A Comprehensive Analysis of The Transcriptome In Human Colorectal Cancer Cells Lacking DNA Methyltransferases

THESIS

Submitted in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

PAWAN KUMAR TIWARY

Under the supervision of
Dr. David Eric Symer MD PhD

Birla Institute of Technology and Sciences
Pilani (Rajasthan) India
2012

Completed at National Cancer Institute –Frederick, Frederick, Maryland, USA under BITS-NCI (NIH) Collaboration

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled ***"A Comprehensive Analysis Of The Transcriptome In Human Colorectal Cancer Cells Lacking DNA Methyltransferases"*** and submitted by ***Pawan Kumar Tiwary,*** ID No. **2001PH29094** for award of Ph.D. Degree of the Institute embodies my original work.

Pawan Kumar Tiwary
ID No.: 2001PH29094
Designation: Pre-doctoral fellow, BITS-Pilani
Manager, Prescient Life Sciences

Date:

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled "***A Comprehensive Analysis Of The Transcriptome In Human Colorectal Cancer Cells Lacking DNA Methyltransferases"*** and submitted by ***Pawan Kumar Tiwary*** ID No **2001PH29094** for award of Ph. D. Degree of the Institute, embodies original work done by him under my supervision.

David E Symer, M.D. Ph.D.
Principal Investigator
National Cancer Institute
Frederick MD 21702

Currently: Assistant Professor
Human Cancer Genetics Program
Departments of Molecular Virology, Immunology and Medical Genetics, Internal Medicine and (adjunct) Biomedical Informatics
The Ohio State University Comprehensive Cancer Center
Columbus, OH 43210

Date:

DEDICATIONS

I dedicate this thesis to my parents, Dr. Yogesh Tiwary and Dr. Laxmi Tiwary. They more than anyone else, have been the best role models I could have hoped for. Any accomplishment of mine is due to their tremendous support and encouragement.

&

To my grandfather Late Prof. Mahesh Tiwary

ACKNOWLEDGEMENTS

Finally…I see the light at the end of the tunnel…it is time to graduate. The experience of graduate education encompasses many people, friends, teachers and colleagues. I take this opportunity to recognize individuals whom I am indebted to in many ways for my successful graduation. I really appreciate all their help.

First and foremost, I wish to thank my mentor Dr. David Symer for giving me this coveted opportunity of working under the collaboration program and supervising my thesis work. This work would not have been possible without his able guidance and support. I thank David for the trenchant critiques, the probing questions, and the remarkable patience throughout. On the personal side, he helped me adjust to the new environment and gave right advice during downs in my personal life. Thank you for all those special moments.

I also would like to acknowledge the constant support and guidance of Dr. Brad St Criox and Dr. Shyam Sharan, Mouse Cancer Genetics Program (NCI Frederic). I thank Dr. Amar Klar for his extremely helpful insights regarding experimental planning, he always gave a new dimension to all the research discussions we had.

I thank Dr S Venkateswaran (Former Vice Chancellor), Dr. R Raghurama (Director, BITS, Pilani), Dr. S K Verma (Dean, Research and Consultancy Division), BITS, Pilani for giving me the opportunity and support to work for my PhD degree at NCI-Frederick. I express my gratitude towards Dr A K Das (Professor, BioSciences) for the guidance and support rendered to me by. I thank Dr Shibashis Chowdhary (Group Leader, Biosciences) for his guidance and administrative support. I express my sincerely thanks to Drs. Uma Dubey and Vishal Saxena for accepting to be in my

Dissertation Advisory Committee and for their guidance. I am thankful to Dr. Dinesh Deshmukh for coordinating this study with BITS; I appreciate his commitment towards the successful running of this program.

I am extremely thankful to Dr. Jingfeng Li, senior colleague and friend, for all his help during my PhD work including help in critical lab work, experimental planing, sharing of reagents and the collaborative experimental work. Dr. Keiko Akagi was very helpful in conducting crucial bioinformatic analysis; the collaborative work with Dr Akagi provided new insight into the study. I am also extremely thankful to my friend and fellow graduate student Manoj Kannan for his constant help and support in the lab; there is so much that I learnt from him both on professional and personal front. I thank my lab mate Dr Elizabeth Flynn for her scientific discussions and technical help whenever I required. I thank my lab colleagues – Amy Graham, Heath Bradley and Anna Trivett for their support in lab.

One meets lot of people as they progress but some friendships can actually predate it. I was surrounded by great friends, who constantly provided intellectual stimulation as well as, were just fun to hang out with, which is of utmost importance for graduate education. I thank my long term friends Rajesh, Sangeetha, Sudhakar, Srilaxmi, Anuj and Sai for all those special moments; I also thank Simanti, Dimple and Guhan for all their valuable advice. Manmohan, I have always enjoyed discussions with you, thanks for being such a great company.

And beyond friends, there is family. My grandfather late Dr Mahesh Tiwary, who was a renowned educationist and an erudite scholar, was the one who showed me the path to higher education and its importance; I will always remain indebt to him. My parents, Drs. Yogesh Tiwary and Laxmi Tiwary, need special mention for all the

love, support and sacrifices they made over the years. My dad, whom I have talked countless times about my lab work, has listened to me very patiently and inspired me to keep marching; my mom always reminded me the importance of good education. It is to them I owe my dissertation the most. I thank my uncle and aunt for their constant support and encouragement. My loving brother, Major Tarun has always inspired me in a way that no one else could; he could always draw a parallel between a war-zone and a research lab; my sister and bro-in-law has always kept me on my toes with their enquiries about my progress. I thank my wife, Dr Shikha for her unconditional love, and support throughout; she has been patient while I vented out my frustration due to work pressure. I also need to thank the family I inherited through my marriage; they have always rendered invaluable advice and support. My loving son Dhruv has brought me smiles and fun…thanks *beta* you kept me going during my ups and downs. I thank all my cousins and other members of my immediate and extended family for their constant support blessings.

TABLE OF CONTANTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ASP	:	**A**nti-**S**ense **P**romoter
DKO	:	**D**ouble **K**nock**o**ut
DNMT	:	**DN**A **M**ethyl **T**ransferase
GIS	:	**G**ene **I**dentification **S**ignature
HERV	:	**H**uman **E**ndogenous **R**etro**v**iruses
IAP	:	**I**ntra-**c**isternal **P**article - **A**
MHC	:	**M**ajor **H**istocompatibility **C**omplex
LINE	:	**L**ong **I**ntersperse **N**uclear **E**lements
LTR	:	**L**ong **T**erminal **R**epeats
MPSS	:	**M**assively **P**arallel **S**ignature **S**equencing
MT	:	**M**etallo**t**hionein
miRNA	:	**Mi**cro **RNA**
ORF	:	**O**pen **R**eading **F**rame
qRT-PCR	:	**Q**uantitative **R**everse **T**ranscriptase **P**olymerase **C**hain **R**eaction
RNAi	:	**RNA i**nterference
SAGE	:	**S**erial **A**nalysis of **G**ene **E**xpression
SAH	:	**S**-**a**denosyl **H**omocystenine
SAM	:	**S**-**a**denosyl **M**ethionine
SINE	:	**S**hort **I**nterspersed **N**uclear **E**lement
SVA	:	**S**INE-**V**NTR-**A**lu
TPM	:	**T**ag **P**er **M**illion
UTR	:	**U**n-**T**ranslated **R**egion

ABSTACT

DNA methylation is a stable but changeable epigenetic mark that is strongly associated with transcriptional repression in most eukaryotes. Abnormal patterns of both genome-wide hypomethylation and localized hypermethylation have been identified in virtually all human malignancies. In this study, we compared transcript levels in HCT116 human colorectal cancer cells and their double knockout (DKO) derivatives, which lack DNA methyltransferases *DNMT1* and *DNMT3b*. This analysis, predominantly using serial analysis of gene expression (SAGE) cDNA library tag counts, revealed a set of possible "signature genes" whose upregulation is inversely linked to their upstream promoters' genomic methylation. These include certain cancer testis genes, interferon-inducibile genes, major histocompatibility complex (MHC) genes, and members of the metallothionein gene cluster. However, repetitive elements including widespread endogenous retrotransposons are not strongly upregulated, despite their large template numbers and dense methylation in the human genome. Upregulated transcripts with the largest changes in SAGE tag frequencies were verified using two additional independent techniques, i.e. Northern blotting and quantitative polymerase chain reaction (qRT-PCR). In addition, the most highly upregulated genes were verified both by identification of similar expression levels in several independently derived DKO clones, and by comparable results from exon microarray assays. The results suggest that a distinctive, reproducible transcriptome signature that results from profound genome-wide hypomethylation can be identified and refined. In addition, these results have helped to elucidate several biologically important pathways whose deregulation by genome-wide hypomethylation may contribute to cancer formation.

Genetic changes contributing to cancer formation, such as point mutations, chromosomal rearrangements and other changes in DNA sequences have been known for a long time. Extensive studies have been conducted to elucidate the basis and impact of such alterations in various steps involved in carcinogenesis. By contrast, epigenetic contributions to carcinogenesis have been appreciated only recently. In the past three decades, research efforts have established numerous connections between various stages of carcinogenesis and epigenetic aberrations.

While epigenetic processes are heritable, they are not encoded in primary DNA sequences in the genome. They include transcriptional gene silencing and post-transcriptional gene silencing. Their molecular mechanisms, including DNA methylation, histone modifications (including methylation, acetylation and phosphorylation) and certain RNA-mediated events, appear to be conserved throughout most eukaryotes. Epigenetic controls of the mammalian genome play fundamentally important roles in regulating gene expression, genomic stability, differentiation and development. While changes occur normally in development, disrupted controls can lead to cancers and other diseases (Baylin and Bestor, 2002; Bestor, 1992; Dean et al., 2005). Two central questions persist in epigenetics: how are the intricate, nonrandom epigenetic marks established and maintained in the genome? And secondly, what are the consequences of changes in them, whether in normal development, in diseases, or in experimental models? In human cancers, both aberrant increases and decreases in DNA methylation at CpG dinucleotides located in different genomic regions have been reported to occur frequently (Figure 1). These involve genome-wide hypomethylation and focal hypermethylation (Smiraglia and Plass, 2002; Sharma et al., 2010; You et al., 2012). The latter often occurs aberrantly at CpG

islands. CpG islands are defined to be >200 nt long, to have >50% GC content, and to have an "observed/ expected" CpG ratio that is higher than expected at > 60%. They normally remain unmethylated in all tissue types throughout development (Antequera and Bird, 1993; Antequera et al., 1990). Aberrant hypermethylation at certain CpG islands has been associated with transcriptional gene silencing of tumor suppressor genes in cancer formation; such genes include *RB* (Greger et al., 1989), *VHL (Herman et al., 1994)*, *p16/INK4a* (Gonzalez-Zulueta et al., 1995; Merlo et al., 1995) and *MLH1 (Baylin et al., 2001)*). Conversely, certain genomic compartments normally are methylated in most somatic tissues, but can become hypomethylated in cancers. These include repetitive elements such as retrotransposons, the inactive X-chromosome in females, pericentromeric sequences, embryonic genes, and imprinted genes (Ehrlich, 2002; Sharma et al., 2010).

Given the deleterious potential consequences of retrotransposon movement, it is surprising to know that relatively a low number of mutations and other harmful effects actually happen due to their movement. Does this suggest that the genome has some kind of defense system that checks the movement of these elements?

Bestor et al proposed that cytosine methylation may serve as a host genome-defense system that helps check the expression of these elements (Bestor and Tycko, 1996; Meehan et al., 1992) by silencing the retrotransposons. L1s are generally silenced except in germ cells and during embryonic development (Branciforte and Martin, 1994; van den Hurk et al., 2007). It is well known that endogenous L1 and other repetitive elements are highly methylated in somatic cells, which is responsible for keeping these elements in a silent state (Bestor and Tycko, 1996; Meehan et al.,

1992). Any decrease in methylation at these transposable elements increases the risk of their transcription and movement in the genome.

Several studies have shown that decreased genomic methylation (either by genetic or pharmacologic manipulation) reactivates genes that are aberrantly silenced in cancer cells (Coral et al., 2002; Ghoshal et al., 2000; Gibbons et al., 2000; Gius et al., 2004; Guillaudeux et al., 1996; Jackson-Grusby et al., 2001; James et al., 2006; Koslowski et al., 2004; Liang et al., 2002; Majumder et al., 1999; Nie et al., 2001). However, only a few studies have looked into the effects of such epigenetic manipulation on normally methylated and silenced genes.

A recent study by Weber et al has shown that decreased genomic methylation of HCT116 cells, due either to pharmacologic or genetic manipulation, caused induction of non-canonical transcription from an intronic L1 antisense promoter (ASP) located in the proto-oncogene cMet, thus creating a fusion transcript L1-cMet. They showed that this L1-cMet transcript is associated with decreased expression of the cMet gene. These results illustrate an unexpected effect of genomic hypomethylation on the expression of genes, resulting in decreased transcript levels even when the *bona fide* promoter of the gene remains unmethylated (Weber et al., 2010).

Normal monoallelic expression of imprinted genes depends on their parent of origin, but this can be disrupted in cancers through loss of imprinting (Feinberg, 2005; Kaneda and Feinberg, 2005; Biliya et al., 2010). Hypomethylation appears to occur in the early stages of development of solid tumors (Goelz et al., 1985; Sharma et al., 2010; You et al., 2012) and can result in genomic instability due to increased

rearrangements and recombination events (Komarova et al., 2002; Sharma et al., 2010; Boland et al., 2010; Berdasco M et al., 2010).

DNA methylation is established in normal development by the essential, *de novo* DNA methyltransferases, *DNMT3a* and *DNMT3b* (Figure 2) (Okano et al., 1999; Okano et al., 1998). Both are expressed at various levels in adult somatic tissues, suggesting that they continue to play functional roles in differentiated cells (Okano et al., 1999). By contrast, *DNMT1* is the major maintenance methyltransferase (Bestor et al., 1988), but also has some *de novo* activity (Siedlecki and Zielenkiewicz, 2006). *DNMT1* is expressed ubiquitously, with high levels of expression in dividing cells. *Dnmt1* mouse knockouts show embryonic lethality (Bestor, 1992; Lei et al., 1996; Li et al., 1992), as they lose monoallelic expression of most imprinted genes (Antequera and Bird, 1993), show inactivation of the active X-chromosome due to reactivation of *Xist* (Panning and Jaenisch, 1996), express high levels of intra-cisternal particle A (IAP) retrotransposons (Walsh et al., 1998), and exhibit genomic instability in mouse embryonic stem (ES) cells (Chen et al., 1998). More recently, a new modification of cytosine has been identified – hydroxymethylcytosine; Tahiliani et al. demonstrated that the enzyme Tet1, an iron-dependent a-ketoglutarate dioxygenase, catalyzes the formation of 5hmC from 5meC (Tahiliani et al., 2009). Additionally, they suggest that the 5hmC may be an intermediate in the conversion of 5meC to cytosine, thus identifying an enzyme that can potentially be involved in demethylating DNA. These forms of cytosine methylation are the most newly discovered types and are likely to confound and confuse measurements of methylcytosine, and also they appear likely to play significant biological roles including active demethylation.

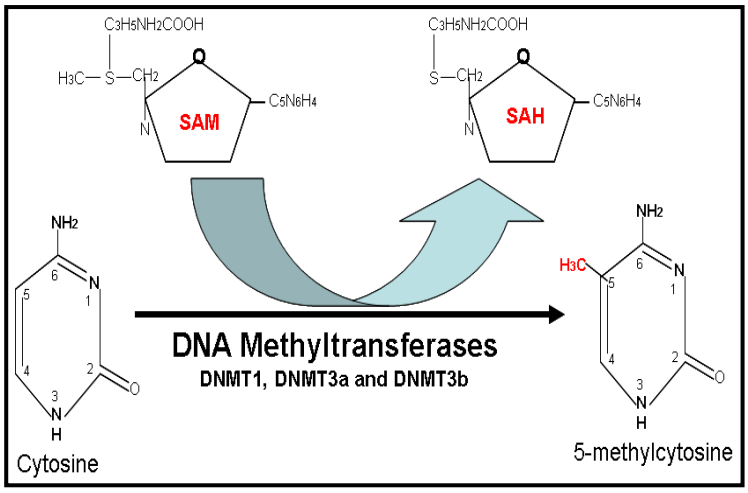

Figure 1: **Formation of 5-methylcytosine**

An active DNA methyltransferase enzyme catalyzes the transfer of a methyl group (CH3) from S-adenosylmethionine (SAM) to (deoxy) cytosine, producing 5-(deoxy) methylcytosine and S-adenosylhomocystine (SAH). Recently a high resolution crystal structure of DNMT1 gave additional new insights into its catalytic mechanism of cytosine methylation (Nie et al.) . Adopted from *Cell. Mol. Life Sci.* 59 (2002) 241-257.

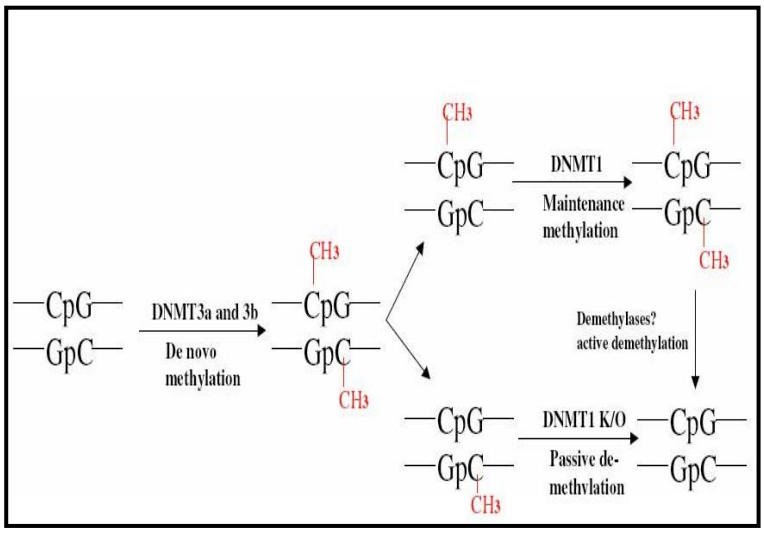

Figure 2: A schematic of the actions of DNA methyltransferases. Depicted here are the roles of two *de novo* DNA methyltransferases, Dnmt3a and Dnmt3b, and the maintenance DNA methyltransferase, Dnmt1, in establishing the DNA methylation patterns and maintaining them during cell division. Although these enzymes have been categorized into *de novo* and maintenance methyltransferases, they may have overlapping functions, e.g. Dnmt1 possesses *de novo* methyltransferase activity. More recently, some evidence suggests a link between rapid demethylation activity and Dnmt3a and Dnmt3b. Adopted from *Pediatric Research*, Vol. 59 (4) 2006.

To gain insights into whether disrupted DNA methylation could result in chromosomal instability (Komarova et al., 2002) or reactivate transcriptionally repressed genes in a nearly diploid, cultured human cancer cell line, Rhee et al. knocked out *DNMT1* in HCT116 cells by homologous recombination (Rhee et al., 2000). Resulting single knockout derivatives are surprisingly viable, dividing at a slightly slower rate than parental cells. Despite a 95% decrease in methyltransferase activity, they exhibit only approximately 20% reductions in genome-wide methylation, with extensive demethylation at pericentromeric satellites but nearly normal methylation persisting at *p16* and short interspersed elements (SINE retrotransposons).

By contrast, subsequent knockout of *DNMT3B* in *DNMT1* single knockout cells, resulting in double knockout (DKO) derivative cells, resulted in 95% reduction in genome-wide methylation and nearly complete abolishment of their DNA methyltransferase activity. DKO cells survive in culture; however, they grow at a significantly slower rate than wild-type cells. DKO cells show significant hypomethylation at repeat sequences such as satellite repeats and Alu elements, a loss of imprinting at the *Igf2* imprinted locus and reactivation of *p16* (Rhee et al., 2002). Establishment of DKO cells demonstrated that the original *DNMT1* knockout allele (Rhee et al., 2000) does indeed substantially affect normal expression and function of this DNA methyltransferase (Rhee et al., 2002).

Testing the original rationale for their derivation, Matsui et al. showed that DKO cells have a high degree of genomic instability; about 90% of cells at passage 25 have at least one novel chromosomal rearrangement (Karpf and Matsui, 2005). Egger et al. subsequently demonstrated that a truncated, hypomorphic DNMT1

protein residually is expressed in the knockout cells, due to previously undetected alternative splicing (Egger et al., 2006). Thus the *DNMT1* knockouts have increased hemi-methylation at specific CpG sites.

RNA interference (RNAi) directed against remaining *DNMT1* transcripts in the knockout cells caused further reductions in their genome-wide methylation and their viability, suggesting that *DNMT1* is essential for cell survival (Egger et al., 2006). More recently, a conditional knockout of all DNMT1 catalytic function has been developed in HCT116 cells, resulting in mitotic catastrophe (Chen et al., 2007). This result again verifies that DNMT1 is an essential cytosine methyltransferase.

In the present investigation, we tested the hypothesis that disruption of DNA methyltransferase genes would affect the transcription of genes, transposable elements and other transcribed elements in the genome. We used HCT116 WT human colorectal cancer cell and a derivative of these cells, DKO (DNMT1 and DNMT3b "double knockout" in HCT116) cells, to compare and study changes in the transcriptome.

Specific aims of the study

1. To study the effect of decreased methylation and methyltransferase activity on the transcriptome;
2. To study the effect of this epigenetic alteration on expression of LINE (L1), SINE and other repetitive elements;
3. To study the regional specificity (if any) of this epigenetic alteration and
4. To identify a transcriptional signature of a demethylated genome

Cell culture

The human colorectal cell line HCT116 and its derivatives, i.e. *DNMT1* single knockout clones 9A and 1C1, *DNMT3B* single knockout cells, and DKO clone 2 late passage cells (all generous gifts from Drs. Ina Rhee, Christopher Lengauer and Bert Vogelstein, Johns Hopkins) and other DKO clones 2 early passage, 3 and 8 (kindly provided by Dr. Bert Vogelstein (Johns Hopkins) and Kurtis Bachman (University of Maryland)) were cultured in McCoy's 5A modified medium with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Gibco) at 37°C, 5% CO_2 and saturated humidity at 70–90% confluency. Cells were passaged by trypsinization, washing by centrifugation, and re-plating in fresh tissue culture flasks.

RNA extraction

Total RNA was extracted using the modified single-step acid guanidinium thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi, 1987) using RNA-Bee extraction solution (Teltest Inc.). Polyadenylated transcripts were purified further as outlined below. RNA quality was evaluated by spectrophotometry and agarose gel electrophoresis and samples were stored at –80°C.

Generation of LongSAGE libraries

Three longSAGE libraries, corresponding to HCT116 (WT), single DNMT1 knockout clone 1C1, and double knockout clone 2 late passage (DKO2L) cells, were prepared using the I-SAGE Long Kit, version A (Invitrogen), following its original description and subsequent modifications (Figure 3) (Lal et al., 1999). Briefly, poly(A)$^+$ RNA was purified from total RNA (50μg) using oligo(dT) magnetic beads, and double stranded cDNA was synthesized using Superscript II reverse transcriptase and E. coli

DNA polymerase. This cDNA product was digested with anchoring enzyme NlaIII and divided into two fractions, each of which was ligated with unique adapters A and B using T4 DNA ligase.

LS Adapter A

5′TTTGGATTTGCTGGTGCAGTACAACTAGGCTTAATATCCGACATG 3′

3′amino(C7)CCTAAACGACCACGTCATGTTGATCCGAATTATAGGCT PO4 5′

LS Adapter B

5′TTTCTGCTCGAATTCAAGCTTCTAACGATGTACGTCCGACATG 3′

3′amino(C7)GACGAGCTTAAGTTCGAAGATTGCTACATGCAGGCT PO4 5′

The adapters contain a 3' recognition site for tagging enzyme MmeI and a PCR primer-binding site. After ligation, cDNAs were digested with MmeI, releasing ~ 60 nt tags, each of which contain ~20-21 nt of unique sequence from transcripts. The two tag fractions with adapters A and B were ligated in tail-to-tail fashion with T4 DNA ligase, yielding ditags of ~130 nt, which were PCR amplified, ethanol precipitated, electrophoresed on 12% polyacrylamide gels, and gel eluted using SNAP columns. Purified 130 nt ditags were digested with anchoring enzyme NlaIII to release 34 nt ditags, which were purified from LS adapters by 12% polyacrylamide gel electrophoresis (PAGE), eluted from the gel, and ethanol precipitated. These 34 nt ditags were then concatenated by ligation using T4 DNA ligase, separated on 8% PAGE gels, and appropriate size ranges (300bp – 2kb) were excised and cloned into linearized pZero-I vector (linearized with SphI) by blunt-end ligation using T4 DNA ligase. A small portion of the resulting size-fractionated longSAGE library was transformed into One Shot TOP10 electrocompetent bacteria (Invitrogen), grown on

LB agar plates with zeocin at 50 μg/mL at 37°C, and individual colonies were picked into 96 well plates with SOB liquid medium + Zeo (50 μg/mL) for overnight growth at room temperature.

One microliter of each bacterial culture was used in "colony PCR" reactions together with M13F (-20) and M13R primers to amplify the cloned concatamers. Products were examined on 1% agarose gels, and ~2 μL (~50 ng) of each PCR product was purified for sequencing using ExoSAP-IT (United States Biologicals). Approximately 25 ng of purified PCR products was used in cycle sequencing reactions using BigDye v.3.1 (Applied Biosystems, ABI). Sequencing reaction products were ethanol precipitated and resuspended in HiDi formamide (ABI). Conventional Sanger sequencing was performed using a 96 well, capillary electrophoresis sequencing machine (Spectrumedix, Transgenomics). More than 100,000 tags were sequenced for each of the WT and DKO2L libraries, in two rounds of about 50,000 tags each.

Analysis of longSAGE tag frequencies

Raw sequencing data were generated and analyzed using Base Spectrum software (Spectrumedix, Transgenomics), yielding sequence chromatograms. Only good quality sequence traces with Phred scores consistently > 20 were analyzed further. To parse and enumerate longSAGE tags, we analyzed concatamer sequences using SAGE-2000 v4.5 (software application was a kind gift of Dr. Ken Kinzler, Johns Hopkins). To allow calculation of ratios between tags, we arbitrarily assigned a raw tag count value of 0.5 to those tags that were absent from one library but were identified in the other. Then, we normalized raw tag numbers per library to tag counts per million. The abundance of each tag was compared in both libraries using

Microsoft Access. Tags were annotated by reference to the NCBI web portal (http://www.ncbi.nlm.nih.gov) and were assigned a corresponding UniGene ID and description. Tags with no match were regarded as potentially novel, i.e. previously unreported or alternately spliced transcripts.

Bioinformatic analyses

Statistical analyses were performed using the software package SPSS (SPSS Inc.). We used Database for Annotation, Visualization, and Integrated Discovery (DAVID: http://david.abcc.ncifcrf.gov) (Dennis et al., 2003) to categorize genes into various pathways, biological and molecular functions, cytobands and chromosomes to facilitate our understanding of biological meaning of the various findings from SAGE analysis. We used BRB-ArrayTools to normalize raw microarray expression data. All bioinformatic analyses were performed in collaboration with Dr. Keiko Akagi (previously Staff Scientist, Mouse Cancer Genetics Program, NCI-Frederick; now Research Assistant Professor, Human Cancer Genetics Program, The Ohio State University Comprehensive Cancer Center).

To compare data from SAGE and other expression-profiling platforms (van Ruissen et al., 2005), we first mapped both SAGE tags and array probes to NCBI Unigene clusters. If multiple SAGE tags identified the same cluster, we selected the 3' most tag as the representative tag. Next, based on the UniGene clusters, data were filtered for gene expression (i.e. we required tag counts > 0 and present calls on microarray platform). To match human cell-line data to mouse model data, the NCBI Homologene database was queried to obtain human/mouse homolog. Finally, to compare expression levels, an expression ratio between mutant (DKO2L) and wildtype (HCT116) transcript

levels was calculated and log-transformed for each platform and each transcript. Based on these ratios, the correspondence between platforms was estimated using Up/Down classifications and the Pearson correlation coefficient. The Pearson correlation is a measure of the linear association between two variables.

LongSAGE tags were predicted using consensus sequences for L1, Alu, HERV-K (accession no. NM_001007236) and SVA retrotransposon (accession no. AC016142) using DNA strider 1.4f3 (CEA, France), by identifying sequences in both orientations adjacent to NlaIII restriction enzyme sites.

Statistical analysis of HCT116 and DKO2L libraries

A total of more than 100,000 tags was sequenced for each library in two rounds of about 50,000 tags each (i.e. batches 1 and 2 of HCT116, and batches 1 and 2 of DKO2L). The purpose of sequencing the libraries in two rounds was to analyze the two halves of the libraries statistically and to check the random sampling of tags in the libraries. This approach would help to determine if there were any sequencing biases between the two halves of the library. Scatter plots were produced for two halves of each SAGE library to visualize the relationship between batches (Fig. 5). Pearson correlation coefficients were used to compare two batches of SAGE libraries.

To determine fold-changes of tag expression, the ratio of the DKO2L to HCT116 normalized tag counts was calculated. The Z-test was used to detect differential gene expression between the two longSAGE libraries. High-frequency tags were scored as more reliable (Kal et al., 1999). Differentially expressed tags with a p-value < 0.005 were considered significant. It has been demonstrated that the Z-test is as effective and efficient as other statistical methods including Fisher's Exact test

and a simulation-based approach (Ruijter et al., 2002). Statistical analyses were performed with the software package SPSS (SPSS Inc., Chicago, IL).

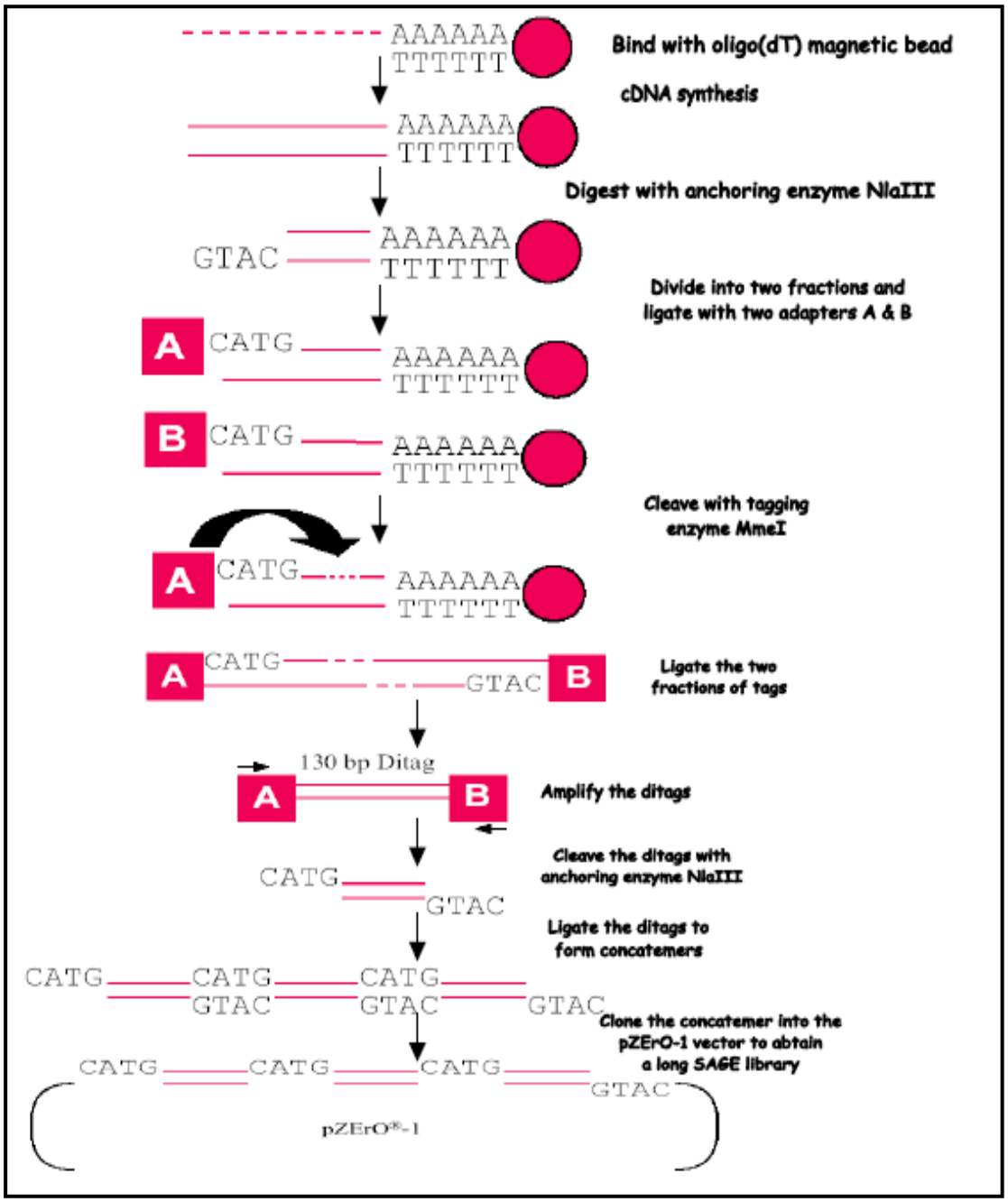

Figure 3: **Schematic of longSAGE technique, adopted from I-SAGE Long kit instruction manual (Invitrogen)**

Northern blots

Ten micrograms of total RNA were brought to constant volumes in loading buffer (MOPS, formaldehyde and formamide) and incubated at 55^0 C for 15 min. RNA samples were loaded on a 1% agarose denaturing gel and electrophoresed at 5 V/cm until the bromophenol blue dye migrated almost two-thirds the length of the gel. RNA bands were visualized on Kodak IS-2000 MM gel documentation system after staining with ethidium bromide. Electrophoresed RNA was hydrolyzed by soaking the gel in 10 gel volumes of 0.05 N NaOH for 20 minutes followed by neutralization by soaking in 10 gel volumes of 20X SSC for 30 minutes. Following this, a transfer stack was constructed by placing a nylon membrane (Hybond-N, Amersham Biosciences) on top of the gel, followed by Whatman 3MM paper of the same size as the gel. The RNA was allowed to transfer onto the nylon membrane by capillary action overnight. This was followed by UV cross-linking using a UV Stratalinker device (Stratagene).

The Northern membrane was prehybridized in prehybridizing mix (1:1 prehybridizing solution and formamide (50% v/v) with denatured salmon sperm DNA (10mg/ml)) for 3 hours at 42^0 C. Simultaneously, a probe was radiolabeled freshly. A double-stranded DNA probe was made by PCR amplification of the specific cDNA, either from commercially available, human fetal brain or testis cDNA libraries using appropriate primer sets (Table 1). This double stranded probe was then radiolabeled using α-^{32}P labeled dCTP (radioactive) using the Megaprime DNA (random hexamer) labeling system (Amersham Biosciences). To characterize incorporation of free nucleotides into the probe, 1µL of radioactive (hot) probe was used for thin layer chromatography, and the radioactivity in another 1µL was measured with a scintillation counter to calculate the percent incorporation of the hot dCTP in the probe.

After prehybridization of the nylon membrane, hot probe was added and was hybridized overnight at 42^0 C. Following the hybridization, the membrane was washed once each with 2XSSC/0.1%SDS, 1XSSC/0.1%SDS and 0.1XSSC/0.1%SDS at 65^0 C, and wrapped in Saran wrap. The membrane was exposed to X-ray film for appropriate times at -80^0 C using intensifying screens, and developed using a Kodak X-OMAT 200A developer. Nylon membranes were stripped by soaking in boiling 0.1% SDS, and then probed for GAPDH (as a loading control) and developed on X-ray film as above.

Table 1: List of primers to make Northern probes by PCR amplification from cDNAs.

Gene	Forward Primer	Reverse primer
IFI27	CTGCTCTCACCTCATCAG	GCTCCTC GCTGGGTC
GAGED2	ATTCTTTCTCCGCTACTGAG	GCTGTATCCTGATCTTCTTC
PRSS21	CACCCGTTACTTCGTATCG	CAAGGCATCAACTGGAATGTG
VIM	GTGAATACCAAGACCTGCTC	TTGTAGGAGTGTCGGTTGTTA
UCHL	CCTGAAGACAGAGCAAAATGC	CATCTGAAGCTCACACCAC
UBE2C	GTCTGGCGATAAAGGGATT	AACAAAACAAAAATACCACAGCTC
G1P3	GGCGGTATCGCTTTTCTTG	CCAAGAAGGAAGAAGAGGTTC
G1P2	GGCAACGAATTCCAGGTG	GCAGGCGCAGATTCATG
BST2	TCACCATCAAGGCCAACAG	CCATAACAACAGGCAGCAC

Quantitative PCR (q-PCR)

Double stranded cDNA synthesis was carried out using a SuperScript double-stranded cDNA synthesis kit (Invitrogen). First strand synthesis was carried out using an oligo(dT) primer (i.e. either DES887, which includes a 5' XhoI site, or DES1141, which includes a 5' M13 F site), followed by second strand synthesis. Synthesized cDNA was extracted with phenol/chloroform and salt precipitated. A small aliquot of the cDNA was run on a gel to check quality and the size range distribution. The cDNA was then quantitated and appropriate dilutions were made for further analyses.

Quantitative PCR (q-PCR) reactions were performed to confirm some of the most differentially expressed transcripts in DKO2L cells, using cDNAs synthesized from HCT116, DKO2L and other DKO clones (i.e. DKO2 (early passage), DKO3 and DKO8, kind gifts of Dr. Kurtis Bachmann, University of Maryland, and Dr. Bert Vogelstein, Johns Hopkins University). SYBR-green was used to quantitate the amplified cDNAs in real time, using BioRad SYBR Green supermix IQ containing PCR buffer, dNTPs and Taq polymerase. Appropriate primer sets (Table 2) and cDNA dilutions were added separately, and the reactions were performed in triplicate. *TSMB10* and *RPL13A* gene transcripts were chosen to be used as loading controls (internal standards) in qPCR analyses, since their expression levels remained unaffected in DKO2L and HCT116 cells. The primers for these two internal standards were previously described (Vandesompele et al., 2002). Primers for differentially up regulated genes were designed using the Operon primer design tool kit (https://www.operon.com/oligos/toolkit.php). The primer pairs were designed to span across spliced intron(s), to avoid amplification of contaminating genomic DNA if any (Table 2). These experiments were carried out using a BioRad iCycler machine. Threshold cycle numbers (Ct) were calculated to determine the relative initial amount of cDNA templates, which are anticipated to reflect transcript levels. In parallel, a 10-fold dilution series of a standard (amplified from a commercially available cDNA library or IMAGE clones using same primers) was made, and qPCR was done to construct a standard curve, calculate PCR efficiency and optimize experimental conditions.

Table 2: Primers used for qPCR experiments

Gene	Forward Primer	Reverse primer
TSMB10	AAGAAAATGGCAGACAAACCAG	TTCAGAGACCCACAGGC
RPL13A	CCTGGAGGAGAAGAGGAAAGAGA	TTGAGGACCTCTGTGTATTTGTCAA
IFI27	CTGCTCTCACCTCATCAG	GCTCCTC GCTGGGTC
GAGED2	ATTCTTTCTCCGCTACTGAG	GCTGTATCCTGATCTTCTTC
PRSS21	CACCCGTTACTTCGTATCG	CAAGGCATCAACTGGAATGTG
VIM	GTGAATACCAAGACCTGCTC	TTGTAGGAGTGTCGGTTGTTA
MT1A	CTTATAGCCTCTCAACTTCTTG	CAGTAAATGGGTCAGGGTTG
CTCFL	CAGGCCCTACAAGTGTAACGACTGCAA	GCATTCGTAAGGCTTCTCACCTGAGTG

Bisulfite sequencing

Genomic DNA was extracted using a DNeasy kit (QIAGEN) or by protease digestion and extraction in phenol/ chloroform followed by ethanol precipitation, and was quantified using an Agilent 8453 spectrophotometer. About 1.2 μg of genomic DNA from HCT116 and DKO2L cells each was bisulfite treated using optimized reagents from the MethylDetector bisulfite modification kit (Active Motif) (Figure 4). Each DNA sample was treated in the dark for 9 hrs. as per the manufacturer's protocol. Genomic DNA was incubated in the proprietary conversion buffer and hydroquinone for 9, 12, 15 or 18 hrs. to convert sequences with variable CpG content, followed by purification over a size exclusion column provided with the kit using desulfonation buffer, and finally was eluted using 50 μL of elution buffer. Treated DNA samples were stored at -20^0 C until analysis within 2 weeks.

Following conversion and purification, bisulfite modified DNA samples were tested for complete conversion of non-CpG cytosines by PCR amplification of the *p16* promoter region using a nested primer set provided with the kit as a

positive control. Amplification of the predicted 190 nt product suggested optimal bisulfite conversion, indicating that the converted DNA sample could be used reliably in downstream applications.

A schematic of bisulfite PCR and sequencing is presented in Figure 4. We used 5 μL of converted DNA eluate in PCR reactions to assay methylation status at promoter regions of interest, using primers that we designed (Table 3) with MethylPrimer Express software (ABI) or from published studies (Esteller, 2002). Resulting PCR products, amplified from bisulfite-treated genomic DNA samples, were gel eluted and cloned into PCR2.1 TOPO TA vector (Invitrogen). Briefly, 4 μL of gel eluted PCR product was incubated with 1μL of TOPA TA vector and 1μL of provided salt solution for 5 min., and transformed into One Shot TOP10 chemically competent E. coli cells (Invitrogen) by adding 1 μL of the TOPO-cloned PCR product to chemically competent cells and incubating at 42^0 C for 40 seconds. Bacteria were grown at 37^O C without antibiotic selection for 1 hr, and then plated on LB-agar-carbenicillin (50 μg/mL) plates with X-gal and IPTG and incubated overnight at 37^O C to select for positive clones (blue/white selection). White colonies, chosen because they should contain PCR products, were picked with a sterile toothpick (i.e. 24 each from HCT116 and DKO2L) and grown on 100 μL of LB-carbenicillin (50 μg/mL) medium for 8 hrs at 37^O C. One microliter of this culture was used for direct colony PCR using M13 (-20) forward and M13 reverse vector primers flanking the insert. PCR products were electrophoresed on 1% agarose gels, and their band intensities were compared with the corresponding size bands of known concentration in a 2-log ladder (New England Biolabs).

To sequence amplified products of bisulfite treated DNAs, about 50 ng of each PCR product was cleaned up using EXOSAP-IT reagent (United States Biologicals). Approximately 25 ng was taken as template for cycle sequencing as described before. Cycle sequencing PCR reactions were performed using Big-Dye v. 3.1 dye terminator chemistry and either M13 (-20) forward or M13 reverse primer as sequencing primer. Sequencing products were purified by precipitation using ammonium acetate and ethanol, or by chromatography using G50 Sephadex (GE Amersham Pharmacia), resuspended in highly deionized (HiDi) formamide (Applied Biosystems), incubated at room temperature for 30 min followed by denaturation at 95° C for 3 min, and loaded on a 96 well-capillary electrophoresis DNA sequencer (Spectrumedix, Transgenomics).

Table 3: Primers used for bisulfite PCR

Gene	Forward primer	Reverse primer
GAGED2	GGAATTTATATATAGAGAGGAG	CAATTAAATCTACCTATAAACC
VIM	GYGGGGTTTAGTTTTTTGTTATTTT	AACATCCTACRATAAAAAAACRAAAACA
GAGE1	GGTAGTGTTGTGTGGTTTTTGT	ACCCCATTCAAAAAATCTACC
UBE2C	TGTAGTTAGAAGGAGGGGTTTA	ACTCTAAATACCAATACCTAAAACATAACA
L1	CCCCCAAAAATAAAACCTACAAAAAC	GTYGAATAGGAATAGTTTYGGTTTATAGTTTTTAG
Alu	GGTGGTTTAYGTTGTAATTTTAGTA	CTCTATCRCCCAAACTAAAATACA

Sequencing data were analyzed using BaseSpectrum software (Spectrumedix, Transgenomics) to generate chromatogram sequence files in Standard Chromatogram Format (scf). Chromatograms were imported into Sequencher 4.2.2 software (Gene Codes) and trimmed for vector sequence and for poor quality Phred score sequences. Trimmed sequences were aligned and manually checked for unconverted non-CpG cytosine bases (Cs). Sequences that contained more than 2% of such unconverted Cs were not used for the analysis. Sequences from all remaining, converted clones were

aligned with unconverted sequences and checked for unconverted (i.e. methylated) cytosines at every CpG dinucleotide site. The percentage of cytosine methylation at every CpG dinucleotide was determined for both HCT116 and DKO2L clones.

