MAPPING EPIGENETIC FEATURES AND CISPLATIN-INDUCED ALTERATIONS IN CANCER GENOMES

by

ALICE NAI CHIN KUO

B.Sc., The University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2009

©Alice Nai Chin Kuo, 2009

ABSTRACT

Since the completion of human genome sequencing, scientists have attempted to decipher the interactions and distribution of functional elements in the genome in hopes of explaining the mechanisms behind cancer development. Recently, researchers have been studying epigenomics, as the reversible nature of epigenomic alterations has greatly aided the understanding of the underlying mechanisms in tumour development. The major epigenetic events include DNA methylation and histone modifications. The first part of my thesis focuses on studying the interactions among DNA methylation and various histone H3K9 modifications. At any given locus, there are 243 possible patterns of these epigenetic marks. Through global epigenetic mapping, I identified non-random distribution of these patterns and discovered that some specific patterns are more prominent than others. The fact that these histone marks are not distributed equally at large scale suggests that histone modification is not only a small scale phenomenon.

To extend studies of modification on DNA, the second part of my thesis focuses on studying cisplatin induced DNA adducts and their clearance. I achieved this by adapting the immunoprecipitation technique that was optimized in the first part of my thesis. My results showed non-random integration of cisplatin on DNA. Furthermore, from global clearance maps, I identified slow drug clearance and rapid drug clearance loci. This implies that there are some adduct formation sites that are being preferentially repaired over others.

With more interest on the interaction between chromatin structure and DNA damage, the global maps generated in my thesis provide a critical step in understanding the distribution of these events in the genome, and will contribute to establishing a biological model investigating effects of cisplatin induced alterations on chromatin modification structure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAS	atomic absorption spectroscopy
BAC	bacterial artificial chromosome
BER	base excision repair
CGH	comparative genomic hybridization
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation on microarray chip
CpG	Cytosine-phosphate-Guanine
C_T	threshold cycle
CXCR7	*chemokine orphan receptor 1*
DDAH1	*dimethylarginine dimethylaminohydrolase 1*
DHFR	*dihydrofolate reductase*
DNMT	DNA methyltransferase
DSBs	double strand breaks
HATs	histone acetyltransferases
HDACs	histone deacetylases
HDACi	histone deacetylase inhibitors
HMTs	histone methyltransferases
HP1	heterochromatin protein
HR	homologous recombination
H3K9	histone 3 lysine residue 9
H3K9ac	histone 3 lysine residue 9 acetylation
H3K9me1	histone 3 lysine residue 9 mono-methylation

H3K9me2	histone 3 lysine residue 9 di-methylation
H3K9me3	histone 3 lysine residue 9 tri-methylation
IN DNA	input reference DNA
IP DNA	immunoprecipitated DNA
kb	kilobase pair
LRRTM4	*leucine rich repeat transmembrane neuronal 4*
MBD2	methyl binding domain protein
MC2R	*melanocortin 2 receptor*
MeCP2	methyl CpG binding domain protein 2
MeDIP	methylated DNA immunoprecipitation
NER	nucleotide excision repair
NHEJ	non-homologous end joining
NSCLC	non small cell lung cancer
PIC	protease inhibitor cocktail
PMSF	phenylmethanesulfonyl fluoride
ProK	proteinase K
SIGMA	system of integrative genomic microarray analysis
SNR	signal to noise ratio
SSBs	single strand breaks
TCFL 5	*transcription factor-like 5 protein*
TF	transcription factor
3'UTR	3' untranslated region
5mC	5’-methylcytosine

ACKNOWLEDGEMENTS

I am most grateful to my supervisor, Dr. Wan Lam, for providing me with an enjoyable work environment and for his generous support towards my thesis project.

Many thanks to my supervisory committee, Dr. Richard Stokes, Dr. Carolyn Brown, and Dr. Calum MacAulay for valuable advice throughout my thesis work.

Thank you to all members of Wan Lam Lab for assistance, specifically, Dr. Bradley Coe and Ian Wilson for many stimulating discussions.

One part of the introduction chapter is modified from a collaborative work towards a published book chapter. I would like to thank the co-authors: Ian Wilson, Emily Vucic, Eric Lee, and Jonathan Davies for manuscript input and Dr. Calum MacAulay, Dr. Carolyn Brown, and Dr. Wan Lam for reviewing the manuscript before submission.

I would also like to thank Dr. Peggy Olive for discussion on the cisplatin project, and Dr. Jürgen Thomale for the R-C18 antibodies through collaboration.

DEDICATION

To Mom, Dad, Tina,
and Steve

CHAPTER 1 INTRODUCTION

1.1 EPIGENOMICS BACKGROUND[1]

Epigenomics refers to the genome-wide study of heritable changes, other than those alterations found in the DNA sequence. DNA methylation and histone modifications are two of the most studied epigenomic events. Epigenomic modifications are integral to the regulation of gene expression, chromatin organization, DNA integrity and stability (Callinan and Feinberg 2006; Esteller and Herman 2002).

1.1.1 DNA methylation

In mammalian cells, DNA methylation involves the cytosine in CpG dinucleotide sequences ("p" refers to the phosphate connecting "C" and "G" on the same strand). The 5-carbon position of cytosine base is modified to become 5'-methylcytosine (5mC). The spontaneous deamination of 5mC to uracil results in an under-representation of CpG dinucleotides in the genome. In normal tissues, three to four percent of all cytosines are methylated (Baylin 2005). CpG islands are regions rich in CpG dinucleotides that are often conserved through evolution and associated with gene promoter regions (Callinan and Feinberg 2006). Cancer cells display abnormal DNA methylation, whereby DNA is globally hypomethylated with focal hypermethylation at some CpG islands (Feinberg and Tycko 2004). While global hypomethylation may lead to genomic instability; hypermethylation of

[1] Section 1.1~1.3 are modified from published work (see Appendix I)

CpG islands is linked with the transcriptional silencing of associated genes (Feinberg and Tycko 2004).

1.1.2 Histone modification

Another significant epigenetic event is post-translational histone modification. Histones are proteins that enable the condensation of double-stranded supercoiled eukaryotic DNA into nucleosomes thus allowing for further folding of the DNA into chromatin structures. The histone core of nucleosomes consists of two copies each of H2A, H2B, H3, and H4 (Shilatifard 2006). Please also refer to Figure 1 in Marks *et al.* 2001 for illustration of the histone structure in nucleosomes as well as the post-translational modification (Marks *et al.* 2001). Post-translational modifications to the histone tails, including acetylation, methylation, and phosphorylation (Figure 1.1), determine whether the chromatin exists as euchromatin or heterochromatin (Schubeler *et al.* 2004; Strahl and Allis 2000). Euchromatin is loosely compacted where DNA is more accessible for transcription, while heterochromatin is tightly compacted and is better associated with transcriptional silencing, as illustrated in Figure 1.1. The level of chromatin compaction is ultimately regulated by modifications to both the protein and DNA components (Santos-Rosa and Caldas 2005). The term, *histone code*, was proposed to describe distinct combinations of histone modifications that regulate specific downstream events (Strahl and Allis 2000).

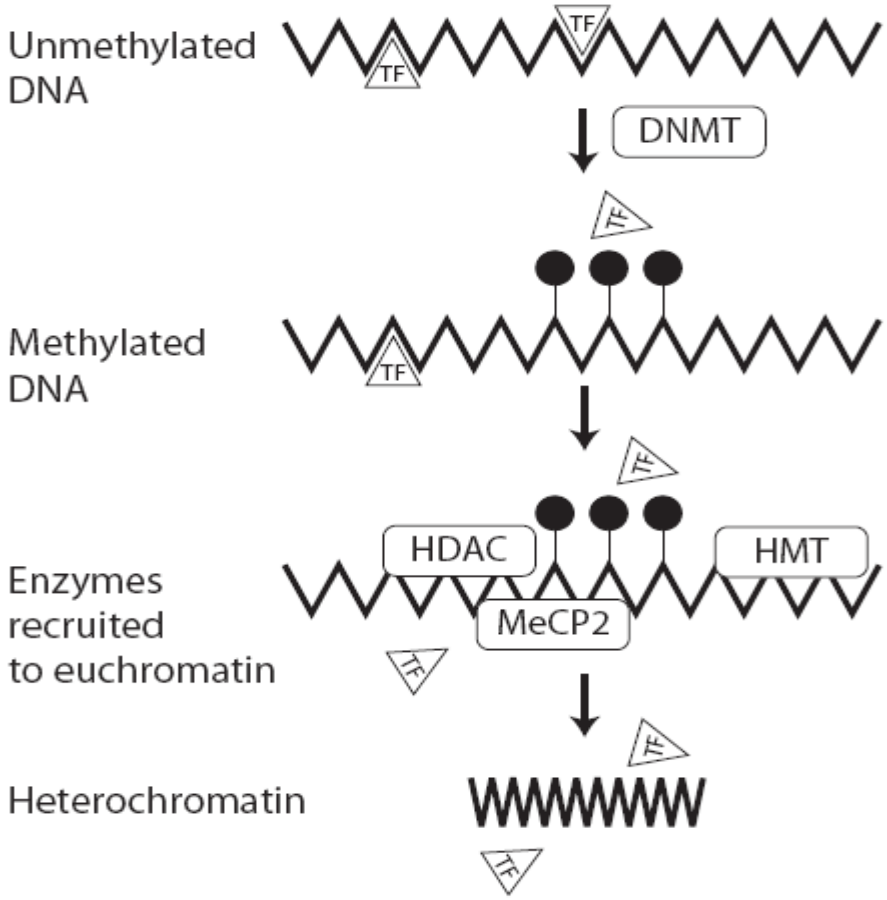


Figure 1.1 DNA methylation and heterochromatin formation.

DNA is methylated (represented as filled lollipops) via DNA methyltransferases (DNMT). Methylated DNA blocks the access of some transcription factors (TFs) to DNA. Methyl CpG binding protein 2 (MeCP2) and enzymes including histone deacetylases (HDACs) and histone methyltransferases (HMTs) are recruited to the loosely compacted DNA (euchromatin) forming a more tightly compacted DNA (heterochromatin). The condensed chromatin blocks TFs, resulting in gene silencing. Figure taken from Kuo *et al.* 2008, and reproduced with permission

1.1.3 Chromatin condensation regulates gene expression

DNA methylation affects histone modifications, which determine the level of chromatin condensation and, in turn, regulates gene transcription (as illustrated in Figure 1.1). DNA is methylated by DNA methyltransferases (DNMTs), and methylated DNA is recognized by methyl binding proteins such as methyl CpG binding domain protein 2 (MeCP2) and methyl binding domain protein (MBD2) (Feinberg and Tycko 2004; Jaenisch and Bird 2003). Heterochromatin is then formed by the removal of acetyl groups from the histone tails by histone deacetylases (HDACs), and the addition of methyl groups by histone methyltransferases (HMTs) with the transcriptional corepressor Sin3a. In contrast, histone acetyltransferases (HATs) are responsible for maintaining the open structure of chromatin for active transcription. Histone modifications can also self-reinforce gene silencing while cooperating with DNA methylations in the maintenance of heterochromatin (Feinberg and Tycko 2004). H3K9 methylated by HMTs is recognized by a heterochromatin protein (HP1), which recruits DNMTs resulting in methylation of DNA (Feinberg and Tycko 2004).

To date, histone H3 and H4 modifications have been most widely studied (Strahl and Allis 2000). Although modifications can be divided in correlation to active or repressive transcription (such as acetylation implicated in active transcription, and methylation implicated in repression), any given modification could potentially activate or repress transcription under any given condition (Kouzarides 2007). For example, removal of the acetyl group on lysine residue 9 of histone H3 (H3K9), gain of methyl group to H3K9 as well as H3K27 are usually associated with heterochromatin formation and transcription silencing (Santos-Rosa and Caldas 2005). In addition, the fact that lysine residues can accept up to

three methyl groups forming mono-, di-, and tri- methylation adds on to the complexity. Trimethylation of H3K4, acetylation of H3K9 (H3K9ac), H3K14ac and phosphorylation of serine residue 10 on H3 are associated with euchromatin and transcriptional initiation (Guenther *et al.* 2007; Schubeler *et al.* 2004). It has become more evident that these modifications act on a gene and location dependent manner regarding its effects on gene transcription (Barski *et al.* 2007).

1.2 CANCER EPIGENOMICS

Epigenetic events such as gene silencing and global hypomethylation can contribute to tumourigenesis.

1.2.1 Gene silencing

Normally, most CpG islands are unmethylated. However, CpG islands can become hypermethylated in cancer cells, which is thought to be involved with gene silencing (Jones and Baylin 2002). Examination of promoter hypermethylation of CpG islands revealed transcriptional silencing of genes in various cancer genomes (Baylin *et al.* 2000; Esteller *et al.* 2000; Shames *et al.* 2006; Yates *et al.* 2007). Aberrant promoter hypermethylation has also been suggested as an early event that may drive tumourigenesis (Fong *et al.* 2003; Suzuki *et al.* 2004; Tlsty *et al.* 2004). The timing of promoter hypermethylation makes CpG islands potential targets for early tumour detection, while tissue-specific methylation patterns may be useful in subclassifying specific tumour types and determining tissue of origin in metastases (Baylin *et al.* 2001; Costello *et al.* 2000; Esteller *et al.* 2001; Esteller 2002). Gain

of H3K9 methylation has also been associated with promoter CpG-island hypermethylation in cancer cells (Fahrner *et al.* 2002).

1.2.2 Global hypomethylation and hypomethylation of parasitic DNA sequences

Global hypomethylation has been associated with early stages of malignant transformation or throughout cancer progression (Eden *et al.* 2003; Jones and Baylin 2002; Wilson *et al.* 2007). This has a particularly large effect on repeat DNA sequences (Ehrlich 2006). With the exception of CpG islands, the CpG dinucleotides throughout the genome are normally methylated. The bulk of 5mC is found in repetitive/parasitic DNA sequences, such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and in centromeric satellite DNA (Ehrlich 2002). Methylation of these sequences is thought to be important for suppressing retrotransposition events, illegitimate recombination events, and inappropriate gene transcription from retroelement promoters/enhancers. The lack of this methylation may lead to increased genomic instability (Ehrlich 2002). Though CpG-island hypermethylation has largely been the focus of cancer research in the past, the importance of global hypomethylation may prove to play a significant role.

1.3 GENOME-WIDE TECHNOLOGIES FOR EPIGENETIC ANALYSIS

Many techniques have been developed for studying methylation at both locus-specific and genome-wide levels. Current methods used to study the epigenome are discussed below.

PCR based methods

Methylated DNA can be differentiated based on susceptibility to digestion by restriction enzymes and their 5mC sensitive isochizomers. A commonly used enzyme pair is *Hpa* II and *Msp* I. *Msp* I is not sensitive to DNA methylation, however *Hpa* II is. Using primers flanking restriction cut sites, PCR will only generate product in methylated samples that are treated with *Hpa* II (Gonzalgo *et al.* 1997). Bisulfite conversion can also differentiate methylated DNA by converting unmethylated cytosine to uracil while methylated cytosine is not affected. Sequencing of untreated and treated DNA identifies the 5mC positions (Frommer *et al.* 1992). Alternatively, this technique can be used in PCR application to distinguish between unmethylated and methylated loci.

Restriction landmark genomic scanning (RLGS)

RLGS combines the use of labeled genomic DNA digested with restriction enzymes and high resolution two dimensional gel electrophoresis. It can measure the DNA methylation level quantitatively in thousands of CpG islands separated based on restriction sites (Hatada *et al.* 1991).

Methylated DNA Immunoprecipitation (MeDIP)

Anti-5mC antibodies are used to enrich for methylated genomic DNA fragments. Methylation status is revealed by comparing the immunoprecipitated DNA (IP) to untreated DNA. DNA methylation profiles are generated by competitive co-hybridization to a whole genome resolution tiling path array (Weber *et al.* 2005; Wilson *et al.* 2006) (see Figure 1.2). This methods will be used for generating DNA methylation profiles in Chapter 2.

Chromatin immunoprecipitation (ChIP)

For histone modification studies, chromatin immunoprecipitation has been widely used by investigators to study the interaction between proteins and DNA in various organisms (Oberley *et al.* 2004; Ren *et al.* 2000; Weinmann *et al.* 2002; Ballestar *et al.* 2003). Antibodies against a specific histone modification state are used to enrich for the DNA sequences associated with the protein-DNA complex of interest. ChIP can be combined with genomic microarray (ChIP-chip) to assess global distribution of histone modification (Wu *et al.* 2006; Kondo *et al.* 2004; Miao and Natarajan 2005; Bernstein *et al.* 2005) (see Figure 1.3). This method will be used for generating global histone profiles in Chapter 2. Alternatively, ChIP DNA can be sequenced and mapped genome-wide (ChIP-Seq).

Both bisulfite treatment and enzyme digestion with methylation sensitive enzymes are suitable for locus specific studies. However, restriction enzyme based methods in particular are limited by the distribution of the recognition site. Immunoprecipitation based methods were developed to circumvent this issue, and were therefore used in my thesis to achieve my goal of generating global epigenetics maps. These methods are nowadays complemented with microarray technologies to allow researchers to evaluate the alterations in a high throughput manner. Even though there are many different selections of arrays, there were not many available at the start of the project. The in-house built whole-genome tiling path array was therefore picked to fulfill my global scale studies. To compare the epigenetic profiles, ChIP products were labeled and hybridized on the same whole genome tiling path arrays used for DNA methylation studies.

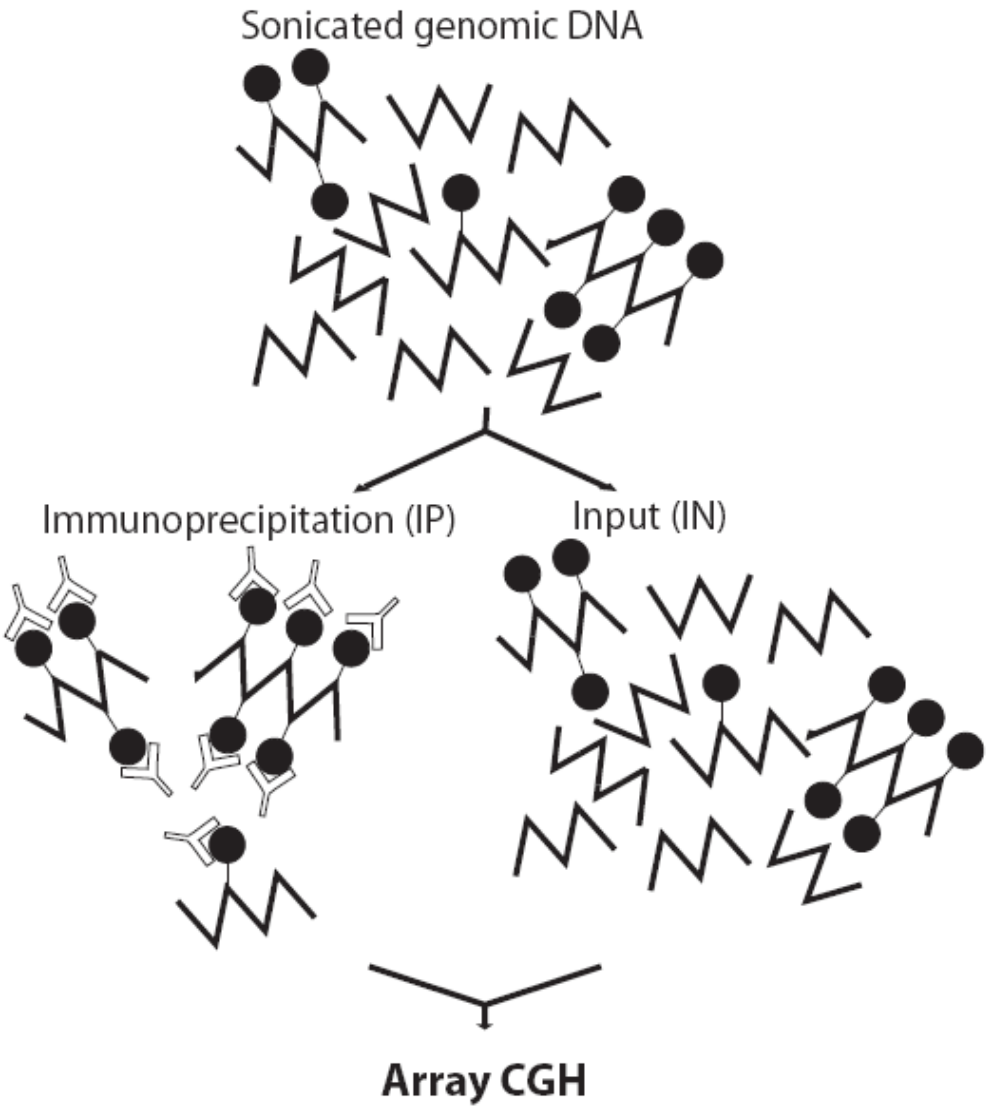


Figure 1.2 Scheme of methylated DNA immunoprecipitation (MeDIP).

