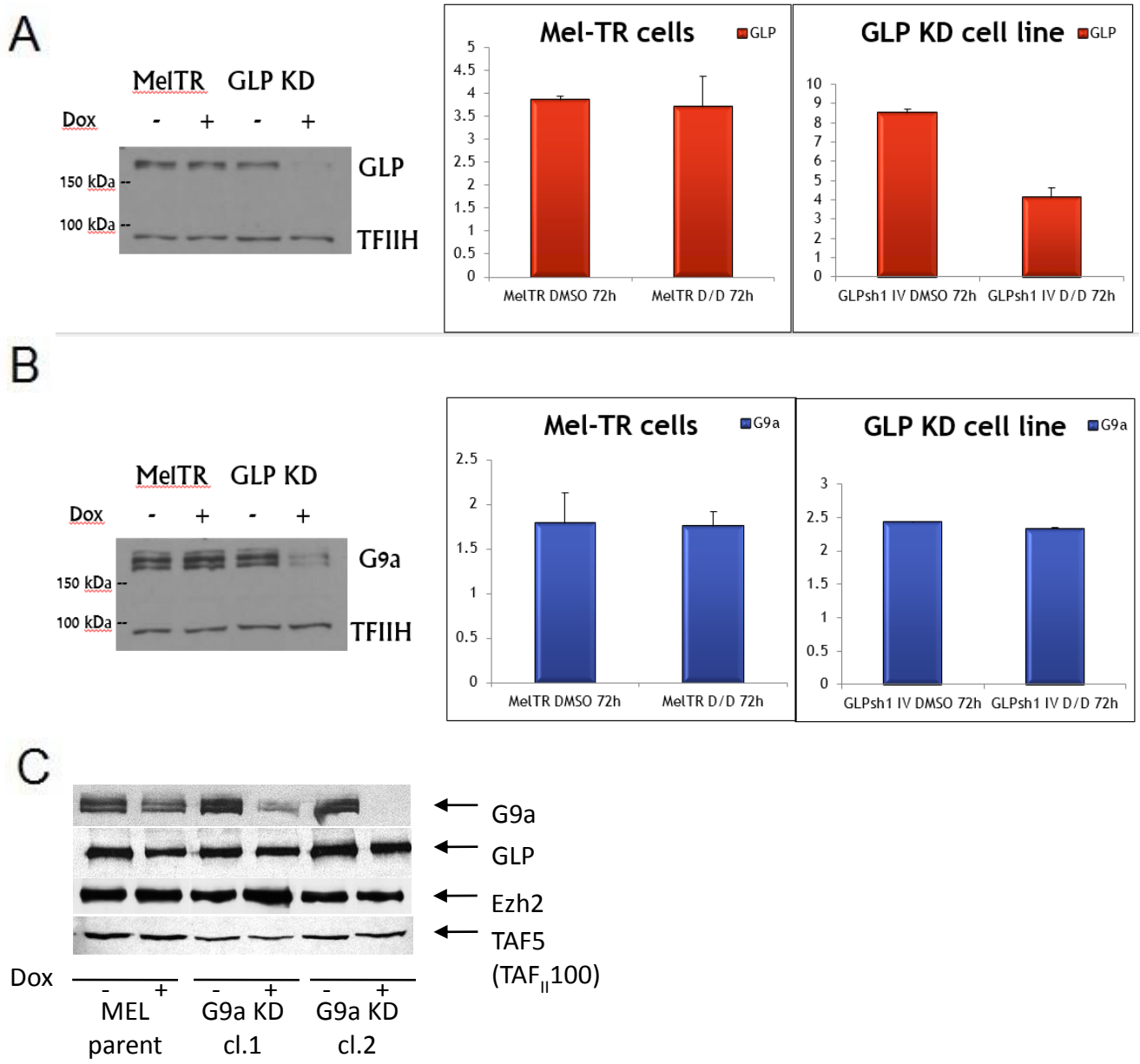


Figure 12. Western Blot results for GLP and G9a KD cell lines.

<u>A. Left panel.</u> Western Blot with MEL-TR (control) and GLP KD nuclear extracts using GLP and TF2H (internal control) antibodies. GLP signal is significantly reduced after treatment with Doxycycline (Dox+) <u>Right panel.</u> qPCR confirmation of GLP knockdown using RNA from GLP KD cells and primers specific to GLP mRNA. <u>B. Left panel</u>.

Western blot demonstrates concomitant decrease of G9a signal in GLP KD cells after treatment with Doxycycline (Dox). Right panel. qPCR indicates no changes in the amount of G9a transcripts in GLP KD cells after induction of GLP knockdown (D/D). C. Western Blot for two G9a KD cell lines demonstrates decrease of G9a signal (but not GLP) following induction of G9a knockdown (Dox+ lane 4 and 6).

Rescue GLP and G9a KD cell lines with exogenous wild-type and mutant GLP and G9a

The next step was the generation of wild-type and mutant G9a and GLP expressing cell lines. As a control cell line we used knockdown cell line with the expression of exogenous wild-type protein with a V5-Flag-Biotin tag at C-terminus. The idea to use the tagged form of GLP and G9a came up due to low amounts of G9a- and GLP-associated DNA after chromatin pull-down using antibodies against endogenous G9a and GLP. The tagged form of proteins would facilitate the pull-down of chromatin using antibodies against V5-tag for subsequent experiments.

To perform rescue assay we cloned two templates (wild-type and mutant) of GLP and G9a coding sequences (CDS) into expression vectors (pcDNA5/TO) containing V5-Flag-Biotin tag and hygromycin resistance gene. The introduction of point mutations (TGG to GCG) was performed by mutagenesis. We substituted tryptophan (TGG) which is forming a cavity in the Ankyrin domain for the H3K9^{me2} mark to alanine (GCG) at the W844 in G9a sequence and W880 in GLP sequence (Collins, Northrop et al. 2008). In order to clone GLP (wild-type and W880A mutant) I designed custom primers to the 5’ and 3’ ends of CDS with added restriction sites (Bgl II from 5’ and Sal I from 3’) (See appendix). To generate the insert, I amplified

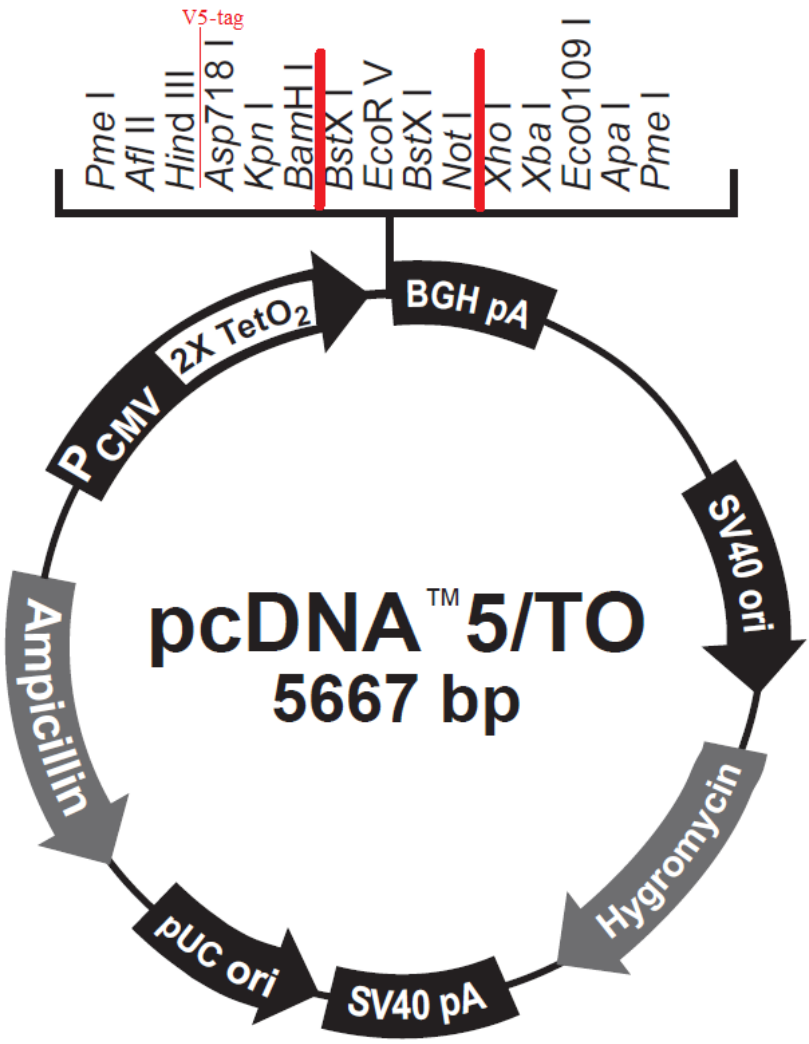


Figure 13. pcDNA 5/TO plasmid map.

Contains CMV promoter with the tetracycline operator site, multiple cloning sites, and resistance genes for Hygromycin and Ampicillin. The tag is inserted downstream of Hind III restriction site; GLP coding sequence is inserted between BamH I and Xho I restriction sites.

(http://tools.lifetechnologies.com/content/sfs/vectors/pcdna5to_map.pdf)

GLP CDS (3747bp) using these primers and purified the target fragment from the agarose gel. Then I digested pcDNA5/TO in two sites: BamHI and XhoI. The "sticky ends" generated by digestion with BamHI and XhoI restriction enzymes are complementary to those which are digested with BglII and SalI (respectively). This allowed me to ligate pcDNA with my GLP insert. The ligation product was amplified in TOP10 Z-competent bacteria cells. Separate colonies were taken to screen the ligation efficiency. To do that, I used the miniprep assay to extract plasmids from bacteria cells and digested them with XbaI restriction enzyme. XbaI target sequence is located approximately at 1800 bp upstream of 3' end of GLP coding sequence (Figure 14) and XbaI site on pcDNA (Figure 13). Thereby by digestion with XbaI restriction enzyme I expected to see two bands on the resolved agarose gel. Results of the digestion with XbaI for plasmids with wild-type GLP inserts (Figure 15, left panel) and W880A mutant GLP inserts (Figure 15, bottom panel) provided me with several positive ligation products: lanes 3, 9 and 10 for wild-type and lanes 7 and 10 of W880A mutant. As expected the size of second fragment is approximately 1800 bp. In order to confirm that the whole insert was ligated into the plasmid I used another restriction enzyme HindIII with the positive plasmids. Restriction

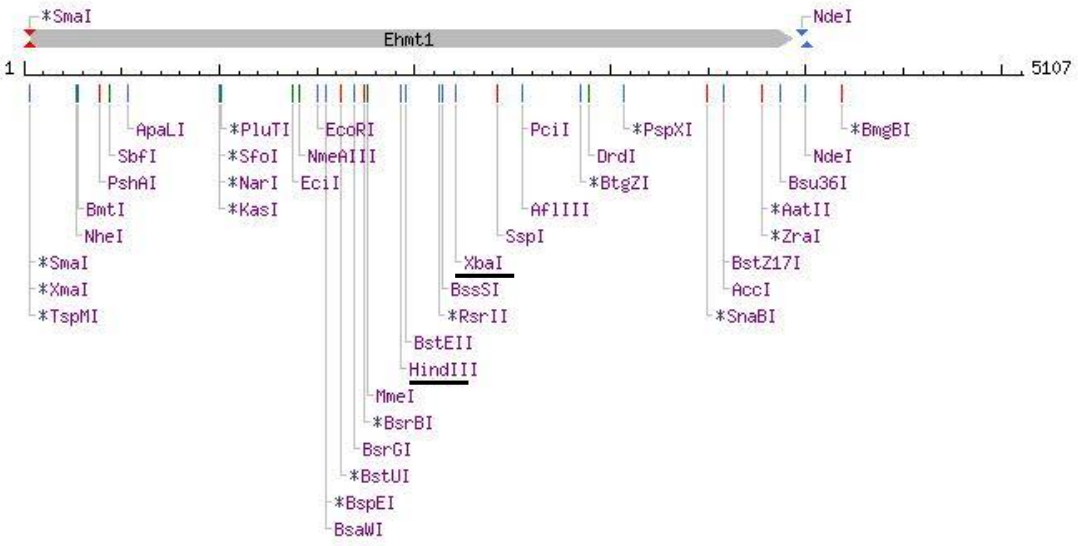


Figure 14. Restriction map for the coding sequence of GLP (Ehmt1).

