12-2010

IDENTIFICATION OF HOLOCARBOXYLASE SYNTHETASE CHROMATIN BINDING SITES IN HUMAN MAMMARY CELL LINES USING THE DAMID TECHNOLOGY

Dipika Singh
University of Nebraska-Lincoln, dipikasingh10@gmail.com

IDENTIFICATION OF HOLOCARBOXYLASE SYNTHETASE CHROMATIN BINDING SITES IN HUMAN MAMMARY CELL LINES USING THE DAMID TECHNOLOGY

by

Dipika Singh

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Agricultural and Biological Systems Engineering

Under the Supervision of Professors Angela K. Pannier and Janos Zempleni

Lincoln, Nebraska

December, 2010

Identification of holocarboxylase synthetase chromatin binding sites in human mammary cell lines using the DamID technology

Dipika Singh, M.S.

University of Nebraska, 2010

Advisers: Angela Pannier and Janos Zempleni

Holocarboxylase synthetase (HCS) is a chromatin protein that is essential for mediating the covalent binding of biotin to histones. Biotinylation of histones plays crucial roles in the repression of genes and repeats in the human genome. The objective of this project was to determine genomic binding sites of holocarboxylase synthetase (HCS) in the human chromatin of breast cell lines using an antibody independent technology. We tested the feasibility of DNA adenine methyltransferase identification (DamID) technology to map HCS binding sites in human mammary cell lines. Full-length HCS was fused to Dam for subsequent transfection into breast cancer (MCF-7) and normal breast (MCF-10A) cells. HCS docking sites in chromatin were identified by using the unique adenine methylation sites established by Dam in the fusion construct; docking sites were unambiguously identified using methylation sensitive digestion, cloning, and sequencing. Fifteen novel HCS binding sites were identified and included sequences within genes coding for inositol polyphosphate-5-phosphatase A, corticotropin hormone precursor, myosin heavy chain 9 non-muscle, cadherins 8 and 11, prolactin, leptin precursor, schwannomin 1, apoptosis inhibitor, netrin-G1 ligand, neuron navigator, SRY-box4, purinergic receptor, and CUB and sushi domains. We conclude that DamID is a useful technology to map HCS binding sites in human chromatin and propose that the

entire set of HCS binding sites could be mapped by combining DamID with microarray technology.

Acknowledgements

Dr. Pannier and Dr. Zempleni, thank you for giving me the opportunity to work in your labs. I am really appreciative of the support you both provided during the last two years. Toshi, and Samudra, thank you for all your help and suggestions throughout my graduate studies. Sarah, thank you for your friendship and I really appreciate all your help. Mary, Tadas, and Tim, thank you for being great friends and co-workers. Bo, I really appreciate all your help with my thesis, so thank you. Connie, and Donna, thank you for all your help with reserving meeting rooms and paperwork.

List of Abbreviations

AMP	Adenosine monophosphate
ANOVA	Analysis of variance
ART	Assisted reproductive technology
AS	Angelman syndrome
BCA	Bicinchoninic acid; protein assay
Bio	Biotinylated
BLAST	Basic Local Alignment Search Tool
BWS	Beckwith-Wiedemann syndrome
cDNA	Complimentary DNA
ChIP	Chromatin immunoprecipitation; assay
CpG	Cytosine – phosphate - Guanine
CRH	Corticotropin releasing hormone
CYP26A1	Cytochrome P450, family 26, subfamily A, polypeptide 1
Dam	DNA adenine methyltransferase
DamID	DNA adenine methyltransferase identification
DMEM/F12	Dulbecco's modified eagle medium with nutrient mixture F-12
DNA	Deoxyribonucleic acid
Ecoli	Escherichia coli
EGFP	Enhanced green fluorescent protein

gDNA	Genomic DNA
H; such as H3	Histone
HAT	Histone acetyltransferase
HCS	Holocarboxylase synthetase
HDAC	Histone deacetylases
INPP5A	Inositol polyphosphate-5-phosphatase A
K; such as K9	Indicates lysine residue
KCNAB1	Potassium voltage-gated channel, shaker-related subfamily, beta member 1
LB	Lysogeny broth
LF	Lipofectamine
Luc	Luciferase
MCF-10A	Normal breast cells
MCF-7	Breast cancer cells
MeCP2	Methyl CpG binding protein
MEGM	Mammary epithelial growth media
MYH9	Homo sapiens myosin, heavy chain 9; protein
NAV3	Neuron navigator 3
Opti-MEM	reduced serum cell media used for complex formation in transfection
PCR	Polymerase chain reaction
qRT-PCR	Quantitative Reverse Transcription PCR
RASSF1A	Rat sarcoma association domain family 1A;

	gene
RLU	Relative Light Units
RNA	Ribonucleic acid
RRS1	Ribosome biogenesis regulatory protein
SCHIP1	Schwannomin interacting protein 1
SNCA	Synuclein, alpha (non A4 component of amyloid precursor)
SRY	Sex determining region Y

Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

The objective of this project was to monitor native binding sites of holocarboxylase synthetase [HCS], a chromatin protein, in human breast cell lines using an antibody independent technology. This project utilizes the DNA adenine methyltransferase identification [DamID] technology developed by Dr. Bas van Steensel from Netherland's Cancer Research Institute [1]. Studies conducted in this project verify the DamID technology in breast cancer and human breast cell lines. The project eliminates the use of ChIP assays, a commonly used technology to study DNA-protein interactions and epigenetic mechanisms. ChIP assays depend heavily on specificity of antibodies and large cell numbers and hence cannot be used to study DNA-protein interactions in small populations of cells, such as oocytes [2, 3]. The verification of DamID technology in human breast tissues confirms the principle behind the technology. Future work with this technology could enable the study of chromatin proteins and aberrant epigenetic marks in oocytes that are artificially cultured for a period of 2-3 days prior to *in-vitro* fertilization.

1.2 Background

1.2.1 Epigenetics

Epigenetics is defined as 'the study of heritable changes in gene expression that

occur independent of changes in the primary DNA sequence' [4]. Most of these heritable changes are established during differentiation and are stably maintained through multiple cycles of cell division, enabling cells to have distinct identities while containing the same genetic information. This heritability of gene expression patterns is mediated by epigenetic modifications, which include methylation of cytosine bases in DNA, and posttranslational modifications, such as methylation, acetylation, ubiquitylation, phosphorylation, and biotinylation, of histone proteins [5, 6]. The complement of these modifications, collectively referred to as the epigenome, provides a mechanism for cellular diversity by regulating what genetic information can be accessed by cellular machinery. Failure of the proper maintenance of heritable epigenetic marks can result in inappropriate activation or inhibition of various signaling pathways and lead to disease states such as cancer [4, 7].

DNA methylation is perhaps the most extensively studied epigenetic modification in mammalian cell lines [8, 9]. It provides a stable gene silencing mechanism that plays an important role in regulating gene expression and chromatin architecture, in association with histone modifications and other chromatin associated proteins. In mammals, DNA methylation primarily occurs by the covalent modification of cytosine residues in CpG dinucleotides [8]. CpG islands are preferentially located at the 5' end of genes and occupy 60% of human gene promoters [8]. Methylation of these gene promoters is generally not detected in normal tissues but hypermethylation of CpG islands resulting in a loss of gene function is a common feature of almost every type of cancer. CpG gene promoter is among the most frequently methylated sites in human breast cancer and its methylation leads to silencing of RASSF1A and several other genes [10]. A screen of

paired invasive ductal carcinomas with adjacent normal tissues and additional breast samples obtained from healthy individuals identified several loci that are preferentially methylated in tumor versus normal tissue; a careful analysis of three such genes [CYP26A1, KCNAB1, and SNCA] suggests that their methylation in the primary tumors can be associated with methylation in the adjacent normal tissues [10].

In addition to epigenetic modification of the DNA, the histone N-terminal tails are crucial to maintaining chromatin stability and they are subject to numerous epigenetic modifications. Most modifications have some role to play in transcriptional regulation and so each has the potential to be oncogenic if deregulated deposition leads, for example, to loss of expression of a tumour suppressor gene [6, 11]. Histone acetyltransferase [HATs], enzyme responsible for histone acetylation, tend to be transcriptional activators whereas histone deacetylases [HDACs] tend to be transcriptional repressors. Many HAT genes are altered in some way in a variety of cancers [6, 11]. For instance, the p300 HAT gene is mutated in gastrointestinal tumours [6]. Similarly, methylation of histone proteins has also been associated with various types of cancer, including lymphoma and gastrointestinal cancer [12]. Lysine methyltransferases that target histone N-terminal tails contain a so-called SET domain. This domain possesses lysine methyltransferase activity and numerous SET domain-containing proteins are implicated in cancer [12, 13]. One example is the Suv39 family of enzymes that catalyse methylation of H3K9. Transgenic mice devoid of these enzymes are susceptible to cancer, especially B cell lymphomas [13]. Aurora kinases that perform phosphorylation of histone H3 at serine residues 10, and 28 during mitosis, a crucial stage of cell cycle, has been linked to pancreatic cancer [11, 13, 14].

Ubiqutination of histones H2A and H2B at aberrant amino acid site has been assocaited with lung cancer [15].