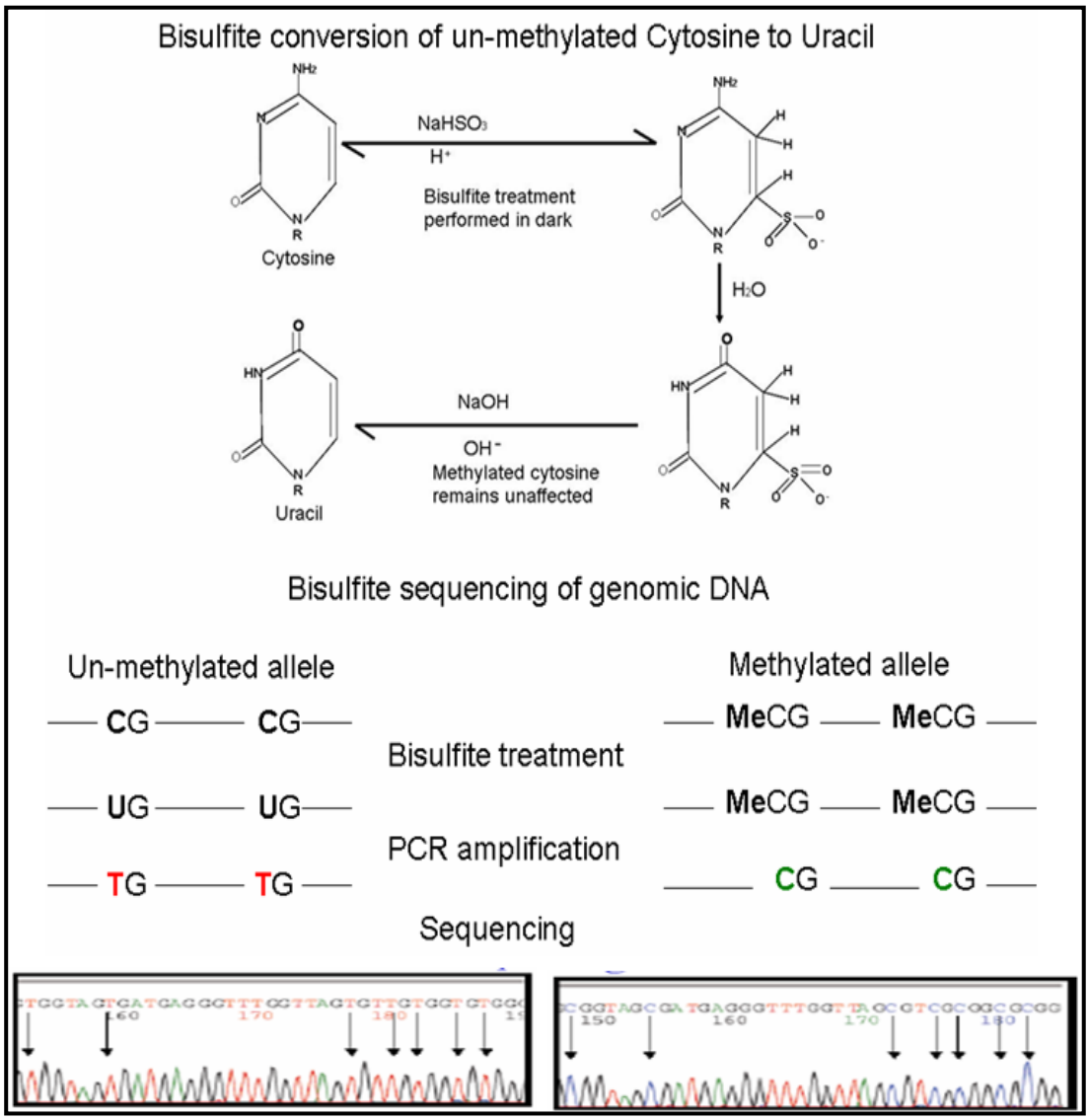

Figure 4: Schematic of the bisulfite sequencing technique.

Simulated longSAGE concatamer sequence files

We constructed simulated longSAGE concatamer sequence files containing ditags with variable lengths (i.e. 29, 30, 31, 32, 33, 34, 35, 36, 37 and 38 nt). These test cases were analyzed using SAGE-2000 v4.5 software to determine how altered ditag lengths affect tag extraction by the software.

Bioinformatic prediction of L1, Alu, HERV-K and SVA long-SAGE tags

We predicted longSAGE tags throughout the length of consensus sequences for L1, Alu, HERV-K (accession no. NM_001007236) and SVA elements (accession no. AC016142) using DNA Strider 1.4f3 (CEA, France). In brief, NlaIII restriction enzyme sites (5'CATG3') were identified throughout consensus L1, Alu, HERV-K and SVA sequences; in each orientation, the 17 bases following NlaIII sites comprised a long-SAGE tag

Comparison of SAGE findings with previous transcriptome studies

Data from SAGE and other expression-profiling platforms were compared (van Ruissen et al., 2005). First, both SAGE tags and array probes were mapped to NCBI Unigene clusters. If there were multiple SAGE tags for a cluster, the tag at the most 3' end was selected as a representative tag. Next, platforms were matched according to UniGene clusters, and data were filtered for the presence of gene expression (i.e., tag counts > 0 for SAGE libraries, and "present" calls on microarray platform).

To match human cell line data to mouse model data, the NCBI Homologene database was queried to obtain human/mouse homologs. To compare expression levels, an expression ratio between wild type and mutant was calculated and log-

transformed for each platform and each transcript. Using these ratios, the correspondence between platforms was estimated, using the Pearson correlation coefficient and Up/Down classification. Statistical analyses were performed with the software packages SPSS (SPSS Inc., Chicago, IL), and NCI's BRB-ArrayTools (http://linus.nci.nig.gov/BRB-arrayTools.html). BRB-$_{\text{ArrayTools}}$ was used to normalize raw microarray expression data.

Introduction

The development of cultured human colorectal cancer cell derivatives that continuously lack normal expression of DNA methyltransferases has provided a valuable, albeit somewhat flawed, tool to study the role of DNA methylation in the transcriptional regulation of genes.

Despite their residual problems of cumulative chromosomal instability and residual *DNMT1* activity, DKO cells have provided a useful *in vitro* system to study genome-wide effects of hypomethylation. For example, using cDNA microarrays, Gius et al. compared the transcriptomes of HCT116, DNMT1k/o, DNMT3b k/o and DKO cells with and without treatment with epigenetic modulators including the demethylating drug, 5-azadeoxycytidine, and a histone deacetylase inhibitor, trichostatin A (Gius et al., 2004). The two drug treatments were reported to have particular effects on HCT116 cells that are similar to those in DKO cells that have genetic knockouts (Gius et al., 2004). They reported that the number of genes with altered expression levels in *DNMT1*$^{-/-}$ cells was more than in DKO cells. This result is somewhat similar to our longSAGE results, described in following sections. As expected, given lower levels of expression of *DNMT3B,* the least affected transcriptome was observed in *DNMT3B*$^{-/-}$ cells (Gius et al., 2004). This study also showed that a 75 kb metallothionein gene cluster was coordinately upregulated, and that hypomethylating agents may have uncharacterized or undesirable "off-target" effects since similar numbers of genes were up- and downregulated.

Other experimental investigations using DKO cells have included the work of Polyak et al., who used DKO cells' hypomethylated DNA to validate methylation

sensitive digital karyotyping, a new method to map genomic DNA methylation patterns (Allegrucci et al., 2005). More recently, elevated microRNA levels were identified in DKO cells compared to their parent HCT116 cells, demonstrating that DNA methylation represses expression of miRNA precursors (Bostick et al., 2007; Yan et al., 2011). Moreover, the hypomethylation caused by the hypomorphic *DNMT1* allele can be considered to be more stable than hypomethylation caused by pharmacological agents which may be variably toxic, cause off-target effects, have heterogeneous uptake and metabolism in a population of cells, etc.

We undertook this project to study changes in the human transcriptome upon genome-wide hypomethylation. Here we constructed and analyzed both long-tag Serial Analysis of Gene Expression (longSAGE) (Saha et al., 2002; Velculescu et al., 1995) libraries from parental HCT116 cells and their DKO derivatives, and analyzed mouse exon microarrays. Long SAGE has several advantages over other methods:

1. Transcripts can be quantified without prior knowledge of their sequence structure;
2. Results are quantitative, even for poorly expressed genes;
3. Results are comparable between platforms and with many previously reported SAGE reference libraries;
4. Novel gene transcripts can be identified based on their tag sequences; and
5. Subtle sequence differences distinguishing transcripts may be specifically and sensitively identified by sequencing rather than by differential hybridization on microarrays (Bourc'his and Bestor, 2004; Saha et al., 2002; Velculescu et al., 1995).

In this study, we identified several classes of genes whose transcription is upregulated strongly in the context of genome-wide hypomethylation. In numerous

cases, differential longSAGE results were verified by Northern blotting, qRT-PCR and exon microarray results, and a direct connection with promoter hypomethylation could be established. As described in Chapter 4, we also studied the effects of genome-wide hypomethylation on the transcription of transposable elements such as L1, *Alu*, HERV and SVA elements, which were not addressed by previous studies. Several previously unreported longSAGE tags were identified, suggesting expression of previously unreported transcripts or splice variants. We statistically compared the findings of our study (LongSAGE and the exon microarray) with various previously accomplished transcriptional profiling studies on a genome-wide scale, and thereby highlighting the strengths and weaknesses of various profiling methods. We identified a potential "transcriptome signature" of genome-wide hypomethylation.

We note that more recently developed methodologies, including additional transcript tagging procedures such as MPSS (Brenner et al., 2000), GIS (Allegrucci et al., 2005) and RNAseq (Mortazavi et al., 2008), present additional opportunities and challenges in cataloging and characterizing the entire transcriptome. For example, GIS can define corresponding 5' ends of transcripts. RNASeq has the potential to characterize alternative splicing junctions, additional variability at the 5' and 3' ends of transcripts, and even RNA editing (Li et al., 2011). However, these methods are technically more challenging, more expensive and not uniformly available. While we started this study long before some of these technologies were available, we believe that future comparisons between results from such distinct, new technologies will shed further light on the cancer cell transcriptome.

RESULTS

We set out to test the hypothesis that disruption of DNA methylation would result in profound alterations in the transcription of many genes and transposable elements. Because of their relatively unbiased measurement of transcripts (whether previously identified or not), longSAGE libraries were generated from related cell lines derived from parental HCT116 human colorectal cancer cells, i.e. HCT116 (WT), clone 1C1 (*DNMT1*$^{-/-}$) and clones of DKO2L cells lacking DNA methyltransferases *DNMT1* and *DNMT3B* (Rhee et al., 2002; Rhee et al., 2000). As described in Table 4, over 100,000 tags each were sequenced by conventional Sanger sequencing in two approximately equal groups from the HCT116 and DKO2L libraries, respectively, and over 50,000 tags from the *DNMT1*$^{-/-}$ clone 1C1. For subsequent comparisons, we normalized raw tag counts to tags per million.

Since DKO2L cells have much more profound genomic hypomethylation than single knockout *DNMT1*$^{-/-}$ 1C1 cells (Rhee et al., 2002; Rhee et al., 2000), we anticipated that more genes in the double knockout cells would undergo pronounced changes in expression. However, while many hundreds of genes were either up- or down-regulated at least 5-fold in both mutant cell lines, more genes in *DNMT1*$^{-/-}$ 1C1 cells (2,103 genes) were upregulated than in DKO2L cells (1,711). In addition, the extent of upregulation of many genes in 1C1 cells e.g. *BST2*, *G1P3*, *BBAP* etc. was greater than in DKO2L cells. Moreover, several genes showed opposite changes in expression in Dnmt1$^{-/-}$ and DKO2L cells, e.g. *VIM*, *ARL6IP*, *PSCD2*, *PSMA*, *ZNF205*. Since we wanted to identify transcripts whose changes in expression were consistent and directly related to methylation changes, we did not fully sequence the *DNMT1*$^{-/-}$1C1 library, and instead we focused on significant differences between the HCT116 and DKO2L libraries.

Generation of longSAGE libraries and quality control

Sequencing of the two longSAGE libraries from WT and DKO2L cells each was completed in two approximately comparable batches. To check the internal consistency and quality of the library sequences, we compared tag frequencies within each batch, as displayed by scatter plots (Fig. 5). This analysis revealed a strong, linear correlation between the batches. Pearson's correlation coefficients were calculated for both libraries as 0.979 (Table 5). This suggested that the tags were randomly sampled during sequencing of both batches, and that sequencing quality was uniformly high.

The parsing software, SAGE-2000 v4.5, can determine the occurrence of duplicate dimers, which can arise due to ligation bias or PCR amplification bias. A high level (>10%) of duplicate dimers would reflect poor library quality (Velculescu et al., 1995). However, in both of our longSAGE libraries, the frequency of duplicate dimers was < 10%, reflecting their high quality. Both libraries also had a similar number of unique tags (Table 4).

SAGE 2000 v4.5 software can extract tags from appropriate length ditags arising from high quality sequence runs.

We tested the ability of SAGE-2000 v4.5 software to determine ditag lengths and to extract tags of appropriate length (17 nt). LongSAGE concatamer sequence files were simulated by altering a real sequencing output file, such that it contained ditags with variable lengths (29-38 nt). This was accomplished by inserting sequence punctuation marks at the specified positions. We verified that SAGE-2000 v4.5 software extracts tags from those ditags whose length falls between 32 and 34 bases. It does not extract tags from ditags that are longer than this range of lengths. When the ditag length is 34 nt, the

software extracts two 17 nt tags from either side as two individual tags. When the ditag length is 32 or 33 nt, the software extracts two individual tags from either side so the two resulting extracted tags share 1 or 2 nt overlapping sequence, respectively. If the ditag length is more than 34 or less than 32 nt, no tags are extracted (not shown).

We also tested if SAGE-2000 v4.5 software is able to distinguish between sequence patches with good vs. poor Phred quality scores (base calls with Phred score less than 20 were taken as poor quality), thereby extracting tags only from good quality sequence patches. We constructed several simulation files containing concatamers of longSAGE tags that had patches of good and poor quality sequences in different lengths and positions. The results verified that the software indeed is able to distinguish poor quality sequence patches, resulting in extraction of tags only from good quality (Phred scores > 20) sequence patches (not shown).

To determine relative changes of tag expression in DKO cells vs. parental HCT116 (WT) cells, we plotted normalized tag counts (per million) on a scatter plot (Fig. 6). Pearson's correlation coefficient between the WT HCT116 and DKO2L libraries is low, i.e. 0.527, suggesting substantial differences between the transcriptomes of the two cell types. To detect significant expression differences between tags in the two libraries, we also calculated the Z-test (Kal et al., 1999), since it has been shown to be as effective and efficient as other methods such as Fisher's exact test and a simulation-based approach (Ruijter et al., 2002). We defined gene expression levels to be significantly different when p-values were less than 0.05 (Table 6 and Fig. 6). For p-values < 0.05, the total number of unique differentially expressed tags is 1,003; and for p-values < 0.01, the total number of unique differentially expressed tags is 117. The most highly up- and down-regulated genes are tabulated in Tables 7 and 8.

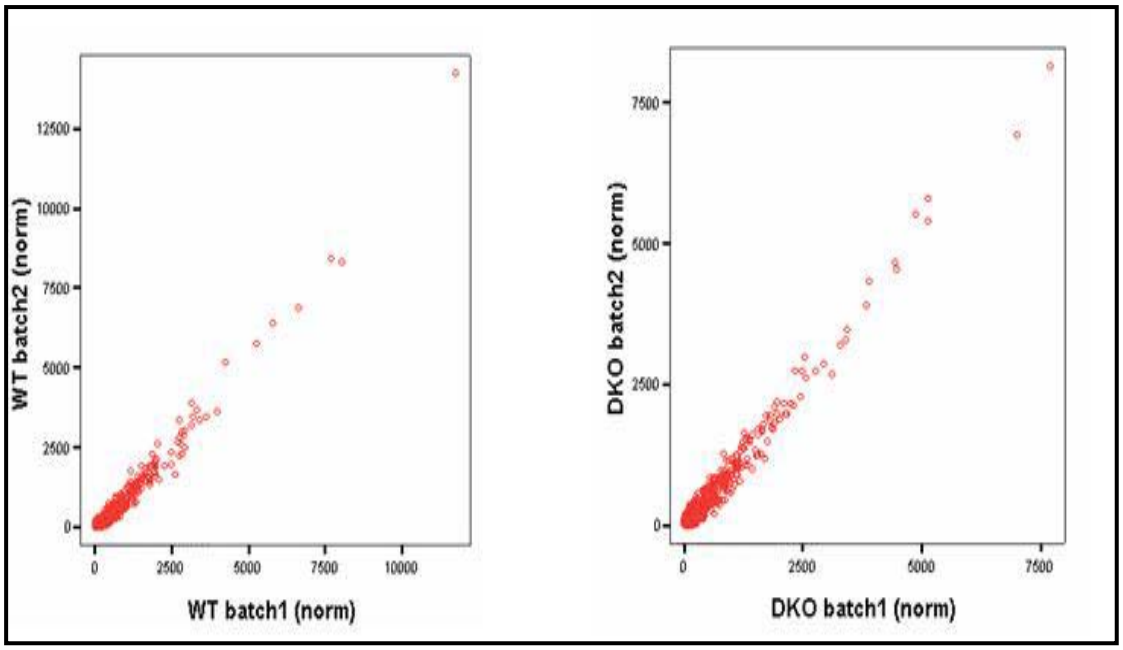

Figure 5: **Scatter plots show strong linear correlations between tag counts from two sequencing batches of longSAGE libraries.** Libraries were prepared from (A) HCT116 WT and (B) DKO2L cells. The tag counts in each batch are normalized to tags per million (TPM).

Table 4: Summary of HCT116 and DKO2L longSAGE libraries. High quality longSAGE libraries were constructed from HCT16 and DKO2L cells. *Column headers* Type: NlaIII, used as anchoring enzyme; total tags: number of tags sequenced in library; % duplicate dimers: percentage of duplicate dimer sequences in library (these ditags are excluded from the total tag count); tags w/ count >= 2: total number of tags with counts at least 2, Unique tags: number of distinct tags appearing in the library.

Library	Type	Total tags	Tags w/ count ≥2	% Duplicate dimers	Unique tags
HCT116	NlaIII	114,438	9,116	9.41	39,518
DKO2L	NlaIII	101,872	9,014	7.85	41,389

Table 5: Comparison between the two sequencing batches of HCT116 and DKO2L libraries. Also see figure 5, scatter plots show strong linear correlations between two sequencing batches of WT and DKO longSAGE libraries.

Comparison	Pearson's coefficient of correlation
HCT116 batch 1 vs. batch 2	0.979
DKO2L batch1 vs. batch 2	0.979

Table 6: Numbers of genes significantly up- or down-regulated in DKO2L vs. WT cells. Hundreds of genes are up- and downregulated in DKO2L cells compared with wildtype HCT116 cells (categorized as per p-value range during statistical analysis).

p-value	Upregulated genes	Downregulated genes	Total # of deregulated genes
P<0.01	53	64	117
P<0.05	527	476	1003

Table 7: The most highly up-regulated tags in DKO2L vs. WT cells. Due to assignment of an arbitrary minimal tag frequency, ratios represent an arbitrary, miminum change in tag frequencies in DKO2L vs. WT cells. *Column headers*: Tag, upregulated tags; description, annotated gene represented by a tag; NormWT, normalized tag counts (tags per million) in HCT116 library; NormDKO, normalized tag counts in DKO2L library; RatioDKO/WT, ratio of tag counts in NormDKO vs. NormWT; Northern, validation performed by Northern blotting; q-PCR, validation performed by quantitative PCR; ND, not done.

Tag	Description	NormWT	NormDKO	RatioDKO/WT	Northern	q-PCR
CCAGGGGAGAAGGCACC	IFI27, Interferon, alpha-inducible protein 27	4.3	1109.2	253.8	Yes	Yes
GAAGGTGATCTGCAAGA	GAGED2, X antigen family, member 1	4.3	834.3	190.9	Yes	Yes
GAAGGTGATCTGCAAGA	GAGED2, X antigen family, member 1	4.3	834.3	190.9	Yes	Yes
CAGTCTAAAATGCTTCA	UCHL1, Ubiquitin carboxyl-terminal esterase L1	4.3	549.7	125.8	Yes	ND
CAGCCTGGGGCCACTGC	PRSS21, Protease, serine, 21 (testisin)	4.3	480.9	110	Yes	Yes
TGCTGCCTGTTGTTATG	BST2, Bone marrow stromal cell antigen 2	4.3	431.9	98.8	Yes	ND
CAGACGGTGGCCCAGCC	LOC147111, Hypothetical protein LOC147111	4.3	323.9	74.1	ND	ND
ACCCTGCCAAATCCCCC	- CDNA FLJ31134 fis, clone IMR322000984	4.3	206.1	47.1	ND	ND
GTGCCTGCCGAGAGGGC	LOC339903, Hypothetical protein LOC339903	4.3	196.3	44.9	ND	ND
CGCCTCCGGGAGAAATT	VIM, Vimentin	4.3	186.5	42.6	Yes	Yes
CTTGTAATCCTACTTGG	G1P3, Interferon, alpha-inducible protein	4.3	186.5	42.6	Yes	ND
AGCTGTGCCAAGTGTGC	MT1S, Metallothionein 1A (functional)	4.3	186.5	42.6	ND	Yes
AGATTTAAATTCTGTGG	CTCFL, CCCTC-binding factor -like	4.3	137.4	31.4	ND	Yes

Table 8: The most highly downregulated sequence tags in DKO2L vs. HCT116 cells. The most downregulated tags in DKO2L vs. WT cells are indicated. Due to assignment of an arbitrary minimal tag frequency, ratios represent a minimum change in tag frequencies in DKO2L vs. WT cells. *Column headers:* Tag, downregulated tag; description, annotated gene represented by a tag; NormWT, normalized tag count (tags per million) in HCT116 library; NormDKO, normalized tag

count (tags per million) in DKO2L library; RatioDKO/WT, ratio of NormDKO to NormWT. Tags which have "zero" counts in one library but a positive number of counts in the other arbitrarily were assigned a value of 0.5 tags, to avoid division by zero. Upon normalization (tags per million), they therefore were assigned an arbitrary value of 4.3 TPM.

Tag	Description	NormWT	NormDKO	RatioDKO/WT
GGTGGTAAAGTACCACC	ST13 Suppression of tumorigenicity 13	78.6	4.9	0.06
CATCTAGAGGGCTCAAT	Unknown, no UniGene associated to this tag	78.6	4.9	0.06
CACTCTAAGATGAGTTC	MGC13017 Similar to RIKEN cDNA A430101B06 gene	78.6	4.9	0.06
CATCCTTGGGCAGGTTG	ST13 Suppression of tumorigenicity 13	78.6	4.9	0.06
GCAGCTAATTTTGTAAC	GK001 GK001 protein	78.6	4.9	0.06
ACTACAAATAGTCCGAA	MGC4825 Hypothetical protein MGC4825	78.6	4.9	0.06
ACTACAAATAGTCCGAA	RPL9 Ribosomal protein L9	78.6	4.9	0.06
AAGCTGTTGTGTGAGGT	DNMT1 DNA (cytosine-5-)-methyltransferase 1	87.3	4.9	0.05
CTGAGGCGCTTCCTGGG	THOP1 Thimet oligopeptidase 1	96.1	4.9	0.05
TACCATCAATAAAGGAC	Unknown, no UniGene associated to this tag	104.8	4.9	0.04

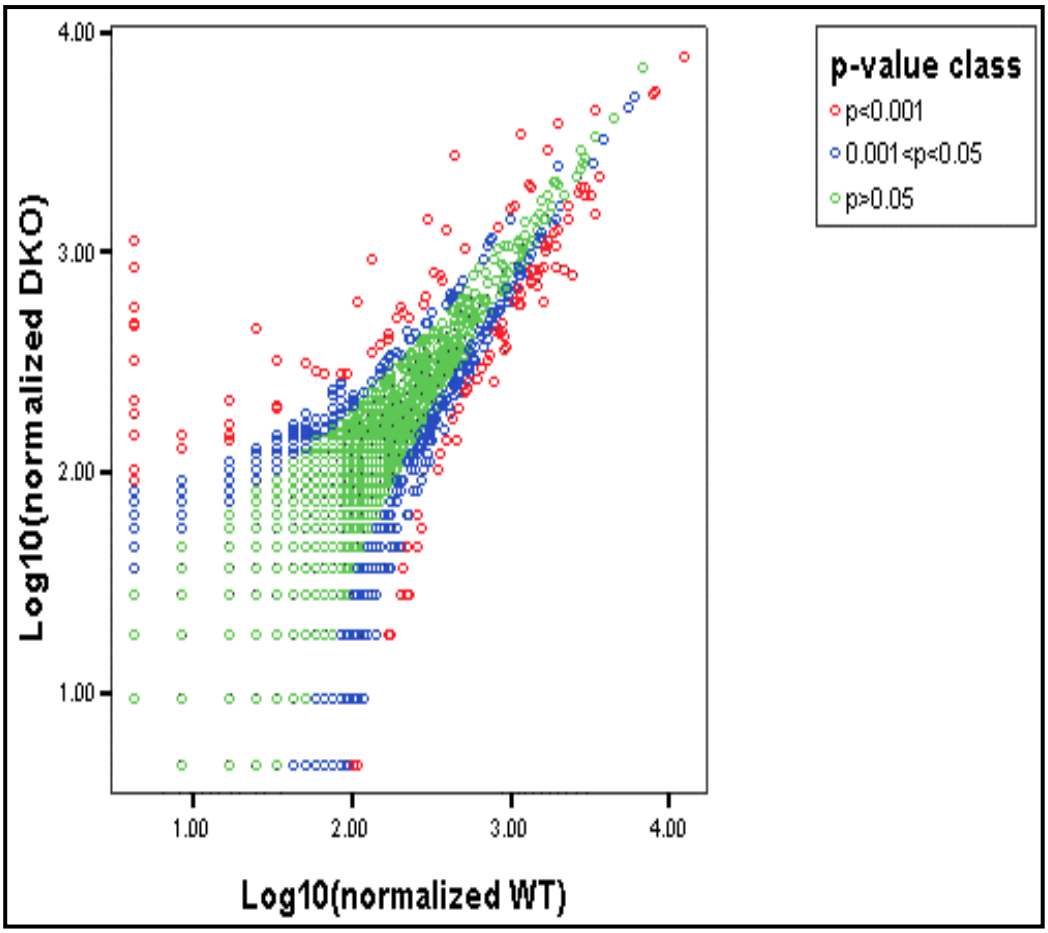

Figure 6: Comparison of transcripts in HCT116 and DKO libraries. A scatter plot of normalized transcript frequencies is presented here. Each dot on this log transformed (log10) plot represents normalized expression levels of individual sequence tags detected in DKO2L (*y*-axis) vs. WT (*x-axis*) cells. The Z test was performed for all data points, as indicated (*key*). *Red:* genes which are highly differentially expressed, p value < 0.01; *blue*: genes which are differentially expressed with p values between 0.01 and 0.05; *green*: genes with p values > 0.05. Tag counts in each library were normalized to tags per million.

As measured by longSAGE tag frequences, the overall numbers of genes that are significantly upregulated or downregulated in DKO2L vs. HCT116 (WT) cells are similar (Table 6). However, the fold increase in significantly upregulated transcripts was substantially more than the fold decrease in downregulated transcripts.

When differentially expressed transcripts measured by Gius et al. (Gius et al., 2004) were compared on a statistical basis with those identified in our analysis of longSAGE libraries, the correlation between these two studies was remarkably low (Pearson's correlation coefficient, 0.140) (Figure 7). This result suggested the possibility of a systematic bias distinguishing between these two assays. However, 58 genes were deregulated as measured by both experimental platforms, i.e. 33 genes were upregulated, while 25 genes were downregulated (Appendix 1.1 and 1.2). Sets of significantly up- and downregulated tags (p values < 0.05, Fishers' exact test) were analyzed separately, using bioinformatics tools on Database for Annotation, Visualization, and Integrated Discovery (DAVID).

By contrast, when we compared differentially expressed genes that we identified here by longSAGE vs. those that we identified from the same cell samples using mouse exon microarrays, the correlation was much higher (Figures 8 and 9) and (Appendices 6.1 and 6.2). Expression analysis was done by SAGE and Affymetrix HuEx-1_0-st-v2 platforms. No technical duplicates were performed using these exon microarrays, because of a high degree of technical reproducibility that has been reported for them. This is likely attributable to the numerous independent measurements of transcripts based on multiple exonic probes per spliced transcript. SAGE tags were annotated by using SAGE Genie and were matched to Affymetrix (Hu-Ex) transcript sets by NCBI Gene ID. 8,782 genes

were matched between two platforms. We focused on a total of 119 differentially expressed genes identified from the SAGE platform (Z-score < -3 or >3: p-value < 0.0026) for comparison with exon microarray data (Fig 8). We plotted the $\log_{10}$ ratios of transcript expression in DKO vs. WT, measured using both Affymetrix exon microarray and SAGE platforms, in scatter plots.

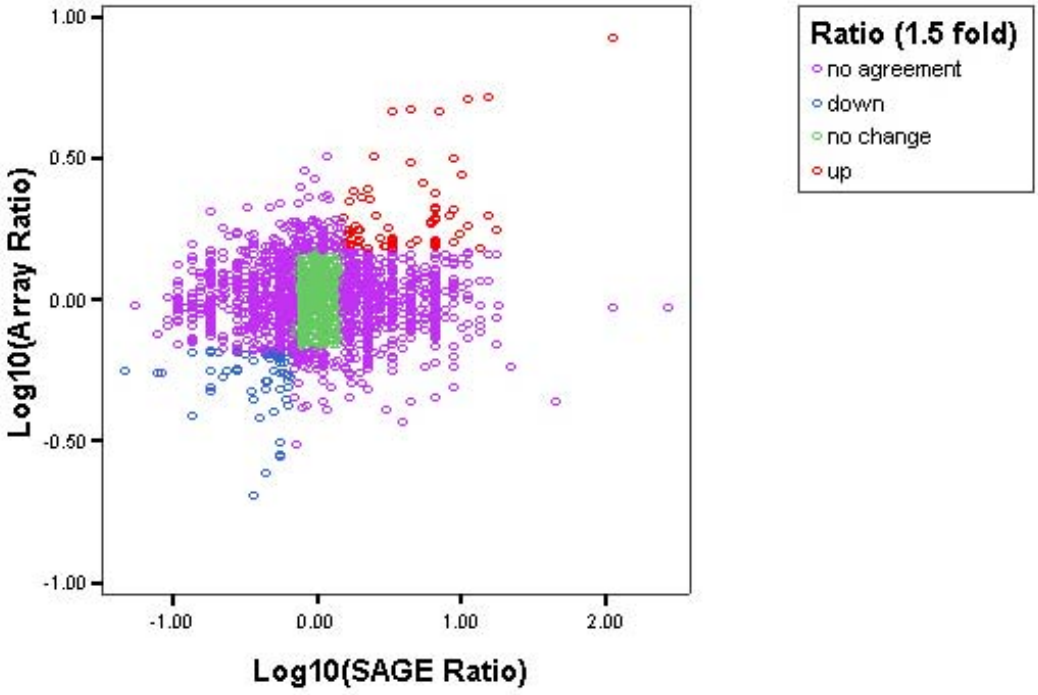

Figure 7: Correlation between transcriptional profiles measured using cDNA microarray vs. longSAGE. Presented here is a scatter plot of ratios of transcript expression levels, comparing transcripts from DKO vs. WT HCT116 cells, as measured by longSAGE (x-axis) vs. cDNA microarray experiments (y- axis) (Gius et al., 2004). Each tag count in longSAGE analysis was normalized to tags per million, ratios of counts in DKO vs. WT were calculated, and values were log transformed (log10). Similar ratios were calculated from exon microarray results. Red circles indicate those genes which are significantly upregulated as measured by both assays; blue circles indicate those genes which are downregulated as per both assays; green circles indicate those genes which are not significantly changed as per either assay; purple circles indicate those genes which show poor agreement between the assays.

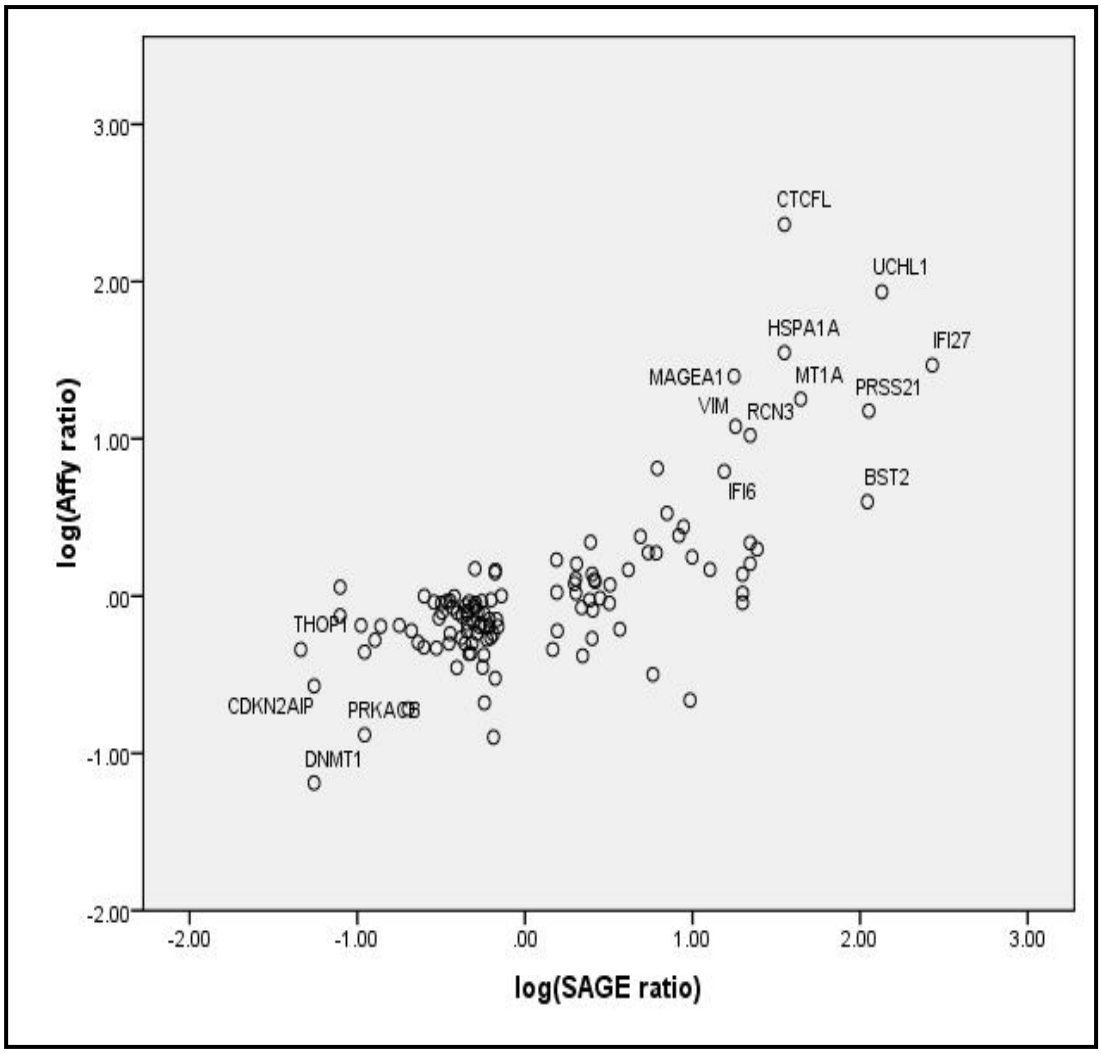

Figure 8: Scatter plot depicting the most highly differentially expressed genes observed using the long SAGE platform with Z-score < -3 or >3 (p-value < 0.0026) vs. their differential expression measured by the Affymetrix platform Z-score: <-3 or > 3, Pearson correlation = 0.742 (N=119 genes). *X*-axis: $\log_{10}$ transformation of the longSAGE DKO/WT ratio; *y*-axis: $\log_{10}$ transformation of the exon array (Affymetrix) DKO/WT ratio. Certain highly upregulated and highly downregulated genes are labeled.

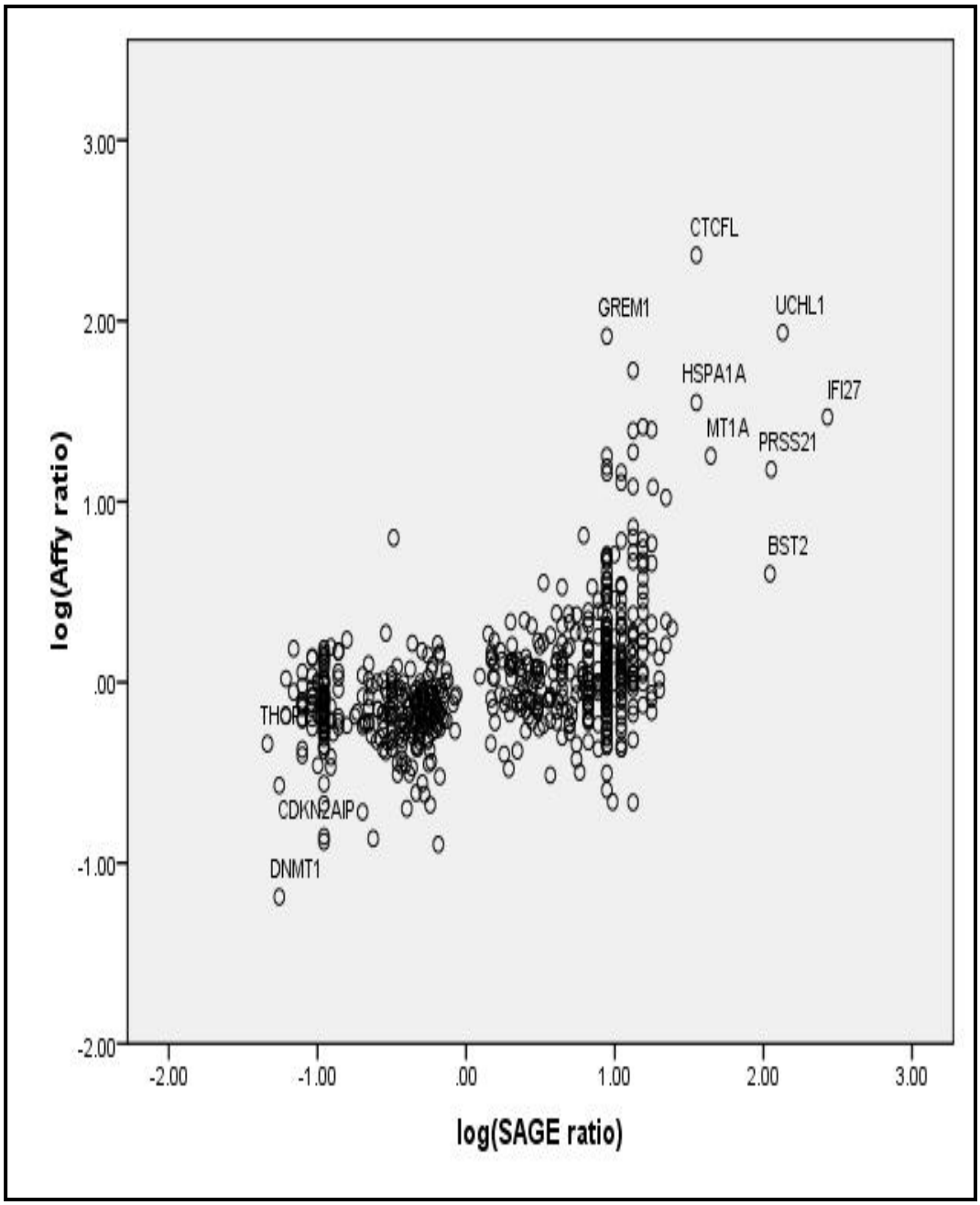

Figure 9: Scatter plot depicting differentially expressed genes (total of 631) in long SAGE platform with Z-score < -2 or >2 and comparing their ratio to those of Affymetrix platform Z-score: <-3 or > 3, Pearson correlation = 0.498 (N=631 genes). *X*-axis: $\log_{10}$ transformation of the longSAGE DKO/WT ratio; *y*-axis: $\log_{10}$ transformation of the exon array (Affymetrix) DKO/WT ratio. Certain highly upregulated and highly downregulated genes are labeled.

Bioinformatic analyses of significantly deregulated tags

Data obtained from comparing transcriptomes of HCT116 and DKO2L cells were subjected to additional bioinformatics analyses. We focused on those tags that were differentially transcribed to a significant extent ($p<0.05$) in DKO2L cells. A total of 1,003 such tags were identified. Of these, 527 tags were upregulated, while 476 tags were downregulated (Appendices 2.1 and 2.2). The sets of up and downregulated tags were analyzed separately using bioinformatic tools on Database for Annotation, Visualization, and Integrated Discovery (DAVID) to categorize the affected genes into various pathways, biological and molecular functions, and chromosomal locations to facilitate our understanding of biological meaning of the various findings in SAGE analysis.

The upregulated tags were categorized according to their cytobands and chromosomes of origin and using tools on DAVID bioinformatics resources that uses Fishers' exact test (p value < 0.05). The analysis showed that tags from chromosome 17, 19 and 1 and cytobands 11q13, 6p21.3, 17q25 and 1p32-31 were most represented amongst the upregulated tags (Appendix 3.1 and 3.2), whereas tags from chromosome 19, 17 and 16 and cytobands 16p13.3, 19p13.3, 22q13 and 11q13 were most represented amongst the most downregulated tags (Appendix 4.1 and 4.2).

In addition, differentially transcribed tags were classified according to various biological processes (GO terms, DAVID) with at least 2 genes in each category and p-value < 0.05. This analysis showed that genes involved in cell organization and biogenesis, DNA metabolism, negative regulation of cellular processes, response to DNA damage stimulus, chromatin assembly and cellular localization were amongst the upregulated genes (Appendix 3.3) while genes related to protein biosynthesis, RNA metabolism and processing, RNA splicing, translation, ribosome biogenesis and assembly

were amongst the downregulated genes (Appendix 4.3). Notably, as expected, the longSAGE tag for *DNMT1* was among the list of most downregulated tags in the DKO2L cells (Rhee et al., 2002; Rhee et al., 2000).

Validation of longSAGE findings

To validate some of the most differentially expressed transcripts that were identified initially by longSAGE, we used two independent techniques, *i.e.* Northern hybridization and quantitative reverse transcriptase - polymerase chain reaction (qRT-PCR). Thus we reassessed the differential expression of *IFI27*, *GAGED2*, *PRSS21*, *VIM*, *UCHL1*, *BST2*, *G1P3*, *G1P2* and *UBE2C*; and of *IFI27*, *GAGED2*, *PRSS21*, *MT1A* and *CTCFL*, respectively. In qRT-PCR experiments, samples were analyzed in triplicate. We measured *TSMB10* and *RPL13A* transcript levels to normalize the data, as these genes showed similar expression levels in both cell types (Vandesompele et al., 2002). Our qRT-PCR and the Northern hybridization analysis corroborated upregulation of all genes originally identified by longSAGE analysis. In general, there was very good agreement among all three independent techniques (Fig. 10 and 11).

As another test of the reproducibility of our longSAGE results, we assayed several independently derived, subcloned DKO cell lines for upregulation of some of the most increased transcripts in DKO2L cells (Table 7). These cell lines include clones 2, 3 and 8 as originally described (Bachman et al., 2003; Rhee et al., 2002). For the DKO2 clone, both early and late passage cells were analyzed (Bachman et al., 2003). The results show that independent DKO clones have comparable up-regulation of the same genes (Fig. 11), with minor variations in the extent of over-expression in different DKO clones. They validate the main findings from our longSAGE libraries. They also suggest that various DKO clones, which showed minor differences in genomic methylation levels (as described originally), show little variations in the gene expression pattern, at least for the most highly upregulated genes we analyzed.