The figure represents the positions of sole restriction sites on the GLP coding sequence. This map was used to pick the appropriate restriction enzymes to cut the plasmid with the GLP insert. XbaI and HindIII restriction sites which I used for digestion are underlined.

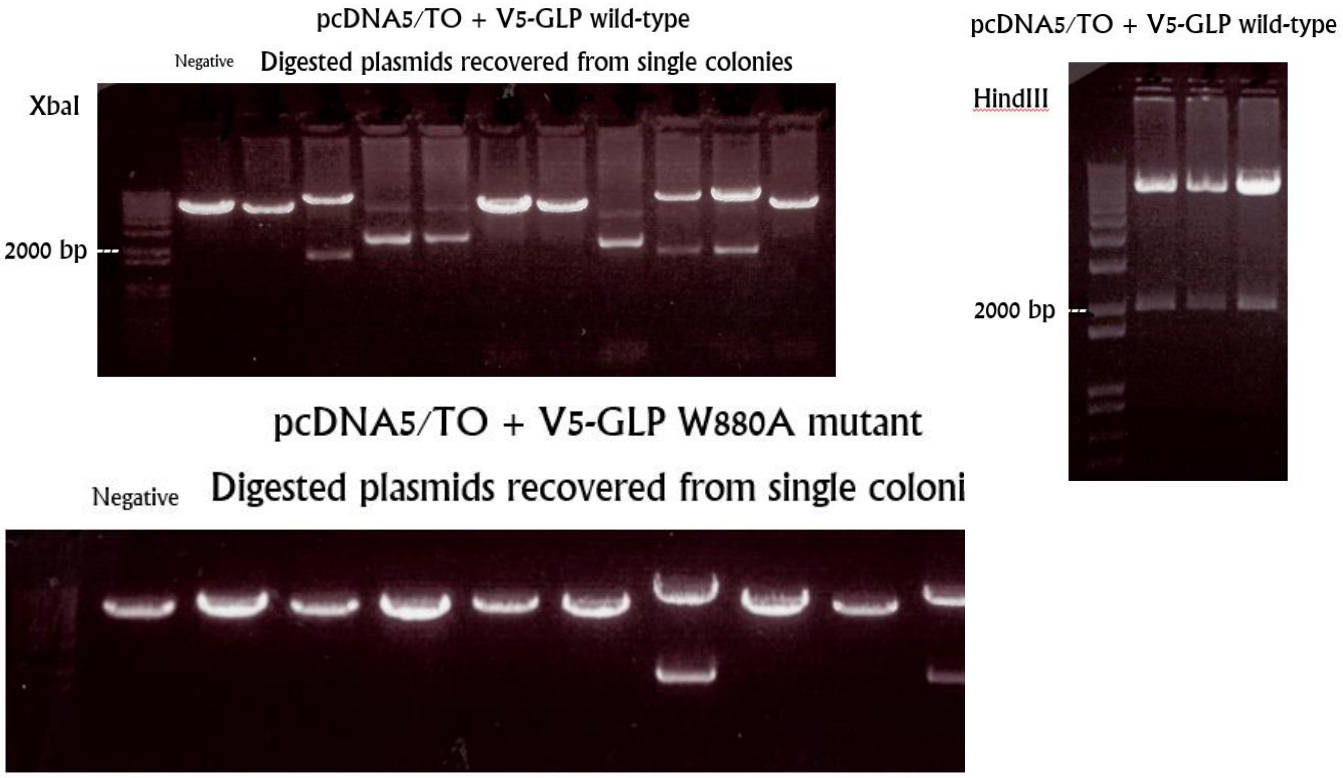


Figure 15. Screening single bacteria colonies for positive V5-GLP ligation products.

Figure represents the results from running digested plasmids on 1% agarose gel. Left and bottom panels demonstrate digestion with XbaI of plasmids ligated with the V5-GLP wild-type and W880A mutant inserts (respectively). Lanes 3, 9 and 10 for wild-type and lanes 7 and 10 for mutant indicate expected two fragments of the plasmid; the second fragment is ~1800 bp. Right panel demonstrates digestion with HindIII of plasmids with wild-type V5-GLP (identified in the previous XbaI digestion); the second fragment is ~2200 bp.

site of HindIII is located at 2000 bp downstream of 5' end of GLP coding sequence and 2200 bp away from the HindIII site on pcDNA. Digestion with HindIII restriction enzyme (Figure 15, right panel) identified the second fragments at the size of ~2200 bp. This result reconfirmed the positive ligation of the insert into the plasmid.

To check the integrity of my insert in the plasmid, I designed 8 custom primers (see Appendix) to cover the template sequence and connection parts with pcDNA and sent it for sequencing. The results from the sequencing confirmed the correct ligation of the insert. Dr. Chandra-Prakash Chaturvedi cloned G9a templates (wild-type and W844A mutant) into pcDNA5/TO. Both constructs contained V5-Flag-Biotin tag to facilitate subsequent G9a pull down in ChIP assays. Dr. Chandra-Prakash Chaturvedi transfected the plasmids into G9a KD cell line via electroporation. The batch cultures were selected in the medium with addition of 0.8 mg/mL Hygromycin for 10 days. Separate clones from the single-cell colonies were obtained through the serial dilution in 96-well plates. Then these clones were screened for the level of V5-G9a expression in the cells via Western Blot. Clones with the best level of V5-G9a expression were chosen to be used in characterization and subsequent experiments. Since Dr. CP Chaturvedi left

the lab by this time, I focused on determination of G9a binding on chromatin *in vivo*. The GLP project was suspended.

I characterized V5-G9a wild-type (V5-G9a WT) and V5-G9a mutant (V5-G9a W844A) cell lines by their growth rate, differentiation level and efficiency of simultaneous knockdown of endogenous G9a and expression of V5-G9a counterpart. The growth curve (Figure 16) indicated no problem in cell growth rate in undifferentiated condition. The cells even reached higher concentration of $2.5*10^6$ cell/mL than parental MelTR cells (see Figure 10 for MelTR growth curve). However, all batch cultures and clones reached the plateau within 24 hours after induction of differentiation by 2% v/v DMSO and stopped further proliferation (Figure 16, right graph). A similar trend in the growth rate was observed with the addition of DMSO + Dox (Figure 16, bottom graph). This might resulted from the abundance of different antibiotics and compounds in the medium to keep cells selected and induced (e.g. selection medium for V5-G9a cell line besides regular 1% Pen/Strep mix contained 20ug/mL Blasticidine, 877ug/mL G418, 800ug/mL Hygromycin and additionally 2% v/v DMSO).

This result might also suggest that these cell are defective in their capability to differentiate. To test this possibility I performed benzidine staining of hemoglobin within the cells. The number of cells producing

hemoglobin by 72 hours of differentiation also decreased two-fold in all the batch cultures and clones (Figure 17). The results indicated an abnormal response to DMSO and defect in cellular processes of proliferation and hemoglobin production. To exclude the possibility of poor quality of DMSO solution I performed the staining for parental Mel-TR cells (Figure 17, compare the last section with other clones). These cells responded well to the compound and had over 80% differentiation level. Taking together, we concluded that the most probable reason was cellular stress after addition of DMSO.

Western Blot assay with G9a (Figure 18, left panel) and V5-tag (Figure 18, right panel) antibodies demonstrated the signal in V5-G9a clones and batch cultures, implying that these cells were producing exogenous V5-G9a with practically complete knockdown of endogenous G9a upon induction with Doxycycline (D/D). However, overall level of G9a and V5-G9a in modified cell lines was lower than that in parental MEL-TR cells.

In order to confirm that the mutated G9a is expressed in newly generated cell line I extracted RNA from the V5-G9a W844A cells and performed first strand cDNA synthesis using M-MLV Reverse Transcriptase and oligoDT_{12-18} primers. To sequence mutation site I amplified a small

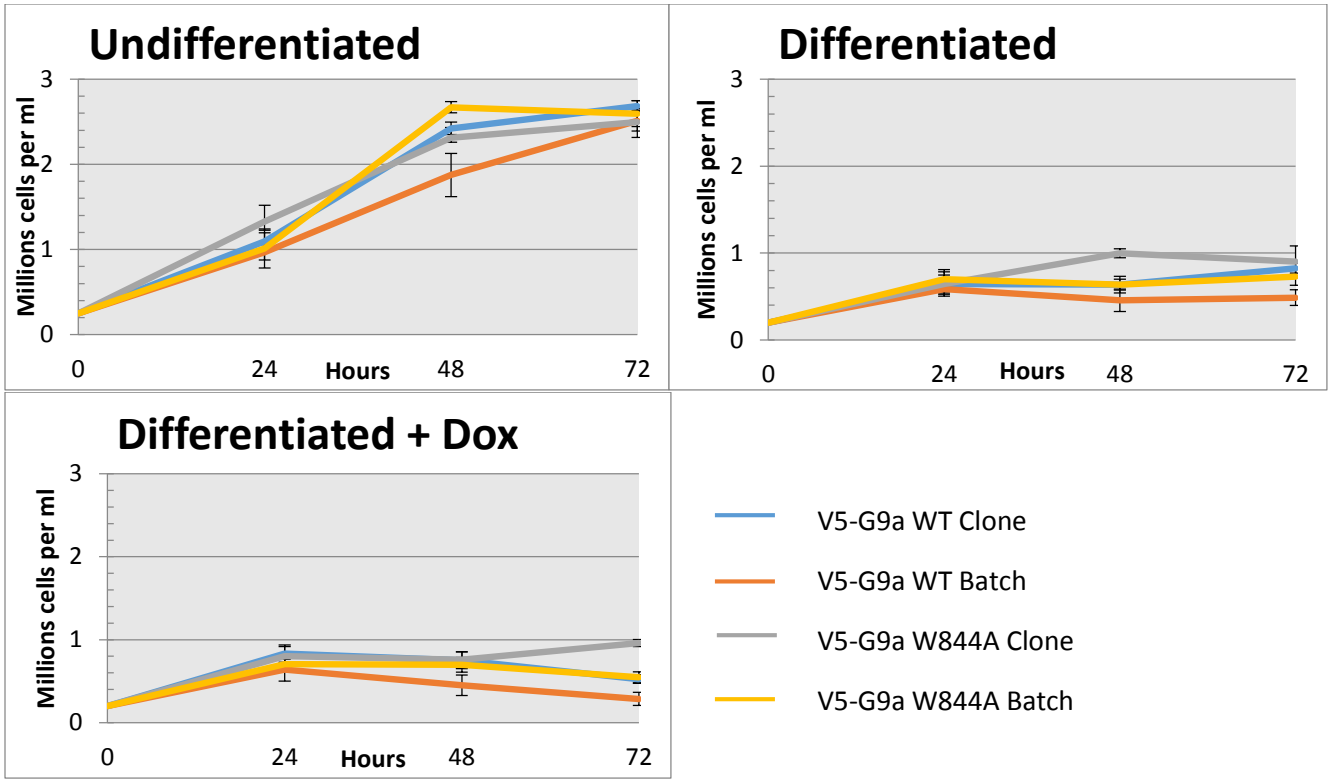


Figure 16. Growth curves for V5-G9a batch cultures and clones.

All cell lines under undifferentiated conditions exhibit high rate of growth, reaching plateau level in around 48 hours; upon induction of differentiation with DMSO (2% v/v) cells completely stop to proliferate. Similar trend is observed with addition of DMSO + 5ug/mL Doxycycline.