Methylated DNA immunoprecipitation (MeDIP) uses anti-5mC antibodies to immunocapture methylated fragments of DNA. The immunoprecipitated DNA (IP DNA) and input reference DNA (IN DNA) are differentially labeled with different cyanine dyes and co-hybridized onto genomic targets on microarrays. Figure taken from Kuo *et al.* 2008, and reproduced with permission.

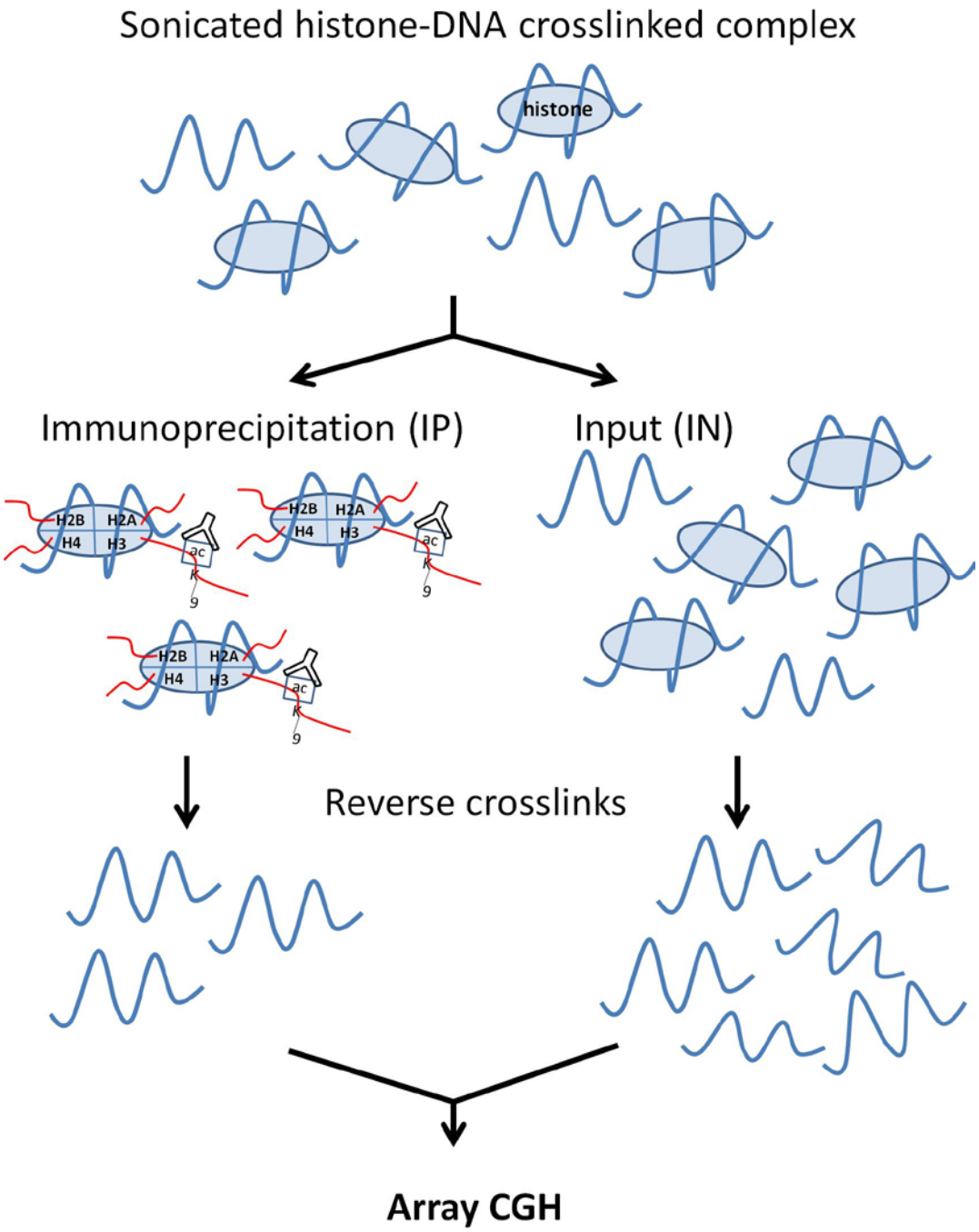


Figure 1.3 Chromatin immunoprecipitation on microarray chip (ChIP-chip).

Antibodies against specific histone residues are used in ChIP. The histone core (blue oval) has DNA wrapped around it. This figure illustrates antibodies that are against the H3K9ac residue (indicated by the box on the red histone tails). The crosslinks between DNA and histone are reversed before labeling the IN and IP DNA for array hybridization.

1.4 CISPLATIN AND DNA DAMAGE

1.4.1 DNA damage and cancer

Genome instability is an important feature of tumour development. DNA damage can lead to genome instability, which can predispose cells to tumourigenesis. DNA damage can result from ionizing radiation and certain chemotherapeutic drugs (exogenous factors) as well as byproducts from metabolism including reactive oxygen species and mechanical stress on the chromosome (endogenous factors) (Khanna and Jackson 2001). Cell cycle arrest or even apoptosis initiation can result from DNA damage (Helleday *et al.* 2008; Hoeijmakers 2001). Repair mechanisms within cells are activated upon the presence of DNA damage.

There are four main repair pathways in mammals: Nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR), and non-homologous end joining (NHEJ). Please refer to Hoeijmakers 2001 for detailed illustrations of the pathways (Hoeijmakers 2001). Both NER and BER are targeted to single strand breaks (SSBs), which happen three times more frequently than double strand breaks (DSBs) (Caldecott 2008). Most NER lesions are derived from exogenous factors whereas BER is mostly involved with damage from endogenous sources. DSBs are the most toxic DNA damage cells can sustain as both strands are affected. HR and NHEJ are the two pathways activated for DSBs repair. HR is an error-free process since it uses a sister chromatid as template for precise repair, and therefore seems to dominate in S and G2 phase when DNA is replicated. NHEJ is more prone to errors as it rejoins the broken DNA ends without sister chromatid as template, and is most prevalent in G1 phase when a second copy is not available (Hoeijmakers 2001).

There has been evidence supporting the idea that unrepaired SSBs in proliferating cells can result in DNA replication fork blockage or collapse, leading to the formation of DSBs (Caldecott 2008; Kuzminov 2001). Although cells can adapt to low levels of non-repairable damage, the consequences can be lethal if one DNA DSB happens to inactivate an essential gene or even activate apoptosis (Khanna and Jackson 2001; Rich *et al.* 2000). Failure to respond to DSBs properly and failure to repair DSBs (i.e. errors in rejoining the broken DNA DSBs) can lead to genomic instability, chromosomal translocation, gene deletions/gains and could enhance the rate of tumourigenesis (Khanna and Jackson 2001).

1.4.2 Cisplatin

Cisplatin is a commonly used therapeutic for several cancer types including ovarian, cervical, and head and neck cancers (Kelland 2007; Rosenberg *et al.* 1969). Cisplatin binds to the purine DNA bases covalently to form DNA adducts and subsequently disrupts DNA function. After cisplatin enters the cell, the chloride ions are replaced by water molecules (Wang and Lippard 2005). The positively charged platinum complex reacts with the nucleophilic sites on DNA. The covalent binding results in cisplatin-DNA adducts via intra-strand and inter-strand crosslinks which prevents DNA, RNA, and protein synthesis (Wang and Lippard 2005). There are two types of adducts: monofunctional and bifunctional adducts, with bifunctional adducts being the majority (Kelland 2007; Liedert *et al.* 2006). These adducts are created via five structures of the crosslinked complex, and have been identified and summarized in Table 1.1 and Figure 1.4 (Kelland 2007; Liedert *et al.* 2006). These outcomes lead to DNA distortion that may interfere with the replication fork and

eventually cause cellular apoptosis (Eastman 1990; Helleday *et al.* 2008; Kelland 2007). In chapter 3, I will describe how antibodies specific to one of the intrastrand adducts were utilized for part of my thesis research.

Table 1.1 Two types of cisplatin induced adducts: monofunctional and bifunctional adducts.

Types of adducts	**Strands**	**Structure**	**Of total adducts (%)**
Monofunctional adducts (via one Chlorine leaving group)		Guanine (Figure 1.4a)	~2%
Bifunctional adducts (via both Chlorine leaving groups)	Intrastrand	GpG 1, 2 intrastrand (Figure 1.4b)	60~65%
	Intrastrand	ApG 1, 2 intrastrand (Figure 1.4c)	20~25%
	Intrastrand	GpXpG 1, 3 intrastrand (Figure 1.4d)	~2%
	Interstrand	Guanine-Guanine on opposite side (Figure 1.4e)	~2%

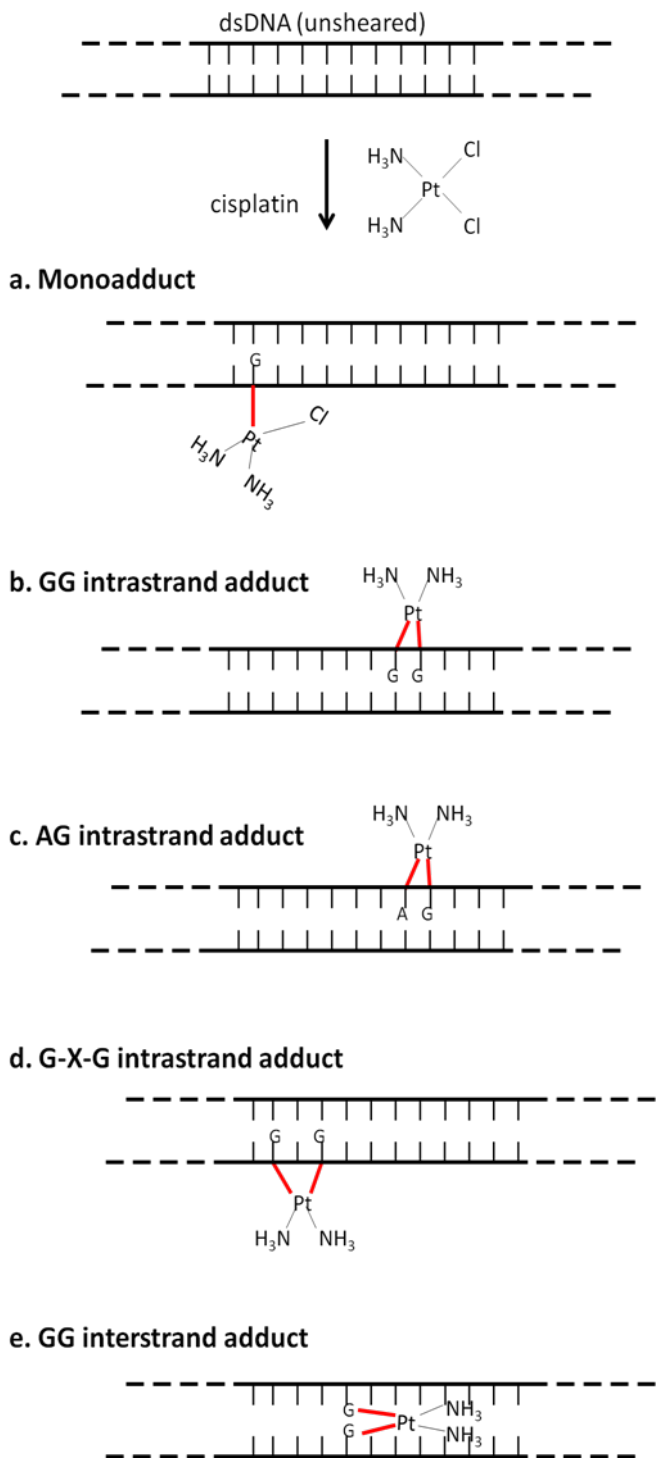


Figure 1.4 Five structures of cisplatin adducts.

Five major types of adducts are shown here. Refer to Table 1.1 for structural name and the frequency of adducts upon exposure to cisplatin.

1.5 PROJECT GOAL, AIMS, AND HYPOTHESIS

Goal: To map epigenetic features and cisplatin-induced alterations in cancer genomes

Hypotheses for this thesis:

Hypothesis 1- There exist non-random patterns of epigenetic modifications in the cancer genome. Integrative analysis of histone modification and DNA methylation will show relationships and patterns among these interactions.

Hypothesis 2- Cisplatin, a chemotherapeutic agent, induces DNA adducts in a non-random manner throughout the genome.

Hypothesis 3- The clearance of DNA adducts induced by cisplatin occurs non-randomly in the human genome.

The following aims have been designed in order to test the hypotheses:

Aim 1. To study the distribution patterns and relationships and among DNA methylation and various H3K9 histone modifications.

Aim 2. To generate spatial and temporal maps of cisplatin clearance.

CHAPTER 2 DELINEATION OF LARGE SCALE H3K9 MODIFICATIONS AND DNA METHYLATION PATTERNS IN LUNG CANCER CELLS

2.1 PROJECT GOAL

In order to understand the interplay between DNA methylation and histone modifications, it is imperative to look at the distribution of patterns among these epigenetic marks. As mentioned in the previous chapter, histone H3 and H4 have been widely studied. The three best characterized covalent modifications associated with a repressed chromatin state are DNA methylation, H3K9 deacetylation, and H3K9 methylation (Fuks 2005). However, the roles of various H3K9 modifications and DNA methylation in genome wide regulation remain unclear. The goal for this project is to map epigenetic features in cancer genomes by investigating the distribution of various H3K9 modifications in relation to DNA methylation in lung cancer cells.

2.2 METHODS

The aim for this project is to study the distribution patterns and relationships among DNA methylation and various H3K9 histone modifications. Chromatin immunoprecipitation profiling with a whole genome tiling path array was optimized, with integrative analysis of DNA methylation and multiple states of histone H3K9 modifications following. See Figure 2.1 for the experimental scheme.

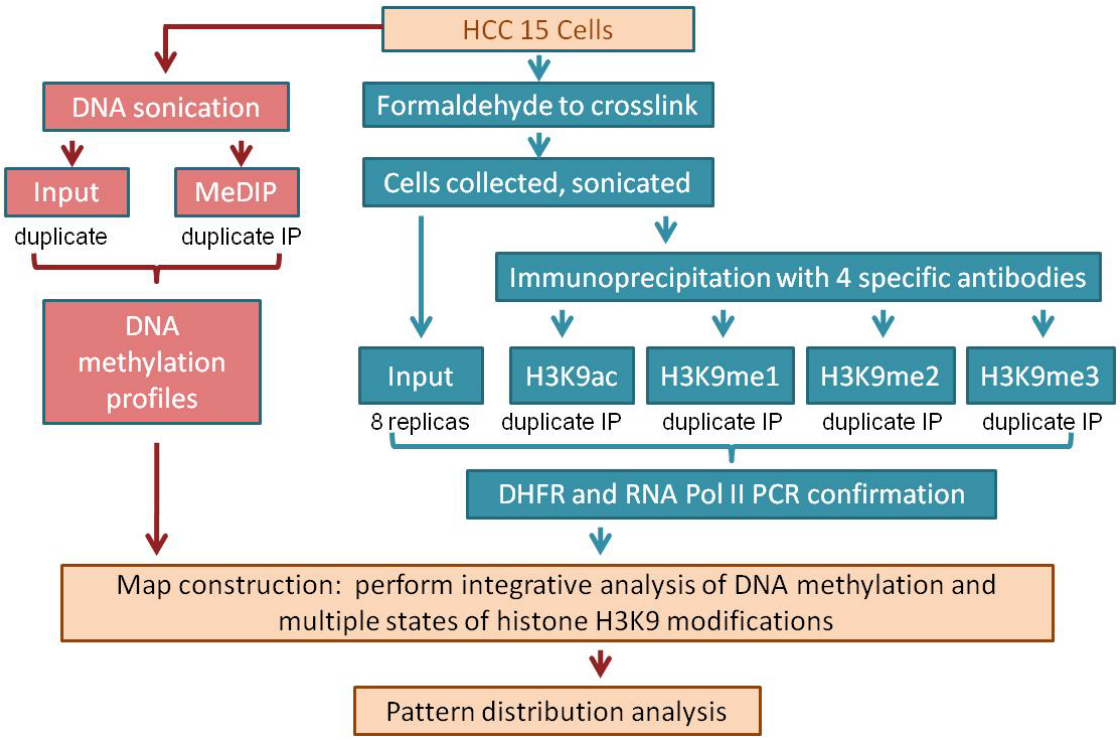


Figure 2.1 Experimental scheme of epigenetics mapping to achieve Aim 1.

Non small cell lung cancer (NSCLC) HCC15 cells were processed via MeDIP (see Section 1.3 and Figure 1.2) to generate DNA methylation profiles using whole genome tiling path array (pink), and via ChIP (blue) to generate H3K9 modification profiles using whole genome tiling path array (see Sections 1.3 and Figure 1.3). Integrative analysis was performed on generated DNA methylation and H3K9 global mapping results, with pattern distribution analysis following.

Methylated DNA Immunoprecipitation (MeDIP)

Sheared genomic DNA was immunoprecipitated with anti-5mC antibodies (CalBiochem) to isolate methylated genomic DNA fragments (Weber *et al.* 2005). . Since the sonication strength may vary with different sonicator brands and models, fragments size was verified on the agarose gel. Please refer to Thu *et al.* for detailed demonstration of the protocol (Thu *et al.* 2009). Briefly, genomic DNA was sonicated into fragments ranging 300~1000 bp in size, and divided into IP and IN portions. Purified methylated DNA fragments enriched by 5mC antibodies were then used to generate global DNA methylation profiles using whole genome tiling path array.

Chromatin Immunoprecipitation (ChIP)

Proteins and DNA were crosslinked by adding formaldehyde to culture medium to a final concentration of 1% formaldehyde for 10 minutes with swirling in cultured HCC15 cells (from Dr. Adi Gazdar, University of Texas Southwestern Medical Center, Dallas, Texas) . 1/10 volume of 1.25 M glycine was slowly added to quench unreacted formaldehyde. Flasks were then gently swirled and left on ice for 5 minutes. Cells were collected after two washes with 1X PBS containing 0.5 mM PMSF protease inhibitor. Cell suspension was subjected to centrifugation at 1,000 rpm for 5 minutes at 4°C. The supernatant was discarded and cell pellets were either subjected to sonication or stored at -80°C.

The cell pellets were resuspended in lysis buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1 mM PMSF, 1X protease

inhibitor cocktail (PIC)) (Bell *et al.* 2001). The suspension was sonicated to fragments ranging from 300 to 1000 bp in size. Fragment sizes were checked on an agarose gel.

The sonicated cross-linked fragments were centrifuged at 13,000 rpm for 10 minutes at 4°C. The pellet was discarded, and the sheared chromatin solution was precleared for 1 hour at 4°C with BSA solution pre-blocked (1XPBS, 0.5%BSA) sheep anti-rabbit IgG Dynabeads (Invitrogen). A small portion of the pre-cleared chromatin was set aside as input (IN). The rest of the portion was divided equally and incubated with 2 μg primary antibody (H3K9ac, H3K9me1, H3K9me2, or H3K9me3, Abcam) at 4°C with gentle rotation overnight. The next day, pre-blocked beads were added to each of the chromatin-primary antibody solutions and incubated with gentle rotation at 4°C for 2 hours. The beads were washed once each with lysis buffer, high salt buffer (50 mM HEPES pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1 mM PMSF, 1X PIC), wash buffer (50 mM HEPES pH 7.9, 250 mM LiCl, 1 mM EDTA, 0.5% NP40, 0.5% Na-deoxycholate, 1X PIC), and TE buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA, 1X PIC). Finally, the beads were suspended in 200μl of elution buffer (50 mM Tris-HCL pH 8.0, 10 mM EDTA, 1% SDS) and left at 65°C overnight. The eluates were treated with 10 μg of ribonuclease A (RNaseA, Invitrogen), 100 μg of proteinase K (Pro K, Invitrogen), and followed by phenol/chloroform extraction and ethanol precipitation to recover the ChIP fragments. Replicates were performed for each of the modifications studied.