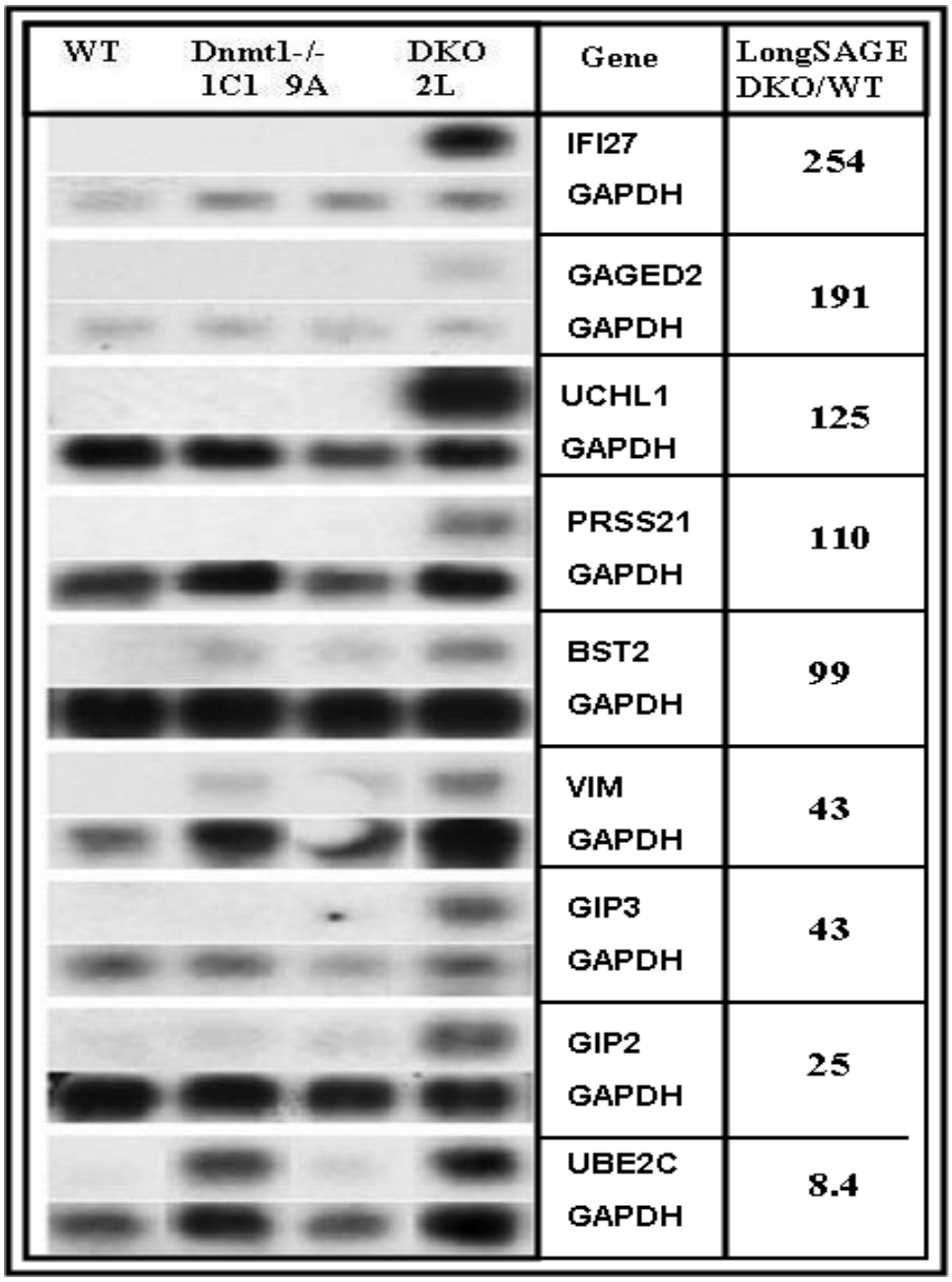

Figure 10: Validation of longSAGE results by Northern hybridization. Nine upregulated genes (IFI27, GAGED-2, G1P3, G1P2, PRSS21, UCHL1, BST2 VIM, and UBE2C) were analyzed. All the genes show significant upregulation in all the DKO clonal cell lines as compared to the HCT116. WT = HCT116, DNMT1k/o ICI and DNMT1k/o 9A are two independent clones of DNMT1 single knockout; DKO2L is the clone of DNMT1 & 3b double knockout derivative HCT116 cells used in construction of longSAGE library

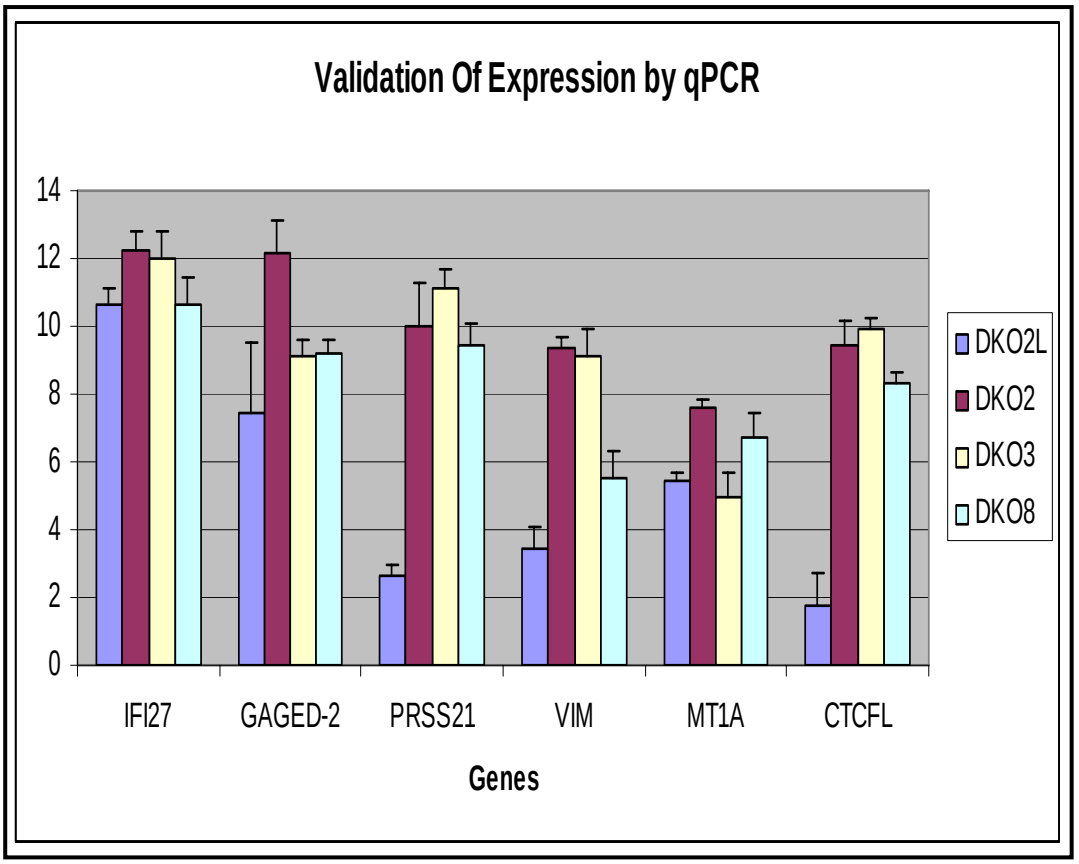

Figure 11: Validation of longSAGE results by q-PCR. We sought to validate the differential expression levels of 6 genes (*IFI27*, *GAGED-2*, *PRSS21*, *VIM*, *MT1A* and *CTCFL*), all shown by LongSAGE to be upregulated in DKO cells, using qRTPCR. Several different independent clones of DKO cells were tested, including DKO2 (DKO2L (late passage) and DKO2 (early passage)), DKO3 and DKO8 (Rhee et al., 2002). All the genes showed significant upregulation in all the DKO clones as compared to the HCT116. The *y*-axis shows the threshold cycle difference with respect to HCT116 sample (delta C_T), and the *x*-axis shows the genes analyzed. The error bars represent standard deviations of the triplicate samples.

Genome-wide hypomethylation activates expression of several potential novel transcripts

Out of 1,003 deregulated tags ($p<0.05$), 172 (17.1%) did not have any corresponding UniGene ID or tag annotation. These tags are apparently novel tags, of which 94 (54%) were upregulated and 80 (85%) were significently upregulated at least 5-fold. Moreover, 69 of these are markedly upregulated, novel tags (86%) which showed no expression in HCT116 cells, while 9 (11.2 %) showed very low expression and 2 (2.5%) showed moderate expression in HCT116 cells. These potentially novel tags could represent alternatively spliced or previously undescribed transcripts (Appendices 5.1 and 5.2).

Promoter methylation is a negative transcriptional regulator

Upregulation of transcripts in the context of genomic hypomethylation could be due to an indirect effect, i.e. mediated by factors in *trans* rather than by direct changes in promoter methylation in *cis*. To establish a direct correlation between observed changes in gene expression and upstream promoter hypomethylation, we performed bisulfite sequencing at the promoters of some of the most upregulated genes in both parental and DKO cells. Bisulfite sequencing analysis is highly quantitative, and several CpG dinucleotides can be assessed in one bisulfite PCR amplicon (Grunau et al., 2001). Bisulfite sequencing was performed at the promoters or 5' ends of *VIM* (-33 to 169), *GAGED-2* (-198 to 103), *GAGE1* (-17 to 359) and *UBE2C* (-276 to 242). The numbers in parentheses show the gene sequence coordinates of the region analyzed, relative to transcription start site at +1. *VIM* showed 77% methylation in HCT116 cells which was reduced to 1% in DKO2L cells, *GAGED-2* showed 79% methylation in HCT116 cells which came down to 5.6% in DKO2L cells and *GAGE1* showed 89% methylation in HCT116 cells which was reduced to 13% in DKO2L cells. In summary, all three genes

analyzed showed a significant reduction in methylation in DKO2L cells as compared to the HCT116 cells, ranging from 85% to 98% (Figure 12, 13 and 14). These results compare favorably with the initial determination of genome-wide hypomethylation in DKO cells of 95% (Rhee et al., 2002).

We also identified a few genes where expression was affected, but did not correlate with a change in methylation at the promoter region. One such example is *UBE2C* (Figure 15), whose expression was upregulated by ~8-fold in DKO2L cells but its promoter was completely unmethylated in both HCT116 and DKO2L cells. These analyses suggest that while transcription of most of the highly upregulated genes inversely correlates with promoter methylation, there are genes whose expression changes do not correlate with promoter methylation changes, suggesting that the genome-wide hypomethylation can affect expression changes in *trans.*

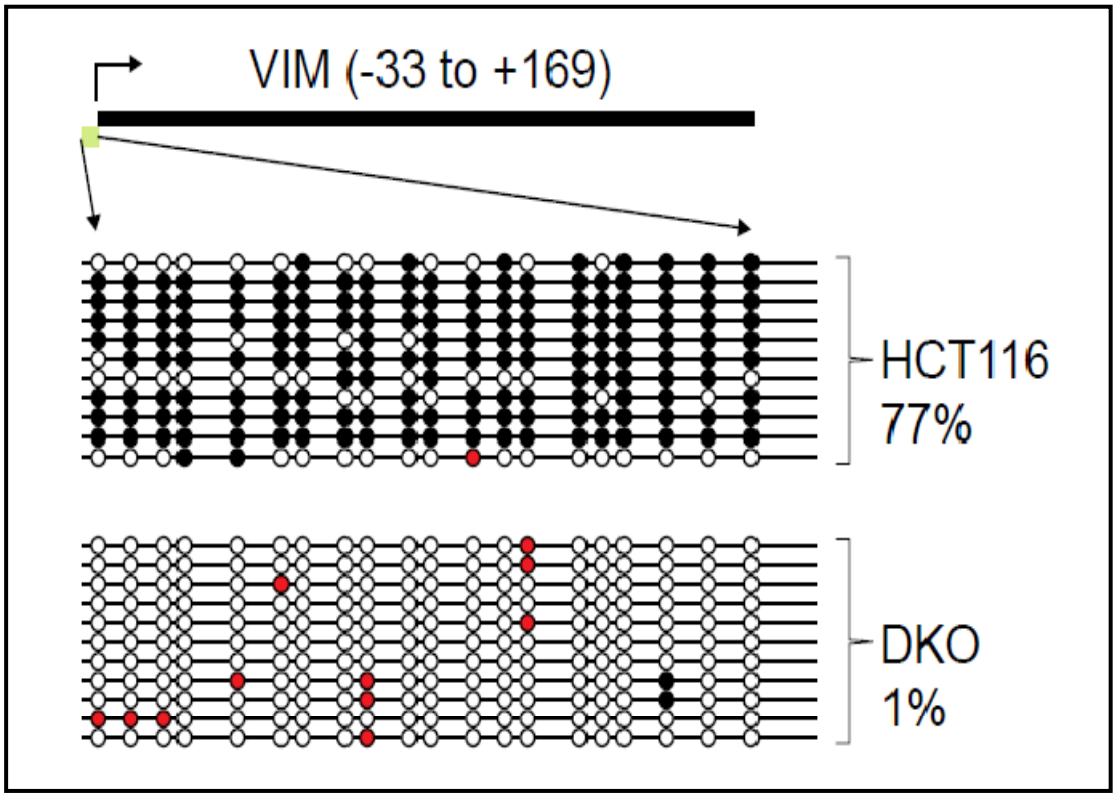

Figure 12: Bisulfite sequencing at the 5' CpG island of VIM (Vimentin) gene. Top numbers in parentheses show the coordinates of the genomic stretch sequenced; horizontal lines represent a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotides; *black*, methylated, *blue*, unmethylated, *red*, poor sequencing; *right:* the number following HCT116 and DKO2L shows the total percent methylation. We analyzed only those sequences where < 2% of non-CpG Cs were unconverted to uracil (thymidine) after bisulfite treatment.

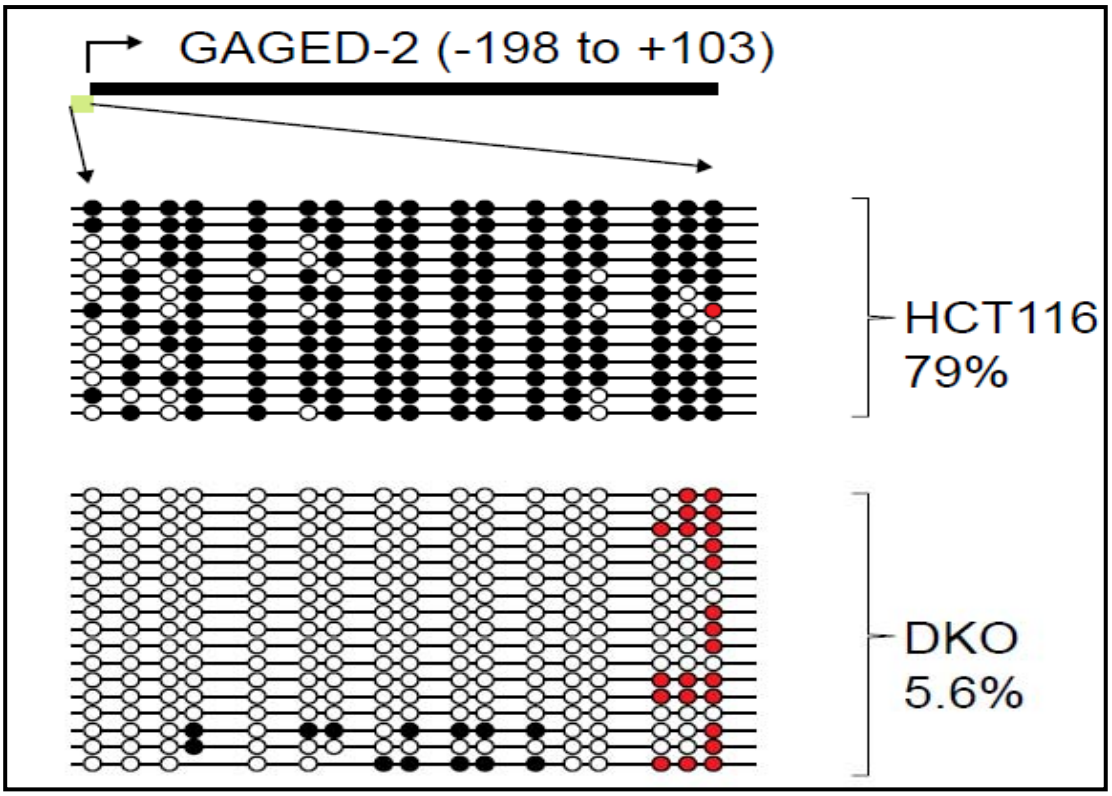

Figure 13: Bisulfite sequencing at the 5' CpG island of *GAGED2* gene. *Top*: Numbers in parentheses show the coordinates of the genomic stretch sequenced; horizontal lines represent a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotides; *black*, methylated, *blue*, unmethylated, *red*, poor sequencing; *right*: the number following HCT116 and DKO2L shows the total percent methylation. We analyzed only those sequences where < 2% of non-CpG Cs were unconverted to uracil (thymidine) after bisulfite treatment.

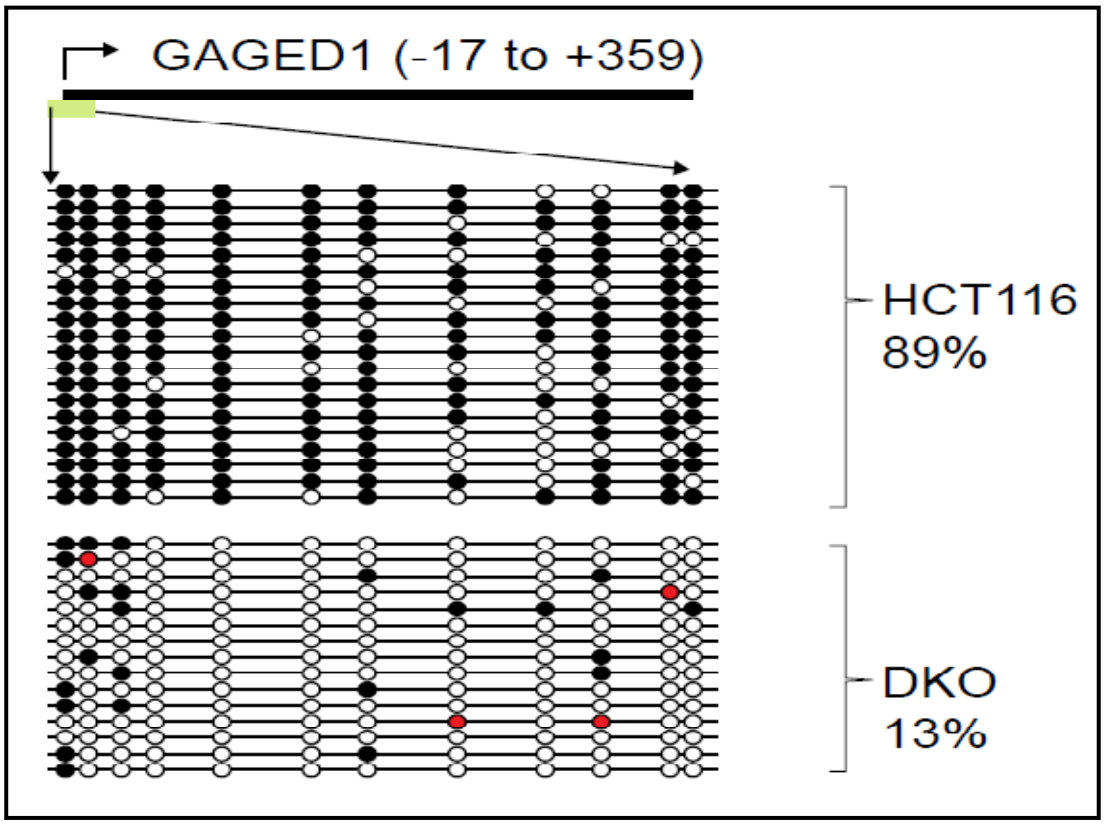

Figure 14: Bisulfite sequencing at the 5' CpG island of *GAGE1* gene. *Top*: Numbers in parentheses show the coordinates of the genomic stretch sequenced; horizontal lines represent a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotides; *black*, methylated, *blue*, unmethylated, *red*, poor sequencing; *right*: the number following HCT116 and DKO2L shows the total percent methylation. We analyzed only those sequences where < 2% of non-CpG Cs were unconverted to uracil (thymidine) after bisulfite treatment.

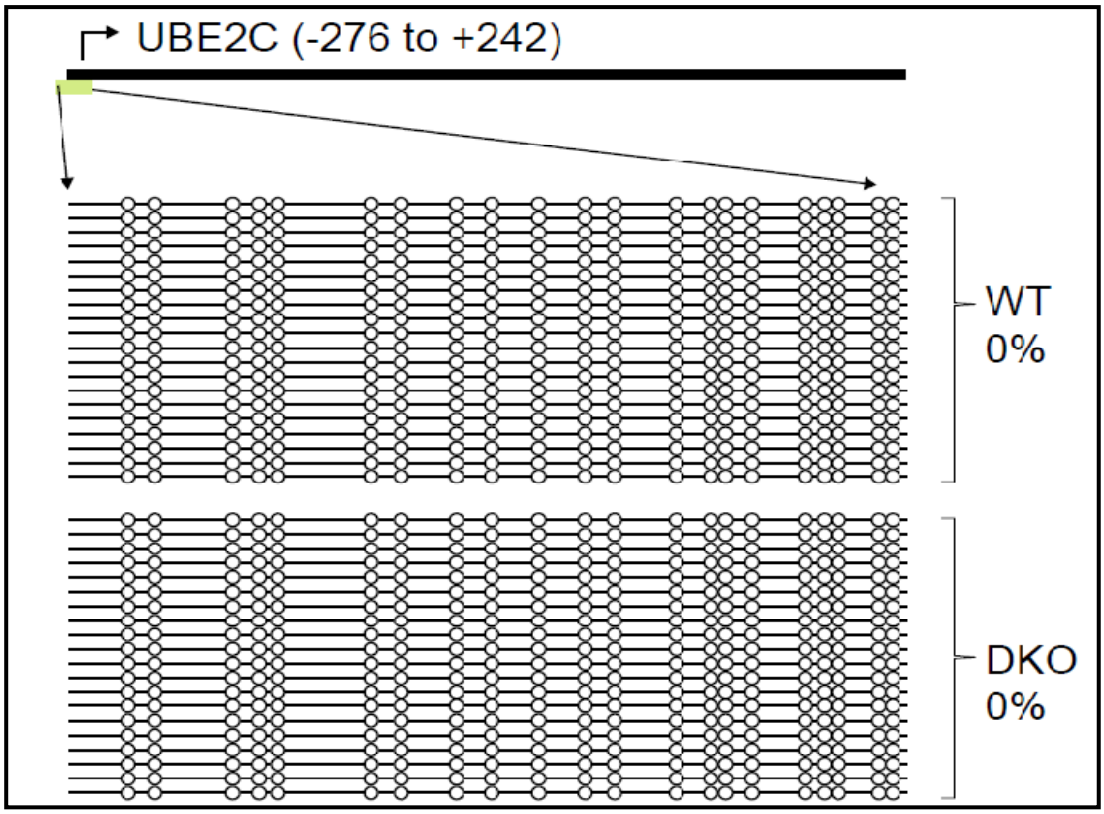

Figure 15: Bisulfite sequencing at the 5' CpG island of *UBE2C* gene. *Top*: Numbers in parentheses show the coordinates of the genomic stretch sequenced; horizontal lines represent a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotides; *black*, methylated, *blue*, unmethylated, *red*, poor sequencing; *right*: the number following HCT116 and DKO2L shows the total percent methylation. We analyzed only those sequences where < 2% of non-CpG Cs were unconverted to uracil (thymidine) after bisulfite treatment.

Interferon-inducible genes over-express in genome-wide hypomethylated cells

One of the aims of our study was to identify the particular classes of genes whose expression is regulated by DNA methylation, and therefore get affected the most by genome-wide hypomethylation. Comparing gene expression profiles of DKO2L cells against HCT116 cells revealed that interferon-inducible genes are one of the most affected classes due to genome-wide hypomethylation in the DKO2L cells (Table 9).

Several of the interferon-inducible genes were upregulated in DKO2L cells from 2-fold to 250-fold, including a few of the most highly upregulated genes such as *IFI27* and *G1P3*. Other important genes of this class which are upregulated in DKO2L cells are *STAT1*, *STAT3* and *IRF7*. In total, 15 genes of this class were upregulated in DKO2L cells more than 2-fold as compared to HCT116. The expression changes of some of these genes were also independently verified by quantitative qPCR and Northern blotting analyses, which corroborated our longSAGE findings.

Table 9: Interferon-inducible genes are overexpressed due to genome-wide hypomethylation. Interferon-inducible genes are upregulated in genome-wide hypomethylated DKO cells. *Column headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; 5' CGI: presence of CpG island at the 5' sequence of the gene; CpG islands were identified using relevant track information from the UCSC genome browser.

Tag	Description	WT	DKO	DKO/WT	5' CGI
CCAGGGGAGAAGGCACC	IFI27 Interferon, alpha-inducible protein 27	4.3	1109.2	253.4	**No**
CTTGTAATCCTACTTGG	G1P3 Interferon, alpha-inducible protein (clone IFI-6-16)	4.3	186.5	42.6	Yes
ACCATTGGATTCATCCT	IFITM1 Interferon induced transmembrane protein 1 (9-27)	8.7	98.1	11.2	No
AATCTGCGCCTGCGGGG	G1P2 Interferon, alpha-inducible protein (clone IFI-15K)	139.8	922.7	6.6	Yes
ATTTAGTCATAATTGTG	IFI44 Interferon-induced protein 44	8.7	29.4	3.4	No
TGATTTAATTCTCCATT	IFIT5 Interferon-induced protein with tetratricopeptide repeats 5	8.7	29.4	3.4	Yes
GGGGGCAGATCCAGTCC	IRF7 Interferon regulatory factor 7	4.3	29.4	6.7	Yes
CCTGTAATCCCAGCACT	IFIT3 Interferon-induced protein with tetratricopeptide repeats 3	17.4	68.7	3.9	No
GCTCCAGCCATACTCCA	ISGF3G Interferon-stimulated transcription factor 3, gamma 48kDa	8.7	29.4	3.4	No
ACCTGTATCCCACGTAC	IFITM3 Interferon induced transmembrane protein 3 (1-8U)	4.3	19.6	4.5	No
ATGCCTTTTTATATTTA	BIK (BCL2-interacting killer)	4.3	19.6	4.49	Yes
GGCAGACTGGCAGAAGC	IFIT1 (Interferon-inducible protein with tetratricopeptide repeat 1	4.3	19.6	4.49	No
GCTTGCAAAAAGTAAAC	SOD2 (manganese superoxide dismutase 2	26.2	49	1.87	Yes
CAAATGCTGTATTCTTC	STAT1 (Signal transducer and activator of transcription 1)	34.9	157	4.49	Yes
TATCTGGTCTTAACTCT	STAT3 (Signal transducer and activator of transcription 3)	8.7	19.6	2.2	Yes

Cancer-Testis genes are overexpressed in hypomethylated cells

Comparison of the expression profiles of DKO2L and HCT116 cells generated by longSAGE shows that markedly increased expression of a large number of cancer-testis (CT) genes in the DKO2L cells. Several members of various cancer-testis gene families are upregulated in DKO2L cells to varying degrees, ranging from 2-fold to 190-fold (Table 10). Statistical analyses also showed that cancer-testis genes are enriched in DKO2L library. Out of 527 significantly upregulated genes in DKO2L library, 9 belong to cancer-testis gene families. Fisher's exact test shows that the p-value of this finding is 7.63*107, which is highly significant. Of a total of 89 annotated cancer-testis genes, there are 29 belonging to 10 different families that were upregulated at least 2-fold in the DKO2L library. We also observed a 31-fold upregulation of *BORIS* (an important epigenetic regulator (Loukinov et al., 2002) in DKO2L cells as compared to HCT116. Most of these genes are located on the X chromosome.

Upregulation of cancer-testis genes correlates with decreased promoter methylation

We carried out bisulfite sequencing at the promoters and/or 5' regions of two of the upregulated cancer-testis genes i.e. (*GAGED-2*, nt -198 to 103; and *GAGE1*, nt -17 to 359) to test the hypothesis that expression changes might be correlated with promoter methylation. Bisulfite sequencing results revealed that promoters of these genes were heavily methylated in HCT116 cells, and became profoundly hypomethylated in DKO2L cells. Considering all CpG dinucleotides in the amplicons tested, *GAGED-2* showed 79% methylation in the HCT116 cells and 5.6% in the DKO2L cells, while *GAGE1* showed 89% methylation in HCT116 cells and 13% in the DKO2L cells (Figures 13 and 14). The upregulation of these genes inversely correlates with the extent of methylation of their promoters, which suggests that transcription of cancer-testis genes is negatively regulated by promoter methylation in the HCT116 cell line.

Table 10: Expression status of cancer-testis genes in HCT116 and DKO2L longSAGE libraries. Cancer-testis genes are upregulated in genome-wide hypomethylated DKO cells. *Column Headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; Ch: chromosome of origin of the tag.

Tag	Description	WT	DKO	DKO/WT	Chr.
GAGE family					
CTGAAATGTTGCAGGCT	GAGE2, G antigen 1	4.3	69	15.7	X
GCTGATGTCACATTGAA	GAGEC1, P antigen family, member 4	4.3	19.6	4.49	X
XAGE family					
GAAGGTGATCTGCAAGA	GAGED2, X antigen family, member 1	4.3	834	191	X
TTCACTGGGCATCTTCC	GAGED3, X antigen family, member 2	4.3	9.8	2.2	X
MAGE Family					
AGGCCCATTCTTCACTC	MAGEA1, Melanoma antigen family A, 1	8.7	147	16.8	X
AGGCCCATTCTTCACTC	MAGEA4, Melanoma antigen family A, 4	8.7	147	16.8	X
TCAAGTTTTTGTTTTCT	MAGEB1, Melanoma antigen family B, 1	4.3	69	15.7	X
TTGTTTTGCTTATCCAT	MAGEA11, Melanoma antigen family A, 11	4.3	29.4	6.7	X
TGGCTGATACAAAGGTC	MAGED2, Melanoma antigen family D, 2	4.3	19.6	4.49	X
ATACAAGGAACTCAAAA	MAGEA2, Melanoma antigen family A, 2	4.3	19.6	4.49	X
ATACAAGGAACTCAAAA	MAGEA2, Melanoma antigen family A, 2B	4.3	19.6	4.49	X
AGATAACTCAAGAAATC	MAGEA6, Melanoma antigen family A, 6	43.6	147.2	3.37	X
GGGGACAAGGATATGCC	MAGEC1, Melanoma antigen family C, 1	4.3	9.8	2.2	X
CATTTTGTGGCAAGGGT	MAGEH1, Melanoma antigen family H, 1	4.3	9.8	2.2	X
CAAGATAAATTTTATTT	MAGED1, Melanoma antigen family D, 1	4.3	9.8	2.2	X
AGTGGGCTTTGAGAGAG	MAGEA6, Melanoma antigen family A, 6	4.3	9.8	2.2	X
CCTCGTGGTCAGAAGAG	MAGEB2, Melanoma antigen family B, 2	17.4	29.4	1.68	X
PAGE family					
TAACAAGATCCCAATCC	PAGE-5, P antigen family, member 5	4.3	39.2	8.9	X
cTAGE family					
TTTGGAGCTTCTCGAGA	CTAGE5, CTAGE family, member 5	4.3	19.6	4.4	18
BORIS family					
AGATTTAAATTCTGTGG	CTCFL, (BORIS)	4.3	137	31.4	20

Several embryonic genes are upregulated in hypomethylated cells

Embryonic genes are expressed during specific stages of normal embryogenesis and are silenced in adult animals. DNA methylation is known to play an important role in transcriptional regulation of embryonic genes (Dean et al., 2005; Eden and Cedar, 1994). We compared the expression profiles of DKO2L cells with HCT116 cells in order to test the hypothesis that developmental/embryonic genes would be deregulated due to genome-wide hypomethylation. Comparison showed that genome-wide hypomethylation in DKO2L affects the expression of several embryonic genes including POU5F (well-known pluripotency marker), TGIF and TNFRSF6**.** Many of the upregulated embryonic genes have CpG islands in their promoter/5' region, suggesting that DNA methylation regulates their transcription. A list of embryonic genes upregulated in DKO2L cells as compared to HCT116 is presented in (Table 11).

Table 11: Embryonic genes are upregulated in DKO2L cells. *Column Headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; 5' CGI: presence of CpG island at the 5' sequence of the gene; CpG islands were detected using tracks in UCSC genome browser.

Tag	Description	WT	DKO	Ratio DKO/WT	5' CGI
TGTGGGTTAAGCGGTTT	POU5F1, POU domain, class 5, transcription factor 1	4.3	29.4	6.7	No
TGGAACAGGATGCCCAC	TGIF, TGFB-induced factor homeobox 1	4.3	68.7	15.7	Yes
AGGGCGCTGGAGGGGCC	TNFRSF6, tumor necrosis factor receptor superfamily, member 6	4.3	19.6	4.49	—
TCCCTTGTCACTGTGGC	NOMO1, NODAL modulator 1	17.4	39.3	2.24	Yes
GGACGTCGTCCTGGTGG	NANOS1, nanos homolog 1	8.7	29.4	3.37	No
CCTAAACTCAAATAAAG	MDS032, uncharacterized hematopoietic stem/progenitor	17.4	49.1	2.8	Yes
ATCTACAAAACAAAATC	DPPA2, developmental pluripotency associated 2	4.3	19.6	4.49	No

MHC class I genes are over-expressed in genome-wide hypomethylated cells

Comparison of the expression profiles of HCT116 and DKO2L cells revealed that several of the MHC class I genes such as MHC-B, MHC-E and MHC-F are upregulated in DKO2L cells as compared to HCT116 (Table 12). Bioinformatic analysis showed that these genes have *bona-fide* CpG islands near their promoters/5' region, thus suggesting a possible role of DNA methylation in regulating expression of these genes.

A cluster of metallothionein genes upregulated in genome-wide hypomethylated cells

One of the aims of our study was to determine if there were localized clusters of genes which get coordinately deregulated upon genome-wide hypomethylation caused by genetic disruption of DNA methyltransferase genes in DKO2L cells. We compared the genes expression profiles of HCT116 and DKO2L cells as obtained by longSAGE and found that several genes e.g. MT1A, MT1L, and MTM from a cluster of metallothionein family genes present on chromosome 16q13 were upregulated in DKO2L cells as compared to the HCT116 cells (Table 13). Bioinformatic analyses revealed the presence of a classical CpG island near the promoter region of these genes suggesting a possible role of DNA methylation in regulation of transcription. The results showed that metallothionein family genes constitute a cluster of genes, which is upregulated in DKO2L cells due to genome-wide hypomethylation.

Table 12: Expression status MHC class I genes in HCT116 and DKO2L longSAGE libraries. MHC class I genes are upregulated in DKO2L cells. *Column Headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library, DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; 5' CGI: presence of CpG island at the 5' sequence of the gene. CpG islands were detected using tracks in UCSC genome browser.

Tag	Description	WT	DKO	Ratio DKO/WT	5' CGI
GTGCACTGAGCTGTAAC	MHC-A, major histocompatibility complex, class I, A	87.3	206	2.3	Yes
CTGACCTGTGTTTCCTC	MHC-B, major histocompatibility complex, class I, B	114	599	5.2	Yes
GTGCACTGAGCTGCAAC	MHC-C, major histocompatibility complex, class I, C	52.4	157	2.9	Yes
CCTGTAATCCCAGCACT	MHC-E, major histocompatibility complex, class I, E	17.4	68.7	3.9	Yes
TGCAGCACGAGGGGCTG	MHC-F, major histocompatibility complex, class I, F	26.2	88.3	3.3	Yes

Table 13: Expression status of metallothionein genes in HCT116 and DKO2L longSAGE libraries. Several metallothionein genes are upregulated in DKO2L cells. *Column Headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; 5' CGI: presence of CpG island at the 5' sequence of the gene. CpG islands were detected using tracks in UCSC genome browser

Tag	Description	WT	DKO	Ratio DKO/WT	5' CGI
AGCTGTGCCAAGTGTGC	MT1A Metallothionein 1A	4.3	186.5	42.6	Yes
GGCTGTGCCAAGTGTGC	MT1L Metallothionein 1L	4.3	88.3	20.2	Yes
GGCTGTGCCAAATGTGC	MTM Metallothionein M	4.3	9.8	2.2	Yes
GATCCCAACTGCTCCTG	MT2A Metallothionein 2A	131	157	1.19	Yes
TACAAACCTGGATTTTT	MT1F Metallothionein	4.3	9.8	2.2	Yes
GGGCTTCTGTTCCTCTG	MT1E Metallothionein 1E	4.3	9.8	2.2	Yes

Genome-wide hypomethylation has minimal or no effect on expression of tumor suppressor genes in DKO2L cells

To study the effect of genome-wide hypomethylation on tumor suppressor genes that are silenced in most tumor cell lines such as HCT116 (Baylin and Bestor, 2002; Baylin et al., 2001), we compared the expression profiles of DKO2L and HCT116 cells by longSAGE and looked for the predicted longSAGE tags of some of the well-known tumor suppressor genes. To our surprise, we did not detect tags corresponding to most of the tumor suppressor genes in either library. Those that were detected showed a minimal increase in expression in DKO2L cells and compared to HCT116 e.g. MTS1 and STK11 (Table 14).

Genome-wide hypomethylation has no effect on expression of imprinted genes in DKO2L cells

Imprinted genes are mono-allelically expressed depending on the parent of origin, due to methylation-induced silencing of one of the alleles. We did not detect tags corresponding to many known imprinted genes and, moreover, those that were detected had no differential expression (Table 15).

Table 14: Expression status of some of the tumor suppressor genes in HCT116 and DKO2L longSAGE libraries. Tumor suppressor genes minimally affected in DKO2L cells. *Column Headers* Gene: the tumor suppressor gene; tag: 3' tag corresponding of the gene; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT.

GENE	TAG	WT	DKO	DKO/WT
ATM	CCATTGCACTCCAGCCT	35	49	1.04
CDKN2A/p16	ACAAGCATTTTGTGAAC	17.5	19.6	1.12
MLH1	CTTGCCTTAGATAGTCC	4.3	9.8	2.2
MSH2	GGCTATCAACTTAATAA	8.7	4.9	0.56
MTS1	ATGTGTAACGAATTCTT	8.7	68.7	7.86
NF2	AGTGGCTGTGTCTTGTA	8.7	4.9	0.56
P53	TTTTGTAGAGATGGGGT	17.4	4.9	0.28
PJ(STK11)	CAGCGCCACCTGGAAGC	174.4	314.1	1.79
PTEN	TATATACCTTTTTGTGT	8.7	4.9	0.56
RB	AATATCATACAAATCAG	8.7	4.9	0.56
SMAD4	TGGGTGAGTTAATTTTA	8.7	4.9	0.56
WRN	TAAAAAATGTAAAATGG	61.16	19.6	0.32
p15	AGTCCTGCTTCTAGCTC	8.7	9.8	1.123

Table 15: Expression status of some of the imprinted genes in HCT116 and DKO2L longSAGE libraries. Expression of imprinted genes is not affected by genome-wide hypomethylation in DKO2L cells. *Column Headers* Gene: the imprinted gene; tag: 3' tag corresponding of the gene; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT; NF: tag not found.

GENE	TAG	WT	DKO	DKO/WT
CDKN1C	TAGCAGCAACCGGCGGC	8.7	9.8	1.12
H19	AAAGAAATGGTGCTACC	4.3	9.8	2.24
IGF2R	AGTGTATTTTTTAAAAT	4.3	9.8	2.24
SLC22A18	CTGGGCCTCTGCGCCTC	4.3	9.8	2.24
ZNF215	GCCCTGTACTGGAATAA	8.9	9.8	1.12

Genomic hypomethylation in DKO2L cells leads to decreased expression of *MET* proto-oncogene

We studied the effect of genome-wide hypomethylation on the expression of the *MET* proto-oncogene, by comparing the tag expression level of the 3' end longSAGE tag of *MET*. We observed that the expression of this gene decreased approximately 4-fold in DKO2L cells as compared to the WT HCT116 cells (table 16). This result corroborates previously established downregulation of this gene, which is correlated with upregulation of an internal fusion transcript between an intronic L1 retrotransposon and the downstream exons (Weber et al., 2010)

Table 16: Expression status of *MET* proto-oncogene in HCT116 and DKO2L longSAGE libraries. The expression of *MET* proto-oncogene was decreased due to genomic hypomethylation. *Column Headers* Tag: 3' tag corresponding of the gene; description: description of the gene corresponding to that tag; Cytoband: the Cytoband corresponding to the gene; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT.

Tag	Description	Cytoband	WT	DKO	Ratio DKO/WT
CAATTAAAGTAAAGTGA	Met proto-oncogene (hepatocyte growth factor receptor)	7q31	33.8	9.3	0.3

Toward a transcriptome signature of genomic hypomethylation

We sought to identify a "transcriptome signature" of genome-wide hypomethylation. Strong correlations have been drawn between interspersed retrotransposon hypomethylation and genome-wide hypomethylation (Yang et al., 2004), but as described below we found that transposon expression is only minimally upregulated despite extensive genomic hypomethylation (Chapter 4). Therefore we surveyed other classes of genes to comprise this transcriptome signature. We chose to define "signature transcripts" as that set of tags/genes that show significant upregulation in DKO2L cells and are either not expressed or poorly expressed in other (normally methylated) cell types. Ideally, other extensively hypomethylated cells would have similarly upregulated transcripts. We used publicly available SAGE libraries data on SAGEmap (NCBI) to compare the abundance of a potential signature tag in those libraries. We required that each signature tag must have at least 5-fold higher expression than the maximum normalized expression level observed in any other public SAGE library.

We identified a set of tags that fit this criterion for signature tags, and compared the relative abundance of the same tags across all the longSAGE libraries. The maximum and average expression level of theses tags were calculated using the tag data available on the NCBI SAGEmap database. We developed a candidate list of eleven tags, which constitutes a preliminary transcriptome signature of DKO2L cells (Table 17). Most genes represented by these tags have a CpG island near or at their promoters, and are either not expressed or poorly expressed in normal colon tissue.

Table 17: Expression status of the candidate tags constituting "SAGE-tag signature" of DKO2L cells in HCT116, DKO2L and other publicly available longSAGE libraries. *Column Headers* Tag: upregulated tag; description: description of the gene corresponding to that tag; WT: normalized tag count (tags per million) in HCT116 library; DKO: normalized tag count (tags per million) in DKO2L library; RatioDKO/WT: ratio of DKO to WT, Long TPM: average count of the tags in all longSAGE libraries on NCBI portal; Our/Lavg: ratio of tag count in DKO library and normalized average count (per million) of the tags in all longSAGE libraries on NCBI portal; Ex. in colon: expression status in normal colon; CGI: presence of CpG island at the 5' sequence of the gene; CpG island were detected using tracks in UCSC genome browser; Ch: chromosome of origin of the tag.

Tag	Description	WT	DKO	DKO /WT	Long TPM	Our/ Lavg	Ex. In Colon	CGI	Ch.
GAAGGTGATCTGCAAGA	GAGED-2 X, antigen family, member 1	4.3	834	191	1.18	698	low	Yes	X
AGATTTAAATTCTGTGG	CTCFL, CCCTC-binding factor (zinc finger protein)-like	4.3	137.4	31.4	0.27	563	low	Yes	20
AGGCCCATTCTTCACTC	MAGEA4/MAGEA1,	8.7	147.2	16.8	0.27	545	not exp	No	X
CAGCCTGGGGCCACTGC	PRSS21, protease, serine, 21 (testisin)	4.3	480	110	2.4	228	low	Yes	16
TCCCCAGCTCTGGGAGG	OXCT2, 3-oxoacid CoA transferase 2	4.3	58.8	13.4	1.62	36	not exp	Yes	1
GGCTGTGCCAAGTGTGC	MT1L, metallothionein 1L	4.3	88.3	20.22	2.7	33	low	Yes	16
CCGTGGCTGGCGCTGCC	NXN, nucleoredoxin	4.3	68.7	15.7	2.16	31	mod/high	Yes	17
AGCTGTGCCAAGTGTGC	MT1S, metallothionein 1S	4.3	186.5	42.5	6.2	19.67	low	Yes	16
CTGAAATGTTGCAGGCT	GAGE2, G antigen 2	4.3	68.7	15.72	0	∞	not exp	Yes	X
TCAAGTTTTTGTTTTCT	MAGEB1, melanoma antigen family B, 1	4.3	68.7	15.7	0	∞	not exp	Yes	X
ACCCGCGTGCTGCAGGG	CGB5, chorionic gonadotropin, beta polypeptide 5	4.3	78.5	17.95	13	6	not exp	No	19

We compared the transcriptomes of HCT116 cells and DKO2L cells with those of various other cell types (including mouse lines) with genome-wide hypomethylation due to either pharmacological or genetic manipulations (Coral et al., 2002; Ghoshal et al., 2000; Gibbons et al., 2000; Gius et al., 2004; Guillaudeux et al., 1996; Jackson-Grusby et al., 2001; James et al., 2006; Koslowski et al., 2004; Liang et al., 2002; Majumder et al., 1999; Nie et al., 2001). Comparing the raw and processed data from these gene expression profiling studies with our SAGE findings revealed that there were few classes of genes that were commonly upregulated in genome-wide hypomethylated cells in all these studies (Table 18). These classes included interferon-inducible genes, cancer-testis genes, MHC class I genes, metallothionein genes and some of the embryonic genes. Most of the genes belonging to these classes have classic CpG islands in their promoter/5'region; promoter methylation-induced silencing of many of these genes has been independently shown in several previous studies (Coral et al., 2002; Ghoshal et al., 2000; Gibbons et al., 2000; Gius et al., 2004; Guillaudeux et al., 1996; Jackson-Grusby et al., 2001; James et al., 2006; Koslowski et al., 2004; Liang et al., 2002; Majumder et al., 1999; Nie et al., 2001). Also, in mammals, cancer-testis genes and embryonic genes are the classes of genes which are known to be expressed only during normal developmental processes like gametogenesis (Suri, 2006) and embryogenesis (Eden and Cedar, 1994; Kafri et al., 1992) when the total genomic methylation level of the cells is much lower than the somatic cells. These results show that coordinated expression of these classes of genes could indicate the genome-wide, hypomethylated state of a cell. These arguments suggest that a "transcriptome-signature" of genome-wide hypomethylated cells can be identified and refined (Table 18).

Table 18: Towards a transcriptome signature of genomic hypomethylation. Classes of genes upregulated in genome-wide hypomethylated cells (as found in previous studies). IFN: interferon; CT: cancer-testis; MT: metallothionein; MHC: major histocompatibility complex. "Yes: indicates upregulation in genome-wide hypomethylated cells; "No" indicates either the study did not look for these classes of genes or they could not find upregulation in genome-wide hypomethylated cells.

Studies	Cells used	Technique used	IFN genes	CT genes	MT genes	MHC class I genes
Our longSAGE analyses	HCT116 & DKO	longSAGE	Yes	Yes	Yes	Yes
Gius et al (2004)	HCT116, HCT116- Dnmt1-/- & DKO	cDNA microarray	Yes	Yes	Yes	Yes
Jackson-Grusby et al (2001)	MEF (Dnmt1-/-)	Oligonucleotide microarray	Yes	No	No	Yes
Liang et al (2002)	T24 & LD419	Oligonucleotide microarray	Yes	Yes	No	No
Koslowaski et al (2004)	HCT116 & DKO		No	Yes	No	No
James et al (2006)	HCT116 & DKO		No	Yes	No	No
Coral et al (2002)	human renal cell carcinoma	RT-PCT	No	Yes	No	No
Majumder et al (1999)	(Ku-80)	RT-PCR	No	No	Yes	No
Ghoshal et al (2000 and 2002)	rat hepatoma	RT-PCR	No	No	Yes	No
Nei et al (2001)	human esophageal squamous cell carcinomas (ESCC)	RT-PCR	No	No	No	Yes
Guillaudeux et al (1996)	meiotic pachytene spermatocytes and postmeiotic round spermatids	RT-PCR	No	No	No	Yes

DISCUSSION

The recent development of cultured human colorectal cancer cell derivatives that continuously lack normal expression of DNA methyltransferases has provided a valuable, albeit somewhat flawed, tool to study the role of DNA methylation in the transcriptional regulation of genes. Genetic disruption of the major maintenance methyltransferase, *DNMT1*, reduced genomic methylation by 20% (Rhee et al., 2000). By contrast, the independent disruption of *DNMT3B*, a *de novo* methyltransferase, by itself led to a 3% decrease in methylation (Rhee et al., 2002). Disruption of both methyltransferases caused 95% reduction in genomic methylation and nearly abolished methyltransferase enzymatic activity. We note that the cell lines used in this study undoubtedly were not isogenic after unknown numbers of passages, due to ongoing chromosomal instability particularly in DKO cells (Karpf and Matsui, 2005). Moreover, the knockout cells possess residual hypomorphic and truncated expression of *DNMT1* (Egger et al., 2006).

In the present study, we extensively analyzed the transcriptomes of HCT116 and DKO2L cells by comparing Serial Analysis of Gene Expression (longSAGE) libraries both with previously published cDNA microarray data (Gius et al., 2004) and with current exon microarray assays of total RNA. We tested the hypothesis that disruption of methyltransferase activity, leading to profound decreases in genomic methylation, would result in pronounced differences in transcript levels, particularly of interspersed retrotransposons whose methylation status have been used as a surrogate for such genome-wide methylation changes (Yang et al., 2004). Our aim has been to identify differentially expressed genes and genetic loci in the two related cell lineages, and to study the methylation at

those genes/loci to find any correlations between differential transcription and underlying methylation status.