Abbreviations:

- **V5-G9a WT Clone – V5-G9a wild-type clone;**
- **V5-G9a WT Batch – V5-G9a wild-type batch culture;**
- **V5-G9a W844A Clone – V5-G9a Ankyrin mutant clone;**
- **V5-G9a W844A Batch – V5-G9a Ankyrin mutant batch culture.**

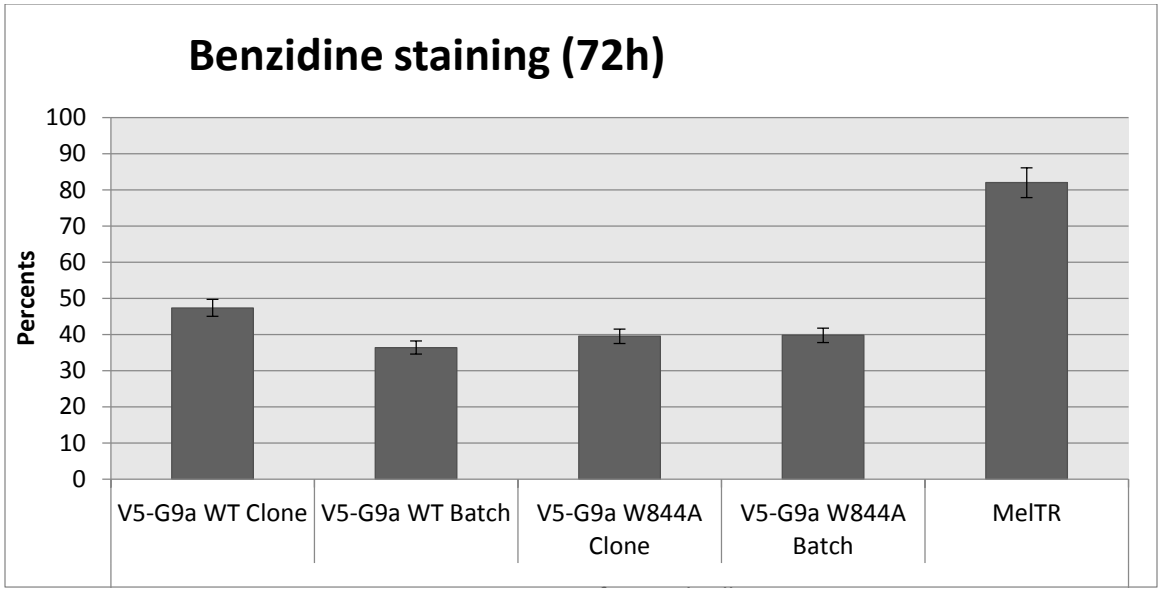


Figure 17. Benzidine staining of hemoglobin for V5-G9a cell lines.

The graph represents the percent of cells producing hemoglobin (stained cells). In comparison to parental Mel-TR cells, all clones and batch cultures have less than half of the cells producing hemoglobin by 72 hours after induction of differentiation.

Abbreviations:

- **MelTR – parental MEL cells expressing tetracycline repressor;**
- **V5-G9a WT Clone – V5-G9a wild-type clone;**
- **V5-G9a WT Batch – V5-G9a wild-type batch culture;**
- **V5-G9a W844A Clone – V5-G9a Ankyrin mutant clone;**
- **V5-G9a W844A Batch – V5-G9a Ankyrin mutant batch culture.**

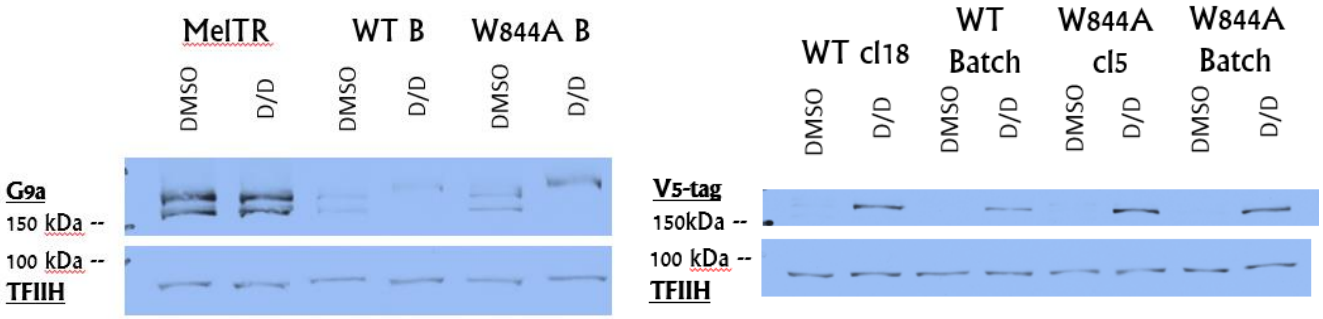


Figure 18. Western Blot assay for V5-G9a wild-type and mutant cell lines.

All G9a-V5 clones and batch cultures (WT and W844A) after doxycycline induction at 72 hours are producing V5-tagged G9a (D/D condition). In comparison to parental Mel-TR (unmodified, as a control) the cells present considerable decrease in overall G9a protein level. Mutated isoform has slightly higher expression than wild type (compare W844A B, D/D with WT B, D/D).

Abbreviations:

- **MelTR – parental MEL cells expressing tetracycline repressor;**
- **WT B – V5-G9a wild-type batch culture;**
- **W844A B – V5-G9a Ankyrin mutant batch culture;**
- **DMSO 2% v/v, D/D – DMSO + 5ug/mL Doxycycline.**

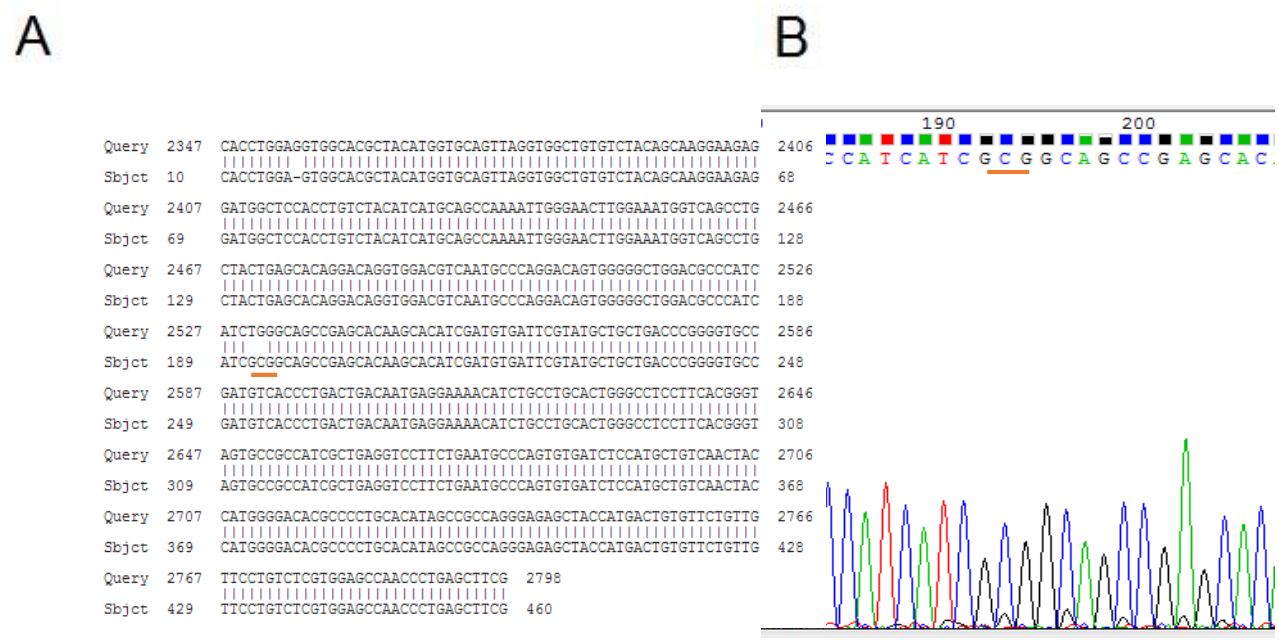


Figure 19. Confirmation of mutation presence in V5-G9a W844A cell line.

<u>A.</u> BLAST results for original G9a sequence (Query) and sequenced fragment overlapping Ankyrin mutation W844A (Sbjct). Sequencing confirmed the substitution of tryptophan (TGG) on Alanine (GCG). <u>B.</u> The graph representing the probability of each nucleotide in the sequence. GCG triplet is prevalent in V5-G9a W844A cells.

fragment of 488bp overlapping the mutation site. The results from sequencing confirmed the presence of point mutation TGG (W) to GCG (A) (Figure 19A) and its dominant expression in the cells (Figure 19B). Thereby cells completely lost the wild-type form of G9a through the knockdown and effectively expressed mutated one.

At this step we had achieved the goal to generate Ankyrin mutant cell line in which we could induce the replacement of endogenous G9a with its mutant counterpart. Doxycycline is an inducer of simultaneous expression of shRNA and V5-G9a.

The schematics on the Figure 20 demonstrates the path of modifications we implemented to MEL cells to produce wild-type and mutant V5-G9a cell lines.

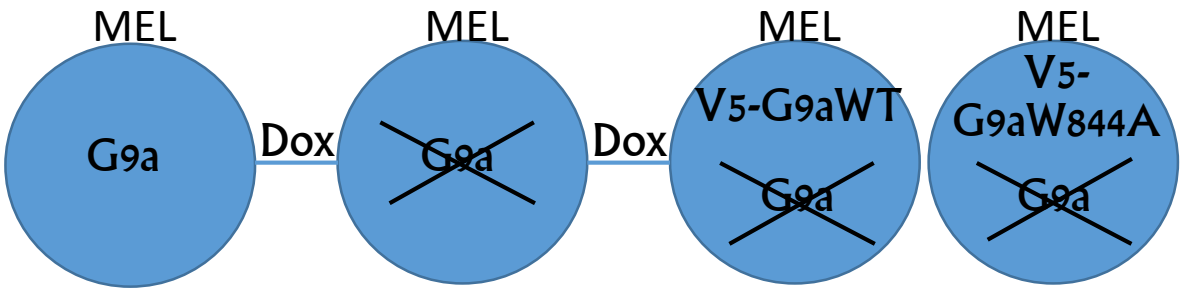


Figure 20. Schematics for MEL cell line modification.

Original MEL cells were transfected with the vector containing shRNA which targets G9a transcripts within the cells. On the next stage, expression vectors with the wild-type and mutant coding sequences of V5-G9a were transfected to G9a knockdown cell line. In order to induce expression of shRNA and V5-G9a we used Doxycycline (Dox). After induction these cells start to produce only V5-G9a isoform with wild-type coding sequence (V5-G9a WT) or coding sequence with the point mutation in the Ankyrin domain W844A, which is preventing the binding with H3K9me2 mark.

ChIP-sequencing and bioinformatic analysis

Then we were ready to approach to our first aim: assess the binding of G9a genome-wide. To assess the binding of proteins to the specific genomic regions we used Chromatin Immunoprecipitation (ChIP) assay which was designed to pull-down target protein crosslinked with its complexes and DNA from the sheared chromatin. Then for qPCR analysis we used purified DNA from the solution which is supposed to be associated with the protein of interest *in vivo*. To locate G9a targets genome-wide Dr. Chandra-Prakash Chaturvedi performed ChIP assay using V5-tag antibody with subsequent sequencing of purified DNA. For this experiment we used undifferentiated and differentiated V5-G9a (wild-type) cells induced by doxycycline. The raw ChIP-seq data was analyzed by Alphonse Chu using SICER tool – a clustering approach for identification of enriched domains, such as spreading G9a. The analysis of ChIP-seq data revealed 23,003 G9a enriched regions before differentiation and 9,825 after differentiation ($p<0.01$) (Figure 21). This analysis revealed a high number of peaks on the genome due to the multiple localization on the anticipated spreading regions of G9a. Only third part of G9a targets (6587) remained after induction of differentiation, while new arose and other disappeared. This result suggested an active role of G9a in regulation of MEL cells differentiation. To identify the biological processes associated with the G9a target genes I used Gene Ontology tool.

Interestingly, among different biological processes before differentiation G9a targets were also associated with erythrocyte homeostasis and differentiation, implying a role of G9a in these processes.