Hybridization to whole genome tiling path array

The array consists of ~27,000 individual BAC clones spotted in duplicate representing complete, overlapping coverage of the sequenced human genome (Watson *et al.*

2007). The immunoprecipitated (MeDIP or ChIP) DNA and the reference input (IN) DNA were each labeled with fluorescent Cyanine (Cy) 3 and Cy5 dye, and competitively co-hybridized on the whole genome tiling path arrays. Refer to Ishkanian *et al*. for detailed procedures on DNA labeling and hybridization as well as array imaging and analysis (Ishkanian *et al.* 2004) and Kennett *et al.* for detailed demonstration of whole genome array CGH (Kennett *et al.* 2008).

PCR

Primers of 3' untranslated region (3'UTR) of *DHFR* gene and *RNA Polymerase II* promoter were used in PCR assays to assess the efficacy of ChIP pull down. 3'UTR of *DHFR* gene: product size 349 bp, forward primer 5'-CTGATGTCCAGGAGGAGAAAGG-3', reverse primer 5'-AGCCCGACAATGTCAAGGACTG-3'. *RNA Polymerase II* promoter: product size 384 bp, forward primer 5'-AGATGAAACCGTTGTCCAAACT-3', reverse primer 5'-AGGTTACGGCAGTTTGTCTCTC-3'. PCR conditions for *RNA Polymerase II* promoter were as follows: 95°C for 5 minutes, 95°C for 30 seconds, then 40 cycles of 57.8°C for 22 seconds and 72°C for 24 seconds, and finally 10 minutes at 72°C. The PCR mixture for *RNA Polymerase II* contained lX buffer with 1.0 mM $MgCl_2$, 30 pmol of each primer, and 0.2 mM dNTPs. PCR conditions for 3'UTR of *DHFR* gene were as follows: 95°C for 5 minutes and 95°C for 30 seconds, then 40 cycles of 67°C for 22 seconds and 72°C for 24 seconds, and finally 10 minutes at 72°C. The PCR mixture for 3'UTR of *DHFR* gene contained lX buffer with 1.5 mM $MgCl_2$, 30 pmol of each primer, and 0.2 mM dNTPs.

Comparing observed vs. expected frequencies of alterations

Normalized Cy3/Cy5 intensity $\log_2$ ratio was first converted to a Z score for comparison across different samples. We first calculated the frequencies of alterations observed in the dataset using different cut-off values. Table 2.1 shows an example of analyzing alteration and calculating frequencies of alterations. We first looked at the frequency of total alterations based on two set parameters (Z at 1.96 and Z at 1.0). For example, for Z at 1.0 as cutoff, a Z score of greater than 1.0 or less than -1.0 would be considered as alterations. Replicate #1 at locus #3 from example Table 2.1 was less than -1.0; therefore, "No" (0) was assigned under the "Z>1.0 (replicate #1)" column, and "yes" (1) was then assigned to the "Z<-1.0 (replicate #1)" column. Numbers were assigned for convenience of tallying and for calculating frequencies of alteration in the next step (Table 2.2). Table 2.2 illustrates calculating observed frequencies for agreement of duplicates based on the tally from Table 2.1. Under the "tally across the same locus" column, "0" would mean no alterations detected in either replicate; "1" would mean 1 alteration detected in one of the replicates; "2" would mean 2 alterations detected in both replicates either in agreement (such as locus #3 and #5), or not in agreement (such as locus #4). The results taken from samples were analyzed using this method and are listed as Table 2.3, which includes the observed frequencies that would have been expected by random chance and the observed frequencies in the dataset.

Table 2.1 Example of analyzing Z score and calculating alteration frequencies.

	Z>1.0 (replicate #1)	**Z>1.0 (replicate #2)**	**Z<-1.0 (replicate #1)**	**Z<-1.0 (replicate #2)**	**Tally across the same locus**
Locus #1	No ("0")	No ("0")	No ("0")	No ("0")	0
Locus #2	No ("0")	Yes("1")	No ("0")	No ("0")	1
Locus #3	No ("0")	No ("0")	Yes("1")	Yes("1")	2
Locus #4	No ("0")	Yes("1")	Yes("1")	No ("0")	2
Locus #5	Yes("1")	Yes("1")	No ("0")	No ("0")	2
Locus #6	No ("0")	No ("0")	Yes("1")	No ("0")	1

Table 2.2 Example of calculating observed frequencies for ChIP-chip experiment duplicate agreement.

Duplicate Agreement	**Observed Frequencies for agreement (%)**
No alteration detected in either replicate	Locus #1 1/6 =16.67 %
Alteration detected in 1 replicate	Locus 2, 6 2/6 = 33.33 %
Alterations detected in replicates, in agreement	Locus 3, 5 2/ 6 = 33.33 %
Alterations detected in replicates, not in agreement	Locus 4 1 / 6 = 16.67 %

Moving average thresholding for integrative analysis

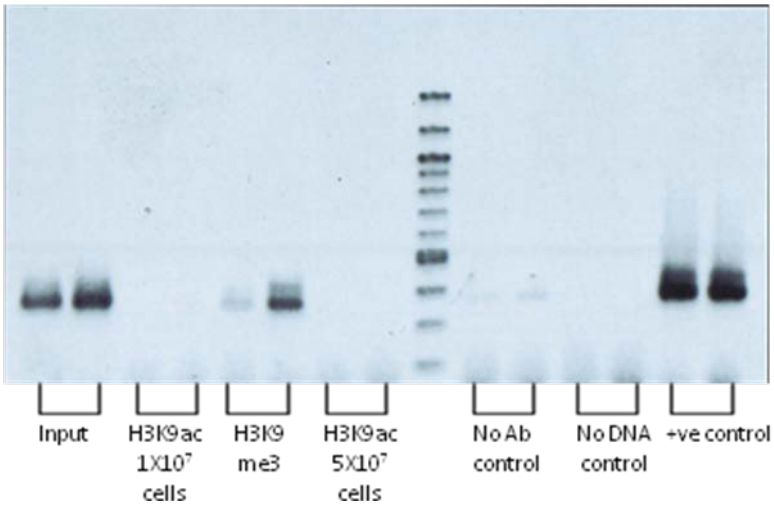
Thresholding of moving average data was applied across the replicates of each of the modifications (DNA methylation, H3K9ac, H3K9me1, H3K9me2, H3K9me3). This involves calculating the average across a sliding window of data points (600 kilobase (kb) windows sliding at 100 kb intervals). Averages were taken for each of the modification replicates. Figure 2.5 shows how these epigenetic profiles can be aligned on the same visualization platform.

2.3 RESULTS

2.3.1 Evaluation of ChIP pull down

Primers were used to check ChIP pull down. Male genomic DNA was used as positive control. Together, the "no antibody" and "no DNA" negative controls confirmed the control PCR experiment. As part of ChIP protocol optimization, there were two lanes of H3K9ac with pull down from different amounts of starting cells. Figure 2.2 shows two examples from the ChIP pull down. Since the 3' untranslated region (3'UTR) is transcriptionally repressed in mammalian systems, we would expect a band on the gel for the methylated silent marks (ie. H3K9me3). At the same time, no band would be observed on the active mark (H3K9ac). Conversely, *RNA polymerase II* gene is present for DNA replication in the genome, and thus a band would be observed with the active mark (H3K9ac), and no band observed for the silent marks (ie. H3K9me3). The experimental evaluation of the silent/active marks confirmed that ChIP pull down was successful.

a. 3'UTR of *DHFR* gene



b. *RNA Polymerase II* promoter

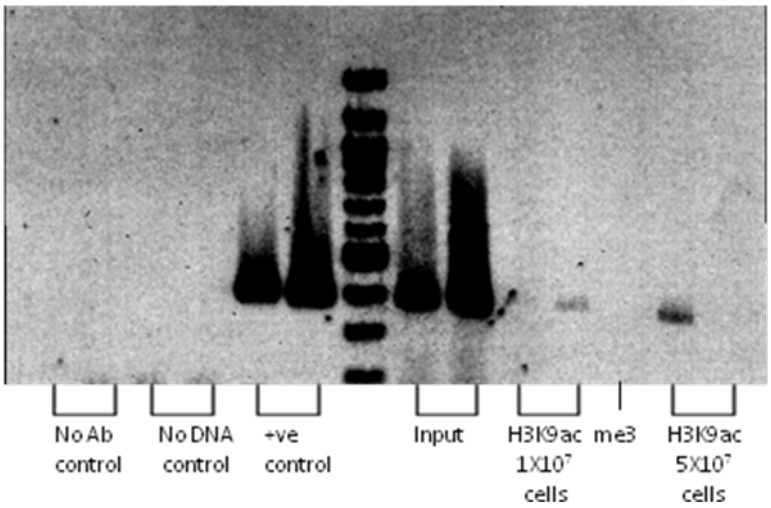


Figure 2.2 Validation of ChIP pull down.

a) ChIP products were tested using primers flanking the 3'UTR of the *DHFR* gene. $5X10^7$ cells were used for ChIP. $1X10^7$ cells for H3K9ac was part of the optimization protocol for starting cell amounts. Male reference DNA was used as positive control. **b)** ChIP products were tested using primers flanking *RNA Polymerase II* promoter.

2.3.2 Generation of DNA methylation and histone H3K9 maps

After successful ChIP pull down had been demonstrated, the IP and IN DNA were each labeled and co-hybridized to the whole genome tiling path array (see Section 2.2). Arrays were scanned and analyzed using the GenePix microarray scanner and the resulting images were analyzed using SoftWoRx Tracker Spot Analysis software. Signal intensity ratios were calculated to determine hyper/hypo DNA methylation, and enrichment/no-enrichment in H3K9ac, H3K9me1, me2, and me3 status. Data were visualized as $\log_2$ ratio plots (Figure 2.3) in custom SeeGH and SIGMA^2 software (Chari *et al.* 2008; Chi *et al.* 2004).

Figure 2.3 HCC15 H3K9ac global map.

The signal intensity ratios were normalized and imported in visualization software to generate the DNA methylation and the various histone H3K9 global mappings. Here, an example of HCC15 H3K9ac profile is shown.

2.3.3 Expected vs. observed frequency of alterations and reproducibility of experiments

The observed frequencies were compared to the frequencies that would be expected by random chance to ensure the datasets contain alterations worth further investigation. Using the same calculation as shown in Table 2.1 and 2.2, the expected frequencies and the observed frequencies were calculated for each of the modifications (Table 2.3). Figure 2.4 includes the observed frequencies in each of the modifications plotted against the expected frequencies if by random chance and if in a perfect replicate scenario. In a perfect replicate scenario, duplicates would always agree: either, "Alterations detected in replicates, in agreement" or, "No alterations detected in either replicate". The perfect frequency of "No alterations detected in either replicate" is determined by adding half of the expected if by random frequency from "Alteration detected in 1 replicate" and half of the frequency from "Alteration detected in replicates, not in agreement" to the original random frequency for "No alterations detected in either replicate" (Table 2.3). Same calculation applies to determining the perfect frequency for "Alterations detected in replicates, in agreement". The results show the observed frequencies were not due to random chance.

We further assessed the correlation coefficient (r) between each of the replicates, and r ranged from 0.71 to 0.87 for the five modifications studied (DNA methylation, H3K9ac, H3K9me1, me2, me3). This correlation is within the acceptable range given that correlation numbers between dissimilar datasets would be fairly low. For example, a profile from this project (chapter 2), and a profile from the cisplatin project (chapter 3), which are not similar,

gave a correlation number of -0.20. As a result, the replicate experiments were considered successful.

Table 2.3 **Frequency of the alterations expected if by random chance, if by perfect replicate scenario, and frequencies of observed alterations in H3K9me1 ChIP-chip experiment duplicates.**

Expected if by Random Chance	
Duplicate Agreement	**Observed Frequencies for agreement (%)**
No alterations detected in either replicate	46.61%
Alteration detected in 1 replicate	43.32%
Alterations detected in replicates, in agreement	5.03%
Alterations detected in replicates, not in agreement	5.03%
Expected if by Perfect Replicate Scenario	
No alterations detected in either replicate	70.79%
Alteration detected in 1 replicate	0%
Alterations detected in replicates, in agreement	29.21%
Alterations detected in replicates, not in agreement	0%
Observed in the H3K9me1 Profiles	
No alterations detected in either replicate	57.57%
Alteration detected in 1 replicate	31.47%
Alterations detected in replicates, in agreement	10.31%
Alterations detected in replicates, not in agreement	0.65%

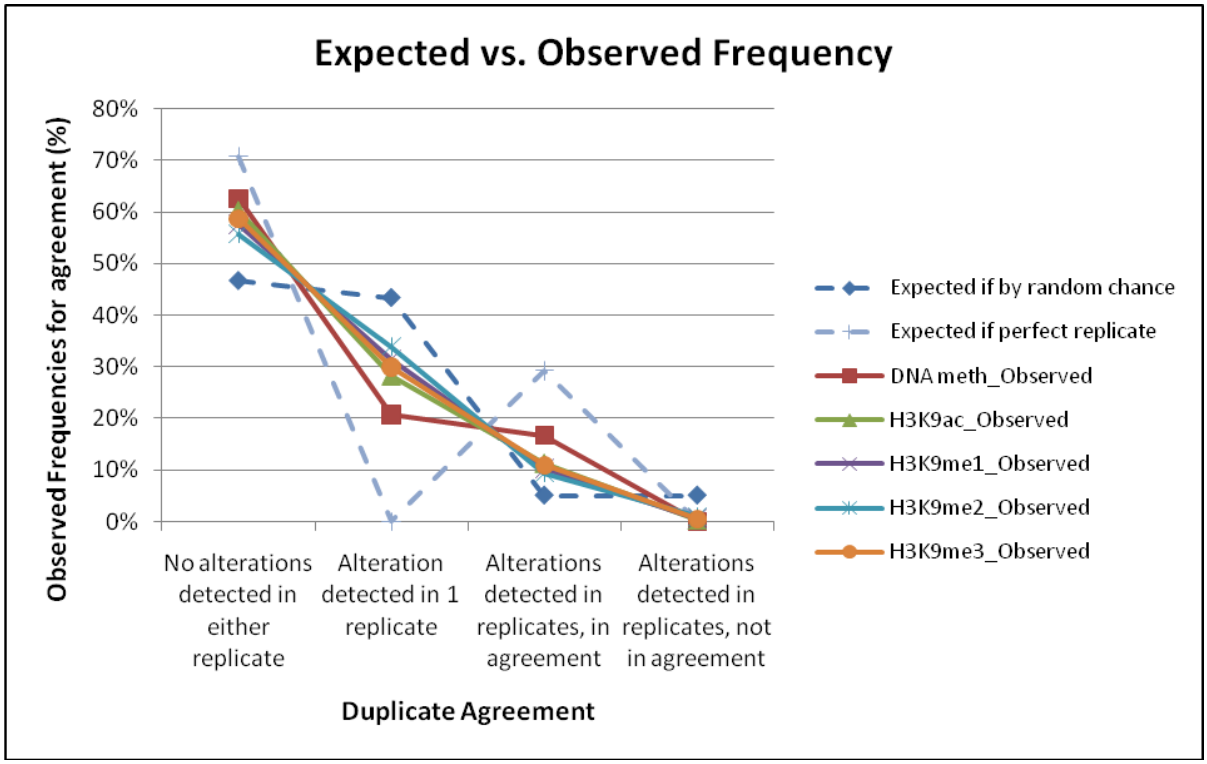


Figure 2.4 Expected vs. observed frequency of alterations for DNA methylation, H3K9me1, H3K9me2, and H3K9me3 modifications.

Frequencies were taken from all the profiles (observed, solid lines) and plotted against the expected frequency (dotted lines). All of the profiles did not follow the trend of random chance, showing that the frequencies of alterations observed from all of my data were not likely due to random chance. Further, a line of "expected if perfect replicate" was also plotted assuming that the perfect duplicate scenario would be either "No alterations detected in either replicate" or "Alterations detected in replicates, in agreement". DNA methylation in particular was the closest to "perfect". P value from binomial test is less than $1X10^{-6}$ for all data points.

2.3.4 Integrative analysis on profiles of DNA methylation and multiple states of histone modification

For integrative analysis, profiles of the various modifications were aligned after moving average thresholding. Peaks to the right of the midline denote enrichment of the modification as indicated; to the left of the midline refers to no-enrichment of the modification. Figure 2.5 shows that these profiles could be aligned and visualized on the same platform.

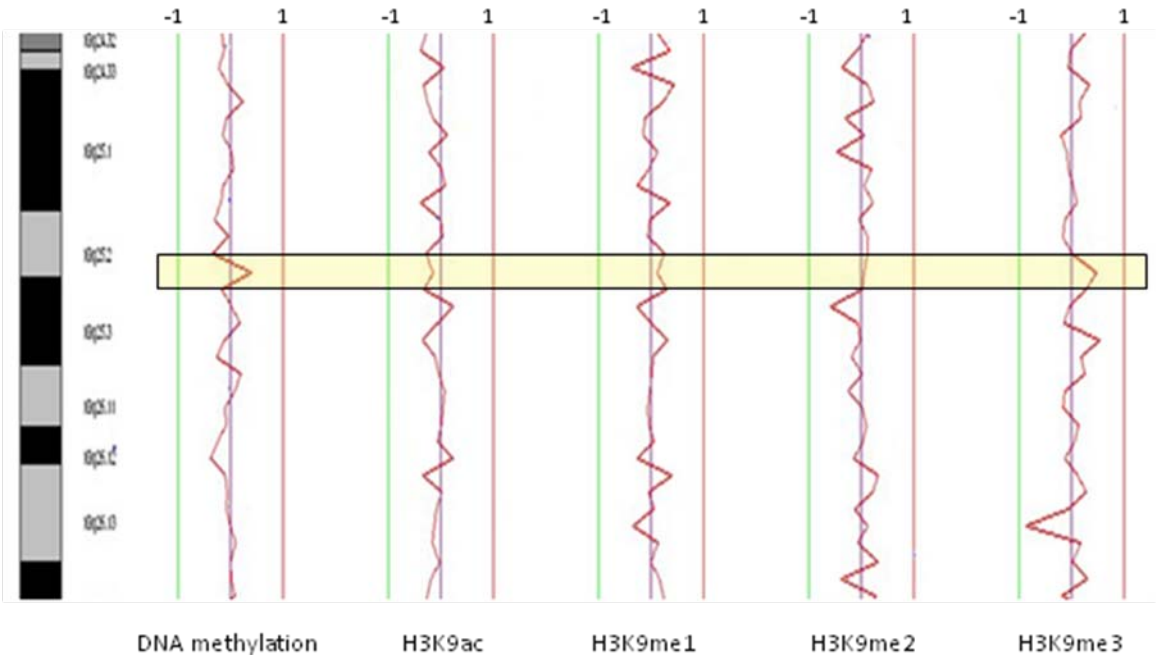


Figure 2.5 Integration of the epigenetic profiles of the HCC15 lung cancer cell line.

From left to right are DNA methylation, H3K9ac, H3K9me1, H3K9me2, and H3K9me3 profiles. These profiles are represented by calculated moving averages from $\log_2$ ratio of data points. Specifically, chromosomal region 10q25 is highlighted illustrating a region of enrichment of DNA methylation, lack of H3K9ac enrichment, and enrichment of H3K9 me1, me2, me3.

2.3.5 Combinatorial patterns of DNA methylation, histone H3K9ac, H3K9me1, me2, and me3

With the same cut off that was applied in calculating alterations (Table 2.1), aligned profiles at a given moving average window have a possibility of a profile showing a Z score greater than 1.0, less than -1.0, or between 1.0 and -1.0. Considering these three possible statuses for each of the five modifications studied (DNA methylation, histone H3K9ac, H3K9me1, me2, and me3), 243 possible combinatorial patterns could be generated. Table 2.4 is the complete table that lists the 243 combinatorial patterns and the number of data points observed for each of the patterns. Figure 2.6 is a visual representation of the observed frequencies for each of the patterns in the dataset. This visual plot reveals the non-random distribution of the 243 patterns with different frequencies observed for each of the patterns. The most represented pattern was the "No alterations detected in all five modifications", which comprised 42.12% of the entire dataset. One other observation was that not all 243 combinatorial patterns were present in the dataset. Only 82 of these 243 combinatorial patterns (34%) were observed, and this shows that these 82 patterns are probably more favored in the HCC15 genome. Table 2.5, in particular, shows in detail the spikes that have higher observed frequencies than expected frequencies in Figure 2.6. Table 2.6 shows patterns that were observed at lower frequencies compared to the expected frequencies.