We observed profound differences between the two cell lines' transcriptomes, as measured by longSAGE library tag counts. Indeed, hundreds of genes are differentially expressed in DKO2L cells compared with parental HCT116 cells (Table 6, Fig. 6), as reflected by a low correlation coefficient between these related cell lines. The upregulated genes include interferon-inducible genes, cancer testis genes, several embryonic genes, HLA genes and metallothionein genes, while the downregulated genes include several ribosomal protein genes, RNA processing and RNA metabolism genes (Tables 9-13). We utilized Database for Annotation, Visualization and Integrated Discovery-2006 (DAVID-2006, NIH) bioinformatics resources to categorize the affected genes comprehensively into various pathways, and biological and molecular functions to facilitate our understanding of the biological meaning of these findings. In addition, we surveyed relationships to physical locations on cytobands and chromosomes. In general, those genes involved in negative regulation of biological and cellular processes, e.g. DNA damage response genes, are amongst the most highly upregulated genes in the DKO2L cells. This general finding could be due to the fact that DKO2L cells show a much slower growth rate (Rhee et al., 2002) and have a high level of genomic stability and DNA damage (Karpf and Matsui, 2005) as compared to HCT116 cells. Genes related to biosynthesis, cellular physiology, RNA metabolism and processing, and translation are amongst the downregulated genes. This finding could be due to slower growth, lower protein synthesis and metabolism rates in DKO2L cells as compared to HCT116.

Two independent techniques for expression profiling, i.e. Northern blotting and qRT-PCR, corroborated the most highly upregulated genes that were identified initially by our longSAGE findings. In addition, exon microarrays corroborated a very large number of highly differentially expressed genes identified first by longSAGE (Figures 8 and 9). All tested genes showed significant upregulation in DKO2L cells when compared to HCT116. In addition, similar upregulation was observed in several independent DKO clones (Figures 11) (Rhee et al., 2002). In contrast to comparisons with previously published cDNA microarray data, which mostly missed the most upregulated genes identified here, these consistent results using a variety of techniques indicate that genomic hypomethylation profoundly disrupts the human transcriptome in specific ways.

Several studies have shown previously that epigenomic reactivation by genetic manipulation or drug treatment deregulates a large number of genes (Karpf, 2007). However, the deregulation of many of the affected genes could be due to an indirect effect, i.e. mediated by factors in *trans* rather than by direct changes in promoter methylation in *cis*. To establish a correlation between transcriptome changes and promoter methylation changes, we carried out methylation analysis at the promoters of genes that are most upregulated in DKO2L cells. Bisulfite sequencing analysis was chosen for these methylation studies because it is highly quantitative, and provides high resolution analysis of several individual CpG sites simultaneously in one bisulfite PCR amplicon. As expected, comparative bisulfite sequencing of the promoters of highly upregulated genes generally showed an inverse correlation between changes in their methylation and changes in their expression (Figures 12 - 14). All genes analyzed

by bisulfite sequencing showed heavy methylation of the promoter in the HCT116 cells that became significantly hypomethylated in DKO2L cells. This result suggests that a major portion of differential gene expression in DKO cells is attributable to hypomethylation *in cis*. However, a few genes were upregulated in DKO2L cells did not appear to have such an inverse correlation with their promoter methylation status. This result suggests that while upregulation of most affected genes resulted directly from promoter demethylation, there could be some indirectly affected genes whose expression was changed due to some other transcriptional control factors such as histone tail modifications or the presence of a crucial transcription factor which may be directly or indirectly regulated by methylation. Also, it is possible that in certain cases, a predicted promoter region is not the actual promoter for a particular gene, and that another cryptic promoter located elsewhere could affect the expression of these genes. Nonetheless, methylation analyses of the highly upregulated genes suggest a usual pattern of negative transcriptional regulation by promoter methylation that fits well with the classical view of promoter methylation as a repressor of transcription (Bestor, 1998; Bird, 1992; Cedar and Razin, 1990).

We found that interferon-inducible genes are one of the most affected classes of genes, including the most highly upregulated gene, *IFI27* (Table 9). In total, 15 genes of this class are significantly upregulated. Several of these genes have CpG islands comprising their promoters. This result corroborates several prior studies in a variety of cell types using microarrays, which documented activation of interferon-inducible genes in response to genome-wide hypomethylation caused by pharmacological treatments or by genetic disruptions

(Jackson-Grusby et al., 2001; Karpf et al., 1999; Liang et al., 2002; Sato et al., 2003). Intriguingly, several interferon alpha-inducible genes also can be activated by expression of double-stranded RNA (Braganca and Civas, 1998).

While recent work has demonstrated that miRNAs are induced in DKO cells (Bostick et al., 2007; Yan et al., 2011), more studies are needed to investigate the possibility that double stranded or antisense RNAs (Yu et al., 2008) also might be upregulated upon genomic hypomethylation.

Our results also corroborate previous findings that other classes of genes are upregulated in the context of genomic hypomethylation, including cancer testis (CT) genes (De Smet et al., 1999; Gure et al., 2002; James et al., 2006; Koslowski et al., 2004; Suri, 2006; Weber et al., 1994; Akers et al., 2010), BORIS (Hong et al., 2005), embryonic genes, metallothionein genes clustered at chromosome 16q13 (Ghoshal et al., 2002; Ghoshal et al., 2000; Gius et al., 2004; Majumder et al., 1999), and MHC class I genes (Gius et al., 2004; Guillaudeux et al., 1996; Nie et al., 2001; Serrano et al., 2001).

While the transcriptional repression of many tumor suppressor genes in cancers associated with localized hypermethylation at their promoters, *e.g. p16*, *Rb*, *MLH1*, *RASSF1*, *VHL*, etc., we did not detect significant upregulation of any of them (Table 14). Their persistent silencing could be mediated by the residual, truncated DNMT1 expressed in the DKO cells, or by repressive histone modifications and chromatin condensation (Bachman et al., 2003; Egger et al., 2006; McGarvey et al., 2006). Vatolin et al. suggested that sustained ectopic expression of BORIS can cause hypermethylation at several CTCF/BORIS-

binding regulatory sequences at the promoters of various tumor suppressor genes (Hong et al., 2005). We observed a 31-fold upregulation of BORIS in DKO2L cells as compared to HCT116 (Table 10), suggesting that BORIS could play a role in persistent silencing of tumor suppressor genes in DKO2L cells. Another possibility is that the colorectal cells may not express required tissue-specific transcription factors to upregulate such TSGs.

Loss of imprinting has been observed in a wide range of cancers (Cui et al., 2002). DNA methylation is a major mechanism implicated in the maintenance of imprinting, implying any faulty methylation in cells might cause loss of imprinting. Rhee et al. showed that imprinting of *IGF2* is disrupted in DKO cells, by identifying its biallelic expression (Rhee et al., 2002). However, we observed no effect on the expression of imprinted genes in DKO2L vs. HCT116 cells (Table 15). One possible reason could be that their transcript levels are below the limit of detection in our longSAGE libraries, despite relatively deep sequencing.

Following an "epigenomic reactivation strategy" (Karpf, 2007), many groups have tried to establish correlations between various alterations in epigenetic controls, induced by pharmacological agents or genetic manipulations, and subsequent changes in the transcriptome. The idea is that we thereby can identify epigenetic controls that are aberrant in cancers, and can identify how cancer cells respond to "resetting" these controls.

We also observed that there could be effects of genomic hypomethylation on the expression of genes even if their *bona fide* promoter is unmethylated. Such effects could be due to the induction of alternate transcripts, fusion transcripts

and/or other cryptic promoters which originate from intronic retrotransposons (sense or antisense promoters). One such example is the paradoxical downregulation of the *MET* proto-oncogene in DKO cells. This has been associated with the hypomethylation-dependent induction of an illegitimate transcript from an antisense promoter of a L1 retrotransposon located in the intronic region of this gene, thus giving rise to the fusion transcripts L1-ASP (Table 16). The exact molecular mechanism by which this fusion transcript is linked to downregulated *MET* expression remains unclear.

Statistical comparisons between our longSAGE study and a previous microarray-based study of the same HCT116 cells and their derivatives (Gius et al., 2004) revealed a relatively poor overall correlation ($r^2 = 0.1$). Some possible explanations for this striking discrepancy between the data sets include differences in the DKO clones or passage numbers (Bachman et al., 2003; Rhee et al., 2002) used for RNA extractions, and/or fundamental differences in the sensitivity and specificity determined by different techniques and platforms used in the studies (Bourc'his and Bestor, 2004). Similarly, there is a similar overall lack of correlation between our findings and other recent studies of transcriptome variation (Jackson-Grusby et al., 2001; Liang et al., 2002). This could be attributed to highly divergent cell types used in these studies, i.e. mouse embryonic fibroblasts vs. a cultured human colorectal cancer cell line. Nevertheless, we validated a very high percentage of the most upregulated longSAGE tags observed in DKO2L cells, by exon microarray, Northern blotting and/or qRT-PCR (Figures 8-11), and verified that the most upregulated transcripts are similarly overexpressed in several, independently derived DKO clones (Figure 11).

Using publicly available transcriptome (SAGE) data and comparing our findings with previous studies of induced hypomethylation, we compiled a set of "signature tags" which may well characterize differential gene expression in the context of genome-wide hypomethylation in human colorectal cells (Table 17). Strikingly, this list does not include tags representing transposons, since despite extensive hypomethylation of those widespread elements, we did not observe substantial upregulation of them (Chapter 4). In our transcriptome signature, which includes interferon-inducible genes, cancer-testis genes, metallothionein gene cluster and MHC class I genes (Tables 18), most genes represented by tags had a corresponding CpG island at or near their promoters, and all are either not expressed or poorly expressed in normal colon tissue. Most of these genes have a testis-restricted expression pattern and significant numbers of these genes are present on X-chromosome and belong to the CT gene family. Together with extensive previous results, our longSAGE data suggest that upregulation of these normally or developmentally restricted classes of genes could reflect genome-wide hypomethylation (Coral et al., 2002; Eden and Cedar, 1994; Ghoshal et al., 2000; Gibbons et al., 2000; Gius et al., 2004; Guillaudeux et al., 1996; Jackson-Grusby et al., 2001; James et al., 2006; Kafri et al., 1992; Koslowski et al., 2004; Liang et al., 2002; Majumder et al., 1999; Nie et al., 2001; Suri, 2006).

An approach to refine and improve this proposed transcriptome signature of genomic hypomethylation in cultured human colorectal cancer cells (Table 18) would be to re-introduce *DNMT1* and *DNMT3B* genes into DKO2L cells, to determine if expression of members of the transcriptome signature returns back to expression levels in the parental cells. Of course, this assumes that karotypic

instability in the DKO cells does not preclude reestablishment of "wildtype" expression patterns. In future experiments, we will attempt to measure the transcriptomes comprehensively in additional clonal cell isolates of HCT116 lacking DNA methyltransferases accomplished either by genetic knockout or knockdown by RNA interference; to use even more comprehensive expression profiling platforms such as RNA-Seq (Allegrucci et al., 2005; Mortazavi et al., 2008); and/or to comparatively study the transcriptome in other hypomethylated cell lines derived from colorectal tumors or other tissues.

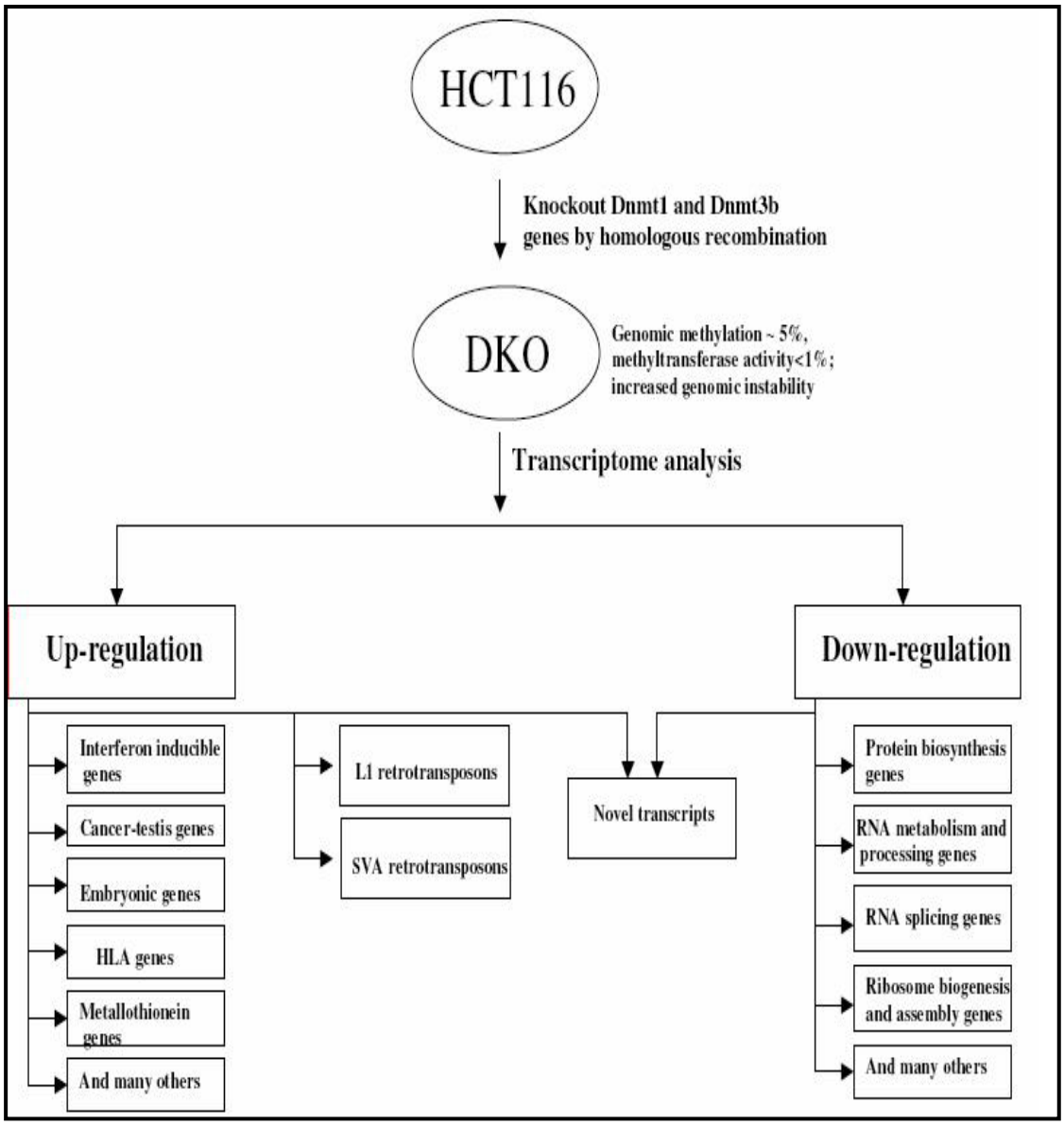

Figure 16: Schematic representation of the effects of genome-wide hypomethylation on the transcriptome

FUTURE SCOPE OF WORK

1. **Correlate the transcriptome changes seen in DKO2L cells with its methylome.** Methylated genomic DNA will be isolated from HCT116 and DKO2L. Library will be made from this fragments of the methylated genomic DNA and will be sequenced using deep sequencing platforms such as 454, Illumina, SoLiD or Ion Torrent. This would let us identify the sequences that are specifically methylated in HCT116 and DKO2L cells, which could be mapped to their genomic loci. We have already mapped the significantly deregulated tags to its genomic loci. Next we would correlate the significantly deregulated longSAGE tags loci with the loci with the mapped methylome of the corresponding cells. This would give us a comprehensive correlation between the transcriptome and the methylome of HCT116 and DKO2L cells.

2. **Reintroduce *DNMT1* and *DNMT3b* genes into DKO2L cells**

   - To study the remethylation pattern in DKO 2L cells (if it resembles HCT116 pattern)
   - To check the expression status of those genes which were most affected by the genome-wide hypomethylation in DKO2L cells and to determine if their expression level is brought to the HCT116 levels.
   - To study the promoter of those genes which were most affected by the genome-wide hypomethylation in DKO2L cells and determine if the resetting of the expression status is accompanied with any change promoter methylation change.
   - To study how does different level of DNMTs expression affect the resetting of methylation marks

3. **To construct reverse SAGE library from HCT116 and DKO2L cells**

   - PCR amplify the transcripts corresponding to the apparent novel tags and confirm the validity of the identified novel tags and characterize the transcript.

4. Compare the effects of genome-wide hypomethylation on the transcriptome in somatic cells vs. embryonic stem cells.

5. Assess transcriptome with other comprehensive high-throughput methods including RNA-Seq, GIS, etc.

Transposable elements

Although previously considered as "junk" DNA, mammalian genomic transposable elements play many possible biological roles that recently have become more clearly recognized. The human and mouse genomes each contain an enormous number of transposable elements, accounting for 45% of genomic content (Jackson-Grusby et al., 2001). These are broadly divided into four classes, namely DNA transposons, Long interspersed elements (LINE), short interspersed elements (SINE) and long terminal repeat-containing (LTR) retrotransposons (Figure 17). Retrotransposons transpose via RNA intermediates (Akagi et al., ; Ostertag and Kazazian, 2001). Most of these elements have accumulated mutations in their sequences and are therefore incapable of moving in the genome. However, active elements that are capable of mobilization still remain in the genome. LINE-1 (L1) retrotransposons are the most abundant and oldest, comprising approximately 17% of the genome. These elements are the most active in mouse (Akagi et al., 2008). Alu elements (SINEs) are most active in human (Jackson-Grusby et al., 2001) and utilize L1 machinery for mobilization. There is a controversy over whether or not HERV-K elements have been mobile recently in the human genome, although mouse ERVs remain very active (Contreras-Galindo et al., 2008). Rampant retrotransposition events could lead to genomic instability (Gilbert et al., 2002; Gilbert et al., 2005; Symer et al., 2002), insertion mutation and interference with transcription of adjoining genes (Weber et al.).

L1 are autonomous, mobile retrotransposons of about 6 kb in length (Figure 17) (Jackson-Grusby et al., 2001; Ostertag and Kazazian, 2001). Although L1 elements are interspersed throughout the genome, they are particularly abundant in AT-rich, gene-poor regions corresponding to G-bands (Korenberg and Rykowski,

1988). L1 elements have an unusually high density on chromosome X (29% of chromosome content) as compared to about 17% of the total genome (Martens et al., 2005). There are about 450,000 L1 elements in the genome categorized into different families; most of the active human L1s belong to Ta (Hs) family. Full-length L1s consist of a 5'UTR containing an internal promoter (Swergold, 1990), two in-frame open reading frames (ORF1 and ORF2) separated by a 63 bp non-coding spacer required for retrotransposition (Chen et al., 2001; Esnault et al., 2000; Feng et al., 1996) and a 3'UTR ending in polyadenylation signal. ORF1 encodes a 40-kDa protein with RNA binding and nucleic acid chaperone activities in vitro (Guy et al., 2001; Kolosha and Martin, 1997, 2003). ORF2 encodes three distinct conserved domains, i.e. an N-terminal endonuclease domain, central reverse transcriptase domain and a C-terminal zinc knuckle-like domain (Fanning and Singer, 1987; Feng et al., 1996; Mathias et al., 1991). L1 is thought to move in the genome by target-primed reverse transcription mechanism (Cost et al., 2002).

SINE (Short Interspersed Element) are the second most abundant retrotransposons comprising 13% of the genomic content and are short (100-400 bp) in length (figure 17). These elements have an internal RNA polymerase III promoter, do not encode for any protein and require L1 machinery in trans for their movement. Alu elements are the most numerous (~1,000,000 copies) and only active family of this class comprising 10% of the genome, are approximately 300 bp in length (Jackson-Grusby et al., 2001). Alu elements have high density of CpG dinucleotides and are highly methylated in somatic tissues.

SVA is a hominid-specific, non-autonomous, composite and youngest of all retrotransposon families. Its components are SINE-R, VNTR and Alu. There are more than 2,500 SVA elements identified in human genome. They are enriched in G+C rich regions. SVA elements are classified into 6 sub-families (SVA-A to SVA-F) (Ross et al., 2005). SVA elements have evolved recently, which is apparent from lack of high level of sequence divergence. Movement of SVA element is facilitated by L1 retrotransposons *in trans* (Hancks et al., 2011). SVA elements are highly methylated in all somatic tissues of adult (Strichman-Almashanu et al., 2002)

LTR (Long Terminal Repeat) retrotransposons are autonomous retrotransposons comprising about 8.3% of the genome (Figure 17). These elements have long terminal repeats at 3' and 5' ends containing the required transcriptional regulatory sequences; between the LTRs, these elements have gag and pol genes encoding protease, reverse transcriptase, RNaseH and integrase. The endogenous retrovirus-K (ERV-K) family of LTR class is a active families and has about 8,000 copies in the mammalian genome (Jackson-Grusby et al., 2001).

Retrotransposition events in the mammalian genome can have several deleterious effects. L1 movement in the genome can promote unequal homologous recombination and/or insertion into genes, thus affecting normal transcription (Takahara et al., 1996). During the retrotransposition process, two single stranded breaks that are created close to each other could act as a double stranded break, thereby increasing the chances of chromosomal breakage, deletion, translocation and illegitimate recombination (Gilbert et al., 2002; Gilbert et al., 2005; Symer et

al., 2002). Once the retrotransposition event has taken place it could cause a variety of transcriptional deregulation of the neighboring genes or transcription units depending upon the context and orientation of the new insertions. There are several documented cases of diseases caused by insertion of L1 elements (Ostertag and Kazazian, 2001). Also, L1 provides the necessary machinery for movement of other non-autonomous elements (Dewannieux et al., 2003; Ostertag et al., 2003), Alu elements are known to be active and require L1 machinery to move. There are several documented cases of Alu insertions causing human diseases (Ostertag and Kazazian, 2001). SVA elements are one of the least studied retrotransposons, however there are at least three-documented cases of SVA insertion-mediated human diseases (Ross et al., 2005). Further, L1s can also give rise to novel genes through shuffling by 3' transduction (Moran et al., 1999). Recent studies suggest that, although previously less emphasized, repetitive elements (retrotransposons) are commonly expressed in a highly tissue-specific manner (especially embryonic tissues), mainly utilizing a previously unknown sense and/or antisense promoter. Retrotransposons which are near to the 5' end of a protein coding region may act as an alternate promoter that may express alternative mRNA and other non-coding RNAs, thus regulating the nearby genes and altering the epigenome (Faulkner et al., 2009).

Given the deleterious potential consequences of retrotransposon movement, it is surprising to know that relatively a low number of mutations and other harmful effects actually happen due to their movement. Does this suggest that the genome has some kind of defense system that checks the movement of these elements?

Most of the L1 elements in the genome are defective due to 5' truncations, point mutations and inversions, thereby leaving them incapable of moving (Ostertag and Kazazian, 2001). There are about 3,000-5,000 full length L1 elements residing in the human genome (Grimaldi et al., 1984) of which only about 80-100 are considered capable of actively moving in the genome (Kazazian, 2004; Sassaman et al., 1997). Bestor et al proposed that cytosine methylation may serve as a host genome-defense system that helps check the expression of these elements (Bestor and Tycko, 1996; Meehan et al., 1992) by silencing the retrotransposons. L1s are generally silenced except in germ cells and during embryonic development (Branciforte and Martin, 1994; van den Hurk et al., 2007). It is well known that endogenous L1 and other repetitive elements are highly methylated in somatic cells, which is responsible for keeping these elements in a silent state (Bestor and Tycko, 1996; Meehan et al., 1992). Any decrease in methylation at these transposable elements increases the risk of their transcription and movement in the genome. One of the initial publications showed using oligonucleotide microarray that genome-wide hypomethylation in mouse embryonic fibroblasts (with disrupted *Dnmt1*) caused increased expression of a particular L1 element (L1Md-Tf14) (Jackson-Grusby et al., 2001). It was shown that in mouse germ cells, disruption of *Dnmt3L* prevents *de novo* methylation of non-LTR and LTR retrotransposons caused high expression of these elements in spermatogonia and spermatocytes (Bourc'his and Bestor, 2004). In a recent study, cancer-specific chimeric transcripts were isolated in cells where L1 retrotransposons were hypomethylated, leading to genomic instability and making them susceptible to cancer progression (Cruickshanks and Tufarelli, 2009).

Another study by Rangwala et al has shown that many L1 elements are expressed in human somatic cells, thus significantly contributing to the transcriptome (Rangwala et al., 2009).

RNA interference (RNAi) due to antisense promoter activity is also thought to play a modest role in regulating expression of human L1 retrotransposons (Soifer et al., 2005; Yang and Kazazian, 2006). Further, it has been suggested that Miwi proteins which interact with small RNAs called piRNAs play a role in regulating expression of L1s (Carmell et al., 2007; O'Donnell and Boeke, 2007); Mili mutant mice testis show expression of L1 and IAP elements. Interestingly, they also have decreased methylation at L1 elements (Aravin et al., 2007). Additional cellular inhibitors involved in checking L1 expression are members of the APOBEC3 protein family (Bogerd et al., 2006; Muckenfuss et al., 2006), which appear to inhibit L1 movement without editing new integrant sequences.

There is evidence for DNA methylation playing a role in regulating the expression of HERVs. One study showed that treating Tera-1 cells with 5-azacytidine increased the expression of HERV-K(HML-2) Gag protein (Gotzinger et al., 1996). Another study on Tera-1 cells supported CpG methylation as an important factor in silencing these elements. However, it was also suggested that CpG methylation is not the only factor needed for silencing the HERV promoter (Lavie et al., 2005). In mice it was shown that disruption of *Dnmt1* causes increased IAP expression (one of the ERV LTR retrotransposon family) (Walsh et al., 1998). Oligonucleotide microarray analysis on Dnmt1-disrupted, p53-

inactivated MEFs showed increased expression of IAP elements (Jackson-Grusby et al., 2001).

To our knowledge, no previous study has used genome-wide expression profiling either by microarray or sequencing based techniques to specifically look into the effects of genome-wide hypomethylation on the expression of transposable elements. In this study, we utilized gene expression profiles of HCT116 and DKO2L cells generated by longSAGE to compare the transcriptomes of HCT16 and DKO2L cells and to investigate the effects of genome-wide hypomethylation on the transcriptional regulation of L1 retrotransposons and other transposable elements such as Alu, HERV and SVA elements.

Class	Length	Copies	Fraction of genome	Sub-families
LINEs (autonomous)	6-8 kb	850,000	21%	3
SINEs (non-autonomous)	100-300 bp	1,500,000	13%	3
Retrovirus-like elements (autonomous)	6-11 kb	450,000	8%	4
Non-autonomous	1.5-3 kb			
DNA transposon (autonomous)	2-3 kb	300,000	3%	7
Non-autonomous	80-3,000 bp			

Figure 17: Classes of interspersed elements in human genome. Number of copies and fraction of genome for these different classes of interspersed repeat in human genome. Adopted and modified from Nature, vol. 409, p 860-921.

RESULTS

Genome-wide hypomethylation causes a modest upregulation of L1 expression

Since most methylated CpG dinucleotides in somatic tissues are located in repetitive elements such as retrotransposons (Yoder et al., 1997), their expression would be expected to increase markedly in the context of genome-wide hypomethylation. Therefore, we assessed frequencies of tags predicted to correspond to transcripts from various widespread retrotransposons. Comparison of longSAGE library tag frequencies of HCT116 and of DKO2L cells shows that the genome-wide hypomethylation in DKO2L affects expression of L1 retrotransposons. We compared the expression of various L1-specific sense and antisense tags in DKO2L and HCT116 libraries. Only 4 of 22 predicted sense-oriented L1 retrotransposon transcript tags were expressed in the DKO2L library; the 3' most tag showed the highest expression in DKO2L cells (78.5 tags per million, TPM), which is slightly higher than the cumulative average expression in public longSAGE libraries, ranging from 0.27 TPM (tags per million tags sequenced) to 76 TPM. The longSAGE tag at the 3' end of L1 templates showed a modest 3-fold upregulation in DKO2L cells vs. HCT116. The third tag from the 5' end showed the second highest expression (49.08 TPM), which is 1.7-fold higher than the cumulative average expression in public long-SAGE libraries and upregulated 11-fold in DKO cells vs. parental cells. The expression of other L1 tags was either very low or not detected in either library (Table 19). We compared the expression of various antisense L1 tags from HCT116 and DKO2L libraries. We did not observe expression of most antisense L1 tags except for the 3' most tags and the sixth tag from 5'end, which showed low level expression (Table 20).

Upregulation of L1 expression correlates with decreased promoter methylation

To verify that genome-wide hypomethylation affects retrotransposons' methylation status, we performed bisulfite sequencing at the 5' UTR region of interspersed L1s (in collaboration with Dr. Jingfeng Li in the Symer laboratory), which contains a sense-orientation promoter and critical binding sites for various transcription factors. Numerous allelic clones were sequenced from both HCT116 and DKO2L samples. The 5' UTR region shows a profound decrease in methylation level in DKO2L cells as compared to the HCT116 cells (Fig. 18). Overall, 74.4% of CpGs in this region were methylated in HCT116 cells, vs. 17.7 % in DKO2L cells. This shows that L1 transcription inversely correlated with promoter methylation, However, despite extensive changes in genomic methylation at their 5' internal promoters, L1 retrotransposons are not extensively upregulated in DKO cells as much as might have been expected given the extent of genomic hypomethylation.

Expression of L1 sense-stranded tags in public long-SAGE libraries

The expression of each of the L1 longSAGE tags was studied in all the publicly available long-SAGE libraries on SAGEmap database (NCBI). The average expression of each predicted L1 tag was calculated from all those libraries. We found that the cumulative average expression of the tags in the public libraries had a wide range from 0.27 TPM (tags per million tags sequenced) to 76 TPM. Mimicking our results from DKO2L cells, the 3' end L1 tag showed the highest average expression of 76.4 TPM followed by the third tags from 5'end, which showed an average expression of 29.7 TPM. Other tags showed low average expression (Table 19).

Comparison of expression status of predicted L1 sense-orientation tags in DKO2L and public long-SAGE libraries on NCBI

We studied the expression of all the bioinformatically predicted L1 sense-orientation tags in our DKO2L long-SAGE library and found that only 4 of the 22 predicted tags were expressed in it. The 3' end tag showed the highest expression in DKO2L cells (78.5 TPM), which is slightly higher than the cumulative average expression in public long-SAGE libraries, the third tag from 5'end showed the second highest expression in the DKO2L cells (49.08 TPM) which is 1.7-fold higher than the cumulative average expression in public long-SAGE libraries (Table 19).

Table 19: Expression status of predicted L1 tags in sense orientation *Column Headers* Tags; predicted L1 tag, WT: normalized tag count (tags per million) in HCT116 library, DKO: normalized tag count (tags per million) in DKO2L library, RatioDKO/WT: ratio of DKO to WT, AvgLong: cumulative average count of the tags in all long-SAGE libraries on NCBI portal, nt. position: nucleotide position of the tags in L1 sequence in sense orientation, NF: tag not found.

Tag	WT	DKO	Ratio DKO/WT	AvgLong	nt Position
GAAAGGAACAACCGGTA	NF	NF	---	2.43	1879
CCAAAATGTAAAGACCA	4.3	9.8	2.2	1.62	1916
GAAACTGAACAACCTGC	4.3	49.08	11.2	29.7	2739
GAGGAACTGGTACCATT	NF	NF	---	0.27	3445
ATCAAGTGGGCTTCATC	NF	NF	---	1.62	3674
ATTATCTCAATAGATGC	NF	NF	---	0.54	3775
CTAAAAACTCTCAATAA	NF	NF	---	0.54	3829
ATTGTATATCTAGAAAA	NF	NF	---	0.27	4102
GGTGAACTCCCATTCGT	NF	NF	---	0	4252
GGTAGGAAGAATCAATA	NF	NF	---	0.54	4407
GTACTGGTACCAAAACA	NF	NF	---	5.6	4653
TCCAAAACACCAAAAGC	NF	NF	---	1.35	4975
GGAGAAAATTTTCGCAA	NF	NF	---	0	5097
AACAGACACTTCTCAAA	NF	NF	---	1.35	5218
AAGAAATGCTCATCATC	NF	NF	---	0.27	5266
CTGCTATAAAGACACAT	NF	NF	---	0	5550
CACACGTATGTTTATTG	NF	NF	---	2.43	5568
GAATACTATGCAGCCAT	NF	NF	---	0.54	5677
GATGAAATTGGAAACCA	NF	NF	---	0.54	5730
GACACAGGAAGGGGAAT	4.3	9.8	2.2	1.08	5834
TATACATATGTAACTAA	NF	NF	NA	5.13	5964
TACCCTAAAACTTAGAG	26.2	78.5	2.9	76.41	6000

Table 20: Expression status of predicted L1 tags in antisense orientation. *Column headers* Tags; predicted L1 tag, WT: normalized tag count (tags per million) in HCT116 library, DKO: normalized tag count (tags per million) in DKO2L library, RatioDKO/WT: ratio of DKO to WT, AvgLong: cumulative average count of the tags in all long-SAGE libraries on NCBI portal, nt. position: nucleotide position of the tags in L1 sequence in anti-sense orientation, NF: tag not found.

Tag	WT	DKO	DKO/WT	AvgLong	nt Position
TGCACATTGTGCAGGTT	NF	NF	---	0.81	65
TGCCATGCTGGTGCGCT	NF	NF	---	0	101
CTGGTGCGCTGCACCCA	NF	NF	---	11.8	108
TGATCTCATTGTTCAAT	NF	NF	---	0.54	231
TCCCTACAAAGGATATG	NF	NF	---	0.27	335
GTGTATATGTGCCACAT	4.3	9.8	2.2	4.3	388
TGTCTTTATAGCAGCAT	NF	NF	---	0	497
ATTTATACTCATTTGGG	NF	NF	---	0	515
TGTTTTTTGGCTGCATA	NF	NF	---	0.81	799
TCCTTCGCCCACTTTTT	NF	NF	---	2.7	847
TTGTAGGTTGCCTGTTC	NF	NF	---	0	968
AAGTCCTTGCCCACGCC	NF	NF	---	0	1090
CTGTTTTGGTTACTGTA	NF	NF	---	3.2	1412
GAATGTTCTTCAGCATG	NF	NF	---	0	1660
GAATGTTCTTCCATTTG	NF	NF	---	0.27	1677
ATTTGGCTCTCTGTTTG	NF	NF	---	0.81	1826
TCGTCTGCAAACAGGGA	NF	NF	---	0.54	1976
AAGGGTTGTTGAATTTT	NF	NF	---	0	2249
TGGTTTTTGTCTTTGGC	NF	NF	---	0.27	2303
GTGGATAAGCTTTTTGA	NF	NF	---	2.7	2404
TACCTCTGGTAGAATTC	NF	NF	---	0	2633
TAGTTGAGCGGCTTTGA	NF	NF	---	0.27	3339
ATTTTGCAGCGGCTGGT	NF	NF	---	0.27	4162
TTTAGCGCTTCCTTCAG	8.7	4.9	0.5	2.16	4199

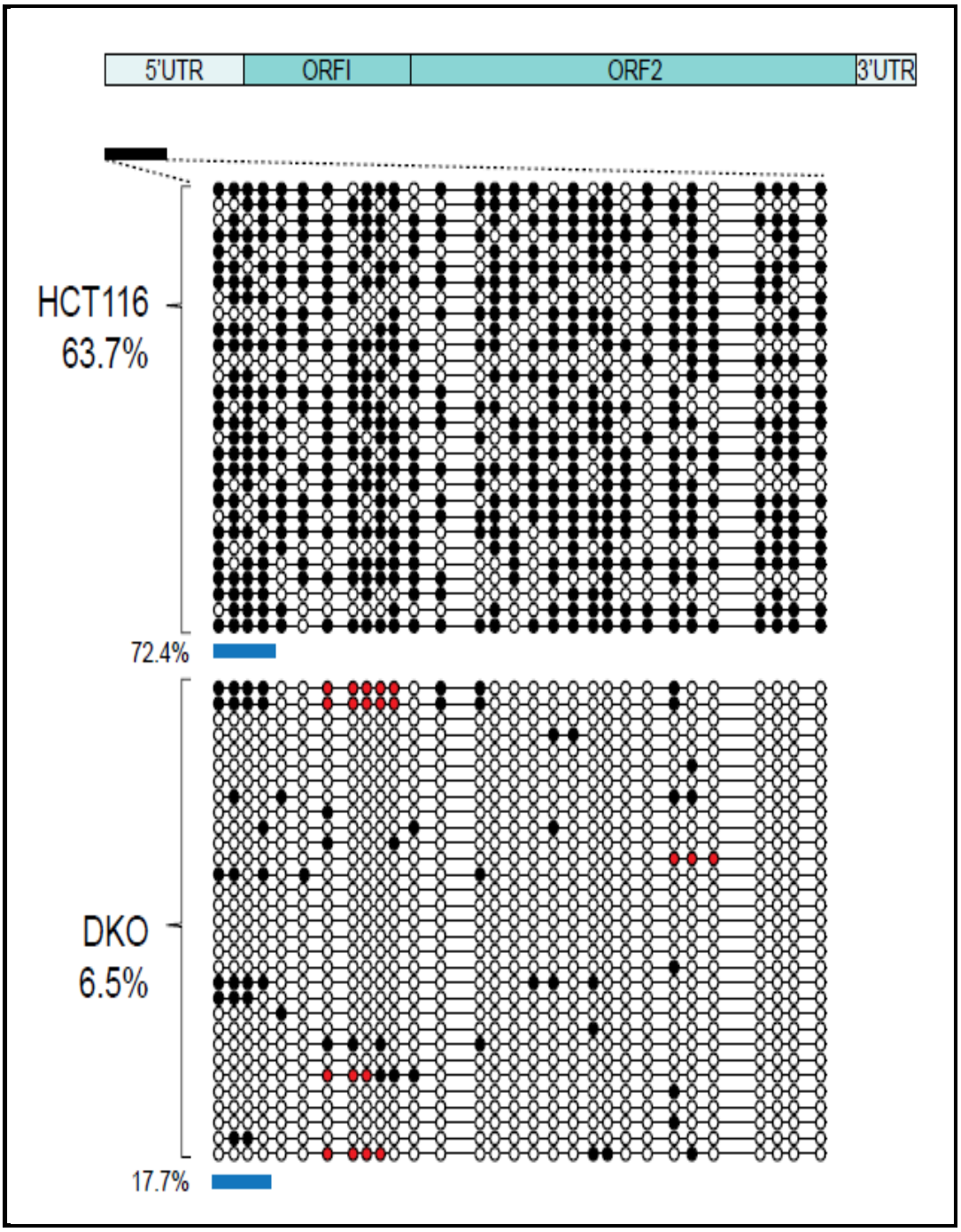

Figure 18: Bisulfite sequencing at the 5' UTR of L1 retrotransposon: Bisulfite sequencing at Hs L1s 5'UTR in HCT116 and DKO2L cells; methylation percentages of all CpGs and four critical CpGs (at +52, +58, +61 and +70 were calculated); horizontal strands represents a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotide, *black*, methylated, *blue*, unmethylated, *red*, poor sequencing.

Effect of genome-wide hypomethylation on Alu, HERV and SVA element transcripts

We conducted a similar analysis of Alu, HERV-K and SVA retrotransposons. We compared the expression of longSAGE tags derived from Alu in HCT116 vs. DKO2L libraries. We observed a subtle 1.3-fold upregulation of this predicted Alu tag in DKO2L library as compared to HCT116. Next, bisulfite sequencing covering a selection of elements was performed (in collaboration with Dr. Jingfeng Li in the Symer laboratory) to correlate differential expression with methylation changes. We found that Alu elements were highly methylated in HCT116 cells which became significantly hypomethylated in DKO2L cells. The total methylation across all the CpG site in Alu elements was 45.1% in HCT116 cells and 6.9% in DKO2L cells (Figure 19). The cumulative average expression on the Alu tags in all publicly available long-SAGE libraries is 581.8 TPM, substantially higher than the expression seen in HCT116 and DKO2L libraries.

We compared the expression of all the bioinformatically predicted HERV-K tags in HCT116 and DKO2L libraries; we did not observe any expression of the predicted HERV-K tags in either library (Table 21). The cumulative average expression of all the predicted HERV-K long-SAGE tags in all the publicly available long-SAGE libraries very low. These results show that either genome-wide hypomethylation in DKO2L cells did not affect the expression of HERV-K elements or the expression of these tags are below the limit of detection of our long-SAGE libraries.

We also compared the expression of all the bioinformatically predicted SVA tags in HCT116 and DKO2L long-SAGE libraries; out of the four predicted SVA tags, three tags from the 3' end were up-regulated in DKO2L cells from 2-fold to 15-fold as compared to HCT116 cells (figure 20 and table 22). This finding suggests that expression of SVA elements is affected by methylation; genome-wide hypomethylation in DKO2L cells was associated with increases in expression of these elements.

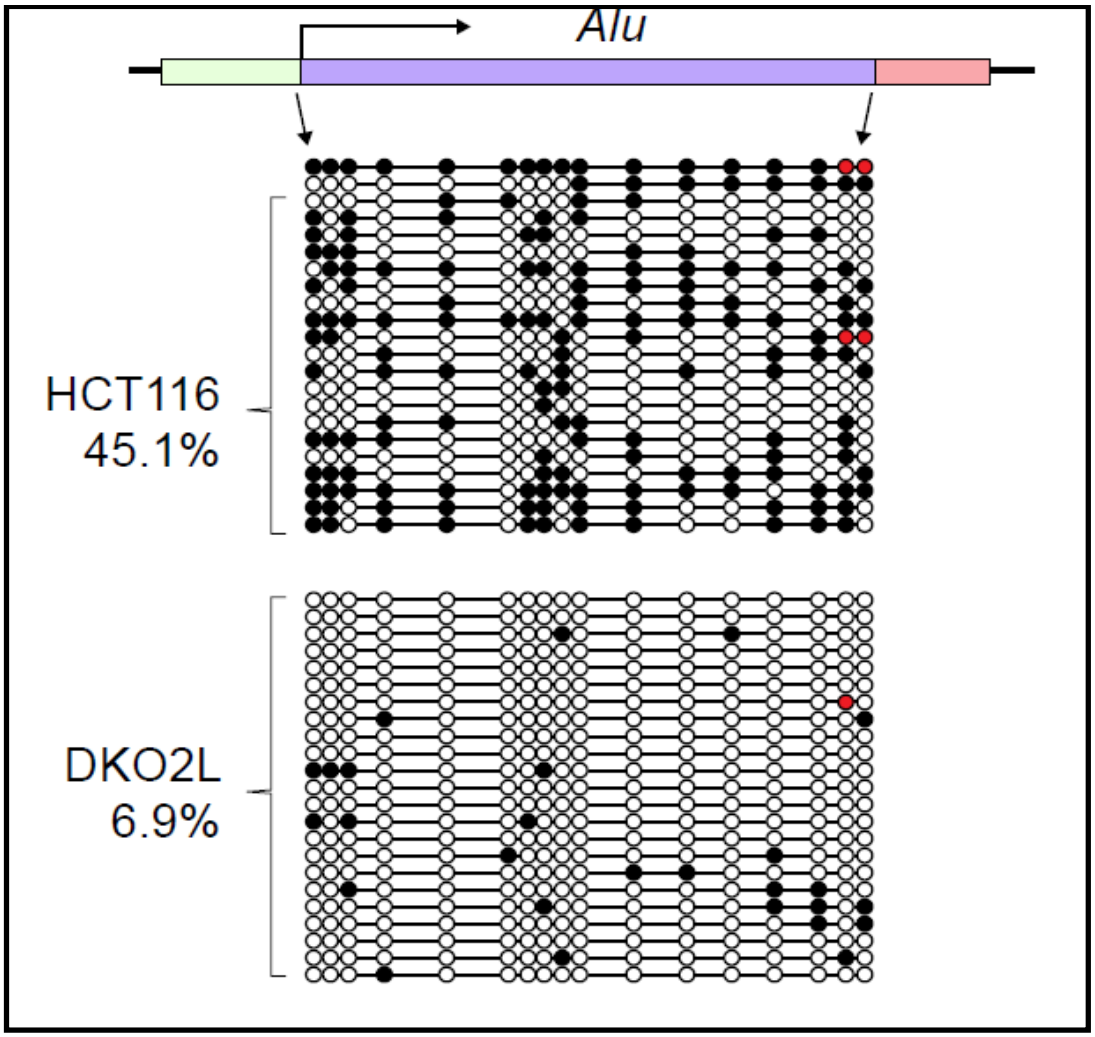

Figure 19: Bisulfite sequencing of Alu sequence: Bisulfite sequencing of entire Alu sequence in HCT116 and DKO2L cells; methylation percentages of all CpGs were calculated; horizontal strands represents a single bisulfite PCR amplicon sequenced, circles represent individual CpG dinucleotide, black=methylated, blue=unmethylated, red=poor sequencing.