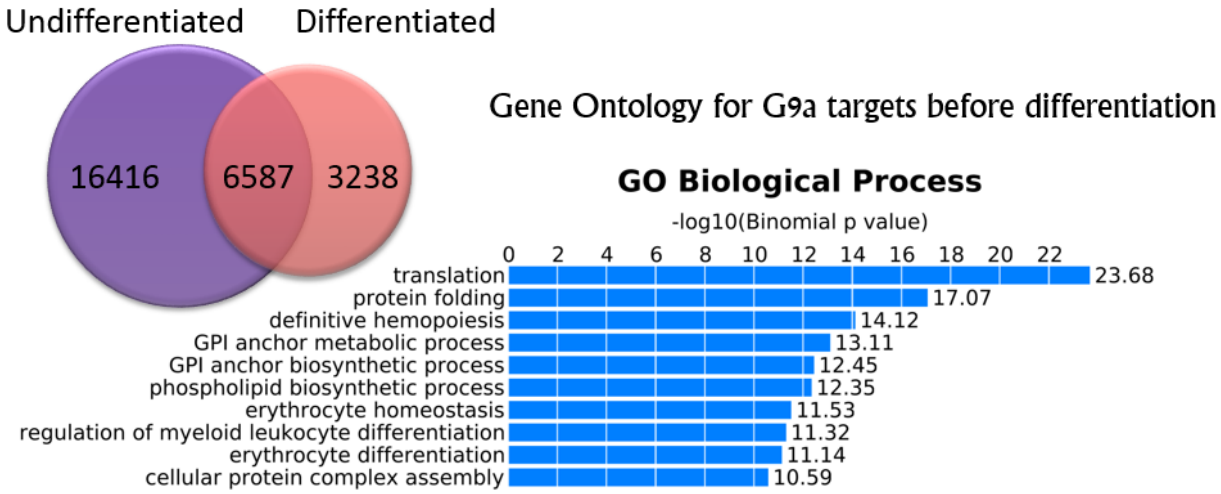


Figure 21. ChIP-seq raw data analysis.

SICER tool for clustering was used to identify 23,003 G9a enriched regions before differentiation and 9,825 after differentiation ($p<0.01$). The biological processes implicated with G9a targets were revealed by Gene Ontology tool.

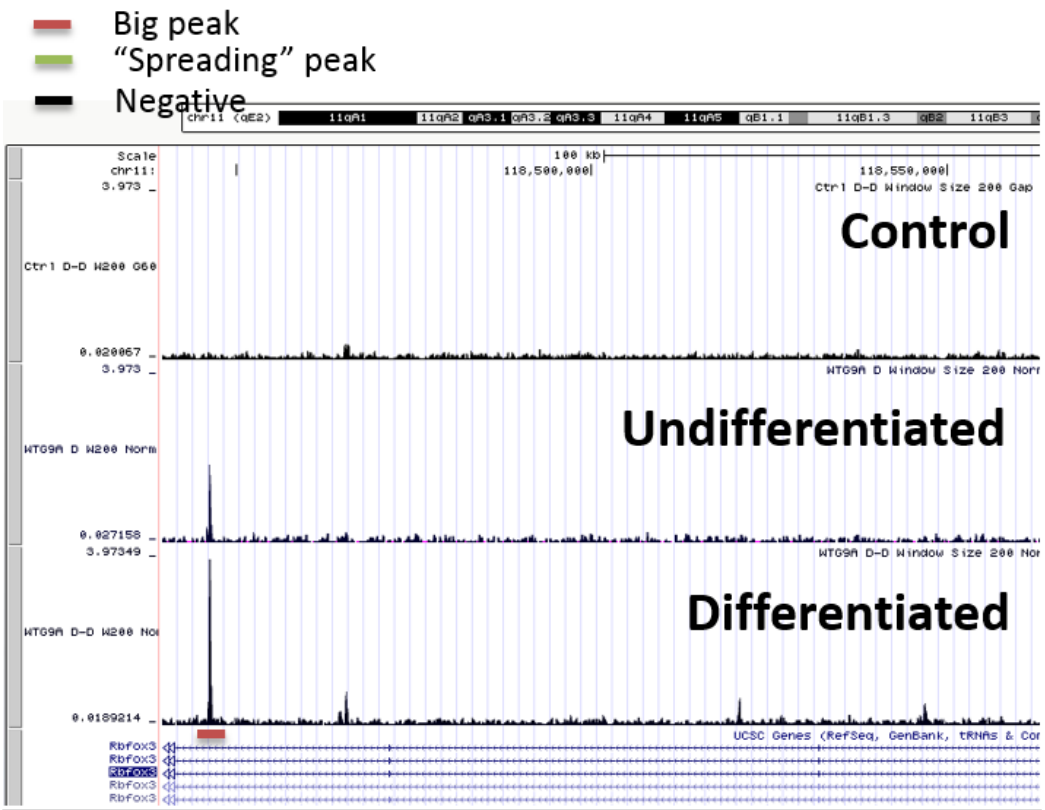
ChIP-seq validation and mutant binding

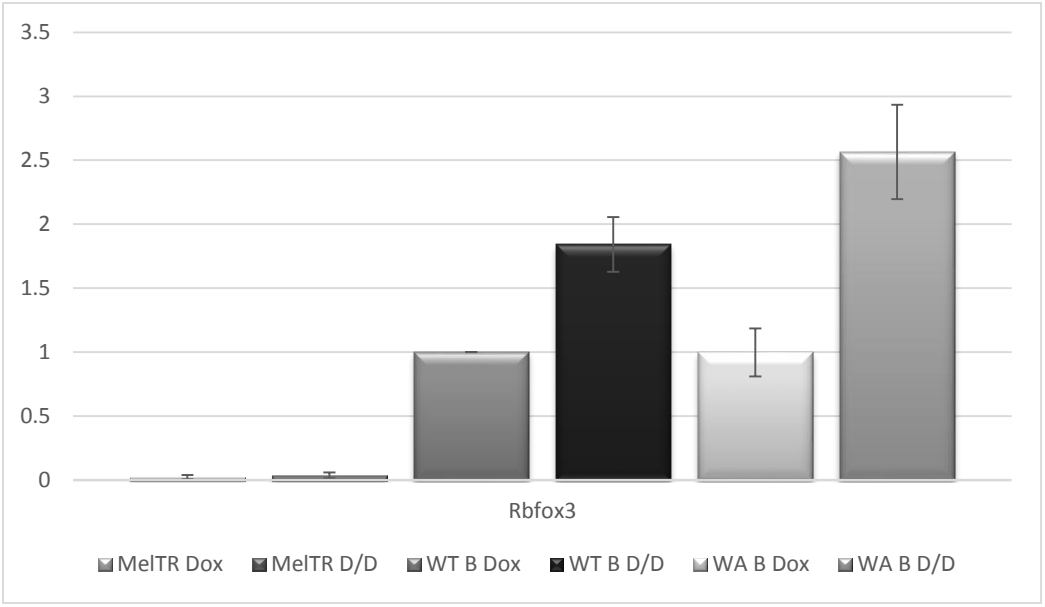
The next step was to identify genomic regions with anticipated G9a spreading potential. The mono and bi-directional spreading of G9a from the recruitment site was the criterion used to assess the validity of the target site. Usually this spreading is identified on the UCSC browser as the presence of a big peak followed by smaller ones on one or both sides of the peak. I used chosen targets to validate the ChIP-seq data and also to test the binding of G9a mutant in spreading regions. I designed custom primers for the peaks of interest (see Appendix) to use them in qPCR analysis following the ChIP assay. To test all the regions I performed series of ChIP experiments and analyzed the results. Totally, 13 different targets were tested in wild type and 6 in mutant cells (Figure 22).

The analysis is presented in two panels: top panel is visualization of ChIP-seq data on the UCSC browser and bottom panel is the qPCR amplification of DNA pulled down together with V5-G9a from the chromatin. The graph represents a relative enrichment of V5-G9a on the target genomic region. Analysis confirmed the results of G9a binding for all tested regions from the ChIP-seq data (Figure 22 A-E), introducing a reliable tool to visualize G9a targets in MEL cells. However, I found that V5-G9a with Ankyrin mutation was still enriched at all tested regions (Figure 22A, 22B, 22F, WA B samples), including beta-globin genes which are predicted

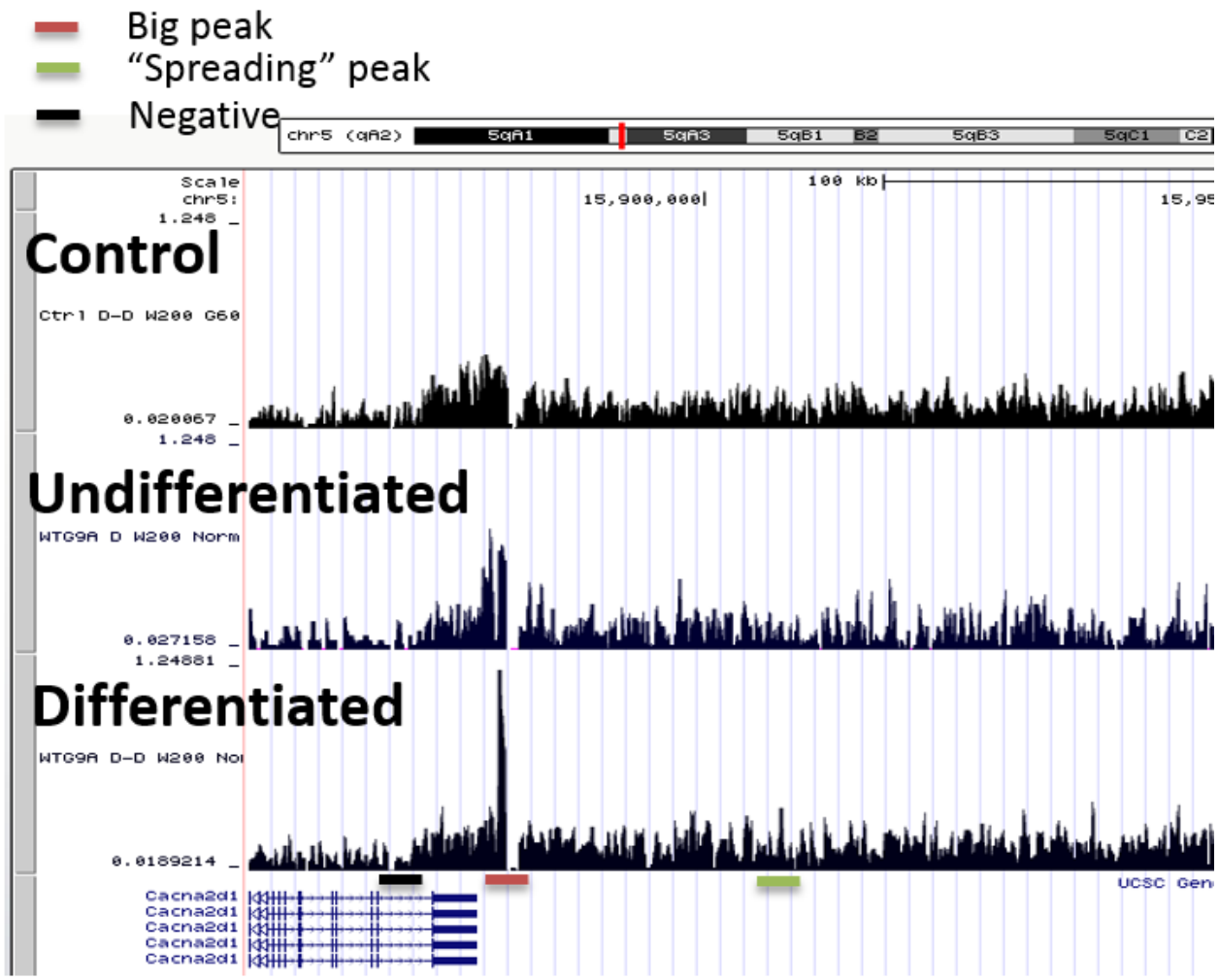
spreading regions of G9a. These results suggested that Ankyrin domain is not the only factor contributing to the binding and stabilization of G9a on chromatin *in vivo*. This function may be complemented by other G9a interacting agents and microenvironment factors.