1.2.2 Holocarboxylase Synthetase and Biotinylation

A less investigated epigenetic modification of the histone proteins is biotinylation. There is increasing evidence that histones are covalently modified by the vitamin biotin, mediated by holocarboxylase synthetase [HCS] and biotinidase [16, 17, 18, 19, 20]. Subsequent studies have revealed that HCS is much more important than biotinidase for biotinylation of histones *in vivo* [19, 20, 21, 22], with biotinidase demonstrating histone biotinyl transferase activity *in vitro* [20, 23, 24]. HCS functions as a biotin-ligase for carboxylases and histones [25]. HCS catalyzes biotinylation of carboxylases and histones in a two-step ATPdependent reaction [25]. In the first step, biotinyl 5'-AMP [B-AMP] is generated; in the second step, biotin is transferred from B-AMP to the lysine residue in the apocarboxylases and histones [25]. Several histone biotinylation sites have been identified using recombinant histone biotin ligases, including HCS and synthetic histone tail peptides. The identified biotinylation sites are lysine (K)-9, K13, K125, K127, and K129 in histone H2A [26]; K4, K9, K18, and perhaps K23 in histone H3 [20, 24]; and K8 and K12, and perhaps K5 and K16 in histone H4 [22, 23, 24]. The important role of HCS and subsequent biotinylation of histones has been confirmed in various studies. For example, phenotypes of HCS knockdown include decreased life span and heat survival in *Drosophila melanogaster* [19], and increased transcriptional activity of retrotransposons in metazoans, predisposing test organisms to chromosomal instability [27]. Biotinylation of histones has also been shown to play a role in processes such as cell proliferation, gene silencing, and the cellular response to UV-induced DNA damage [28, 29].

HCS has a dispersed and distinct distribution in chromatin; however its chromatin binding sites are poorly defined [19]. Therefore, the objective of this study was to generate first insights into HCS binding sites in human chromatin of breast cancer and normal breast epithelial cell lines. It was hypothesized that HCS binding sites of the two cell lines would be largely similar, as both cell lines are epithelial, with a few variations due to the genomic abnormality of the cancer cell line. The investigation was carried out using an antibody independent technology as antibody dependent assays demand specficity and high cell numbers. Methods that are currently used to study epigenetic marks will be described in the following section.

1.2.3 Chromatin Immunoprecipitation Assay

The chromatin immunoprecipitation [ChIP] assay is a powerful tool for identifying proteins, including histone proteins and non-histone proteins, associated with specific regions of the genome by using antibodies that specifically recognize the protein/s under investigation. The most common approach taken by researchers to study epigenetic mechanisms is ChIP assays combined with DNA arrays [7]. ChIP assays together with DNA arrays have been demonstrated to successfully map epigenetic marks in the entire genome of various species. This procedure is also an excellent tool to map the enzymes that mediate epigenetic modifications of DNA and DNA-binding proteins in chromatin. Briefly, the following steps are commonly involed with ChIP/DNA array method **(Fig. 1)**: [1] DNA and proteins in chromatin are cross-linked using formaldehyde, and chromatin is sheared to produce fragments of 500–1500 base pairs in size; [2] chromatin is immunoprecipitated using an antibody that binds the epigenetic mark of interest; the supernatant does not contain the mark of interest and is discarded; [3] DNA is purified

from the precipitated chromatin; [4] DNA is chemically labeled [e.g., using fluorophores] and is hybridized to DNA microarrays; and [5] the epigenetic mark of interest is mapped in the genome, based on its relative enrichment by antibody precipitation [2]. The information obtained from the DNA microarray map indicates the relative enrichment of a given epigenetic mark in a distinct region of the chromatin. This information can then be correlated to transcriptional activity of genes, such as gene upregulation or downregulation, and gene silencing, residing in that region.

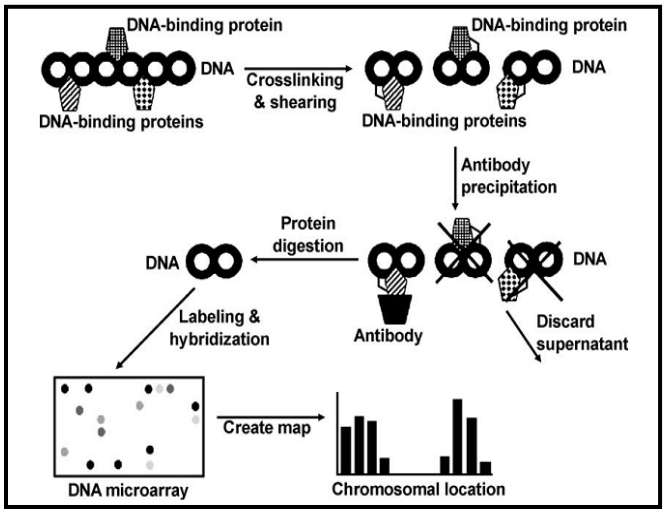

Figure 1: ChIP assay/DNA microarray Methodology [2]

The combination of ChIP assays and DNA microarrays holds great promise for the identification of epigenetic modifications and protein involved, but there are also various challenges associated with this technology that are discussed below. A major requirement of ChIP assays are antibodies. For each protein or epigenetic mechanism under investigation there is a need for an antibody specific for the protein or epigenetic mark of interest. Several common antibodies are commercially available but problem

arises when the antibodies lack specificity or are not commercially available. In both situations antibodies have to be synthesized in the lab, which can take anywhere from two to four months for polyclonal antibodies, and 8-12 months for monoclonal antibodies. Another drawback with ChIP assays is the requirement for large cell numbers [in the 10^6–10^7 range], which limits the applicability of this assay to rare cell samples, such as small stem cell batches, oocytes and tissue biopsies that are not available in large numbers [3]. Aside from the need for specific antibodies and large cell numbers, there is an ethical issue associated with synthesizing antibodies as they are raised in animals that are sacrificed at the end of the synthesis process. Finally, a challenge remains with the lack of appropriate DNA microarrays. The vast majorities of DNA microarrays contain probes for coding regions in genomes but do not contain probes for noncoding regions. Noncoding regions contain the majority of regulatory sequences of genes and, hence, are the prime target for many epigenetic modifications [7].

Therefore, to overcome the problems associated with the ChIP/DNA array technique, an antibody-independent technology, specifically DNA adenine methyltransfearse identification [DamID] technology, was used in this study to monitor HCS in human breast cell lines. This technology would also enable the study of chromatin proteins in small cell batches as it is independent of cell number.

1.2.4 DNA Adenine Methyltransfearse Identification [DamID] Technology

Prior to the development of the DamID technique, it had been demonstrated that a DNA cytosine methyltransferase can be targeted in vitro to a pre-determined specific DNA sequence by tethering it to a DNA-binding protein, such as zinc finger protein [30]. Therefore, it was assessed whether a similar approach could be used to target a

Escherichia coli DNA adenine methyltransfearse **[**Dam] to a specific DNA locus in vivo in *Drosophila melanogaster* [31]. Dam methylates the N6-position of adenine in the sequence GATC, which occurs on average every 200–300bp in the fly genome. Dam was chosen over other methyltransferases essentially because endogenous methylation of adenines is absent in the DNA of most eukaryotes. When expressed in mammalian cells, yeast, and *Drosophila*, Dam has been shown to have no harmful effects on cell viability or development [31], in contrast to certain cytosine methyltransferases [31]. DamID technology is based on the creation of a fusion protein consisting of *Escherichia coli* DNA adenine methyltransferase [Dam] and a chromatin protein or transcription factor of interest. Upon expression of the fusion protein in cultured cells or in an intact organism such as *Drosophila*, Dam is targeted to the native binding sites of the associated chromatin protein resulting in local methylation of adenine residues [1, 31]. Hence, the sequences near a binding site of the protein are marked with a unique methylation tag, which can be detected using Southern blot-, PCR- and microarray based assays that take advantage of restriction enzymes that are methylation sensitive [1].

With the development of the DamID technology, it became clear that the entire human genome could be analyzed for specific DNA-protein interactions and epigenetic mechanisms using DamID technology in combination with cDNA or oligonucleotide arrays. DamID technique has a number of advantages over other methods that are currently used to identify target sequences. First, it eliminates the use of ChIP assays that call for large cell numbers and heavily depend on the availability and quality of antibodies. Furthermore, it allows for the detection of protein-DNA interactions as they occur in living cells unlike ChIP assays that require lysis of cells to obtain chromatin.

ChIP assays also need crosslinking agents that may induce alterations in the chromatin structure under investigation leading to identification of unspecific sequences [1, 31]. The DamID procedure minimizes the detection of non-target sequences as it has been shown that the technique works best when the Dam fusion protein is expressed at very low levels, making it unlikely that the fusion protein itself interferes with the functions of the endogenous protein or its targets.

This study is a proof of principle where feasibility of the DamID technology was demonstrated in human mammary epithelial cell lines. For this project, Dam was fused with HCS to map the native binding sites of HCS in human breast and breast cancer cell genomes in an attempt to identify similarities and differences in the HCS binding regions between the two cell lines. HCS-Dam fusion plasmid was expressed in the two cell lines using non-viral gene delivery methods.

1.2.5 Non-Viral Gene Delivery

Gene delivery is the introduction of exogenous DNA or genes into cells for genetic modification. There are several methods to deliver DNA into cells, including viral and nonviral techniques [32]**.** Initial research conducted on gene delivery primarily focused on using viral carriers, including both retroviruses and adenoviruses, for gene delivery. Viral vectors exhibited high efficiency at delivering both DNA and RNA to numerous cell lines [33]. However, fundamental problems associated with viral vector systems, including toxicity, immunogenicity, and limitations with respect to scale-up procedures, encouraged the investigation of other potential systems, such as non-viral gene delivery, to transfer foreign genes into targeted tissue [32].

Nonviral vector systems, including cationic lipids, polymers, dendrimers, and peptides, all offer potential ways for compacting DNA for delivery to cells [34]. However, unlike viral analogues that have evolved means to overcome cellular barriers and immune defense mechanisms, nonviral gene carriers consistently exhibit significantly reduced transfection efficiency as they are hindered by numerous extra- and intracellular obstacles. However, biocompatibility and potential for large-scale production make these compounds increasingly attractive for gene delivery [34]. The most common and successful nonviral gene delivery method is complexation with a delivery agent, where negatively charged plasmid DNA is condensed through electrostatic interactions with cationic lipids to form complexes termed lipoplexes [33, 34] or cationic polymers to form polyplexes [35, 36]. Complexation provides the DNA protection from degradation by nucleases and other serum components [32] and enhances cellular uptake by reducing the effective size of DNA and promoting interactions between positively charged DNA complexes and the negatively charged cellular membrane [32]. These complexation agents can also facilitate intracellular trafficking, while dissociating from the DNA to allow expression [35, 36]. There are two methods in which lipoplexes or polyplexes are delivered to cells, including the bolus delivery approach and substrate-mediated gene delivery approach. In the conventional bolus delivery approach, cells are first plated on a tissue culture substrate and allowed to adhere for a period of 18-20 hours; complexes are then added and allowed to diffuse through the bulk media to the cells. In the substrate-mediated gene delivery approach, complexes are first immoblized on a tissue culture substrate or biomaterial and cells are then seeded on top of the immoblized complexes.