Table 2.4 Complete table of the possible 243 combinatorial patterns [a].

Category	Pattern [b]	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected [c]	Difference [d]
No Modification Statuses Detected	00000	0	0	0	0	0	11720	42.12%	14.83%	27.29%
1 Modification Status Detected	10000	1	0	0	0	0	1782	6.40%	3.45%	2.96%
	01000	0	1	0	0	0	406	1.46%	3.45%	-1.99%
	00100	0	0	1	0	0	447	1.61%	3.45%	-1.84%
	00010	0	0	0	1	0	424	1.52%	3.45%	-1.92%
	00001	0	0	0	0	1	547	1.97%	3.45%	-1.48%
	0000-1	0	0	0	0	-1	192	0.69%	3.45%	-2.76%
	000-10	0	0	0	-1	0	316	1.14%	3.45%	-2.31%
	00-100	0	0	-1	0	0	228	0.82%	3.45%	-2.63%
	0-1000	0	-1	0	0	0	199	0.72%	3.45%	-2.73%
	-10000	-1	0	0	0	0	1835	6.60%	3.45%	3.15%
2 Modification Statuses Detected	11000	1	1	0	0	0	110	0.40%	0.80%	-0.41%
	10100	1	0	1	0	0	113	0.41%	0.80%	-0.39%
	10010	1	0	0	1	0	57	0.20%	0.80%	-0.60%
	10001	1	0	0	0	1	72	0.26%	0.80%	-0.54%
	1000-1	1	0	0	0	-1	21	0.08%	0.80%	-0.73%
	100-10	1	0	0	-1	0	58	0.21%	0.80%	-0.59%
	10-100	1	0	-1	0	0	8	0.03%	0.80%	-0.77%
	1-1000	1	-1	0	0	0	32	0.12%	0.80%	-0.69%
	01100	0	1	1	0	0	163	0.59%	0.80%	-0.22%
	01010	0	1	0	1	0	206	0.74%	0.80%	-0.06%
	0100-1	0	1	0	0	-1	0	0.00%	0.80%	-0.80%
	010-10	0	1	0	-1	0	0	0.00%	0.80%	-0.80%
	01-100	0	1	-1	0	0	0	0.00%	0.80%	-0.80%
	00110	0	0	1	1	0	252	0.91%	0.80%	0.10%
	00101	0	0	1	0	1	149	0.54%	0.80%	-0.27%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
2 Modification Statuses Detected	0010-1	0	0	1	0	-1	0	0.00%	0.80%	-0.80%
	001-10	0	0	1	-1	0	0	0.00%	0.80%	-0.80%
	00011	0	0	0	1	1	183	0.66%	0.80%	-0.14%
	0001-1	0	0	0	1	-1	0	0.00%	0.80%	-0.80%
	000-11	0	0	0	-1	1	0	0.00%	0.80%	-0.80%
	000-1-1	0	0	0	-1	-1	133	0.48%	0.80%	-0.32%
	00-110	0	0	-1	1	0	0	0.00%	0.80%	-0.80%
	00-101	0	0	-1	0	1	0	0.00%	0.80%	-0.80%
	00-10-1	0	0	-1	0	-1	69	0.25%	0.80%	-0.55%
	00-1-10	0	0	-1	-1	0	138	0.50%	0.80%	-0.30%
	0-1100	0	-1	1	0	0	0	0.00%	0.80%	-0.80%
	0-1010	0	-1	0	1	0	0	0.00%	0.80%	-0.80%
	0-1001	0	-1	0	0	1	0	0.00%	0.80%	-0.80%
	0-100-1	0	-1	0	0	-1	112	0.40%	0.80%	-0.40%
	0-10-10	0	-1	0	-1	0	75	0.27%	0.80%	-0.53%
	0-1-100	0	-1	-1	0	0	90	0.32%	0.80%	-0.48%
	-11000	-1	1	0	0	0	7	0.03%	0.80%	-0.78%
	-10100	-1	0	1	0	0	1	0.00%	0.80%	-0.80%
	-10010	-1	0	0	1	0	0	0.00%	0.80%	-0.80%
	-10001	-1	0	0	0	1	7	0.03%	0.80%	-0.78%
	-1000-1	-1	0	0	0	-1	83	0.30%	0.80%	-0.50%
	-100-10	-1	0	0	-1	0	66	0.24%	0.80%	-0.56%
	-10-100	-1	0	-1	0	0	169	0.61%	0.80%	-0.19%
	-1-1000	-1	-1	0	0	0	96	0.35%	0.80%	-0.46%
3 Modification Statuses Detected	11100	1	1	1	0	0	57	0.20%	0.19%	0.02%
	11010	1	1	0	1	0	30	0.11%	0.19%	-0.08%
	11001	1	1	0	0	1	40	0.14%	0.19%	-0.04%
	1100-1	1	1	0	0	-1	0	0.00%	0.19%	-0.19%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
3	110-10	1	1	0	-1	0	0	0.00%	0.19%	-0.19%
Modification	11-100	1	1	-1	0	0	0	0.00%	0.19%	-0.19%
Statuses Detected	10110	1	0	1	1	0	57	0.20%	0.19%	0.02%
Cont'd	10101	1	0	1	0	1	26	0.09%	0.19%	-0.09%
	1010-1	1	0	1	0	-1	0	0.00%	0.19%	-0.19%
	101-10	1	0	1	-1	0	0	0.00%	0.19%	-0.19%
	10011	1	0	0	1	1	50	0.18%	0.19%	-0.01%
	1001-1	1	0	0	1	-1	1	0.00%	0.19%	-0.18%
	100-11	1	0	0	-1	1	0	0.00%	0.19%	-0.19%
	100-1-1	1	0	0	-1	-1	23	0.08%	0.19%	-0.10%
	10-110	1	0	-1	1	0	0	0.00%	0.19%	-0.19%
	10-101	1	0	-1	0	1	0	0.00%	0.19%	-0.19%
	10-10-1	1	0	-1	0	-1	0	0.00%	0.19%	-0.19%
	10-1-10	1	0	-1	-1	0	9	0.03%	0.19%	-0.15%
	1-1100	1	-1	1	0	0	0	0.00%	0.19%	-0.19%
	1-1010	1	-1	0	1	0	0	0.00%	0.19%	-0.19%
	1-1001	1	-1	0	0	1	0	0.00%	0.19%	-0.19%
	1-100-1	1	-1	0	0	-1	21	0.08%	0.19%	-0.11%
	1-10-10	1	-1	0	-1	0	5	0.02%	0.19%	-0.17%
	1-1-100	1	-1	-1	0	0	6	0.02%	0.19%	-0.16%
	01110	0	1	1	1	0	201	0.72%	0.19%	0.54%
	01101	0	1	1	0	1	226	0.81%	0.19%	0.63%
	0110-1	0	1	1	0	-1	0	0.00%	0.19%	-0.19%
	011-10	0	1	1	-1	0	0	0.00%	0.19%	-0.19%
	01011	0	1	0	1	1	306	1.10%	0.19%	0.91%
	0101-1	0	1	0	1	-1	0	0.00%	0.19%	-0.19%
	010-11	0	1	0	-1	1	0	0.00%	0.19%	-0.19%
	010-1-1	0	1	0	-1	-1	0	0.00%	0.19%	-0.19%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
3	01-110	0	1	-1	1	0	0	0.00%	0.19%	-0.19%
Modification	01-101	0	1	-1	0	1	0	0.00%	0.19%	-0.19%
Statuses Detected	01-10-1	0	1	-1	0	-1	0	0.00%	0.19%	-0.19%
Cont'd	01-1-10	0	1	-1	-1	0	0	0.00%	0.19%	-0.19%
	00111	0	0	1	1	1	214	0.77%	0.19%	0.58%
	0011-1	0	0	1	1	-1	0	0.00%	0.19%	-0.19%
	001-11	0	0	1	-1	1	0	0.00%	0.19%	-0.19%
	001-1-1	0	0	1	-1	-1	0	0.00%	0.19%	-0.19%
	00-111	0	0	-1	1	1	0	0.00%	0.19%	-0.19%
	00-11-1	0	0	-1	1	-1	0	0.00%	0.19%	-0.19%
	00-1-11	0	0	-1	-1	1	0	0.00%	0.19%	-0.19%
	00-1-1-1	0	0	-1	-1	-1	126	0.45%	0.19%	0.27%
	0-1110	0	-1	1	1	0	0	0.00%	0.19%	-0.19%
	0-1101	0	-1	1	0	1	0	0.00%	0.19%	-0.19%
	0-110-1	0	-1	1	0	-1	0	0.00%	0.19%	-0.19%
	0-11-10	0	-1	1	-1	0	0	0.00%	0.19%	-0.19%
	0-1011	0	-1	0	1	1	0	0.00%	0.19%	-0.19%
	0-101-1	0	-1	0	1	-1	0	0.00%	0.19%	-0.19%
	0-10-11	0	-1	0	-1	1	0	0.00%	0.19%	-0.19%
	0-10-1-1	0	-1	0	-1	-1	211	0.76%	0.19%	0.57%
	0-1-110	0	-1	-1	1	0	0	0.00%	0.19%	-0.19%
	0-1-101	0	-1	-1	0	1	0	0.00%	0.19%	-0.19%
	0-1-10-1	0	-1	-1	0	-1	238	0.86%	0.19%	0.67%
	0-1-1-10	0	-1	-1	-1	0	104	0.37%	0.19%	0.19%
	-11100	-1	1	1	0	0	0	0.00%	0.19%	-0.19%
	-11010	-1	1	0	1	0	3	0.01%	0.19%	-0.18%
	-11001	-1	1	0	0	1	1	0.00%	0.19%	-0.18%
	-1100-1	-1	1	0	0	-1	0	0.00%	0.19%	-0.19%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
3 Modification Statuses Detected Cont'd	-110-10	-1	1	0	-1	0	0	0.00%	0.19%	-0.19%
	-11-100	-1	1	-1	0	0	0	0.00%	0.19%	-0.19%
	-10110	-1	0	1	1	0	0	0.00%	0.19%	-0.19%
	-10101	-1	0	1	0	1	0	0.00%	0.19%	-0.19%
	-1010-1	-1	0	1	0	-1	0	0.00%	0.19%	-0.19%
	-101-10	-1	0	1	-1	0	0	0.00%	0.19%	-0.19%
	-10011	-1	0	0	1	1	0	0.00%	0.19%	-0.19%
	-1001-1	-1	0	0	1	-1	0	0.00%	0.19%	-0.19%
	-100-11	-1	0	0	-1	1	0	0.00%	0.19%	-0.19%
	-100-1-1	-1	0	0	-1	-1	22	0.08%	0.19%	-0.11%
	-10-110	-1	0	-1	1	0	0	0.00%	0.19%	-0.19%
	-10-101	-1	0	-1	0	1	0	0.00%	0.19%	-0.19%
	-10-10-1	-1	0	-1	0	-1	47	0.17%	0.19%	-0.02%
	-10-1-10	-1	0	-1	-1	0	42	0.15%	0.19%	-0.04%
	-1-1100	-1	-1	1	0	0	0	0.00%	0.19%	-0.19%
	-1-1010	-1	-1	0	1	0	0	0.00%	0.19%	-0.19%
	-1-1001	-1	-1	0	0	1	0	0.00%	0.19%	-0.19%
	-1-100-1	-1	-1	0	0	-1	59	0.21%	0.19%	0.03%
	-1-10-10	-1	-1	0	-1	0	23	0.08%	0.19%	-0.10%
	-1-1-100	-1	-1	-1	0	0	82	0.29%	0.19%	0.11%
4 Modification Statuses Detected	11110	1	1	1	1	0	120	0.43%	0.04%	0.39%
	11101	1	1	1	0	1	56	0.20%	0.04%	0.16%
	1110-1	1	1	1	0	-1	0	0.00%	0.04%	-0.04%
	111-10	1	1	1	-1	0	0	0.00%	0.04%	-0.04%
	11011	1	1	0	1	1	59	0.21%	0.04%	0.17%
	1101-1	1	1	0	1	-1	0	0.00%	0.04%	-0.04%
	110-11	1	1	0	-1	1	0	0.00%	0.04%	-0.04%
	110-1-1	1	1	0	-1	-1	0	0.00%	0.04%	-0.04%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
4	11-110	1	1	-1	1	0	0	0.00%	0.04%	-0.04%
Modification	11-101	1	1	-1	0	1	0	0.00%	0.04%	-0.04%
Statuses Detected	11-10-1	1	1	-1	0	-1	0	0.00%	0.04%	-0.04%
Cont'd	11-1-10	1	1	-1	-1	0	0	0.00%	0.04%	-0.04%
	10111	1	0	1	1	1	102	0.37%	0.04%	0.32%
	1011-1	1	0	1	1	-1	0	0.00%	0.04%	-0.04%
	101-11	1	0	1	-1	1	0	0.00%	0.04%	-0.04%
	101-1-1	1	0	1	-1	-1	0	0.00%	0.04%	-0.04%
	10-111	1	0	-1	1	1	0	0.00%	0.04%	-0.04%
	10-11-1	1	0	-1	1	-1	0	0.00%	0.04%	-0.04%
	10-1-11	1	0	-1	-1	1	0	0.00%	0.04%	-0.04%
	10-1-1-1	1	0	-1	-1	-1	7	0.03%	0.04%	-0.02%
	1-1110	1	-1	1	1	0	0	0.00%	0.04%	-0.04%
	1-1101	1	-1	1	0	1	0	0.00%	0.04%	-0.04%
	1-110-1	1	-1	1	0	-1	0	0.00%	0.04%	-0.04%
	1-11-10	1	-1	1	-1	0	0	0.00%	0.04%	-0.04%
	1-1011	1	-1	0	1	1	0	0.00%	0.04%	-0.04%
	1-101-1	1	-1	0	1	-1	0	0.00%	0.04%	-0.04%
	1-10-11	1	-1	0	-1	1	0	0.00%	0.04%	-0.04%
	1-10-1-1	1	-1	0	-1	-1	18	0.06%	0.04%	0.02%
	1-1-110	1	-1	-1	1	0	0	0.00%	0.04%	-0.04%
	1-1-101	1	-1	-1	0	1	0	0.00%	0.04%	-0.04%
	1-1-10-1	1	-1	-1	0	-1	20	0.07%	0.04%	0.03%
	1-1-1-10	1	-1	-1	-1	0	5	0.02%	0.04%	-0.03%
	01111	0	1	1	1	1	900	3.23%	0.04%	3.19%
	0111-1	0	1	1	1	-1	0	0.00%	0.04%	-0.04%
	011-11	0	1	1	-1	1	0	0.00%	0.04%	-0.04%
	011-1-1	0	1	1	-1	-1	0	0.00%	0.04%	-0.04%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
4	01-111	0	1	-1	1	1	0	0.00%	0.04%	-0.04%
Modification	01-11-1	0	1	-1	1	-1	0	0.00%	0.04%	-0.04%
Statuses Detected	01-1-11	0	1	-1	-1	1	0	0.00%	0.04%	-0.04%
Cont'd	01-1-1-1	0	1	-1	-1	-1	0	0.00%	0.04%	-0.04%
	0-1111	0	-1	1	1	1	0	0.00%	0.04%	-0.04%
	0-111-1	0	-1	1	1	-1	0	0.00%	0.04%	-0.04%
	0-11-11	0	-1	1	-1	1	0	0.00%	0.04%	-0.04%
	0-11-1-1	0	-1	1	-1	-1	0	0.00%	0.04%	-0.04%
	0-1-111	0	-1	-1	1	1	0	0.00%	0.04%	-0.04%
	0-1-11-1	0	-1	-1	1	-1	0	0.00%	0.04%	-0.04%
	0-1-1-11	0	-1	-1	-1	1	0	0.00%	0.04%	-0.04%
	0-1-1-1-1	0	-1	-1	-1	-1	1369	4.92%	0.04%	4.88%
	-11110	-1	1	1	1	0	5	0.02%	0.04%	-0.03%
	-11101	-1	1	1	0	1	2	0.01%	0.04%	-0.04%
	-1110-1	-1	1	1	0	-1	0	0.00%	0.04%	-0.04%
	-111-10	-1	1	1	-1	0	0	0.00%	0.04%	-0.04%
	-11011	-1	1	0	1	1	1	0.00%	0.04%	-0.04%
	-1101-1	-1	1	0	1	-1	0	0.00%	0.04%	-0.04%
	-110-11	-1	1	0	-1	1	0	0.00%	0.04%	-0.04%
	-110-1-1	-1	1	0	-1	-1	0	0.00%	0.04%	-0.04%
	-11-110	-1	1	-1	1	0	0	0.00%	0.04%	-0.04%
	-11-101	-1	1	-1	0	1	0	0.00%	0.04%	-0.04%
	-11-10-1	-1	1	-1	0	-1	0	0.00%	0.04%	-0.04%
	-11-1-10	-1	1	-1	-1	0	0	0.00%	0.04%	-0.04%
	-10111	-1	0	1	1	1	0	0.00%	0.04%	-0.04%
	-1011-1	-1	0	1	1	-1	0	0.00%	0.04%	-0.04%
	-101-11	-1	0	1	-1	1	0	0.00%	0.04%	-0.04%
	-101-1-1	-1	0	1	-1	-1	0	0.00%	0.04%	-0.04%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
4 Modification Statuses Detected Cont'd	-10-111	-1	0	-1	1	1	0	0.00%	0.04%	-0.04%
	-10-11-1	-1	0	-1	1	-1	0	0.00%	0.04%	-0.04%
	-10-1-11	-1	0	-1	-1	1	0	0.00%	0.04%	-0.04%
	-10-1-1-1	-1	0	-1	-1	-1	75	0.27%	0.04%	0.23%
	-1-1110	-1	-1	1	1	0	0	0.00%	0.04%	-0.04%
	-1-1101	-1	-1	1	0	1	0	0.00%	0.04%	-0.04%
	-1-110-1	-1	-1	1	0	-1	0	0.00%	0.04%	-0.04%
	-1-11-10	-1	-1	1	-1	0	0	0.00%	0.04%	-0.04%
	-1-1011	-1	-1	0	1	1	0	0.00%	0.04%	-0.04%
	-1-101-1	-1	-1	0	1	-1	0	0.00%	0.04%	-0.04%
	-1-10-11	-1	-1	0	-1	1	0	0.00%	0.04%	-0.04%
	-1-10-1-1	-1	-1	0	-1	-1	74	0.27%	0.04%	0.22%
	-1-1-110	-1	-1	-1	1	0	0	0.00%	0.04%	-0.04%
	-1-1-101	-1	-1	-1	0	1	0	0.00%	0.04%	-0.04%
	-1-1-10-1	-1	-1	-1	0	-1	159	0.57%	0.04%	0.53%
	-1-1-1-10	-1	-1	-1	-1	0	80	0.29%	0.04%	0.24%
5 Modification Statuses Detected	11111	1	1	1	1	1	614	2.21%	0.01%	2.20%
	1111-1	1	1	1	1	-1	0	0.00%	0.01%	-0.01%
	111-11	1	1	1	-1	1	0	0.00%	0.01%	-0.01%
	111-1-1	1	1	1	-1	-1	0	0.00%	0.01%	-0.01%
	11-111	1	1	-1	1	1	0	0.00%	0.01%	-0.01%
	11-11-1	1	1	-1	1	-1	0	0.00%	0.01%	-0.01%
	11-1-11	1	1	-1	-1	1	0	0.00%	0.01%	-0.01%
	11-1-1-1	1	1	-1	-1	-1	0	0.00%	0.01%	-0.01%
	1-1111	1	-1	1	1	1	0	0.00%	0.01%	-0.01%
	1-111-1	1	-1	1	1	-1	0	0.00%	0.01%	-0.01%
	1-11-11	1	-1	1	-1	1	0	0.00%	0.01%	-0.01%
	1-11-1-1	1	-1	1	-1	-1	0	0.00%	0.01%	-0.01%

Category	Pattern	DNA meth	H3K9ac	H3K9me1	H3K9me2	H3K9me3	Counts	Observed	Expected	Difference
5	1-1-111	1	-1	-1	1	1	0	0.00%	0.01%	-0.01%
Modification	1-1-11-1	1	-1	-1	1	-1	0	0.00%	0.01%	-0.01%
Statuses Detected	1-1-1-11	1	-1	-1	-1	1	0	0.00%	0.01%	-0.01%
Cont'd	1-1-1-1-1	1	-1	-1	-1	-1	158	0.57%	0.01%	0.56%
	-11111	-1	1	1	1	1	0	0.00%	0.01%	-0.01%
	-1111-1	-1	1	1	1	-1	0	0.00%	0.01%	-0.01%
	-111-11	-1	1	1	-1	1	0	0.00%	0.01%	-0.01%
	-111-1-1	-1	1	1	-1	-1	0	0.00%	0.01%	-0.01%
	-11-111	-1	1	-1	1	1	0	0.00%	0.01%	-0.01%
	-11-11-1	-1	1	-1	1	-1	0	0.00%	0.01%	-0.01%
	-11-1-11	-1	1	-1	-1	1	0	0.00%	0.01%	-0.01%
	-11-1-1-1	-1	1	-1	-1	-1	0	0.00%	0.01%	-0.01%
	-1-1111	-1	-1	1	1	1	0	0.00%	0.01%	-0.01%
	-1-111-1	-1	-1	1	1	-1	0	0.00%	0.01%	-0.01%
	-1-11-11	-1	-1	1	-1	1	0	0.00%	0.01%	-0.01%
	-1-11-1-1	-1	-1	1	-1	-1	0	0.00%	0.01%	-0.01%
	-1-1-111	-1	-1	-1	1	1	0	0.00%	0.01%	-0.01%
	-1-1-11-1	-1	-1	-1	1	-1	0	0.00%	0.01%	-0.01%
	-1-1-1-11	-1	-1	-1	-1	1	0	0.00%	0.01%	-0.01%
	-1-1-1-1-1	-1	-1	-1	-1	-1	989	3.55%	0.01%	3.54%

[a.] This table shows counts of data points corresponding with each of the 243 possible combinatorial patterns, observed percentage of the pattern, calculated expected percentage of each pattern, and how each pattern was categorized.

