

**THE ANALYSIS OF THE MCF7 CANCER MODEL SYSTEM AND THE EFFECTS
OF 5-AZA-2'-DEOXYCYTIDINE TREATMENT ON THE CHROMATIN STATE
USING A NOVEL MICROARRAY-BASED TECHNOLOGY FOR HIGH
RESOLUTION GLOBAL CHROMATIN STATE MEASUREMENT**

APPROVED BY SUPERVISORY COMMITTEE

Harold Garner, Ph.D.

Elliott Ross, Ph.D.

Thomas Kodadek, Ph.D.

John Minna, M.D.

Keith Wharton, M.D., Ph.D.

DEDICATION

Omnibus qui adiuverunt

**THE ANALYSIS OF THE MCF7 CANCER MODEL SYSTEM AND THE EFFECTS
OF 5-AZA-2'-DEOXYCYTIDINE TREATMENT ON THE CHROMATIN STATE
USING A NOVEL MICROARRAY-BASED TECHNOLOGY FOR HIGH
RESOLUTION GLOBAL CHROMATIN STATE MEASUREMENT**

by

MICHAEL RYAN WEIL

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

March, 2006

Copyright

by

Michael Ryan Weil

All Rights Reserved

**THE ANALYSIS OF THE MCF7 CANCER MODEL SYSTEM AND THE EFFECTS
OF 5-AZA-2'-DEOXYCYTIDINE TREATMENT ON THE CHROMATIN STATE
USING A NOVEL MICROARRAY-BASED TECHNOLOGY FOR HIGH
RESOLUTION GLOBAL CHROMATIN STATE MEASUREMENT**

Publication No. _____

MICHAEL RYAN WEIL, B.S.

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Professor: Harold Ray (Skip) Garner, Ph.D.

A microarray method to measure the global chromatin state of the human genome was developed in order to provide a novel view of gene regulation. The 'chromatin array' employs traditional methods of chromatin isolation, microarray technology, and advanced data analysis, and was applied to a cancer model system. Chromatin is first separated by its condensation state using chromatin fractionation. By probing with a comparative genomic hybridization-style microarray, the chromatin condensation state of thousands of individual loci in an MCF7 tumor model cell line was determined and correlated with transcriptional activity. The chromatin array showed a significant portion (>3,000) of the genes were in a condensation state that was neither condensed or relaxed as a result of heterogeneity in the

condensation states in the population. The utility of the chromatin array in deciphering gene regulation was demonstrated in a MCF7 cell line treated with 5 Aza dC, which disrupts genome methylation, and as a result causes global relaxation of chromatin structure. 5 Aza dC treatment results in strong changes in expression, and a normalized global chromatin relaxation of two-fold. A significant subset of 378 genes was condensed by 5 Aza dC treatment, indicating that a mechanism of chromatin regulation exists that can resist the effects of 5 Aza dC treatment. The genes with the largest changes in response to 5 Aza dC treatment showed a strong correlation with CpG island-based regulation ($p < 0.0001$), and a restoration of transcription patterns associated with normal mammary tissue. Analysis using splice-form specific microarray probes demonstrated that the chromatin state was not uniform across a gene. These findings indicate that certain gene regions exhibit differential sensitivity to 5 Aza dC treatment, and therefore may be regulated independently. Using functional annotation, expression microarray, and comparative genomic hybridization data, this work should provide a framework through which the biological implications of the relationship between chromatin accessibility and expression may be deciphered.

TABLE OF CONTENTS

Chapter One: Background

1-1 Introduction	20
1-2 The basics of chromatin.....	22
1-2-1 The history of chromatin research: structural versus	24
1-2-2 Modern chromatin research: The unification of structure and function.....	26
1-2-3 Chromatin ImmunoPrecipitation.....	29
1-3 Microarray technology	30
1-3-1 A history of transcriptomics.....	30
1-3-2 A history of microarray technology	32
1-3-3 Spotted versus light-directed synthesis	35
1-3-4 Expression microarray methodology	38
1-3-5 Microarray-based comparative genomic hybridization	42
1-3-6 Chromatin immunoprecipitation on microarrays	44
1-4 Thesis and organization.....	47
1-5 References	48

Chapter Two: Development of techniques for robust hybridization of human genomic DNA to cDNA microarrays

2-1 Specificity and sensitivity of cDNA versus BAC microarrays	54
2-2 Measurement of human genomic DNA coverage on cDNA microarrays.....	56
2-3 Refinements to the high-resolution aCGH method	61
2-4 Materials and methods.....	64

2-4-1 Genomic DNA Labeling	64
2-4-2 RNA Labeling	65
2-4-3 Microarray Production	65
2-4-4 Microarray Clean Up and Hybridization.....	66
2-4-5 Washing and Scanning.....	66
2-4-6 Analysis.....	67
2-5 References	68

Chapter Three: Development of a microarray-based assay to measure chromatin condensation state at a subgene resolution

3-1 Development of methods to recover chromatin by its condensation state in preparation for hybridization	70
3-1-1 Fragment Length Selection	71
3-1-2 Chromatin Fractionation	74
3-2 Validation of the microarray-based method	77
3-2-1 Validation by traditional methods of chromatin state measurement.....	77
3-2-2 Validation of the chromatin array method	78
3-2-3 Cross validation of the two chromatin state fractionation techniques	82
3-3 Insight into transcriptional regulation revealed by global high resolution chromatin studies	83
3-3-1 Combining chromatin array and expression data increases the robustness of the functional classification	84

3-3-2 Novel information provided by high-throughput global chromatin state measurements of synchronized cell culture to resolve heterogeneity issues	87
3-3-3 Novel information resulting from chromatin state measurements of genes with expression state too low for microarray-based measurement	89
3-3-4 The chromatin state data provides significant insight into the relationship between chromosomal alteration and transcription	91
3-4 Materials and methods.....	94
3-4-1 Cell culture	94
3-4-2 Cell cycle synchronization by serum starvation and population analysis.....	95
3-4-3 DNase I digestion and fragment length selection.....	95
3-4-4 Micrococcal nuclease digestion and the chromatin fractionation assay	96
3-4-5 Microarray characteristics	96
3-4-6 DNA labeling	97
3-4-7 Chromatin hybridization	97
3-4-8 RNA extraction and hybridization	98
3-4-9 Data extraction, normalization and analysis	98
3-4-10 Data interpretation.....	99
3-5 References	100

Chapter Four: Studies of alterations in the chromatin state of the MCF7 cancer model cell line induced by the drug 5-Aza-2'-Deoxycytidine

4-1 Selection of an epigenetic modifying drug.....	103
--	-----

4-2 Experimental details and validation	107
4-2-1 Examination of the microarray data variance	109
4-2-2 Examination of genes known to be altered by 5 Aza dC treatment	111
4-3 The effects of 5 Aza dC treatment on the chromatin state of MCF7.....	114
4-3-1 Discoveries emerging from the experiment: Functional classification view .	114
4-3-2 Discoveries emerging from the experiment: gene view	120
4-3-3 Novel information derived from genes with expression levels too low to be measured with microarray based methods	128
4-4 Chromatin shows intragenic variation.....	132
4-4-1 Chromatin regulation of the <i>hTert</i> locus	135
4-4-2 Chromatin regulation of the <i>PRDM2</i> locus.....	141
4-5 Materials and methods.....	146
4-5-1 5 Aza dC treatment.....	146
4-5-2 Chromatin fractionation	147
4-5-3 Microarray characteristics	147
4-5-4 Chromatin hybridization	148
4-5-5 RNA extraction and hybridization	148
4-5-6 Data extraction, normalization and analysis	148
4-5-7 Functional classification of the data.....	150
4-5-8 Selection and analysis of the genes with largest change.....	151
4-5-9 Alternative splicing analysis	152
4-5-10 Expanded analysis of the <i>hTert</i> locus.....	152

4-5-11 Chromatin quantitative PCR analysis	153
4-6 References	154
Chapter Five: Improvements to the chromatin array method and its potential for expanded application	
5-1 Development of an optimized microarray platform for the study of chromatin	160
5-2 Development of methods to enable the analysis of the chromatin state of tissue samples.....	161
5-3 Surveys to identify signatures of various cellular states.....	165
5-3-1 Breast cancer surveys	166
5-3-2 Development of a chromatin biosignature for other cancers	169
5-4 Correlating ChIP on chip results with chromatin array	170
5-4-1 The SouthWestern Array.....	173
5-5 References	174
APPENDICES	177
VITAE.....	

PRIOR PUBLICATIONS

Weil M.R., Rosenblatt K.P., Garner H.R., The use of DNA microarray analysis tools to perform metadata guided analyses of serum protein patterns. (2005) *Cellular and Molecular Proteomics*, (Submitted).

Weil M.R., Garner H.R., 5 Aza dC induced chromatin modulation shows differential regulation at the sub-gene level. (2005) *Genes, Chromosomes and Cancer*, (Submitted).

Mizumoto, N., Hui F., Edelbaum D., Weil M.R., Wren J.D., Shalhevet D., Matsue H., Liu L., Garner H.R., Takashima A., Differential activation profiles of multiple transcription factors during dendritic cell maturation. (2005) *Investigative Dermatology* 124:718-724.

Beloludtsev Y.Y., Bowerman D., Weil R., Marthandan N., Balog R., Luebke K., Lawson J., Johnson S.A., Lyons C.R., O'Brien K., Garner H.R., Powdril T.F. Organism Identification Using a Genome Sequence-Independent Universal Microarray Probe Set. (2004) *BioTechniques*, 37:654-658.

Weil M.R., Widlak P., Minna J.D., Garner H.R. Global survey of chromatin packing state using DNA Microarrays. (2004) *Genome Research*, 14:1374-1381. (Featured in the October 2004 issue of MIT Technology Report.)

Weil M.R., Macatee T., and. Garner H.R., Towards a universal standard: Comparing two methods for standardizing spotted microarray data. (2002) *Biotechniques*, 32:1310-1314.

LIST OF FIGURES

FIGURE 1.1 Chromatin condensation annotated with the transcriptional state of the various packing levels.....	24
FIGURE 1.2 The 2.8 Å crystal structure of the nucleosome bound to DNA	29
FIGURE 1.3 A diagram of mask-based lithographic oligonucleotide synthesis	36
FIGURE 1.4 A schematic of a maskless lithographic synthesizer	37
FIGURE 1.5 A simplified depiction of the microarray work flow.....	40
FIGURE 1.6 The data extraction process for a typical spotted microarray.....	41
FIGURE 1.7 A diagram of the Chromatin Immunoprecipitation microarray work flow.....	45
FIGURE 2.1 The log ₁₀ of the background-subtracted spot intensities sorted by value shows the minimal difference between the two samples	58
FIGURE 2.2 A portion of a DOC microarray with probes for over 8,000 human genes, used for aCGH.....	62
FIGURE 2.3 CoT1 DNA is able to bind and compete off the MCF7 DNA even using probes designed to be sequence specific.	63
FIGURE 3.1 A diagram showing the chromatin array method	71
FIGURE 3.2 A diagram of the fragment length selection method	72
FIGURE 3.3 A fragment length selection gel showing the distribution of DNase I cleaved fragments with the concentration of DNase I increasing from left to right.....	74
FIGURE 3.4 A diagram showing a model of how the chromatin fraction assay works with example gels of the DNA, and Protein for each fraction.....	76

FIGURE 3.5 A comparison of replicate chromatin microarrays shows the reproducibility of the method.....	80
FIGURE 3.6 The log scale scatter plot of the co-normalized illustrates the relationship between condensation and the absolute RNA expression	81
FIGURE 4.1 A diagram showing the methylation of normal 2'-deoxycytidine with the methyl group added to the 5 position in presence of DNA methyl transferase and the structure of 5-aza-2'-deoxycytidine.....	106
FIGURE 4.2 A scatter plot of the unfiltered raw intensity data from the replicate chromatin arrays of the 5 Aza dC treated MCF7 cells.....	110
FIGURE 4.3 A: 5 Aza dC responsive genes exhibit chromatin state changes as measured by the chromatin array. B: The expression state of the 5 Aza dC responsive genes.....	113
FIGURE 4.4 A micrograph of several MCF7 cells after treatment.....	133
FIGURE 4.5 A graphic of the structure of DMD showing the areas interrogated by the chromatin and expression arrays and their response to treatment	134
FIGURE 4.6 Intensity changes across probes specific to distinct <i>hTert</i> regions show that the chromatin state varies over the length of the loci	138
FIGURE 4.7 A histogram of the intragenic CpG islands associated with the α and β splice forms of <i>hTert</i>	140
FIGURE 4.8 Isoforms of PRDM2 probed on the chromatin array.....	143
FIGURE 4.9 The amplification curves of the 1ng and 10 ngs of DNA per reaction tests of primers and probes for the two different <i>PRDM2</i> isoforms (called long and short)	145

LIST OF TABLES

TABLE 3.1 The validation of a selection of genes from the BL60 chromatin fractionation	77
TABLE 3.2 Measurements of the Chromatin Array method's reproducibility	79
TABLE 3.3 The functional signatures of the primary groups	87
TABLE 3.4 The functional signatures of chromosomal alterations measured by accessibility state.....	92
TABLE 4.1 Functional classes based on expression change after 5 Aza dC treatment	115
TABLE 4.2 Functional classes based on chromatin change after 5 Aza dC treatment	116
TABLE 4.3 Functional classifications of genes with 2 fold increased expression after treatment further sub-divided by chromatin state	118
TABLE 4.4 Functional classifications of genes with 2 fold decreased expression after treatment further sub-divided by chromatin state	120
TABLE 4.5 A selection of genes with the largest treatment induced expression changes.	124
TABLE 4.6 A selection of genes with the largest treatment induced chromatin changes .	127
TABLE 4.7 Genes with measurable chromatin state but not measurable expression	131
TABLE 4.8 Selected genes with intragenic chromatin variation	136
TABLE 4.9 <i>hTert</i> exons 5 and 6 bisulfite sequencing results	141
TABLE 4.10 Fold changes in PRDM2 after treatment by platform	146
TABLE 4.11 TaqMan® primers and probes for the PRDM2 CQ-PCR validation	154