The latter approach enhances gene transfer as it elevates DNA concentration within cells local microenvironment [].37

Although the substrate- mediated approach results in enhanced gene transfer, bolus delivery is commonly used to transfect cells in order to determine transfection efficiencies. Bolus delivery can report on the ability of a cell line to take up foreign DNA, and also advises on the need for the substrate-mediated gene delivery approach. If a cell line is easily transfected using the bolus delivery approach there is usually no need for transfecting cells via substrate-mediated approach. Substrate mediated approach is commonly used in studies where higher transfection [>40%] is required or ideal.

For this project, the bolus delivery method was first used to deliver a pEGFP-Luc reporter gene plasmid complexed with commercially available cationic lipid formations, Lipofectamine 2000 and FugeneHD, to human breast and breast cancer cells. The DNA to lipid ratio was optimized, to determine the optimum transfection condition, and transfection efficiencies were determined. The substrate-mediated approach was not utilized for the purposes of this project as bolus delivery resulted in transfection efficiencies suitable for the working of Dam fusion plasmids.

1.3 Description of Studies

The purpose of this project was to establish the ability of DamID technology, an antibody independent technique, to monitor chromatin protein binding sites in mammalian cell lines. This research was conducted in an attempt to eliminate the use of ChIP assays in studying DNA-protein interactions in mammalian cell lines. HCS, a nuclear protein known to bind to chromosomes [19], was used as a model protein for this project. DamID technology was utilized to map the native binding sites of HCS in breast cancer and normal breast cells. HCS and HCS-Dam plasmids were cloned, breast cells were then transfected with these plasmids using nonviral delivery techniques, genomic DNA [gDNA] was isolated, genomic sequences methylated by Dam were digested and cloned into the pBlueScript II sk(+) cloning vector, which were then further sequenced. As the cloning approach was utilized for screening the binding sites, only a few sites were revealed. Future studies should utilize microarray technology, such as cDNA microarrays, to allow for genome wide analysis of HCS binding sites. Chapter 2 elaborates on purpose, materials/methods, and results of this project. Chapter 3 highlights the conclusion and future of this project.

References

[1] B.v. Steensel, DamID. Available from: FAQ.http://research.nki.nl/vansteensellab/DamID%20info/FAQ.htm

[2] A.M. Oommen, J.B. Griffin, G. Sarath, and J. Zempleni, Roles for nutrients in epigenetic events. J Nutr Biochem 16 (2005) 74-7.

[3] J.A. Dahl, and P. Collas, MicroChIP--a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. Nucleic Acids Res 36 (2008) e15.

[4] G. Egger, G. Liang, A. Aparicio, and P.A. Jones, Epigenetics in human disease and prospects for epigenetic therapy. Nature 429 (2004) 457-63.

[5] J.S. Stanley, J.B. Griffin, and J. Zempleni, Biotinylation of histones in human cells. Effects of cell proliferation. Eur J Biochem 268 (2001) 5424-9.

[6] H. Santos-Rosa, and C. Caldas, Chromatin modifier enzymes, the histone code and cancer. Eur J Cancer 41 (2005) 2381-402.

[7] P.A. Jones, and S.B. Baylin, The fundamental role of epigenetic events in cancer. Nat Rev Genet 3 (2002) 415-28.

[8] S. Sharma, T.K. Kelly, and P.A. Jones, Epigenetics in cancer. Carcinogenesis 31 27-36.

[9] D. Takai, and P.A. Jones, Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99 (2002) 3740-5.

[10] K. Visvanathan, S. Sukumar, and N.E. Davidson, Epigenetic biomarkers and breast cancer: cause for optimism. Clin Cancer Res 12 (2006) 6591-3.

[11] K. Zhang, and S.Y. Dent, Histone modifying enzymes and cancer: going beyond histones. J Cell Biochem 96 (2005) 1137-48.

[12] A.H. Lund, and M. van Lohuizen, Epigenetics and cancer. Genes Dev 18 (2004) 2315-35.

[13] R. Schneider, A.J. Bannister, and T. Kouzarides, Unsafe SETs: histone lysine methyltransferases and cancer. Trends Biochem Sci 27 (2002) 396-402.

[14] P.S. Espino, S. Pritchard, H.H. Heng, and J.R. Davie, Genomic instability and histone H3 phosphorylation induction by the Ras-mitogen activated protein kinase pathway in pancreatic cancer cells. Int J Cancer 124 (2009) 562-7.

[15] S. Lipkowitz, The role of the ubiquitination-proteasome pathway in breast cancer: ubiquitin mediated degradation of growth factor receptors in the pathogenesis and treatment of cancer. Breast Cancer Res 5 (2003) 8-15.

[16] J. Hymes, K. Fleischhauer, and B. Wolf, Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency. Biochem Mol Med 56 (1995) 76-83.

[17] Y.I. Hassan, H. Moriyama, L.J. Olsen, X. Bi, and J. Zempleni, N- and C-terminal domains in human holocarboxylase synthetase participate in substrate recognition. Mol Genet Metab 96 (2009) 183-8.

[18] J. Hymes, K. Fleischhauer, and B. Wolf, Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency. Biochem Mol Med 56 (1995) 76-83.

[19] G. Camporeale, E. Giordano, R. Rendina, J. Zempleni, and J.C. Eissenberg, Drosophila melanogaster holocarboxylase synthetase is a chromosomal protein required for normal histone biotinylation, gene transcription patterns, lifespan, and heat tolerance. J Nutr 136 (2006) 2735-42.

[20] K. Kobza, G. Camporeale, B. Rueckert, A. Kueh, J.B. Griffin, G. Sarath, and J. Zempleni, K4, K9 and K18 in human histone H3 are targets for biotinylation by biotinidase. FEBS J 272 (2005) 4249-59.

[21] Y.C. Chew, J.T. West, S.J. Kratzer, A.M. Ilvarsonn, J.C. Eissenberg, B.J. Dave, D. Klinkebiel, J.K. Christman, and J. Zempleni, Biotinylation of histones represses transposable elements in human and mouse cells and cell lines and in Drosophila melanogaster. J Nutr 138 (2008) 2316-22.

[22] K.A. Kobza, K. Chaiseeda, G. Sarath, J.M. Takacs, and J. Zempleni, Biotinyl-methyl 4-(amidomethyl)benzoate is a competitive inhibitor of human biotinidase. J Nutr Biochem 19 (2008) 826-32.

[23] G. Camporeale, E.E. Shubert, G. Sarath, R. Cerny, and J. Zempleni, K8 and K12 are biotinylated in human histone H4. Eur J Biochem 271 (2004) 2257-63.

[24] Y.C. Chew, A.S. Raza, G. Sarath, J. Zempleni, Biotinylation of K8 and K12 co-occurs with acetylation and mono-methylation in human histone H4. J FASEB 20 (2006) A610

[25] M.A. Narang, R. Dumas, L.M. Ayer, and R.A. Gravel, Reduced histone biotinylation in multiple carboxylase deficiency patients: a nuclear role for holocarboxylase synthetase. Hum Mol Genet 13 (2004) 15-23.

[26] Y.C. Chew, G. Camporeale, N. Kothapalli, G. Sarath, and J. Zempleni, Lysine residues in N-terminal and C-terminal regions of human histone H2A are targets for biotinylation by biotinidase. J Nutr Biochem 17 (2006) 225-33.

[27] N. Kothapalli, G. Sarath, and J. Zempleni, Biotinylation of K12 in histone H4 decreases in response to DNA double-strand breaks in human JAr choriocarcinoma cells. J Nutr 135 (2005) 2337-42.

[28] Y.C. Chew, G. Sarath, and J. Zempleni, An avidin-based assay for histone debiotinylase activity in human cell nuclei. J Nutr Biochem 18 (2007) 475-81.

[29] N. Kothapalli, G. Sarath, and J. Zempleni, Biotinylation of K12 in histone H4 decreases in response to DNA double-strand breaks in human JAr choriocarcinoma cells. J Nutr 135 (2005) 2337-42.

[30] G.L. Xu, and T.H. Bestor, Cytosine methylation targetted to pre-determined sequences. Nat Genet 17 (1997) 376-8.

[31] B. van Steensel, and S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18 (2000) 424-8.

[32] M.A. Mintzer, and E.E. Simanek, Nonviral vectors for gene delivery. Chem Rev 109 (2009) 259-302.

[33] W.F. Anderson, Human gene therapy. Nature 392 (1998) 25-30.

[34] P. Behr, Gene transfer with synthetic cationic amphiphiles: prospects for gene therapy. Bioconjug Chem 5 (1994) 382-9.

[35] C.M. Varga, K. Hong, and D.A. Lauffenburger, Quantitative analysis of synthetic gene delivery vector design properties. Mol Ther 4 (2001) 438-46.

[36] T. Niidome, and L. Huang, Gene therapy progress and prospects: nonviral vectors. Gene Ther 9 (2002) 1647-52.

[37] A.K. Pannier, J.A. Wieland, and L.D. Shea, Surface polyethylene glycol enhances substrate-mediated gene delivery by nonspecifically immobilized complexes. Acta Biomater 4 (2008) 26-39.