Table 21: Expression status of predicted HERV-K tags. *Column headers* Tags; predicted HERV-K tag, WT: normalized tag count (tags per million) in HCT116 library, DKO: normalized tag count (tags per million) in DKO2L library, RatioDKO/WT: ratio of DKO to WT, AvgLong: cumulative average count of the tags in all long-SAGE libraries on NCBI portal. HERV-K (accession no. NM_001007236)

Tag	WT	DKO	RatioDKO/WT	AvgLong
GTTTCCAGAACAAGGAA	NF	NF	---	1.89
AAATTATTGATAAATCA	NF	NF	---	0
GCAATTCCCAGTAACGT	NF	NF	---	2.16
GACATAGACTCATTCCT	NF	NF	---	0.27
CCCGCTCCATCATATAG	NF	NF	---	0
ACCAAGATGGGATATAT	NF	NF	---	0.81
GGGCCTCTCCAACCCGG	NF	NF	---	0
ATCCCAAAAGATTGGCC	NF	NF	---	2.16
TCAAATTTGTTCTCTAT	NF	NF	---	0.27
GAAGATTGGTCTTGCTA	NF	NF	---	0
GATGATCAGTTAAACCA	NF	NF	---	1.35
CTTTGACTCATGTAAAT	NF	NF	---	0
TAAATGCAGCAGGATTA	NF	NF	---	0.27
GAAACAGGCAAAAGATA	NF	NF	---	0
TACCTTCATTTGGAAGA	NF	NF	---	0
TTAAAAAACATTTATTG	NF	NF	---	0.27
AAGGAAAACTAATTTGG	NF	NF	---	0.27
GGAAATAGGGAAGGTGA	NF	NF	---	0.81

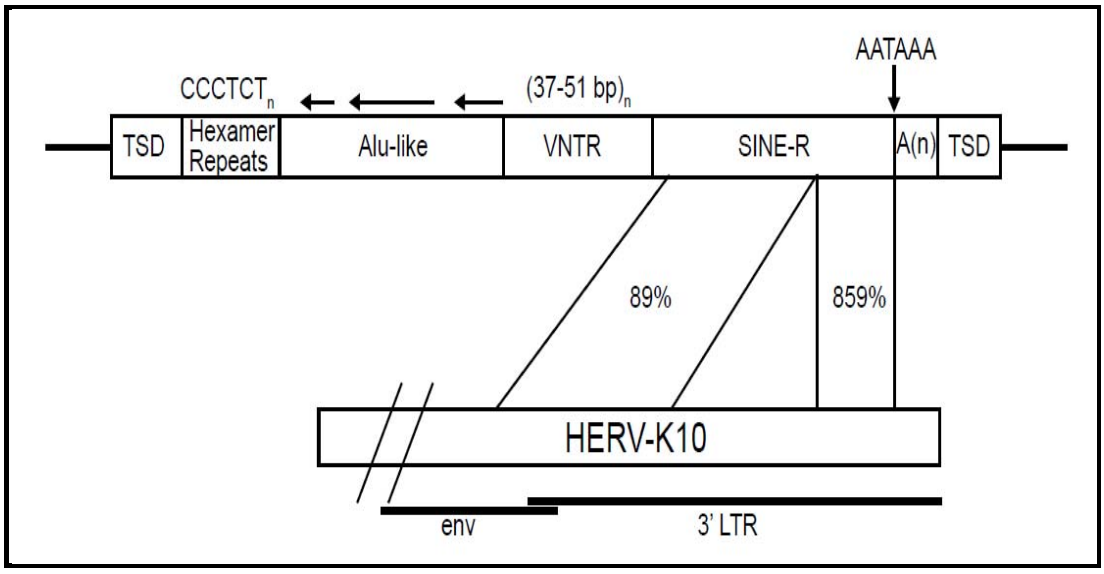

Figure 20: Schematic of SVA elements. The figure depicts the composition of a typical SVA elements

Table 22: Expression status of predicted SVA tags. *Column headers* Tags: predicted SVA tag, WT: normalized tag count (tags per million) in HCT116 library, DKO: normalized tag count (tags per million) in DKO2L library, RatioDKO/WT: ratio of DKO to WT, AvgLong: cumulative average count of the tags in all long-SAGE libraries on NCBI portal. SVA elements (accession no. AC016142).

Tag	WT	DKO	Ratio DKO/WT	AvgLong
ATGACAATGGCGGTTTT	NF	NF	---	0.81
GGAGACTTTTCATTTTG	4.3	68.7	15.7	9.45
TGCTGTGTCCACTCAGG	4.3	49.1	11.2	5.67
CTCGTTAAGAGTCATCA	26.2	58.9	2.2	34.56

DISCUSSION

According to the "genome defense" model proposed by Bestor et al (Yoder et al., 1997), CpG methylation is believed to be an important factor in silencing of transposable elements and repetitive sequence. In fact, an enormous portion of the human and mouse genomes (~ 45%) consists of transposable elements and most DNA methylation is focused on such elements. Activity of these transposable elements can lead to genomic instability (Symer et al., 2002), insertional mutagenesis (Ostertag and Kazazian, 2001) and/or activation or inhibition of cancer-causing genes or oncogenes. For example, expression of IAP elements in mouse is kept under control by cytosine methylation and it was shown that in Dnmt1 hypomorphs ($Dnmt1^{chip/-}$), the centromeric repeats and IAP elements are hypomethylated and expressed (Gaudet et al., 2003).

One of the major focuses of our study was to determine the effect of genomic demethylation in DKO2 cells on transcription of retrotransposons and endogenous retroviral elements. Our longSAGE results showed only a modest 3-fold increase in expression of human L1 retrotransposons in DKO2L cells. Bisulfite sequencing at the 5' UTR promoter region of L1 retrotransposons showed decreased methylation in DKO2L cells as compared to HCT116 suggesting that DNA methylation plays an important role in silencing of these parasitic elements. These observations fit the genome defense model proposed by Bestor et al. However despite profound hypomethylation there was only a modest increase in expression of L1, suggesting that DNA methylation might not be the only mechanism playing a role in the expression or silencing these transposable elements. Other mechanisms such as histone tail modifications and RNAi and cellular inhibitors including members of APOBEC family of proteins possibly could be involved in silencing of these elements. In addition, it is possible that transcription factors required for

expression of these endogenous transposable elements are absent or limiting in somatic, human colorectal cancer cells.

Analysis of transposon expression is a complex undertaking, because of their highly repetitive nature genome-wide. For example, microarrays typically exclude probes for such Repeat Masker-identified sequences, because it would be impossible to identify which element(s) out of thousands could give rise to transcripts. An additional complexity factor is that the elements frequently are degenerate, due to nucleotide substitution, recombination events, etc. over time. Moreover, unlike single copy genes, thousands of repetitive elements could template transcripts, posing a challenge about normalization of transcript counts to template copies.

These problems are illustrated by L1 elements in the human genome, which have integrated over time as member of successful primate-specific or human-specific L1 subfamilies. Moreover, genomic L1 structures frequently are truncated from their 5' ends, so most templates for sense-strand would lack L1-specific promoter but include 3' L1 sequences. L1 transcripts undergo premature polyadenylation and termination, and alternative splicing (Belancio et al., 2007). Recognizing that L1 genomic templates of many shapes, ages, sizes and numbers can give rise to complicated distribution of transcripts, we analyzed predicted longSAGE tags frequencies in our libraries to every possible tag along the consensus "young" L1.3 sequence. This assumes transcripts' taggable 3' ends could occur anywhere along the L1 template in either orientation. We also recognize that active L1 variants might have different sequences at some of the tag positions. Such variant tags are not analyzed by our work here.

We studied the expression of sense-strand L1 tags in all publicly available long-SAGE libraries (http://www.ncbi.nlm.nih.gov/SAGE/). Our survey showed that most L1 tags are weakly expressed across various long-SAGE libraries. However, the 3' most tag and the 3rd tag from the 5'end showed substantial expression in those libraries. We also compared the expression of these predicted L1 tags in HCT116 and DKO2L libraries with the cumulative average expression of all the publicly available long-SAGE libraries. HCT116 library did not show expression of any on these tags. However, DKO2L library showed expression of 4 of predicted L1 tags which included the 3' tag and the 3rd tag from 5'end. The expression of these tags in DKO2L was slightly higher than the cumulative average expression in all the long-SAGE libraries previously sequenced. These comparisons suggest that various predicted L1 sense-orientation tags have higher expression in DKO2L libraries than other publicly available libraries.

We were interested in studying the effect of genome-wide hypomethylation on the expression of the other classes of transposons, including Alu elements. The comparison of expression profiles of HCT116 and DKO2L cells showed that there was no significant change in the expression level of Alu elements between the two cell types. However, bisulfite sequencing across the entire Alu sequence revealed significant decreases in cytosine methylation in the DKO2L cells as compared to HCT116 cells. These results show that despite significant demethylation at Alu sequences, there is no effect on the expression of these elements. Although Alu elements have high CpG densities and are highly methylated in somatic cells, decreases in DNA methylation may not be sufficient for their expression, unlike RNA polymerase II transcribed elements. Tissue specific-factors and/or lack of effect of CpG methylation on RNA polymerase III may play a role here.

We were unable to detect the expression of any of the predicted HERV-K tags in both HCT116 and DKO2L libraries, suggesting that either there was no effect of genome-wide hypomethylation on the expression of these elements or the expression was below the limit of detection of our long-SAGE libraries.

SVA elements are composite retrotransposons which are highly methylated in all somatic tissues, suggesting that DNA methylation might play an important role in regulating expression of these elements. Upon comparing the expression of all the predicted SVA tags in HCT116 and DKO2L libraries, we found that genome-wide hypomethylation in DKO2L caused increases expression of SVA elements from 2-fold to 15-fold as compared to the HCT116 cells.

Collectively these results suggest that DNA methylation may play variable roles in the regulation of expression of various classes of retrotransposons. Different classes appear to show different levels of effects due to changes in methylation status. Thus DNA methylation is not the sole mechanism regulating expression of transposons at least in these cultured cells. There could be multiple overlapping regulatory mechanisms such as histone tail modifications, regulatory RNAs, and cellular inhibitors like APOBEC proteins, or limiting transcription factors affecting certain classes. The expression of a particular class of retrotransposons would depend on the interplay of these multiple regulatory mechanisms. The reason behind multiple regulatory mechanisms could be to have very tight governance over the expression of this huge compartment of genome lest any rampant expression of these elements could lead to increased genomic instability and diseases.

FUTURE SCOPE OF WORK

1. Reintroduce Dnmt1 and Dnmt3b gene into DKO2L cells

    A. To study the remethylation pattern transposable elements in DKO2L cells with reintroduced Dnmt1 and 3b genes (to check if it recapitulates HCT116 pattern)

    B. To check the expression status of those transposable elements, which were affected by the genome-wide hypomethylation in DKO2L cells, and to determine if their expression level is brought to the HCT116 levels

    C. To study how different levels of DNMT expression may affect the resetting of methylation marks.

2. To study the effect of various histone tail modification on the expression of transposable elements.

3. Compare the effects of genome-wide hypomethylation on the expression of transposable elements in somatic cell versus embryonic stem cells.

DNA methylation is a stable epigenetic mark that is strongly associated with transcriptional repression. Abnormal patterns of both genome-wide hypomethylation and localized hypermethylation have been identified in virtually all human malignancies. The studies described above validate that DNA methylation plays an important role in the transcriptional regulation and expression of genes and various other transcribed elements.

As described in Chapter 3, we verified that genetic disruption of DNA methyltransferase genes (Dnmt1 and Dnmt3b) can affect gene transcription. As described previously (Toth, M., et al., 1989; Qian, X., et al., 1996), our study demonstrates an inverse correlation between promoter methylation and gene expression. We identified several thousand genes and transcribed elements that show profound changes in gene expression, both upregulated and downregulated. In addition, we validated many of these genes using independent methods and cell lineages. Our study also suggests that there are certain classes of genes that are overexpressed due to decreased promoter methylation and genome-wide hypomethylation. These include Interferon alpha- inducible genes, cancer-testis genes, embryonic genes, MHC genes and metallothionein genes. We compared the transcriptomes of HCT116 cells and DKO2L cells with those of various other cell types (including mouse lines) that have genome-wide hypomethylation due to either pharmacological or genetic manipulations. This comparison revealed that there are very few classes of genes that are commonly upregulated in genome-wide hypomethylated cells in all these studies. These classes included interferon-inducible genes, cancer-testis genes, MHC class I genes, metallothionein genes and some of the embryonic genes. Most of the genes have classic CpG islands in their

promoter/5'region and show inverse correlation between promoter methylation and expression. These results demonstrate that coordinated expression of these classes of genes could indicate the genome-wide, hypomethylated state of a cell. These arguments suggest that a "transcriptome-signature" of genome-wide hypomethylated cells can be identified and refined.

We also developed a candidate list of eleven tags, which constitutes a preliminary transcriptome signature of DKO2L cells. An approach to refine and improve this proposed transcriptome signature of genomic hypomethylation in cultured human colorectal cancer cells would be to re-introduce DNMT1 and DNMT3B genes into DKO2L cells, to determine if expression of members of the transcriptome signature reverts back to expression levels of the parental cells. Of course, this assumes that karotypic instability in the DKO cells does not preclude reestablishment of "wildtype" expression patterns.

The results presented in Chapter 4 suggest that DNA methylation may contribute only modestly to the regulation of expression of various classes of retrotransposons, at least in the colorectal cells studied here. Different classes appear to show different levels of effects due to changes in their methylation status. Thus DNA methylation is not the sole mechanism regulating expression of transposons at least in these cultured cells. There are likely multiple overlapping regulatory mechanisms such as histone tail modifications, regulatory RNAs, and cellular inhibitors like APOBEC proteins, or limiting transcription factors affecting certain classes. The expression of a particular class of retrotransposons would depend on the interplay of these multiple regulatory mechanisms. The reason behind multiple regulatory mechanisms could be to have very tight governance over the expression of

this huge compartment of genome lest any rampant expression of these elements could lead to increased genomic instability and diseases.

Together, all these findings help to validate the hypothesis that DNA methylation plays a very important role in gene expression, and that decreased genomic methylation affects various compartments of transcriptome. The results suggest that a distinctive, reproducible transcriptome signature of profound genome-wide hypomethylation can be identified and refined. In addition, these results have helped to elucidate several biologically important pathways whose deregulation by genome-wide hypomethylation may contribute to cancer formation.

REFERENCES (Chapter 3)

Akers SN, Odunsi K, Karpf AR (2010). Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy. Future Oncol. 2010 May;6(5):717-32

Allegrucci, C., Thurston, A., Lucas, E., and Young, L. (2005). Epigenetics and the germline. Reproduction *129*, 137-149.

Antequera, F., and Bird, A. (1993). CpG islands. Exs *64*, 169-185.

Antequera, F., Boyes, J., and Bird, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell *62*, 503-514.

Argeson, A.C., Nelson, K.K., and Siracusa, L.D. (1996). Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. Genetics *142*, 557-567.

Bachman, K.E., Park, B.H., Rhee, I., Rajagopalan, H., Herman, J.G., Baylin, S.B., Kinzler, K.W., and Vogelstein, B. (2003). Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell *3*, 89-95.

Baylin, S., and Bestor, T.H. (2002). Altered methylation patterns in cancer cell genomes: cause or consequence? Cancer Cell *1*, 299-305.

Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., and Herman, J.G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet *10*, 687-692.

Berdasco M, Esteller M. (2010). Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010 Nov 16;19(5):698-711

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol *203*, 971-983.

Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. Embo J *11*, 2611-2617.

Bestor, T.H. (1998). Gene silencing. Methylation meets acetylation. Nature *393*, 311-312.

Biliya S, Bulla LA Jr.(2010) Genomic imprinting: the influence of differential methylation in the two sexes. Exp Biol Med (Maywood). 2010 Feb;235(2):139-47

Bird, A. (1992). The essentials of DNA methylation. Cell *70*, 5-8.

Boland CR, Goel A (2010). Microsatellite instability in colorectal cancer. Gastroenterology. 2010 Jun;138(6):2073-2087.e3

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Braganca, J., and Civas, A. (1998). Type I interferon gene expression: differential expression of IFN-A genes induced by viruses and double-stranded RNA. Biochimie *80*, 673-687.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630-634.

Cedar, H., and Razin, A. (1990). DNA methylation and development. Biochim Biophys Acta *1049*, 1-8.

Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L., and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. Nature *395*, 89-93.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet *39*, 391-396.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem *162*, 156-159.

Coral, S., Sigalotti, L., Altomonte, M., Engelsberg, A., Colizzi, F., Cattarossi, I., Maraskovsky, E., Jager, E., Seliger, B., and Maio, M. (2002). 5-aza-2'-deoxycytidine-induced expression of functional cancer testis antigens in human renal cell carcinoma: immunotherapeutic implications. Clin Cancer Res *8*, 2690-2695.

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C.L., and Feinberg, A.P. (2002). Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. Cancer Res *62*, 6442-6446.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V., and Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol *19*, 7327-7335.

Dean, W., Lucifero, D., and Santos, F. (2005). DNA methylation in mammalian development and disease. Birth Defects Res C Embryo Today *75*, 98-111.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. Curr Opin Genet Dev *4*, 255-259.

Egger, G., Jeong, S., Escobar, S.G., Cortez, C.C., Li, T.W., Saito, Y., Yoo, C.B., Jones, P.A., and Liang, G. (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc Natl Acad Sci U S A *103*, 14080-14085.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. Oncogene *21*, 5400-5413.

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene *21*, 5427-5440.

Feinberg, A.P. (2005). A genetic approach to cancer epigenetics. Cold Spring Harb Symp Quant Biol *70*, 335-341.

Ghoshal, K., Majumder, S., and Jacob, S.T. (2002). Analysis of promoter methylation and its role in silencing metallothionein I gene expression in tumor cells. Methods Enzymol *353*, 476-486.

Ghoshal, K., Majumder, S., Li, Z., Dong, X., and Jacob, S.T. (2000). Suppression of metallothionein gene expression in a rat hepatoma because of promoter-specific DNA methylation. J Biol Chem *275*, 539-547.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet *24*, 368-371.

Gius, D., Cui, H., Bradbury, C.M., Cook, J., Smart, D.K., Zhao, S., Young, L., Brandenburg, S.A., Hu, Y., Bisht, K.S., *et al.* (2004). Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. Cancer Cell *6*, 361-371.

Goelz, S.E., Vogelstein, B., Hamilton, S.R., and Feinberg, A.P. (1985). Hypomethylation of DNA from benign and malignant human colon neoplasms. Science *228*, 187-190.

Gonzalez-Zulueta, M., Bender, C.M., Yang, A.S., Nguyen, T., Beart, R.W., Van Tornout, J.M., and Jones, P.A. (1995). Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res *55*, 4531-4535.

Greger, V., Passarge, E., Hopping, W., Messmer, E., and Horsthemke, B. (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet *83*, 155-158.

Grunau, C., Clark, S.J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res *29*, E65-65.

Guillaudeux, T., Gomez, E., Onno, M., Drenou, B., Segretain, D., Alberti, S., Lejeune, H., Fauchet, R., Jegou, B., and Le Bouteiller, P. (1996). Expression of HLA class I genes in meiotic and post-meiotic human spermatogenic cells. Biol Reprod *55*, 99-110.

Gure, A.O., Wei, I.J., Old, L.J., and Chen, Y.T. (2002). The SSX gene family: characterization of 9 complete genes. Int J Cancer *101*, 448-453.

Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M., *et al.* (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A *91*, 9700-9704.

Hong, J.A., Kang, Y., Abdullaev, Z., Flanagan, P.T., Pack, S.D., Fischette, M.R., Adnani, M.T., Loukinov, D.I., Vatolin, S., Risinger, J.I., *et al.* (2005). Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. Cancer Res *65*, 7763-7774.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., *et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet *27*, 31-39.

James, S.R., Link, P.A., and Karpf, A.R. (2006). Epigenetic regulation of X-linked cancer/germline antigen genes by DNMT1 and DNMT3b. Oncogene *25*, 6975-6985.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev *6*, 705-714.

Kal, A.J., van Zonneveld, A.J., Benes, V., van den Berg, M., Koerkamp, M.G., Albermann, K., Strack, N., Ruijter, J.M., Richter, A., Dujon, B., *et al.* (1999). Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. Mol Biol Cell *10*, 1859-1872.

Kaneda, A., and Feinberg, A.P. (2005). Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. Cancer Res *65*, 11236-11240.

Karpf, A.R. (2007). Epigenomic reactivation screening to identify genes silenced by DNA hypermethylation in human cancer. Curr Opin Mol Ther *9*, 231-241.

Karpf, A.R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. Cancer Res *65*, 8635-8639.

Karpf, A.R., Peterson, P.W., Rawlins, J.T., Dalley, B.K., Yang, Q., Albertsen, H., and Jones, D.A. (1999). Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. Proc Natl Acad Sci U S A *96*, 14007-14012.

Komarova, N.L., Lengauer, C., Vogelstein, B., and Nowak, M.A. (2002). Dynamics of genetic instability in sporadic and familial colorectal cancer. Cancer Biol Ther *1*, 685-692.

Koslowski, M., Bell, C., Seitz, G., Lehr, H.A., Roemer, K., Muntefering, H., Huber, C., Sahin, U., and Tureci, O. (2004). Frequent nonrandom activation of germ-line genes in human cancer. Cancer Res *64*, 5988-5993.

Lal, A., Lash, A.E., Altschul, S.F., Velculescu, V., Zhang, L., McLendon, R.E., Marra, M.A., Prange, C., Morin, P.J., Polyak, K., *et al.* (1999). A public database for gene expression in human cancers. Cancer Res *59*, 5403-5407.

Lei, H., et al., (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122, 3195-3205

Li, E., Bestor, T. H., and Jaenisch, R. (1992) Targeted mutation of DNA methyltransferase gene results in embryonic lethality. Cell 69, 915-926

Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. Widespread RNA and DNA Sequence Differences in the Human Transcriptome. Science.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol *22*, 480-491.

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., *et al.* (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic

reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A *99*, 6806-6811.

Majumder, S., Ghoshal, K., Li, Z., and Jacob, S.T. (1999). Hypermethylation of metallothionein-I promoter and suppression of its induction in cell lines overexpressing the large subunit of Ku protein. J Biol Chem *274*, 28584-28589.

McGarvey, K.M., Fahrner, J.A., Greene, E., Martens, J., Jenuwein, T., and Baylin, S.B. (2006). Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer Res *66*, 3541-3549.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., and Sidransky, D. (1995). 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med *1*, 686-692.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Nie, Y., Yang, G., Song, Y., Zhao, X., So, C., Liao, J., Wang, L.D., and Yang, C.S. (2001). DNA hypermethylation is a mechanism for loss of expression of the HLA class I genes in human esophageal squamous cell carcinomas. Carcinogenesis *22*, 1615-1623.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Qian X, von Wronski MA, Brent TP. (1995). Localization of methylation sites in the human O6-methylguanine-DNA methyltransferase promoter: correlation with gene suppression. Carcinogenesis 16(6):1385-90.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., *et al.* (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552-556.

Rhee, I., Jair, K.W., Yen, R.W., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B., and Schuebel, K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. Nature *404*, 1003-1007.

Ruijter, J.M., Van Kampen, A.H., and Baas, F. (2002). Statistical evaluation of SAGE libraries: consequences for experimental design. Physiol Genomics *11*, 37-44.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Using the transcriptome to annotate the genome. Nat Biotechnol *20*, 508-512.

Sato, N., Fukushima, N., Maitra, A., Matsubayashi, H., Yeo, C.J., Cameron, J.L., Hruban, R.H., and Goggins, M. (2003). Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. Cancer Res *63*, 3735-3742.

Serrano, A., Tanzarella, S., Lionello, I., Mendez, R., Traversari, C., Ruiz-Cabello, F., and Garrido, F. (2001). Rexpression of HLA class I antigens and restoration of antigen-specific CTL response in melanoma cells following 5-aza-2'-deoxycytidine treatment. Int J Cancer *94*, 243-251.

Sharma S, Kelly TK, Jones PA. (2010). Epigenetics in cancer. Carcinogenesis. 2010 Jan;31(1):27-36

Siedlecki, P., and Zielenkiewicz, P. (2006). Mammalian DNA methyltransferases. Acta Biochim Pol *53*, 245-256.

Smiraglia, D.J., and Plass, C. (2002). The study of aberrant methylation in cancer via restriction landmark genomic scanning. Oncogene *21*, 5414-5426.

Suri, A. (2006). Cancer testis antigens--their importance in immunotherapy and in the early detection of cancer. Expert Opin Biol Ther *6*, 379-389.

Toth M, Lichtenberg U, and Doerfler W. (1989) Genomic sequencing reveals a 5-methylcytosine-free domain in active promoters and the spreading of preimposed methylation patterns. Proc Natl Acad Sci U S A. 86(10):3728-32

Toung, J.M., Morley, M., Li, M., and Cheung, V.G. RNA-sequence analysis of human B-cells. Genome Res *21*, 991-998.

Tahiliani,M.(2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 324, 930–935

van Ruissen, F., Ruijter, J.M., Schaaf, G.J., Asgharnegad, L., Zwijnenburg, D.A., Kool, M., and Baas, F. (2005). Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. BMC Genomics *6*, 91.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol *3*, RESEARCH0034.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484-487.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Weber, B., Kimhi, S., Howard, G., Eden, A., and Lyko, F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene *29*, 5775-5784.

Weber, J., Salgaller, M., Samid, D., Johnson, B., Herlyn, M., Lassam, N., Treisman, J., and Rosenberg, S.A. (1994). Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine. Cancer Res *54*, 1766-1771.

Yan H, Choi AJ, Lee BH, Ting AH. (2011). Identification and functional analysis of epigenetically silenced microRNAs in colorectal cancer cells. PLoS One. 6(6):e20628.

Yang, A.S., Estecio, M.R., Doshi, K., Kondo, Y., Tajara, E.H., and Issa, J.P. (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res *32*, e38.

You JS, Jones PA (2012). Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell. 2012 Jul 10;22(1):9-20.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature *451*, 202-206.

REFERENCES (CHAPTER 4)

Akagi, K., Li, J., Stephens, R.M., Volfovsky, N., and Symer, D.E. (2008). Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. Genome Res *18*, 869-880.

Akagi, K., Stephens, R.M., Li, J., Evdokimov, E., Kuehn, M.R., Volfovsky, N., and Symer, D.E. (2010). MouseIndelDB: a database integrating genomic indel polymorphisms that distinguish mouse strains. Nucleic Acids Res *38*, D600-606.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761-764.

Belancio, V.P., Whelton, M., and Deininger, P. (2007). Requirements for polyadenylation at the 3' end of LINE-1 elements. Gene *390*, 98-107.

Bestor, T.H., and Tycko, B. (1996). Creation of genomic methylation patterns. Nat Genet *12*, 363-367.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 8780-8785.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Branciforte, D., and Martin, S.L. (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol *14*, 2584-2592.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet *27*, 327-331.

Contreras-Galindo, R., Kaplan, M.H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., Giusti, F., Dosik, M.H., Hayes, D.F., Gitlin, S.D., *et al.* (2008). Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. J Virol *82*, 9329-9336.

Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. (2002). Human L1 element target-primed reverse transcription in vitro. Embo J *21*, 5899-5910.

Cruickshanks, H.A., and Tufarelli, C. (2009). Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics *94*, 397-406.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41-48.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat Genet *24*, 363-367.

Esteller, M., Fraga, M.F., Paz, M.F., Campo, E., Colomer, D., Novo, F.J., Calasanz, M.J., Galm, O., Guo, M., Benitez, J., *et al.* (2002). Cancer epigenetics and methylation. Science *297*, 1807-1808; discussion 1807-1808.

Fanning, T., and Singer, M. (1987). The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res *15*, 2251-2260.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., *et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat Genet *41*, 563-571.

Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905-916.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.

Gilbert, N., Lutz-Prigge, S., and Moran, J.V. (2002). Genomic deletions created upon LINE-1 retrotransposition. Cell *110*, 315-325.

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol *25*, 7780-7795.

Gotzinger, N., Sauter, M., Roemer, K., and Mueller-Lantzsch, N. (1996). Regulation of human endogenous retrovirus-K Gag expression in teratocarcinoma cell lines and human tumours. J Gen Virol *77 (Pt 12)*, 2983-2990.

Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. Embo J *3*, 1753-1759.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet *27*, 322-326.

Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E., and Kazazian, H.H., Jr. (2011). Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet *20*, 3386-3400.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E.*, et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet *27*, 31-39.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Kolosha, V.O., and Martin, S.L. (1997). In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. Proc Natl Acad Sci U S A *94*, 10155-10160.

Kolosha, V.O., and Martin, S.L. (2003). High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). J Biol Chem *278*, 8112-8117.

Korenberg, J.R., and Rykowski, M.C. (1988). Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell *53*, 391-400.

Lavie, L., Kitova, M., Maldener, E., Meese, E., and Mayer, J. (2005). CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). J Virol *79*, 876-883.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. Science *254*, 1808-1810.

Meehan, R., Lewis, J., Cross, S., Nan, X., Jeppesen, P., and Bird, A. (1992). Transcriptional repression by methylation of CpG. J Cell Sci Suppl *16*, 9-14.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530-1534.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G., and Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem *281*, 22161-22172.

O'Donnell, K.A., and Boeke, J.D. (2007). Mighty Piwis defend the germline against genome intruders. Cell *129*, 37-44.

Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian, H.H., Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet *73*, 1444-1451.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Biology of mammalian L1 retrotransposons. Annu Rev Genet *35*, 501-538.

Rangwala, S.H., Zhang, L., and Kazazian, H.H., Jr. (2009). Many LINE1 elements contribute to the transcriptome of human somatic cells. Genome Biol *10*, R100.

Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., *et al.* (2005). The DNA sequence of the human X chromosome. Nature *434*, 325-337.

Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian, H.H., Jr. (1997). Many human L1 elements are capable of retrotransposition. Nat Genet *16*, 37-43.

Soifer, H.S., Zaragoza, A., Peyvan, M., Behlke, M.A., and Rossi, J.J. (2005). A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. Nucleic Acids Res *33*, 846-856.

Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., and Feinberg, A.P. (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome Res *12*, 543-554.

Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol *10*, 6718-6729.

Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. (2002). Human l1 retrotransposition is associated with genetic instability in vivo. Cell *110*, 327-338.

Takahara, T., Ohsumi, T., Kuromitsu, J., Shibata, K., Sasaki, N., Okazaki, Y., Shibata, H., Sato, S., Yoshiki, A., Kusakabe, M., *et al.* (1996). Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. Hum Mol Genet *5*, 989-993.

van den Hurk, J.A., Meij, I.C., Seleme, M.C., Kano, H., Nikopoulos, K., Hoefsloot, L.H., Sistermans, E.A., de Wijs, I.J., Mukhopadhyay, A., Plomp, A.S., *et al.* (2007). L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet *16*, 1587-1592.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Weber, B., Kimhi, S., Howard, G., Eden, A., and Lyko, F (2010). Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene *29*, 5775-5784.

Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol *13*, 763-771.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Appendix 1 (Commonly de-regulated genes, longSAGE vs. cDNA Microarray)

Appendix 1.1 Commonly upregulated genes in longSAGE vs. cDNA microarray comparison of expression profiles of HCT116 and DKO2L cells. Gene: upregulated gene, description: description of the gene, SAGE ratio= ratio of NormDKO to Norm WT, Array fold change: ratio of expression in DKO to HCT116.

Gene	Description	SAGE Ratio	Array Fold Change
BST2	Bone marrow stromal cell antigen 2	110.49	8.50
NUDC	Nuclear distribution gene C homolog (A. nidulans)	17.68	1.77
IFI6	Interferon, alpha-inducible protein 6	15.47	5.22
BSCL2	Bernardinelli-Seip congenital lipodystrophy 2 (sei	15.47	1.77
CBR1	Carbonyl reductase 1	15.47	2.00
LGALS3B	Lectin, galactoside-binding, soluble, 3 binding pr	13.26	1.54
HLA-A	Major histocompatibility complex, class I, A	11.05	2.27
CASP7	Caspase 7, apoptosis-related cysteine peptidase	11.05	1.85
IFITM1	Interferon induced transmembrane protein 1 (9-27)	11.05	5.11
PRAME	Preferentially expressed antigen in melanoma	9.94	2.77
HIST1H1	Histone cluster 1, H1c	9.67	1.72
EN2	Engrailed homolog 2	8.84	3.17
COL6A3	Collagen, type VI, alpha 3	8.84	2.08
TP53AP1	TP53 activated protein 1	8.84	1.61
UBE2C	Ubiquitin-conjugating enzyme E2C	8.29	2.00
ISG15	ISG15 ubiquitin-like modifier	7.04	4.65
EZH2	Enhancer of zeste homolog 2 (Drosophila)	6.63	1.95
PNMA1	Paraneoplastic antigen MA1	6.08	1.85
CTSL	Cathepsin L	6.08	1.89
HLA-B	Major histocompatibility complex, class I, B	5.44	2.61
ACAA2	Acetyl-Coenzyme A acyltransferase 2 (mitochondrial	4.97	1.64
STK17A	Serine/threonine kinase 17a (apoptosis-inducing)	4.42	1.58
STAT1	Signal transducer and activator of transcription 1	4.42	4.72
DUSP6	Dual specificity phosphatase 6	3.13	1.81
P4HB	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	3.01	1.56
STMN1	Stathmin 1/oncoprotein 18	2.59	2.01
B2M	Beta-2-microglobulin	2.46	3.24
CALR	Calreticulin	2.04	2.32
DYNLL1	Dynein, light chain, LC8-type 1	2.03	2.33
TACC3	Transforming, acidic coiled-coil containing protei	1.70	1.67
STIP1	Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organ	1.55	1.60
PSMC5	Proteasome (prosome, macropain) 26S subunit, ATPas	1.55	1.59
ANXA5	Annexin A5	1.51	1.98

Appendix 1.2. Commonly downregulated genes in longSAGE vs. cDNA microarray comparison of expression profiles of HCT116 and DKO2L cells. Gene: downregulated gene, description: description of the gene, SAGE ratio= ratio of NormDKO to Norm WT, Array fold change: ratio of expression in DKO to HCT116.

Gene	Description	SAGE Ratio	Array Fold
THOP1	Thimet oligopeptidase 1	0.05	0.56
FAM3C	Family with sequence similarity 3, member C	0.08	0.55
SLC25A6	Solute carrier family 25 (mitochondrial carrier; a	0.08	0.56
KRT8	Keratin 8	0.09	0.42
EIF5A	Eukaryotic translation initiation factor 5A	0.10	0.61
PRR5	Proline rich 5 (renal)	0.16	0.64
TNNT1	Troponin T type 1 (skeletal, slow)	0.21	0.66
CRABP2	Cellular retinoic acid binding protein 2	0.24	0.56
TXNRD1	Thioredoxin reductase 1	0.35	0.48
MALL	Mal, T-cell differentiation protein-like	0.40	0.38
RPL10	Ribosomal protein L10	0.44	0.52
KRR1	KRR1, small subunit (SSU) processome component, ho	0.46	0.52
NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex	0.46	0.66
TPM2	Tropomyosin 2 (beta)	0.48	0.65
CYC1	Cytochrome c-1	0.49	0.66
RPL23	Ribosomal protein L23	0.50	0.40
RPL27A	Ribosomal protein L27a	0.51	0.65
EEF2	Eukaryotic translation elongation factor 2	0.51	0.56
RPL29	Ribosomal protein L29	0.55	0.62
CYB5B	Cytochrome b5 type B (outer mitochondrial membrane	0.55	0.28
TOMM7	Translocase of outer mitochondrial membrane 7 homo	0.56	0.61
NME1	Non-metastatic cells 1, protein (NM23A) expressed	0.59	0.60
RPL12	Ribosomal protein L12	0.61	0.54
RPL30	Ribosomal protein L30	0.61	0.66
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	0.63	0.49

Appendix 2. Differentially regulated genes (pvalue<0.05)

Appendix 2.1 Differentially upregulated genes (pvalue<0.05)

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CCAGGGGAGAAGGCACC	Hs.532634	IFI27	14q32	4.23	1139.10	269.60	0.00E+00
GAAGGTGATCTGCAAGA	Hs.112208	XAGE1	Xp11.22	4.23	868.33	205.52	0.00E+00
CAGTCTAAAATGCTTCA	Hs.518731	UCHL1	4p14	4.23	569.55	134.80	2.22E-16
CAGCCTGGGGCCACTGC	Hs.72026	PRSS21	16p13.3	4.23	476.18	112.70	6.02E-14
TGCTGCCTGTTGTTATG	Hs.118110	BST2	19p13.2	4.23	466.84	110.49	1.06E-13
CAGACGGTGGCCCAGCC	Hs.106137	NOTUM	17q25.3	4.23	326.79	77.34	5.00E-10
ACCCTGCCAAATCCCCC	Hs.231895	CECR2	22q11.2	4.23	214.75	50.83	4.62E-07
AGCTGTGCCAAGTGTGC	Hs.513626	MT1A	16q13	4.23	186.74	44.20	2.59E-06
CGCCTCCGGGAGAAATT	Hs.642813	VIM	10p13	4.23	186.74	44.20	2.59E-06
CTTGTAATCCTACTTGG	Hs.608710			4.23	186.74	44.20	2.59E-06
GTGCCTGCCGAGAGGGC	Hs.146346	C3orf41	3p22.1	4.23	186.74	44.20	2.59E-06
AGATTTAAATTCTGTGG	Hs.131543	CTCFL	20q13.3	4.23	149.39	35.36	2.61E-05
CAGAGATGAATTTATAC	Hs.520028	HSPA1A	6p21.3	4.23	149.39	35.36	2.61E-05
AACACGGTGCTCAGGGG				4.23	102.71	24.31	4.90E-04
CACTGTGACCTTGGGGG	Hs.513491	YPEL3	16p11.2	4.23	102.71	24.31	4.90E-04
ATTGTAGACAATGAGGG	Hs.462777	MYO1D	17q11-q	4.23	93.37	22.10	8.87E-04
CCTGTCCAGCCTGGGCA	Hs.516966	BCL2L1	20q11.2	4.23	93.37	22.10	8.87E-04
GGGTTGCTGTAAGGATT				4.23	93.37	22.10	8.87E-04
TTTGTGGGCAGTCAGGC	Hs.567550	RCN3	19q13.3	4.23	93.37	22.10	8.87E-04
GGCTGTGCCAAGTGTGC	Hs.647358	MT1L	16q13	4.23	84.03	19.89	1.61E-03
TGCACTTCACCGCCCTG	Hs.504249	DCPS	11q24.2	4.23	84.03	19.89	1.61E-03
TTCAGAACAAAAGTGCA	Hs.373763	HNRPR	1p36.12	4.23	84.03	19.89	1.61E-03
TCCAAATCGATGTGGAT	Hs.642813	VIM	10p13	25.35	457.51	18.05	1.50E-11
ACCCGCGTGCTGCAGGG	Hs.590955	CGB	19q13.3	4.23	74.70	17.68	2.95E-03
CACGCCAGCCCCACGGG	Hs.287717	PEMT	17p11.2	4.23	74.70	17.68	2.95E-03
CAGACCTTAATAAATAC				4.23	74.70	17.68	2.95E-03
CCAGGGGGAGAAGGCAC				4.23	74.70	17.68	2.95E-03
GCTCAGCAGCACGAGGG	Hs.263812	NUDC	1p35-p3	4.23	74.70	17.68	2.95E-03
GGATGCGCAGGGGAGGC	Hs.459211	AKAP13	15q24-q	4.23	74.70	17.68	2.95E-03
GGGGGCAAGCGGGACCC	Hs.591290	PODXL2	3q21.3	4.23	74.70	17.68	2.95E-03
GTGCAAGGGTATTGTCG	Hs.646351	ECAT8	19q13.1	4.23	74.70	17.68	2.95E-03
GTGCCAGCCCTCCTGGG	Hs.119177	ARF3	12q13	4.23	74.70	17.68	2.95E-03
GTGTGGGAGATTGCAGA	Hs.518201	DTX3L	3q21.1	4.23	74.70	17.68	2.95E-03
TTTATCCCAAACTAATC	Hs.611057	C1orf77	1q21.3	4.23	74.70	17.68	2.95E-03
AGGCCCATTCTTCACTC	Hs.72879	MAGEA1	Xq28	8.45	149.39	17.68	1.19E-04