LIST OF APPENDICES

APPENDIX A 177

APPENDIX B 180

LIST OF DEFINITIONS

aCGH - array Comparative Genomic Hybridization

ChIP - Chromatin ImmunoPrecipitation

ChIP on chip - Chromatin ImmunoPrecipitation on microarray

DNA - Deoxyribonucleic Acid

PCR – Polymerase Chain Reaction

Å – Ångstrom

RNA – Ribonucleic Acid

mRNA – messenger RNA

RT – Reverse Transcription

cDNA - complementary DNA

SSC – Saline sodium citrate

DMSO – Dimethyl Sulfoxide

UV – Ultraviolet

DMD – Digital Mutlimirror Device

DLP – Digital Light Processor

SNP - Single Nucleotide Polymorphisms

LOH - Loss of Heterozygosity

FISH - Fluorescent *In Situ* Hybridization

CGH - comparative genomic hybridization

BAC - Bacterial Artificial Chromosomes

YAC - Yeast Artificial Chromosomes

LOWESS – LOcally WEighted Scatterplot Smoothing

SOM - Self Organizing Maps

SVM - Support Vector Machines

QT – Quality Threshold

PCA - Principle Component Analysis

GO - Gene Ontology

NCBI – National Center for Biotechnology Information

EBI – European Bioinformatics Institute

GEO – Gene Expression Omnibus

MGED - Microarray Gene Expression Data

MIAME - minimum information about a microarray experiment

XML - Extensible Markup Language

MAGE-ML - Microarray Gene Expression Markup Language

gDNA – Genomic DNA

DOC - Digital Optical Chemistry

CyDye – Cyanine Dye

HMG - High Mobility Group

HnRNP - Heterogeneous RiboNucleoproteins Particles

rDNA – ribosomal DNA

C/E – Condensation/Expression

DNMT- DNA MethylTransferase

TSA - Trichostatin A

5 Aza dC - 5-aza-2'-deoxycytidine

CpG – A dinucleotide with the sequence Cytidine Guanidine

MMNase – Micrococcal Nuclease

MAGE – Melanoma Antigen

MEME – A software package to discover motifs in biopolymers

TRANSFAC – A database of eukaryotic cis-acting regulatory elements and transacting factors

Q-PCR – Quantitative PCR

CQ-PCR – Chromatin Quantitative PCR

cT – Cycle Threshold

UTR – UnTranslated Region

SCLC – Small Cell Lung Cancer

NCI - National Cancer Institute

ER (+/-) – Estrogen Receptor positive or negative

CHAPTER ONE

Introduction

1.1 Introduction

The invention of nucleic acid microarray technology by Southern in 1988 and the subsequent development of expression microarrays using spotted polynucleotides by Shalon and Brown in 1995 signaled the start of a fundamental change in the way biomedical science is performed (Southern 1989; Schena, Shalon et al. 1995; Shalon 1997). The first way is through the increased number of survey-type experiments where the inherent hypothesis is that by systematically studying enough variations of the variable of interest (drug treatment, disease, etc) and gathering of huge amounts of data about these variations, the biological nature of the variable of interest can be determined.

The most traditional means of doing a survey-type experiment is by using expression microarrays to compare the gene expression state of the normal or wild type samples to that of diseased or treated samples (Schena, Shalon et al. 1995). By comparing normal samples to a variety of similarly altered samples, the stable “signal” can be separated from the variable “noise”, and the biological signature of the variable of interest isolated. An example of this is the expression survey of lung cancer cell lines done as a multi-lab study including the Minna Lab at UTSW to create a molecular pathology for lung cancer based on expression differences (Dobbin, Beer et al. 2005). The study compared normal lung epithelial cells to nearly a hundred different lung cancer model cell lines. This allowed them to determine a gene expression signature for the disease. More importantly, the flexibility of the dataset permitted it to be used repeatedly to pursue other questions about lung cancer such as *de novo* identification of genes conferring resistance to drug treatments (Dowell and Minna 2004).

While expression microarray technology is the most familiar to scientists, there are many alternative nucleic acid microarray platforms. The two platforms that are most relevant to this work are array-based comparative genomic hybridization (aCGH) and chromatin immunoprecipitation on microarray (ChIP on chip). aCGH measures chromosomal changes that are associated with diseases like cancer, where sections of chromosomes as large as several megabases or even whole chromosomes can be amplified or deleted. ChIP on chip is a method used to study proteins that bind to DNA and thereby comprise chromatin. Both microarray types will be discussed in detail in section 1.3 of this chapter. Moreover, the use of microarray technology has expanded so much that nearly any biological material can be arrayed, from small molecule libraries to tissue sections several millimeters in diameter (Belosludtsev, Bowerman et al. 2004; Matsuzaki, Loi et al. 2004; Tzankov, Went et al. 2005; Yamamoto, Clark et al. 2005). Once the samples are arrayed, they can be interrogated using much the same experimental approach employed in expression microarray protocols. Using these microarray-based and/or other survey methods (like mass spectrometry), the search for biosignatures or biomarkers is now common place (Carr, Rosenblatt et al. 2004). But several areas of biological research do not have a standard method by which surveys are preformed. One of the most important areas is direct chromatin state measurement, since the chromatin condensation state indicates transcriptional state of the region (Rauscher 2005). In the past, the measurement of chromatin state was either performed using high-resolution, yet low-throughput PCR based methods or low-resolution and high-throughput DNA staining methods (Banerjee and Hulten 1994; Aasland and Stewart

1999). These methods and chromatin in general will be discussed in section 1.2 of this chapter.

The rapid increase in the use of survey-style methods led to a second area of impact on how science is executed: the massive increase in data volume (Brazma, Robinson et al. 2000). In the past, slot blot and other traditional methods yielded datasets with hundreds or at most a few thousand data points (Shalon 1997). This data was largely analyzed manually and because the experimental model was hypothesis-driven, only a few of the data points needed to be considered to test the hypothesis. Now the need for statistical power means survey-type experiments must generate large amounts of data to isolate the key variable (Schena, Shalon et al. 1995). For example, in the previously discussed lung cancer survey, nearly a hundred different cell lines were used, each cell line was run at least in duplicate, and each microarray had approximately 40,000 features (genes) (Dobbin, Beer et al. 2005). This translates to over seven million data points generated by this survey alone. In order for the experimental model to identify the signature of the variable of interest, all the data must be analyzed. The large data volume means computers are required to supplement the work of individual scientists and probabilities replace absolute measurements.

1.2 The basics of chromatin

In 1882 Walther Flemming used the word chromatin for the first time when he described the results of experiments involving staining cells with aniline dyes (Flemming 1882). Chromatin is the name he gave the nuclear material that bound the dye strongly. Beyond its ability to bind basophilic dyes, chromatin is the means by which DNA

molecules that far exceed the length of the cell are packed into the nucleus (Li 1975). This packing (condensation) also serves other functions (figure 1.1), including limiting access between certain proteins and DNA, to controlling the transcription of genes, and allowing segregation of the chromosomes to occur during replication.

The traditional view of chromatin packing is that during interphase the chromatin (and therefore the DNA) is divided into two types. These consist of densely packed, and darkly staining heterochromatin which is not transcribed, and loosely packed, lightly staining, euchromatin which is transcriptionally available (Bender 2004). The inverse relationship between transcription and condensation is the central dogma of chromatin-based transcriptional regulation (Rose and Garrard 1984).

This simple model is, however, in conflict with some experimental evidence since a number of genes are known to be expression out of very dense heterochromatin (Taddei, Hediger et al. 2004; Dimitri, Corradini et al. 2005). While it is very possible that high resolution chromatin state studies will reconcile this by demonstrating that the genes “expressed out of heterochromatin” are really expressed from localized regions of euchromatin. However, since the definitions of eu- and hetero-chromatin conveys very precise biochemical meanings if the discrepancy cannot be reconciled the terms may be retired in favor of new descriptors. (Gilbert, Boyle et al. 2004; Dimitri, Corradini et al. 2005). Also, given the very precise meanings of eu- and hetero-chromatin I will refrain from using those terms, instead the chromatin state will simply be referred to as condensed or relaxed because these terms are descriptive and immune to any changes in nomenclature.

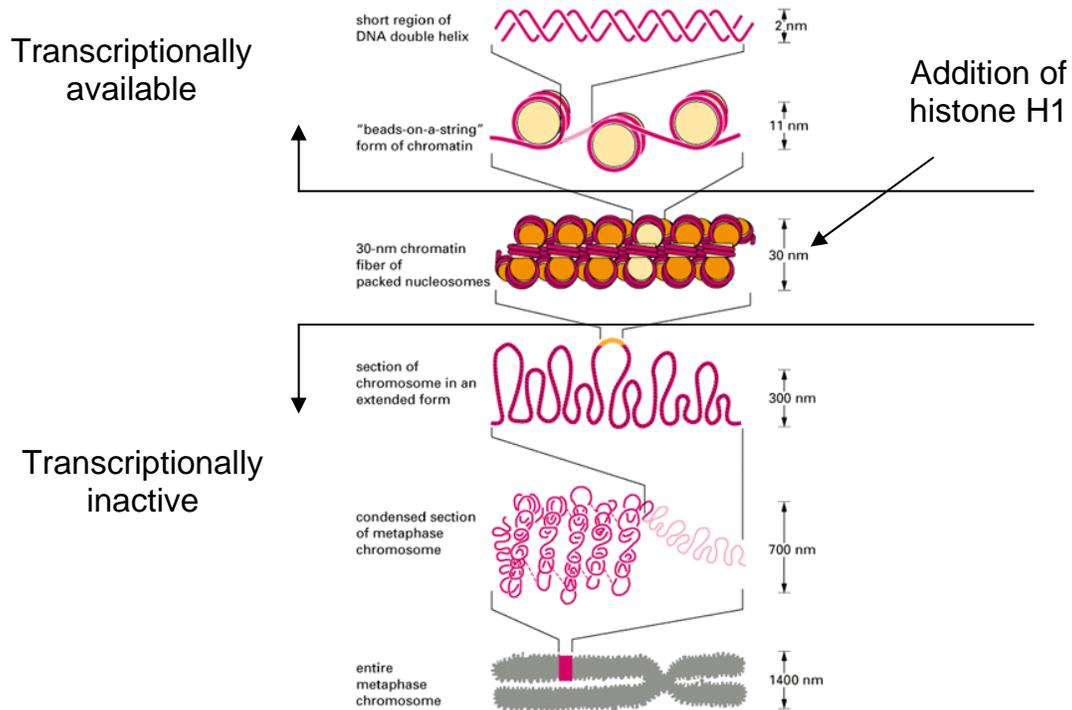


Figure 1.1: Chromatin condensation annotated with the transcriptional state of the various packing levels. Note the addition of the H1 or packing histone marks the transition to transcriptional suppression. Modified from: Alberts, et al, *Molecular Biology of the Cell*. 3rd ed. New York and London: Garland Publishing; 1994 (Alberts, Johnson et al. 1994)

1.2.1 The history of chromatin research: structural versus functional

Chromatin research, since the discovery of DNA, has been divided into lines of inquiry: structural and functional. Structural chromatin studies focus on the structure of the chromatin components and how these influence chromatin organization and packing. Functional chromatin studies focus on how chromatin organization influences the DNA's accessibility to protein binding which in turn regulates transcription. Modern chromatin research was born in 1976 when Camerini-Otero *et al.* discovered that the nucleosome is an octamer of 4 types of histones (figure 1.2) (Camerini-Otero, Sollner-Webb et al. 1976)

and Weintraub discovered that chromatin shows various levels of resistance to digestion based on its transcription state (Weintraub and Groudine 1976). By combining Weintraub's method of generating cleaved chromatin with Southern's method of transferring DNA from gels to membranes for probing (Southern and Mitchell 1971), DNA footprinting was possible (Wu, Wong et al. 1979). Using footprinting, it was possible to visualize the opening and closing of the chromatin, since active DNA is hypersensitive to cleavage as bands on the Southern blot that disappeared or appeared, respectively (Keene and Elgin 1981). The study of hypersensitive sites in active genes led to a better understanding of the nature of promoters and transcriptional regulation (Keene, Corces et al. 1981).

While footprinting is a very useful means to measure the chromatin state, condensation state is only a part of the information available from the chromatin. Chromatin is more than the four histone dimers in the core histone and the DNA: there is a fifth histone called H1, or the packing histone, and a variety of other accessory proteins (Bakayev, Melnickov et al. 1975). Chromatin regulation occurs through the interaction of these proteins with the core nucleosome, and the alterations these interactions modulate in the availability and the molecular environment of the DNA. Structural studies of chromatin accessory proteins investigated the protein's characteristics or used synthetic chromatin constructs to measure the molecular interactions between proteins and DNA. Functional studies extracted and identified the chromatin accessory proteins from living cells in an attempt to understand the biological implications of the interactions (Burckard, Mazen et al. 1975; Gadski and Chae 1976). Since the groups operated separately, the functional groups proved, for example, that the binding of the H1 histone to the core

nucleosome was linked to transcriptional suppression, while separately structural groups demonstrated that the binding of the H1 histone to the core nucleosome induced chromatin condensation (Li 1975; Varshavsky and Bakayev 1975). However, the consequences of these two results, that binding of the H1 histone to the core nucleosome is associated with condensed chromatin, and that the chromatin condensation suppresses transcription, was left uninterpreted for more than a year (Niedzwiecka and Kalinski 1977).

Around the same time that much of this work was being carried out, the discovery of DNA restriction endonucleases as an expedient means of genetic manipulation was beginning to over-shadow the role of chromatin in transcriptional regulation research (Rougeon, Kourilsky et al. 1975; Camerini-Otero, Sollner-Webb et al. 1976). As the new DNA-centric view of molecular biology gained favor, the deep rift between structural and functional chromatin researchers proved detrimental to advancing the understanding of chromatin function as a “function” of chromatin structure (Urnov and Wolffe 2001).