Chapter 2

Identification of holocarboxylase synthetase chromatin binding sites in human mammary cell lines using the DamID technology

Abstract

Holocarboxylase synthetase (HCS) is a chromatin protein that is essential for mediating the covalent binding of biotin to histones. Biotinylation of histones plays crucial roles in the repression of genes and repeats in the human genome. We tested the feasibility of DNA adenine methyltransferase identification (DamID) technology to map HCS binding sites in human mammary cell lines. Full-length HCS was fused to Dam for subsequent transfection into breast cancer (MCF-7) and normal breast (MCF-10A) cells. HCS docking sites in chromatin were identified by using the unique adenine methylation sites established by Dam in the fusion construct; docking sites were unambiguously identified using methylation sensitive digestion, cloning, and sequencing. Fifteen novel HCS binding sites were identified and included sequences within genes coding for inositol polyphosphate-5-phosphatase A, corticotropin hormone precursor, myosin heavy chain 9 non-muscle, cadherins 8 and 11, prolactin, leptin precursor, schwannomin 1, apoptosis inhibitor, netrin-G1 ligand, neuron navigator, SRY-box4, purinergic receptor, and CUB and sushi domains. We conclude that DamID is a useful technology to map HCS binding sites in human chromatin and propose that the entire set of HCS binding sites could be mapped by combining DamID with microarray technology.

2.1 Introduction

Holocarboxylase synthetase (HCS) is a chromatin protein [1] that plays a pivotal role in catalyzing histone biotinylation [1, 2]. Biotinylation is unique as it serves as an essential metabolic co-factor (biotin) and also functions as a histone modification. Initially, it was believed that biotinylation of histones is catalyzed by biotinidase [3]. However, subsequent studies revealed that HCS is much more important than biotinidase for biotinylation of histones *in vivo* [1, 4, 5, 6, 7], with biotinidase demonstrating histone biotinyl transferase activity *in vitro* [4, 8, 9]. Histone biotinylation is a recently discovered epigenetic mark [10]. Several histone biotinylation sites have been identified using recombinant histone biotin ligases, including HCS and synthetic histone tail peptides. The identified biotinylation sites are lysine (K)-9, K13, K125, K127, and K129 in histone H2A [9]; K4, K9, K18, and perhaps K23 in histone H3 [4, 7]; and K8 and K12, and perhaps K5 and K16 in histone H4 [8, 11, 7]. *In vitro* biotinylation of K8, K12, and K16 in histone H4 has been confirmed by mass spectrometric analysis of mammalian histones [11]. The existence of biotinylated histones was recently questioned [12], but three independent laboratories [14, 15, 16], in addition to ours [2, 4, 7, 8, 10, 13], confirmed that biotinylation is a natural histone modification.

Histone biotinylation is a comparably rare event (<0.1% of histones are biotinylated) [10, 14], but the abundance of an epigenetic mark is no marker for its importance. For example, only ~3% of cytosines are methylated, but the role of DNA methylation in gene regulation is undisputed [17]. Likewise, serine-14 phosphorylation in histone H2B and histone poly(ADP-ribosylation) are detectable only after induction of apoptosis and major DNA damage, respectively, but the role of these epigenetic marks in

cell death is unambiguous [18, 19, 20]. Moreover, low abundance of an epigenetic mark in bulk histone extracts compared with its high enrichment at specific loci is consistent with regulatory function. Evidence suggests that about one out of three histone H4 molecules are biotinylated at K12 in telomeric repeats [21]. K12-biotinylated histone H4 (H4K12bio), K9-biotinylated H3 (H3K9bio), and K18-biotinylated H3 (H3K18bio) are enriched at transcriptionally repressed loci in mammalian genomes, suggesting a role for histone biotinylation in gene regulation [5, 6, 13, 22]. The important role of HCS and subsequent biotinylation of histones has been confirmed in various studies. For example, phenotypes of HCS knockdown include decreased life span and heat survival in *Drosophila melanogaster* [1], and increased transcriptional activity of retrotransposons in metazoans, predisposing test organisms to chromosomal instability [6].

HCS has a dispersed and distinct distribution in chromatin, however its chromatin binding sites are poorly defined [1]. Therefore, the objective of this study was to adopt DNA adenine methyl transferase identification (DamID) technology [23] to generate first insights into HCS binding sites in human chromatin, and to pave the way for more comprehensive mapping efforts in future studies. Briefly, DamID technology is based on fusing DNA adenine methyltransferase (Dam) from *Escherichia coli* to a chromatin protein or transcription factor of interest [23], e.g., HCS. Upon expression of the fusion protein in cultured cells or in an intact organism such as *Drosophila*, HCS guides Dam to HCS binding sites, where Dam creates a unique adenine methylation tag, which is absent in most eukaryotes. Adenine methylation sites can be identified by methylation sensitive sequencing procedures.

In this proof-of-concept study, DamID technology was used to map, for the first time, genomic HCS- binding sites in the human chromatin using mammary epithelial cell lines, human breast (MCF-10A) and breast cancer (MCF-7), as aberrant epigenetic marks have already been shown in breast cancer cells [24].

2.2 Materials and Methods

2.2.1 Cloning HCS-Dam fusion plasmid

The pIND (V5) EcoDam [23, 25] vector was obtained from Dr. Bas van Steensel's laboratory at the Netherlands Cancer Institute. pIND (V5) EcoDam codes for *E.coli* Dam, and contains a multiple cloning site upstream of the Dam open reading frame. An HCS-Dam fusion construct was generated as follows. Full-length human HCS was PCR-amplified using HCS-pET41a as a template [2], AccuPrime™ Pfx DNA polymerase SuperMix (Invitrogen, Carlsbad, CA) and the following primers: 5'-GTTCGAATTCATGGAAGATAGACTCCACATGG-3' (forward, EcoRI site underlined) and 5'-GTTTCTCGAGCCGCCGTTTGGGGAGGATGAGG-3' (reverse, XhoI site underlined). Following amplification, HCS and pIND (V5) EcoDam vector were digested with EcoRI and XhoI (Fermentas, Glen Burnie, MD), ligated using Fast-Link™ DNA Ligation Kit (Epicenter Biotechnologies, Madison, WI), and transformed into MAX Efficiency® DH5α™ Competent Cells (Invitrogen). The HCS-Dam fusion plasmid was sequenced (Eurofin MWG Operon, Huntsville, AL) to confirm its identity and was denoted "HCS-Dam." Two control plasmids were used in HCS mapping studies. (i) Plasmid pIND V5 EcoDam [23], codes for Dam only and was used to identify artifactual binding of Dam to chromatin that was not mediated by HCS. (ii) Plasmid

"HCS" codes for full-length human HCS and was used to identify artifactual adenine methylation in the absence of Dam. Plasmid HCS was generated as described above for HCS-Dam, but the following reverse primer was substituted for the original reverse primer: 5'GTTTTCTAGATTACCGCCGTTTGGGGAGGATGAGG-3' (reverse; XbaI site underlined). The stop codon in this reverse primer terminates translation after the HCS open reading frame.

2.2.2 Cell culture

MCF-7 and MCF-10A cells (ATCC, Manassas, VA) were used for all studies. MCF-7 cells were cultured in DMEM/F12 media (ATCC) with the following supplements (v/v): 1% L-glutamine (Invitrogen), 10% fetal bovine serum (Invitrogen), 1% antibiotic/antimycotic (Invitrogen), 1% nonessential amino acids (Invitrogen), and 0.06% bovine insulin (Sigma-Aldrich, St. Louis, MO). MCF-10A cells were cultured in mammary epithelial growth media (MEGM) + bullet kit (Lonza, Basel, Switzerland), 1% (v/v) fetal bovine serum, and 0.05% (v/v) cholera toxin (Sigma-Aldrich).

2.2.3 Transfection Optimization using Reporter gene pEGFP-Luc

Transfection was optimized using a plasmid containing reporter genes, pEGFP-Luc (Clonetech, Mountain View, CA) in MCF-7 and MCF10A cell lines. pEGFP-Luc encodes for a fusion of enhanced green fluorescent protein (EGFP) and luciferase (Luc) from the firefly *Photinus pyralis*. Briefly, 20,000 cells/well were seeded in a 48-well plate and cells were allowed to adhere for 18 hours. After 18 h, cells were transfected with pEGFP-LUC-transfection reagent complexes formulated in serum-free opti-MEM media (Invitrogen). For MCF-7 cells, pEGFP-Luc was complexed with Lipofectamine

(LF) 2000 (Invitrogen,) and the complexes were incubated for 20 min. For MCF10A cells, pEGFP-Luc was complexed with FugeneHD (Roche, Madison, WI) and the complexes were allowed to incubate for 15 minutes. After incubation, complexes were delivered to the two cell lines. Cells were assayed for EGFP expression using fluoroescence microscopy (Leica DMI 3000B, Bannockburn, IL) at 24 and 48 h post-transfection. At 48 h, cells were lysed and cell lysate was assayed for luciferase expression, using the Luciferase Assay System (Promega, Madison, WI) and a luminometer (Turner Biossytem, Sunnyvale, CA), which returned luciferase activity in RLUs (Relative Light Units). Luciferase activity was normalized to the total protein amount, in the cell lysate, determined with the BCA protein assay (Pierce, Rockford, IL) and transfection levels were reported as RLUs/mg of protein. Each transfection experiment was performed in duplicate.

DNA (pEGFP-Luc) amount and DNA to transfection reagent ratio was varied to determine an optimum transfection condition for both cell types. Conditions were optimized based on the following factors i) transfection efficiency (percentage of EGFP-positive cells assayed by fluorescence microscopy); ii) gene activity (RLUs/mg of protein); and iii) cell viability and morphology (analyzed through phase microscopy using Leica DMI 3000B).