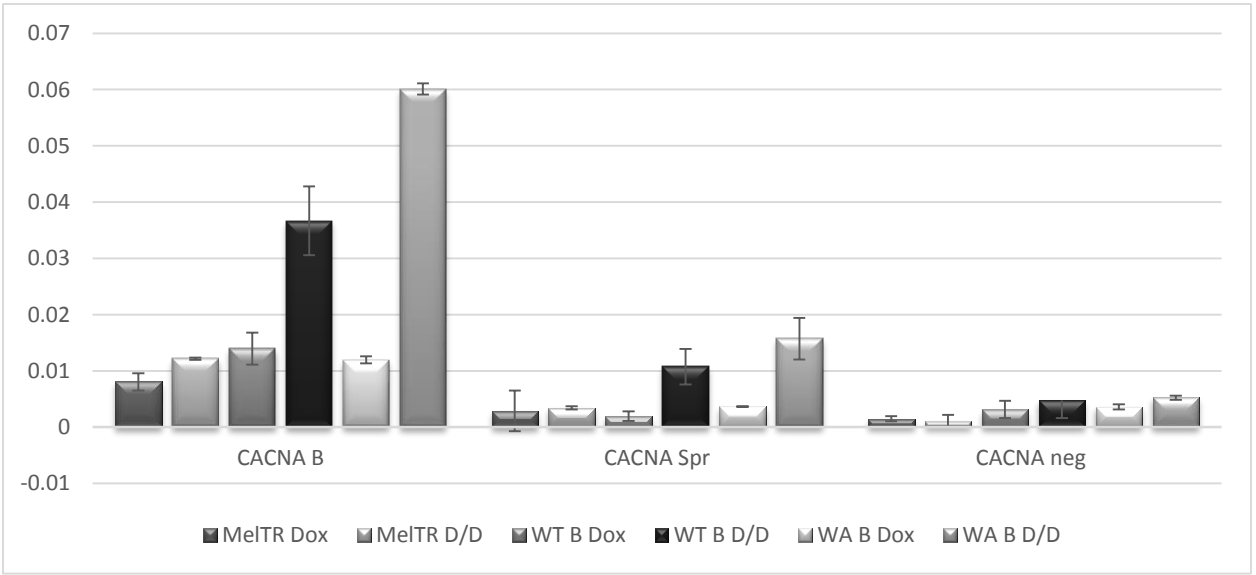
A


Big peak
"Spreading" peak
Negative
Control
Undifferentiated
Differentiated

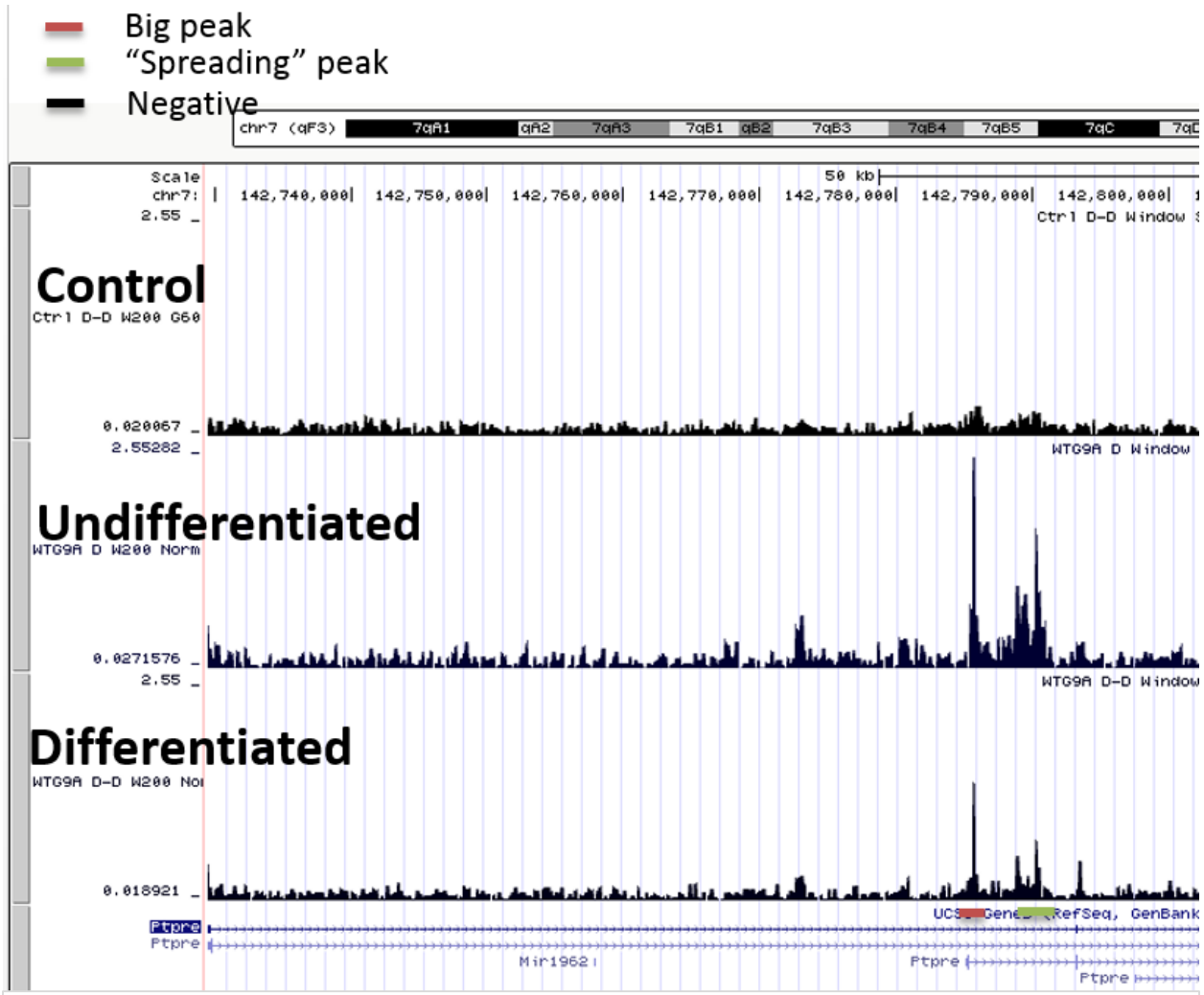

3.5
3
2.5
2
1.5
1
0.5
0
Rbfox3
MelTR Dox
MelTR D/D
WT B Dox
WT B D/D
WA B Dox
WA B D/D

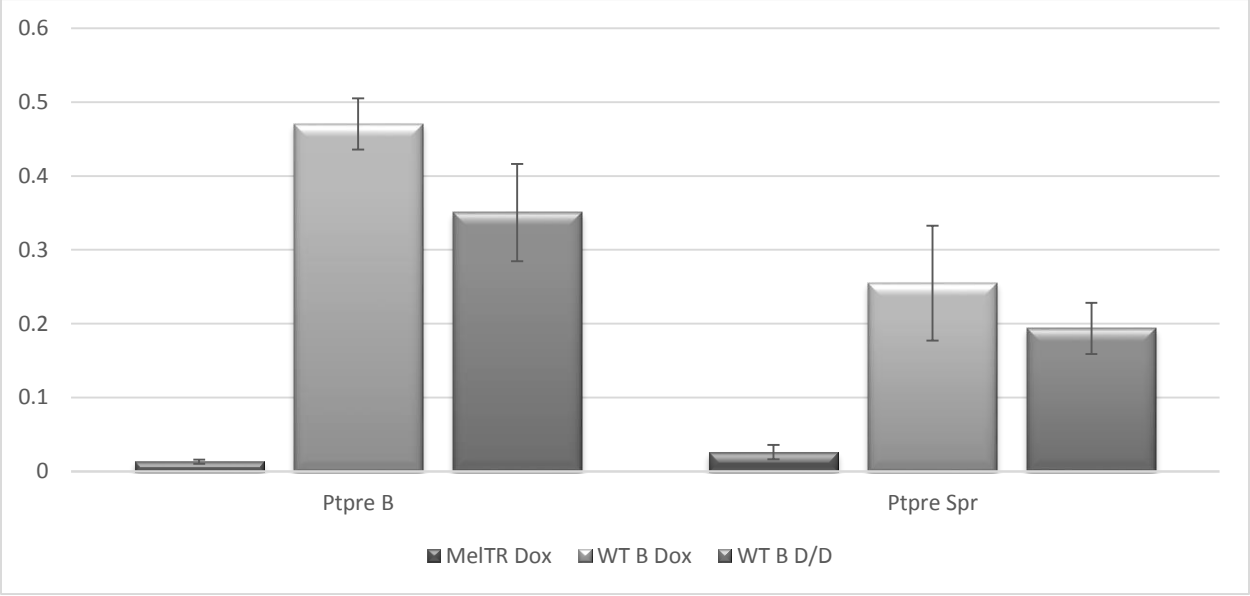
B


Big peak
"Spreading" peak
Negative
chr5 (qA2)
5qA1
5qA3
5qB1
B2
5qB3
5qC1
C2
Scale
chr5:
1.248
15,900,000
100 kb
Control
Ctrl D-D W200 G60
0.020067
1.248
Undifferentiated
WTG9A D W200 Norm
0.027158
1.24681
Differentiated
WTG9A D-D W200 Nor
0.0189214
Cacna2d1
Cacna2d1
Cacna2d1
Cacna2d1
Cacna2d1
UCSC Gen

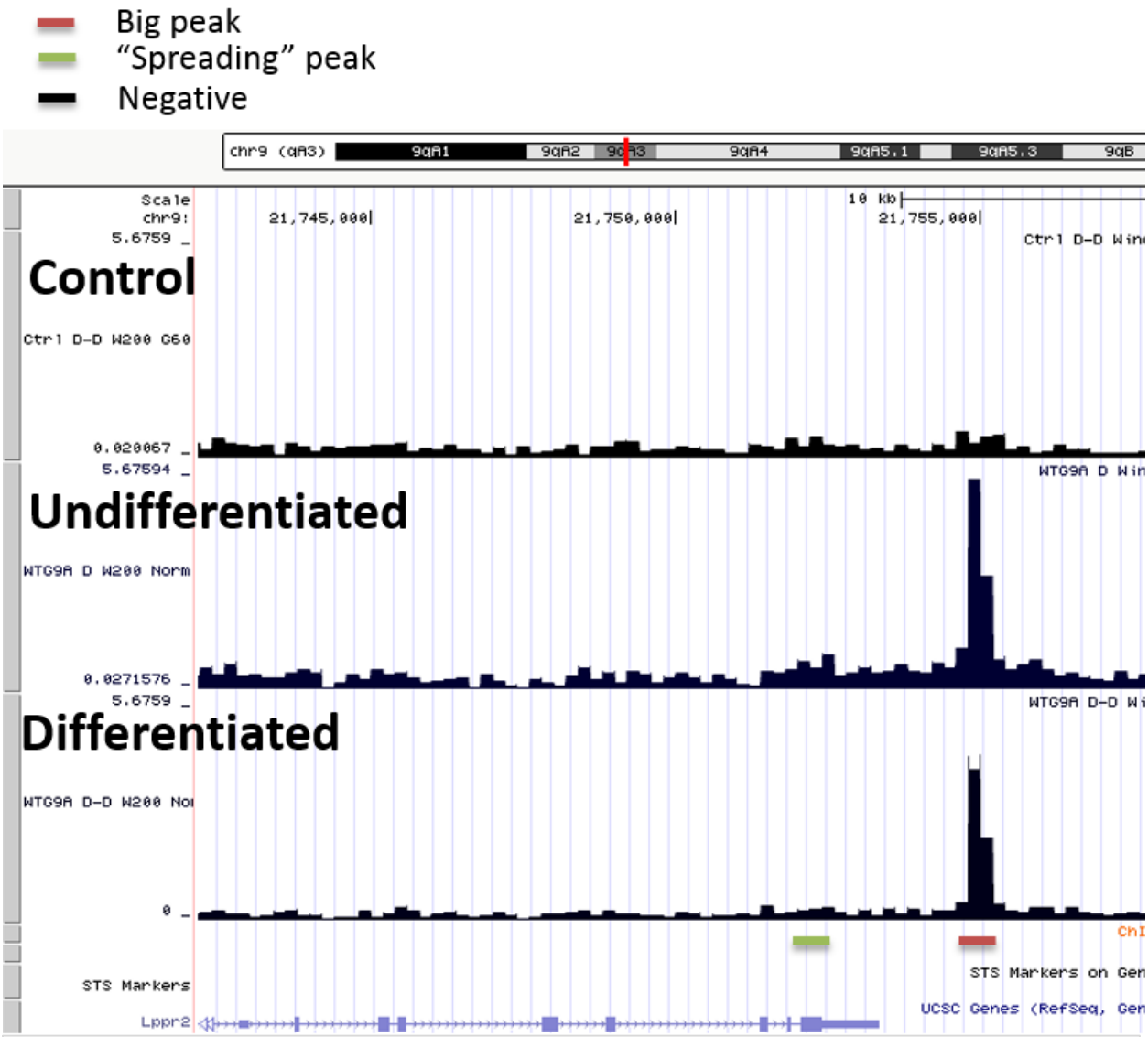

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.01
CACNA B
CACNA Spr
CACNA neg
MelTR Dox
MelTR D/D
WT B Dox
WT B D/D
WA B Dox
WA B D/D