1.2.2 Modern chromatin research, the unification of structure and function

Fortunately, chromatin research did not cease as DNA-based methods gained popularity and strides continued to be made to unify structure with function. Of these advancements, the one that is of most direct interest to my efforts was the creation of a method called chromatin fractionation (Rose and Garrard 1984). While this method will be covered extensively in Chapter 3, a brief introduction is still warranted. The initial steps of chromatin fractionation are similar to footprinting, but after cleavage of the DNA by nucleases, instead of purifying the DNA, the varying solubilities of the chromatin and the bound proteins are exploited to separate the chromatin into three fractions that

correlated with the chromatin condensation state. Because this method separates the chromatin by its condensation state, and leaves chromatin associated proteins largely unperturbed, it is possible to study from the same original sample both the cleavage pattern of the DNA and the associated proteins, all in the context of the three fractions (Huang, Barnard et al. 1986). This means that structural and functional studies could be conducted from the same sample, greatly facilitating linking structure and function. However, this method does not crosslink the proteins to the DNA, so it was not possible to directly address what proteins, bound to what piece of DNA.

The discovery that really sparked the revival of interest in chromatin was the 1994 discovery by Wolffe of the Swi/Snf enzymes, which actively remodel chromatin (Wolffe 1994). This discovery was significant for two reasons. First, active chromatin regulation garnered the interest of researchers outside the chromatin field, since it allows chromatin to play an active role in transcriptional regulation. Second, the discovery provided a link between chromatin structure and function that was needed to unify the field. Structure and function were further unified by future advances in understanding histone structure and post-translational modification.

While it had been known since the early 1960's that chromatin was modified by histone acetylases and that hyper-acetylation of the histones is strongly associated with transcriptionally active chromatin (Allfrey, Faulkner et al. 1964), the means by which this modification altered the function was not well understood. For many years, the accepted model assumed electrostatic effects between the DNA and the histones, which were "altered" by the post-translational modifications. However this model could not be verified experimentally (Struhl 1998). In 1997, Luger *et. al.* solved the structure of the

complete nucleosome associated with DNA at a 2.8 Å resolution (figure 1.2) (Luger, Mader et al. 1997). The knowledge gained from the structure suggested that the histone tails had dual functions. The H2A and H2B histone tails were bound largely in the minor groove to stabilize the histone/DNA interaction leaving the tails of histones H3 and H4 free in solution possibly to interact with chromatin binding proteins (not shown in figure 1.2 because they are disordered in the crystal) (Kayne, Kim et al. 1988; Nakatani 2001). Proof that the structure and function of the nucleosome were inseparable hastened the progress in the field and a coherent picture of chromatin regulation began to emerge.

With the structure of the nucleosome solved, it became much simpler to study the functional effects of structural changes caused by histone post-translational modifications such as acetylation. The understanding of histone post-translational modification expanded rapidly. By 2001, when the method to study these interactions in context of the DNA was finally patented (ChIP on chip), more than six different classes of modifications were known, each with a novel function (Wyrick 2001). In 2001 Jenuwein and Allis proposed the “Histone Code” which explained how a number of the modifications interact to alter the chromatin state and, therefore, transcriptional regulation (Jenuwein and Allis 2001). This work has been expanded upon greatly and histone post-translational modification has been linked to processes ranging from cell cycle control to apoptosis (Fu, Wang et al. 2004; Wang, Fu et al. 2004).

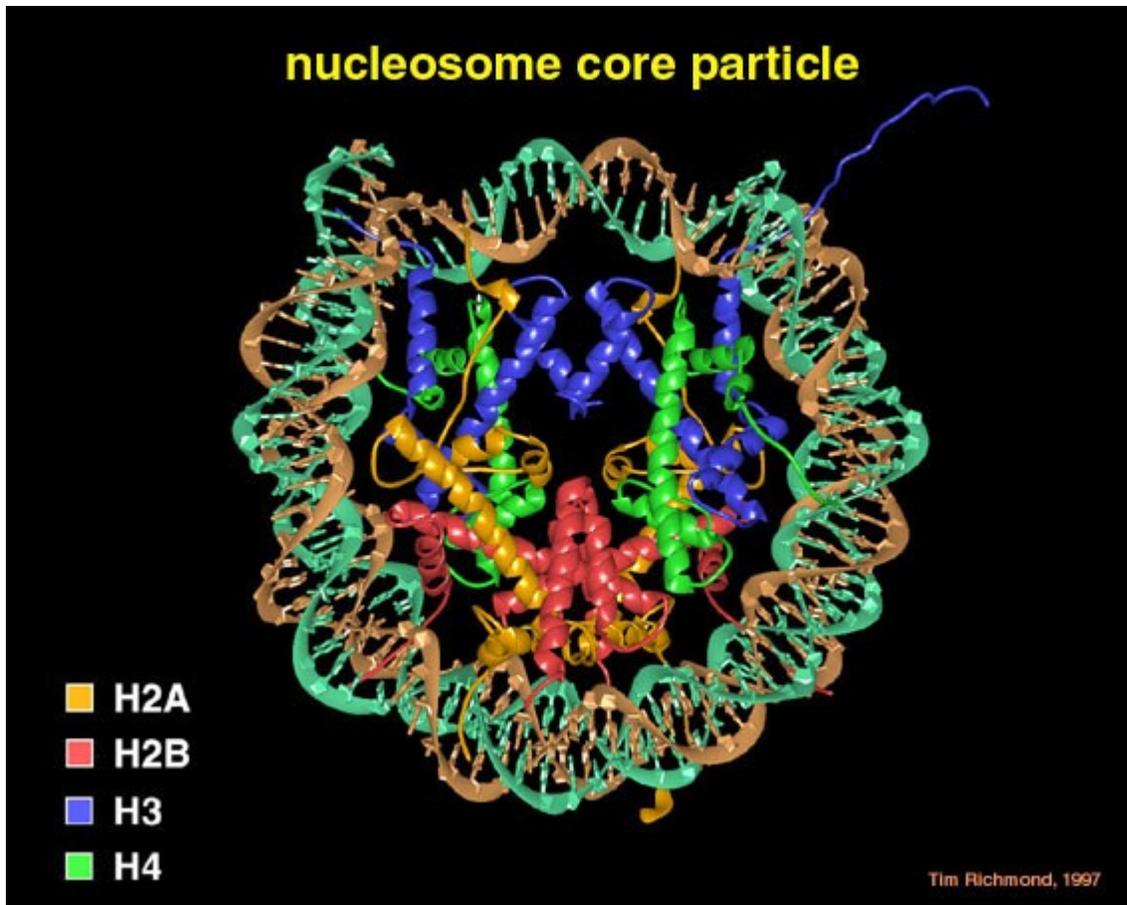


Figure 1.2: The 2.8 Å crystal structure of the nucleosome bound to DNA. This figure was kindly provided by Dr. Richmond.

1.2.3 Chromatin Immunoprecipitation

A method used to probe the proteins bound to the DNA component of chromatin is called Chromatin Immunoprecipitation (ChIP). ChIP is performed by reversibly cross-linking the proteins to the DNA in living cells. The cells are then lysed and the DNA cleaved by enzymatic digestion to uniform length. The lysate is run through a column that contains antibodies against to a protein of interest, thereby retaining the protein and the bound DNA. The column is eluted and the crosslink reversed by mild heat treatment, freeing the DNA. The identity of the resulting DNA fragments, are read out

using either PCR-based methods or Southern blotting (Wyrick 2001; Das, Ramachandran et al. 2004).

While ChIP-and a high-throughput variation of ChIP called ChIP on chip, which will be discussed in 1.3.6, facilitated the study of DNA-protein interactions, each method's unable to directly measure the chromatin condensation state. Accordingly, chromatin condensation studies are still performed using hypersensitivity mapping or chromatin fractionation and read out with Southern Blots or PCR-based methods, in much the same way they have been done since the late 1980's (Weintraub and Groudine 1976; Huang, Barnard et al. 1986; Aasland and Stewart 1999). These methods yield high-resolution views of the chromatin state however, they are very low-throughput since each probe or primer pair used for reading out a region requires a separate reaction.

Conversely, existing methods to measure the chromatin state in a high-throughput fashion are very low-resolution since they use chromatin staining to measure an average chromatin condensation state of the whole genome (Banerjee and Hulten 1994). Full modernization of chromatin condensation state research requires a new high-throughput/high-resolution method to complement ChIP.

1.3 Microarray technology

1.3.1 A history of transcriptomics

RNA, specifically messenger RNA (mRNA), was discovered in 1956 by Elliot "Ken" Volkin and Lazarus Astrachan at Oak Ridge National Laboratory (Volkin and Astrachan 1956). In 1958 Francis Crick published the central dogma of molecular biology: Information moves from DNA to RNA to protein, which spawned an increased interest in

RNA, since it was a central player in the genetic “information pathway” (Crick 1958). On Good Friday in 1960 Jacob, Brenner, and Crick sat down together at a conference and shared what each knew about RNA and came to the realization that RNA was the messenger of genetic information (Meselson, Nomura et al. 1964). With the help of Meselson and his work in ribosome subunit research, by the summer of 1960 they had proven that RNA was the genetic messenger (Meselson, Nomura et al. 1964).

The study of RNA was confined mostly to structural and viral models until 1970, when Temin and Baltimore co-discovered reverse transcriptase (RT) enzymes while studying how RNA-based viruses function (Baltimore 1970; Temin and Mizutani 1970). This class of enzymes synthesizes a DNA strand from a RNA template, which not only shattered the Central Dogma, but opened the way for modern transcriptional studies since unstable RNA could now be converted to stable DNA. In 1973, Shimada *et al.* demonstrated a method to separate cellular RNA using gel electrophoresis, so that it was possible to separate RNA by size and therefore to observe changes in “transcription levels” of a particular band (message) (Shimada, Sekikawa et al. 1973).

By the late 1970’s, an RNA version of Southern’s Blot (Southern and Mitchell 1971) called the Northern Blot was in common use (Bonitz, Coruzzi et al. 1980). Soon after, blot technology and reverse transcription were combined to create spotted membrane arrays (Kafatos, Jones et al. 1979). The difference between a Northern blot and this method was that the probes were bound to the membrane and the sample was reverse transcribed (RT) into complementary DNA (cDNA) and labeled, so that multiple genes could be probed simultaneously from a single sample (Dobner, Kawasaki et al. 1981). However, the method, also called “dot blotting” or “slot blotting” depending on how the

blot was produced, was completely manual, making it difficult, costly, and irreproducible. Because of these problems, Northern Blots and RT coupled PCR remained the standard for determining the level of a transcript in a sample (Shalon 1997).

1.3.2 A history of microarray technology

Who actually invented the microarray is a subject of great debate, since there is evidence from the early 1980's that semiconductor researchers had experimented with arraying and/or embedding biomolecules or whole cells onto silicon-based substrates (Kucheria 1986). However, Edwin Southern's patent (#EP0373203) "Method and apparatus for analyzing polynucleotide sequences" which was filed with the European Patent Office in 1989, marks the accepted invention date and inventor of microarray technology (Southern 1989; Dickson 2000). To paraphrase the major claims of the patent, Southern describes a means to use oligonucleotide probes adhered in an array, to a glass slide, or other substrate; to analyze the sequence of polynucleotides of known or unknown sequence (the target). This was accomplished by labeling the polynucleotides to be analyzed and applying them to the array under hybridizing conditions, and then measuring the amount of hybridization (Southern 1989).

The microarray is so named because it is an array (or grid) of microscopic spots of probe (Southern 1989; Schena, Shalon et al. 1995). However, Southern's early arrays were not like the microarrays commonly used today, since the probes were grown *in situ* on macroscopic areas of the slide and the arrays were used for resequencing in order to find mutations in the target sequences. (Southern, Maskos et al. 1992). The word "microarray" was coined by Schena *et al.* in their 1995 paper entitled "Quantitative monitoring of gene expression patterns with a complementary DNA microarray"

(Schena, Shalon et al. 1995). In addition to a name for the method, this paper also discussed two major advancements in microarray technology, since the microarray preparation was completely different from Southern's method and it was the first time anyone had published on using microarrays to study gene expression patterns.

The Southern method to produce arrays used *in situ* synthesis of the oligonucleotides from monomers, which at the time was an extremely expensive way to make DNA (>\$1 per base). The arrays were made one at a time, so the per microarray cost was thousands of dollars. Conversely, the method patented by Shalon and Brown used a capillary dispenser to deposit a known volume of a reagent, at a known position on the substrate, thus printing or spotting the microarray. The density of printed or spotted microarrays is greater than a thousand spots per square centimeter, which made their microarrays high density (with spot center to center distances from 50-200 microns) (Shalon 1997). Using robotic spotting it was possible to manufacture hundreds of replicate arrays, with thousands of unique spots using less than a nanoliter of probe per spot. The probe of choice in the early years was cDNAs, since they could be cloned into a vector, grown up in bacteria, sequence verified and then PCR amplified for spotting (Schena, Shalon et al. 1995; Duggan, Bittner et al. 1999) thereby allowing entire collections of clone libraries to be made and sold commercially as frozen bacteria cultures. By using these cDNA libraries it was possible for numerous institutions to produce very similar microarray probe sets, and keep the per microarray cost at around 100 to 200 dollars (Brown and Botstein 1999; Duggan, Bittner et al. 1999).

Technical problems plagued microarray technology in the early years with the most prominent being the choice of substrate and spotting buffer (Wrobel, Schlingemann

et al. 2003). In some facilities DMSO-based spotting buffers worked very well, but in others the spots did not dry fast enough and merged with neighboring spots. An SSC-based spotting buffer produced small, uniform spots in some labs, while in others it evaporated too fast from the pin and slides that were spotted towards the end of the cycle were “dry tapped” (i.e. no solution was deposited) since the pin was dry when it touched the substrate. While no solution has been found to date, the local humidity of the area seems to be the key, with dry areas favoring DMSO buffer and humid areas favoring SSC. Even seasonal differences matter, since in the winter the air is much drier (McQuain, Seale et al. 2003).