2.2.4 Transfection of cells with HCS-Dam and control plasmids

For transfection with HCS-Dam, EcoDam and HCS plasmids, 3.5×10^6 cells were seeded in T-75 flasks and allowed to adhere for 18 h, after which time cells were transfected with one of the three plasmids as follows. DNA was complexed with Lipofectamine (LF) 2000 (Invitrogen) or FugeneHD (Roche) in serum-free opti-MEM

media (Invitrogen). For MCF-7 cells, 6.75 µg of each plasmid was complexed with 10.2 µl of LF 2000 in a 1:1.5 DNA to lipid ratio (µg of DNA to µl of LF2000). These complexes were allowed to incubate for 20 min and were then added to each designated flask. For MCF-10A cells, 8.25 µg of each plasmid was complexed with 20.625 µl of FugeneHD in a 1:2.5 DNA to Fugene ratio (µg of DNA to µl of FugeneHD). These complexes were incubated for 15 min and were then delivered to each designated flask. Forty-eight hours after transfection, cells were collected and lysed using protocols specific for isolation of RNA or genomic DNA.

2.2.5 Quantitative Real-time PCR (qRT-PCR)

Total RNA was collected using the RNeasy mini kit (Qiagen, Valencia, CA) and reverse transcribed using the ImProm-II Reverse Transcription System (Promega). The cDNA was used to confirm successful transfection of cells with plasmids HCS-Dam, EcoDam, and HCS, using the following primers and SYBR green (Qiagen) in qRT-PCR expression analysis: (i) HCS-Dam = 5'-GCAACTCCTTCGACATGCTGAGAAA-3' (forward, positioned in HCS region) and 5'-TCTTCATGCCGGTACGCGTAGAAT-3' (reverse, positioned in Dam region); (ii) Dam = 5'-TCTGGTTGAGCGCCATATTCCAGT-3' (forward) and 5'-TGTACAAAGCCAGCAGTTCGTCCA-3' (reverse); and (iii) HCS = 5'-ATGGAAGATAGACTCCACAT-3' (forward) and 5'-TGAGACCTGATCCTTAACTTCC-3' (reverse). Glyceraldehyde 3-phosphate dehydrogenase was used as the reference gene for qRT-PCR normalization, using primers 5'-TCCACTGGCGTCTTCACC-3' (forward) and 5'-GGCAGAGATGATGACCCTTT-

3’ (reverse) [6]. The cycle threshold values were used to calculate amplicon abundance [31].

2.2.6 Mapping HCS-binding sites of adenine-methylated Sequences

Genomic DNA was purified 48 h after transfection by using the DNeasy mini kit (Qiagen). DNA was digested with *DpnI* (Fermentas, Glen Burnie, MD), which is sensitive to adenine methylation. The digested fragments were purified using a PCR purification kit (Qiagen) and ligated into the pBlueScript II sk(+) vector after linearization with *BamHI* (Fermentas); the Fast-Link™ DNA Ligation Kit was used for ligation. Note that digestion with *DpnI* and *BamHI* produces compatible overlaps. The ligation mixture was transformed into MAX Efficiency® DH5α™ Competent Cells and the cells were plated on LB-Ampicillin plates. White colonies were randomly selected for sequencing by using the T7 promoter primer (Eurofins MWG Operon, Huntsville, AL).

2.2.7 Statistics

The GraphPad Prism 5.0 program (La Jolla, CA) was used to plot transfection optimization and qRT-PCR graphs. Data are presented as the mean ± SEM. Multiple comparisons were evaluated by one-way ANOVA, followed by post hoc Tukey test. Values were considered statistically significant if $P < 0.05$.

2.3. Results and Discussion

2.3.1 Transfection Optimization

Work in *Drosophila* [26] suggests that it is important to keep the expression level of Dam-fusion proteins very low to avoid saturating methylation levels [23]. The pIND (V5) EcoDam vector consists of an inducible promoter; however it has been shown that leaky expression of Dam fusion plasmids is sufficient for specific methylation of target sequences [26, 27]. Induction of the vector leads to high background methylation, which makes the identification of target sequences difficult [27]. Ideally a transfection efficiency of 30% or higher is needed for transient transfection with Dam fusion proteins [27], although good binding maps have also been obtained with lower efficiency transfections [27]. Transfection was hence optimized using pEGFP-Luc reporter plasmid in the two breast cell lines in order to determine the maximum transfection or expression level that could be achieved with adequate cell viability. It was found that 0.11 µg of pEGFP-Luc in a 1:1.5 DNA: LF2000 ratio was the optimum transfection condition for MCF-7 cells. This condition resulted in 16 ± 3.5 % transfection efficiency **(Fig. 2A)** and a luciferase activity of 3×10^9 RLUs/mg of protein **(Fig. 2B).** Similarly, for MCF-10A cells, 0.11 µg of pEGFP-Luc using 1:2.5 DNA to FugeneHD ratio was determined to be the optimum condition. This condition resulted in 29 ± 2.4% transfection efficiency **(Fig. 3A)** and a luciferase activity of 4×10^9 RLUs/mg of protein **(Fig. 3B)**. The optimal conditions were then used for the DamID experiments. Although, transfection efficiency was less than the desired 30% for MCF-7 cells, expression of Dam plasmids were confirmed, *in vitro*, using qRT-PCR (results discussed below), prior to the conduction of the DamID assay.

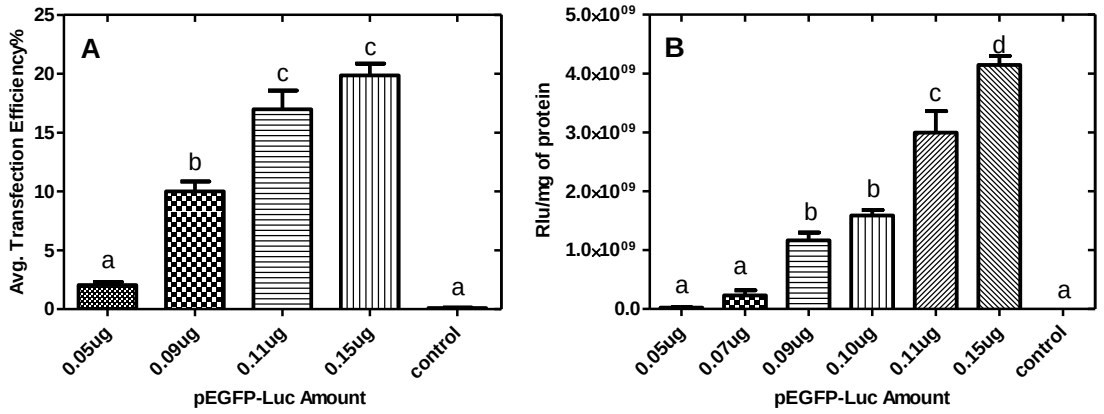

Figure 2. Complex conditions (DNA to transfection reagent ratio and DNA amount) used for transfection in **MCF-7** cells used for DamID assays were selected based on optimization of transfection conditions in which ratio and DNA amounts were compared to transfection efficiency (percentage of EGFP-positive cells) (A) and luciferase gene activity (B). This figure represents an optimum ratio of 1:1.5 DNA: LF 2000 at different pEGFP-Luc concentrations. Each bar represents the mean ± SEM, n = 2. [a,b,c, d]Bars with different letters indicate significant differences ($P < 0.05$) between treatments.

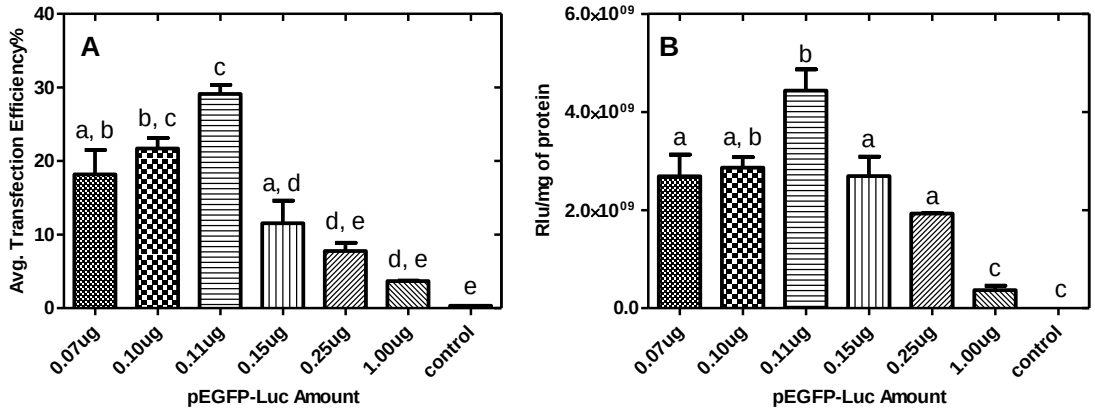

Figure 3. Complex conditions (DNA to transfection reagent ratio and DNA amount) used for transfection in **MCF-10A** cells used for DamID assays were selected based on optimization of transfection conditions in which ratio and DNA amounts were compared to transfection efficiency (percentage of EGFP-positive cells) (A) and luciferase gene activity (B). This figure represents an optimum ratio of 1:2.5 DNA: Fugene at different pEGFP-Luc concentrations. Each bar represents the mean ± SEM, n = 2. [a,b,c]Bars with different letters indicate significant differences ($P < 0.05$) between treatments.

2.3.2 Plasmid expression

Transfection of mammary cells with plasmids Dam-HCS, EcoDam, or HCS produced the expected expression patterns compared with non-transfected control cells. When cDNA was analyzed using PCR primers for HCS, the transcript abundance was 170-fold and 62-fold greater in MCF-7 cells transfected with HCS and HCS-Dam, respectively, than in non-transfected control cells (**Fig. 4A**); the abundance of HCS transcript was not altered by transfection with EcoDam (negative control) as it did not contain HCS. When cDNA was analyzed using PCR primers for the HCS-Dam fusion protein, the transcript abundance was 2000-fold greater in MCF-7 cells transfected with HCS-Dam than in non-transfected control cells (**Fig. 4B**); the abundance of HCS-Dam fusion transcript was not altered by transfection with EcoDam and HCS (negative controls) as these controls lacked HCS-Dam fusion. When cDNA was analyzed using PCR primers for Dam, the transcript abundance was 660-fold and 800-fold greater in MCF-7 cells transfected with EcoDam and HCS-Dam, respectively, than in non-transfected control cells (**Fig. 4C**); the abundance of Dam transcript was not altered by transfection with HCS (negative control) as it did not contain Dam. Results were comparable in MCF-10A cells (**Fig. 5**).