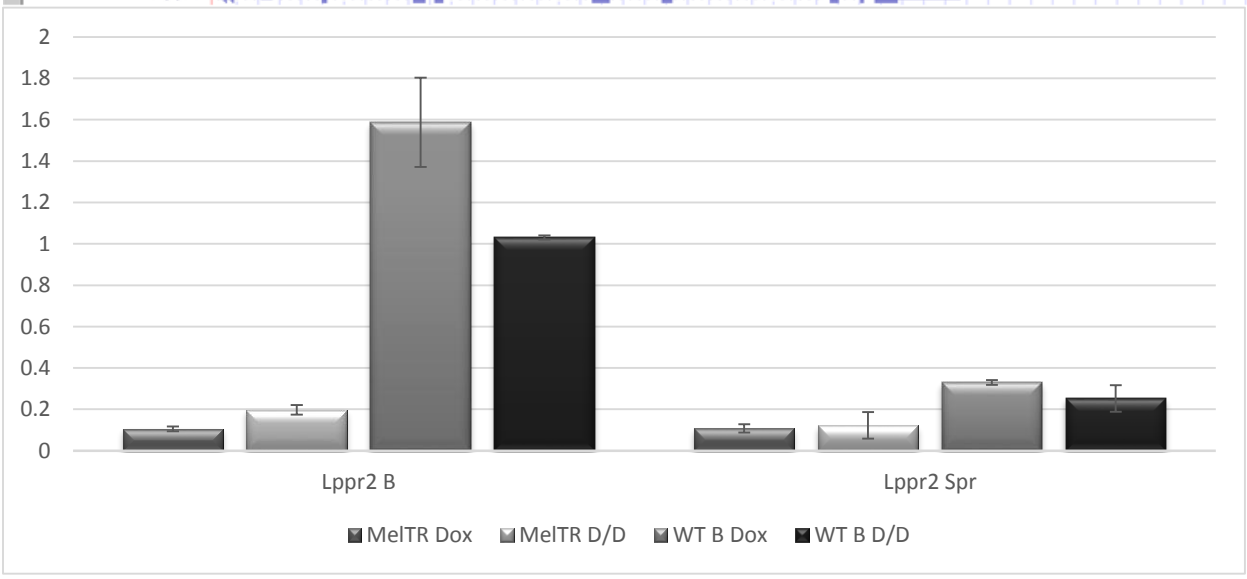
C


Big peak
"Spreading" peak
Negative
chr7 (qF3)
Scale
chr7:
142,740,000
142,750,000
142,760,000
142,770,000
142,780,000
142,790,000
142,800,000
50 kb
Control
Ctrl D-D W200 G60
Undifferentiated
WTG9A D W200 Norm
Differentiated
WTG9A D-D W200 Nor
RefSeq, GenBank
Ptpre
Mir1962
Ptpre


0.6
0.5
0.4
0.3
0.2
0.1
0
Ptpre B
Ptpre Spr
MelTR Dox
WT B Dox
WT B D/D

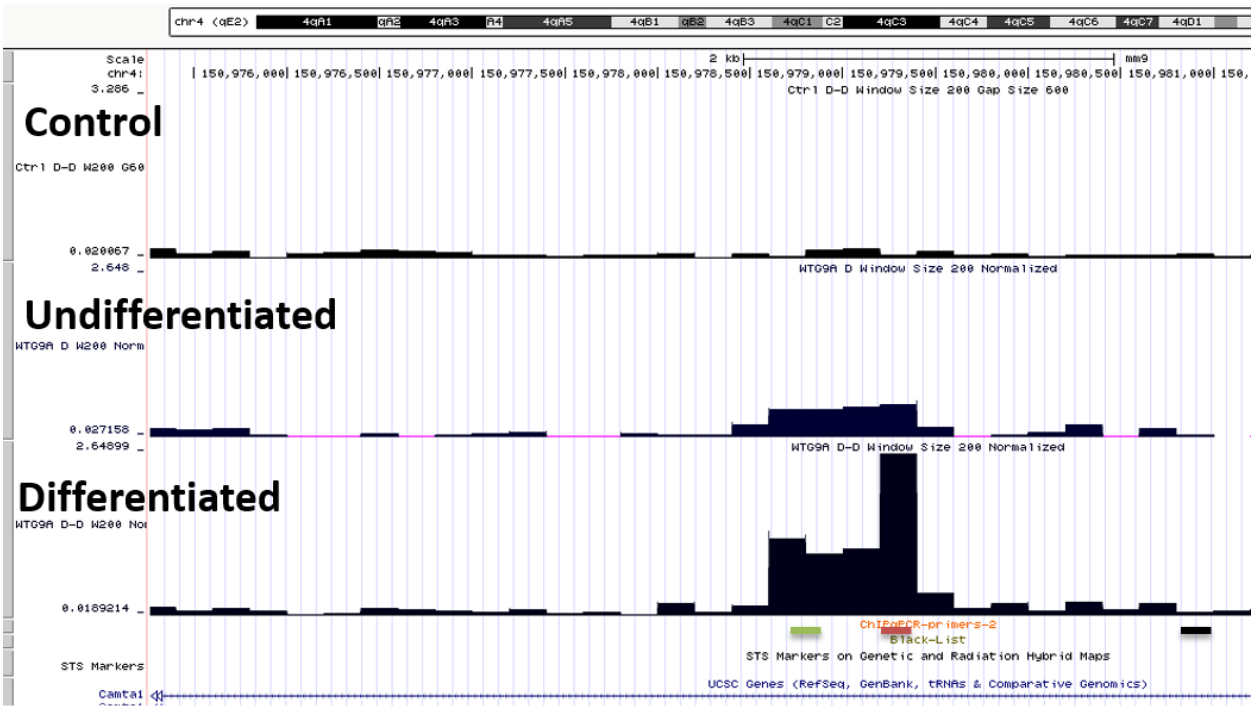
D

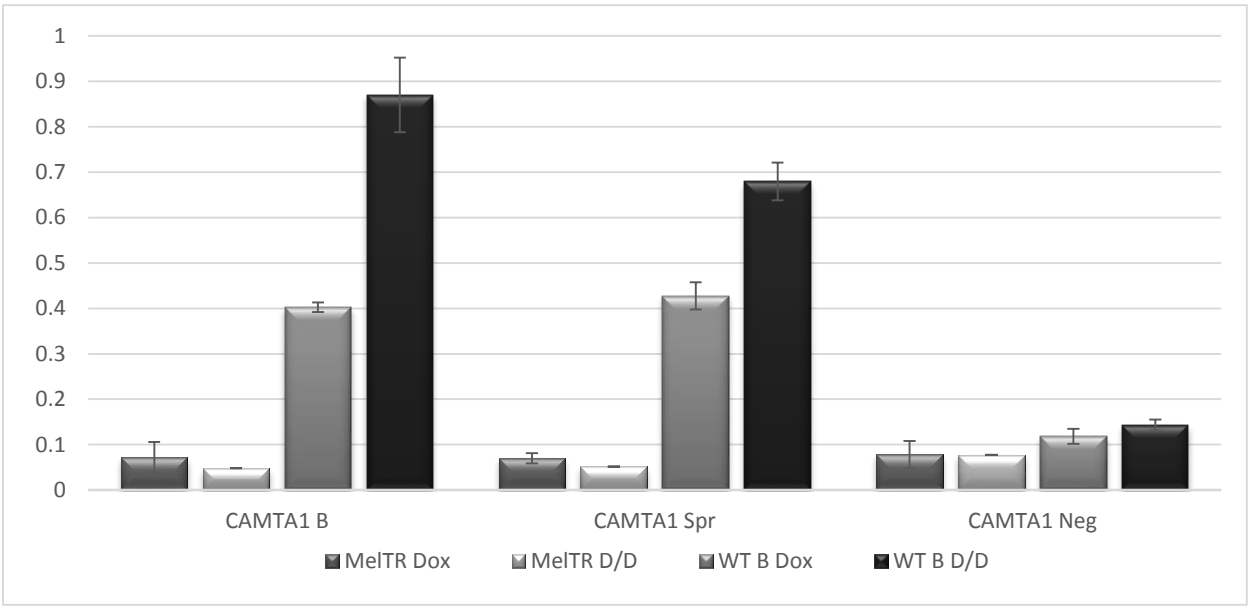

Big peak
"Spreading" peak
Negative
chr9 (qA3)
9qA1
9qA2
9qA3
9qA4
9qA5.1
9qA5.3
9qB
Scale
chr9:
21,745,000
21,750,000
21,755,000
10 kb
Control
Ctrl D-D W200 G60
Undifferentiated
WTG9A D W200 Norm
Differentiated
WTG9A D-D W200 Nor
STS Markers
Lppr2
STS Markers on Gen
UCSC Genes (RefSeq, Gen


2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Lppr2 B
Lppr2 Spr
MelTR Dox
MelTR D/D
WT B Dox
WT B D/D

E

Big peak
"Spreading" peak
Negative





F

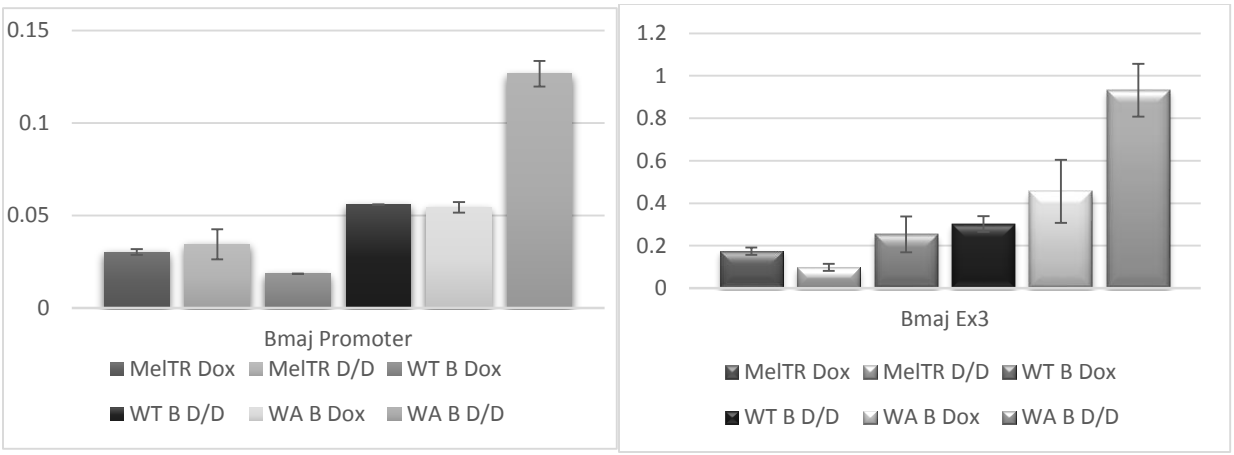


Figure 22. V5-G9a wild-type and Ankyrin mutant binding analysis.

The analysis is presented in two panels: top panel is visualization of ChIP-seq data on the UCSC browser and bottom panel is the qPCR amplification of DNA pulled down together with V5-G9a from the chromatin. The graph represents a relative (to input) enrichment of V5-G9a on the target genomic region.