The issue of substrate remains controversial. In the early years most microarrays were printed on poly-L-lysine-coated slides since the strong negative charge of the DNA was attracted to the positively charged surface. While these slides cost less than a dollar to produce, many felt that the non-covalent interaction with the substrate and that the DNA being parallel to the substrate surface made the hybridization sub-optimal. As cost of synthetic oligonucleotides decreased, many groups switched from cDNA to oligonucleotide probes. These oligonucleotides could include various chemically reactive groups to facilitate covalent linkages of the probe to a chemically active surface. Aldehyde and hydroxyl-modified surfaces had problems as well, and given the increased cost of the substrate and probe modification, they never gained wide acceptance (Dufva, Petronis et al. 2004). Recently, lysine coating has been replaced with amino-silylate-modified surface since the coating is more uniform, and covalently attached to the substrate, but still binds the DNA electrostatically (Franssen-van Hal, Vorst et al. 2002). Thus, the optimal surface chemistry of microarrays remains a matter of opinion.

1.3.3 Spotted versus light-directed synthesis

While spotting is the most common non-commercial means of making a microarray, it is not the only means. In fact it was not even the first method, since the original Southern arrays used *in situ* synthesis to grow the octomeric probes from base monomers (Southern 1989). In 1989 a method to use mask-based light directed synthesis to create synthetic oligonucleotide probes on a glass substrate at high density was created by McGall, Fodor, and Sheldon (McGall 1992). The technology borrowed heavily from semiconductor manufacturing since photolithography uses quartz masks with transparent areas at specific regions so that the positive photoresist is made soluble by illumination, allowing the area to be etched. Conversely areas that are not illuminated during that step are coated in metal, and thus the region is masked. Using light directed synthesis (illustrated in figure 1.3) the probes are grown stepwise *in situ* by shining ultraviolet light through transparent regions in the mask onto the regions where a base monomer is to be added while the mask shades the rest of the substrate. The light causes the chemical protecting groups on the surface to be cleaved and the base monomer can then couple to the reactive surface. The uncoupled monomers are washed off and the process is repeated until the synthesis is complete (the oligonucleotides produced by this method are generally between 21-25 bases) (Lockhart, Dong et al. 1996) While the method is very reproducible and can achieve more than 10 times the feature density of spotted arrays, its cost is quite high since currently a set of lithographic masks costs \$50,000 and lithographic plant represents a capital investment in excess of a billion dollars (Barone, Beecher et al. 2001).

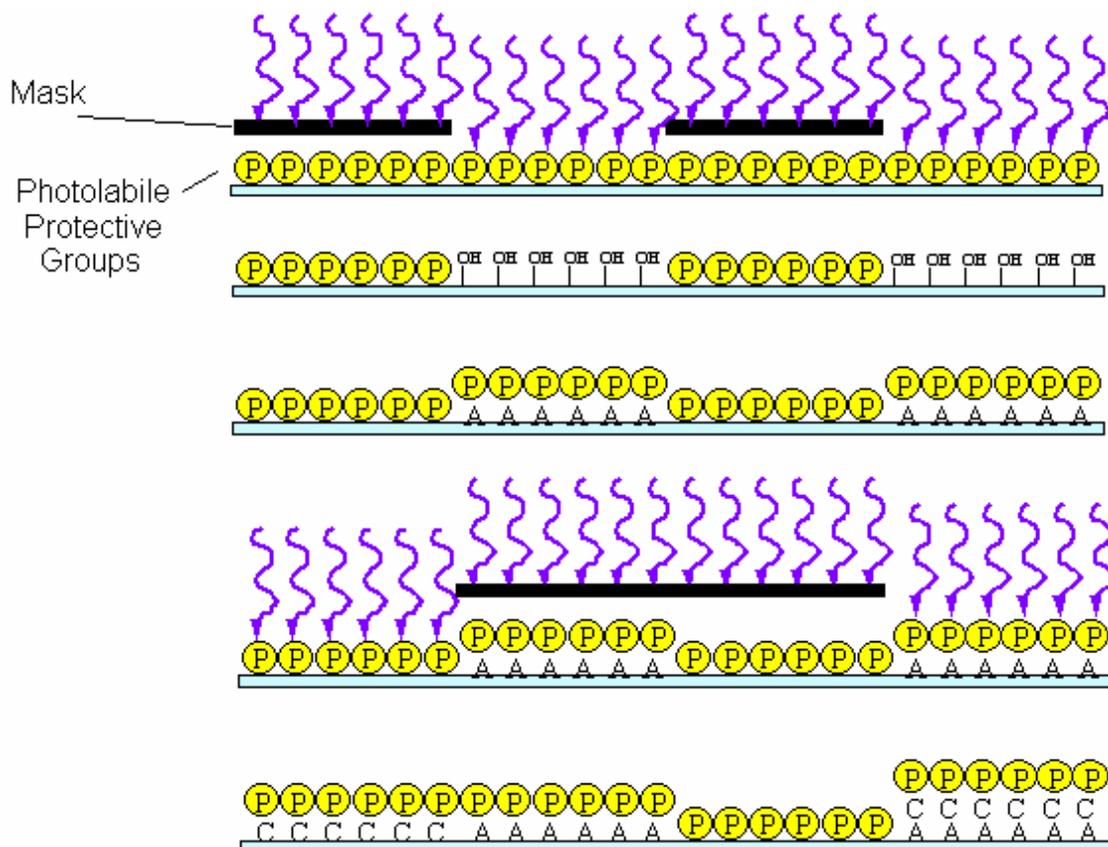


Figure 1.3: A diagram of mask-based lithographic oligonucleotide synthesis, showing two cycles to illustrate the how the process can be optimized to reduce the number of addition steps. The UV light represented by the wavy blue lines shines on the substrate through the mask. In the illuminated areas the protecting group is cleaved exposing, reactive hydroxyls, which can be coupled to base monomers. After washing to remove uncoupled monomers, the process is repeated, until the oligonucleotides are completely synthesized. This figure was kindly provided by Dr. Luebke.

In 1999, Harold Garner filed a patent on a means to use the Digital Micromirror Device (DMD) technology from Texas Instruments for maskless lithographic light directed oligonucleotide synthesis (Garner 2001). The DMD which is commonly used in video projectors has approximately one million 17 micron mirrors that can be individually addressed. As shown in figure 1.4 the mask is replaced by an image file that when projected on the surface of the substrate illuminates certain areas, and leaves others dark.

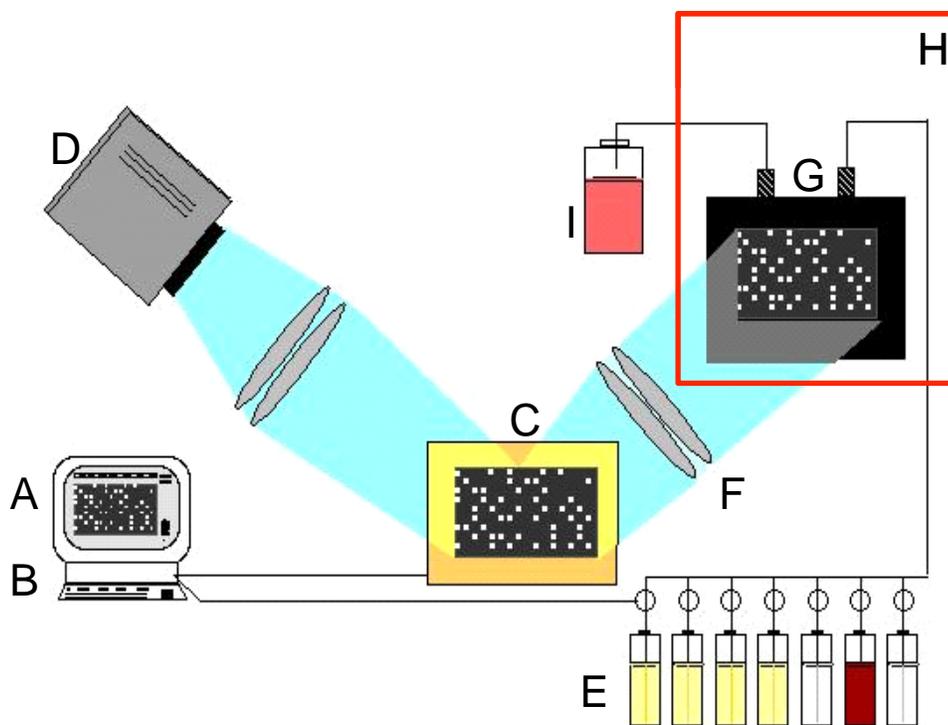


Figure 1.4: A schematic of a maskless lithographic synthesizer A: Control PC. B: DLP, valve and microwave controller interface. C: DLP on which computer generated pattern is placed. D: Mercury UV lamp. E: Monomer and solvent reservoirs. F: 6-element imaging optics. G: Array synthesized on a microscope slide held in a sealed liquid holder on which pattern of UV light is projected. H: A light tight housing containing a solenoid shutter under computer control (open during UV illumination, closed during monomer coupling). I: Waste. This figure was kindly provided by Dr. Garner

While this technology cannot match the productivity of traditional lithography-based synthesis, the equipment to produce the maskless arrays has a very low capital cost, so it is well suited to custom “discovery-guided” microarray work. This is especially true because the array is produced by digital masks, so each chip can be different from the previous with little additional cost (Stengele, Buhler et al. 2005).

1.3.4 Expression microarray methodology

Regardless of the technology used to produce microarrays, the vast majority are used for expression studies (Duyk 2002). There are two major variations of expression microarray technology. First there are experiments where a single sample is hybridized to the microarray or “one color microarray”. Alternately, two samples can be simultaneously hybridized to the microarray which is called “two color microarray”. Each type has its own advantages and disadvantages, so when a microarray-based experiment is designed, the selection is normally a function of the user’s preferences, budget and sample availability. In general, spotted microarrays are run as two color experiments to compensate for spot to spot variability, and *in situ* synthesized microarrays as one color because they are manufactured with tighter quality controls and the spots or features are more uniform. This is not a hard and fast rule, so both types deserve further consideration.

The basic scheme of a two color microarray experiment (figure 1.5) consists of a control or untreated sample (target 1) labeled with one color of fluorophore and the experimental sample (target 2) being labeled with another fluorophore. Both targets are competitively hybridized to the microarray which has any number of probes (a.k.a. spots, or features). After hybridization, the microarrays are washed to remove unbound target and scanned to measure the relative intensity of each of the fluorophores bound to the different probes. Each fluorophore is measured using a different excitation and emission spectrum (i.e. channel) so a single microarray generates two images. After scanning, the features are identified to define the area of the spot, and then the data quality of each spot is individually assessed in a process called flagging. Flagging is necessary since washing

artifacts or printing problems during the creation of the microarray can alter the intensity of spots in ways that are not related to biological processes of interest, and so this data must be eliminated before the data can be analyzed. Following gridding and flagging, data is extracted from the microarray. Extraction takes an average pixel intensity of the spot as defined by the grid, and for some experiments the background value is also extracted by taking an average pixel intensity of a defined region centered on the spot. Following extraction, if the targets are analyzed separately, the intensity of the features gives a semi-quantitative measure of the expression level of each transcript, which can be referenced to other genes to estimate transcript copy number. By using a ratio of the two targets (channels), the difference between conditions can be determined semi-quantitatively, and expressed as a fold change (figure 1.6).

For a one color microarray the process is similar to that described above, but each sample is run on a separate microarray. The advantage is that the control sample is run only once instead of being run on each microarray, which can reduce sample cost significantly. Yet the disadvantage is that without a control sample as a reference on each array, comparing arrays can be difficult unless high quality arrays are used.

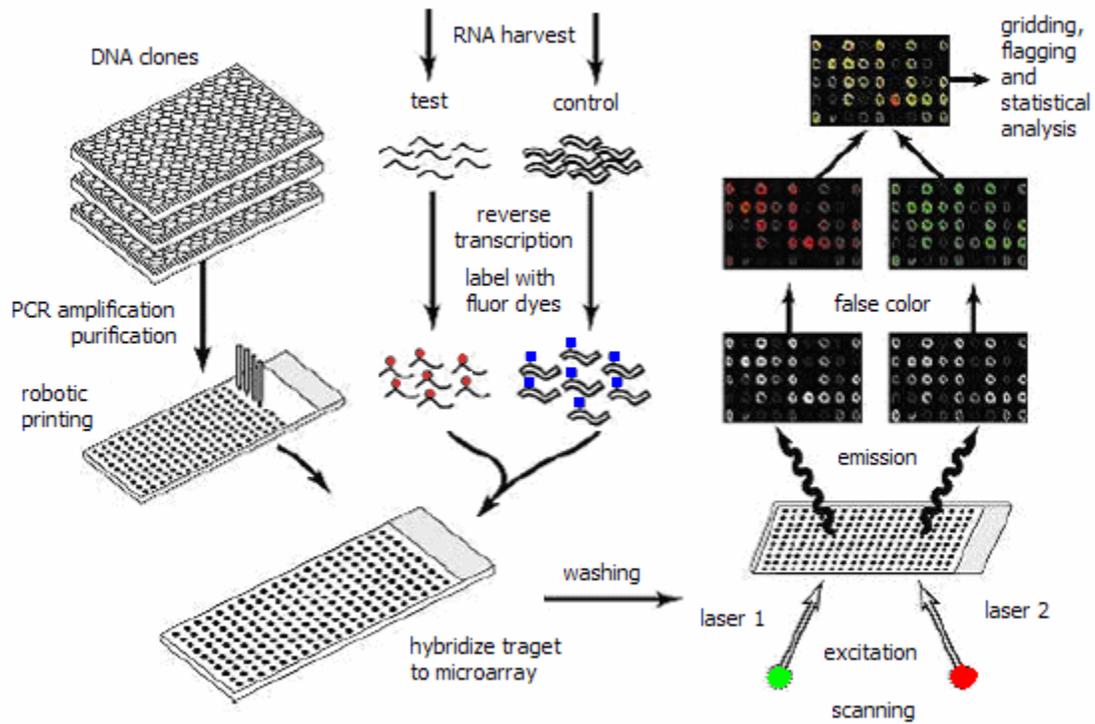


Figure 1.5: A simplified depiction of the microarray work flow. The microarrays are robotically spotted, using DNA clones or oligonucleotides. The nucleic acid samples are extracted, and labeled with fluors then cohybridized to the microarray. After hybridization the microarray is washed and scanned. The two images are then over laid so the image can be analyzed and the data can be extracted. Modified from Duggan, D.J., et. al. (Duggan, Bittner et al. 1999)