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
AATTAAAAAAAAAAAAA				4.23	65.36	15.47	5.42E-03
ACATTTCAATTCATCTC	Hs.524250	GABARAP	12p13.2	4.23	65.36	15.47	5.42E-03
AGGGAGGGGCCCAAAGC	Hs.386793	GPX3	5q23	4.23	65.36	15.47	5.42E-03
CAGGAAAGGCCAGGGGC	Hs.444106	TOR2A	9q34.11	4.23	65.36	15.47	5.42E-03
CCGTGGCTGGCGCTGCC	Hs.527989	NXN	17p13.3	4.23	65.36	15.47	5.42E-03
CGCCGACGATGCCCAGA	Hs.523847	IFI6	1p35	8.45	130.72	15.47	3.79E-04
CTGAAATGTTGCAGGCT	Hs.632815	GAGE1	Xp11.4-	4.23	65.36	15.47	5.42E-03
GCGTGGTATTTATTGTG	Hs.530749	PPFIA1	11q13.3	4.23	65.36	15.47	5.42E-03
GGAGACTTTTCATTTTG	Hs.151761	FLJ2500	17q11.2	4.23	65.36	15.47	5.42E-03
GGCCCCATTTTGTACCT	Hs.88778	CBR1	21q22.1	4.23	65.36	15.47	5.42E-03
GGGACGAGTGACGGCAG	Hs.351316	TM4SF1	3q21-q2	4.23	65.36	15.47	5.42E-03
TAAATATTCCAAAGCAG	Hs.517941	SETD2	3p21.31	4.23	65.36	15.47	5.42E-03
TAGTCATTGAGTGAGGG	Hs.28020	EPM2AIP	3p22.1	4.23	65.36	15.47	5.42E-03
TCAAGTTTTTGTTTTCT	Hs.73021	MAGEB1	Xp21.3	4.23	65.36	15.47	5.42E-03
TCCAGGAAACTTGTAAC	Hs.11590	CTSF	11q13	4.23	65.36	15.47	5.42E-03
TCTGCCAGGGTGCTCCC	Hs.633863	LRCH4	7q22	8.45	130.72	15.47	3.79E-04
TGGAACAGGATGCCCAC	Hs.373550	TGIF	18p11.3	4.23	65.36	15.47	5.42E-03
TGGTTCCAAACTACCCC	Hs.301696	KIAA153	9p13.3	4.23	65.36	15.47	5.42E-03
TTCTTGGTCACCATTTC	Hs.533709	BSCL2	11q12-q	4.23	65.36	15.47	5.42E-03
TTGAAGGGAAGAGGGGA				4.23	65.36	15.47	5.42E-03
TTGTGATGTAAATTGTG	Hs.642877	MALAT1	11q13.1	4.23	65.36	15.47	5.42E-03
TTTTAACTTATTTGAGC	Hs.253903	STOM	9q34.1	4.23	65.36	15.47	5.42E-03
AAGAACGGACAGGTCCG	Hs.380386	C1QL4	12q13.1	4.23	56.02	13.26	1.00E-02
ACTGGGATTTAGAGGAG	Hs.10848	BMS1L	10q11.2	4.23	56.02	13.26	1.00E-02
AGCCTTGACCGTCAGGC	Hs.217409	C10orf8	10q24.1	4.23	56.02	13.26	1.00E-02
AGTACCACCCTGCCTGC	Hs.129867	CIB2	15q24	4.23	56.02	13.26	1.00E-02
ATGCTCCCTGAGGAGCT	Hs.514535	LGALS3B	17q25	4.23	56.02	13.26	1.00E-02
ATGGAAAAATAAAAACC	Hs.173094	TSPYL5	8q22.1	4.23	56.02	13.26	1.00E-02
ATGGCCAGAAAGATGAA	Hs.515487	CALM3	19q13.2	4.23	56.02	13.26	1.00E-02
CAGATAAACCATTTAAA	Hs.389037	MCM3AP	21q22.3	4.23	56.02	13.26	1.00E-02
CAGCTCAGCTGGGAGCT	Hs.58414	FLNC	7q32-q3	4.23	56.02	13.26	1.00E-02
CCAGGGGAAGAAGGCAC				4.23	56.02	13.26	1.00E-02
CCCGTGCAGTTCAAACC	Hs.100914	CEP192	18p11.2	4.23	56.02	13.26	1.00E-02
CTAACGTTGCAGACCCC				4.23	56.02	13.26	1.00E-02
CTCACATTTGATTTCTG	Hs.150107	BIRC6	2p22-p2	4.23	56.02	13.26	1.00E-02
CTGAGTCTCCCAAGGCT	Hs.77269	GNAI2	3p21	4.23	56.02	13.26	1.00E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CTGCCCTCCCATCACCC	Hs.89560	IDUA	4p16.3	4.23	56.02	13.26	1.00E-02
GAAACTGAACAACCTGC	Hs.240321	DKFZp54	4q22.1	4.23	56.02	13.26	1.00E-02
GACGCGGCTGAAACGTG	Hs.652223	NCF1	7q11.23	4.23	56.02	13.26	1.00E-02
GACTCGCTCCAGACCGT	Hs.69517	LY6K	8q24.3	4.23	56.02	13.26	1.00E-02
GAGTGAGACCCAGGAGC	Hs.651190	THY1	11q22.3	4.23	56.02	13.26	1.00E-02
GCCCGTTCTCAATGAGC	Hs.477144	CCDC52	3q13.2	4.23	56.02	13.26	1.00E-02
GCGCTGGTACGTAAATA	Hs.528634	OAS3	12q24.2	4.23	56.02	13.26	1.00E-02
GCTCTGTCACAATAGGG				4.23	56.02	13.26	1.00E-02
GGCTTCACTGCGGCGGG	Hs.532634	IFI27	14q32	4.23	56.02	13.26	1.00E-02
GGGACTTGGGGGTGGGG	Hs.380857	VPS13D	1p36.22	4.23	56.02	13.26	1.00E-02
GGTTATCTGTATGGACT	Hs.518200	PARP9	3q13-q2	4.23	56.02	13.26	1.00E-02
GTCCCTAGCAAAATGCT	Hs.368982	CASP2	7q34-q3	4.23	56.02	13.26	1.00E-02
GTCCGGTGGTTTGTAGT	Hs.632480	DENND4B	1p36.13	4.23	56.02	13.26	1.00E-02
GTGTGTTTGTAATAATA	Hs.369397	TGFBI	5q31	4.23	56.02	13.26	1.00E-02
TAACACAAGCTCACAGC				4.23	56.02	13.26	1.00E-02
TACAGCCTCCAAAGGGC	Hs.507122	SETD1B	12q24.3	4.23	56.02	13.26	1.00E-02
TATGATTGAATGGGTGC	Hs.528634	OAS3	12q24.2	4.23	56.02	13.26	1.00E-02
TCCCCAGCTCTGGGAGG	Hs.472491	OXCT2	1p34	4.23	56.02	13.26	1.00E-02
TCCTGGGGCAGGGGCGG	Hs.535378	CLSTN3	12p13.3	4.23	56.02	13.26	1.00E-02
TGGATGATTTGGTGGGA	Hs.511143	ZFP106	15q15.1	4.23	56.02	13.26	1.00E-02
TTCTCGAGATGGGTGGC	Hs.25669	NCOA5	20q12-q	4.23	56.02	13.26	1.00E-02
TTGCTTTTTTCAAACAG	Hs.529006	C18orf8	18q11.2	4.23	56.02	13.26	1.00E-02
TTTACAAACCTCAAGCC	Hs.651258	STAT1	2q32.2	4.23	56.02	13.26	1.00E-02
TTTTTAGAATTGCTGTG	Hs.474783	TST	22q13.1	4.23	56.02	13.26	1.00E-02
GGAGATAGTGACTACAC	Hs.634882	ARL6IP1	16p12-p	16.90	214.75	12.71	8.39E-06
AAAAAATGAAAAGTAGA	Hs.487036	MYO5C	15q21	4.23	46.68	11.05	1.87E-02
AAATGACAAGTCCTTGG				4.23	46.68	11.05	1.87E-02
AACCCGGGAGGCGGAGC	Hs.121593	COX19	7p22.3	4.23	46.68	11.05	1.87E-02
AAGAAAAAAAAATTTGG	Hs.61188	XRCC6BP	12q14.1	4.23	46.68	11.05	1.87E-02
AATGTCCAGTAAAAATC	Hs.516813	D2HGDH	2q37.3	4.23	46.68	11.05	1.87E-02
ACAAGAACAAAACCACT	Hs.9216	CASP7	10q25	4.23	46.68	11.05	1.87E-02
ACATTAAAATATTACTC				4.23	46.68	11.05	1.87E-02
ACCAGCATAGTGCTTTG	Hs.632296	PDAP1	7q22.1	4.23	46.68	11.05	1.87E-02
ACCTTCCTAGTTTCCTG	Hs.631639	GIPC1	19p13.1	4.23	46.68	11.05	1.87E-02
ACGCAGCACGGGCAGGG				4.23	46.68	11.05	1.87E-02
ACTGTGGGAAAGTGGGC				4.23	46.68	11.05	1.87E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
AGGGTCTGCCCAAGCCC	Hs.181244	HLA-A	6p21.3	4.23	46.68	11.05	1.87E-02
ATAAAGTATCTATACAG	Hs.602900	C1D	2p13-p1	4.23	46.68	11.05	1.87E-02
ATCCTCTGCGTGGGAGG	Hs.434875	CAMK1	3p25.3	4.23	46.68	11.05	1.87E-02
ATGACTGCTGTTTGACT	Hs.150651	FAM111A	11q12.1	4.23	46.68	11.05	1.87E-02
CAACCATCATCTTCCAC				4.23	46.68	11.05	1.87E-02
CACCACGGTGTCTGTGT	Hs.125867	EVL	14q32.2	4.23	46.68	11.05	1.87E-02
CCAGGGGAGAAGGCCCC				4.23	46.68	11.05	1.87E-02
CCCACCTGCCCACACCC	Hs.25524	PTPN23	3p21.3	4.23	46.68	11.05	1.87E-02
CCGTGACAATTTTCTTG	Hs.192371	DSCR8	21q22.2	4.23	46.68	11.05	1.87E-02
CCTTTGTAAGTTATTTC	Hs.525704	JUN	1p32-p3	4.23	46.68	11.05	1.87E-02
CTACTTTTTAATTGGCC	Hs.506759	ATP2A2	12q23-q	4.23	46.68	11.05	1.87E-02
CTGGTCTGACTTATGCT	Hs.308332	RPL10L	14q13-q	4.23	46.68	11.05	1.87E-02
CTTGATTCCCACGCTAC	Hs.518374	QSCN6	1q24	4.23	46.68	11.05	1.87E-02
CTTTTCTTCAAGGAAAG	Hs.231750	C3orf9	3q13.33	4.23	46.68	11.05	1.87E-02
GAAAGGGGCTGCCCTCC	Hs.121605	HNRNPG-	11p15	4.23	46.68	11.05	1.87E-02
GAGAAGCAGCCATTGTC	Hs.594708	SH3PXD2	10q24.3	4.23	46.68	11.05	1.87E-02
GAGGGATGGCGTTGGCA	Hs.9347	RGS14	5q35.3	4.23	46.68	11.05	1.87E-02
GATGTGTGCTTCATTTG	Hs.488189	H2AFV	7p13	4.23	46.68	11.05	1.87E-02
GCACTACTCGACACCTG	Hs.352548	TMEM149	19q13.1	4.23	46.68	11.05	1.87E-02
GCATTCCTCTTTAACTG	Hs.9788	NDFIP1	5q31.3	4.23	46.68	11.05	1.87E-02
GCTCTCGGCGGCCTTCG	Hs.512417	FLJ4544	19p13.3	4.23	46.68	11.05	1.87E-02
GCTGGAATGAAGGTATC	Hs.59584	C1orf17	1p34.2	4.23	46.68	11.05	1.87E-02
GGACAGCCCACAGCCCC	Hs.633863	LRCH4	7q22	4.23	46.68	11.05	1.87E-02
GGCACAGAGGCTGGGGA				4.23	46.68	11.05	1.87E-02
GGCCCATCCCTTGATGG	Hs.42853	TNXB	6p21.3	4.23	46.68	11.05	1.87E-02
GGCGACACGAGCCGATC				4.23	46.68	11.05	1.87E-02
GGCGCCCTCATCATCCA	Hs.514631			4.23	46.68	11.05	1.87E-02
GGGGAGGGGGTACTGCC				4.23	46.68	11.05	1.87E-02
GGGGGGGGGTTTGATGT				4.23	46.68	11.05	1.87E-02
GGGTAGGGTGAGAAGGG	Hs.524484	ITGA7	12q13	4.23	46.68	11.05	1.87E-02
GGTTTCCCAGCTGTTTG	Hs.591964	RCE1	11q13	4.23	46.68	11.05	1.87E-02
GTCACAGTAGAGATCCC	Hs.286221	ARF1	1q42	4.23	46.68	11.05	1.87E-02
GTCCCTGGAGCCGAGCC	Hs.185677	NEDD4L	18q21	4.23	46.68	11.05	1.87E-02
GTCCTGCTCCTCAAGGG	Hs.64746	CLIC3	9q34.3	4.23	46.68	11.05	1.87E-02
GTCCTGGTGGTGGGGGG	Hs.652162	FOXH1	8q24.3	4.23	46.68	11.05	1.87E-02
GTGTTGGGGGTGCTGGT	Hs.55016	EPS8L2	11p15.5	4.23	46.68	11.05	1.87E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GTTCTCTTCTCTGCAGG				4.23	46.68	11.05	1.87E-02
TCAGCTGGCCCTCCAGG	Hs.465885	ILF3	19p13.2	4.23	46.68	11.05	1.87E-02
TCCTGCAATTCTGAAGG				4.23	46.68	11.05	1.87E-02
TGAGAATATAAATATTC				4.23	46.68	11.05	1.87E-02
TGATTTCACTTACCACT				4.23	46.68	11.05	1.87E-02
TGCTGACTCCCCCCATC	Hs.527971	NES	1q23.1	4.23	46.68	11.05	1.87E-02
TGCTGTGTCCACTCAGG	Hs.639263			4.23	46.68	11.05	1.87E-02
TGGCACTTCAAAAGGCA	Hs.287714	RAB32	6q24.3	4.23	46.68	11.05	1.87E-02
TTCCTTGTAATCAAAGA	Hs.208267	B3GNT5	3q28	4.23	46.68	11.05	1.87E-02
TTGAATAAATTTGTGAG				4.23	46.68	11.05	1.87E-02
TTGCCAACACCTTGAGA	Hs.80919	SYPL1	7q22.2	4.23	46.68	11.05	1.87E-02
TTTCAGAGAGCACAGGG	Hs.232751	NEK5	13q14.3	4.23	46.68	11.05	1.87E-02
TTTTAAGATAATTAAGT	Hs.133998	CRIPT	2p21	4.23	46.68	11.05	1.87E-02
TTTTCTTTAGGAATTCA	Hs.19385	ABHD5	3p21	4.23	46.68	11.05	1.87E-02
ACCATTGGATTCATCCT	Hs.458414	IFITM1	11p15.5	8.45	93.37	11.05	3.94E-03
GGCTTAGGATGTGAATG	Hs.233955	RASIP1	19q13.3	8.45	93.37	11.05	3.94E-03
TATAATAAATGGCATTG	Hs.350194	ZMAT2	5q31.3	8.45	93.37	11.05	3.94E-03
GTGTACCGGATCTCGGC	Hs.144011	PSCD2	19q13.3	16.90	168.06	9.94	1.41E-04
GATGGTGGAGTACGGCC	Hs.522087	OPRS1	9p13.3	8.45	84.03	9.94	7.13E-03
GGCTAGAAGAGGAGAGG	Hs.579243	CHRNE	17p13-p	8.45	84.03	9.94	7.13E-03
TAGGAGTTAATCCCTGT	Hs.30743	PRAME	22q11.2	8.45	84.03	9.94	7.13E-03
TATGGTGAGCGAGAGGC	Hs.631593	PPP1R15	19q13.2	8.45	84.03	9.94	7.13E-03
TCCGAGACTGCTCCTGC	Hs.7644	HIST1H1	6p21.3	33.80	326.79	9.67	1.28E-07
AAAAAAAAAGCCAATCC				4.23	37.35	8.84	3.55E-02
AAAAGCAGAAATCGGTT	Hs.508234	KLF5	13q22.1	4.23	37.35	8.84	3.55E-02
AAAATAAACCTGGTGGC	Hs.192233	PPL	16p13.3	4.23	37.35	8.84	3.55E-02
AAACTCACGCCAGGTGC	Hs.600163			4.23	37.35	8.84	3.55E-02
AAAGTTATTTACAATGA	Hs.651204	FOXJ1	17q22-1	4.23	37.35	8.84	3.55E-02
AAATAAAAAGTGCCTGA	Hs.123198	MYO9B	19p13.1	4.23	37.35	8.84	3.55E-02
AAATAGATCCACCTGCT				4.23	37.35	8.84	3.55E-02
AAATATTAAACATTTCG	Hs.213198	PSPC1	13q12.1	4.23	37.35	8.84	3.55E-02
AAATTCAGGTCTAGCTG	Hs.473877	FAM3B	21q22.3	4.23	37.35	8.84	3.55E-02
AACAGCTCTGAGATCCT	Hs.67397	HOXA1	7p15.3	4.23	37.35	8.84	3.55E-02
AACCCAGGAGGCGGAGC	Hs.135904	GK5	3q23	4.23	37.35	8.84	3.55E-02
AAGAACGCCAGGGAGCT				4.23	37.35	8.84	3.55E-02
AAGGAAGATGGTTGGGT	Hs.180062	PSMB8	6p21.3	4.23	37.35	8.84	3.55E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
AAGGATTACACTAGTCC				4.23	37.35	8.84	3.55E-02
AATCTGTAACTAGAATT	Hs.40098	GREM1	15q13-q	4.23	37.35	8.84	3.55E-02
AATGGATTAGAAATGGG				4.23	37.35	8.84	3.55E-02
AATTCCACGCGCACCTG	Hs.616282			4.23	37.35	8.84	3.55E-02
ACAAAACTGAAAGAAGC	Hs.191219	CPNE3	8q21.3	4.23	37.35	8.84	3.55E-02
ACAAAATCAATGGAAAG	Hs.552700	DKFZp54	18p11.2	4.23	37.35	8.84	3.55E-02
ACCAAAAACCAAAAGTG	Hs.649756	COL1A1	17q21.3	4.23	37.35	8.84	3.55E-02
ACCCCAGGTTCCAGTGT	Hs.514435	SF3B3	16q22.1	4.23	37.35	8.84	3.55E-02
ACCCCCAAACAGAACCC				4.23	37.35	8.84	3.55E-02
ACCCCTGCTCCAGCAAC				4.23	37.35	8.84	3.55E-02
ACCGGGACTTCATTCGG				4.23	37.35	8.84	3.55E-02
ACCTCCGAGAAGAGCCA	Hs.1897	POMC	2p23.3	4.23	37.35	8.84	3.55E-02
ACTTTAGATGGGAAGCC	Hs.233240	COL6A3	2q37	4.23	37.35	8.84	3.55E-02
AGAAGCAAGAAGTATCA				4.23	37.35	8.84	3.55E-02
AGAGCTCACTATTTTAA	Hs.13845	SLC25A2	3p21.31	4.23	37.35	8.84	3.55E-02
AGCCTCAAAACTGAATG	Hs.433442	KIFAP3	1q24.2	4.23	37.35	8.84	3.55E-02
AGCTACGGAAACAGGCC				4.23	37.35	8.84	3.55E-02
AGGAAAAGATGCTCTCC	Hs.523829	POLD4	11q13	8.45	74.70	8.84	1.29E-02
AGGCTGCCCAGAGAAAG	Hs.374043	ASXL1	20q11.1	4.23	37.35	8.84	3.55E-02
AGGGGCTGCCGAAGCCC	Hs.77961	HLA-B	6p21.3	4.23	37.35	8.84	3.55E-02
AGGTATTTCTTCCTTCC				4.23	37.35	8.84	3.55E-02
AGTGTCTGTGAGAGGCA	Hs.8867	CYR61	1p31-p2	4.23	37.35	8.84	3.55E-02
AGTTGTCACTTCTTGTT	Hs.142442	HP1BP3	1p36.12	4.23	37.35	8.84	3.55E-02
ATACAGATTGGTTTTGC				8.45	74.70	8.84	1.29E-02
ATGCTGGGGAGCTTGGC	Hs.274329	TP53AP1	7q21.1	4.23	37.35	8.84	3.55E-02
ATTTACAACAGTTGAGG	Hs.538075			4.23	37.35	8.84	3.55E-02
ATTTAGTCATAATTGTG	Hs.82316	IFI44	1p31.1	4.23	37.35	8.84	3.55E-02
CAACATAAAAAAGACAC	Hs.492128			4.23	37.35	8.84	3.55E-02
CAAGCGCTCTAATTCCT	Hs.459759	CREBBP	16p13.3	4.23	37.35	8.84	3.55E-02
CACAGTTCTCTTATACC				4.23	37.35	8.84	3.55E-02
CACGACAGTCCTTGTTC	Hs.514713	MPPE1	18p11.2	4.23	37.35	8.84	3.55E-02
CACTCAATAAAGAATGA	Hs.79361	KLK6	19q13.3	4.23	37.35	8.84	3.55E-02
CAGCTGGCCATCACCGG	Hs.24601	FBLN1	22q13.3	4.23	37.35	8.84	3.55E-02
CAGGAACCACAGTGGCT				4.23	37.35	8.84	3.55E-02
CAGGCTCAGCAGAGGGG	Hs.42964			4.23	37.35	8.84	3.55E-02
CATTCTCCATTGATAAG	Hs.652159	FLJ4505	8p23.3	4.23	37.35	8.84	3.55E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CATTTTAATACCCTAAG	Hs.369017	RAB2	8q12.1	4.23	37.35	8.84	3.55E-02
CCAATGCTATGTCCACC	Hs.519804	RMND5B	5q35.3	4.23	37.35	8.84	3.55E-02
CCATTGAAACTAAGAGC	Hs.497636	LAMB3	1q32	4.23	37.35	8.84	3.55E-02
CCCCCTGGATCAGGCCC				4.23	37.35	8.84	3.55E-02
CCCTGGGTCCTGCCCGC				4.23	37.35	8.84	3.55E-02
CCGCCTTTAAGAACTGC				4.23	37.35	8.84	3.55E-02
CCTCTGCCGGGTGCCGG	Hs.410817	RPL13	16q24.3	4.23	37.35	8.84	3.55E-02
CCTGGAAGGAACAAGGC				4.23	37.35	8.84	3.55E-02
CCTTACTTTATCAAATG				4.23	37.35	8.84	3.55E-02
CTACTTTGGAAACTGCC	Hs.302754			4.23	37.35	8.84	3.55E-02
CTATACTAATGCTGTTG				4.23	37.35	8.84	3.55E-02
CTATTTGGTCAAGGCCT	Hs.647791	GGPS1	1q43	4.23	37.35	8.84	3.55E-02
CTGAAATTCGGTGCAGC	Hs.474833	CSNK1E	22q13.1	4.23	37.35	8.84	3.55E-02
CTGGCTGTGGCCGGGGC	Hs.645243	ARVCF	22q11.2	4.23	37.35	8.84	3.55E-02
CTGTATACTTAAGAGGA				4.23	37.35	8.84	3.55E-02
GAATATGGCTACATTGC	Hs.214247	C7orf25	7p14-p1	4.23	37.35	8.84	3.55E-02
GAATTCAGCCCGATGGC				4.23	37.35	8.84	3.55E-02
GAATTCTACAGAAACCA				4.23	37.35	8.84	3.55E-02
GACCCAATCCTCTGAGC	Hs.112160	PIF1	15q22.3	8.45	74.70	8.84	1.29E-02
GACCCCAAGGCCGCCGA	Hs.523852	CCND1	11q13	8.45	74.70	8.84	1.29E-02
GACTGTATTTATTTTCA	Hs.3352	HDAC2	6q21	4.23	37.35	8.84	3.55E-02
GAGGCCAGGTCTCAGTC	Hs.134989	EN2	7q36	4.23	37.35	8.84	3.55E-02
GAGGGCCATTCCTGCCC	Hs.59804	SECISBP	9q22.2	4.23	37.35	8.84	3.55E-02
GAGGTCACCAGCACGTT	Hs.292177	SLC27A5	19q13.4	4.23	37.35	8.84	3.55E-02
GAGTAAACTGGTACCTG				4.23	37.35	8.84	3.55E-02
GAGTTGTTCAACCTGCC				4.23	37.35	8.84	3.55E-02
GATCTTCGTACCAAGAG	Hs.494691	PFN1	17p13.3	8.45	74.70	8.84	1.29E-02
GATGCAGCAGCCTGAAA	Hs.203559	MRPL44	2q36.1	4.23	37.35	8.84	3.55E-02
GATTTTAATGTTTTTCT	Hs.7370	PITPNB	22q12.1	8.45	74.70	8.84	1.29E-02
GCACGACACGAGCCGAT				4.23	37.35	8.84	3.55E-02
GCCAAGGAAGGCATTGC				4.23	37.35	8.84	3.55E-02
GCCACACACGATGAGGC				4.23	37.35	8.84	3.55E-02
GCCAGCCTCCTGGAGGA	Hs.269898	SERTAD1	19q13.1	4.23	37.35	8.84	3.55E-02
GCCCCAGTTCACTATTC				8.45	74.70	8.84	1.29E-02
GCCGCCATCCGCAGGGC				4.23	37.35	8.84	3.55E-02
GCGGCCTAACCACCCGC	Hs.582627	CCDC59	12q21.3	4.23	37.35	8.84	3.55E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GCTCTGACTGCTGTGGG	Hs.193842	TDRD7	9q22.33	16.90	149.39	8.84	4.38E-04
GCTGGCTGTTTTGTGAC	Hs.632238	SKIP	17p13.3	4.23	37.35	8.84	3.55E-02
GCTTTGATGATAAACGA	Hs.89649	EPHX1	1q42.1	4.23	37.35	8.84	3.55E-02
GCTTTTTAGGATACCGG				4.23	37.35	8.84	3.55E-02
GGAAAAATGTTGGAATG	Hs.510334	SERPINA	14q32.1	4.23	37.35	8.84	3.55E-02
GGAAGTGATCTGCAAGA				4.23	37.35	8.84	3.55E-02
GGAGCTTGAGGCAGTAA	Hs.520046	GPSM3	6p21.3	4.23	37.35	8.84	3.55E-02
GGAGGGCTGGGCACAGG	Hs.567524	C10orf1	10q24.2	4.23	37.35	8.84	3.55E-02
GGCGGCTGCCAGATCCA				4.23	37.35	8.84	3.55E-02
GGCTGATTTTCATTTTT	Hs.377830	MBOAT1	6p22.3	4.23	37.35	8.84	3.55E-02
GGGACGGGTGCCTGTAA				4.23	37.35	8.84	3.55E-02
GGTGAAGACAAGCAAGT	Hs.493796	RUSC2	9p13.3	4.23	37.35	8.84	3.55E-02
GTACTGTAGCAGGGGAA	Hs.265829	ITGA3	17q21.3	8.45	74.70	8.84	1.29E-02
GTAGATTAAAATTAAGC	Hs.444200	BPTF	17q24.3	4.23	37.35	8.84	3.55E-02
GTCAGTCACTTAATACA	Hs.601540			8.45	74.70	8.84	1.29E-02
GTCTGGAAGCGGGGGGG				4.23	37.35	8.84	3.55E-02
GTGCCGCCGAGCCCGGG	Hs.516370	CHST10	2q11.2	4.23	37.35	8.84	3.55E-02
GTGCTGGACCTGAAGGC				4.23	37.35	8.84	3.55E-02
GTGCTTTCCTGTTTCTC	Hs.75527	ADSL	22q13.1	4.23	37.35	8.84	3.55E-02
GTTCTTCACTGAGAGCC	Hs.310453	NUPL1	13q12.1	4.23	37.35	8.84	3.55E-02
GTTGCCCTGGCCCGTGC				4.23	37.35	8.84	3.55E-02
GTTGCGGTTAATCTGGT				4.23	37.35	8.84	3.55E-02
GTTTGAAGGGACTGAAC	Hs.652269	PSMD7	16q23-q	4.23	37.35	8.84	3.55E-02
GTTTTTCATTTCTGTCT	Hs.26663	HERC5	4q22.1	4.23	37.35	8.84	3.55E-02
TAAATAATAAAAGAGAG				4.23	37.35	8.84	3.55E-02
TAAATCCCCACTGGGAC	Hs.297413	MMP9	20q11.2	4.23	37.35	8.84	3.55E-02
TAACAAGATCCCAATCC	Hs.293317	PAGE5	Xp11.21	4.23	37.35	8.84	3.55E-02
TAAGCACTCTACTGCTT	Hs.479656	COX7B2	4p12	4.23	37.35	8.84	3.55E-02
TAAGTAATTGAAGTCCC	Hs.571424			4.23	37.35	8.84	3.55E-02
TACTCTATAACTTAAAA				4.23	37.35	8.84	3.55E-02
TACTTACTAATATGTTG				4.23	37.35	8.84	3.55E-02
TAGGCAACACGAGCAGG				8.45	74.70	8.84	1.29E-02
TATTTTCTTTTTTCCTT	Hs.525163	ANKRD10	13q34	4.23	37.35	8.84	3.55E-02
TCCAGCCCCTGAAGTTG	Hs.24956	C14orf1	14q32.3	4.23	37.35	8.84	3.55E-02
TCCTGAGTGCCAGTGTA	Hs.590914	LOC1507	2q11.2	4.23	37.35	8.84	3.55E-02
TCGGTGCCCGCTGGGGA	Hs.326953	NUMBL	19q13.1	8.45	74.70	8.84	1.29E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TCTCTACCCACTATGCA	Hs.370247	APLP2	11q23-q	4.23	37.35	8.84	3.55E-02
TCTTCTCACAAGCCAGC	Hs.656	CDC25C	5q31	4.23	37.35	8.84	3.55E-02
TGAAACGTGCTTGAAGG	Hs.302513	BEST4	1p33-p3	4.23	37.35	8.84	3.55E-02
TGAAAGTAACAAGCAAA				4.23	37.35	8.84	3.55E-02
TGACCAGGGTCCCAAGG	Hs.590876	TBC1D20	20p13	4.23	37.35	8.84	3.55E-02
TGACCATCAATAAAGTA				4.23	37.35	8.84	3.55E-02
TGAGGACTCAATGAGGT				4.23	37.35	8.84	3.55E-02
TGAGTTGGGCAAATTGT	Hs.592095	SLC16A5	17q25.1	4.23	37.35	8.84	3.55E-02
TGATTTTTTTTCCTCC	Hs.373550	TGIF	18p11.3	4.23	37.35	8.84	3.55E-02
TGGAACCCTGATGCAGC				4.23	37.35	8.84	3.55E-02
TGGACAAGCTAAGTGGG				4.23	37.35	8.84	3.55E-02
TGGAGGATGAATCTGCT	Hs.591692	LOC9082	4q31.23	4.23	37.35	8.84	3.55E-02
TGGGCAGGGACAAATAA				4.23	37.35	8.84	3.55E-02
TGGGCCTTCCCCAGGAG	Hs.533712	RBM4	11q13	4.23	37.35	8.84	3.55E-02
TGGGGTCCCCAGCCTCG	Hs.461727	LOC1973	16q24.3	4.23	37.35	8.84	3.55E-02
TGTGTGGCTTTAGTGGG	Hs.652265	CD2BP2	16p11.2	4.23	37.35	8.84	3.55E-02
TTACGACTTGTCTCCTC				4.23	37.35	8.84	3.55E-02
TTACTGACAAATTTGTG	Hs.598295			4.23	37.35	8.84	3.55E-02
TTCCCCTTCCTGGGAGG	Hs.591936	SRPR	11q24.3	4.23	37.35	8.84	3.55E-02
TTGATTGAGTGAATCGT	Hs.593327			4.23	37.35	8.84	3.55E-02
TTTTGTTAATCCGCTTC	Hs.368855	GMPR2	14q12	4.23	37.35	8.84	3.55E-02
TTTTTATTTTTATTGCT	Hs.440263	TP53RK	20q13.2	4.23	37.35	8.84	3.55E-02
CTGCCGAGCTCTGGAAA	Hs.93002	UBE2C	20q13.1	16.90	140.05	8.29	7.72E-04
AAGAAAACTGTGTTTCA	Hs.591502	KIAA152	1p35.1	8.45	65.36	7.73	2.35E-02
ACCAAATATTTTGTATC	Hs.631618	TPM4	19p13.1	8.45	65.36	7.73	2.35E-02
AGTGCCCTGAGCAGGCC	Hs.444046	NETO2	16q11	8.45	65.36	7.73	2.35E-02
ATGTGTAACGAATTCTT	Hs.557609	S100A4	1q21	8.45	65.36	7.73	2.35E-02
ATTGCTTTTGAGGGCCC	Hs.525527	RER1	1pter-q	8.45	65.36	7.73	2.35E-02
CCTGATGAAGAGTTTAG	Hs.169615	C8orf33	8q24.3	8.45	65.36	7.73	2.35E-02
CTATGGTAATGCACTTG	Hs.556296	DNM1L	12p11.2	8.45	65.36	7.73	2.35E-02
CTGGTGATGGCTCCCCT	Hs.469171	C1orf16	1p36.11	8.45	65.36	7.73	2.35E-02
GCAAGCCCCAGGCTAGC	Hs.295563	C12orf4	12q23.3	8.45	65.36	7.73	2.35E-02
GGAAGAGCACTGGTGTG	Hs.591947	ST3GAL4	11q23-q	8.45	65.36	7.73	2.35E-02
TAAATTAGAAGGGACAG				8.45	65.36	7.73	2.35E-02
TATGTAATATGCTTTCT	Hs.435122	PPAP2A	5q11	8.45	65.36	7.73	2.35E-02
TATTGAATGAGTGAACC	Hs.279912	CP110	16p12.3	8.45	65.36	7.73	2.35E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TGTGAACCTACAACACC	Hs.468972	ARID1A	1p35.3	8.45	65.36	7.73	2.35E-02
TTTCTGCCTTCTTTATG	Hs.532357	TRIM21	11p15.5	8.45	65.36	7.73	2.35E-02
AATCTGCGCCTGCGGGG	Hs.458485	ISG15	1p36.33	135.20	952.36	7.04	0.00E+00
TTAATTTCTCAGCCCCC	Hs.514920	CALCOCO	17q21.3	16.90	112.04	6.63	4.20E-03
AAATCCTTCTATGGGGG	Hs.608420			8.45	56.02	6.63	4.29E-02
AAATGTAATTTACTTGG	Hs.525419	LIMA1	12q13	8.45	56.02	6.63	4.29E-02
AAGCTGAGAAACGAGAG	Hs.209983	STMN1	1p36.1-	8.45	56.02	6.63	4.29E-02
AAGGGAGGAAATTTTGG	Hs.546261	HNRPA1	12q13.1	8.45	56.02	6.63	4.29E-02
AAGGTAGATGTGGGTGG	Hs.652404	H1F0	22q13.1	8.45	56.02	6.63	4.29E-02
AAGTTCCAGAACCAGAA	Hs.567567	NCAPG	4p15.33	8.45	56.02	6.63	4.29E-02
AATAAAAGTGGATTTCA	Hs.591953	PLCB3	11q13	8.45	56.02	6.63	4.29E-02
AATGGAGACTTCTAATT	Hs.652150	SRP19	5q21-q2	8.45	56.02	6.63	4.29E-02
ATTTTGTGTCAAGATGA	Hs.512815	AP3D1	19p13.3	8.45	56.02	6.63	4.29E-02
CAAAATACTGCAGATTT	Hs.652133	VEZF1	17q22	8.45	56.02	6.63	4.29E-02
CACGCTCACTCTGCTTC	Hs.339809	CPNE2	16q13	8.45	56.02	6.63	4.29E-02
CAGTTTGAAATTCTGAA	Hs.444082	EZH2	7q35-q3	8.45	56.02	6.63	4.29E-02
CCTGTAGATGGATTGGG				8.45	56.02	6.63	4.29E-02
CGCCTGCCGCGCCGGGC	Hs.211282	CRELD2	22p13	8.45	56.02	6.63	4.29E-02
CGGATTCTAGTGACAGG	Hs.404119	TSTA3	8q24.3	8.45	56.02	6.63	4.29E-02
CTACTGTACCACAGATT	Hs.602559			8.45	56.02	6.63	4.29E-02
CTGGTCCTCCTGGCCCA	Hs.557655	C19orf2	19p13.3	8.45	56.02	6.63	4.29E-02
GACGACACCAGCCGATC				8.45	56.02	6.63	4.29E-02
GCACAGTAACCAAATCC				8.45	56.02	6.63	4.29E-02
GGACCTTCCTCCGGCGG	Hs.611909			8.45	56.02	6.63	4.29E-02
GGGCCAGGAAAGTCTGG	Hs.513829	MC1R	16q24.3	8.45	56.02	6.63	4.29E-02
GGGGCTCCAGCCTCAGG				8.45	56.02	6.63	4.29E-02
GTCCCTCTCAAGTATCT	Hs.632184	CARHSP1	16p13.2	8.45	56.02	6.63	4.29E-02
GTCGGGCCTCTGACAGC	Hs.73769	FOLR1	11q13.3	8.45	56.02	6.63	4.29E-02
GTGGCCAGAGGTGTCAC	Hs.1420	FGFR3	4p16.3	8.45	56.02	6.63	4.29E-02
TAAATAAAATTTCAAGG				8.45	56.02	6.63	4.29E-02
TAAGATTGAGCATTTTG	Hs.514016	C17orf5	17p11.2	8.45	56.02	6.63	4.29E-02
TAATAAATAAATTAGCC	Hs.449880	LOC4404	17q12	8.45	56.02	6.63	4.29E-02
TACCAGCGGAAGGAAGG	Hs.532872	FLJ4012	19q13.3	8.45	56.02	6.63	4.29E-02
TAGGTCACAGGTTGAGG	Hs.207157	ANKDD1A	15q22.3	8.45	56.02	6.63	4.29E-02
TCACACTGGCTATCAAA	Hs.525198	STIL	1q32|1p	8.45	56.02	6.63	4.29E-02
TCCAAAGTAATGGAGAT	Hs.88556	HDAC1	1p34	8.45	56.02	6.63	4.29E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TCCCACGTTCTCTGCTG	Hs.534465	PSENEN	19q13.1	8.45	56.02	6.63	4.29E-02
TGATTATCGACCATTCG	Hs.234775	NRTN	19p13.3	8.45	56.02	6.63	4.29E-02
TGCTTTAGTGGGCTTTG	Hs.353175	AGPAT4	6q26	8.45	56.02	6.63	4.29E-02
TTACTTTGAGATGCTAG	Hs.533440	WWP1	8q21	8.45	56.02	6.63	4.29E-02
TACCCCAGAACTTAAAG	Hs.132526	LOC3750	1q12	50.70	317.45	6.26	2.05E-06
TGGGGTTCTTGGTGTGG	Hs.272499	DHRS2	14q11.2	447.86	2763.72	6.17	0.00E+00
CTGGAGACTCTGCGGGA	Hs.434044	PRKCDBP	11p15.4	16.90	102.71	6.08	7.38E-03
CTGGCAGTCCTCACAGC	Hs.569018			16.90	102.71	6.08	7.38E-03
GAGAGGTTGATATGCAC	Hs.194709	PNMA1	14q24.3	16.90	102.71	6.08	7.38E-03
GATGTTAATTGAGAGCC	Hs.171501	USP11	Xp11.23	16.90	102.71	6.08	7.38E-03
GGATGCAAGGCCGAAGC	Hs.534341	MAP4K2	11q13	16.90	102.71	6.08	7.38E-03
GGAGGAATTCATCTTCA	Hs.418123	CTSL	9q21-q2	33.80	205.41	6.08	1.51E-04
AGGTCAGGAGATCGAGA	Hs.123072	RAB3B	1p32-p3	33.80	196.07	5.80	2.58E-04
CCCGGCCTTAAAATGCC	Hs.81907	C5orf33	5p13.2	16.90	93.37	5.52	1.29E-02
GAAAGTGCAGAAAGAGG	Hs.631890	VRK2	2p16-p1	16.90	93.37	5.52	1.29E-02
GAGACTCCTGCCCTGTT	Hs.473721	SLC2A1	1p35-p3	16.90	93.37	5.52	1.29E-02
GTACGTCCCACCCTGTC	Hs.631582	SLC1A5	19q13.3	16.90	93.37	5.52	1.29E-02
TGCTTGGGCACTGGTGG	Hs.523004	PSAP	10q21-q	16.90	93.37	5.52	1.29E-02
TTACCATTGGTTTATTC	Hs.194718	ZRANB2	1p31	16.90	93.37	5.52	1.29E-02
CTGACCTGTGTTTCCTC	Hs.77961	HLA-B	6p21.3	109.85	597.56	5.44	3.89E-10
CCTCAGGATACTCCTCA				25.35	130.72	5.16	4.01E-03
TCAACAGCCAGACAGGG	Hs.128420	VPS4A	16q22.1	25.35	130.72	5.16	4.01E-03
TCAATAAAGAACAGCTA	Hs.79322	QARS	3p21.3-	25.35	130.72	5.16	4.01E-03
AAAGTCTAGAAATAAAA	Hs.523852	CCND1	11q13	16.90	84.03	4.97	2.27E-02
CTAGAAGTACATTCTCT	Hs.200136	ACAA2	18q21.1	16.90	84.03	4.97	2.27E-02
GAGACCAGCGCGAAGGC	Hs.642874	RAB8A	19p13.1	16.90	84.03	4.97	2.27E-02
GCCTCCTGAGTGATGGG	Hs.443258	SREBF2	22q13	16.90	84.03	4.97	2.27E-02
GGATTTCACTTCCACTC				16.90	84.03	4.97	2.27E-02
GTCACCCAAAAGCAACT	Hs.469018	STAMBP	2p13.1	16.90	84.03	4.97	2.27E-02
TGCTAGATTGGAGTGGG	Hs.584654	MLLT7	Xq13.1	16.90	84.03	4.97	2.27E-02
TGTTGTTACAATTTTCT	Hs.513145	NGRN	15q26.1	16.90	84.03	4.97	2.27E-02
TTTGGAGTTTGTAGTGG	Hs.643279	EIF4EBP	10q21-q	16.90	84.03	4.97	2.27E-02
TCTCCTGCATAGCTTTT	Hs.424414	MSX1	4p16.3-	59.15	289.44	4.89	2.60E-05
GTGACTGAGTCTATGGG	Hs.584807	TCF19	6p21.3	25.35	121.38	4.79	6.87E-03
TAGTCATCTTCAAAAAG	Hs.500645	ALDH18A	10q24.3	25.35	121.38	4.79	6.87E-03
CACTACTCACCAGACGC	Hs.631491			304.21	1409.87	4.63	0.00E+00

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TCTGCCTGGGGCCCTTC	Hs.24379	TRAPPC1	17p13.1	25.35	112.04	4.42	1.17E-02
AACTCAGTGTTTATTCC	Hs.435004	SEC23IP	10q25-q	16.90	74.70	4.42	3.96E-02
ACCGCTTGTTTTTCTCC	Hs.65758	ITPR3	6p21	16.90	74.70	4.42	3.96E-02
AGGGTACGGAAACAGGC				16.90	74.70	4.42	3.96E-02
CAACCCACGCTCGGTCC	Hs.26593	HDAC10	22q13.3	16.90	74.70	4.42	3.96E-02
CACTGCCTTGGTGACCA	Hs.652247	BEST1	11q13	16.90	74.70	4.42	3.96E-02
GACTGCCGCCCACTGCC	Hs.2430	VPS72	1q21	16.90	74.70	4.42	3.96E-02
GCCCATTGGAGCACCTG	Hs.631988	DDR1	6p21.3	16.90	74.70	4.42	3.96E-02
GGAGAACCTAGGGGAGG				16.90	74.70	4.42	3.96E-02
GGATTTGGAGTTAGGTG	Hs.388116	DVL3	3q27	16.90	74.70	4.42	3.96E-02
GTGGAGGTGCGCAAGGA	Hs.122523	SND1	7q31.3	16.90	74.70	4.42	3.96E-02
TAACTGGAGGATGTGCT	Hs.502378	LENG8	19q13.4	16.90	74.70	4.42	3.96E-02
TACTGGAAGTGGATAAC	Hs.268887	STK17A	7p12-p1	16.90	74.70	4.42	3.96E-02
TAGCTAATATTTTTTGG	Hs.309090	SFRS7	2p22.1	16.90	74.70	4.42	3.96E-02
TAGCTTAAAAGGAAACC	Hs.516859	PANK2	20p13	16.90	74.70	4.42	3.96E-02
TATCTGGAGATAGGTAG	Hs.598672			16.90	74.70	4.42	3.96E-02
CAAATGCTGTATTCTTC	Hs.651258	STAT1	2q32.2	33.80	149.39	4.42	3.61E-03
TCCCTGGCTGTTGAGGC	Hs.523004	PSAP	10q21-q	67.60	280.11	4.14	1.04E-04
CGTTCCTGCGGACGATC	Hs.504609	ID1	20q11	33.80	140.05	4.14	6.06E-03
TCGTAACGAGGTGGGCC	Hs.408427	COMMD7	20q11.2	33.80	140.05	4.14	6.06E-03
AGACAAGCTGGGGAAGT	Hs.632326	SFRS5	14q24	25.35	102.71	4.05	1.99E-02
AGGAAGAAACAAAAGCC	Hs.98133	PUS3	11q24.2	25.35	102.71	4.05	1.99E-02
GAAATACAGTGGGAAAA	Hs.67201	NT5C	17q25.1	25.35	102.71	4.05	1.99E-02
GAGAAGCGGCGCCGAGC	Hs.154029	HES4	1p36.33	25.35	102.71	4.05	1.99E-02
GTGGGGACGGCTGGAGG	Hs.132753	FBXO2	1p36.22	25.35	102.71	4.05	1.99E-02
TAAAAAAGGGTTGGGGG	Hs.520182	TRAM2	6p21.1-	25.35	102.71	4.05	1.99E-02
TTTTATAAACAGGACCC				25.35	102.71	4.05	1.99E-02
CTTGTGAACTGCACAAC	Hs.532803	HN1	17q25.1	42.25	168.06	3.98	3.14E-03
ATGCAGGCGCTGCTGGG	Hs.644596	TNNI3	19q13.4	33.80	130.72	3.87	1.01E-02
GCTGGCAGGCCAGAGCC	Hs.439777	CPT1B	22q13.3	42.25	158.73	3.76	5.18E-03
CCGGTTGGCAATTGTCA	Hs.353163	TMEM99	17q21.2	25.35	93.37	3.68	3.37E-02
CTGGGATCATCGGGGGA	Hs.213541	LOC5528	12q21	50.70	186.74	3.68	2.67E-03
GAAAGTGGCTGTCCTGG	Hs.627467			25.35	93.37	3.68	3.37E-02
TAGCTGCCTTTGTTACT	Hs.118640	DVL2	17p13.2	25.35	93.37	3.68	3.37E-02
TCAGAACAGTCCAGACT	Hs.309763	GRSF1	4q13	25.35	93.37	3.68	3.37E-02
TTCTTGAACAATCAGGT	Hs.527861	OS9	12q13	25.35	93.37	3.68	3.37E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TCTACTGTTAGGTGAGG	Hs.437338	NDRG3	20q11.2	33.80	121.38	3.59	1.68E-02
AGATAACTCAAGAAATC	Hs.417816	MAGEA3	Xq28	42.25	149.39	3.54	8.48E-03
GCCGCAGACACGCCGGG	Hs.154029	HES4	1p36.33	42.25	149.39	3.54	8.48E-03
CCAAGGAATGGAATTTC	Hs.534679	LOC1300	2q21.1	33.80	112.04	3.31	2.77E-02
CCTGTAATCCCAGCTAC	Hs.181301	CTSS	1q21	33.80	112.04	3.31	2.77E-02
CGCAGTGTCCTAGTGCC	Hs.389107	ATP6V0C	16p13.3	84.50	280.11	3.31	4.97E-04
TCCCCATAAGCAGAGCT	Hs.486507	TBPL1	6q22.1-	33.80	112.04	3.31	2.77E-02
CAGTGGGTGTGGGGGGG	Hs.533709	BSCL2	11q12-q	50.70	168.06	3.31	6.99E-03
AGCTCTATGATCTGGAG	Hs.531330	CBWD1	9p24.3	42.25	140.05	3.31	1.38E-02
CCCCCTGGATCAGGCCA	Hs.275243	S100A6	1q21	397.16	1279.15	3.22	2.42E-13
GGGCAGGCGTGGGGAGC	Hs.501629	IER2	19p13.1	76.05	242.76	3.19	1.51E-03
GGTACCCATTTGATAAG	Hs.298654	DUSP6	12q22-q	50.70	158.73	3.13	1.12E-02
GTGCACTGAGCTGCAAC	Hs.77961	HLA-B	6p21.3	50.70	158.73	3.13	1.12E-02
GGAGGTGTGGTTTATTG	Hs.143733	CCDC34	11p14.1	42.25	130.72	3.09	2.23E-02
GTGTGCTTAGAAACTGC	Hs.72071	KCTD9	8p21.1	42.25	130.72	3.09	2.23E-02
TACCCTAAAACTTAAAG	Hs.276252	ATRN	20p13	84.50	261.43	3.09	1.23E-03
AGCACTGCAGCGATTGC	Hs.532790	NMT1	17q21.3	33.80	102.71	3.04	4.52E-02
CATTGCCTTCATTTATT	Hs.542262			33.80	102.71	3.04	4.52E-02
CGGGCCGTGCGCAGGGA	Hs.157394	HAGH	16p13.3	33.80	102.71	3.04	4.52E-02
CTCTCCGCCCTCGCCCG	Hs.520967	MDH2	7p12.3-	33.80	102.71	3.04	4.52E-02
GAGATGTAAAATGAGTT	Hs.435001	KLF10	8q22.2	33.80	102.71	3.04	4.52E-02
GAGGTGTTCTGTTTGGG	Hs.269944	MTCH2	11p11.2	33.80	102.71	3.04	4.52E-02
GGTTTGGAAGACCCAGC	Hs.562083	ICMT	1p36.21	33.80	102.71	3.04	4.52E-02
TCAATAAATGTTCTTCT	Hs.514950	SCPEP1	17q22	33.80	102.71	3.04	4.52E-02
TCAGAAAATTCAGAGGG	Hs.119882	CDK6	7q21-q2	33.80	102.71	3.04	4.52E-02
TTGAGCCAGCCAGAGAA	Hs.91142	KHSRP	19p13.3	33.80	102.71	3.04	4.52E-02
TGATTTCACTTCCACTC				1140.78	3454.65	3.03	0.00E+00
CCTGGAAGAGGAGCTGG	Hs.464336	P4HB	17q25	92.95	280.11	3.01	9.99E-04
GGGTTGGCTTGAAACCA				59.15	177.40	3.00	9.03E-03
AGACCCACAACAAATAG				84.50	252.10	2.98	1.92E-03
GTGCTGATTCTGGGGGG	Hs.476218	COL7A1	3p21.1	50.70	149.39	2.95	1.78E-02
TAAGTAGCAAACAGGGC	Hs.652143	ITM2B	13q14.3	76.05	224.09	2.95	3.72E-03
AGTGCCACGGGGAAGGG	Hs.435621	ROBO3	11q24.2	42.25	121.38	2.87	3.57E-02
CCTGTTCTCCTTCCTTG	Hs.640836	C6orf48	6p21.3	42.25	121.38	2.87	3.57E-02
CGAGGGGCCAGCAGAGG	Hs.270291	ACTN4	19q13	202.81	569.55	2.81	7.49E-06
GACCTAAAGCCAGAACC	Hs.471403	RNF25	2q35	84.50	233.42	2.76	4.60E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CCAAGGATTGGCTTTGG	Hs.9003	C16orf5	16p11.2	50.70	140.05	2.76	2.81E-02
TGGGTGAGCCAGTGGAA	Hs.520898	CTSB	8p22	50.70	140.05	2.76	2.81E-02
TACATAATTACTAATCA	Hs.523789	TncRNA	11q13.1	59.15	158.73	2.68	2.21E-02
AATGCTTTGTTACAGAC	Hs.524390	TUBA3	12q12-1	67.60	177.40	2.62	1.74E-02
GAGGACCCAACAGGAGG	Hs.109	CDK10	16q24	135.20	354.80	2.62	7.66E-04
AAGCTGAGGTCTTGAAG	Hs.209983	STMN1	1p36.1-	194.36	504.19	2.59	7.01E-05
CCTTGGGCCTAGAATAC	Hs.213724	SUPT16H	14q11.2	76.05	196.07	2.58	1.36E-02
TCCAAAAGGAAAATAAG	Hs.309763	GRSF1	4q13	50.70	130.72	2.58	4.40E-02
GGAGGGGGCTTGAAGCC	Hs.594444	LMNA	1q21.2-	169.00	429.50	2.54	3.06E-04
TGTGCTCGGGGTTGGCT	Hs.595071	GANAB	11q12.3	84.50	214.75	2.54	1.07E-02
GCAAGCCAACGCCACTT				59.15	149.39	2.53	3.41E-02
GAAATACAGTTGTTGGC	Hs.121575	CTSD	11p15.5	211.26	532.20	2.52	6.58E-05
TCCCTGTAGTCGGTAGG	Hs.9788	NDFIP1	5q31.3	152.10	382.81	2.52	7.20E-04
GTTGTGGTTAATCTGGT	Hs.534255	B2M	15q21-q	329.56	812.31	2.46	1.28E-06
ACAGGGTGACCCTGGAG	Hs.174050	EDF1	9q34.3	169.00	410.82	2.43	6.65E-04
CCTGTACCCCAGATGGG	Hs.406534	HMG20B	19p13.3	67.60	158.73	2.35	4.02E-02
CTGTGCAGCAAGAACCC	Hs.76244	SRM	1p36-p2	67.60	158.73	2.35	4.02E-02
TCAAATCACATTGAAGC	Hs.143703	EHD4	15q11.1	67.60	158.73	2.35	4.02E-02
GACTCACTTTTGTAACA	Hs.434937	PPIB	15q21-q	76.05	177.40	2.33	3.11E-02
GATTTTTCATCTTCTAC	Hs.475125	ATXN10	22q13.3	76.05	177.40	2.33	3.11E-02
AGGGACATAAATGGGCC	Hs.134846	C16orf4	16p13.3	84.50	196.07	2.32	2.41E-02
GTGCACTGAGCTGTAAC	Hs.181244	HLA-A	6p21.3	84.50	196.07	2.32	2.41E-02
ACTGAAGAATTAACAGC	Hs.452319			354.91	793.64	2.24	1.16E-05
ATCCCTCAGTGCATAAA	Hs.496487	ATF4	22q13.1	92.95	205.41	2.21	2.75E-02
CCACAGGAGAATTCGGG	Hs.520421	PERP	6q24	228.16	504.19	2.21	5.53E-04
GTGCACCGAGTGATTTC	Hs.505924	HMGA2	12q15	84.50	186.74	2.21	3.56E-02
TCGAAGCCCCCATCGCT				76.05	168.06	2.21	4.62E-02
GTGCCTGAGAGGCAGGC	Hs.594444	LMNA	1q21.2-	101.40	224.09	2.21	2.13E-02
AAGCCAGGACAGAGGCC	Hs.10326	COPE	19p13.1	295.76	644.25	2.18	1.19E-04
GCCTATGGTCCTGCCCA	Hs.409834	PHPT1	9q34.3	101.40	214.75	2.12	3.10E-02
TCATCTTCAACTACAAG	Hs.515162	CALR	19p13.3	219.71	448.17	2.04	2.79E-03
GACTGTGCCACACACCC	Hs.5120	DYNLL1	12q24.2	371.81	756.29	2.03	1.08E-04
CCCTTAGCTTTACAGCT	Hs.190086	MRCL3	18p11.3	101.40	205.41	2.03	4.46E-02
CAGGAGTTCAAAGAAGG	Hs.529303	ARPC2	2q36.1	287.31	578.89	2.01	8.05E-04
CTTCCAGCTAACAGGTC	Hs.511605	ANXA2	15q21-q	515.46	1036.40	2.01	7.70E-06
CAGGGCGGGTTTGGAGG	Hs.23978	SAFB	19p13.3	169.00	336.13	1.99	1.19E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TGCTTGTCCCTGTGCTC	Hs.286221	ARF1	1q42	160.55	317.45	1.98	1.52E-02
GCAGCCATCCGCAGGGC	Hs.652114	RPL28	19q13.4	1960.45	3874.81	1.98	0.00E+00
GCCTGGCCATCTTGTTG	Hs.426359	PRR13	12q12	118.30	233.42	1.97	3.79E-02
GAGGCCGCGGGGTGGGG				152.10	298.78	1.96	1.95E-02
GCAACAGCAATAGGATT	Hs.488282	SEC61G	7p11.2	152.10	298.78	1.96	1.95E-02
AAAGTGAAGATCTGGCT	Hs.356766	C20orf1	20q13.1	177.45	345.47	1.95	1.30E-02
CCACTGCACTCCAGCCT	Hs.107003	CCNB1IP	14q11.2	135.20	261.43	1.93	3.21E-02
TGAGGGGTGAACCTTGG	Hs.268530	GPS1	17q25.3	152.10	280.11	1.84	3.75E-02
TGGTGACAGTTGTGTGT	Hs.488189	H2AFV	7p13	152.10	280.11	1.84	3.75E-02
CCCTCCTGGACAAGGCT	Hs.518805	HMGA1	6p21	194.36	354.80	1.83	2.07E-02
CAGCGCCACCTGGAAGC	Hs.515005	STK11	19p13.3	169.00	308.12	1.82	3.14E-02
GCTAAGGAGATTGGTGC	Hs.413812	RAC1	7p22	228.16	410.82	1.80	1.46E-02
TCAGCCTTCTGATGATC	Hs.179986	FLOT1	6p21.3	236.61	410.82	1.74	2.08E-02
GACGACACGAGCCGATC	Hs.322473	RPS28	19p13.2	1715.40	2922.45	1.70	2.11E-09
ACTCAATAAAAGTTTCC	Hs.104019	TACC3	4p16.3	202.81	345.47	1.70	3.97E-02
TAATAAATGCTGCAGCC	Hs.437594	TSPAN4	11p15.5	253.51	429.50	1.69	2.30E-02
CTCCCTCCTCTCCTACC	Hs.515122	TK1	17q23.2	321.11	541.54	1.69	1.13E-02
CAACTTAGTTTCACAGC	Hs.464472	MRLC2	18p11.3	295.76	476.18	1.61	2.85E-02
AAGTTGCTATTAAATGG	Hs.523004	PSAP	10q21-q	304.21	485.52	1.60	2.95E-02
GTAGGGGTAAAAGGAGG				1292.88	2054.12	1.59	8.95E-06
ACTAACACCCTTAATTC				828.12	1307.17	1.58	4.69E-04
CACCTAATTGGAAGCGC				1005.58	1587.27	1.58	1.16E-04
ATTTGTCCCAGCCTGGG	Hs.518805	HMGA1	6p21	1064.73	1661.97	1.56	1.14E-04
TTATGGGGAGGGGAGGG	Hs.337295	STIP1	11q13	752.07	1167.11	1.55	1.40E-03
TCCATCAAGAAATTATG	Hs.79387	PSMC5	17q23-q	422.51	653.58	1.55	1.75E-02
GTGCTGGACCTGAGGGC	Hs.512410	PSME2	14q11.2	743.62	1148.44	1.54	1.70E-03
TTGCTTTTGTTATCAGC	Hs.652183	ARF4	3p21.2-	388.71	588.22	1.51	3.14E-02
ATACTTTAATCAGAAGC	Hs.480653	ANXA5	4q26-q2	414.06	625.57	1.51	2.70E-02
GGCCCTGAGCGTCCTAC	Hs.441072	POLR2L	11p15	473.21	700.27	1.48	2.55E-02
TCTGCAAATTAGGAGGG	Hs.193491	TUBB6	18p11.2	726.72	1073.74	1.48	5.84E-03
AGGTCCTAGCCCCTGGC	Hs.523836	GSTP1	11q13	430.96	634.91	1.47	3.53E-02
ATGTACTAAAGTAGAGC	Hs.518123	TFG	3q12.2	507.01	746.95	1.47	2.24E-02
TTGGGGTTTCCTTTACC	Hs.524910	FTH1	11q13	1343.59	1970.08	1.47	2.43E-04
TGCATCTGGTGTAGGAA	Hs.605502	HSPA5	9q33-q3	676.02	952.36	1.41	2.11E-02
TAGGTAGCTCATTCAGG				1005.58	1409.87	1.40	5.58E-03
TTCATACACCTATCCCC				3371.64	4500.38	1.33	1.79E-05
GCCCCCAATAAAGGCAG	Hs.445351	LGALS1	22q13.1	2011.15	2483.61	1.23	1.77E-02