[b.] To interpret the pattern, the order of these five digit numbers denotes detection status of DNA methylation, H3K9ac, H3K9me1, H3K9me2, H3K9me3 as the breakdown shown in column 3 to column 7. For example, "10001" pattern from the 2 modification status category refers to enrichment of DNA methylation, no alteration detected for H3K9ac, H3K9me1,me2, enrichment of H3K9me3).

The yellow highlighted portions are the "canonical patterns" explained in Section 2.3.7; light grey highlighted portions are the "non canonical patterns" explained in Section 2.3.7.

[c.] The expected frequencies were calculated by multiplying the probabilities: 0.68269 for "0"; 0.158655 for "1"; 0.158655 for "-1".

[d.] Difference was calculated from "Observed" minus "Expected" (see Section 2.3.5 for details)

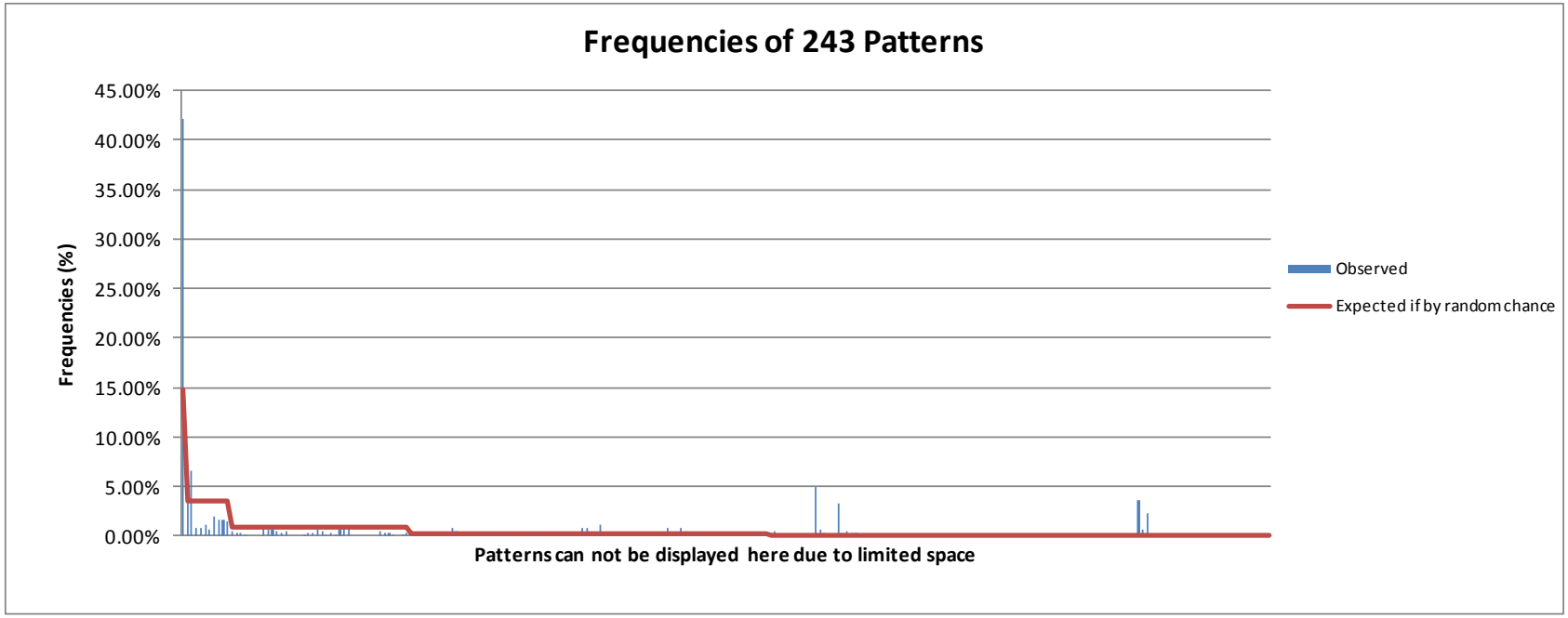


Figure 2.6 Non-random distributions of 243 combinatorial patterns.

Certain patterns were observed more than expected if by random chance. Due to space limitations, this figure only serves as a visual representation of the non random distribution of 243 combinatorial patterns. Please refer back to Table 2.4 for details.

Table 2.5 Selective patterns that showed more than 1.0% positive difference between the observed and the expected frequencies calculated from Table 2.4[a].

Category	Pattern	Observed	Expected	Difference (Observed minus Expected)[b]
No Modification Statuses Detected	00000	42.12%	14.83%	27.29%
1 Modification	-10000	6.60%	3.45%	3.15%
Status Detected	10000	6.40%	3.45%	2.96%
4 Modification	01111	3.23%	0.04%	3.19%
Statuses Detected	0-1-1-1-1	4.92%	0.04%	4.88%
5 Modification	11111	2.21%	0.01%	2.20%
Statuses Detected	-1-1-1-1-1	3.55%	0.01%	3.54%

[a.] The order of these five digit numbers denotes detection status of DNA methylation, H3K9ac, H3K9me1, H3K9me2, and H3K9me3. Refer to Table 2.4 for more details on the numerical representation.

[b] P value from binomial test is less than $1X10^{-6}$ for all data points.

Table 2.6 Selective patterns that showed more than 1.0% negative difference between the observed and the expected frequencies calculated from Table 2.4 [a].

Category	Pattern	Observed	Expected	Difference (Observed minus Expected)[b]
1 Modification Status Detected	0000-1	0.69%	3.45%	-2.76%
	0-1000	0.72%	3.45%	-2.73%
	00-100	0.82%	3.45%	-2.63%
	000-10	1.14%	3.45%	-2.31%
	01000	1.46%	3.45%	-1.99%
	00010	1.52%	3.45%	-1.92%
	00100	1.61%	3.45%	-1.84%
	00001	1.97%	3.45%	-1.48%

[a.] The order of these five digit numbers denotes detection status of DNA methylation, H3K9ac, H3K9me1, H3K9me2, and H3K9me3. Refer to Table 2.4 for more details on the numerical representation.

[b] P value from binomial test is less than $1X10^{-6}$ for all data points.

2.3.6 Categorizing combinatorial patterns

In order to study these 243 possible combinatorial patterns in a more convenient manner, the patterns were broken down into six categories (Table 2.4): no modification statuses detected (total of 1 possible pattern fitting in this category), one modification status detected (total of 10 possible patterns), two modification statuses detected (total of 40 possible patterns), three modification statuses detected (total of 80 possible patterns), four modification statuses detected (80 possible patterns), all five modification statuses detected (total of 32 possible patterns). Figure 2.7 shows the observed vs. expected distribution according to these six categories.

Since most of the categories involved more than 1 single pattern, we also decided to investigate further which pattern was best-represented in each of the categories mentioned above (Figure 2.8). Some prominent patterns that we observed in each of the categories include no enrichment of DNA methylation (pattern -10000), and enrichment of DNA methylation (pattern 10000) each takes up 29% and 28%, respectively, of all the data in the "one modification status detected" category.

The other prominent ones were the combinatorial pattern of no detection of DNA methylation status, and no enrichment of all the H3K9 histone marks (45%) from the "four modification statuses detected" category; as well as no enrichment of all the marks (56%) in the "five modification statuses detected" category.

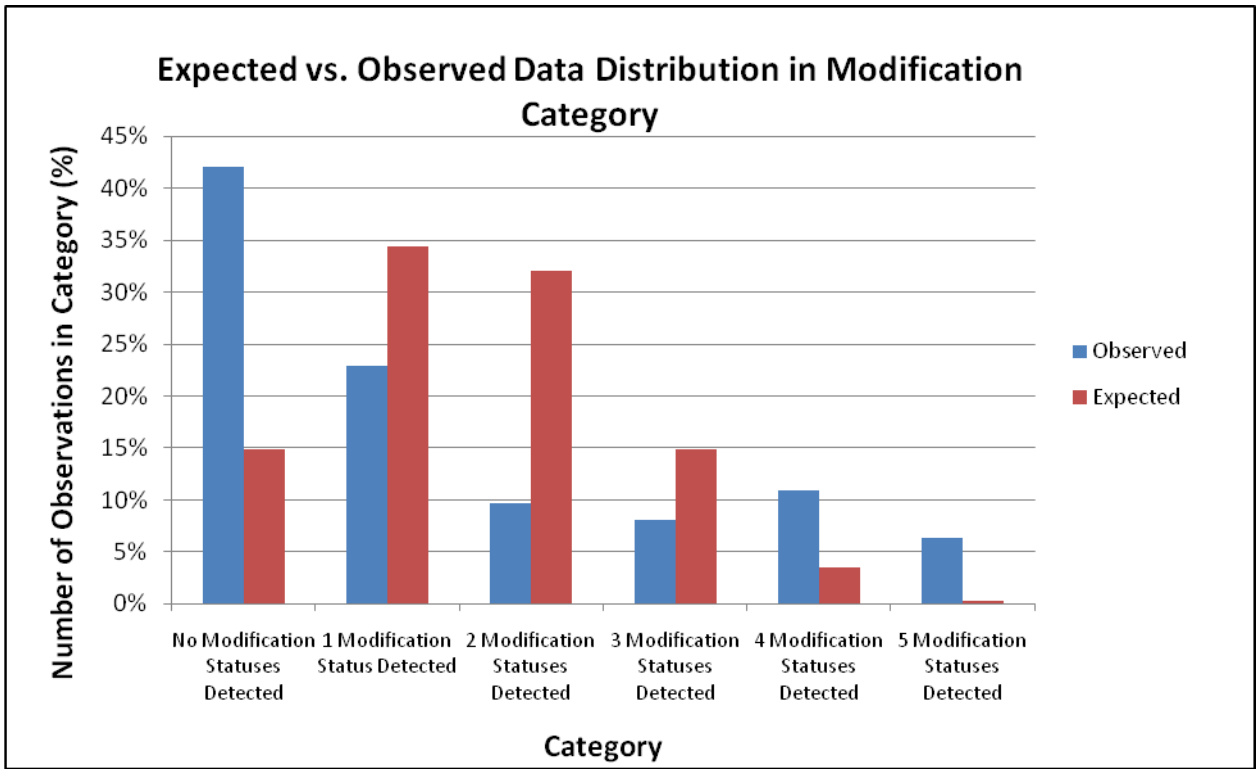
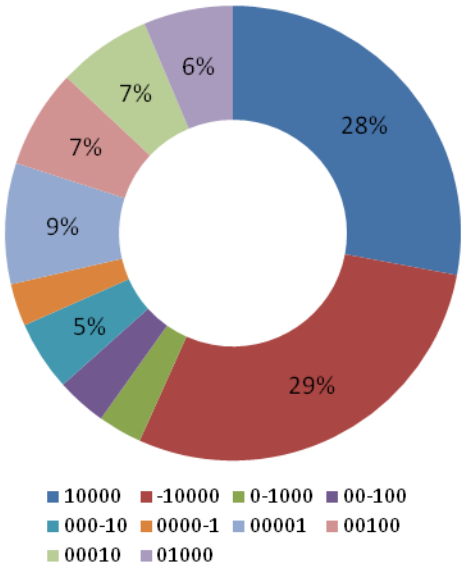


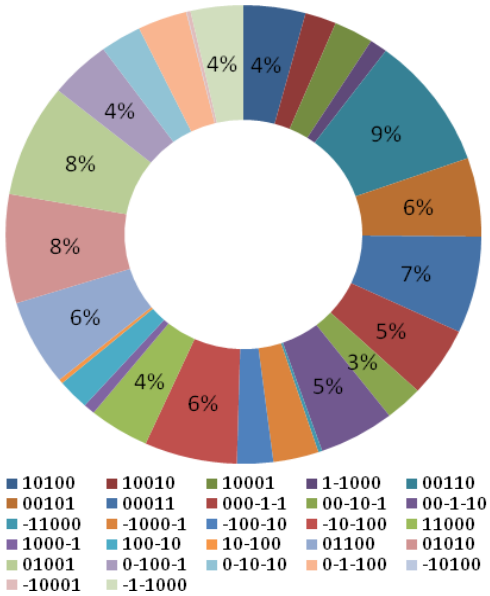
Figure 2.7 Data distribution according to number of modifications status detected.

Taking data from Table 2.4, the frequencies were added up for each category. The number of observations in each category for observed and expected if by random chance were shown as percentages and plotted as bar graph. "No modification statuses detected" took up 42% of the entire dataset, followed by any 1 modification of the five modification statuses detected (23%), any four modification statuses detected (11%), any two modification (10%), any 3 modification (8%), and all five modification statuses detected (6%). We can conclude that the distribution of the dataset by modification category classification was not likely due to random chance.

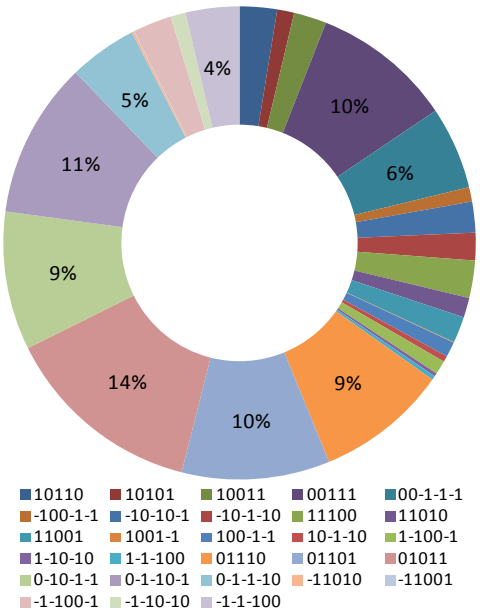
a. Any one modification status detected



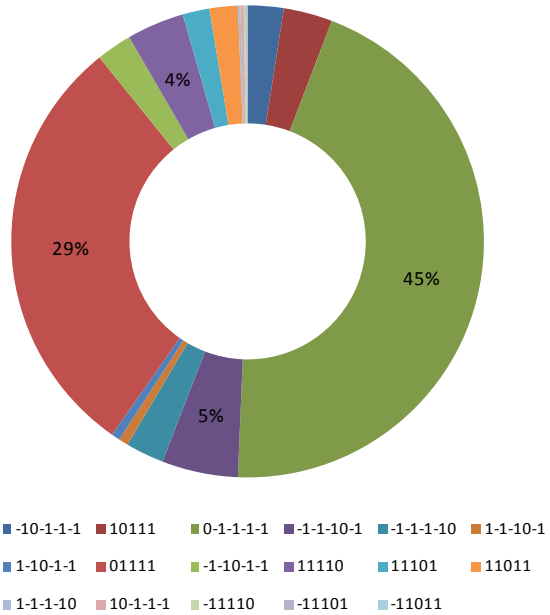
b. Any two modification statuses detected



c. Any three modification statuses detected



d. Any four modification statuses detected



e. Five modification statuses detected

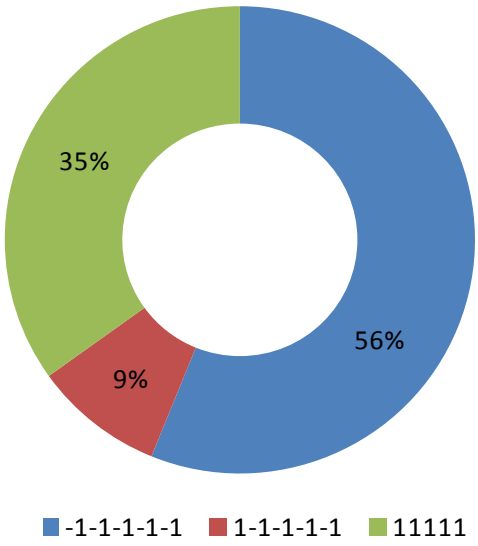


Figure 2.8 Patterns of histone modification distributions categorized by modification status detected [2].

Numbers are shown as percentages, and patterns that have 0% in distributions are excluded in these diagrams. Only the ones that took up four percent of the whole data or above were indicated with numbers on the color coded regions. (a) Percentage of pattern distribution for any one modification status detected. All of the 10 possible patterns were observed, (10/10 observed). (b) Any two modifications status detected, 27/40 observed. (c) Any three modifications status detected, 28/80 observed. (d) Any four modifications status detected, 17/80 observed. (e) All five modifications status detected, 3/32 observed.

[2] The order of these five digit numbers denotes detection status of DNA methylation, H3K9ac, H3K9me1, H3K9me2, and H3K9me3. Refer to Table 2.4 for more details on the numerical representation.

2.3.7 Canonical and non-canonical patterns

Since we are studying alterations at large scale, each region could harbor a mixture of marks as they could act on different parts of a single BAC clone. The prominent patterns, such as no enrichment of all the marks in the "all five modification status category", were in fact the patterns that were not typically expected (referred to as "non-canonical patterns" here). Those typically expected phenomena are referred to as "canonical patterns" here. We decided to further examine each of the categories by the distribution of canonical patterns and non-canonical patterns (Figure 2.9).

a. Canonical Patterns **b. Non-Canonical Patterns**

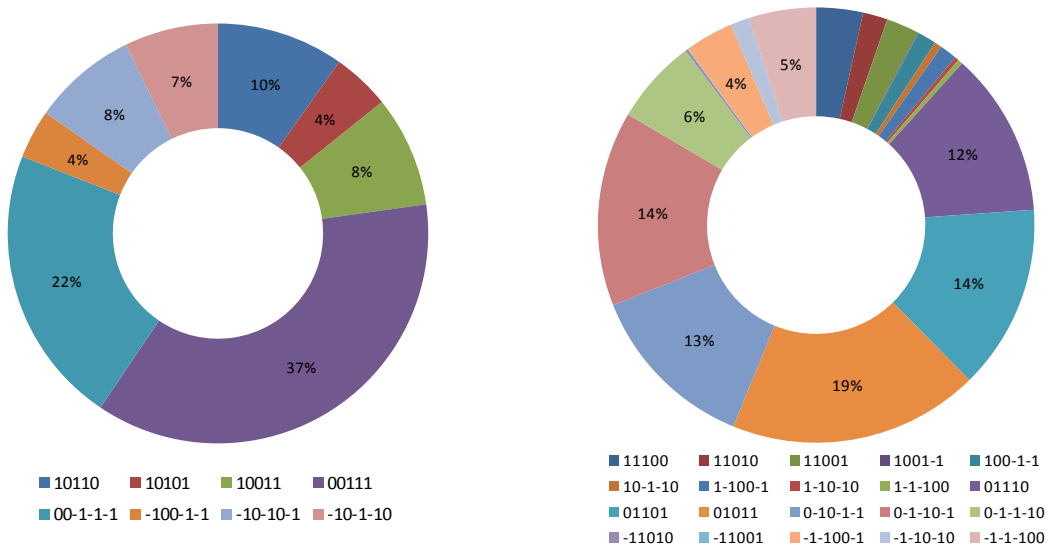


Figure 2.9 Distribution of canonical and non-canonical patterns of the combinatorial patterns from the "three modification statuses detected" category [3].