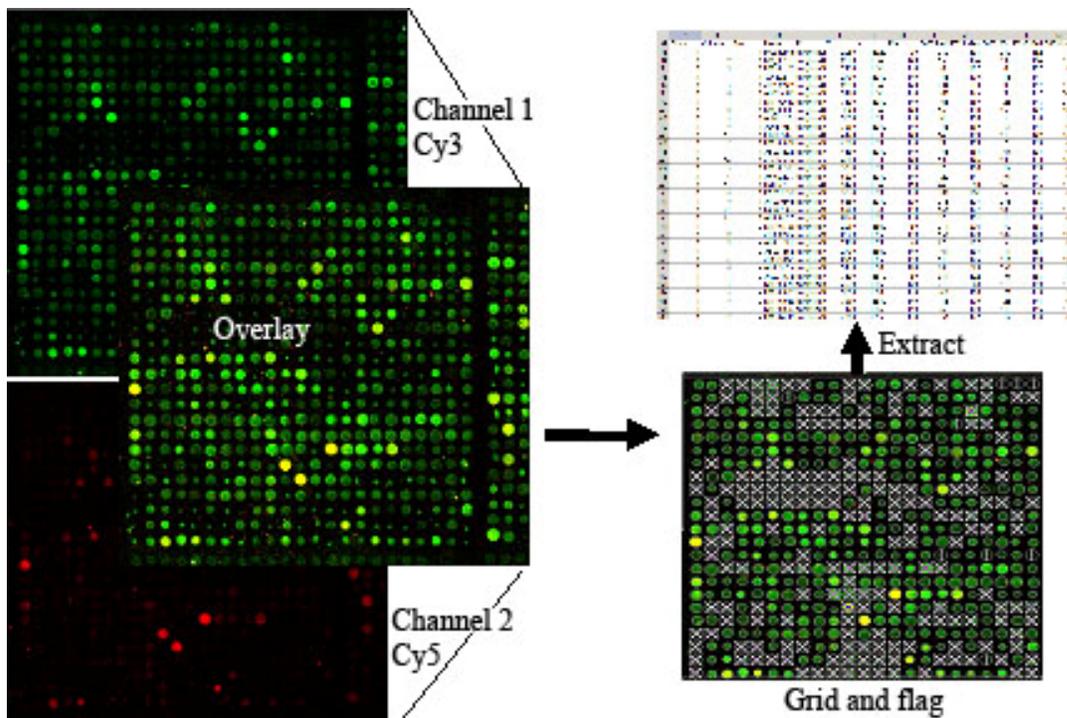


Figure 1.6: The data extraction process for a typical spotted microarray. The panels in the figure show 1 of 48 blocks on the microarray. After the signal between the two channels is balanced during scanning, the two channels (left bottom to top) Cy5 experimental and Cy3 control) are overlaid (left middle). Then a grid is placed on the microarray to measure identify the features, then the microarray is manually flagged to classify the spots as good (“O”), bad (“X”) or marginal (“|”) (bottom right). The data is then extracted as a spreadsheet file that records the intensities for each spot in both channels and a multitude of other parameters.

When expression microarray technology was in its earliest stages of development, (i.e. before the commercialization of Affymetrix GeneChips®) researchers were primarily concerned with getting the technology to work. Early spotting robots were unpredictable and the array hybridization process itself required a great deal of optimization. Most of the early experiments involved verifying the expression of a few hundred genes with known expression values in a “high-throughput fashion” (Schena, Shalon et al. 1995). However, since a microarray can include many probes, it was not long before the primary use of microarray technology was hypothesis-generation and

discovery biology (Brown and Botstein 1999). Researchers began to use microarray studies to find signatures or changes in transcription levels that correlated with an experimental variable.

In 2006 the entire known human transcriptome can be interrogated using a single microarray (Hardiman 2004). Commercially produced microarrays from Affymetrix or bead-based arrays from Illumina have reduced the need for spotted expression microarrays (Lockhart, Dong et al. 1996; Kuhn, Baker et al. 2004). However, since spotted arrays are relatively easy to make, new types of microarrays are being created to address novel biological lines of enquiry in a high-throughput manner.

There are now microarrays for myriad applications, from resequencing individual bases for finding single nucleotide polymorphisms (SNPs) to biosignature arrays that measure the unique hybridization characteristics of an organism for identification purposes (Belosludtsev, Bowerman et al. 2004; Matsuzaki, Loi et al. 2004). Even what material is spotted on the array is no longer limited to nucleic acids; biomolecules from serum, antibodies, small molecules, and even tissues are spotted on microarrays (Tzankov, Went et al. 2005; Yamamoto, Clark et al. 2005). The alternate microarray platforms of greatest relevance to my work are array-based Comparative Genomic Hybridization or aCGH (Forozan, Karhu et al. 1997; Pollack, Perou et al. 1999) and the Chromatin ImmunoPrecipitation microarray or ChIP on chip (Baetz, Moffat et al. 2001).

1.3.5 Microarray-based comparative genomic hybridization

It has long been known that there exists a strong association between cancer and loss of heterozygosity (LOH) or other chromosomal abnormalities (Girard, Zochbauer-Muller et al. 2000). The traditional method to study these processes is Fluorescent *In Situ*

Hybridization (FISH) (Cheung, Tishler et al. 1977). However, FISH is very low throughput since the readout of the experiment is microscopic visualization of numerous metaphase chromosomes interrogated with very few probes or a single probe. In order to study these chromosomal changes *en masse* Kallioniemi *et al.* created comparative genomic hybridization (CGH) which compares an experimental genome labeled with one fluorophore to a reference genome (normal karyotype) labeled with a second fluorophore via simultaneous hybridization to a normal chromosomal spread (Kallioniemi, Kallioniemi et al. 1992). The differences between hybridization levels of the two genomes can then be used to identify regions of amplification or deletion associated with the disease. However, this method sacrifices the ability to associate a specific sequence with the differential regions in cases where the chromosomes are rearranged. Microarray-based approaches are well suited to this technique and could provide the resolution of FISH and the throughput of CGH.

Microarray-based CGH (aCGH) works similarly to a standard microarray, only with cells of unknown karyotype labeled in one color and a cell with a normal karyotype used as a reference in the other color (Pollack, Perou et al. 1999). While this is a very straightforward technique and the application seems obvious, there was an almost year long delay between the suggestion and publication of aCGH for humans (Forozan, Karhu et al. 1997; Pinkel, Segev et al. 1998). This delay was a consequence of the need to design a new type of probe for the arrays. At the time, probes for transcription microarrays were cDNAs, which were not believed to be capable of binding the long nick-translated fragments of low complexity human genomic DNA with the required sensitivity and specificity. Since the standard for CGH was megabase resolution,

microarrays for aCGH were spotted with the same large clones (Bacterial or Yeast Artificial Chromosomes, BACs and YACs respectively) used for standard CGH (Albertson, Collins et al. 2003). However, before they could be used, it was necessary to validate that these clones were genomically unique, complete and free of cloning errors caused by recombination during the creation of the clones, or the inclusion of host DNA (Forozan, Mahlamaki et al. 2000). Despite these issues, the use of aCGH has grown rapidly and has greatly increased our understanding of how chromosomal abnormalities contribute to disease (Snijders, Pinkel et al. 2003; Vissers, Veltman et al. 2005).

1.3.6 Chromatin immunoprecipitation on microarrays

From the chromatin perspective, ChIP on chip was an evolutionary step from low throughput ChIP protocols to high throughput readout, using the microarray technologies used for aCGH (Robyr, Suka et al. 2002). As shown in figure 1.7, ChIP on chip is similar to ChIP methods, except the recovered DNA is labeled and hybridized to a microarray instead of being probed by Southern blotting.

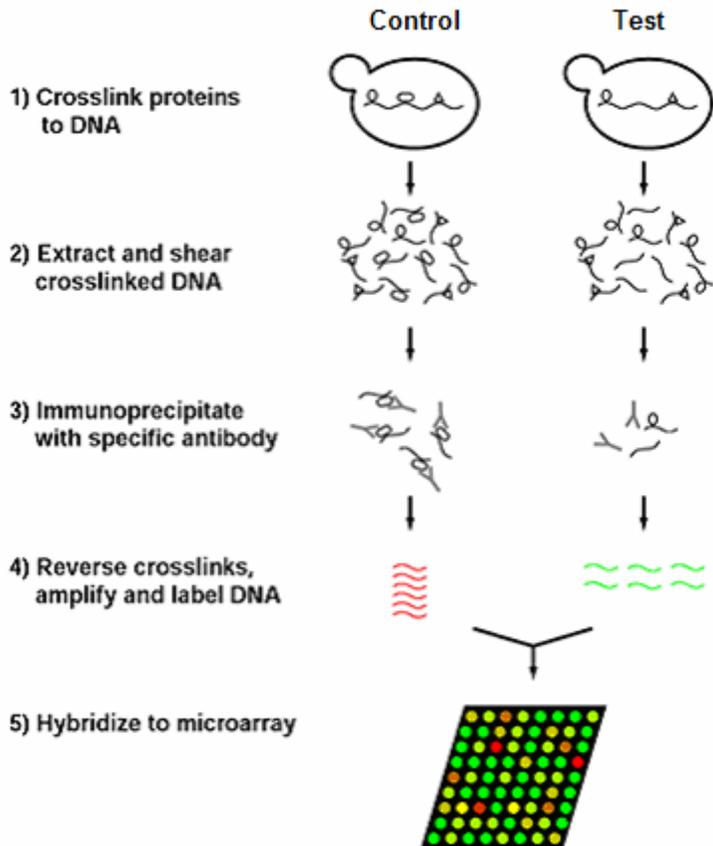


Figure 1.7: A diagram of the Chromatin Immunoprecipitation microarray work flow. ChIP on chip and ChIP share the same initial steps since the DNA and protein must be cross-linked, the DNA sheared, the protein of interest is then immunoprecipitated with an antibody, and then the cross-linking reversed. Once the DNA is free in solution the two methods differ since for regular ChIP PCR or blotting methods are used for the readout, and for ChIP on chip the DNA is labeled and hybridized to a microarray.

While ChIP on chip can probe far more regions than could be studied using Southern blotting, there were complications related to probe design. Since the creation of cDNA from RNA is 3' biased, the microarrays designed for expression use probes that are biased to the 3' end of genes (unless they are for organisms with small genomes like yeast which are simply tiled across the genome). While aCGH has a similar problem, the resolution required for an aCGH experiment is at the gene (kilobase) level or greater, so

expression microarrays are satisfactory for this purpose (Albertson, Collins et al. 2003). However, ChIP on chip produces DNA targets that requires sub-gene resolution, and since these experiments are often focused on promoter regions is strongly 5' biased (Das, Ramachandran et al. 2004). This means that microarrays for ChIP on chip often have to be custom made for the application, and are not compatible with traditional expression studies. Because of the cost and expertise required to design and make a new kind of microarray, the technique is limited to labs with the resources to developed these microarrays and the focus to use the resource for sufficient duration to justify the investment.

As of 2006, companies like Nimblegen and Affymetrix are beginning to remedy ChIP on chips probe design problem with the introduction of tiling arrays for the mouse and human genomes (Bertone, Trifonov et al. 2005). Tiling arrays use either overlapping probes to give complete coverage of the genomic region of interest, or use probes spaced at fixed intervals across the genome so an entire genome may be covered within the limits of the microarrays probe density. Even with these limitations, since it is now possible to study protein/DNA interactions across the whole genome, one interaction at a time, ChIP on chip has uncovered a wealth of knowledge about nuclear regulation. Recent ChIP on chip studies for example have decribed the global nucleosome modification state in yeast (Pokholok, Harbison et al. 2005) as well as mapped unknown but active promoters in humans in order to better understand expression patterning (Kim, Barrera et al. 2005). However, these studies do not address how chromatin condensation state contributes to the regulation of chromatin/protein interactions.

Traditional high resolution chromatin state measurement techniques are too resource-intensive to follow up on every relationship discovered as part of a ChIP on chip survey. So, without a complementary microarray platform to measure the chromatin condensation state, a great deal of information gathered by ChIP on chip experiments will be missed or misinterpreted. For example, in a ChIP on chip survey, if a promoter element is found to be insufficient for transcriptional activation of a particular gene for a particular set of conditions based solely on ChIP results, it is unlikely this result will be investigated further. However, knowing the condensation state of the element could indicate that the element could be sufficient but only if a required transcription factor can access it (Renaud, Loukinov et al. 2005). So, under other conditions that relax the chromatin containing the DNA element, the element might be able to activate transcription. Thus, the combination of ChIP on chip and chromatin state data might be used to more fully understand the transcriptional regulation of the gene. A microarray technique that measured the chromatin condensation state would not just be a companion to ChIP on chip, as chromatin condensation state data alone would address fundamental questions about transcriptional control and chromatin architecture. Thus, creation of a novel microarray technique to measure the chromatin condensation state is a worthwhile endeavor.

1.4 Thesis and organization

The remainder of my dissertation will describe experiments designed to develop methods for high resolution aCGH and the creation of a novel type of a microarray technique to directly measure the chromatin state at a subgene resolution. The exploitation of this technique for application to a cancer model system at baseline and

after drug-induced epigenetic modification will allow a cell's chromatin architecture to be linked to transcriptional control and, therefore, information about the disease and the drug can be extracted.