Appendix 2.2 Differentially downregulated genes (pvalue<0.05)

Tag	UniGene ID	Gene symbol	cytoband	wt_norm	dko_norm	ratio	twotailp
CTGGGTTAATAAATTGC	Hs.438429	RPS19	19q13.2	5965.86	5135.29	0.86	8.15E-03
TGGTGTTGAGGAAAGCA	Hs.627414	RPS18	6p21.3	3870.20	3258.58	0.84	1.52E-02
TGTGTTGAGAGCTTCTC	Hs.644639	EEF1A1	6q14.1	5416.60	4537.73	0.84	3.13E-03
GCGCTGGAGTGAGATGG	Hs.110695	SF3B5	6q24.2	2036.50	1661.97	0.82	3.92E-02
TTCAATAAAAAGCTGAA	Hs.356502	RPLP1	15q22	1571.74	1241.81	0.79	3.74E-02
CCAGTGGCCCGGAGCTG	Hs.546288	RPS9	19q13.4	3287.14	2586.32	0.79	2.20E-03
AGAACAAAACCTCTTCT	Hs.180909	PRDX1	1p34.1	1132.33	868.33	0.77	4.84E-02
GATCCCAACATTGTTGG	Hs.406510	ATP5B	12q13.1	1461.89	1101.75	0.75	1.74E-02
ATTATTTTTCTAAGCTG	Hs.571841	RPL7	8q21.11	1571.74	1176.45	0.75	1.17E-02
CTGCTATACGAGAGAAT	Hs.532359	RPL5	1p22.1	946.43	700.27	0.74	4.26E-02
GAAACAAGATGAAATTC	Hs.652416	PGK1	Xq13	1926.65	1419.21	0.74	3.36E-03
GAAGTTATGAAGATGCT	Hs.363137	TCP1	6q25.3-	1842.15	1335.18	0.72	2.64E-03
TTGGTGAAGGAAGAAGT	Hs.522584	TMSB4X	Xq21.3-	1360.49	980.37	0.72	8.66E-03
GAAGCTTTGCAGGCTGG	Hs.523560	HSP90AA	14q32.3	1309.79	943.03	0.72	9.82E-03
GCCGAGGAAGGCATTGC	Hs.546289	RPS12	6q23.2	2763.22	1988.76	0.72	1.72E-04
CTCAACATCTCCCCCTT	Hs.546285	RPLP0	12q24.2	1174.58	840.32	0.72	1.28E-02
GTGAAGGCAGTAGTTCT	Hs.356572	RPS3A	4q31.2-	2340.71	1661.97	0.71	3.36E-04
AGAAAGATGTCTATGTA	Hs.494173	ANXA1	9q12-q2	752.07	532.20	0.71	4.04E-02
GCTGGCTGGCTGCTGGG	Hs.368149	CCT7	2p13.2	1132.33	793.64	0.70	9.94E-03
GCCGTTCTTAGTTGGTG	Hs.631491			760.52	532.20	0.70	3.39E-02
AGCAGGGCTCCTCGTGC	Hs.474596	LIMK2	22q12.2	1149.23	802.97	0.70	8.86E-03
GGCAAGAAGAAGATCGC	Hs.514196	RPL27	17q21.1	2704.07	1876.72	0.69	4.40E-05
GAATTAACATTAAACTT	Hs.513851	YWHAE	17p13.3	633.77	438.83	0.69	4.69E-02
AACGCGGCCAATGTGGG	Hs.407995	MIF	22q11.2	2704.07	1867.38	0.69	3.54E-05
GGAGAAGATGAGAAGCC	Hs.492516	PFDN2	1q23.3	853.47	588.22	0.69	1.97E-02
GGCACAGTAAAGGTGGC	Hs.500874	CUEDC2	10q24.3	1842.15	1251.14	0.68	3.86E-04
TAAATAATACATTGTTC	Hs.389649	EIF4A3	17q25.3	785.87	532.20	0.68	1.97E-02
GAGGCCATCCCCAACCC	Hs.512610	LSM7	19p13.3	566.17	382.81	0.68	4.69E-02
TAATAAAGGTGTTTATT	Hs.512675	RPS8	1p34.1-	2957.58	1979.42	0.67	3.30E-06
AAGGAGATGGGAACTCC	Hs.469473	RPL31	2q11.2	1723.85	1148.44	0.67	3.37E-04
GTTCCCTGGCCCGTGCT	Hs.387208	FAU	11q13	743.62	494.86	0.67	1.83E-02
GAAAAATGGTTGATGGA	Hs.449909	RPSA	3p22.2	8103.77	5387.39	0.66	4.88E-15
TGAGGGAATAAACCTGG	Hs.524219	TPI1	12p13	7909.41	5247.33	0.66	7.99E-15
GCCTGTATGAGAAGAAA	Hs.356794	RPS24	10q22-q	549.26	364.14	0.66	4.09E-02
CGGATAACCAGTGGTCC	Hs.524498	PA2G4	12q13	1073.18	709.60	0.66	4.05E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GCGGTGAGGTGTGTGCC	Hs.203910	SGTA	19p13	718.27	466.84	0.65	1.48E-02
CGCCGCCGGCTCAACAA	Hs.182825	RPL35	9q34.1	1250.63	812.31	0.65	1.28E-03
CTAACTAGTTACGCGAC				980.23	634.91	0.65	4.14E-03
AGGAAAGCTGCTGCCAA	Hs.408018	RPL36	19p13.3	1673.15	1083.08	0.65	1.76E-04
ATTAAGAGGGACGGCCG	Hs.380689	LOC5542	2q21.2	870.37	560.21	0.64	6.22E-03
TACTAATAAAAGAAAGT	Hs.472564	C20orf5	20q11.2	1081.63	690.93	0.64	1.96E-03
AAGAAGATAGAAGACAA	Hs.419463	RPL23A	17q11	659.12	420.16	0.64	1.53E-02
GTGACAGAAGAAGACAA	Hs.129673	EIF4A1	17p13	954.88	606.90	0.64	3.31E-03
CAATAAATGTTCTGGTT	Hs.558601	RPL37	5p13	2002.70	1269.82	0.63	1.91E-05
GGCCGCGTTCGCACCAA	Hs.433427	RPS17	15q	692.92	438.83	0.63	1.18E-02
TACCATCAATAAAGTAC	Hs.544577	GAPDH	12p13	12472.54	7880.34	0.63	0.00E+00
ATTGGCTTAAAGTGAAG	Hs.514303	PHB	17q21	845.02	532.20	0.63	4.93E-03
GTGACCTCCTTGGGGGT	Hs.433901	COX8A	11q12-q	1352.04	849.66	0.63	3.54E-04
TCCCCGTGGCTGTGGGG	Hs.498727	DHCR24	1p33-p3	566.17	354.80	0.63	2.02E-02
GGAATGTACGTTATTTC	Hs.429	ATP5G3	2q31.1	2898.43	1811.36	0.62	1.24E-07
TACCAGTGTACTGCTTT	Hs.595053	HSPD1	2q33.1	599.97	373.48	0.62	1.55E-02
GTGACCACGGGTGACGG	Hs.430589	CBLB	3q13.11	3574.45	2222.18	0.62	3.03E-09
CTCCTCACCTGTATTTT	Hs.523185	RPL13A	19q13.3	1115.43	690.93	0.62	8.65E-04
TACATTTTCATATTAGA	Hs.631639	GIPC1	19p13.1	980.23	606.90	0.62	1.78E-03
CCAGAACAGACTGGTGA	Hs.400295	RPL30	8q22	2323.81	1428.54	0.61	1.08E-06
GTGATGGTGTAGCCCTC	Hs.292493	XRCC6	22q13.2	1673.15	1027.06	0.61	3.36E-05
GGTCCAGTGTTCATCTG	Hs.632918	PGAM1	10q25.3	777.42	476.18	0.61	4.56E-03
ACATCATCGATGACATC	Hs.408054	RPL12	9q34	1394.29	849.66	0.61	1.26E-04
CTGTTGATTGCTAAATG	Hs.546261	HNRPA1	12q13.1	507.01	308.12	0.61	2.02E-02
ACATTAAAATAAATTTG				464.76	280.11	0.60	2.41E-02
GGAATCCAATCTGTTGC	Hs.350966	PTTG1	5q35.1	439.41	261.43	0.59	2.51E-02
ACTGGGTCTATGAATAA	Hs.463456	NME1	17q21.3	785.87	466.84	0.59	2.66E-03
AGGGCCCTCAGGAGGGG	Hs.369759	TRPM2	21q22.3	363.36	214.75	0.59	3.94E-02
CTGCACTTACTCCTTTT	Hs.438720	MCM7	7q21.3-	380.26	224.09	0.59	3.43E-02
ATAGTAGCTTCAAACTG	Hs.118400	FSCN1	7p22	523.91	308.12	0.59	1.27E-02
GCACCTCAGCCAGGGGT	Hs.525899	C6orf49	6p21.31	523.91	308.12	0.59	1.27E-02
GGGTTTGAACGGATTTT	Hs.367854	PRMT5	14q11.2	414.06	242.76	0.59	2.59E-02
CTGTTAATAAATACTGG	Hs.431307	MRPL40	22q11.2	574.62	336.13	0.58	8.45E-03
TAATGGTAACTTGGACT	Hs.401903	COX5A	15q25	430.96	252.10	0.58	2.26E-02
GGGATCAAGGAGACCCG	Hs.418233	MRPL24	1q21-q2	481.66	280.11	0.58	1.50E-02
AAAAAGCAGATGACTCG				321.11	186.74	0.58	4.70E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TGTAATCAATAAACGAT	Hs.546261	HNRPA1	12q13.1	338.01	196.07	0.58	4.07E-02
TGTGCTAAATGTGTTCG	Hs.438227	RPL34	4q25	709.82	410.82	0.58	2.92E-03
TAGAAAAATAAAGATGC	Hs.466471	GPI	19q13.1	811.22	466.84	0.58	1.33E-03
ATGGCTGGTATCGATGA	Hs.506997	RPS2	16p13.3	3177.29	1820.69	0.57	1.56E-10
AGAATTTGCAACAGGGG	Hs.459927	PTMA	2q35-q3	456.31	261.43	0.57	1.54E-02
TCTAAGTACGCACGGCC	Hs.586920			1064.73	606.90	0.57	1.91E-04
GCAAAACCAGCTGGTGG	Hs.125113	CCT8	21q22.1	1901.30	1083.08	0.57	5.97E-07
GGCAGGCACAAGAAGGG	Hs.30619	EBP	Xp11.23	1149.23	653.58	0.57	1.00E-04
TAATTCTTCTCTATTGT	Hs.491494	CCT3	1q23	574.62	326.79	0.57	5.96E-03
CCGACGGGCGCTGACCC	Hs.586920			1335.14	756.29	0.57	2.47E-05
GGCAGAGGACCAGGCTG	Hs.463456	NME1	17q21.3	346.46	196.07	0.57	3.15E-02
CCCCAGCCAGTCCCCAC	Hs.546286	RPS3	11q13.3	549.26	308.12	0.56	6.09E-03
GAAGATGTGTGCTCTGG	Hs.112318	TOMM7	7p15.3	667.57	373.48	0.56	2.40E-03
TGCACGTTTTCTGTTTA	Hs.265174	RPL32	3p25-p2	1335.14	737.61	0.55	1.23E-05
GTGAAGCTCGCGAAGGC	Hs.79110	NCL	2q12-qt	388.71	214.75	0.55	1.84E-02
AAGGTCGAGCTGTGCAG	Hs.477028	RPL24	3q12	287.31	158.73	0.55	4.26E-02
GAGTCTGTTCGTGACTC	Hs.309231	C6orf15	6p21.1	371.81	205.41	0.55	2.11E-02
GATGAACACTGGAGGTG	Hs.32826	HDDC2	6q13-q2	811.22	448.17	0.55	6.56E-04
GTGGGCCGCTTGAATGA	Hs.30345	TRAP1	16p13.3	321.11	177.40	0.55	3.21E-02
TAGTTGAAGTCTGTGGA	Hs.131255	UQCRB	8q22	473.21	261.43	0.55	9.27E-03
AGAATAAAATACTGGCG	Hs.461131	CYB5B	16q22.1	591.52	326.79	0.55	3.62E-03
GGGCTGGGGTCCTCCTG	Hs.425125	RPL29	3p21.3-	1537.94	849.66	0.55	2.70E-06
TTTTGTAACTGTAGAGG				456.31	252.10	0.55	1.06E-02
TGCGCTGGCCCCCTGGC	Hs.289019	LTBP3	11q12	270.41	149.39	0.55	4.92E-02
TAAATAATTTCCATATT	Hs.1197	HSPE1	2q33.1	1461.89	802.97	0.55	3.99E-06
GAAACTGTGAGGAGGGG	Hs.544578			549.26	298.78	0.54	4.19E-03
TCTGCAATGAAGAGATT	Hs.482526	TINP1	5q13.3	329.56	177.40	0.54	2.45E-02
CTTGAGCAATAAAGTGG	Hs.524183	FKBP4	12p13.3	887.27	476.18	0.54	2.10E-04
TCCGGCCGCGAAGGTGG	Hs.356440	CCDC72	3p21.31	1630.89	858.99	0.53	2.59E-07
CTTAAATATCAAAGCAG	Hs.437403	PPA1	10q11.1	338.01	177.40	0.52	1.86E-02
GGCATTTTGTTTATGAG				338.01	177.40	0.52	1.86E-02
CTTCGAAACTCTGACAG	Hs.464572	NDUFV2	18p11.3	321.11	168.06	0.52	2.13E-02
ACCCGCCGGGCAGCTTC				1123.88	588.22	0.52	1.64E-05
TGACCCCACAGTGGGGC	Hs.356578	MRPL54	19p13.3	540.81	280.11	0.52	2.46E-03
TAACAAAAATGTATTTT	Hs.522752	PSMD10	Xq22.3	236.61	121.38	0.51	4.28E-02
CCTTGGTTCAAGGGATG	Hs.644639	EEF1A1	6q14.1	219.71	112.04	0.51	4.92E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GTTTGGCGTCAGAAGTC	Hs.469264	RPIA	2p11.2	219.71	112.04	0.51	4.92E-02
TTGTAAACATTTCTTTC	Hs.346868	EBNA1BP	1p35-p3	439.41	224.09	0.51	5.42E-03
AGCACCTCCAGCTGTAC	Hs.515070	EEF2	19pter-	532.36	270.77	0.51	2.13E-03
GAGGGAGTTTCATTAAA	Hs.523463	RPL27A	11p15	1157.68	588.22	0.51	5.75E-06
CTCACTTCTTAAAAGTC	Hs.530412	SERBP1	1p31	312.66	158.73	0.51	1.83E-02
GGGGACGGGAGGAGGGG	Hs.53066	HSPBP1	19q13.4	405.61	205.41	0.51	7.05E-03
TGTGATCAGACTGCTAT	Hs.486360	ATP5L	11q23.3	405.61	205.41	0.51	7.05E-03
ATTCTCCAGTATATTTG	Hs.406300	RPL23	17q	853.47	429.50	0.50	8.21E-05
ACTAACTGTGTAAGTGC	Hs.647652	RBBP4	1p35.1	278.86	140.05	0.50	2.41E-02
ATCAAGAATCCTGCTCC	Hs.14623	IFI30	19p13.1	261.96	130.72	0.50	2.76E-02
CAACTTCAACATATAGA	Hs.482038	PAIP1	5p12	245.06	121.38	0.50	3.17E-02
TGGCAGCTTTTGTTTGA	Hs.463456	NME1	17q21.3	245.06	121.38	0.50	3.17E-02
AAAATAAAGAGCCATAG	Hs.73722	APEX1	14q11.2	473.21	233.42	0.49	2.70E-03
GTTGTAAATAAAGGTTT	Hs.92033	ILKAP	2q37.3	228.16	112.04	0.49	3.63E-02
TGTCTGTGCCGTGCGAC	Hs.109059	MRPL12	17q25	304.21	149.39	0.49	1.56E-02
GCCGTGTCCGCCTGCTA	Hs.408073	RPS6	9p21	895.72	438.83	0.49	3.18E-05
AGAAGTATGACAACAGC	Hs.544577	GAPDH	12p13	287.31	140.05	0.49	1.79E-02
GGAATAAATTAATTTTC	Hs.289271	CYC1	8q24.3	1461.89	709.60	0.49	7.87E-08
GAACCCTGGGATTACCC	Hs.157160	MRPS34	16p13.3	346.46	168.06	0.49	8.92E-03
GGAATATGCAGAATTTC	Hs.224764	CRLS1	20p13-p	270.41	130.72	0.48	2.04E-02
CACAGGCAAAATGTATT	Hs.355983	BZW1	2q33	194.36	93.37	0.48	4.78E-02
GGTGTATATGGAGCCCT	Hs.3439	STOML2	9p13.1	194.36	93.37	0.48	4.78E-02
GACCAGGCCCTCAAGTC	Hs.300772	TPM2	9p13.2-	312.66	149.39	0.48	1.16E-02
GATGAGTCTCGATGTGT	Hs.233952	PSMA7	20q13.3	549.26	261.43	0.48	7.77E-04
TTGGAGATCTCTATTGT	Hs.50098	NDUFA4	7p21.3	278.86	130.72	0.47	1.50E-02
CAATGTGTTATGTAGTG	Hs.534168	NDUFA1	Xq24	743.62	345.47	0.46	6.06E-05
TGGCTGGGAAACTGTTG	Hs.534373	VAMP8	2p12-p1	261.96	121.38	0.46	1.70E-02
TTAGCAATAAATGATGT	Hs.408236	TXNL5	17p13.2	523.91	242.76	0.46	7.39E-04
TCCTGAAATAAATATTG	Hs.503222	RAB6A	11q13.3	202.81	93.37	0.46	3.46E-02
TGACATCATTAAAATAG	Hs.645517	KRR1	12q21.2	202.81	93.37	0.46	3.46E-02
GACCAGCCTTCAGATGG	Hs.462913	MRPL45	17q21.2	304.21	140.05	0.46	9.65E-03
CTCGAGGAGGAGAGGCA	Hs.3254	MRPL23	11p15.5	245.06	112.04	0.46	1.93E-02
AGAAATACCAAGAAATT	Hs.380933	RPL22L1	3q26.2	921.07	420.16	0.46	5.47E-06
GGGGTAAGAAAAGCTGG	Hs.652392	PEBP1	12q24.2	659.12	298.78	0.45	1.09E-04
TGATAATTCAATTTGTA	Hs.500921	USMG5	10q24.3	515.46	233.42	0.45	6.16E-04
CTGTCATTTGTAATATG	Hs.405144	SFRS3	6p21	414.06	186.74	0.45	2.06E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GCATAGGCTGCAACCCA	Hs.12084	TUFM	16p11.2	726.72	326.79	0.45	4.24E-05
GTTTTTCATTGAGTAGA	Hs.624731			312.66	140.05	0.45	7.04E-03
AGGGCTTCCAATGTGCT	Hs.534404	RPL10	Xq28	1935.10	858.99	0.44	1.32E-11
ATGCAGCCATATGGAAG	Hs.467701	ODC1	2p25	211.26	93.37	0.44	2.49E-02
TACCATCAATAAAGTCC				211.26	93.37	0.44	2.49E-02
TTCATATTAAAGTTGTC	Hs.406607	LOC3913	2p23.3	211.26	93.37	0.44	2.49E-02
TTGATGCCATTTCAATA				211.26	93.37	0.44	2.49E-02
GGGGACTGAAGAGTTCG	Hs.146602	UQCRQ	5q31.1	616.87	270.77	0.44	1.14E-04
GGACCACTGAAGAAAGA	Hs.119598	RPL3	22q13	3422.34	1493.90	0.44	0.00E+00
AGAAGTACTGATAGGAC	Hs.558393	RRM1	11p15.5	321.11	140.05	0.44	5.11E-03
GACTCTGGTGCTTCCAA	Hs.370504	RPS15A	16p	236.61	102.71	0.43	1.58E-02
GAAAAGGGTTTTCTTTT	Hs.492314	LAPTM4B	8q22.1	194.36	84.03	0.43	2.81E-02
CCCGTCCGGAACGTCTA	Hs.410817	RPL13	16q24.3	464.76	196.07	0.42	5.20E-04
AAAGGAAAATAAAAATT	Hs.55028	CENPN	16q23.2	177.45	74.70	0.42	3.17E-02
TTGTAATCGTGCAAATA	Hs.446427	OAZ1	19p13.3	380.26	158.73	0.42	1.54E-03
AAACTGATTGTAAAGCT	Hs.58488	CTNNAL1	9q31.2	202.81	84.03	0.41	1.99E-02
CAGATTTTGGTGCTTTC	Hs.438227	RPL34	4q25	430.96	177.40	0.41	6.44E-04
AGGCTACGGAAAACAGG				160.55	65.36	0.41	3.56E-02
GTGCTCTTGATATATAA	Hs.545512	LOC3922	8p22	160.55	65.36	0.41	3.56E-02
TCGGTGCAGGTGCCTGG	Hs.31714	WDR8	1p36.3	160.55	65.36	0.41	3.56E-02
TATAATAAATACATCTC	Hs.528222	NDUFS4	5q11.1	253.51	102.71	0.41	8.01E-03
CAAATAAAAGTTGTTTG	Hs.185055	MALL	2q13	304.21	121.38	0.40	3.27E-03
TTACCATATCAAGCTGA	Hs.300141	RPL39	Xq22-q2	912.62	364.14	0.40	3.49E-07
ATAGACATAAAATTGGT	Hs.555866	C1QBP	17p13.3	946.43	373.48	0.39	1.66E-07
ATAATTCTTTGTATATA	Hs.156367	RPS29	14q	2163.26	849.66	0.39	2.00E-15
TGTCTTTGCTCTTTCTG	Hs.9589	UBQLN1	9q22|9q	143.65	56.02	0.39	3.97E-02
AAGGTAATGCTAGTCTT	Hs.27222	NOLA2	5q35.3	169.00	65.36	0.39	2.47E-02
GAAGACGAATTTGGGTT	Hs.189716	NDUFAB1	16p12.1	169.00	65.36	0.39	2.47E-02
CGGCCCAACGCCAAGAA	Hs.20521	PRMT1	19q13.3	194.36	74.70	0.38	1.55E-02
GACAGTGACGCAAGGAC	Hs.477273	GRRP1	1p36.11	194.36	74.70	0.38	1.55E-02
GTAAAAGTTCTAAATCT	Hs.523238	NOLC1	10q24.3	295.76	112.04	0.38	2.55E-03
TGAAATAAAACTCAGTA	Hs.557550	NPM1	5q35	1613.99	597.56	0.37	7.50E-13
AAACCCGAAGACAAGAA	Hs.3100	KARS	16q23-q	126.75	46.68	0.37	4.38E-02
AAGCCTAAAATACCAAG	Hs.79136	SLC39A6	18q12.2	126.75	46.68	0.37	4.38E-02
CAAAAGGAATCGTGCTG	Hs.527193	RPS23	5q14.2	126.75	46.68	0.37	4.38E-02
CAGGTCATACGTTACTC	Hs.8715	C4orf14	4q12	126.75	46.68	0.37	4.38E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CCTGCCAAAGACGTGTC	Hs.514303	PHB	17q21	126.75	46.68	0.37	4.38E-02
GATGCATTAGATAAAGT	Hs.283734	MRPL47	3q26.33	126.75	46.68	0.37	4.38E-02
GAGTATAAATGATTATT	Hs.521800	CAPN1	11q13	152.10	56.02	0.37	2.72E-02
GGGAGCCCCTGGTGGGA	Hs.20961	MGC1125	7p22.3	152.10	56.02	0.37	2.72E-02
GGGCCATATATCTTGGG				177.45	65.36	0.37	1.71E-02
GGGGCAGGGCCCTCCCA				312.66	112.04	0.36	1.25E-03
CTGAGTTAGGTAGGTGG	Hs.522826	SLC10A3	Xq28	287.31	102.71	0.36	1.95E-03
TATTTTTGTTACAAATC	Hs.567352	TXNRD1	12q23-q	397.16	140.05	0.35	2.36E-04
CAGCACAGACCAATTAG	Hs.355307	CD27	12p13	160.55	56.02	0.35	1.86E-02
GCCAGGAGCTAAGTGCC	Hs.519035	LAD1	1q25.1-	135.20	46.68	0.35	2.96E-02
TTTCCTATCCAAGAGAC	Hs.177530	ATP5E	20q13.3	135.20	46.68	0.35	2.96E-02
TAGATTCAACTAGAATC	Hs.427232	VKORC1L	7q11.21	245.06	84.03	0.34	3.27E-03
GGGTTTTTATTATTTTT	Hs.473583	YBX1	1p34	354.91	121.38	0.34	3.92E-04
GTGCTGGTCAGAGAGCC	Hs.631966	LOC6439	5q14.1	354.91	121.38	0.34	3.92E-04
AGAATATCAGTGATACT	Hs.531614	BTBD14B	19p13.1	109.85	37.35	0.34	4.77E-02
CATTTAAGTTTAAGTGA	Hs.491695	UBE2V2	8q11.21	109.85	37.35	0.34	4.77E-02
CCAGGGAGATCTTTGAC	Hs.517145	ENO1	1p36.3-	109.85	37.35	0.34	4.77E-02
GAAAAAAGGTTGATGGA				109.85	37.35	0.34	4.77E-02
GAGTTAGTGAGGCGCTC				109.85	37.35	0.34	4.77E-02
GATGCGCTTGTGAATGT	Hs.469473	RPL31	2q11.2	109.85	37.35	0.34	4.77E-02
ACAGATATCAACAGGGC	Hs.220594	CCDC58	3q21.1	278.86	93.37	0.33	1.45E-03
TATTTTTCAAGCTAACC	Hs.6799	FAM98B	15q14	169.00	56.02	0.33	1.26E-02
CCCATCCGAAAGGATGA	Hs.652201	RPL26	17p13	794.32	261.43	0.33	5.54E-08
GATGAGAAGAAGAAGGA	Hs.512465	SURF4	9q34.2	143.65	46.68	0.32	1.99E-02
GCTCAGCTGGAGGCCTG	Hs.333388	EEF1D	8q24.3	143.65	46.68	0.32	1.99E-02
TGCTGTAAAGGAATTCA	Hs.643493	FASTKD3	5p15.3-	143.65	46.68	0.32	1.99E-02
TGTGATCACAAAGACTG	Hs.9661	PSMB10	16q22.1	261.96	84.03	0.32	1.54E-03
GCATTTAAATAAAAGAT	Hs.421608	EEF1B2	2q33-q3	2467.47	784.30	0.32	0.00E+00
CACCAGCATTGGAGATA	Hs.255973	EID1	15q21.1	118.30	37.35	0.32	3.17E-02
CTCCCAGGTCATTGTCA	Hs.344400	MPHOSPH	16q23.3	118.30	37.35	0.32	3.17E-02
GGCTGCCGAGTCCTGCC	Hs.490991	PINX1	8p23	118.30	37.35	0.32	3.17E-02
GTTTGCAAGTGAACAGA	Hs.151787	EFTUD2	17q21.3	118.30	37.35	0.32	3.17E-02
TAACCATTTTAACTCTC	Hs.591332	FUCA2	6q24	118.30	37.35	0.32	3.17E-02
TCTCAATTCTTTGTATA	Hs.597524	CDC42	1p36.1	118.30	37.35	0.32	3.17E-02
TGTACTACTTAAGTTTA	Hs.518774	PAICS	4q12	118.30	37.35	0.32	3.17E-02
CCCGGGGCCTCCCTTCC	Hs.90691	NPM3	10q24.3	177.45	56.02	0.32	8.51E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GTGTCTCATCGTTCGGG	Hs.517145	ENO1	1p36.3-	177.45	56.02	0.32	8.51E-03
TTTGTAGATGGGGCAGA	Hs.184233	HSPA9	5q31.1	177.45	56.02	0.32	8.51E-03
TTAAACTTTGATTTGTC	Hs.530412	SERBP1	1p31	152.10	46.68	0.31	1.33E-02
CCTTCCAAATTGTGGGT	Hs.520967	MDH2	7p12.3-	456.31	140.05	0.31	1.81E-05
TTCCATCAATAAAGTAC				219.71	65.36	0.30	2.49E-03
AGCTCTCCCTGCCACAT	Hs.374588	RPL17	18q21	346.46	102.71	0.30	1.41E-04
ATCGTTGTAATTGTTGA	Hs.5683	DHX15	4p15.3	126.75	37.35	0.29	2.09E-02
CAAGTTTGCTGAGCTGA	Hs.644639	EEF1A1	6q14.1	126.75	37.35	0.29	2.09E-02
GAGAAATATATTAAATA	Hs.503597	HSPC148	11q21	126.75	37.35	0.29	2.09E-02
ACAACTACTTTAAGGGG	Hs.525600	HSP90AA	14q32.3	194.36	56.02	0.29	3.83E-03
GCCAAGAATCATTATGA	Hs.503047	MRPL21	11q13.2	194.36	56.02	0.29	3.83E-03
GCAAGGTCAGTGTGGAG	Hs.558447	EMG1	12p13	228.16	65.36	0.29	1.67E-03
TGCCCCCGGGTTCCTTC	Hs.174134	CCDC44	17q23.3	135.20	37.35	0.28	1.38E-02
TGGACACAAGCATAAGT	Hs.506215	RARS	5q35.1	135.20	37.35	0.28	1.38E-02
ACAGTGGGGATTTTTTT	Hs.50425	PTGES3	12q13.3	101.40	28.01	0.28	3.29E-02
CAGGATCCAGAAGTTAT	Hs.652409	ST13	22q13.2	101.40	28.01	0.28	3.29E-02
GCGACTTTTTCTGGCCC	Hs.367842	MKI67IP	2q14.3	101.40	28.01	0.28	3.29E-02
GGAGTGTGCGTGGACTG	Hs.4944	C9orf58	9q34.13	101.40	28.01	0.28	3.29E-02
GTTGACTTACAAGTGTC	Hs.355177	C5orf15	5q31.1	101.40	28.01	0.28	3.29E-02
TGATGGGCATTGAGCCA	Hs.503716	DCUN1D5	11q22.3	101.40	28.01	0.28	3.29E-02
TGCAGGCCTGGAAGAGC	Hs.497599	WARS	14q32.3	101.40	28.01	0.28	3.29E-02
ATTGTCAGGGAGGTGCC	Hs.652101	SNX5	20p11	177.45	46.68	0.26	3.88E-03
TATATTTTTGTCTCCCC	Hs.630800			143.65	37.35	0.26	9.01E-03
AAAAATAAACCAGGTCC	Hs.523238	NOLC1	10q24.3	109.85	28.01	0.25	2.12E-02
AGCCCCCTGAGGCCCCA	Hs.557550	NPM1	5q35	109.85	28.01	0.25	2.12E-02
TGGGAGAAGTGAGTTAG	Hs.410977	SIDT2	11q23.3	109.85	28.01	0.25	2.12E-02
TTACTAAATGGTGTTAC	Hs.651169	CANX	5q35	109.85	28.01	0.25	2.12E-02
AAAAATAAAGGTTCCAT	Hs.298280	ATP5A1	18q12-q	185.91	46.68	0.25	2.56E-03
TTAAGAAATGCATTAAG	Hs.278277	OXCT1	5p13.1	261.96	65.36	0.25	3.30E-04
TACCATCAATAAAGTTC				152.10	37.35	0.25	5.88E-03
CTCACCGCCCTGCTTCA	Hs.405662	CRABP2	1q21.3	118.30	28.01	0.24	1.37E-02
CAGCCGTGATCAAGGGA	Hs.532790	NMT1	17q21.3	160.55	37.35	0.23	3.82E-03
AGGTGCAGAGGCCCACC	Hs.517543	PES1	22q12.1	202.81	46.68	0.23	1.10E-03
TTGGGGTTTGTGGGTCG				202.81	46.68	0.23	1.10E-03
GAAAAATGGTTGATGGG				169.00	37.35	0.22	2.48E-03
TAGGACCCTGCAGGGGG	Hs.76873	HYAL2	3p21.3	126.75	28.01	0.22	8.79E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TTGTACAACAGATACCT	Hs.444969	C2orf4	2p22-p2	126.75	28.01	0.22	8.79E-03
CCTATGTAAGGAAAGTG	Hs.380118	RBMX	Xq26.3	84.50	18.67	0.22	3.24E-02
GAAACCCTCACTCCTGG	Hs.75859	MRPL49	11q13	84.50	18.67	0.22	3.24E-02
GCATCCTGCTCGTGTAG	Hs.119598	RPL3	22q13	84.50	18.67	0.22	3.24E-02
GCCAAACTTGGAGGTGG	Hs.610373			84.50	18.67	0.22	3.24E-02
GCCTCAGTTCCGAAAAC	Hs.380689	LOC5542	2q21.2	84.50	18.67	0.22	3.24E-02
TAAGAAATATCGTCAGT	Hs.330663	C12orf4	12q23.2	84.50	18.67	0.22	3.24E-02
TATCCTCAATGCCCGGG	Hs.446149	LDHB	12p12.2	84.50	18.67	0.22	3.24E-02
TGATTAAAACAAGTTGC	Hs.520819	INSIG1	7q36	84.50	18.67	0.22	3.24E-02
AATGTCATTGGAGAACC	Hs.406510	ATP5B	12q13.1	219.71	46.68	0.21	4.67E-04
TGCTTTGGGATGAGTGT	Hs.423968	FIS1	7q22.1	177.45	37.35	0.21	1.61E-03
GGGGAGGAACAGCTCCG	Hs.631558	TNNT1	19q13.4	135.20	28.01	0.21	5.63E-03
AAAATACTAGCTTATTT	Hs.40499	DKK1	10q11.2	278.86	56.02	0.20	5.88E-05
ATTAACAAAGCAACCTT	Hs.125898	GNAS	20q13.3	92.95	18.67	0.20	2.04E-02
GCCGTGAGCAGAACTGT				92.95	18.67	0.20	2.04E-02
GCCTCCTCTTTGCTGAT				92.95	18.67	0.20	2.04E-02
TGGCAAGATGAGAGTAA				92.95	18.67	0.20	2.04E-02
TGGCGTTGAGGAAAGCA				92.95	18.67	0.20	2.04E-02
TGTACAAAATCTACTTC	Hs.556017	C1orf13	1q42.2	92.95	18.67	0.20	2.04E-02
TCAAAAAAAAGAAAGGT				143.65	28.01	0.19	3.60E-03
CACAAGATGATTAATAC	Hs.644628	RPL4	15q22	101.40	18.67	0.18	1.28E-02
GAAAAATGGCTGATGGA				101.40	18.67	0.18	1.28E-02
TGGCCCCAGGTGCCACC	Hs.110675	TOMM40	19q13	261.96	46.68	0.18	5.35E-05
TCCATCAATAAAGTACC				211.26	37.35	0.18	2.77E-04
TCTGTATTTGGAGATGG	Hs.191582	LOC4404	18q23	109.85	18.67	0.17	8.04E-03
AAATGCCCTCCTAAATG	Hs.151777	EIF2S1	14q23.3	59.15	9.34	0.16	4.74E-02
AAGGAATCGGGAGTGGA	Hs.612604			59.15	9.34	0.16	4.74E-02
AAGGACAAGATAAAAGA	Hs.250758	PSMC3	11p12-p	59.15	9.34	0.16	4.74E-02
ACCTGCAAATTGGAGAG	Hs.531106	RBM25	14q24.3	59.15	9.34	0.16	4.74E-02
ATGCTTTCTTTTGTTTC				59.15	9.34	0.16	4.74E-02
CATTCCAGAGGCCCCCG	Hs.490874	MTX1	1q21	59.15	9.34	0.16	4.74E-02
CGAAGGCTGTATATATT	Hs.79334	NFIL3	9q22	59.15	9.34	0.16	4.74E-02
CGCACTGCATTCTCCGG				59.15	9.34	0.16	4.74E-02
GACAAGGAAGGCAATGT	Hs.523299	SFXN4	10q26.1	118.30	18.67	0.16	5.04E-03
GACATTTAACATCCTTC	Hs.274267	CXorf57	Xq22.3	59.15	9.34	0.16	4.74E-02
GACTTTTAAGGATACCG				59.15	9.34	0.16	4.74E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GAGAGGGAAGCGGGAGG				59.15	9.34	0.16	4.74E-02
GATGAAAACTAAAGGCT				59.15	9.34	0.16	4.74E-02
GCCCAGGGCCGCTGGGA	Hs.10326	COPE	19p13.1	59.15	9.34	0.16	4.74E-02
GCTTAACCTGGTGATAA	Hs.500409	GLUD1	10q23.3	59.15	9.34	0.16	4.74E-02
GGAACAGGGGAGCAGGG	Hs.102336	PRR5	22q13	59.15	9.34	0.16	4.74E-02
GGAAGTTCAAACAAATA	Hs.237536	NT5C3L	17q21.2	59.15	9.34	0.16	4.74E-02
GGCCGCGTTCGCCCCAA				59.15	9.34	0.16	4.74E-02
GGGGTAAATGCCAATCC	Hs.85962	CTF8	16q22.1	59.15	9.34	0.16	4.74E-02
GGTGCAGGGAGGGGTGG				59.15	9.34	0.16	4.74E-02
TAACAGAAAGGGCAACA	Hs.388927	YY1	14q	59.15	9.34	0.16	4.74E-02
TAACCATCACTGGAATC	Hs.161429	GSTCD	4q24	59.15	9.34	0.16	4.74E-02
TATTCTCAATAGGCTGT	Hs.465224	NARS	18q21.2	59.15	9.34	0.16	4.74E-02
TGACCAAAACAATAGAA	Hs.191219	CPNE3	8q21.3	59.15	9.34	0.16	4.74E-02
TGAGGGAATAAAACCTG				59.15	9.34	0.16	4.74E-02
TGATGTGATCAGATGCT	Hs.12272	BECN1	17q21	59.15	9.34	0.16	4.74E-02
TGGGGAAAAGTAGTCTC	Hs.645489	TRA2A	7p15.3	118.30	18.67	0.16	5.04E-03
CATCTAGAGGGCTCATA				126.75	18.67	0.15	3.17E-03
TAAAGGTTTTTTTTTC				126.75	18.67	0.15	3.17E-03
AATAAAGCAATAAAATT	Hs.518326	SERP1	3q25.1	67.60	9.34	0.14	2.88E-02
AGCCTCCTCAACAAGCC	Hs.515472	SNRPD2	19q13.2	67.60	9.34	0.14	2.88E-02
CAAATGATGCCTCTGCC				67.60	9.34	0.14	2.88E-02
CTTCTAAATATAATAGC	Hs.527909	EDG7	1p22.3-	67.60	9.34	0.14	2.88E-02
GGACTTTGAGAAGAGGG	Hs.444947	TRIB1	8q24.13	202.81	28.01	0.14	1.52E-04
GTAAAGACTGCTTTATT				67.60	9.34	0.14	2.88E-02
GTTCATAGTATTCACTG	Hs.368783	NAT5	20p11.2	67.60	9.34	0.14	2.88E-02
TAAATGAATAAAGACAT	Hs.378808	EIF2A	3q25.1	67.60	9.34	0.14	2.88E-02
TAGTTTCAACTTGTAAC	Hs.490551	UBAP2L	1q21.3	67.60	9.34	0.14	2.88E-02
TCCCATCAATAAAGTCC				67.60	9.34	0.14	2.88E-02
TGGAAAGAGCCTTACTA	Hs.9043	NGDN	14q11.2	67.60	9.34	0.14	2.88E-02
TGGTTTTGGCAGCAATA	Hs.593405			67.60	9.34	0.14	2.88E-02
ATGGCGATCTATCTTCA				143.65	18.67	0.13	1.25E-03
GAGTAGAGAAAAGAGAC	Hs.632416	LOC6468	1p22.1	143.65	18.67	0.13	1.25E-03
CGACCCCACGCCACCCC	Hs.110675	TOMM40	19q13	219.71	28.01	0.13	6.13E-05
ATTTTGCTTGGGTAATC	Hs.563204			76.05	9.34	0.12	1.75E-02
CAGTGTATATATTGAGA	Hs.172755	BRP44L	6q27	76.05	9.34	0.12	1.75E-02
GCCGTTCTTAGTTGGGG				76.05	9.34	0.12	1.75E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GCTATGAAGAAGGACCC	Hs.591731	HNRPAB	5q35.3	76.05	9.34	0.12	1.75E-02
GCTGTCATCAGGAAAAT	Hs.356654	PSMC1	14q32.1	76.05	9.34	0.12	1.75E-02
TACCATCAATAAATACC				76.05	9.34	0.12	1.75E-02
TAGACCCCTTGAAGAGG	Hs.544577	GAPDH	12p13	228.16	28.01	0.12	3.89E-05
TATAAATTAAAATGTTT	Hs.369614	COPS2	15q21.2	76.05	9.34	0.12	1.75E-02
TCTAAAGAGTGAGCTTC	Hs.433422	PIGL	17p12-p	76.05	9.34	0.12	1.75E-02
TGAAAGTGATTTCTTCC	Hs.98594	ARHGEF1	8p23	76.05	9.34	0.12	1.75E-02
TGAGGGAATAACCCTGG				76.05	9.34	0.12	1.75E-02
AAATGACTATAAATGGT	Hs.43071	HDGF2	19p13.3	42.25	4.67	0.11	3.34E-02
AAGGGGCTGGGAACCGA	Hs.511605	ANXA2	15q21-q	42.25	4.67	0.11	3.34E-02
AATTGAATAATTGTTTT	Hs.492612	THRAP6	8q24.11	42.25	4.67	0.11	3.34E-02
AATTGCATTAAGACTCT	Hs.166463	HNRPU	1q44	42.25	4.67	0.11	3.34E-02
ACATACATACGAAAACC	Hs.444321	PTER	10p12	42.25	4.67	0.11	3.34E-02
ACATTTCATTTGGTAAC				42.25	4.67	0.11	3.34E-02
ACTGGAGTTTGCTTTGT	Hs.253726	PAPOLA	14q32.3	42.25	4.67	0.11	3.34E-02
AGCAAATATGTCAAGGG	Hs.487325	PRKACB	1p36.1	169.00	18.67	0.11	3.08E-04
AGGTCAGAGGGAGCTAG	Hs.652477			42.25	4.67	0.11	3.34E-02
ATAATCTCCACTTGGTC				42.25	4.67	0.11	3.34E-02
ATGCTGATCCACTTGGG				42.25	4.67	0.11	3.34E-02
ATGGTTCGCGAGCTTGG	Hs.398178	C4orf9	4p16.3	42.25	4.67	0.11	3.34E-02
ATTTCAAGATGTTATAT	Hs.155097	CA2	8q22	42.25	4.67	0.11	3.34E-02
CACAAACGGCAGTTTTG				42.25	4.67	0.11	3.34E-02
CAGAAGCTGAGCACACC	Hs.474596	LIMK2	22q12.2	42.25	4.67	0.11	3.34E-02
CAGCCGAGGCCCGGGCC	Hs.150319	CRB3	19p13.3	42.25	4.67	0.11	3.34E-02
CATTTAGATTTATATTT	Hs.271695	NOB1	16q22.3	169.00	18.67	0.11	3.08E-04
CCACTCTCAGAGGTTGG				42.25	4.67	0.11	3.34E-02
CCCCCGCGGAGCCGCAG	Hs.466507	LSR	19q13.1	42.25	4.67	0.11	3.34E-02
CCGAACTTTCTTGTGAC	Hs.469459	PDCL3	2q11.2	42.25	4.67	0.11	3.34E-02
CTAGCCTCACGAAAACT				42.25	4.67	0.11	3.34E-02
CTCACACAGGCTATGGG	Hs.530314	SSNA1	9q34.3	42.25	4.67	0.11	3.34E-02
CTGCTATATATAATCAG	Hs.593076			42.25	4.67	0.11	3.34E-02
CTTACGTGATTTTTATT	Hs.652389	LOC5722	12q13.1	42.25	4.67	0.11	3.34E-02
GAAATAACATAGGGTTC				42.25	4.67	0.11	3.34E-02
GAAGGAGTAAGAGCTGG	Hs.530538	TATDN3	1q32.3	42.25	4.67	0.11	3.34E-02
GAGAGCAAGGAGAAGGA	Hs.593928	FDFT1	8p23.1-	42.25	4.67	0.11	3.34E-02
GAGATCAAGATTATATC	Hs.652388	TEGT	12q12-q	42.25	4.67	0.11	3.34E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GAGCAGTTCCAGAAGGA				42.25	4.67	0.11	3.34E-02
GAGCCTGACCTTAGTGC	Hs.515146	ECSIT	19p13.2	42.25	4.67	0.11	3.34E-02
GATCAGCAGCAAGTCCA				42.25	4.67	0.11	3.34E-02
GCCACAGAAATGGAGGG	Hs.435063	ARHGAP2	10q11.2	42.25	4.67	0.11	3.34E-02
GCCTTCCGTGTCCCCAC	Hs.544577	GAPDH	12p13	42.25	4.67	0.11	3.34E-02
GCGATGGGGGAGGGCGA	Hs.592088	ZNF205	16p13.3	42.25	4.67	0.11	3.34E-02
GCTGACACATTCCTGGA	Hs.534770	PKM2	15q22	42.25	4.67	0.11	3.34E-02
GCTGTGGTCCTAGTTAC	Hs.585728	NIP7	16q22.1	42.25	4.67	0.11	3.34E-02
GCTTCCATCTGAGCCAC	Hs.467279	LENG4	19q13.4	42.25	4.67	0.11	3.34E-02
GGAAACTCTGTTTATGC				42.25	4.67	0.11	3.34E-02
GGAAGTTAAGTATTTAG	Hs.10862	AK3L1	1p31.3	42.25	4.67	0.11	3.34E-02
GGCAAGAAAAAAATCGC				42.25	4.67	0.11	3.34E-02
GGCAGGCCCAAGAAGGG				42.25	4.67	0.11	3.34E-02
GGGCTAACTCCAGAATC	Hs.470601	LOC7285	10q11.2	42.25	4.67	0.11	3.34E-02
GGGGGTAACTACGGGGA	Hs.513522	FUS	16p11.2	42.25	4.67	0.11	3.34E-02
GGTACCACTGAAAGAAA				42.25	4.67	0.11	3.34E-02
GGTGTCCCTTGAGTCTG	Hs.252229	MAFG	17q25.3	42.25	4.67	0.11	3.34E-02
GTCAACTGCTTCAGCTT	Hs.293563	C1orf10	1p34.3	42.25	4.67	0.11	3.34E-02
GTGACCACGGGTGAACG				42.25	4.67	0.11	3.34E-02
GTGGCCCGGGTGTGGTC	Hs.356626	P117	19p13.3	42.25	4.67	0.11	3.34E-02
GTGGCTGCTGTTTGTCA	Hs.631971	MRPS36	5q13.2	42.25	4.67	0.11	3.34E-02
TAACAGAGCCAGGAACC	Hs.387567	ACLY	17q12-q	42.25	4.67	0.11	3.34E-02
TACCCTGGCATTGCCGA	Hs.520640	ACTB	7p15-p1	42.25	4.67	0.11	3.34E-02
TACTCGAATATACAGTT	Hs.652301	SOCS4	14q22.3	42.25	4.67	0.11	3.34E-02
TCAAGATTCAGCTAGTG	Hs.652416	PGK1	Xq13	42.25	4.67	0.11	3.34E-02
TCGGGGAGGGGAGGGGA				42.25	4.67	0.11	3.34E-02
TGAGTAAGATAAATGTA	Hs.190520	LSM6	4q31.22	42.25	4.67	0.11	3.34E-02
TGGTGAAGAACTCAAGG	Hs.194121	RCL1	9p24.1-	42.25	4.67	0.11	3.34E-02
TGTGGGTCTTCAGATAC				42.25	4.67	0.11	3.34E-02
TTCAAATGTTTGAAATC	Hs.184492	ELAVL1	19p13.2	42.25	4.67	0.11	3.34E-02
TTGATTTAATAAAGTCA				42.25	4.67	0.11	3.34E-02
TTGGAAAATAAATATGA	Hs.479386	LOC3892	4p15.2	42.25	4.67	0.11	3.34E-02
TTGGGGAAACAAGAAAA	Hs.488143	BLVRA	7p14-ce	42.25	4.67	0.11	3.34E-02
TTGTGTGTACCCAGATA				42.25	4.67	0.11	3.34E-02
TTTATAACTATTCAGTT	Hs.75066	TSN	2q21.1	42.25	4.67	0.11	3.34E-02
TTTGCGGCAGAGGTCTC				42.25	4.67	0.11	3.34E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TTTGTTGTATGTAAAGA	Hs.584957	TOR3A	1q25.2	42.25	4.67	0.11	3.34E-02
TTTTCCTTTTGTTTTTG				42.25	4.67	0.11	3.34E-02
ATGTTAACTTTGTGTGG	Hs.525600	HSP90AA	14q32.3	84.50	9.34	0.11	1.07E-02
CAGACGCTCCGATTTCT	Hs.416207	MRPS33	7q32-q3	84.50	9.34	0.11	1.07E-02
CCTTCTGGTGGACATTT	Hs.189075	TWF1	12q12	84.50	9.34	0.11	1.07E-02
CTGCCTCCTTACATCCT	Hs.472847	DBNDD2	20q13.1	84.50	9.34	0.11	1.07E-02
GAGGAGGAGGTTGATCC				84.50	9.34	0.11	1.07E-02
GCGGCTTTCCGCAGTGT	Hs.567405	SCO2	22q13.3	84.50	9.34	0.11	1.07E-02
GGGGGAGAAGTGATATG	Hs.174249	CDX2	13q12.3	84.50	9.34	0.11	1.07E-02
TATACCAATCACAATGG	Hs.379858	DDAH1	1p22	84.50	9.34	0.11	1.07E-02
TCAGAAGTTTTGTTTAA	Hs.377155	MTDH	8q22.1	84.50	9.34	0.11	1.07E-02
TTGAGAGATGATTCTTT	Hs.530823	COPS7A	12p13.3	84.50	9.34	0.11	1.07E-02
TCCCTTCGCGGGAAGGC	Hs.546285	RPLP0	12q24.2	177.45	18.67	0.11	1.93E-04
TTACATTTTAGGTACCT				177.45	18.67	0.11	1.93E-04
CAGCAATTCCAGAACCC	Hs.467701	ODC1	2p25	92.95	9.34	0.10	6.58E-03
GAAGTTTTTTTAAATAA	Hs.548868	THOC3	5q35.2	92.95	9.34	0.10	6.58E-03
GAATGAGATTACAGGGC	Hs.651167	C4orf20	4q35.1	92.95	9.34	0.10	6.58E-03
GAGAAAAAGTGAACTGG	Hs.385865	WDR53	3q29	92.95	9.34	0.10	6.58E-03
TAAGATCGTCGAGATGT	Hs.534314	EIF5A	17p13-p	92.95	9.34	0.10	6.58E-03
TACAATTTGCCACTGGG				92.95	9.34	0.10	6.58E-03
AAACAATACACCTTGGA	Hs.525796	C15orf2	15q15.1	50.70	4.67	0.09	1.98E-02
ACAGATTAAATAAATGG	Hs.225968	MYH15	3q13.13	50.70	4.67	0.09	1.98E-02
ACATTTTATTAATCCTC	Hs.529778	CENPK	5p15.2-	50.70	4.67	0.09	1.98E-02
AGATGTGTGGGTGGTTG	Hs.534639	HADHB	2p23	50.70	4.67	0.09	1.98E-02
AGGCCTTGGTTCAGAGC	Hs.109672	ST6GALN	9q34.11	50.70	4.67	0.09	1.98E-02
ATGGCCGGTATCGATGA				50.70	4.67	0.09	1.98E-02
CCGTAGGTGGGCACAGG				50.70	4.67	0.09	1.98E-02
CGCTTTCTTCTGAGGGT	Hs.381219	RPL15	3p24.2	50.70	4.67	0.09	1.98E-02
GCAGATCGGGAATACCT	Hs.646417	LOC4017	12q21.3	50.70	4.67	0.09	1.98E-02
GGCTGAAGTTGGCATTT				50.70	4.67	0.09	1.98E-02
GTGACCACGGGCGACGG				50.70	4.67	0.09	1.98E-02
TAAGATTAGAAGTTCTC	Hs.118631	TIMELES	12q12-q	50.70	4.67	0.09	1.98E-02
TACCAGATCAAGTATGA	Hs.533782	KRT8	12q13	50.70	4.67	0.09	1.98E-02
TCAACTAACAATAAAGG				50.70	4.67	0.09	1.98E-02
TCCCAGCCCACATAGAT	Hs.301540	SPR	2p14-p1	50.70	4.67	0.09	1.98E-02
TCCGTGTGTCATCTTGG	Hs.652406	SMCR7L	22q13	50.70	4.67	0.09	1.98E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
TGGAAGCTTTCCTTTCG	Hs.5308	UBA52	19p13.1	50.70	4.67	0.09	1.98E-02
TTGGGGTTTCCTTTTCC				50.70	4.67	0.09	1.98E-02
TTGGTGAAGGAAGAATG				50.70	4.67	0.09	1.98E-02
TTTAAAAGGTCTTAGCC				50.70	4.67	0.09	1.98E-02
ATAGAGGCAATGCATTA	Hs.326387	MORF4L2	Xq22	101.40	9.34	0.09	4.04E-03
TACCATCAATAAAAGTA				101.40	9.34	0.09	4.04E-03
TGCGGGGGGTGGGGTCC	Hs.333579	HSPC152	11q13.1	101.40	9.34	0.09	4.04E-03
TTCAAAGTCTTATTGAC	Hs.368084	LRPPRC	2p21	101.40	9.34	0.09	4.04E-03
GGTGAGACACTCCAGTA	Hs.350927	SLC25A6	Xp22.32	109.85	9.34	0.08	2.49E-03
AAAATAAAGCTCACTGT	Hs.293818	NEIL2	8p23.1	59.15	4.67	0.08	1.18E-02
AGGCGAGATCAATCCCT	Hs.233952	PSMA7	20q13.3	118.30	9.34	0.08	1.54E-03
ATCCAAAGGAACCAGAG	Hs.516539	HNRPA3	2q31.2	59.15	4.67	0.08	1.18E-02
CAGAAATATATATGTGT	Hs.262823	IARS2	1q41	59.15	4.67	0.08	1.18E-02
CTCATCGCAATGGTGGA	Hs.299002	FBL	19q13.1	59.15	4.67	0.08	1.18E-02
CTTTACGGGCTTGTAGG				59.15	4.67	0.08	1.18E-02
GAATAATAGGGATTTTA	Hs.434053	FAM3C	7q31	59.15	4.67	0.08	1.18E-02
GGAAATGGGGTGGATTA	Hs.405925	PSRC1	1p13.3	59.15	4.67	0.08	1.18E-02
GGGAATAAACCAGCATT	Hs.252457	MVD	16q24.3	118.30	9.34	0.08	1.54E-03
GGTACTCGATGTGTAAT	Hs.332422	ASPH	8q12.1	59.15	4.67	0.08	1.18E-02
TAATACTTTTGCAAGTG	Hs.397729	HMGCS1	5p14-p1	118.30	9.34	0.08	1.54E-03
TATAAGGTGGCTCCATA	Hs.576875	DDX21	10q21	59.15	4.67	0.08	1.18E-02
TCTGTAACACCTGTCAA	Hs.512973	PTPLAD1	15q22.2	59.15	4.67	0.08	1.18E-02
TGACTGGCCATTTGTGC				59.15	4.67	0.08	1.18E-02
TTTTCTGAGCAGGCTGC	Hs.306791	POLD2	7p13	59.15	4.67	0.08	1.18E-02
ATTGGGCAGTTTGGTGT	Hs.509736	HSP90AB	6p12	67.60	4.67	0.07	7.13E-03
GAGGAGATGGCCAGGGC	Hs.546454	DKFZp43	22q11.2	67.60	4.67	0.07	7.13E-03
GCGAAGAGGATGAGGAA	Hs.466975			67.60	4.67	0.07	7.13E-03
GTGAAAATACGTCTTCC	Hs.153088	TAF1A	1q42	67.60	4.67	0.07	7.13E-03
TTTATTCCTCAGGAATA	Hs.466662	NBL1	1p36.13	67.60	4.67	0.07	7.13E-03
CATCCTTGGGCAGGTTG	Hs.652409	ST13	22q13.2	76.05	4.67	0.06	4.32E-03
CATCTAGAGGGCTCAAT				76.05	4.67	0.06	4.32E-03
GCAGCTAATTTTGTAAC	Hs.202011	CCDC47	17q23.3	76.05	4.67	0.06	4.32E-03
GGTGGTAAAGTACCACC	Hs.652409	ST13	22q13.2	76.05	4.67	0.06	4.32E-03
TGCCATCTGTACATAAA				76.05	4.67	0.06	4.32E-03
TGGGGGATATCCACGGC				76.05	4.67	0.06	4.32E-03
AAGCTGTTGTGTGAGGT	Hs.202672	DNMT1	19p13.2	84.50	4.67	0.06	2.63E-03
CACTCTAAGATGAGTTC	Hs.156506	MGC1301	5q31.1	84.50	4.67	0.06	2.63E-03
ACTACAAATAGTCCGAA				92.95	4.67	0.05	1.60E-03
CTGAGGCGCTTCCTGGG	Hs.78769	THOP1	19q13.3	101.40	4.67	0.05	9.82E-04
TACCATCAATAAAGGAC				109.85	4.67	0.04	6.03E-04