The patterns shown here were observed at least once in the "three modification statuses detected" category. Only the ones that took up four percent of the whole data or above were indicated with numbers on the colour coded regions. Specifically, some prominent patterns were also observed within each sub-category. This could help to further explain that the non-randomness observed from Figure 2.6 and 2.7 was not due to either the "canonical" or "non-canonical" patterns alone.

[3] The order of these five digit numbers denotes detection status of DNA methylation, H3K9ac, H3K9me1, H3K9me2, and H3K9me3. Refer to Table 2.4 for more details on the numerical representation.

2.4 DISCUSSION

In order to reach the goal of global mapping of epigenetic features in the cancer genome, immunoprecipitation based assays were used for mapping DNA methylation (MeDIP) and histone H3K9 modifications (ChIP) coupled with whole genome tiling path array. ChIP pull down was confirmed successfully by the experimental evaluation of the silent and active marks before generating global histone H3K9 profiles. Since there were duplicate profiles for each of the modifications studied (DNA methylation, H3K9ac, H3K9me1, me2, and me3), we assessed not only the correlation coefficient, but also if the alterations were detected in agreement between the replicates. From the results (Figure 2.4), the observed alterations were not likely due to random chance. Also, DNA methylation gave the most robust duplicate result as it is closest to the perfect replicate scenario. The correlation coefficient numbers between the replicates of the five modifications ranged from 0.71 to 0.87, which is an acceptable range for reproducibility after quality control. Alignment of these profiles for integrative and pattern distribution analysis was demonstrated.

From integrative analysis of the five modifications, 243 possible combinatorial patterns could be generated. We have shown that the distribution of these patterns was not by random chance by comparing the frequencies that would have been expected by random chance with the observed frequencies. Some patterns in particular were observed at higher frequencies than expected whereas most patterns were observed at lower frequencies than expected. Table 2.5 and 2.6 listed patterns that were observed more than one percent different from the expected frequencies. The no modification status detected for all of the

five modifications (pattern 00000) was observed 27% more than expected by random chance (Table 2.5). Pattern "0000-1", no enrichment of H3K9me3 but without detection of the rest of the modifications, was observed almost 3 percent lower compared to the expected frequencies (Table 2.6). It seems that with any one modification status detected, DNA methylation alterations had the greatest differences from the expected ratios. In addition, DNA methylation alterations (pattern 10000, -10000) were observed at higher frequencies compared to the expected. The rest of the patterns under the "one modification status detected" category that involved histone modifications were observed at a lower frequency than expected by random chance. While it seems that DNA methylation is a more frequent event, it is worth pointing out that only the H3K9 modifications were studied in this project. Taking from the histone code concept, it might be interesting to assess interactions of H3K9 with other modifications at different histone residues and then compare with frequency of DNA methylation alterations. Furthermore, it seems that DNA methylation is likely to be an independent event if when four of five the epigenetic marks were detected. However, if all of the five modifications were detected, DNA methylation alteration direction tends to follow with the H3K9ac, H3K9me1, H3K9me2, and H3K9me3 modifications (Table 2.5).

To further examine these 243 combinatorial patterns, these patterns were broken down into categories by the number of modification statuses detected (no modification detected, any one modification status detected, any two, three, four, and all five modification statuses detected). There were some prominent combinatorial patterns such as DNA methylation alterations (pattern -10000, and 10000); no detection of DNA methylation status, but with no enrichment of all the H3K9 histone marks (0-1-1-1-1); and lastly no enrichment of all the marks (-1-1-1-1-1). It is worth noting here that these alterations were detected at

large scale, and thus could be harbouring a mixture of different alterations from each of the marks. Nevertheless, the fact that some of the unexpected combinatorial patterns were more represented (such as no enrichment of all five modification status) prompted the thought to further break down the patterns within each of the category based on canonical and non-canonical patterns (Figure 2.9). The pattern that topped the canonical patterns within the "three modification detected" category with 37% was enrichment of H3K9me1, me2, me3 (00111); whereas the pattern that had the greatest percentage was enrichment of H3K9ac and enrichment of H3K9me2, me3 (01011) with 19%. The fact that there were more non-canonical patterns than canonical patterns probably explains the observation that pattern 01011 topped the list with only 19% of occurrence (Figure 2.9b). The analysis has also revealed non-random distribution of the alteration patterns, and might suggest that certain patterns are favoured or some modifications are dependent on another.

My results supported my hypothesis 1: there exist non-random patterns of epigenetic modifications in the cancer genome. Integrative analysis of histone modification and DNA methylation will show relationships and patterns among these interactions. Future work is still required to further understand the biological meanings behind the interaction among each of these epigenetic modifications.

2.5 CONCLUSION

In my study, I investigated histone H3K9 modifications and DNA methylation at a large scale in a cancer genome and identified non-random distribution of the possible 243 combinatorial patterns. Histone modification is often considered a small scale phenomenon,

and one would speculate that the smaller scale changes would eventually average out and show an equally random distribution at the large scale. Since this is not the case with what was observed in this dataset, this could suggest that histone modification is not only a small scale phenomenon. This novel way of looking at histone modification patterns at large scale could complement other epigenetic findings and further elucidate the crosstalk between DNA methylation and histone modifications.

CHAPTER 3 SPATIAL AND TEMPORAL MAPS OF CISPLATIN CLEARANCE

3.1 PROJECT GOAL

DNA damage can lead to cell cycle arrest and cell death, and therefore, DNA damaging agents have become a major class of chemotherapeutics targeting proliferating cancer cells (Helleday *et al.* 2008; Kelland 2007). DNA lesions can interfere with replication fork progression via creating DNA adducts. This can lead to replication-associated DNA double strand breaks (DSBs), the most toxic damage of all DNA lesions. Cisplatin, a commonly used chemotherapeutic drug, was employed in this project as DNA modifications are induced by this drug. It would be interesting to know if cisplatin is truly global in effects or if it affects DNA adducts at some regions more than others. In addition, we were interested in investigating whether cells clear the damaging agents with any preference for specific genomic regions. Antibodies specific to cisplatin induced DNA adducts were available through collaboration; therefore, the goal of this project was to map cisplatin adducts in cancer genomes by applying the IP technique developed in the previous chapter to assist us in understanding the dynamics of cisplatin clearance.

3.2 METHODS

The aim for this project was to create a genome wide map of cisplatin induced adducts and to study cisplatin adduct clearance. In order to generate these maps, we selected the cancer cell line, SiHa, a cervical cancer cell line, as it is well characterized and not resistant to cisplatin. After a model system had been established, I attempted to adapt the IP technique developed in the previous chapter for mapping cisplatin modifications in a cancer genome. Cells treated with cisplatin allowed us to generate an adduct formation map. Cells allowed to recover for a series of time after treatments were used to generate clearance maps. The experiment scheme is illustrated on Figure 3.1.

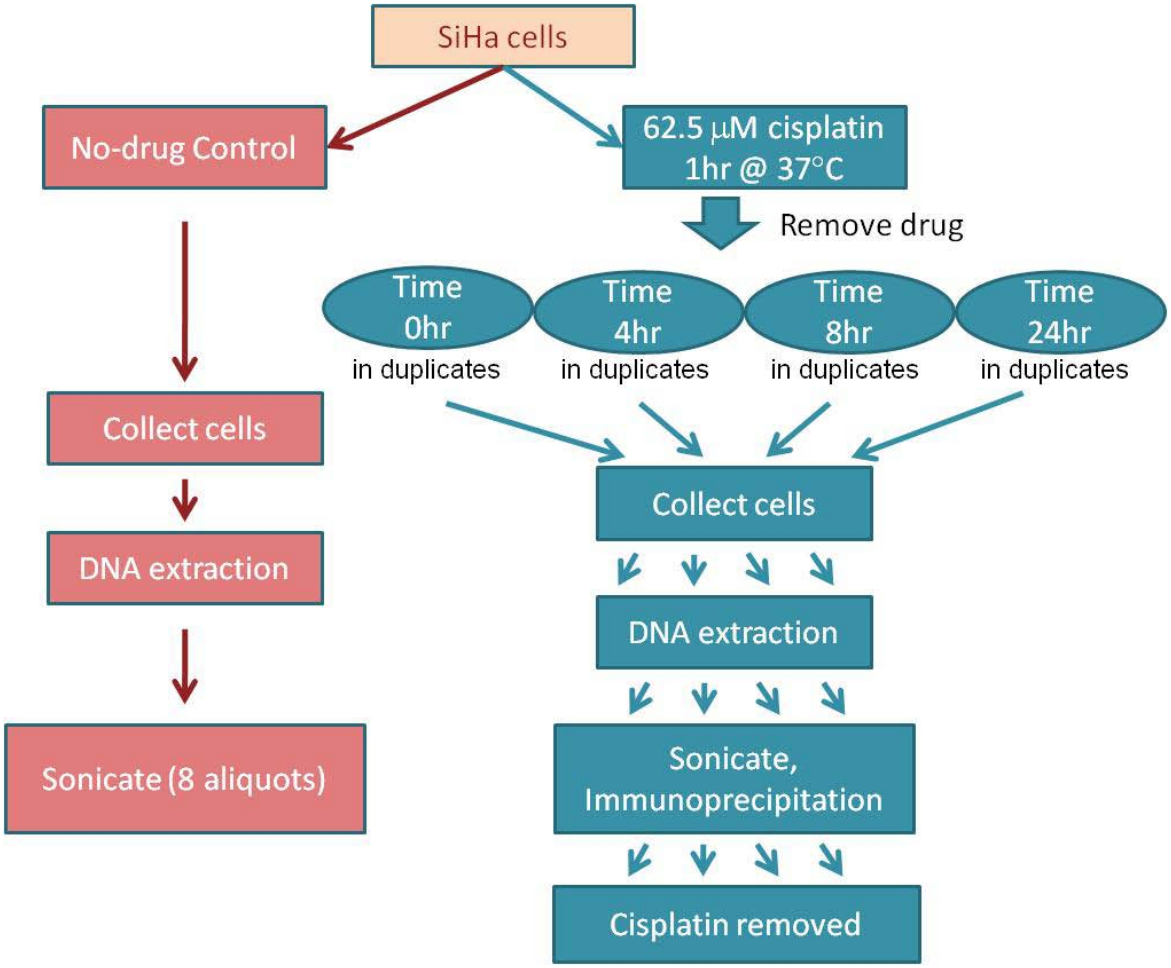


Figure 3.1 Experimental scheme on cisplatin treatment on SiHa cells to achieve Aim 2.

All cells obtained for this experiment are from the same batch of culture. SiHa cells were exposed to 62.5 μM cisplatin for 1 hour at 37°C. Cells were collected at the indicated hours. After DNA extraction, the DNA was sonicated, and then carried on to immunoprecipitation.

Cell Culture

All cells obtained for this experiment were from the same batch of culture. SiHa cells, (from Dr. Peggy Olive, British Columbia Cancer Agency, Canada), were maintained in MEM medium with 10% fetal bovine serum and grown at 37°C in humidified atmosphere containing 5% CO_2.

Determination of Cisplatin Concentration for Treatment

Cisplatin contains platinum and platinum can be measured by atomic absorption spectroscopy (AAS) to provide indication of amount of cisplatin bound to DNA. We needed to determine a cisplatin concentration for treatment that is neither too high that kills all cells nor too low that prevents enough platinum ions bound on DNA for detection of cisplatin integration on DNA using AAS (PerkinElmer AAnalyst 600) and immunoprecipitation. 500 μM cisplatin (Mayne Pharma Inc.) was first introduced to cell culture as an initial concentration for optimization. We tried lowering the drug concentration to 250 μM, 125 μM and 62.5 μM. 62.5 μM was selected as the treatment concentration because this was a concentration that yielded just enough platinum detected by AAS.

A standard curve was generated using platinum atomic absorption standard solution (Sigma). Based on the amount of platinum detected by AAS per given volume, we calculated the amount of platinum ions that could be detected per 500 base pairs (see *Calculation 1* on next page), the fragment size for immunoprecipitation. Further, we calculated that about six percent of the input DNA would have at least one platinum ion bound on the DNA fragments after IP enrichment (see *Calculation 2* on next page). This

calculation is important as this helped us to identify the DNA requirements for IP and that the enriched portion after IP would yield enough DNA for DNA labeling on microarray.

Calculation 1:

From uM of platinum detected by AAS to amount of Platinum molecule/μg DNA

Amount of DNA in a 250 μl measuring cup size: 281.7 ng/μl * 250 μl = 70425 ng DNA

Cisplatin molecular weight: 300.67 g/mole

Platinum level detected by AA (ng/ml) = 16.71 ng/ml

16.71 ng/ml x 250 μl / 70425 ng DNA = 0.06 Platinum/DNA (ng/μg)

Platinum (mole) /DNA (μg) = $0.06e^{-9}$ / 300.67 = $1.97e^{-3}$ mole Platinum /μgDNA = 0.20 ρmol platinum/μgDNA

Total platinum in the sample cup = 0.20 * 70.425μg = 13.89 ρmol

bp / platinum = (70.425 * $9.09e^{14}$) / ($13.89e^{-12}$*$6.06e^{23}$) = 7652.01

Platinum molecule / 500 bp = 500/ 7652.01 = 0.07 platinum molecule/500 bp

Calculation 2:

Calculations of 6.3% IP pull down

500 base pair size – 0.07 platinum molecule / base pair = 499.93

Chances of not having platinum molecule on the fragments: 499.93 / 500 = 0.999869

$0.999869e^{500} = 9.37e^{-1}$

$1- 9.37e^{-1} = 6.33e^{-2}$

$6.33e^{-2}$ *100% = 6.3%

Cisplatin treatment and DNA extraction for IP

Growing cells were exposed to 62.5 μM cisplatin for 1 hour at 37°C. After washing with PBS, cells were transferred to fresh MEM medium, and either collected (0 hour) or further incubated for 4, 8, or 24 hour at 37°C prior to harvesting for DNA extraction. DNA quantities were determined using a spectrophotometer (Nanodrop ND-1000).

R-C18 monoclonal antibodies

The antibodies were contributed by Dr. Jürgen Thomale (University of Duisburg-Essen Medical School, Germany) through collaboration. These antibodies were developed by the Thomale lab specific for the cisplatin induced guanine-guanine intrastrand adducts (Liedert *et al.* 2006)

Optimization of IP

3.5 μg of DNA was sheared and immunoprecipitated with various amounts of R-C18 antibody (0, 0.44, 0.88, 1.75, and 3.5 μg). More sensitive dsDNA quantification was achieved when IP products were stained with fluorescent PicoGreen dye and combined with a fluorospectrometer (Nanodrop ND-3300).

Cisplatin Immunoprecipitation

DNA was sheared by sonication to yield fragments of average 500 bp in size. Since the sonication strength may vary with different sonicator brands and models, fragments size was verified on the agarose gel. Fragmented DNA samples were separated into two fractions: one half was set aside to serve as the reference input (IN), and the other half was

used in the IP reaction with the R-C18 antibodies. The IP fractions with Mab R-C18 were incubated with STE-BSA buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 500 ug BSA/ml) for 12 hours at 4°C. Sheep anti-rat IgG Dynabeads (Invitrogen) were used as secondary antibodies. Cisplatin bound DNA fragments were recovered by adding 10 μg of RNaseA and 100 μg of Pro K followed by phenol/chloroform extraction and ethanol precipitation. To reverse the platinum-DNA crosslinks, the cisplatin IP products were dialyzed (Slide-A-Lyzer Mini Dialysis Unit, Fisher) against 1M Thiourea for 24 hours at 37°C. Thiourea was removed by dialysis against 20 mM potassium phosphate buffer at 4°C for 18 hours.

Coupling immunoprecipitation with whole genome tiling path array

In order to generate a genome wide map of cisplatin adducts, genomic DNA from cisplatin treated SiHa cell culture was collected and cisplatin induced guanine-guanine adducts were enriched via immunoprecipitation as described above. IN and IP DNA samples were differentially labeled with fluorescent dyes (Cy5 and Cy3 respectively) and were co-hybridized to the whole genome tiling path array. DNA labeling, hybridization, array imaging, and analysis were performed as described in Ishkanian *et al* (Ishkanian *et al.* 2004).

Comparing observed vs. expected frequencies of alterations

Normalized Cy3/Cy5 intensity $\log_2$ ratios were scale normalized by conversion to Z scores to allow use of thresholds for comparison across samples. We then calculated the frequencies of alterations observed in the dataset. Refer to Table 2.1 for example of analyzing alterations and calculating frequencies of alterations.

Moving average thresholding

Thresholding of moving average data was applied across the replicates of each of the time points (0, 4, 8, and 24 hour). This involves calculating the average across a sliding window of data points (600 kb windows sliding at 100 kb intervals). Averages were taken for each of the time point replicate experiments. Figure 3.5 and 3.6 show how aligned profiles help to assess slow/rapid patterns in terms of cisplatin clearing.

Quantitative PCR

qPCR experiments and data analysis were performed using standard ABI Taqman protocols with ABI 7500 Fast System. Triplicates were performed for each time point with each of the primers selected. Selected Taqman primers were: *DDAH1*, *MC2R*, *LRRTM4*, *DHFR*, *CXCR7*, and *TCFL* (Applied Biosystems).

Efficiency of the Taqman primers was first tested for accurate C_T adjustment. We investigated several relative quantification methods to decide a suitable C_T correction method for the purpose of this study. Livak and Schmittgen use the $2^{-\Delta\Delta CT}$ method to assess relative expression data (Livak and Schmittgen 2001). Squire's group use a data normalization equation $KC_{ti} = [(AC_{tR}-C_{tRi})/S_R]*S_T + C_{tTi}$ (Weksberg *et al.* 2005). However, they provide different results when tried on the data from this project as one works with log ratio (Livak and Schmittgen) and the other is based on linearized data (Weksberg). The other concern we have on applying these methods was that the cisplatin changes in reference to the normal could be changing with the different time points, and therefore using the above described relative calculation may not be applicable providing the fact that each of the cisplatin treatment time point stands alone by itself, having its own reference. We here propose a

method for adjusting the C_T based on primer efficiency in a non-relative/independent way: $C_{T,Expected} = \log_2(\text{Primer Efficiency}^{\wedge\ CT,\ Observed})$. This equation means assuming the primer is working at 100% efficiency and with doubling with every cycle, $C_{T,Expected}$ will be the expected threshold cycle.

3.3 RESULTS

3.3.1 Determining cisplatin incorporation in DNA after treatment

Various concentrations of cisplatin were used to treat SiHa cell cultures as part of an optimization process to identify a concentration that is neither too high leading to complete cell death nor too low that would prevent AAS detection of cisplatin integration on DNA after treatment. A linear dose dependent relationship was observed between platinum molecules/500 bp detected and concentration of cisplatin (Figure 3.2). Based on these results, it was calculated that 3.5 μg of DNA is required for cisplatin immunoprecipitation to generate 200 ng of enriched products, fulfilling the DNA requirements for microarray experiments.

a. Calculated platinum molecules/500 bp for treated cisplatin concentration

Cisplatin (μM)	Platinum molecules /500 bp
62.5 μM	0.07
125 μM	0.13
250 μM	0.28

b. Chart plotted using results from the chart above

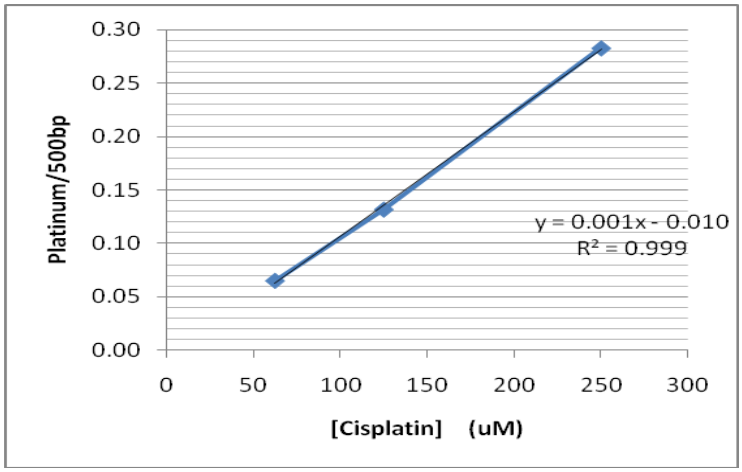


Figure 3.2 Linear drug response observed from cisplatin treatment on SiHa cells.