A. ***Rbfox3*** **intron peak. Test region. Strong binding of G9a increased after differentiation. ChIP reproduced the results of ChIP-sequencing, W844A mutant had a higher enrichment of G9a accordingly to the higher vector expression. B. Peaks downstream of** ***Cacna2d1*** **promoter. W844A G9a mutant was bound on both big and spreading peaks. C.** ***Ptpre*** **intron peaks. The binding of G9a decreased after differentiation. ChIP reproduced the results of ChIP-sequencing on both targets. D. Spreading peak on** ***Lppr2*** **promoter and big peak downstream of**

promoter. G9a binding decreased after differentiation, what was confirmed by ChIP. E. *Camta1* intron peaks. There was a weak G9a enrichment in undifferentiated cells which resulted in two different peaks after differentiation. The data was reproduced for big, spreading and negative targets. F. β_{maj} promoter and exon 3. The mutant G9a bound on both targets.

Conditions: Dox – Doxycycline (Induced, Undifferentiated); D/D – DMSO + Doxycycline (Induced, Differentiated).

Cell lines: Mel TR – parental clone, expressing tetracycline repressor; WT B – V5-G9a wildtype batch cells; WA B – V5-G9a W844A batch cells.

Region name abbreviations: B – big peak; Spr – spreading peak, Neg – negative region.

Discussion

G9a Ankyrin domain is not the only factor contributing to *in vivo* binding on chromatin

The studies of intermolecular interactions and mechanisms which govern them are as much fascinating as challenging. An approach to discover the biochemical mechanism is more like solving the equation without knowing the variables. However, detecting these variables will eventually contribute to discovery of the fundamental mechanism. The goal of the present project is to elucidate the event of generation of repressive chromatin domains. This event includes heterochromatinization which is preceded by the spreading of repressor proteins over the large chromatin domains. Among these proteins are G9a and GLP methyltransferases.

The fact that heterochromatin protein 1 (HP1) recognizes methylated H3K9 marks (Lachner, O'Carroll et al. 2001) and also interacts with self-methylation site of G9a and GLP (Sampath, Marazzi et al. 2007) lead to idea of the spreading mechanism of G9a/GLP complexes. Despite the extensive research, the way of G9a/GLP spreading over the chromatin was not yet determined. However, the recent studies of crystal structures of G9a and GLP discovered the property of Ankyrin repeat domain to recognize and bind H3K9^{me1-2} marks (Collins, Northrop et al. 2008). This made us wonder if it is the way of G9a/GLP complexes to stabilize on nucleosomes *in vivo*.

This hypothesis is also supported by the fact that H3K9^{me1-2} marks are the products of SET domain what could allow G9a/GLP complexes to act independently from the DNA-binding activators. Previously, the studies of β-globin locus in our lab discovered that G9a is recruited to LCR in a NF-E2/p45-dependent manner and then propagates along the whole locus in the absence of NFE2/p45 binding (Chaturvedi, Hosey et al. 2009). Another H3K4^{me3} methyltransferase MLL2 possesses the same properties of binding on the β-globin locus: it is recruited to LCR and then it propagates throughout the whole locus (Demers, Chaturvedi et al. 2007). That suggests that spreading may apply also to creation of large domains of active chromatin. Based on that finding the model of G9a/GLP spreading was proposed whereby G9a is recruited by DNA-binding activator followed by methylation of nearby nucleosome and recognition of that mark by Ankyrin domain. This process could repeat for the following nucleosomes until the complex reach modified H3 which prevents the modification or other proteins with gate-keeping function. Therefore, to test this hypothesis we generated the cell line with the expression of exogenous G9a containing mutation which prevents Ankyrin domain to bind to H3K9^{me1-2} marks. We performed ChIP-sequencing to locate G9a targets genome-wide with the characteristics of spreading mechanism: 1) G9a should be recruited to the

gene; 2) G9a should propagate mono- or bidirectionally from the site of recruitment. In order to visualize properly the regions of G9a spreading we used SICER tool for ChIP-seq raw data analysis which allows identification of enriched domains, such as histone modifications. We found significant changes in G9a target distribution between states before and after differentiation suggesting an active role of G9a in erythroid differentiation by regulating gene transcription. Moreover, validation of ChIP-seq data confirmed all tested regions enriched with G9a providing a reliable tool to visualize G9a targets in MEL cells. We also identified genomic regions with anticipated G9a spreading potential. These regions and β-like globin genes were tested for mutant G9a binding *in vivo* via Chromatin Immunoprecipitation assay. However, all tested regions (including β-globin locus) were enriched with mutant G9a. This result leads to the conclusion that Ankyrin domain is not significantly contributing to the *in vivo* binding of G9a on chromatin suggesting that other proteins domains/mechanisms could have this role. For instance, it is possible that GLP could compensate for the binding property of G9a with its own Ankyrin domain. Another possibility is that other G9a/GLP-interacting proteins may participate in binding and stabilization of the complex on chromatin *in vivo*. Another

interesting question is whether the mechanism of G9a/GLP spreading is conserved on other G9a/GLP-dependent developmental genes.

There is high possibility that the Ankyrin domain of GLP can also contribute to the spreading mechanism of the complex. The fact that GLP and G9a form a heteromeric complex *in-vivo* supports this idea. To prove the hypothesis that both GLP and G9a participate in binding to H3K9^{me2} marks we would need a double knockout of wild-type GLP and G9a and double re-expression of differentially tagged mutant forms. It is also possible that other GLP/G9a-ineracting partners can participate in the complex stabilization on the chromatin. Some of them are LSD1 and TRIM28. Preliminary results obtained by me (data not shown) have suggested that these proteins are also located on the G9a spreading regions of the β-globin locus. The further investigation of their cross-talk with GLP/G9a complex may shed light on G9a/GLP spreading mechanism. Lastly, we cannot be sure that the mechanism of GLP/G9a spreading is the same on all G9a-dependent developmental genes. However, the general features may be similar, such as initial recruitment by DNA-binding activators. The difference may be in the co-factors participating in maintenance of this mechanism on different genes.

In order to address the question whereby GLP is complementing G9a binding function, we could use a novel RNA-guided endonuclease technology for genome editing – CRISPR/Cas9. Type II CRISPR system consists of two major elements. First element is a guide RNA (gRNA), a combination of crRNA complementary to the target sequence and tracrRNA serving as a scaffold for Cas9 (Jinek, Chylinski et al. 2012). And the second element is Cas9 endonuclease itself capable of creating Double Strand Breaks (DSB) at the target site (Jinek, Jiang et al. 2014). These DSBs can be repaired by nonhomologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ pathway is more error-prone and usually causes random nucleotide insertions or deletions leading to frameshift mutations or premature stop codons able to disrupt the open reading frame of the target gene. HDR pathway takes place if the repair template is accessible in the cell and often used within CRISPR/Cas9 system to create modifications at the Cas9 induced DSB. The template with specific modifications can be introduced to the cells together with CRISPR/Cas9 system using transient plasmid transfection (Hsu, Lander et al. 2014).

This system has many applications, but we will use it to introduce mutations to the Ankyrin domains of endogenous G9a and GLP. Due to the multiplexed targeting of Cas9 (Cong, Ran et al. 2013) we will be able to

modify both genes using different gRNAs covering W844 and W880 environment. The templates for HDR will contain a substitution of tryptophan (TGG) to alanine (GCG). The plasmids with Cas9 and gRNA sequences will be transfected to the MEL cells along with HDR-plasmids containing modified templates. After experimentally confirmed transfection and gene modification we will be able to assess the binding of mutant G9a and GLP.

This approach will help us answer the questions: 1) GLP and G9a are able to complement each other's binding function via Ankyrin domains; 2) the spreading mechanism of GLP/G9a is dependent on the Ankyrin domain interaction with H3K9$_{me2}$ marks.

References

Antoniou, M. (1991). "Induction of Erythroid-Specific Expression in Murine Erythroleukemia (MEL) Cell Lines." Methods Mol Biol 7: 421-434.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire and T. Kouzarides (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature 410(6824): 120-124.

Baron, R., P. Setny and J. A. McCammon (2010). "Water in cavity-ligand recognition." J Am Chem Soc 132(34): 12091-12097.

Blus, B. J., K. Wiggins and S. Khorasanizadeh (2011). "Epigenetic virtues of chromodomains." Crit Rev Biochem Mol Biol 46(6): 507-526.

Bongrand, P. and B. Malissen (1998). "Quantitative aspects of T-cell recognition: from within the antigen-presenting cell to within the T cell." Bioessays 20(5): 412-422.

Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase

A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell 84(6): 843-851.

Chaturvedi, C. P., A. M. Hosey, C. Palii, C. Perez-Iratxeta, Y. Nakatani, J. A. Ranish, F. J. Dilworth and M. Brand (2009). "Dual role for the methyltransferase G9a in the maintenance of beta-globin gene transcription in adult erythroid cells." Proc Natl Acad Sci U S A 106(43): 18303-18308.

Chaturvedi, C. P., B. Somasundaram, K. Singh, R. L. Carpenedo, W. L. Stanford, F. J. Dilworth and M. Brand (2012). "Maintenance of gene silencing by the coordinate action of the H3K9 methyltransferase G9a/KMT1C and the H3K4 demethylase Jarid1a/KDM5A." Proc Natl Acad Sci U S A 109(46): 18845-18850.

Collins, R. and X. Cheng (2010). "A case study in cross-talk: the histone lysine methyltransferases G9a and GLP." Nucleic Acids Res 38(11): 3503-3511.

Collins, R. E., J. P. Northrop, J. R. Horton, D. Y. Lee, X. Zhang, M. R. Stallcup and X. Cheng (2008). "The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules." Nat Struct Mol Biol 15(3): 245-250.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science 339(6121): 819-823.

Demers, C., C. P. Chaturvedi, J. A. Ranish, G. Juban, P. Lai, F. Morle, R. Aebersold, F. J. Dilworth, M. Groudine and M. Brand (2007). "Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus." Mol Cell 27(4): 573-584.

Friend, C., W. Scher, J. G. Holland and T. Sato (1971). "Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide." Proc Natl Acad Sci U S A 68(2): 378-382.

Ganguly, S. and A. I. Skoultchi (1985). "Absolute rates of globin gene transcription and mRNA formation during differentiation of cultured mouse erythroleukemia cells." J Biol Chem 260(22): 12167-12173.

Hosey, A. M., C. P. Chaturvedi and M. Brand (2010). "Crosstalk between histone modifications maintains the developmental pattern of gene expression on a tissue-specific locus." Epigenetics 5(4): 273-281.

Hsu, P. D., E. S. Lander and F. Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering." Cell 157(6): 1262-1278.

Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet 33 Suppl: 245-254.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science 337(6096): 816-821.

Jinek, M., F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales and J. A. Doudna (2014). "Structures of Cas9 endonucleases reveal RNA-mediated conformational activation." Science 343(6176): 1247997.

Lachner, M., D. O'Carroll, S. Rea, K. Mechtler and T. Jenuwein (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature 410(6824): 116-120.

Luger, K., M. L. Dechassa and D. J. Tremethick (2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" Nat Rev Mol Cell Biol 13(7): 436-447.

Marks, P. A. and R. A. Rifkind (1978). "Erythroleukemic differentiation." Annu Rev Biochem 47: 419-448.

Moon, A. M. and T. J. Ley (1990). "Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions." Proc Natl Acad Sci U S A 87(19): 7693-7697.

Musselman, C. A., M. E. Lalonde, J. Cote and T. G. Kutateladze (2012). "Perceiving the epigenetic landscape through histone readers." Nat Struct Mol Biol 19(12): 1218-1227.

Nishigaki, K., C. Hanson, T. Ohashi, A. Spadaccini and S. Ruscetti (2006). "Erythroblast transformation by the friend spleen focus-forming virus is associated with a block in erythropoietin-induced STAT1 phosphorylation and DNA binding and correlates with high expression of the hematopoietic phosphatase SHP-1." J Virol 80(12): 5678-5685.