Chapter 2 details the development of a method to hybridize low complexity genomic DNA to probes designed for use in expression microarray experiments for the purpose of increasing the resolution of aCGH microarray technology. The focus of Chapter 3 is of the chromatin array, a novel type of microarray that measures the condensation state of the chromatin. Knowledge of the condensation state of the chromatin provides an additional way to understand transcriptional control even if the region is not transcribed. Chapter 4 describes the exploitation of the chromatin array to investigate chromatin state alterations induced by treatment with an epigenetic modifying drug on a tumor model cell line. By comparing treated and untreated chromatin state and expression data, it is possible to make discoveries that are potentially relevant to understanding the cancer and the mechanism of action of the drug. Finally, Chapter 5 offers ideas on how to expand the use of the chromatin array to understand transcriptional control and how the chromatin state is related to disease processes.

1.5 References:

- Aasland, R. and A. F. Stewart (1999). "Analysis of DNaseI hypersensitive sites in chromatin by cleavage in permeabilized cells." *Methods Mol Biol* **119**: 355-62.
- Alberts, B., A. Johnson, et al. (1994). *Molecular Biology of the Cell*. New York: 352.
- Albertson, D. G., C. Collins, et al. (2003). "Chromosome aberrations in solid tumors." *Nat Genet* **34**(4): 369-76.
- Allfrey, V. G., R. Faulkner, et al. (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." *Proc Natl Acad Sci U S A* **51**: 786-94.

- Baetz, K., J. Moffat, et al. (2001). "Transcriptional coregulation by the cell integrity mitogen-activated protein kinase Slt2 and the cell cycle regulator Swi4." Mol Cell Biol **21**(19): 6515-28.
- Bakayev, V. V., A. A. Melnickov, et al. (1975). "Studies on chromatin. II. Isolation and characterization of chromatin subunits." Nucleic Acids Res **2**(8): 1401-19.
- Baltimore, D. (1970). "RNA-dependent DNA polymerase in virions of RNA tumour viruses." Nature **226**(5252): 1209-11.
- Banerjee, S. and M. A. Hulten (1994). "Sperm nuclear chromatin transformations in somatic cell-free extracts." Mol Reprod Dev **37**(3): 305-17.
- Barone, A. D., J. E. Beecher, et al. (2001). "Photolithographic synthesis of high-density oligonucleotide probe arrays." Nucleosides Nucleotides Nucleic Acids **20**(4-7): 525-31.
- Belosludtsev, Y. Y., D. Bowerman, et al. (2004). "Organism identification using a genome sequence-independent universal microarray probe set." Biotechniques **37**(4): 654-8, 660.
- Bender, J. (2004). "Chromatin-based silencing mechanisms." Curr Opin Plant Biol **7**(5): 521-6.
- Bertone, P., V. Trifonov, et al. (2005). "Design optimization methods for genomic DNA tiling arrays." Genome Res.
- Bonitz, S. G., G. Coruzzi, et al. (1980). "Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase." J Biol Chem **255**(24): 11927-41.
- Brazma, A., A. Robinson, et al. (2000). "One-stop shop for microarray data." Nature **403**(6771): 699-700.
- Brown, P. O. and D. Botstein (1999). "Exploring the new world of the genome with DNA microarrays." Nat Genet **21**(1 Suppl): 33-7.
- Burckard, J., A. Mazen, et al. (1975). "Non-histone chromosomal proteins easily extractible from chick erythrocytes." Biochim Biophys Acta **405**(2): 434-41.
- Camerini-Otero, R. D., B. Sollner-Webb, et al. (1976). "The organization of histones and DNA in chromatin: evidence for an arginine-rich histone kernel." Cell **8**(3): 333-47.
- Carr, K. M., K. Rosenblatt, et al. (2004). "Genomic and proteomic approaches for studying human cancer: prospects for true patient-tailored therapy." Hum Genomics **1**(2): 134-40.
- Cheung, S. W., P. V. Tishler, et al. (1977). "Gene mapping by fluorescent in situ hybridization." Cell Biol Int Rep **1**(3): 255-62.
- Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-63.
- Das, P. M., K. Ramachandran, et al. (2004). "Chromatin immunoprecipitation assay." Biotechniques **37**(6): 961-9.
- Dickson, D. (2000). "Affymetrix loses first round of patent battle." Nature **404**(6779): 697.
- Dimitri, P., N. Corradini, et al. (2005). "The paradox of functional heterochromatin." Bioessays **27**(1): 29-41.
- Dobbin, K. K., D. G. Beer, et al. (2005). "Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays." Clin Cancer Res **11**(2 Pt 1): 565-72.

- Dobner, P. R., E. S. Kawasaki, et al. (1981). "Thyroid or glucocorticoid hormone induces pre-growth-hormone mRNA and its probable nuclear precursor in rat pituitary cells." Proc Natl Acad Sci U S A **78**(4): 2230-4.
- Dowell, J. E. and J. D. Minna (2004). "Cancer chemotherapy targeted at reactivating the expression of epigenetically inactivated genes." J Clin Oncol **22**(8): 1353-5.
- Dufva, M., S. Petronis, et al. (2004). "Characterization of an inexpensive, nontoxic, and highly sensitive microarray substrate." Biotechniques **37**(2): 286-92, 294, 296.
- Duggan, D. J., M. Bittner, et al. (1999). "Expression profiling using cDNA microarrays." Nat Genet **21**(1 Suppl): 10-4.
- Duyk, G. M. (2002). "Sharper tools and simpler methods." Nat Genet **32** Suppl: 465-8.
- Flemming, W. (1882). Zell-substanz, Kern und Zelltheilung
- Forozan, F., R. Karhu, et al. (1997). "Genome screening by comparative genomic hybridization." Trends Genet **13**(10): 405-9.
- Forozan, F., E. H. Mahlamaki, et al. (2000). "Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data." Cancer Res **60**(16): 4519-25.
- Franssen-van Hal, N. L., O. Vorst, et al. (2002). "Factors influencing cDNA microarray hybridization on silylated glass slides." Anal Biochem **308**(1): 5-17.
- Fu, M., C. Wang, et al. (2004). "Acetylation of nuclear receptors in cellular growth and apoptosis." Biochem Pharmacol **68**(6): 1199-208.
- Gadski, R. A. and C. B. Chae (1976). "Mode of reconstitution of chicken erythrocyte and reticulocyte chromatin." Biochemistry **15**(17): 3812-7.
- Garner, H. R. (2001). Digital optical chemistry micromirror imager. U. P. Office, Board of Regents, The University of Texas System
- Gilbert, N., S. Boyle, et al. (2004). "Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers." Cell **118**(5): 555-66.
- Girard, L., S. Zochbauer-Muller, et al. (2000). "Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering." Cancer Res **60**(17): 4894-906.
- Hardiman, G. (2004). "Microarray platforms--comparisons and contrasts." Pharmacogenomics **5**(5): 487-502.
- Huang, S. Y., M. B. Barnard, et al. (1986). "The active immunoglobulin kappa chain gene is packaged by non-ubiquitin-conjugated nucleosomes." Proc Natl Acad Sci U S A **83**(11): 3738-42.
- Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.
- Kafatos, F. C., C. W. Jones, et al. (1979). "Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure." Nucleic Acids Res **7**(6): 1541-52.
- Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-21.
- Kayne, P. S., U. J. Kim, et al. (1988). "Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast." Cell **55**(1): 27-39.

- Keene, M. A., V. Corces, et al. (1981). "DNase I hypersensitive sites in *Drosophila* chromatin occur at the 5' ends of regions of transcription." Proc Natl Acad Sci U S A **78**(1): 143-6.
- Keene, M. A. and S. C. Elgin (1981). "Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure." Cell **27**(1 Pt 2): 57-64.
- Kim, T. H., L. O. Barrera, et al. (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-80.
- Kucheria, C. S., Wells; R.E. (1986). Bonding of bioactive glass coatings U. P. Office.
- Kuhn, K., S. C. Baker, et al. (2004). "A novel, high-performance random array platform for quantitative gene expression profiling." Genome Res **14**(11): 2347-56.
- Li, H. J. (1975). "A model for chromatin structure." Nucleic Acids Res **2**(8): 1275-89.
- Lockhart, D. J., H. Dong, et al. (1996). "Expression monitoring by hybridization to high-density oligonucleotide arrays." Nat Biotechnol **14**(13): 1675-80.
- Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution." Nature **389**(6648): 251-60.
- Matsuzaki, H., H. Loi, et al. (2004). "Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array." Genome Res **14**(3): 414-25.
- McGall, G. H., Fodor, S.P.A., Sheldon, E.L. (1992). Spatially-addressable immobilization of oligonucleotides and other biological polymers on surfaces U. P. Office, Affymax Technologies N.V.
- McQuain, M. K., K. Seale, et al. (2003). "Effects of relative humidity and buffer additives on the contact printing of microarrays by quill pins." Anal Biochem **320**(2): 281-91.
- Meselson, M., M. Nomura, et al. (1964). "Conservation of Ribosomes During Bacterial Growth." J Mol Biol **65**: 696-711.
- Nakatani, Y. (2001). "Histone acetylases--versatile players." Genes Cells **6**(2): 79-86.
- Niedzwiecka, A. and A. Kalinski (1977). "[Role of histone H1 in the structure and function of chromatin (author's transl)]." Postepy Biochem **23**(2): 175-88.
- Pinkel, D., R. Seagraves, et al. (1998). "High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays." Nat Genet **20**(2): 207-11.
- Pokholok, D. K., C. T. Harbison, et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." Cell **122**(4): 517-27.
- Pollack, J. R., C. M. Perou, et al. (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet **23**(1): 41-6.
- Rauscher, F. J., 3rd (2005). "It is time for a Human Epigenome Project." Cancer Res **65**(24): 11229.
- Renaud, S., D. Loukinov, et al. (2005). "CTCF binds the proximal exonic region of hTERT and inhibits its transcription." Nucleic Acids Res **33**(21): 6850-60.
- Robyr, D., Y. Suka, et al. (2002). "Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases." Cell **109**(4): 437-46.
- Rose, S. M. and W. T. Garrard (1984). "Differentiation-dependent chromatin alterations precede and accompany transcription of immunoglobulin light chain genes." J Biol Chem **259**(13): 8534-44.

- Rougeon, F., P. Kourilsky, et al. (1975). "Insertion of a rabbit beta-globin gene sequence into an E. coli plasmid." Nucleic Acids Res **2**(12): 2365-78.
- Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science **270**(5235): 467-70.
- Shalon, T. A., Brown, P.O. (1997). Methods for fabricating microarrays of biological samples U. S. P. Office, The Board of Trustees of the Leland Stanford Junior University: 49.
- Shimada, K., K. Sekikawa, et al. (1973). "A new device of preparative polyacrylamide gel electrophoresis and its application to analysis of cellular RNA." Anal Biochem **51**(2): 456-65.
- Snijders, A. M., D. Pinkel, et al. (2003). "Current status and future prospects of array-based comparative genomic hybridisation." Brief Funct Genomic Proteomic **2**(1): 37-45.
- Southern, E. M. (1989). Method and apparatus for analyzing polynucleotide sequences. E. P. Office: 15.
- Southern, E. M., U. Maskos, et al. (1992). "Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models." Genomics **13**(4): 1008-17.
- Southern, E. M. and A. R. Mitchell (1971). "Chromatography of nucleic acid digests on thin layers of cellulose impregnated with polyethyleneimine." Biochem J **123**(4): 613-7.
- Stengele, K. P., J. Buhler, et al. (2005). "Recent highlights on photolithic oligonucleotide array in situ synthesis." Nucleosides Nucleotides Nucleic Acids **24**(5-7): 891-6.
- Struhl, K. (1998). "Histone acetylation and transcriptional regulatory mechanisms." Genes Dev **12**(5): 599-606.
- Taddei, A., F. Hediger, et al. (2004). "The Function of Nuclear Architecture: A Genetic Approach." Annu Rev Genet.
- Temin, H. M. and S. Mizutani (1970). "RNA-dependent DNA polymerase in virions of Rous sarcoma virus." Nature **226**(5252): 1211-3.
- Tzankov, A., P. Went, et al. (2005). "Tissue microarray technology: principles, pitfalls and perspectives-lessons learned from hematological malignancies." Exp Gerontol **40**(8-9): 737-44.
- Urnov, F. D. and A. P. Wolffe (2001). "Above and within the genome: epigenetics past and present." J Mammary Gland Biol Neoplasia **6**(2): 153-67.
- Varshavsky, A. J. and V. V. Bakayev (1975). "Studies on chromatin. III. v-Bodies and free DNA in chromatin lacking histone H1." Mol Biol Rep **2**(3): 209-17.
- Vissers, L. E., J. A. Veltman, et al. (2005). "Identification of disease genes by whole genome CGH arrays." Hum Mol Genet **14 Spec No. 2**: R215-23.
- Volkin, E. and L. Astrachan (1956). "Phosphorus incorporation in Escherichia coli ribonucleic acid after infection with bacteriophage T2." Virology **2**(2): 149-61.
- Wang, C., M. Fu, et al. (2004). "Histone acetylation/deacetylation as a regulator of cell cycle gene expression." Methods Mol Biol **241**: 207-16.
- Weintraub, H. and M. Groudine (1976). "Chromosomal subunits in active genes have an altered conformation." Science **193**(4256): 848-56.
- Wolffe, A. P. (1994). "Transcriptional activation. Switched-on chromatin." Curr Biol **4**(6): 525-8.

- Wrobel, G., J. Schlingemann, et al. (2003). "Optimization of high-density cDNA-microarray protocols by 'design of experiments'." Nucleic Acids Res **31**(12): e67.
- Wu, C., Y. C. Wong, et al. (1979). "The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity." Cell **16**(4): 807-14.
- Wyrick, J., Young, R.A., Bing, R., Francois, R. (2001). Chromosome-wide analysis of protein-DNA interactions U. P. Office, Whitehead Institute for Biomedical Research (Cambridge, MA)
- Yamamoto, M., J. D. Clark, et al. (2005). "Regulation of oxidative stress by the anti-aging hormone Klotho." J Biol Chem.