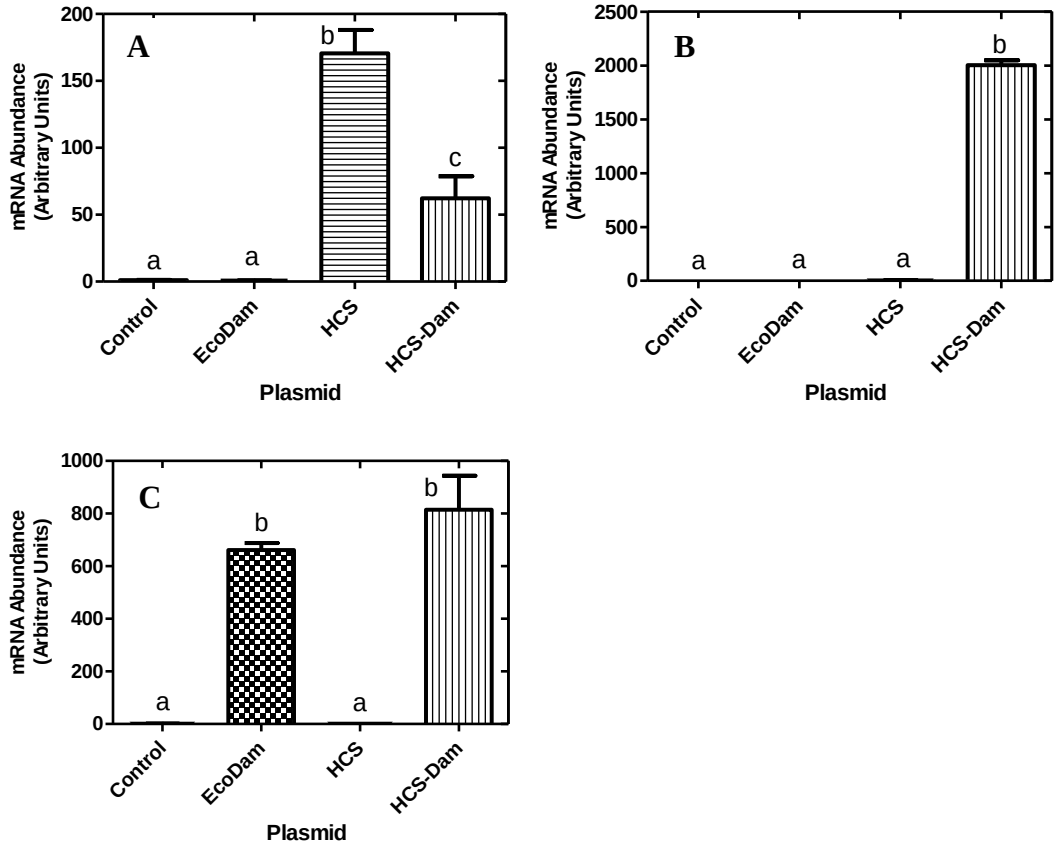

Figure 4. Transcript abundance in MCF-7 cells transfected with plasmids HCS-Dam, EcoDam and HCS. Transcripts were quantified by qRT-PCR using gene-specific primers for HCS (A), HCS-Dam fusion (B), and EcoDam (C). Each bar represents the mean $\pm$ SEM, n = 3. ^{a,b,c}Bars with different letters indicate significant differences ($P < 0.05$) between treatments.

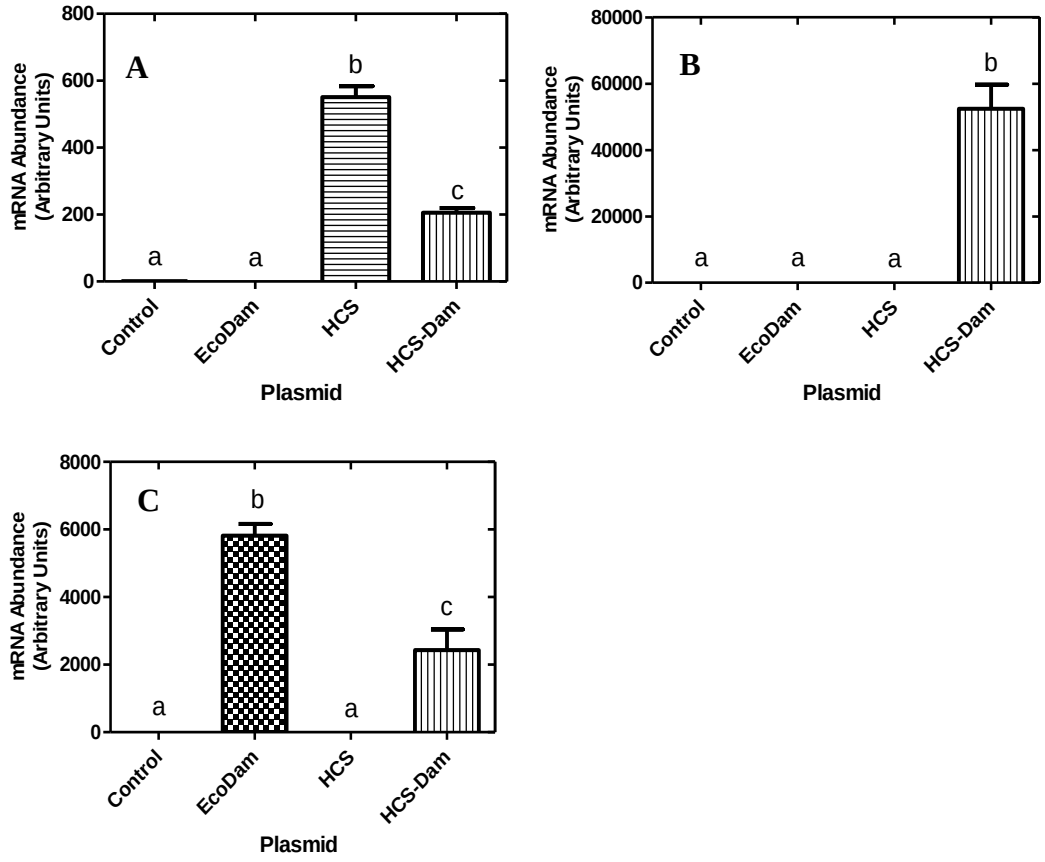

Figure 5. Transcript abundance in MCF-10A cells transfected with plasmids HCS-Dam, EcoDam and HCS. Transcripts were quantified by qRT-PCR using gene-specific primers for HCS (A), HCS-Dam fusion (B), and EcoDam (C). Each bar represents the mean $\pm$ SEM, n = 3. [a,b,c]Bars with different letters indicate significant differences ($P < 0.05$) between treatments.

2.3.3 HCS-docking sites

Following the validation of expression of fusion and control plasmids, these plasmids were expressed in MCF-7 and MCF-10A cell lines, genomic sequences methylated by Dam were cloned into the pBlueScript II sk(+) cloning vector, positive clones were obtained and sequenced. When five clones from HCS-Dam positive MCF-7 cells were annotated by NCBI's basic local alignment search tool (BLAST), seven unique HCS-binding sequences, within specific genes, were detected (**Table 1**). Each clone returned more than one sequence, because of the complimentary ends created during digestion with *DpnI* and the random ligation of these fragments during ligation and cloning. Transfection with plasmid HCS (negative control) did not produce any clones in MCF-7 cells. We conclude that HCS greatly enhances the binding of Dam at distinct genomic loci. Results in MCF-10A cells were comparable to those in MCF-7. Five clones from HCS-Dam positive MCF-10A cells detected eight unique sequences when annoted by BLAST, (**Table 2**).

Table 1

Genomic HCS binding sites in MCF-7 cells.

MCF-7 HCS-Binding Loci	**Genes**	**Chromosome**	**GenBank Acession Number**
5'-CCACAGAATCAGGGGATAACGCAGGAAAGAACAT- 3'	Inositol polyphosphate-5-phosphatase A	10	NM_005539.3
5'-CAGGAAAGAACATGTGAGCAAAAG- 3'	Corticotropin releasing hormone precursor (CRH), and Homolog of yeast ribosome biogenesis regulatory protein (RRS1)	8	NM_000756.2 NM_015169.3
5'-ATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA- 3'	Cadherin 8 and, Cadherin 11 Type2 preproproteins	16	NM_001796.2 NM_001797.2
5'-ACCATGCTCTTTCCCCCTTGCAACCTCACTCATTCTCTCTCCTATTC- 3'	Schwannomin interacting protein 1 (SCHIP1)	3	NM_014575.2
5'-GGTTTTTATCTTTGCAAGCAGCAGAT-3'	Neuron navigator 3 (NAV3)	12	NM_014903.4
5'-ATTCCACACAACATACGAGC-3'	Leptin precursor	7	NM_000230.2
5'-AAAGGGATTTTGGGGCATGAGATCATC- 3'	Apoptosis inhibitor 5 isoform b	11	NM_006595.3
5'-AAAGGGATTTTGGGGCATGAGATCATC- 3'	Netrin-G1 ligand	11	NM_020929.1

Table 2

Genomic HCS binding sites in MCF-10A cells.