Appendix 3. Categories of differentially upregulated genes

Appendix 3.1 Differentially upregulated genes categorized according to chromosome of origin (pvalue < 0.05)

Category	Term	Count	%	Pvalue
CHROMOSOME	17	45	8.52%	6.11E-05
CHROMOSOME	19	37	7.01%	6.56E-04
CHROMOSOME	1	68	12.88%	1.13E-03
CHROMOSOME	20	18	3.41%	3.32E-02

Appendix 3.2 Differentially upregulated genes categorized according to cytoband of origin (pvalue < 0.05)

Category	Term	Count	%	Pvalue
CYTOBAND	11q13	10	1.89%	7.60E-10
CYTOBAND	6p21.3	12	2.27%	9.76E-09
CYTOBAND	17q25	6	1.14%	3.52E-06
CYTOBAND	1p32-p31	3	0.57%	4.30E-04
CYTOBAND	8q24.3	9	1.70%	8.85E-04
CYTOBAND	7p22	3	0.57%	2.21E-03
CYTOBAND	19p13.3	9	1.70%	5.43E-03
CYTOBAND	3p21	3	0.57%	5.62E-03
CYTOBAND	1p34	3	0.57%	6.32E-03
CYTOBAND	16p13.3	8	1.52%	9.26E-03
CYTOBAND	1q32	3	0.57%	1.28E-02
CYTOBAND	5q11	2	0.38%	1.39E-02
CYTOBAND	17q21	3	0.57%	1.82E-02
CYTOBAND	19p13.1	3	0.57%	2.25E-02
CYTOBAND	12q13	3	0.57%	2.57E-02
CYTOBAND	11p15.5-p15.4	2	0.38%	3.43E-02
CYTOBAND	10q25	2	0.38%	3.43E-02
CYTOBAND	22q13.1	4	0.76%	3.51E-02
CYTOBAND	15q22.31	3	0.57%	4.55E-02
CYTOBAND	17p13.3	5	0.95%	4.91E-02

Appendix 3.3 Differentially upregulated genes categorized according to biological process (pvalue < 0.05)

Term	Count	%	PValue
cell organization and biogenesis	73	13.83%	2.00E-07
organelle organization and biogenesis	45	8.52%	5.70E-06
chromosome organization and biogenesis	22	4.17%	4.38E-05
chromosome organization and biogenesis (sensu Eukaryota)	21	3.98%	5.44E-05
establishment and/or maintenance of chromatin architecture	18	3.41%	1.74E-04
DNA packaging	18	3.41%	2.55E-04
DNA metabolism	34	6.44%	2.72E-04
biopolymer metabolism	102	19.32%	3.58E-04
negative regulation of biological process	36	6.82%	6.60E-04
negative regulation of cellular process	34	6.44%	7.22E-04
cell cycle	34	6.44%	7.96E-04
response to DNA damage stimulus	17	3.22%	9.96E-04
negative regulation of cellular physiological process	31	5.87%	1.16E-03
response to endogenous stimulus	17	3.22%	1.92E-03
negative regulation of physiological process	31	5.87%	1.99E-03
regulation of progression through cell cycle	24	4.55%	2.16E-03
regulation of cell cycle	24	4.55%	2.23E-03

Appendix 4. Categories of differentially downregulated genes

Appendix 4.1 Differentially downregulated genes categorized according to chromosome of origin (pvalue < 0.05)

Category	Term	Count	%	Pvalue
CHROMOSOME	19	36	7.56%	1.99E-04
CHROMOSOME	17	40	8.40%	2.24E-04
CHROMOSOME	16	30	6.30%	7.21E-03

Appendix 4.2 Differentially downregulated genes categorized according to cytoband of origin (pvalue < 0.05)

Category	Term	Count	%	PValue
CYTOBAND	16p13.3	10	2.10%	3.02E-04
CYTOBAND	19p13.3	10	2.10%	7.58E-04
CYTOBAND	22q13	3	0.63%	3.55E-03
CYTOBAND	11q13	4	0.84%	7.46E-03
CYTOBAND	19q13.3	4	0.84%	1.03E-02
CYTOBAND	20p13	5	1.05%	1.04E-02
CYTOBAND	10q25-q26	2	0.42%	1.88E-02
CYTOBAND	1p34.3	4	0.84%	2.80E-02
CYTOBAND	14q	2	0.42%	3.11E-02
CYTOBAND	1q21	3	0.63%	3.18E-02
CYTOBAND	12q24.2	2	0.42%	3.42E-02
CYTOBAND	Xq13	2	0.42%	4.63E-02
CYTOBAND	12q24.1	2	0.42%	4.93E-02

Appendix 4.3 Differentially downregulated genes categorized according to biological process (pvalue < 0.05)

Term	Count	%	PValue
biosynthesis	83	17.44%	1.74E-13
cellular physiological process	315	66.18%	3.74E-13
macromolecule biosynthesis	60	12.61%	7.57E-13
macromolecule metabolism	177	37.18%	1.02E-12
cellular biosynthesis	75	15.76%	3.07E-12
protein biosynthesis	55	11.55%	4.88E-12
cellular metabolism	249	52.31%	5.14E-11
primary metabolism	240	50.42%	4.28E-10
metabolism	257	53.99%	1.71E-09
RNA metabolism	33	6.93%	5.66E-08
RNA processing	27	5.67%	7.03E-07
cellular macromolecule metabolism	117	24.58%	2.66E-06
cellular protein metabolism	115	24.16%	3.73E-06
biopolymer metabolism	103	21.64%	9.30E-06
protein metabolism	120	25.21%	1.21E-05
mRNA processing	17	3.57%	3.82E-05
translation	16	3.36%	3.91E-05
mRNA metabolism	18	3.78%	4.98E-05

Appendix 5 Differentially regulated Potential novel tags (p<0.05)

Appendix 5.1 Differentially upregulated potential novel tags (p<0.05)

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
AACACGGTGCTCAGGGG	unknown	unknown	unknown	4.23	102.71	24.31	4.90E-04
GGGTTGCTGTAAGGATT	unknown	unknown	unknown	4.23	93.37	22.10	8.87E-04
CAGACCTTAATAAATAC	unknown	unknown	unknown	4.23	74.70	17.68	2.95E-03
CCAGGGGGAGAAGGCAC	unknown	unknown	unknown	4.23	74.70	17.68	2.95E-03
AATTAAAAAAAAAAAAA	unknown	unknown	unknown	4.23	65.36	15.47	5.42E-03
TTGAAGGGAAGAGGGGA	unknown	unknown	unknown	4.23	65.36	15.47	5.42E-03
CCAGGGGAAGAAGGCAC	unknown	unknown	unknown	4.23	56.02	13.26	1.00E-02
CTAACGTTGCAGACCCC	unknown	unknown	unknown	4.23	56.02	13.26	1.00E-02
GCTCTGTCACAATAGGG	unknown	unknown	unknown	4.23	56.02	13.26	1.00E-02
TAACACAAGCTCACAGC	unknown	unknown	unknown	4.23	56.02	13.26	1.00E-02
AAATGACAAGTCCTTGG	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
ACATTAAAATATTACTC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
ACGCAGCACGGGCAGGG	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
ACTGTGGGAAAGTGGGC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
CAACCATCATCTTCCAC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
CCAGGGGAGAAGGCCCC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
GGCACAGAGGCTGGGGA	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
GGCGACACGAGCCGATC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
GGGGAGGGGGTACTGCC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
GGGGGGGGGTTTGATGT	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
GTTCTCTTCTCTGCAGG	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
TCCTGCAATTCTGAAGG	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
TGAGAATATAAATATTC	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
TGATTTCACTTACCACT	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
TTGAATAAATTTGTGAG	unknown	unknown	unknown	4.23	46.68	11.05	1.87E-02
AAAAAAAAAGCCAATCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AAATAGATCCACCTGCT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AAGAACGCCAGGGAGCT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AAGGATTACACTAGTCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AATGGATTAGAAATGGG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
ACCCCCAAACAGAACCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
ACCCCTGCTCCAGCAAC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
ACCGGGACTTCATTCGG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AGAAGCAAGAAGTATCA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AGCTACGGAAACAGGCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
AGGTATTTCTTCCTTCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
ATACAGATTGGTTTTGC	unknown	unknown	unknown	8.45	74.70	8.84	1.29E-02
CACAGTTCTCTTATACC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CAGGAACCACAGTGGCT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CCCCCTGGATCAGGCCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CCCTGGGTCCTGCCCGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CCGCCTTTAAGAACTGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CCTGGAAGGAACAAGGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CCTTACTTTATCAAATG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CTATACTAATGCTGTTG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
CTGTATACTTAAGAGGA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GAATTCAGCCCGATGGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GAATTCTACAGAAACCA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GAGTAAACTGGTACCTG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GAGTTGTTCAACCTGCC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GCACGACACGAGCCGAT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GCCAAGGAAGGCATTGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GCCACACACGATGAGGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GCCCCAGTTCACTATTC	unknown	unknown	unknown	8.45	74.70	8.84	1.29E-02
GCCGCCATCCGCAGGGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GCTTTTTAGGATACCGG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GGAAGTGATCTGCAAGA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GGCGGCTGCCAGATCCA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GGGACGGGTGCCTGTAA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GTCTGGAAGCGGGGGGG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GTGCTGGACCTGAAGGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GTTGCCCTGGCCCGTGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
GTTGCGGTTAATCTGGT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TAAATAATAAAAGAGAG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TACTCTATAACTTAAAA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TACTTACTAATATGTTG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TAGGCAACACGAGCAGG	unknown	unknown	unknown	8.45	74.70	8.84	1.29E-02
TGACCATCAATAAAGTA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TGAGGACTCAATGAGGT	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TGGAACCCTGATGCAGC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TGGACAAGCTAAGTGGG	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TGGGCAGGGACAAATAA	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TTACGACTTGTCTCCTC	unknown	unknown	unknown	4.23	37.35	8.84	3.55E-02
TAAATTAGAAGGGACAG	unknown	unknown	unknown	8.45	65.36	7.73	2.35E-02
CCTGTAGATGGATTGGG	unknown	unknown	unknown	8.45	56.02	6.63	4.29E-02
GACGACACCAGCCGATC	unknown	unknown	unknown	8.45	56.02	6.63	4.29E-02
GCACAGTAACCAAATCC	unknown	unknown	unknown	8.45	56.02	6.63	4.29E-02
GGGGCTCCAGCCTCAGG	unknown	unknown	unknown	8.45	56.02	6.63	4.29E-02
TAAATAAAATTTCAAGG	unknown	unknown	unknown	8.45	56.02	6.63	4.29E-02
CCTCAGGATACTCCTCA	unknown	unknown	unknown	25.35	130.72	5.16	4.01E-03

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
GGATTTCACTTCCACTC	unknown	unknown	unknown	16.90	84.03	4.97	2.27E-02
AGGGTACGGAAACAGGC	unknown	unknown	unknown	16.90	74.70	4.42	3.96E-02
GGAGAACCTAGGGGAGG	unknown	unknown	unknown	16.90	74.70	4.42	3.96E-02
TTTTATAAACAGGACCC	unknown	unknown	unknown	25.35	102.71	4.05	1.99E-02
TGATTTCACTTCCACTC	unknown	unknown	unknown	1140.78	3454.65	3.03	0.00E+00
GGGTTGGCTTGAAACCA	unknown	unknown	unknown	59.15	177.40	3.00	9.03E-03
AGACCCACAACAAATAG	unknown	unknown	unknown	84.50	252.10	2.98	1.92E-03
GCAAGCCAACGCCACTT	unknown	unknown	unknown	59.15	149.39	2.53	3.41E-02
TCGAAGCCCCCATCGCT	unknown	unknown	unknown	76.05	168.06	2.21	4.62E-02
GAGGCCGCGGGGTGGGG	unknown	unknown	unknown	152.10	298.78	1.96	1.95E-02
GTAGGGGTAAAAGGAGG	unknown	unknown	unknown	1292.88	2054.12	1.59	8.95E-06
ACTAACACCCTTAATTC	unknown	unknown	unknown	828.12	1307.17	1.58	4.69E-04
CACCTAATTGGAAGCGC	unknown	unknown	unknown	1005.58	1587.27	1.58	1.16E-04
TAGGTAGCTCATTCAGG	unknown	unknown	unknown	1005.58	1409.87	1.40	5.58E-03
TTCATACACCTATCCCC	unknown	unknown	unknown	3371.64	4500.38	1.33	1.79E-05

Appendix 5.2 Differentially downregulated potential novel tags ($p<0.05$)

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
CTAACTAGTTACGCGAC	unknown	unknown	unknown	980.23	634.91	0.65	4.14E-03
ACATTAAAATAAATTTG	unknown	unknown	unknown	464.76	280.11	0.60	2.41E-02
AAAAAGCAGATGACTCG	unknown	unknown	unknown	321.11	186.74	0.58	4.70E-02
TTTTGTAACTGTAGAGG	unknown	unknown	unknown	456.31	252.10	0.55	1.06E-02
GGCATTTTGTTTATGAG	unknown	unknown	unknown	338.01	177.40	0.52	1.86E-02
ACCCGCCGGGCAGCTTC	unknown	unknown	unknown	1123.88	588.22	0.52	1.64E-05
TACCATCAATAAAGTCC	unknown	unknown	unknown	211.26	93.37	0.44	2.49E-02
TTGATGCCATTTCAATA	unknown	unknown	unknown	211.26	93.37	0.44	2.49E-02
AGGCTACGGAAAACAGG	unknown	unknown	unknown	160.55	65.36	0.41	3.56E-02
GGGCCATATATCTTGGG	unknown	unknown	unknown	177.45	65.36	0.37	1.71E-02
GGGGCAGGGCCCTCCCA	unknown	unknown	unknown	312.66	112.04	0.36	1.25E-03
GAAAAAAGGTTGATGGA	unknown	unknown	unknown	109.85	37.35	0.34	4.77E-02
GAGTTAGTGAGGCGCTC	unknown	unknown	unknown	109.85	37.35	0.34	4.77E-02
TTCCATCAATAAAGTAC	unknown	unknown	unknown	219.71	65.36	0.30	2.49E-03
TACCATCAATAAAGTTC	unknown	unknown	unknown	152.10	37.35	0.25	5.88E-03
TTGGGGTTTGTGGGTCG	unknown	unknown	unknown	202.81	46.68	0.23	1.10E-03
GAAAAATGGTTGATGGG	unknown	unknown	unknown	169.00	37.35	0.22	2.48E-03
GCCGTGAGCAGAACTGT	unknown	unknown	unknown	92.95	18.67	0.20	2.04E-02
GCCTCCTCTTTGCTGAT	unknown	unknown	unknown	92.95	18.67	0.20	2.04E-02
TGGCAAGATGAGAGTAA	unknown	unknown	unknown	92.95	18.67	0.20	2.04E-02
TGGCGTTGAGGAAAGCA	unknown	unknown	unknown	92.95	18.67	0.20	2.04E-02
TCAAAAAAAAGAAAGGT	unknown	unknown	unknown	143.65	28.01	0.19	3.60E-03
GAAAAATGGCTGATGGA	unknown	unknown	unknown	101.40	18.67	0.18	1.28E-02
TCCATCAATAAAGTACC	unknown	unknown	unknown	211.26	37.35	0.18	2.77E-04
ATGCTTTCTTTTGTTTC	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
CGCACTGCATTCTCCGG	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
GACTTTTAAGGATACCG	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
GAGAGGGAAGCGGGAGG	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
GATGAAAACTAAAGGCT	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
GGCCGCGTTCGCCCCAA	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
GGTGCAGGGAGGGGTGG	unknown	unknown	unknown	59.15	9.34	0.16	4.74E-02
CATCTAGAGGGCTCATA	unknown	unknown	unknown	126.75	18.67	0.15	3.17E-03
TAAAGGTTTTTTTTTTC	unknown	unknown	unknown	126.75	18.67	0.15	3.17E-03
CAAATGATGCCTCTGCC	unknown	unknown	unknown	67.60	9.34	0.14	2.88E-02
GTAAAGACTGCTTTATT	unknown	unknown	unknown	67.60	9.34	0.14	2.88E-02
TCCCATCAATAAAGTCC	unknown	unknown	unknown	67.60	9.34	0.14	2.88E-02
ATGGCGATCTATCTTCA	unknown	unknown	unknown	143.65	18.67	0.13	1.25E-03
GCCGTTCTTAGTTGGGG	unknown	unknown	unknown	76.05	9.34	0.12	1.75E-02
TACCATCAATAAATACC	unknown	unknown	unknown	76.05	9.34	0.12	1.75E-02
TGAGGGAATAACCCTGG	unknown	unknown	unknown	76.05	9.34	0.12	1.75E-02

Tag	UniGene ID	Gene symbol	cytoband	WT	DKO2L	DKO2L/WT	twotailp
ACATTTCATTTGGTAAC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
ATAATCTCCACTTGGTC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
ATGCTGATCCACTTGGG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
CACAAACGGCAGTTTTG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
CCACTCTCAGAGGTTGG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
CTAGCCTCACGAAAACT	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GAAATAACATAGGGTTC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GAGCAGTTCCAGAAGGA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GATCAGCAGCAAGTCCA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GGAAACTCTGTTTATGC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GGCAAGAAAAAAATCGC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GGCAGGCCCAAGAAGGG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GGTACCACTGAAAGAAA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GTGACCACGGGTGAACG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TCGGGGAGGGGAGGGGA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TGTGGGTCTTCAGATAC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TTGATTTAATAAAGTCA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TTGTGTGTACCCAGATA	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TTTGCGGCAGAGGTCTC	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
TTTTCCTTTTGTTTTTG	unknown	unknown	unknown	42.25	4.67	0.11	3.34E-02
GAGGAGGAGGTTGATCC	unknown	unknown	unknown	84.50	9.34	0.11	1.07E-02
TTACATTTTAGGTACCT	unknown	unknown	unknown	177.45	18.67	0.11	1.93E-04
TACAATTTGCCACTGGG	unknown	unknown	unknown	92.95	9.34	0.10	6.58E-03
ATGGCCGGTATCGATGA	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
CCGTAGGTGGGCACAGG	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
GGCTGAAGTTGGCATTT	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
GTGACCACGGGCGACGG	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
TCAACTAACAATAAAGG	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
TTGGGGTTTCCTTTTCC	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
TTGGTGAAGGAAGAATG	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
TTTAAAAGGTCTTAGCC	unknown	unknown	unknown	50.70	4.67	0.09	1.98E-02
TACCATCAATAAAAGTA	unknown	unknown	unknown	101.40	9.34	0.09	4.04E-03
CTTTACGGGCTTGTAGG	unknown	unknown	unknown	59.15	4.67	0.08	1.18E-02
TGACTGGCCATTTGTGC	unknown	unknown	unknown	59.15	4.67	0.08	1.18E-02
CATCTAGAGGGCTCAAT	unknown	unknown	unknown	76.05	4.67	0.06	4.32E-03
TGCCATCTGTACATAAA	unknown	unknown	unknown	76.05	4.67	0.06	4.32E-03
TGGGGGATATCCACGGC	unknown	unknown	unknown	76.05	4.67	0.06	4.32E-03
ACTACAAATAGTCCGAA	unknown	unknown	unknown	92.95	4.67	0.05	1.60E-03
TACCATCAATAAAGGAC	unknown	unknown	unknown	109.85	4.67	0.04	6.03E-04

Appendix 6. List of top 10 up and down regulated genes commonly identified by longSAGE and exon array (Affymetrix) platforms.

Appendix 6.1 List of top 10 upregulated genes commonly identified by longSAGE and exon array (Affymetrix) platforms.

Gene	LongSAGE Tag	Description	SAGE ratio	Affy ratio
IFI27	CCAGGGGAGAAGGCACC	Interferon, alpha-inducible protein 27	269.60	29.35
UCHL1	CAGTCTAAAATGCTTCA	Ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	134.80	85.99
PRSS21	CAGCCTGGGGCCACTGC	Protease, serine, 21 (testisin)	112.70	15.04
BST2	TGCTGCCTGTTGTTATG	Bone marrow stromal cell antigen 2	110.49	3.98
MT1A	AGCTGTGCCAAGTGTGC	Metallothionein 1A	44.20	17.85
HSPA1A	CAGAGATGAATTTATAC	Heat shock 70kDa protein 1A	35.36	35.31
CTCFL	AGATTTAAATTCTGTGG	CCCTC-binding factor (zinc finger protein)-like	35.36	230.65
YPEL3	CACTGTGACCTTGGGGG	Yippee-like 3 (Drosophila)	24.31	1.99
RCN3	TTTGTGGGCAGTCAGGC	Reticulocalbin 3, EF-hand calcium binding domain	22.10	10.51
MYO1D	ATTGTAGACAATGAGGG	Myosin ID	22.10	1.61

Appendix 6.2 List of top 10 downregulated genes commonly identified by longSAGE and exon array (Affymetrix) platforms.

Gene	LongSAGE Tag	Description	SAGE ratio	Affy ratio
TRIB1	GGACTTTGAGAAGAGGG	Tribbles homolog 1 (Drosophila)	0.14	0.64
APOE	CGACCCCACGCCACCCC	Apolipoprotein E	0.13	0.52
NOB1	CATTTAGATTTATATTT	NIN1/RPN12 binding protein 1 homolog (S. cerevisiae)	0.11	0.44
PRKACB	AGCAAATATGTCAAGGG	Protein kinase, cAMP-dependent, catalytic, beta	0.11	0.13
RPLP0	TCCCTTCGCGGGAAGGC	Ribosomal protein, large, P0	0.11	0.65
MVD	GGGAATAAACCAGCATT	Mevalonate (diphospho) decarboxylase	0.08	0.75
HMGCS1	TAATACTTTTGCAAGTG	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	0.08	1.14
CDKN2AIP	CACTCTAAGATGAGTTC	CDKN2A interacting protein N-terminal like	0.06	0.27
DNMT1	AAGCTGTTGTGTGAGGT	DNA (cytosine-5-)-methyltransferase 1	0.06	0.06
THOP1	CTGAGGCGCTTCCTGGG	Thimet oligopeptidase 1	0.05	0.46

Manuscripts under preparation

1. Tiwary, PK, Li, J, Akagi, K, Symer DE: ***Toward a transcriptome signature of hypomethylated colorectal cancer cells.*** 2012

2. Tiwary, PK, Li, J, Symer, DE: ***Epigenetic regulation of L1 retrotransposition*** 2012

List of conference Presentations

Oral Presentation:

1. Laboratory of Immunobiology retreat, Harpers Ferry, West Virginia, May 12, 2003
2. Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute, Frederick, MD, USA. Feb. 10, 2004
3. Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute, Frederick, MD, USA. March 10, 2005
4. Mouse Cancer Genetics Program, National Cancer Institute, Frederick, MD, USA. June 11, 2006
5. Department of Microbiology, Uniformed Services University of Health Science, Bethesda, MD, USA. March 27, 2007
6. Mouse Cancer Genetics Program, National Cancer Institute, Frederick, MD, USA. June 25, 2007

Poster Presentation:

1. NCI fellows and young investigator retreat, Williamsburg, VA, USA. March 2, 2005
2. Center for Excellence in Chromosome Biology retreat, Washington DC, USA, May 12, 2006
3. Genetics Retreat 3, Cumberland MD, USA, June19, 2006
4. FARE award ceremony, NIH, Bethesda, MD, USA, Sept. 25, 2006
5. NCI fellows and young investigator retreat, Ocean City, MD, USA, March 1, 2007
6. Keystone meeting (Epigenetics: Regulation of Chromatin Structure in Development and Disease (D3), Breckenridge, Colorado, USA, April 16, 2007

Biography

Pawan Kumar Tiwary B. Pharm (Hons.), ME (Biotech)

Pawan earned his B Pharm (Hons) from India Institute of Technology, Banaras Hindu University in 2001. He attended BITS-Pilani from 2001-2003 and graduated with Master of Engineering degree in biotechnology. Pawan moved to National Cancer Institute –Frederick (a part of NIH, USA) to pursue his doctoral research work under the guidance of Dr, David Symer; he received Fellow's Award for Research Excellence (FARE) in 2007 granted by NIH, Maryland, USA.for his research work at NCI. Pawan continued his graduate studies at NCI-Frederick through 2008 and, subsequently moved to India. In India, he worked as a Staff Scientist/Project Manager at Life Technologies (formerly Invitrogen) from 2008 to 2010. At Invitrogen, he managed projects related to developing tools for microRNA research such as microRNA inhibitors and mimetic. He also managed projects related to development and validation of monoclonal antibodies for research use. Pawan is currently employed as a Senior Analyst and Project Manager at Prescient Life Sciences, India (a part of Business Research Group, UK). At Prescient, he has gained experience in strategic consulting, market research and business intelligence in pharmaceutical and biotech sector. His area of expertise are oncology and biosimilars.

Biography

Dr. David Eric Symer MD, PhD

Dr. David E. Symer graduated *Summa cum laude* with an A.B. degree in mathematics from Dartmouth College in Hanover, New Hampshire in 1984. He attended the Johns Hopkins University School of Medicine in Baltimore, Maryland in the Medical Scientist Training Program from 1984 to 1993, and received the W. Barry Wood student research award in 1992. His dissertation research was performed in the laboratory of Dr. Howard M. Dintzis. He graduated with M.D. and Ph.D. degrees in biophysics and biophysical chemistry in 1993. He subsequently was a medical intern and resident in Medicine at the Brigham and Women's Hospital, Harvard Medical School until 1995. He served as a clinical fellow in Hematology and Medical Oncology at the Brigham and Women's Hospital from 1995 until 1996, and at the Johns Hopkins Hospital and Johns Hopkins Oncology Center through 2002. Dr. Symer completed his postdoctoral research fellowship in the laboratory of Dr. Jef D. Boeke from 1997 to 2002. Subsequently he was a Principal Investigator in the Center for Cancer Research, National Cancer Institute, Frederick, Maryland from 2002 to 2009. In 2009, Dr. Symer's research group was recruited to the Ohio State University Comprehensive Cancer Center, where he directs his research laboratory, teaches students, sees patients with chronic lymphocytic leukemia and serves on the College of Medicine admissions committee. He has co-authored more than 20 research papers and editorials, including more than 8 first or senior authorships, in numerous high impact journals.

The Symer research group studies retrotransposons and other genomic elements and molecular processes that contribute to phenotypic diversity in mouse and man. They aim to understand some of the causes and consequences of changing epigenetic states upon the transposons, the genes and transcripts they affect, and on genomic stability. In particular, specific examples of transcriptional and phenotypic variation that are due to existing or *de novo* endogenous retrotransposon integrants or due to variants in other noncoding genomic sequences are being identified and characterized. Since in most cases the molecular basis for phenotypic variation between individuals or between different tissues is not known, these questions are centrally important. They use and develop state-of-the-art molecular techniques and assays using conventional and next generation sequencing, exon microarrays, mouse models, and bioinformatics tools. The Symer lab group uses a range of modern molecular and bioinformatics tools to study the movement, the epigenetic controls and the consequences of retrotransposons and other genomic variants in normal development and aging and in certain cancers such as chronic lymphocytic leukemia.

Recently, applying expertise in molecular biology and bioinformatics, the Symer group discovered several disease-causing mutations resulting in the devastating human developmental disorder, microcephalic osteodysplastic primordial dwarfism type I (MOPD I). They conducted targeted deep sequencing at linked chromosomal loci in affected patients and family members, and identified biallelic mutations at *RNU4ATAC,* a small non-protein coding small nuclear RNA (snRNA) gene involved in minor spliceosomal functions in pre-mRNA splicing. More recently they identified high numbers of low frequency allelic variants within and near other

snRNA genes in human populations. They have identified striking examples of substantial transcriptional and phenotypic variation between mouse strains and in human cancers that are due to existing or *de novo* endogenous retrotransposon integrants and due to variable epigenetic controls that regulate them. They are characterizing the detailed molecular mechanisms by which these innate transposons can contribute to genomic and functional variation, by exploiting a robust toolkit of human and mouse genetics and genomics techniques. The group's long-term goal is to translate resulting observations on the mechanisms of aberrant transcriptional splicing, functional variation and disease pathogenesis to improve our understanding of human diversity and variable susceptibility to diseases including cancers and leukemias.