Chapter 2

Development of techniques for robust hybridization of human genomic DNA to cDNA microarrays

2.1 Specificity and sensitivity of cDNA versus BAC microarrays

The first use of a labeled genomic DNA (gDNA)-based target for an aCGH application was performed by Shalon *et al.*, who arrayed lambda clones representing ~75% of the *S. cerevisiae* genome and then hybridized two yeast chromosomes labeled with separate fluorophores (Shalon, Smith *et al.* 1996). This experiment was critical to the invention of aCGH, since it proved that complex DNA samples could be reproducibly hybridized to microarrays. However, the size and complexity of the human genome made the creation of human aCGH methods much more difficult than the creation of that for yeast (Goffeau, Barrell *et al.* 1996; Forozan, Karhu *et al.* 1997; Pinkel, Seagraves *et al.* 1998; Maglott, Ostell *et al.* 2005).

The first implementation of human aCGH used BAC probes with ~40 kilobase resolution, but it left 100 kilobase gaps between the probed regions and only covered Chromosome 20 (Pinkel, Seagraves *et al.* 1998). While this was a major leap forward from the >20 megabase resolution of standard CGH methods, it was not yet capable of the subgene resolution needed to study microdeletions and other genetic abnormalities (Kallioniemi, Kallioniemi *et al.* 1994; Smith, Prasad *et al.* 1995). The ~40 kilobase resolution was selected to prevent any issues with the specificity and the sensitivity of the measurements, that are often associated with using high-resolution readouts to study low complexity human genomic DNA (Albertson, Ylstra *et al.* 2000; Snijders, Nowak *et al.* 2001; Albertson, Collins *et al.* 2003). Specificity, (which is a function of the genomic uniqueness of the probe), was achieved with these microarrays because the probes can

bind a variety of different targets along its length (Albertson, Collins et al. 2003). Why this is important is that sensitivity, (which is the ability to measure the signal of the fluorophores in the target bound to the probe), was achieved using these microarrays because numerous targets could be bound to a single molecule of probe. Sensitivity is a problem for aCGH because the number of fluorophores that can be integrated into a target is limited and the locations of the insertions random (Pinkel, Segev et al. 1998). Therefore blocking with unlabeled CoT1 DNA to prevent cross-hybridization of repetitive elements would effectively reduce the amount of target DNA by more than half, which reduces the available signal by the same amount (Venter, Adams et al. 2001). Long probes however can bind multiple targets simultaneously, which means that enough fluorophores can be bound to give a signal that is strong enough to be measured reliably (Eisen and Brown 1999; Snijders, Nowak et al. 2001).

The first example of human aCGH at a subgene resolution used radiation-hybrid mapped cDNA clones as probes to pinpoint genomic changes in several breast cancer cell lines, such as 10 fold amplification of *ERBB2* or the deletion of *TP53* (Pollack, Perou et al. 1999). Specificity issues were handled by using sequence-verified clones that were genomically unique, and free of repetitive elements, but sensitivity was still a problem. Because both probes and targets were short, getting enough labeled target (fluorophores) bound to the probe to have a detectable signal was difficult. Pollack countered this by developing a high efficiency labeling protocol, as well as using novel microarray scanner hardware to achieve sensitivity twice that of anything previously reported (Pollack, Perou et al. 1999).

Pollack had now proven that it was possible to use cDNA spotted microarrays for aCGH, but only on his own equipment. In order to be able to reproduce his work, it was necessary to validate that the probes were acceptable for genomic DNA hybridization and that the hybridization produced sufficient signal to resolve differences in copy number. Because of these problems, many groups chose low-resolution BAC spotted microarrays in the following years. The higher printing densities and better validation of these probes permitted sub-megabase resolution across the entire human genome (Snijders, Nowak et al. 2001). However, subgene resolution remains beyond the reach of BAC spotted microarray technology.

2.2 Measurement of human genomic DNA coverage on cDNA microarrays

The original focus of the work that led to developing a protocol for high-resolution aCGH on cDNA or oligonucleotide spotted microarrays was to create a universal reference standard for expression microarray analysis (Weil, Macatee et al. 2002). The characteristics of a perfect reference standard are very demanding, since it must be truly universal, easily obtainable, batch invariant, low cost, and most importantly, complete with regard to the probes on the microarray. At the time I began work on this problem there was no commercial product to meet this need, so given the stated requirements, the best solution for this problem was the genomic DNA from the organism itself. It was first necessary to prove that the microarrays and techniques used locally had the sensitivity (as a function of coverage) required for use in high resolution aCGH-style experiments. In order to determine the hybridization characteristics of the genomic DNA and how this compared to hybridizing the microarray with an RNA-based

target, a second sample consisting of pooled RNAs was used. The pooled RNA sample was supposedly genomically complete and since the sensitivity and coverage of RNA-based targets was not in question, this sample was used as a positive control (Novoradovskaya N. 2000; Kim, Zhao et al. 2002; Yang, Chen et al. 2002).

The cDNA library used for spotting the microarrays used in these experiments had been previously sequence-verified by Research Genetics and the IMAGE consortium, so only minimal resequencing was necessary to validate the identity and purity of the clones before spotting (Wang and Rowley 1998). Additionally, in order to minimize development time, I used the Pollack hybridization protocol as a starting point for the new protocol (Pollack, Perou et al. 1999). The complete materials and methods are discussed in section 2.4.1. The validation dataset consisted of nine arrays (seven 4,000 spot microarrays and two 10,800 spot microarray) co-hybridized with both genomic DNA and pooled RNA target. The correlation between the microarrays, including the interplatform correlation, was greater than 90% once the coverage differences were accounted for. Given the high correlation between the results from the two microarray designs (4,000 spot and 10,800 spot) the analysis to follow will be focused on the more complete 10,800 spot microarray experiments.

The metric for determining if it was possible to use these microarrays for an aCGH-based application was coverage, measured by the double positives rate--the number of spots above the signal threshold in both targets. With 8,846 genes (82%) hybridizing above threshold (see figure 2.1 for the threshold values) in both channels, genomic DNA could be hybridized to the microarray with the sensitivity similar to RNA-based targets. However, to more fully understand the hybridization characteristics of

genomic DNA, the number of spots below threshold (the failure rate) in both samples (1,786) was measured. In addition, the number of spots where either the RNA pool was positive and the genomic DNA negative (71; consisting primarily of tissue-specific, low copy number genes) or vice versa (97; primarily expressed sequence tags and hypothetical proteins) was determined to provide a measure of the difference in the coverage between the two samples. In figure 2.1 it can be seen from the behavior of the curves that neither sample shows significantly better sensitivity or coverage than the other.

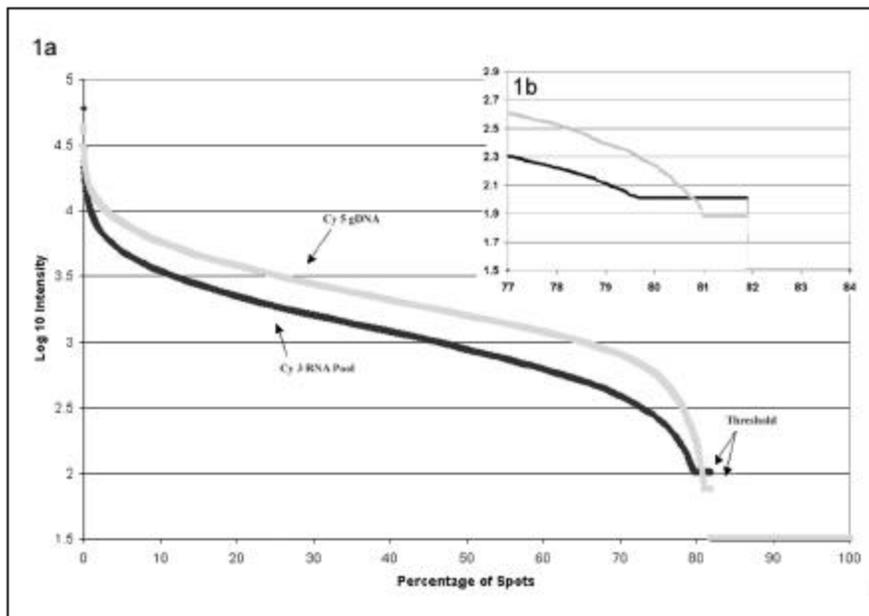


Figure 2.1: The \log_{10} of the background-subtracted spot intensities sorted by value shows the minimal difference between the two samples. 100% corresponds to 10,800 spots on the microarray. (A) The gray line represents the genomic DNA labeled in Cy5 and the black line represents the RNA pool labeled in Cy3. (B) A magnification of the low-intensity threshold region of the curve better illustrates the small difference between the two samples. The flattening of the curve and the length of the tails are measures of the number of spots above the threshold and the reproducibility of the samples, respectively. The line at the bottom at 1.5 \log_{10} intensity is where neither sample is above threshold in either channel and therefore is a measure of the coverage failure.

Because the microarrays can be treated as a single experiment with each spot representing an independent trial ($n = 10,800$), the McNemar test with a Yate's correction was chosen to show that the samples had equal coverage (Rosner 1995). With one degree of freedom, the null hypothesis that the samples had identical coverage could not be rejected, given a χ^2 value of 0.24 and $P = 0.62$, indicating that both samples behaved identically. From this result I concluded that the microarrays and methods used locally had the sensitivity required for use in high-resolution aCGH applications, like CHIP on chip or chromatin array experiments. However, there was still more that could be learned from these results, since a significant number of probes did not show signal above signal threshold for either sample.

Failure rate is an important measure of the sample hybridization characteristics and microarray quality. To determine the cause of the failures, a survey of each of the failed spots and the reason for its failure to hybridize was undertaken. 127 of the missing spots were negative controls, buffer blanks, and empty wells (i.e. no liquid was deposited during printing) that were not expected to be above threshold, leaving 1,659 spots unexplained. To determine which spots were low quality and which were not hybridizing for other reasons, microarray data was gathered from other experiments using the same microarray print run and then analyzed by the same means used for the genomic DNA and pooled RNA samples. With 5 additional microarrays including dye reversals, and the use of indirect labeling, 427 of the 1,659 spots were identified as not being above threshold. Using reproducibility as a measure of success this number increases to 873 spots, if the minimum requirement for inclusion is defined as being above the threshold in more than half of the experiments. Because such a large portion (8%) of the spots

never had signal above the threshold, further experiments were required to understand the reason why so many probes failed.

The labeled clone amplification primers were then employed as a third independent sample to determine the quality of the spots on the microarray. While the primer does not provide a direct measure of the hybridization process, every spot on the microarray has a copy of both primers. Therefore, any fluorescence means that probe was spotted. Two additional microarrays were hybridized with the primer labeled with both fluorophores. The primer sample, which is complete in the context of the microarray because it is incorporated into every DNA spot, had 1,865 spots that did not exceed the threshold value in any of the four trials (two arrays times two fluorophores). When the results were compared across all the microarrays and the three types of samples, only 158 spots did not hybridize above the threshold in any experiment. These results indicate that the failure to get signal above threshold did not lie in the hybridization technique but in the hybridization characteristics of the sequences themselves (Yang, Chen et al. 2002).

These preliminary experiments showed conclusively that the locally-produced microarrays and methods had the coverage and reproducibility required for high-resolution aCGH style applications. However, the use of cDNA-based microarrays at UTSW was short-lived because cDNA libraries were error-prone and subject to much more cross-hybridization than originally expected, even when the clone was found to be specific using computational methods (Yue, Eastman et al. 2001; Wren, Kulkarni et al. 2002). Therefore, it was necessary to extend the method to higher resolutions using the newly introduced oligonucleotide microarrays.

2.3 Refinements to the high-resolution aCGH method

In 2002 I worked with the lab's Digital Optical Chemistry group to demonstrate that using the protocol I developed, aCGH was possible using microarrays with oligonucleotide features as short as 21 bases, which also has single base resolution (figure 2.2). Because this microarray was a proof-of-principle for single base resolution aCGH on a microarray designed for expression microarray, the probes were designed to maximize the gene coverage, not genomic completeness. Accordingly, in order to demonstrate the ability to resolve genomic copy number changes, the human lung cancer cell line H740 was used, since it has a homozygous deletion spanning 1.5 megabases of the 3p21 region. This deletion removes a region which contains over twenty-two genes, including *CACNA2D2*, the *HYAL* and *FUS* families of genes, as well as other important tumor suppressor genes (Girard, Zochbauer-Muller et al. 2000). Analysis of genes with probes on the microarray in the deleted region showed that the signal for the deleted genes was below detection limits in the deleted cell line when compared with signal generated by the normal lung epithelial DNA. In regions that were not altered, the two samples have similar signal levels, which are seen in figure 2.2 as the high percentage of yellow features. Thus all data were consistent.

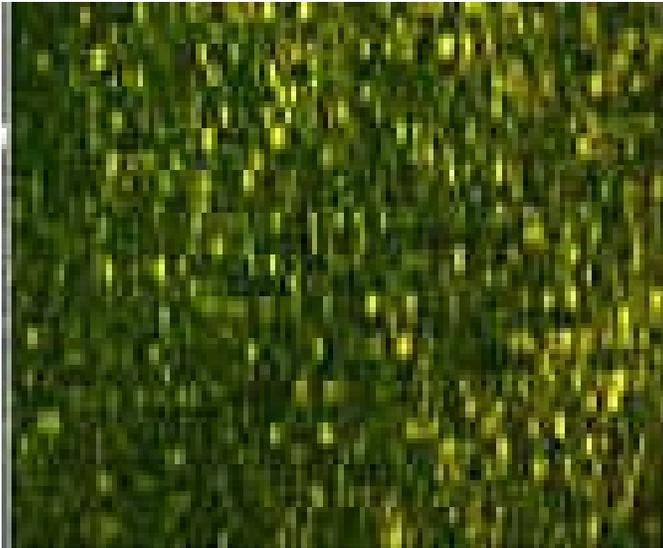


Figure 2.2: A portion of a DOC microarray with probes for over 8,000 human genes, used for aCGH. The samples are the lung cancer cell line H740 (red) versus normal lung epithelial DNA (green). Unlike spotted microarrays the features are square and have no dead space between them. The yellow tinge of this microarray visually indicates equal hybridization of the two cell lines for most sequences, which was expected given the known karyotypes.