MCF-10A HCS-Binding Loci	**Genes**	**Chromosome**	**GenBank Acession Number**
5'-CCACAGAATCAGGGGATAACGCAGGAAAGAACAT- 3'	Inositol polyphosphate-5-phosphatase A	10	NM_005539.3
5'-TCGCCCTATAGTGAGTCGTAT-3'	Homo sapiens myosin, heavy chain 9, non-muscle (MYH9)	22	NM_002473.4
5'-CAGGAAAGAACATGTGAGCAAAAG- 3'	Corticotropin releasing hormone precursor (CRH), and Homolog of yeast ribosome biogenesis regulatory protein (RRS1)	8	NM_000756.2 NM_015169.3
5'-TGAGCTAACTCACATTAATTGCGTTG- 3'	SRY (sex determining region Y)-box 4, and Prolactin	6	NM_003107.2 NM_001163558.1
5'-CCTGTTTTGCAGGGTGCTTT- 3'	Purinergic receptor P2X5 isoform A and B	17	NM_002561.2
5'-GAGCTCCAGCTTTTTTTCTCTTTAG- 3'	Tetratricopeptide repeat domain 7B	14	XM_510118.2
5'-CCTGGGGTGCCTAATGAGTG-3'	CUB and Sushi multiple domains 1	8	Q96PZ7-1
5'-ATTCCACACAACATACGAGC-3'	Leptin precursor	7	NM_000230.2

Please also note the following observations. First, when MCF-10A cells were transfected with plasmid HCS, one clone with one single sequence was identified compared with the seven clones in cells transfected with HCS-Dam, indicating that the false positive rate due to these artifacts might be up to 14% in Dam-based assays. Second, for the MCF-10A cell line, a locus within X-linked neuroligin4 gene was disregarded as a potential HCS docking site, as it appeared due to artefactual methylation by pIND (V5) EcoDam control. No such artifactual methylation by Dam was observed for MCF-7 cells. Third, four HCS-docking sites were found to be common between MCF-7 and MCF-10A cells, indicating a 28% overlap in the HCS binding regions identified in the two epithelial cell lines, further increasing confidence that the results produced by HCS-Dam technology are real. Note that clones were selected randomly in this proof-of-concept study and, therefore, one must not expect a 100% overlap in clones. Comprehensive maps can only be generated by using microarray-based technologies or high-throughput sequencing and may show an expected overlap of 100% between the identified genes, as both cell lines used in this study are of epithelial origin or alternatively could allow for determination of non-overlapping loci specific for either cell type. Also note that the following gene sets share a common HCS-docking locus: *CRH* and *RRS1* (Tables 1 and 2), cadherin 8 and cadherin 11 (Table 1) and, SRY-box4 and prolactin (Table 2).

The overlapping HCS binding regions were found within four genes: *INPP5A* on chromosome 10, and *CRH* and *RRS1* on chromosome 8, and leptin precursor on chromosome 7. INPP5A is a member of the INPP enzyme family [28, 29]. INPPs regulate diverse cellular processes such as protein trafficking, phagocytosis, and synaptic

vesicle recycling [29]. Loss of genetic material in the chromosome that includes INPP5A has been associated with brain tumors and leukemia [30]. It has also been demonstrated that loss or lower expression of INPP5 is an early mechanism in the development of cutaneous squamous cell carcinoma [30]. Similarly, CRH is a polypeptide hormone and is involved in stress response [31]. CRH has been implicated in various stress-induced abnormalities, including changes in both human and rat gastrointestinal functions such as motility and gastric acid secretion [31]. CRH gene has been shown to be highly expressed in MCF-7 cells [32], thus providing the basis to postulate a mechanism according to which CRH is secreted by tumor cells and exerts local – paracrine or autocrine – inhibitory action over breast cell growth [32]. Also, on chromsome 8, a locus of ribosome biogenesis regulatory protein (RRS1) was determined to be HCS-docking site in the two breast cell lines. RRS1 has been shown to regulate ribosome biosynthesis pathway in growing cells [33]. Leptin precursor gene, on chromosome 7, is known to be an integral component of a physiological signalling system that regulates fuel stores and energy balance in specific areas of the brain [34].The finding that *INPP5A, CRH, RRS1,* and leptin precursor genes possess potential HCS binding sites in human breast tissues underlines their importance in the interactions with the chromatin that further lead to biotinylation of histone proteins.

This study proves feasibility of the DamID technology to identify, for the first time, potential HCS binding sites in human mammary epithelial cell lines. DamID technology eliminates the need for large cell numbers and antibodies and thus represents a viable alternative method to the chromatin immunoprecipitation (ChIP) assay, a frequently used tool to investigate DNA-chromatin interactions. With the establishment

of the abilty of DamID technology to identify chromatin protein binding sites, the future goal of this project is to use the technology along with cDNA microarray sequencing to monitor chromatin proteins in small cell samples, such as oocytes. While the combination of techniques employed in this study only provided a snapshot of events occurring at the genomic level, it proves the DamID technology's utility and when combined with microarray sequencing, has the potential for full genome analysis.

References

[1] G. Camporeale, E. Giordano, R. Rendina, J. Zempleni, and J.C. Eissenberg, Drosophila melanogaster holocarboxylase synthetase is a chromosomal protein required for normal histone biotinylation, gene transcription patterns, lifespan, and heat tolerance. J Nutr 136 (2006) 273542.

[2] B. Bao, V. Pestinger, Y.I. Hassan, G.E. Borgstahl, C. Kolar, and J. Zempleni, Holocarboxylase synthetase is a chromatin protein and interacts directly with histone H3 to mediate biotinylation of K9 and K18. J Nutr Biochem

[3] J. Hymes, K. Fleischhauer, and B. Wolf, Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency. Biochem Mol Med 56 (1995) 76-83.

[4] K. Kobza, G. Camporeale, B. Rueckert, A. Kueh, J.B. Griffin, G. Sarath, and J. Zempleni, K4, K9 and K18 in human histone H3 are targets for biotinylation by biotinidase. FEBS J 272 (2005) 4249-59.

[5] M. Gralla, G. Camporeale, and J. Zempleni, Holocarboxylase synthetase regulates expression of biotin transporters by chromatin remodeling events at the SMVT locus. J Nutr Biochem 19 (2008) 400-8.

[6] Y.C. Chew, J.T. West, S.J. Kratzer, A.M. Ilvarsonn, J.C. Eissenberg, B.J. Dave, D. Klinkebiel, J.K. Christman, and J. Zempleni, Biotinylation of histones represses transposable elements in human and mouse cells and cell lines and in Drosophila melanogaster. J Nutr 138 (2008) 231622.

[7] K.A. Kobza, K. Chaiseeda, G. Sarath, J.M. Takacs, and J. Zempleni, Biotinyl-methyl 4 (amidomethyl) benzoate is a competitive inhibitor of human biotinidase. J Nutr Biochem 19(2008) 826-32.

[8] G. Camporeale, E.E. Shubert, G. Sarath, R. Cerny, and J. Zempleni, K8 and K12 are biotinylated in human histone H4. Eur J Biochem 271 (2004) 2257-63.

[9] Y.C. Chew, G. Camporeale, N. Kothapalli, G. Sarath, and J. Zempleni, Lysine residues in N terminal and C-terminal regions of human histone H2A are targets for biotinylation by biotinidase. J Nutr Biochem 17 (2006) 225-33.

[10] J.S. Stanley, J.B. Griffin, and J. Zempleni, Biotinylation of histones in human cells. Effects of cell proliferation. Eur J Biochem 268 (2001) 5424-9.

[11] Y.C. Chew, A.S. Raza, G. Sarath, J. Zempleni, Biotinylation of K8 and K12 co-occurs with acetylation and mono-methylation in human histone H4. J FASEB 20 (2006) A610

[12] S. Healy, B. Perez-Cadahia, D. Jia, M.K. McDonald, J.R. Davie, and R.A. Gravel, Biotin is not a natural histone modification. Biochim Biophys Acta 1789 (2009) 719-33

[13] V. Pestinger, S.S. Wijeratne, R. Rodriguez-Melendez, and J. Zempleni, Novel histone biotinylation marks are enriched in repeat regions and participate in repression of transcriptionally competent genes. J Nutr Biochem.

[14] L.M. Bailey, R.A. Ivanov, J.C. Wallace, and S.W. Polyak, Artifactual detection of biotin on histones by streptavidin. Anal Biochem 373 (2008) 71-7.

[15] R. Takechi, A. Taniguchi, S. Ebara, T. Fukui, and T. Watanabe, Biotin deficiency affects the proliferation of human embryonic palatal mesenchymal cells in culture. J Nutr 138 (2008) 680-4.

[16] S. Ghosh. Physiology, regulation, and pathogenesis of nitrogen metabolism in opportunistic fungal pathogen *Candida albicans*

[17] D. Li, L. Da, H. Tang, T. Li, and M. Zhao, CpG methylation plays a vital role in determining tissue- and cell-specific expression of the human cell-death-inducing DFF45-like effector A gene through the regulation of Sp1/Sp3 binding. Nucleic Acids Res 36 (2008) 330-41.

[18] G. Graziani, G. Ferrandina, G. Pozzoli, M. Vergati, A. Muzi, F. Legge, L. Tentori, G. Scambia, and P. Navarra, Corticotropin-releasing hormone receptor-1 in human endometrial cancer. Oncol Rep 15 (2006) 375-9.

[19] L.M. Bailey, R.A. Ivanov, J.C. Wallace, and S.W. Polyak, Artifactual detection of biotin on histones by streptavidin. Anal Biochem 373 (2008) 71-7.

[20] R. Takechi, A. Taniguchi, S. Ebara, T. Fukui, and T. Watanabe, Biotin deficiency affects the proliferation of human embryonic palatal mesenchymal cells in culture. J Nutr 138 (2008) 680-4.

[21] S.S. Wijeratne, G. Camporeale, and J. Zempleni, K12-biotinylated histone H4 is enriched in telomeric repeats from human lung IMR-90 fibroblasts. J Nutr Biochem 21 310-6.

[22] G. Camporeale, A.M. Oommen, J.B. Griffin, G. Sarath, and J. Zempleni, K12 biotinylation histone H4 marks heterochromatin in human lymphoblastoma cells. J Nutr Biochem 18 (2007) 760-8.