Nishigaki, K., D. Thompson, C. Hanson, T. Yugawa and S. Ruscetti (2001). "The envelope glycoprotein of friend spleen focus-forming virus

covalently interacts with and constitutively activates a truncated form of the receptor tyrosine kinase Stk." J Virol 75(17): 7893-7903.

Paul, R., S. Schuetze, S. L. Kozak, C. A. Kozak and D. Kabat (1991). "The Sfpi-1 proviral integration site of Friend erythroleukemia encodes the ets-related transcription factor Pu.1." J Virol 65(1): 464-467.

Peters, A. H., S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. Martens and T. Jenuwein (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin." Mol Cell 12(6): 1577-1589.

Prockop, L. D. and R. I. Chichkova (2007). "Carbon monoxide intoxication: an updated review." J Neurol Sci 262(1-2): 122-130.

Raghavan, S., R. A. Desai, Y. Kwon, M. Mrksich and C. S. Chen (2010). "Micropatterned dynamically adhesive substrates for cell migration." Langmuir 26(22): 17733-17738.

Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis and T. Jenuwein (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature 406(6796): 593-599.

Rice, J. C., S. D. Briggs, B. Ueberheide, C. M. Barber, J. Shabanowitz, D. F. Hunt, Y. Shinkai and C. D. Allis (2003). "Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains." Mol Cell 12(6): 1591-1598.

Sampath, S. C., I. Marazzi, K. L. Yap, S. C. Sampath, A. N. Krutchinsky, I. Mecklenbrauker, A. Viale, E. Rudensky, M. M. Zhou, B. T. Chait and A. Tarakhovsky (2007). "Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly." Mol Cell 27(4): 596-608.

Shinkai, Y. and M. Tachibana (2011). "H3K9 methyltransferase G9a and the related molecule GLP." Genes Dev 25(8): 781-788.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature 403(6765): 41-45.

Tachibana, M., Y. Matsumura, M. Fukuda, H. Kimura and Y. Shinkai (2008). "G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription." EMBO J 27(20): 2681-2690.

Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato and Y. Shinkai (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3

lysine 9 methylation and is essential for early embryogenesis." Genes Dev 16(14): 1779-1791.

Tachibana, M., J. Ueda, M. Fukuda, N. Takeda, T. Ohta, H. Iwanari, T. Sakihama, T. Kodama, T. Hamakubo and Y. Shinkai (2005). "Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9." Genes Dev 19(7): 815-826.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science 272(5260): 408-411.

Teif, V. B. and K. Bohinc (2011). "Condensed DNA: condensing the concepts." Prog Biophys Mol Biol 105(3): 208-222.

Teif, V. B. and K. Rippe (2010). "Statistical-mechanical lattice models for protein-DNA binding in chromatin." J Phys Condens Matter 22(41): 414105.

Ueda, J., M. Tachibana, T. Ikura and Y. Shinkai (2006). "Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP." J Biol Chem 281(29): 20120-20128.

Vincent, J. A., T. J. Kwong and T. Tsukiyama (2008). "ATP-dependent chromatin remodeling shapes the DNA replication landscape." Nat Struct Mol Biol 15(5): 477-484.

Wickstead, B. and K. Gull (2011). "The evolution of the cytoskeleton." J Cell Biol 194(4): 513-525.

Yoshimura, A., A. D. D'Andrea and H. F. Lodish (1990). "Friend spleen focus-forming virus glycoprotein gp55 interacts with the erythropoietin receptor in the endoplasmic reticulum and affects receptor metabolism." Proc Natl Acad Sci U S A 87(11): 4139-4143.

Zang, C., D. E. Schones, C. Zeng, K. Cui, K. Zhao and W. Peng (2009). "A clustering approach for identification of enriched domains from histone modification ChIP-Seq data." Bioinformatics 25(15): 1952-1958.

Appendix

GLP shRNA sequences

GLP shRNA1 Forward

gatcccGGCATATACCTTTCTAAATTTCAAGAGAATTTAGAAAGGTATATGCCttttttggaaa

GLP shRNA1 Reverse

ggcctttccaaaaaGGCATATACCTTTCTAAATTCTCTTGAAATTTAGAAAGGTATATGCCgg

GLP shRNA2 Forward

gatcccCCTATATCTAGCCCTATATTTCAAGAGAATATAGGGCTAGATATAGGttttttggaaa

GLP shRNA2 Reverse

ggcctttccaaaaaCCTATATCTAGCCCTATATTCTCTTGAAATATAGGGCTAGATATAGGgg

GLP shRNA3 Forward

gatcccGCGAGAAAGCCTTAGATGATTCAAGAGATCATCTAAGGCTTTCTCGCttttttggaaa

GLP shRNA3 Reverse

ggcctttccaaaaaGCGAGAAAGCCTTAGATGATCTCTTGAATCATCTAAGGCTTTCTCGCgg

Primers to amplify GLP CDS (Added BlgII and SalI restriction sequences)

V5_GLP Forward:

5′ TCGAAGATCTATGGCCGCCGCTGATGCT 3′

V5_GLP Reverse:

5′ GATCGTCGACTCATAGGGGGTCAGCAGCG 3′

Primers for GLP CDS integrated into pcDNA5/TO

GLP Region_1 Forward

CGCAAATGGGCGGTAGGCGTG

GLP Region_2 Forward

AATGCACCCAAACACACTCAG

GLP Region_3 Forward

GCAGCTTTTACCAACCTTCCC

GLP Region_4 Forward

CAAGAAGAAATTTCTCAAGAGG

GLP Region_5 Forward

GTAACTTCATGGAATGCCAGC

GLP Region_6 Forward

CAAAGTAAGCGTTCCCCATTAC

GLP Region_7 Forward

AGGGAGAATCGCTACGACTG

GLP Region_8 Forward

TGGTGCAAAATGGTCTCAG

Custom primers specific to G9a binding peaks in MEL cells

Rbfox3_B:

Forward:
GAATGTCAGGTTGTTGTTACTGTC

Reverse:
CACGTTCTGCTTTAGGGATCT

Cacna_B:

Forward:
CACCATAGCTAGTCAGCCAAA

Reverse:
GCATGAACATCATGGAGGTAAAC

Cacna_Spr:

Forward:
CAGCAAGTACAACACCTCAATAAAG

Reverse:
ACAATCTACACCAGGAAAGTAAGG

Cacna_neg:

Forward:
GCTGCATCACAGAACAAACC

Reverse:
CTTAACACAAGGGCACAGAGTA

Ptpre_B:

Forward:
AGATGCAGCTAGGGAGATACT

Reverse:
TGTGTCCTCCAGACTTCCT

Ptpre_Spr:

Forward
GCTAGCTGTGAGGTGTGAAG

Reverse:
AAAGCCAGCATCCTCCAAA

Ptpre_neg:

Forward:
GTGGAGCAGGGCATACTTT

Reverse:
TCTGTGAACTCAGCAGGAATC

Camta1_B:

Forward
GGGCTGTGATTTATAGGAGTGAG

Reverse:
TCTGCAGGTGTGTGCATAG

Camta1_Spr:

Forward
TCCCAAATGTTCGCTTGGA

Reverse:
CAGTGACCGTTCAGAACCAG

Camta1_neg:

Forward:
CCCTTGTTGGTAAGCATTAAA

Reverse:
CCCAGAAAGTCAGATGGAAG

Lppr2_B:

Forward:
GAGCTTCTATCTTACCCGACCA

Reverse:
AACTGTGATTCCCTCCCTGA

Lppr2_Spr:

Forward:
GTCAGAAGATGGCTGGAAGG

Reverse:
GGGCCCACCTGGTTATTT

Buffers for Chromatin Immunoprecipitation

In red: Extemp: Add freshly while using the buffers for experiments

Glycine (2.5M)	100ml	(autoclaved, keep at RT)
Powder (75.07 g/mole)	18.77g	
H_2O	up to 100ml	

Swelling Buffer	50 ml	(Keep at 4^0C)
Hepes pH7.9 (25 mM)	2.5 ml of (0.5M) stock	
$MgCl_2$ (1.5 mM)	75 µl of (1M) stock	
KCl (10mM)	167 µl of (3M) stock	
NP-40 (0.1%)	0.5 ml of (10%) stock	
Extemp:		
DTT (1 mM)	50 µl	
PIC	use 1X final concentration	

Sonication Buffer no SDS	500ml	(Keep at 4^0C,protected from light)
Hepes pH7.9 (50mM)	25 ml of (1M) stock	
NaCl (140mM)	17.5ml of (4M) stock	
EDTA (1mM)	1ml of (0.5M) stock	
Triton X-100 (1%)	5ml of (100%) stock	
Na-deoxycholate (0.1%, w/v)	0.5g	
Extemp:		
PIC	use 1X final concentration	

Sonication Buffer 0.1% SDS	100ml	(Keep at 4^0C,protected from light)
Hepes pH7.9 (50mM)	10 ml of (0.5M) stock	
NaCl (140mM)	3.5 ml of (4M) stock or 2.8 ml of (5M) stock	
EDTA (1mM)	200 µl of (0.5M) stock	
Triton X-100 (1%)	1 ml of (100%) stock	
Na-deoxycholate (0.1%, w/v)	0.1 g	
SDS (0.1%, w/v)	0.1 g	
Extemp:		
PIC	use 1X final concentration	

Sonication Buffer 1% SDS 100ml (Keep at 4^0C,protected from light)

Hepes pH7.9 (50mM)	10 ml of (0.5M) stock
NaCl (140mM)	3.5 ml of (4M) stock or 2.8 ml of (5M) stock
EDTA (1mM)	200 µl of (0.5M) stock
Triton X-100 (1%)	1 ml of (100%) stock
Na-deoxycholate (0.1%, w/v)	0.1 g
SDS (1%, w/v)	1 g
Extemp:	
PIC	use 1X final concentration

Sonicated λ DNA (0.5ug/ul) 0.5ml (Keep at -20^0C)

250ug (liquid) in 485ul	485ul
H_2O	15ul

Sonicate 10X for 30 sec. at Max. setting (1 min. on ice between each sonication)

Ovalbumine (100mg/ml) 1ml (keep at -20^0C)

Powder (Grade V from SIGMA)	0.1g
TE	up to 1ml

Wash Buffer A 100ml (Keep at 4^0C,protected from light)

Hepes pH.9 (50mM)	5 ml of (1M) stock
NaCl (500mM)	10ml of (5M) stock
EDTA (1mM)	0.2ml of (0.5M) stock
Triton X-100 (1%)	1ml of (100%) stock
Na-deoxycholate (0.1%, w/v)	0.1g
SDS (0.1%, w/v)	0.1g
Extemp:	
PIC	use 1X final concentration, to 49 ml buffer add
1ml of	50X stock

Wash Buffer B 100ml (Keep at 4^0C)

Tris pH8.0 (20mM)	1 ml of (2M) stock
EDTA (1mM)	0.2ml of (0.5M) stock
LiCl (250mM)	1.06g

(MW 42.39)
NP40 (0.5%) 5ml of (10%) stock
Na-deoxycholate (0.5%, w/v) 0.5g
Extemp:

PIC use 1X final concentration, to 49 ml buffer add 1ml of 50X stock

IP Buffer 100mM/150mM/500mM/1M

IP 100 and IP 500 buffers	100 mM	500 mM
25 mM Tris pH 7.9	25 mL	25 mL of 1 M
5 mM MgCl2	5 mL	5 mL of 1 M
10% Glycerol	100 mL	100 mL of 100%
X mM KCl	33.3 mL	166.67 mL of 3 M
0.1% NP40	10 mL	10 mL of 10%

- Make up to 1 L with water
Extemp:
DDT(0.3mM) 3ul/10ml of 1M stock
PIC use 1X final concentration, to 49 ml buffer add 1ml of 50X stock

Elution Buffer 100ml (Keep at RT)

Tris pH8.0 (50mM) 2.5 ml of (2M) stock
EDTA (1mM) 0.2ml of (0.5M) stock
SDS (1%) 10ml of (10%) stock

Tris pH7.4 (10mM) 10ml (Keep at RT)

Tris pH7.4 (10mM) 50ul of (2M) stock
H_2O up to 10ml