Platinum content in DNA was detected using AAS. a) platinum content was calculated as platinum molecule/500 bp using calculation 1 in Section 3.2. b) The results from a) are plotted as chart.

3.3.2 To assess the optimal antibody concentration for immunoprecipitation

It is important to know if we were saturating the antibody because beads not saturated with antibodies would lead to increased background signal due to non-specific binding. On the other hand, we did not want to be wasteful using too much antibodies beyond the saturation point. Therefore, IPs with different amount of antibodies were assayed. The amount of DNA pulled down was quantified by flurospectrometer (Figure 3.3). The titration curve generated showed that 1.75 µg of antibody would be the optimal IP amount.

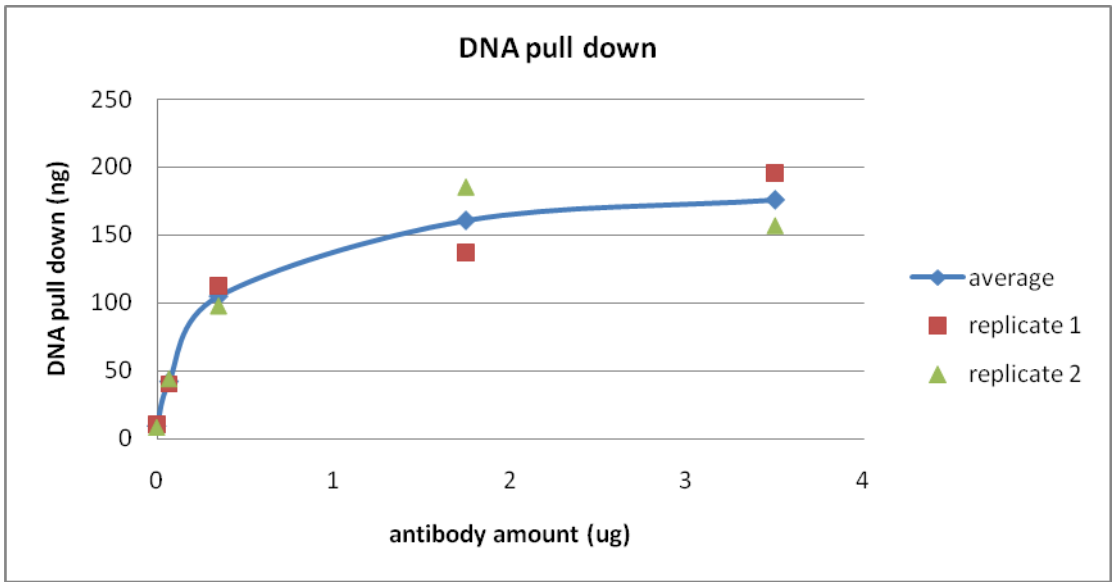


Figure 3.3 Mab R-C18 antibody titration curve.

0 to 3.5 µg of antibodies were added to 3.5 µg of DNA. This titration curve showed that 1.75 µg of antibody in a 50 µl reaction volume would give an optimal working concentration in the IP experiment.

3.3.3 Expected vs. observed frequency of alterations and reproducibility of experiments

Similar to Section 2.3.3, we examined the observed frequencies of alterations and frequencies that would have been expected due to random chance. Figure 3.4 shows the observed frequencies in each of the time points with the expected frequencies by random chance and if in a perfect replicate scenario. In a perfect replicate scenario, duplicates would always agree: either "no alterations detected in either replicate" or "alteration detected in both replicates". Please refer to Section 2.3.3 and Table 2.3 for details on calculating these frequencies for perfect replicate scenario. The results showed the observed frequencies of alterations were not likely due to random chance, and thus our data represents non-random distribution of cisplatin induced DNA adducts.

The correlation coefficient of the replicates within each time point (0, 4, 8, and 24 hour post-incubation upon cisplatin administration) range from r =0.65 to r =0.91.

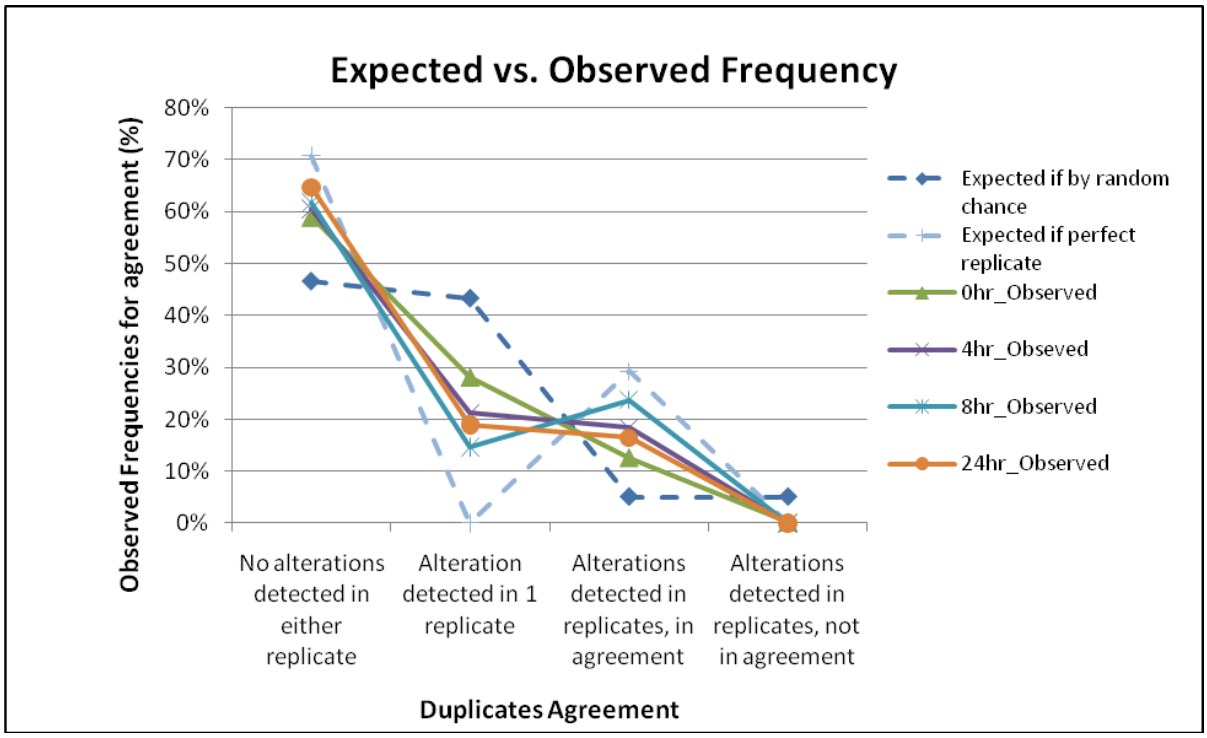


Figure 3.4 Expected vs. observed frequency of alterations for 0, 4, 8, 24 hour post cisplatin treatment on SiHa cells.

Frequencies of alterations were taken from all the profiles (observed, solid lines) and plotted against the frequencies that would have been expected by random chance (dotted lines). All of the profiles did not follow the trend of random chance, showing that the frequencies of alterations observed from all of my data were not by random chance. Further, a line of "expected if perfect replicate" was also plotted assuming that the perfect scenario would be either "No alteration detected in either replicate" or "Alterations detected in replicates, in agreement". 8 hour post-incubation, in particular, was the closet to "perfect". (P value from binomial test is less than $1X10^{-6}$ for all data points).

3.3.4 Cisplatin adduct formation/clearance map

In order to understand the clearance of cisplatin adducts, profiles of treated cells with different recovery time were aligned. The 0 hour profile provides a baseline of the cisplatin adduct formation, and the aligned profiles of different recovery time revealed slow and rapid clearance patterns in the genome upon cisplatin adduct formation. Profiles visualized with $SIGMA^2$ software showed that the end of 2q and 20q tend to have increasing patterns of residual platinum detected throughout the time frame. This could point to a slow clearing process of platinum in the adducts (see Figure 3.5 and 3.6).

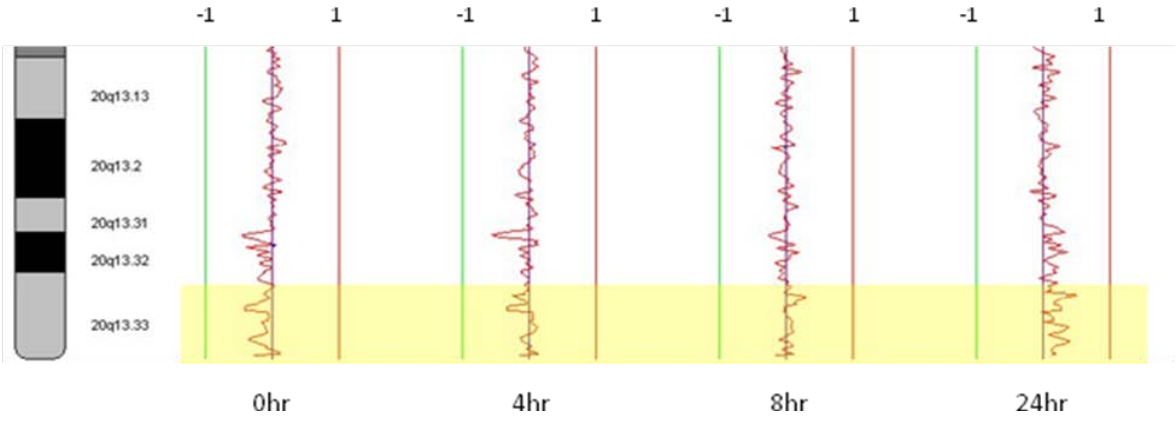
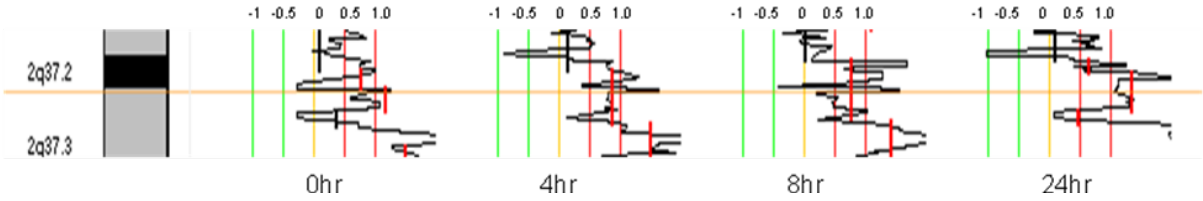


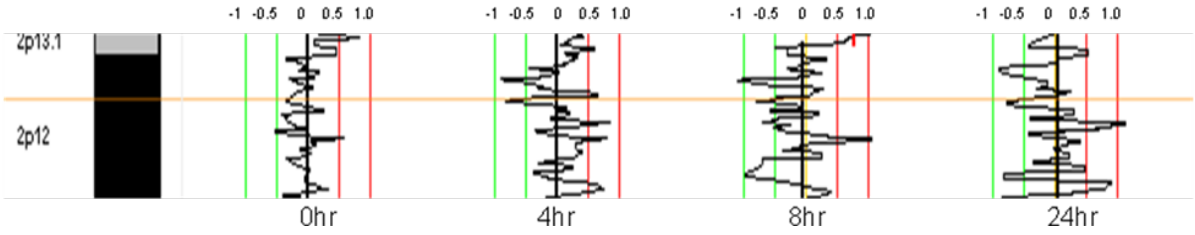
Figure 3.5 Cisplatin global adduct profile and clearance map of SiHa cells.

The midline of each of the profile is the average platinum content in the genome ($\log_2$ ratio of zero). By looking at the 0 hour profile, non-random cisplatin adduct distribution is observed. By examining the recovery time profiles (4, 8, 24 hour), it appears that 20q13 contained less platinum than the rest of the genome after initial treatment. However as the platinum was cleared from the rest of the genome, 20q13 appeared to be more slowly cleared and by 24 hours contained more platinum than the majority of the genome.

Slow Clearing



Neutral



Rapid Clearing

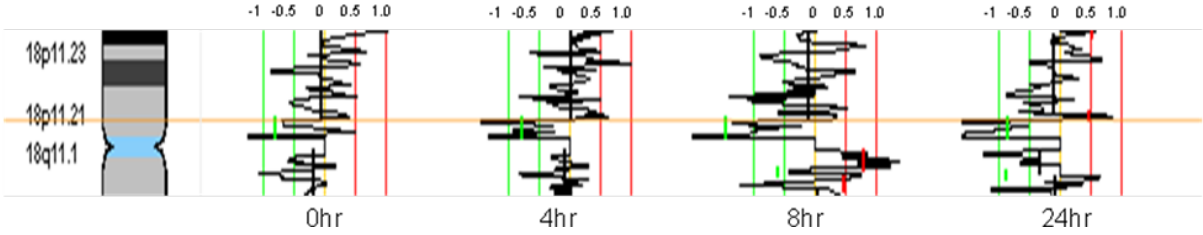


Figure 3.6 Slow and rapid cisplatin clearance.

Platinum was being constantly cleared out throughout time. For slow clearing, by the end of 24 hours, more platinum was retained compared to the rest of the genome. For rapid clearing, by the end of 24 hours, the region highlighted with an orange line still had less platinum content compared to the rest of the genome. The midline of each of the profiles is the average platinum content in the genome ($\log_2$ ratio of zero).

3.3.5 Quantitative PCR validation of selected loci of fast and slow clearance

Genomic regions that demonstrated slow/rapid cisplatin clearance throughout the recovery period were selected for validation. Two Taqman primers (Applied Biosystems) were selected from each of the clearance patterns: Rapid clearance (*DDAH1*, *MC2R*), Neutral (*LRRTM4*, *DHFR*), Slow clearance (*CXCR7*, *TCFL5*). The limitation in using this method for validation is that primers assay only a ~80 bp fragment, whereas the BAC size on our genomic microarray is in the order of 100 kb (see Figure 3.6).

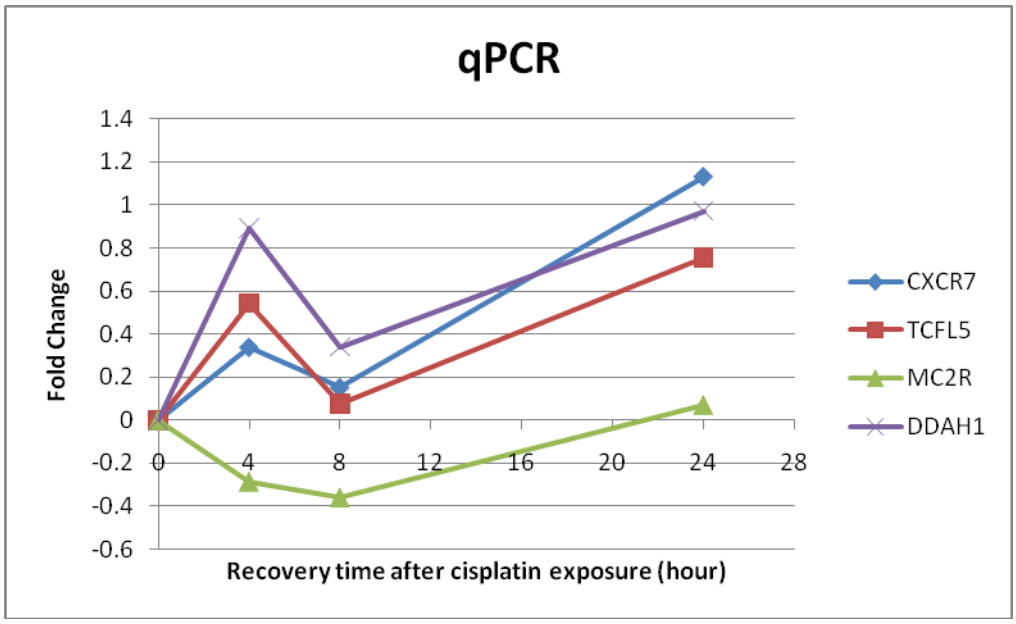


Figure 3.7 qPCR results from selected loci of fast and slow clearance.

Genomic copies amplified at *CXCR7*, *TCFL5*, *MC2R*, and *DDAH1* relative to those amplified at the selected loci of neutral pattern (*LRRTM4*, *DHFR*) (Fold changes) with the respective recovery time point are shown here.

3.4 DISCUSSION

The IP technique developed in Aim 1 (Chapter 2) was adapted to generate cisplatin adduct formation and clearance maps in a cancer genome. Prior to generating the cisplatin adduct maps, the IP protocol was optimized by: 1) Treating cells with various concentrations of cisplatin to assess DNA integration of the drug. From the data, a dose dependent linear relationship between the concentration of cisplatin and the amount of adduct detected was observed. We further calculated, based on the platinum content detected, the amount of DNA fragments that would have at least one platinum adduct, allowing IP. This information is particularly important as we wanted to ensure that we have enough DNA going through the IP to generate sufficient amount of products for microarray labeling. 2) Antibody titration was performed to assess the optimal working concentration. After the amounts of DNA and antibodies were optimized, global cisplatin adduct and clearance maps were generated using the whole genome tiling path arrays.

To assess the reproducibility of the technique, we applied similar criteria to that used in Aim 1, namely we compared the observed frequency of alterations to the frequency of alterations that would have been expected by random chance. From Figure 3.4, we can see that the observed alterations were not due to random chance. Furthermore, the 8 hour post treatment profile was shown to be the most robust as this profile is closest to the "perfect replicate scenario" situation and had the highest correlation coefficient (r=0.91) among others.

The 0 hour profile represented the baseline cisplatin adduct map. From the profile, we observed non-random hits on the genome by the drug (Figure 3.5), and we have already

shown that these alterations did not occur as a result of random chance by comparing the expected and observed frequencies of alterations (Figure 3.4). Moreover, cells that were treated with cisplatin and allowed to recover for various length of time enabled us to study the dynamics of cisplatin clearance. When the profiles from 0, 4, 8, and 24 hour recovery time post treatment were aligned, we identified slow and rapid cisplatin clearance patterns in reference to the average platinum content in the genome (Figure 3.5 and 3.6). qPCR was attempted to validate selected loci of fast and slow clearance, however qPCR has its own limitation for this purpose (Section 3.2.6). My data suggested that the fine scale qPCR results do not correspond to the findings from my large scale analysis highlighting the complexity of cisplatin clearance. Other than the limitation of using qPCR to validate such a phenomenon, one other explanation could be the regions picked from the reference loci may not be the optimal reference, causing the upward shift on all of the four regions from 8 hour to 24 hour. Higher resolution technologies that are more comprehensive than the currently employed arrays (such as oligonucleotide arrays) may be required to further study these clearance patterns.

Lastly, while trying to understand the biology behind these patterns, we attempted to assess the correlation between the platinum content signal ratios to GC content percentage and found not much correlation at the genome level. Correlation coefficients (r) are 0.49, 0.46, 0.38, and 0.45 for 0, 4, 8, and 24 hour profile. Based on the low correlation coefficient, we could not conclude here whether GC rich regions would harbour more platinum content as these fragments were enriched based on the specificity of antibodies towards GG adducts. Additionally, one other factor contributing to the low correlation could be the fact that the GC content data and array CGH data were taken from different resolution.

From these results, global adduct profiles generated (0 hour profile) revealed non-random integration of cisplatin, this supports hypothesis 2, that DNA adduct induced by cisplatin is non-random throughout the genome. These slow/rapid clearing patterns suggested that some regions were preferentially repaired supporting hypothesis 3, that clearance of cisplatin induced DNA adduct occurred non-randomly in the human genome.