[23] B. van Steensel, and S. Henikoff, (van Steensel and Henikoff 2000). Nat Biotechnol 18 (2000) 4248.

[24] N. Dumont, Y.G. Crawford, M. Sigaroudinia, S.S. Nagrani, M.B. Wilson, G.C. Buehring, G. Turashvili, S. Aparicio, M.L. Gauthier, C.A. Fordyce, K.M. McDermott, and T.D. Tlsty, Human mammary cancer progression model recapitulates methylation events associated with breast premalignancy. Breast Cancer Res 11 (2009) R87.

[25] B.v. Steensel, DamID Plasmid. http://research.nki.nl/vansteensellab/damid.htm.

[26] C. Moorman, L.V. Sun, J. Wang, E. de Wit, W. Talhout, L.D. Ward, F. Greil, X.J. Lu, K.P. White, H.J. Bussemaker, and B. van Steensel, Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc Natl Acad Sci U S A 103 (2006) 12027-32.

[27] B.v. Steensel, DamID. Available from: FAQ.http://research.nki.nl/vansteensellab/DamID%20info/FAQ.htm

[28] M.V. Astle, K.A. Horan, L.M. Ooms, and C.A. Mitchell, The inositol polyphosphate 5-phosphatases: traffic controllers, waistline watchers and tumour suppressors? Biochem Soc Symp (2007) 161-81.

[29] L.M. Ooms, K.A. Horan, P. Rahman, G. Seaton, R. Gurung, D.S. Kethesparan, and C.A. Mitchell, The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. Biochem J 419 (2009) 29-49.

[30] A. Sekulic, S.Y. Kim, G. Hostetter, S. Savage, J.G. Einspahr, A. Prasad, P. Sagerman, C. Curiel-Lewandrowski, R. Krouse, G.T. Bowden, J. Warneke, D.S. Alberts, M.R. Pittelkow, D. DiCaudo, B.J. Nickoloff, J.M. Trent, and M. Bittner, Loss of inositol polyphosphate 5-phosphatase is an early event in development of cutaneous squamous cell carcinoma. Cancer Prev Res (Phila) 3 1277-83.

[31] J. Santos, P.R. Saunders, N.P. Hanssen, P.C. Yang, D. Yates, J.A. Groot, and M.H. Perdue, Corticotropin-releasing hormone mimics stress-induced colonic epithelial pathophysiology in the rat. Am J Physiol 277 (1999) G391-9.

[32] G. Graziani, G. Ferrandina, G. Pozzoli, M. Vergati, A. Muzi, F. Legge, L. Tentori, G. Scambia, and P. Navarra, Corticotropin-releasing hormone receptor-1 in human endometrial cancer. Oncol Rep 15 (2006) 375-9.

[33] A. Tsuno, K. Miyoshi, R. Tsujii, T. Miyakawa, and K. Mizuta, RRS1, a conserved essential gene, encodes a novel regulatory protein required for ribosome biogenesis in Saccharomyces cerevisiae. Mol Cell Biol 20 (2000) 2066-74.

[34] B. Morash, A. Li, P.R. Murphy, M. Wilkinson, and E. Ur, Leptin gene expression in the brain and pituitary gland. Endocrinology 140 (1999) 5995-8.

Chapter 3

Conclusions and Future Goals

3.1 Introduction

Studies conducted in this project verified the DNA adenine methyl transferase identification (DamID) technology in human breast (MCF-10A) and breast cancer (MCF-7) cell lines (Chapter 2). The technology was successfully used to map the binding sites of holocarboxylase synthetase (HCS) chromatin protein in MCF-7 and MCF-10A cells. Fifteen novel HCS binding sites were identified and included sequences within genes coding for inositol polyphosphate-5-phosphatase A, corticotropin hormone precursor, myosin heavy chain 9 non-muscle, cadherins 8 and 11, prolactin, and several others. While it was shown in this study that DamID technology can be employed to investigate chromatin proteins in human breast tissues, the future goal of this project is to study chromatin proteins and epigenetic marks in oocytes that are artificially cultured for the purpose of assisted reproductive technology (ART). DamID was established in this project because more traditional and commonly used chromatin immunoprecipitation (ChIP) assays require large cell numbers (typically in the 10^6–10^7 range) and highly specific antibodies [1, 2]. ChIP assays, therefore, cannot be utilized for studying chromatin in oocytes as they are hard to obtain in large batches. In addition, DamID technology has several other advantages over ChIP assays. It eliminates the need for target-specific antibodies making the investigation of any chromatin protein, such as DNA methyltransferases and histone acetyltransferases, relatively easy. Unlike ChIP assays, DamID technology also allows for the detection of protein-DNA interactions as they occur in living cells and does not require cell lysis to acquire the chromatin [3]. As

opposed to the DamID technology, ChIP assays also require crosslinking agents that can induce alterations in the chromatin structure under investigation leading to the identification of unspecific-target sequences [3].

3.2 Assisted Reproductive Technology (ART) and Birth Defects

According to 1992 Fertility Clinic Success Rate and Certification Act, ART is defined as fertility treatments in which both eggs or oocytes and sperms are handled in the laboratory [4]. Although the use of ART has become a widely accepted and implemented therapy for some forms of infertility, there have been concerns about the long-term safety of removing and handling germline cells [4]. Studies conducted between the years 1991-1998 demonstrate a correlation between ART and low birth weight (LBW) and multiple births [5, 6, 7]. Since low birth weight is in part epigenetically controlled, an epigenetic connection seems plausible for birth defects in ART population [4].

Several human disorders involving birth defects have now been shown to be caused by epigenetic alterations. The cardinal features of Beckwith-Wiedemann syndrome (BWS), which is linked to a cluster of imprinted genes on chromosme 11p15.5, are prenatal overgrowth, abdominal wall defects, neonatal hypoglycemia, and macroglossia [4]. Prader-Willi syndrome, which is linked to a cluster of imprinted genes on chromosome 15p11-13 normally expressed from the paternal allele, is characterized by muscular hypotonia, obesity, mental retardation, and hypogonadotrophic hypogonadism. Another neurological birth defect syndrome, Angelman syndrome (AS), is also linked to chromosome 15p11-13; in this case, to a gene normally expressed from

the maternal allele. The characteristic features of AS include severe developmental delay, absent speech, seizures, ataxia, hyperreflexia, and hypotonia [4]. Genetic mutations of the epigenetic machinery also cause birth defects. Rett syndrome is caused by mutations in the X-linked gene encoding MeCP2, a methyl-CpG-binding protein [4].

The cause of the link between ARTs and epigenetic alterations is currently unknown. It could be due to some aspect of the ARTs techqniue. There is a wide range of different ARTs, such as in vitro fertilization, which are now routinely used within clinics. Techniques utilized for ART, duration of artificial culture, and mechanical manipulation can expose one or both of the germ cells to an altered hormonal regime [8]. Any alterations to the normal environment of the oocyte or sperm could result in changes to some aspect of their imprinting mechanisms [8]. For example, studies in humans and mice suggest that superovulation can lead to the production of oocytes without their correct primary imprint, and aberrant DNA methylation [9]. It is unclear whether the cause is hormonal stimulation, gamete manipulation, or culture conditions [8]. It has been suggested that in vitro culture of gametes, zygotes, and embryos, representing a common feature of all ART techniques, results in an accumulation of epigenetic alterations leading to an enhancement in fetal growth and heavier newborn cattle, known as large offspring syndrome [8].

Evidences mentioned above suggest a link between artificial culture of oocytes, used in ART, and aberrant epigenetics and birth defects. Therefore the future goal of this project is to establish biotinylation marks in artificially manipulated mice oocytes, before proceeding on to human oocytes. Biotinylation is a recently discovered epigenetic mark and is mediated by HCS chromatin protein [10]. Genomic binding sites of HCS would be

mapped in artificially cultured oocytes, using the DamID technology. This work would lay the foundation for succeeding studies where the aberrancy and normalcy of biotinylation marks in cultured oocytes, prior to fertilization, would be investigated.

References

[1] A.M. Oommen, J.B. Griffin, G. Sarath, and J. Zempleni, Roles for nutrients in epigenetic events. J Nutr Biochem 16 (2005) 74-7.

[2] J.A. Dahl, and P. Collas, MicroChIP--a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. Nucleic Acids Res 36 (2008) e15.

[3] B. van Steensel, and S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18 (2000) 424-8.

[4] E.L. Niemitz, and A.P. Feinberg, Epigenetics and assisted reproductive technology: a call for investigation. Am J Hum Genet 74 (2004) 599-609.

[5] U.B. Wennerholm, P.O. Janson, M. Wennergren, and I. Kjellmer, Pregnancy complications and short-term follow-up of infants born after in vitro fertilization and embryo transfer (IVF/ET). Acta Obstet Gynecol Scand 70 (1991) 565-73. (IVF/ET). Acta Obstet Gynecol Scand 70:565–573

[6] P. Rufat, F. Olivennes, J. de Mouzon, M. Dehan, and R. Frydman, Task force report on the outcome of pregnancies and children conceived by in vitro fertilization (France: 1987 to 1989). Fertil Steril 61 (1994) 324-30.

[7] F. Olivennes, V. Kerbrat, P. Rufat, V. Blanchet, R. Fanchin, A. Hazout, M. Glissant, H. Fernandez, and R. Frydman, [Follow-up of a cohort of 422 children aged from 6 to 13 and conceived by in vitro fertilization]. Contracept Fertil Sex 25 (1997) XIII-XVIII.

[8] K. Biermann, and K. Steger, Epigenetics in male germ cells. J Androl 28 (2007) 466-80.

[9] Sato, E. Otsu, H. Negishi, T. Utsunomiya, and T. Arima, Aberrant DNA methylation of imprinted loci in superovulated oocytes. Hum Reprod 22 (2007) 26-35.

[10] J.S. Stanley, J.B. Griffin, and J. Zempleni, Biotinylation of histones in human cells. Effects of cell proliferation. Eur J Biochem 268 (2001) 5424-9.