3.5 CONCLUSION

One of the aspects of cancer is genome instability which can be caused from DNA damage. In my study, I studied the global mapping of cisplatin induced adducts and examined if there were preferential sites of cisplatin integration. We observed regions that show higher and lower platinum content when compared to the rest of the genome. After cisplatin exposure, some regions appeared to lose platinum at different rates compared to the genome on average. Regions that retained platinum adducts could be more difficult to repair or less essential to the function of the cell. Conversely, sites that were repaired rapidly could harbor essential genes, and thus, suggested that DNA repair activity after cisplatin exposure is non-random. This brings to our current knowledge a new concept that cisplatin affects certain regions more than others instead of global effects, and implicates cisplatin specific integration sites. Understanding the cisplatin specific integration sites in a timely manner would greatly help in designing new cancer treatment strategy including targeting cellular functions at sites that react closest to DNA adduct formation, and also lowering cisplatin related toxicity.

CHAPTER 4 OVERALL SUMMMARY, DISCUSSION, FUTURE DIRECTIONS, AND CONCLUSION

4.1 OVERALL SUMMARY

In my thesis, IP techniques were employed to study modification on DNA from epigenetic marks and the chemotherapeutic drug cisplatin. MeDIP and ChIP were each coupled with whole genome tiling path arrays to generate DNA methylation and histone H3K9 modification profiles (H3K9ac, H3K9me1, H3K9me2, and H3K9me3) respectively. These profiles were aligned on the same visualization software for integrative analysis. We observed different interactions among these profiles from the integrative analysis, which prompted us to study the distribution of the possible 243 combinatorial patterns. We first examined the observed frequencies of each represented pattern and compared this to the frequencies that would have occurred by random chance. The result revealed non-random patterns of epigenetic modifications, supporting hypothesis 1: "There exist non-random patterns of epigenetic modifications in the cancer genome. Integrative analysis of histone modification and DNA methylation will show relationships and patterns among these interactions." We further broke down the 243 patterns into categories based on the number of modification detected out of the five modifications. Again, some patterns tend to be more prominent than others, and we have identified patterns that tend not to be expected (referred as non-canonical patterns) as H3K9 methylations are generally associated with transcriptional silencing and H3K9ac is associated with active transcription (Bernstein *et al.*

2007). The results showed the epigenetic marks among these five modifications were not distributed equally throughout the genome at large scale.

Several studies have been carried out using ChIP-chip to study histone modifications (Kondo *et al.* 2004; Miao and Natarajan 2005; Bernstein *et al.* 2005; Heintzman *et al.* 2007; Wang *et al.* 2008; Guenther *et al.* 2007; Barski *et al.* 2007). By studying H3K9 methylations to H3K9 acetylation ratio, Kondo *et al*. found H3K9 modifications and DNA methylation coexist frequently at silenced loci in colon cancer cells (Kondo *et al.* 2004). A study focusing on histone methylation patterns in coding regions was carried out and found poor correlation between euchromatin marks (H3K9/K14 di-acetylation, H3K4me2, H3K36me2, and H3K79me2) and heterochromatin marks (H3K9me2, H3K9me3, H3K27me2, and H4K20me2) (Miao and Natarajan 2005). These heterochromatin marks were found to be associated with hypoacetylation (Miao and Natarajan 2005). In a different study, H3K4me3 was shown to be correlated with H3K9/K14 deacetylation at transcription start sites of active genes (Bernstein *et al.* 2005; Guenther *et al.* 2007). The specific chromatin structures could also help to characterize and predict specific transcription regulating elements. Heintzman *et al*. identified high levels of H3K4me1 and low levels of H3K4me3 as characteristics of enhancers (Heintzman *et al.* 2007). However, high levels of H3K4me3 have also been found at many other enhancers in human $CD4^+$ T cells (Barski *et al.* 2007). The difference in the findings might be because the former study only focused on 1% of the genome while the latter study was carried on a whole genome scale. Histone methylation profiling by Barski *et al*. was further complemented by reports on histone acetylation (Wang *et al.* 2008). Together, they concluded that regulatory elements of transcription can be associated with multiple patterns of histone modifications (Wang *et al.* 2008). My study did not address

specific patterns in relation to regulatory elements of transcription; however, we provided a spatial map in particular to H3K9 modifications to understand the distribution of the combinatorial patterns. The non-random distribution of these combinatorial patterns supports, and, with future incorporation of expression data and investigation across different cell types, will contribute to our understanding of the histone code.

The IP technique developed was adapted for mapping cisplatin induced adduct in a cancer genome. Antibodies used in this IP were obtained through collaboration with Thomale Lab in Germany. The cisplatin treatment and IP protocol have been optimized for quantities of input DNA and antibodies to generate IP products fulfilling DNA amount requirement for arrays. Global maps revealed non random integration of cisplatin on DNA and supported hypothesis 2: "Cisplatin, a chemotherapeutic agent, induces DNA adducts in a non-random manner throughout the genome." Through studying the profiles of cells that were allowed to recover for a series of time post-treatment (0, 4, 8, and 24 hour), we identified some regions that tend to retain cisplatin throughout time (slow clearing) and regions that throughout time have cleared cisplatin contents (fast clearing). This observation supports my hypothesis 3: "The clearance of DNA adducts induced by cisplatin occurs non-randomly in the human genome." The genome wide mapping of cisplatin clearance I have generated showed there are non-random effects of cisplatin on the human genome. Further studies of specific regions might help us to understand if there are cisplatin specific integration sites, and if yes, whether these damage sites are truly caused by cisplatin or whether there are cisplatin specific fragile sites (Buttel *et al.* 2004).

Mappings of global epigenetic alterations and mappings of cisplatin induced adduct have revealed combinatorial epigenetic patterns and preferential cisplatin clearance.

4.2 OVERALL DISCUSSIONS

4.2.1 Cisplatin induced DNA damage and histone modification

It is a well accepted view that modifications on DNA alters chromatin structure, and could influence DNA repair, transcription and other processes in different ways (Wang and Lippard 2004). Cisplatin has been found to induce histone H3 serine 10 phosphorylation in some cancer cell lines as well as H3 serine 28 phosphorylation and histone H4 acetylation in HeLa cells (Wang and Lippard 2004). Since acetylation and phosphorylation of histones are linked with active transcription, this observation suggests that nucleosome structural changes might increase accessibility of other proteins to facilitate DNA repair. Even though there is no evidence to date on specific interaction of H3K9 modifications with cisplatin induced damage, given the current understanding of how this residue is involved with transcription, H3K9 modifications may contribute to DNA adduct formation associated mechanism via an indirect role.

4.2.2 Chromatin proteins and cisplatin damage

In general, three models have been suggested for possible interactions between chromatin proteins and cisplatin damage (Widlak *et al.* 2006). Damage shielding model: upon cisplatin damage, the chromatin protein bound onto the damaged sites, shielding the damage from repair. This has led to increased sensitivity of cells to cisplatin as suggested by several reports (He *et al.* 2000; Zamble *et al.* 2002). Repair enhancement model: alternatively, since the chromatin proteins control the fluidity of the chromatin, they could

relax the tight complex to allow the damaged sites to be more accessible to repair factors. Damage signaling model: another model proposed was that the presence of specific chromatin proteins serves as a signal that can be recognized by repair factors.

4.2.3 Combining histone deacetylase inhibitors (HDACi) to cisplatin treatment

The effectiveness of DNA damaging agents can be impaired due to inaccessibility to DNA, which can be altered by histone modifications. Since HDACi increases acetylation of histones, leading to an open chromatin structure that is more accessible to DNA targeting chemotherapeutic drugs. Kim *et al.* pretreated cancer cell lines with HDACi prior to cisplatin treatment, and found increased cytotoxicity of cisplatin (Kim *et al.* 2003). Furthermore, it has been shown that coupling inhibitors of DNA methyltransferases and histone deacetylase together with cytotoxics such as cisplatin re-expressed DNA repair associated genes that were found methylated before treatment (Steele *et al.* 2009). The level of re-expression was found to be higher than that caused from either of epigenetic inhibitors alone (Steele *et al.* 2009). These results might suggest chromatin configuration as another important feature for cisplatin adduct formation and as well as cytotoxicity regulation.

4.3 FUTURE DIRECTION

In order to further study cisplatin clearance patterns and to investigate how transcription is affected by cisplatin treatment, I would like to collect more cisplatin treated SiHa cells to generate gene expression profiles. Arrays such as oligonucleotide or customed arrays that offer smaller target size would be considered to study these alterations at finer scales.

Furthermore, it would be interesting if I could incorporate the experiment and analysis techniques from the epigenomics project with clearance maps after we have gained more understanding of the cisplatin clearance patterns. Integrating genomic and epigenomic approach would allow us to understand the mechanisms behind changes in gene expression. Therefore, it would help to identify genes and pathways involved in chromatin structure due to DNA modifications from cisplatin and benefit more understandings of genomic treatment response.

4.4 CONCLUSION

This work contributes to a greater understanding of epigenomic landscapes in cancer and in response to drug treatment. I have identified large scale epigenetic alterations in cancer genomes. Furthermore, I have identified specific regions of the genome associated with fast and slow platinum clearance in cancer. These findings may be complemented by integration of loci specific fine scale studies and utilization of higher resolution technologies to further elucidate the underlying mechanisms of genome instability and DNA adduct formation in cancer.

BIBLIOGRAPHY

Ballestar E, Paz MF, Valle L, Wei S, Fraga MF, Espada J, Cigudosa JC, Huang TH, Esteller M. 2003. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. EMBO Journal 22:6335-45.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. Cell 129:823-837.

Baylin SB. 2005. DNA methylation and gene silencing in cancer. Nature: Clinical Practice Oncology 2 Supplement 1:S4-11.

Baylin SB, Belinsky SA, Herman JG. 2000. Aberrant methylation of gene promoters in cancer---concepts, misconcepts, and promise. Journal of the National Cancer Institute 92:1460-1461.

Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. 2001. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Human Molecular Genetics 10:687-692.

Bell B, Scheer E, Tora L. 2001. Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function. Molecular Cell 8:591-600.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ,3rd, Gingeras TR, Schreiber SL, Lander ES. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120:169-181.

Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. Cell 128:669-681.

Buttel I, Fechter A, Schwab M. 2004. Common fragile sites and cancer: targeted cloning by insertional mutagenesis. Annals of the New York Academy of Sciences 1028:14-27.

Caldecott KW. 2008. Single-strand break repair and genetic disease. Nature Reviews Genetics 9:619-631.

Callinan PA and Feinberg AP. 2006. The emerging science of epigenomics. Human Molecular Genetics 15 Spec No 1:R95-101.

Chari R, Coe BP, Wedseltoft C, Benetti M, Wilson IM, Vucic EA, MacAulay C, Ng RT, Lam WL. 2008. SIGMA2: a system for the integrative genomic multi-dimensional analysis of cancer genomes, epigenomes, and transcriptomes. BMC Bioinformatics 9:422.

Chi B, DeLeeuw RJ, Coe BP, MacAulay C, Lam WL. 2004. SeeGH--a software tool for visualization of whole genome array comparative genomic hybridization data. BMC Bioinformatics 5:13.

Cohen SM and Lippard SJ. 2001. Cisplatin: from DNA damage to cancer chemotherapy. Progress in Nucleic Acid Research And Molecular Biology 67:93-130.

Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC, Schuller DE, Yu L, Bloomfield CD, Caligiuri MA, Yates A, Nishikawa R, Su Huang H, Petrelli NJ, Zhang X, O'Dorisio MS, Held WA, Cavenee WK, Plass C. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nature Genetics 24:132-8.

Eastman A. 1990. Activation of programmed cell death by anticancer agents: cisplatin as a model system. Cancer Cells 2:275-280.

Eden A, Gaudet F, Waghmare A, Jaenisch R. 2003. Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300:455.

Ehrlich M. 2006. Cancer-linked DNA hypomethylation and its relationship to hypermethylation. Current Topics in Microbiology and Immunology 310:251-274.

Ehrlich M. 2002. DNA methylation in cancer: too much, but also too little. Oncogene 21:5400-13.

Esteller M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene 21:5427-40.

Esteller M, Corn PG, Baylin SB, Herman JG. 2001. A gene hypermethylation profile of human cancer. Cancer Research 61:3225-9.

Esteller M and Herman JG. 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. Journal of Pathology 196:1-7.

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG. 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. Journal of the National Cancer Institute 92:564-569.

Fahrner JA, Eguchi S, Herman JG, Baylin SB. 2002. Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. Cancer Research 62:7213-7218.

Feinberg AP and Tycko B. 2004. The history of cancer epigenetics. Nature Reviews Cancer 4:143-53.

Fong KM, Sekido Y, Gazdar AF, Minna JD. 2003. Lung cancer. 9: Molecular biology of lung cancer: clinical implications. Thorax 58:892-900.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America 89:1827-1831.

Fuks F. 2005. DNA methylation and histone modifications: teaming up to silence genes. Current Opinion in Genetics & Development 15:490-495.

Gonzalgo ML, Liang G, Spruck CH,3rd, Zingg JM, Rideout WM,3rd, Jones PA. 1997. Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. Cancer Research 57:594-9.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130:77-88.

Hatada I, Hayashizaki Y, Hirotsune S, Komatsubara H, Mukai T. 1991. A genomic scanning method for higher organisms using restriction sites as landmarks. Proceedings of the National Academy of Sciences of the United States of America 88:9523-7.

He Q, Liang CH, Lippard SJ. 2000. Steroid hormones induce HMG1 overexpression and sensitize breast cancer cells to cisplatin and carboplatin. Proceedings of the National Academy of Sciences of the United States of America 97:5768-5772.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and

predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics 39:311-318.

Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. 2008. DNA repair pathways as targets for cancer therapy. Nature Reviews Cancer 8:193-204.

Hoeijmakers JH. 2001. Genome maintenance mechanisms for preventing cancer. Nature 411:366-374.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. Nature Genetics 36:299-303.

Jaenisch R and Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genetics 33 Supplement:245-254.

Jones PA and Baylin SB. 2002. The fundamental role of epigenetic events in cancer. Nature Reviews Genetics 3:415-28.

Kelland L. 2007. The resurgence of platinum-based cancer chemotherapy. Nature Reviews Cancer 7:573-584.

Kennett JY, Watson SK, Saprunoff H, Heryet C, Lam WL. 2008. Technical demonstration of whole genome array comparative genomic hybridization. Journal of Visualized Experiments 18.

Khanna KK and Jackson SP. 2001. DNA double-strand breaks: signaling, repair and the cancer connection. Nature Genetics 27:247-254.

Kim MS, Blake M, Baek JH, Kohlhagen G, Pommier Y, Carrier F. 2003. Inhibition of histone deacetylase increases cytotoxicity to anticancer drugs targeting DNA. Cancer Research 63:7291-7300.

Kondo Y, Shen L, Yan PS, Huang TH, Issa JP. 2004. Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. Proceedings of the National Academy of Sciences of the United States of America 101:7398-7403.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128:693-705.

Kuo ANC, Wilson IM, Vucic E, Lee EHL, Davies JJ, MacAulay C, Brown CJ, Lam WL. 2008. “Comparative cancer epigenomics” in Comparative genomics: Fundamental and applied perspectives (Brown JR, ed.), CRC Press / Taylor and Francis, LLC. Boca Raton, FLA. pp. 261-279.

Kuzminov A. 2001. Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proceedings of the National Academy of Sciences of the United States of America 98:8241-8246.

Liedert B, Pluim D, Schellens J, Thomale J. 2006. Adduct-specific monoclonal antibodies for the measurement of cisplatin-induced DNA lesions in individual cell nuclei. Nucleic Acids Research 34:e47.

Livak KJ and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-8.

Marks P, Rifkind RA, Richon VM, Breslow R, Miller T, Kelly WK. 2001. Histone deacetylases and cancer: causes and therapies. Nature Reviews Cancer 1:194-202.

Miao F and Natarajan R. 2005. Mapping global histone methylation patterns in the coding regions of human genes. Molecular and Cellular Biology 25:4650-61.

Oberley MJ, Tsao J, Yau P, Farnham PJ. 2004. High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. Methods in Enzymology 376:315-334.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. 2000. Genome-wide location and function of DNA binding proteins. Science 290:2306-2309.

Rich T, Allen RL, Wyllie AH. 2000. Defying death after DNA damage. Nature 407:777-783.

Rosenberg B, VanCamp L, Trosko JE, Mansour VH. 1969. Platinum compounds: a new class of potent antitumour agents. Nature 222:385-386.

Santos-Rosa H and Caldas C. 2005. Chromatin modifier enzymes, the histone code and cancer. European Journal of Cancer 41:2381-2402.

Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & Development 18:1263-1271.

Shames DS, Girard L, Gao B, Sato M, Lewis CM, Shivapurkar N, Jiang A, Perou CM, Kim YH, Pollack JR, Fong KM, Lam CL, Wong M, Shyr Y, Nanda R, Olopade OI, Gerald W, Euhus DM, Shay JW, Gazdar AF, Minna JD. 2006. A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. PLoS Medicine 3:e486.

Shilatifard A. 2006. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. Annual Review of Biochemistry 75:243-269.

Steele N, Finn P, Brown R, Plumb JA. 2009. Combined inhibition of DNA methylation and histone acetylation enhances gene re-expression and drug sensitivity in vivo. British Journal of Cancer 100:758-763.

Strahl BD and Allis CD. 2000. The language of covalent histone modifications. Nature 403:41-5.

Suzuki H, Watkins DN, Jair KW, Schuebel KE, Markowitz SD, Chen WD, Pretlow TP, Yang B, Akiyama Y, Van Engeland M, Toyota M, Tokino T, Hinoda Y, Imai K, Herman JG, Baylin SB. 2004. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nature Genetics 36:417-422.

Thu KL, Vucic EA, Kennett JY, Heryet C, Brown CJ, Lam WL, Wilson IM. 2009. Methylated DNA immunoprecipitation. Journal of Visualized Experiments 23.

Tlsty TD, Crawford YG, Holst CR, Fordyce CA, Zhang J, McDermott K, Kozakiewicz K, Gauthier ML. 2004. Genetic and epigenetic changes in mammary epithelial cells may mimic early events in carcinogenesis. Journal of Mammary Gland Biology and Neoplasia 9:263-274.

Wang D and Lippard SJ. 2005. Cellular processing of platinum anticancer drugs. Nature Reviews Drug Discovery 4:307-320.

Wang D and Lippard SJ. 2004. Cisplatin-induced post-translational modification of histones H3 and H4. The Journal of Biological Chemistry 279:20622-20625.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nature Genetics 40:897-903.

Watson SK, deLeeuw RJ, Horsman DE, Squire JA, Lam WL. 2007. Cytogenetically balanced translocations are associated with focal copy number alterations. Human Genetics 120:795-805.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genetics 37:853-62.

Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes & Development 16:235-244.

Weksberg R, Hughes S, Moldovan L, Bassett AS, Chow EW, Squire JA. 2005. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. BMC Genomics 6:180.

Widlak P, Pietrowska M, Lanuszewska J. 2006. The role of chromatin proteins in DNA damage recognition and repair. Histochemistry and Cell Biology 125:119-126.

Wilson AS, Power BE, Molloy PL. 2007. DNA hypomethylation and human diseases. Biochimica et Biophysica Acta 1775:138-162.

Wilson IM, Davies JJ, Weber M, Brown CJ, Alvarez CE, MacAulay C, Schubeler D, Lam WL. 2006. Epigenomics: mapping the methylome. Cell Cycle 5:155-8.

Wu B, Droge P, Davey CA. 2008. Site selectivity of platinum anticancer therapeutics. Nature Chemical Biology 4:110-112.

Wu J, Smith LT, Plass C, Huang TH. 2006. ChIP-chip comes of age for genome-wide functional analysis. Cancer Research 66:6899-6902.

Yates DR, Rehman I, Abbod MF, Meuth M, Cross SS, Linkens DA, Hamdy FC, Catto JW. 2007. Promoter hypermethylation identifies progression risk in bladder cancer. Clinical Cancer Research 13:2046-2053.

Zamble DB, Mikata Y, Eng CH, Sandman KE, Lippard SJ. 2002. Testis-specific HMG-domain protein alters the responses of cells to cisplatin. Journal of Inorganic Biochemistry 91:451-462.

APPENDIX I List of Publication

Kuo ANC, Wilson IM, Vucic E, Lee EHL, Davies JJ, MacAulay C, Brown CJ, Lam WL. 2008. "Comparative cancer epigenomics" in Comparative genomics: Fundamental and applied perspectives (Brown JR, ed.), CRC Press / Taylor and Francis, LLC. Boca Raton, FLA. pp. 261-279.