By 2003, aCGH using oligonucleotide microarrays was fairly commonly performed and a number of companies, including MWG (MWG A.G., Highpoint NC), were advertising “no blocking aCGH” which meant that it was unnecessary to add unlabeled CoT1 DNA to the hybridization buffer. The basis of the claim was that the oligonucleotide probes were specially designed using their proprietary “Oligos4Array” design algorithm to be sequence-specific, meaning that they would not cross hybridize with repetitive elements. Before using the MWG microarrays for an aCGH style application, I had to test this assertion. Labeled CoT1 DNA (Invitrogen, Carlsbad, CA) (green) was co-hybridized with labeled genomic DNA (red) at equal concentrations on the MWG “no block” microarray using conditions specified in the MWG protocol. In figure 2.3 the strong green coloring of the microarray demonstrates that the probes designed not to bind repetitive sequences are still binding them. Furthermore, the

suppression of the red signal shows that the CoT1 is competing off the genomic DNA. After working with the manufacturer, suitable blocking protocols were developed and blocking was reinstated in the recommended protocol. The best signal is obtained by labeling 2 μg of genomic DNA for each target and the repetitive elements are blocked by the addition 50 μg of unlabeled CoT1 DNA to the hybridization buffer (Weil, Macatee et al. 2002). These new protocols allowed these microarrays to be used for high-resolution aCGH applications. To my knowledge this was the first time a commercially available microarray had been validated for aCGH use (Weil, Widlak et al. 2004). The use of these microarrays in aCGH-style applications including the materials and methods will be covered in detail in Chapter 3.

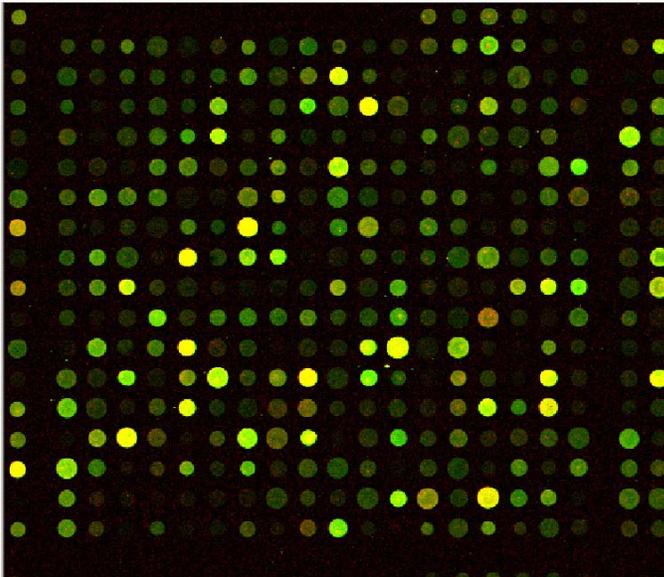


Figure 2.3: CoT1 DNA is able to bind and compete off the MCF7 DNA even using probes designed to be sequence specific. CoT1 DNA was labeled with Cy3 (green) and MCF7 total genomic DNA was labeled with Cy5 (red) and hybridized. The level of nonspecific binding is seen in how green the image is.

In addition to my work in validating high-resolution human aCGH, I collaborated with Dr. Talaat on his efforts to use aCGH style microarrays which tile across small bacterial genomes to identify genomic regions previously not known to be coding (Talaat, Howard et al. 2002). As part of this work, Dr. Talaat showed that nick translation-based labeling methods, (part of the original Pollack protocol), were the best way to label the genomic DNA since the yield was very reproducible, but random priming using Klenow methods was also acceptable. In 2003 Promega discontinued their nick translation kit, expanding the value and usefulness of Dr. Talaat's results. Without a commercial nick translation kit available, most labs had to switch to labeling by random priming using a modified version of Invitrogen's Bioprime kit.

2.4 Materials and methods

2.4.1 Genomic DNA Labeling

High molecular weight genomic DNA was extracted from normal human lung fibroblasts (GM 01604) or was purchased from Invitrogen (Carlsbad, CA). The labeling was done using the Nick Translation kit (Promega, Madison, WI) with 2 μg genomic DNA. This DNA concentration was experimentally determined (concentrations from 0.5 to 5 μg were tested) to minimize DNA-based suppression of the RNA signal while maximizing the number of spots above threshold (Talaat, Howard et al. 2002; Weil, Macatee et al. 2002). The protocol supplied with the kit was followed, with two exceptions. First, no unlabeled dCTP was used in the 10 mmol dNTP mixture; instead, 1 μL of the appropriate cyanine (Cy) dye-conjugated dCTP (Amersham Biosciences,

Piscataway, NJ) was added to each 50 μ L reaction (Talaat, Hunter et al. 2000; Talaat, Howard et al. 2002). Second, the reaction was allowed to run for 7.5 hours, instead of the specified 15 min, to increase the incorporation of dye and reduce the average fragment size.

2.4.2 RNA Labeling

The total RNA pool was purchased from Stratagene (Stratagene, La Jolla, CA). The pool consists of 10 proprietary cell lines representative of most of the body's tissues and therefore representative of most of the expressed sequences (Novoradovskaya N. 2000). Twenty micrograms of the total RNA pool were reverse-transcribed using SUPERScript [®]II and Oligo (dT) 12–18 from Invitrogen. The labeling reaction was performed using a standard protocol (Eisen 1999). The variations from this protocol were a reduction in the reaction volume to 40 μ L and a 1 hour primer annealing step before the addition of reverse transcriptase to increase the specificity.

2.4.3 Microarray Production

The microarrays used were produced in the University of Texas Southwestern (UTSW) Array Core using the PCR products of sequence-verified clones (Research Genetics, Huntsville, AL) and the universal vector primers (forward, 5'-CTGCAAGGCGATTAAGTTGGGTAAC-3', and reverse, 5'-GTGAGCGGATAACAATTTACACAGGAAACAGC-3'). A complete clone list is available from <http://microarray.swmed.edu>. After the size and purity (single band) were verified using gel electrophoresis, the PCR products were resuspended in 7% DMSO and

printed at a pitch of 0.28 mm on poly-L-lysine-coated slides using MAGNA™, a custom-built spotting robot available from Bioautomation (Plano, TX). The detailed protocols, microarray analysis template files, and the materials list are available at <http://innovation.swmed.edu>. The DOC microarray was custom synthesized using light directed synthesis (Luebke, Balog et al. 2003). The microarrays for the CoT1 experiment were MWG human 20K A microarrays purchased from MWG AG (High Point, NC).

2.4.4 Microarray Clean Up and Hybridization

The labeling reactions were purified using 30K Microcon® spin columns (Millipore, Bedford, MA) with two 400- μ L TE rinses. The elution volume was less than or equal to 16 μ L, with any missing volume replaced with TE to bring the final combined volume to 32 μ L. The hybridization buffer consisted of 8 μ L 20 \times SSC and 2 μ L 10% SDS passed through a 0.22- μ m filter, to which 3.2 μ L 10 mg/mL yeast tRNA were added for a total volume of 45.2 μ L. After boiling for 2 min, the solution was pipetted onto one of the UTSW microarrays, either a 10,800- or 4,000-member microarray, and covered with a 24 \times 60 mm coverslip. The microarray was placed in a sealed hybridization chamber (Telechem, Sunnyvale, CA) with 10 μ L 3 \times SSC in the wells to maintain the humidity. The chamber was then sealed and placed in a 61°C water bath for 16 h.

2.4.5 Washing and Scanning

After the hybridization chamber was removed from the water bath, the chamber was immediately opened and the microarray was placed in the first wash buffer (2 \times SSC, 0.1% SDS.) The cover slide was floated off in the first wash buffer and then the container

was gently agitated for 5 min. The microarray was washed in the next two wash buffers (0.4× and 0.2× SSC, respectively) as before. The microarray was dried by centrifugation at 155x g for 3 min at room temperature and scanned on a GenePix™ 4000b two-color scanner (Axon Instruments, Union City, CA) at 10 µm resolution. The pixel-intensity extraction and spot flagging were done with the GenePix 3.0 software package (Axon Instruments, Union City, CA), and the output was converted to a tab-delimited text file.

2.4.6 Analysis

The data files were imported into a custom Microsoft® Excel® worksheet developed at UTSW. The values of the mean signal intensity from each spot were subtracted from the local mean background intensity. Spots that were not above the threshold (calculated by the mean of the blanks plus one standard deviation of the blanks) were filtered out. The data was then analyzed for the number of spots above this threshold, reproducibility, and the threshold value. Annotation of the gene lists was done manually using the SOURCE database which was part of the Stanford Microarray Database (Sherlock, Hernandez-Boussard et al. 2001). The DAG STAT statistical package for Excel (The Mental Health Research Institute, Victoria, Australia), was used to run the McNemar test with a Yate's correction, to determine if there was a significant difference between the two samples (Rosner 1995).

2.5 References

- Albertson, D. G., C. Collins, et al. (2003). "Chromosome aberrations in solid tumors." Nat Genet **34**(4): 369-76.
- Albertson, D. G., B. Ylstra, et al. (2000). "Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene." Nat Genet **25**(2): 144-6.
- Eisen, M. B. and P. O. Brown (1999). "DNA arrays for analysis of gene expression." Methods Enzymol **303**: 179-205.
- Eisen, M. B. a. P. O. B. (1999). DNA arrays for analysis of gene expression. Methods Enzymology. **303**: 179-205.
- Forozan, F., R. Karhu, et al. (1997). "Genome screening by comparative genomic hybridization." Trends Genet **13**(10): 405-9.
- Girard, L., S. Zochbauer-Muller, et al. (2000). "Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering." Cancer Res **60**(17): 4894-906.
- Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.
- Kallioniemi, O. P., A. Kallioniemi, et al. (1994). "Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors." Genes Chromosomes Cancer **10**(4): 231-43.
- Kim, H., B. Zhao, et al. (2002). "Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses." Biotechniques **33**(4): 924-30.
- Luebke, K. J., R. P. Balog, et al. (2003). "Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts." Nucleic Acids Res **31**(2): 750-8.
- Maglott, D., J. Ostell, et al. (2005). "Entrez Gene: gene-centered information at NCBI." Nucleic Acids Res **33**(Database issue): D54-8.
- Novoradovskaya N., P. T., Novoradovsky A., Braman J., Chin N., Pergamenschikov A., Fero M., Botstein D. (2000). "Pooled reference RNA for human microarrays." Stratagene Strategies(13): 121-122.
- Pinkel, D., R. Seagraves, et al. (1998). "High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays." Nat Genet **20**(2): 207-11.
- Pollack, J. R., C. M. Perou, et al. (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet **23**(1): 41-6.
- Rosner, B. (1995). Fundamentals of Biostatistics. Belmont, CA, Wadsworth Publishing Company.
- Shalon, D., S. J. Smith, et al. (1996). "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization." Genome Res **6**(7): 639-45.
- Sherlock, G., T. Hernandez-Boussard, et al. (2001). "The Stanford Microarray Database." Nucleic Acids Res **29**(1): 152-5.

- Smith, A., M. Prasad, et al. (1995). "Comparison of high resolution cytogenetics, fluorescence in situ hybridisation, and DNA studies to validate the diagnosis of Prader-Willi and Angelman's syndromes." Arch Dis Child **72**(5): 397-402.
- Snijders, A. M., N. Nowak, et al. (2001). "Assembly of microarrays for genome-wide measurement of DNA copy number." Nat Genet **29**(3): 263-4.
- Talaat, A. M., S. T. Howard, et al. (2002). "Genomic DNA standards for gene expression profiling in Mycobacterium tuberculosis." Nucleic Acids Res **30**(20): e104.
- Talaat, A. M., P. Hunter, et al. (2000). "Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis." Nat Biotechnol **18**(6): 679-82.
- Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.
- Wang, S. M. and J. D. Rowley (1998). "A strategy for genome-wide gene analysis: integrated procedure for gene identification." Proc Natl Acad Sci U S A **95**(20): 11909-14.
- Weil, M. R., T. Macatee, et al. (2002). "Toward a universal standard: comparing two methods for standardizing spotted microarray data." Biotechniques **32**(6): 1310-4.
- Weil, M. R., P. Widlak, et al. (2004). "Global survey of chromatin accessibility using DNA microarrays." Genome Res **14**(7): 1374-81.
- Wren, J. D., A. Kulkarni, et al. (2002). "Cross-hybridization on PCR-spotted microarrays." IEEE Eng Med Biol Mag **21**(2): 71-5.
- Yang, I. V., E. Chen, et al. (2002). "Within the fold: assessing differential expression measures and reproducibility in microarray assays." Genome Biol **3**(11): research0062.
- Yue, H., P. S. Eastman, et al. (2001). "An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression." Nucleic Acids Res **29**(8): E41-1.

Chapter 3

Development of a microarray-based assay to measure chromatin condensation state at a subgene resolution

3.1 Development of methods to recover chromatin by its condensation state in preparation for hybridization

The study of epigenetic diseases such as Rett's and Angelman's Syndromes has revitalized interest in epigenetics in recent years (El-Osta and Wolffe 2000; Scarano, Strazzullo et al. 2005). Interest escalated once it was recognized that common afflictions such as cancer and cardiac disease also have epigenetic components (Esteller and Herman 2002; Robertson 2005). Recently there have been calls for an Epigenome Project, which would map epigenetic alterations in a fashion similar to the sequencing of the human genome (Jones and Martienssen 2005; Rauscher 2005). To accomplish this goal, methods must be developed for high throughput measurement of every aspect of the epigenome. One of the most critical pieces of information is the measurement of the chromatin condensation state (Jones and Martienssen 2005). My goal to create a microarray-based translation of the chromatin condensation state by merging aCGH with traditional chromatin methods (figure 3.1) required me to overcome significant technical challenges. Foremost among these challenges was separating the chromatin by state in a manner compatible with the microarray labeling protocols.

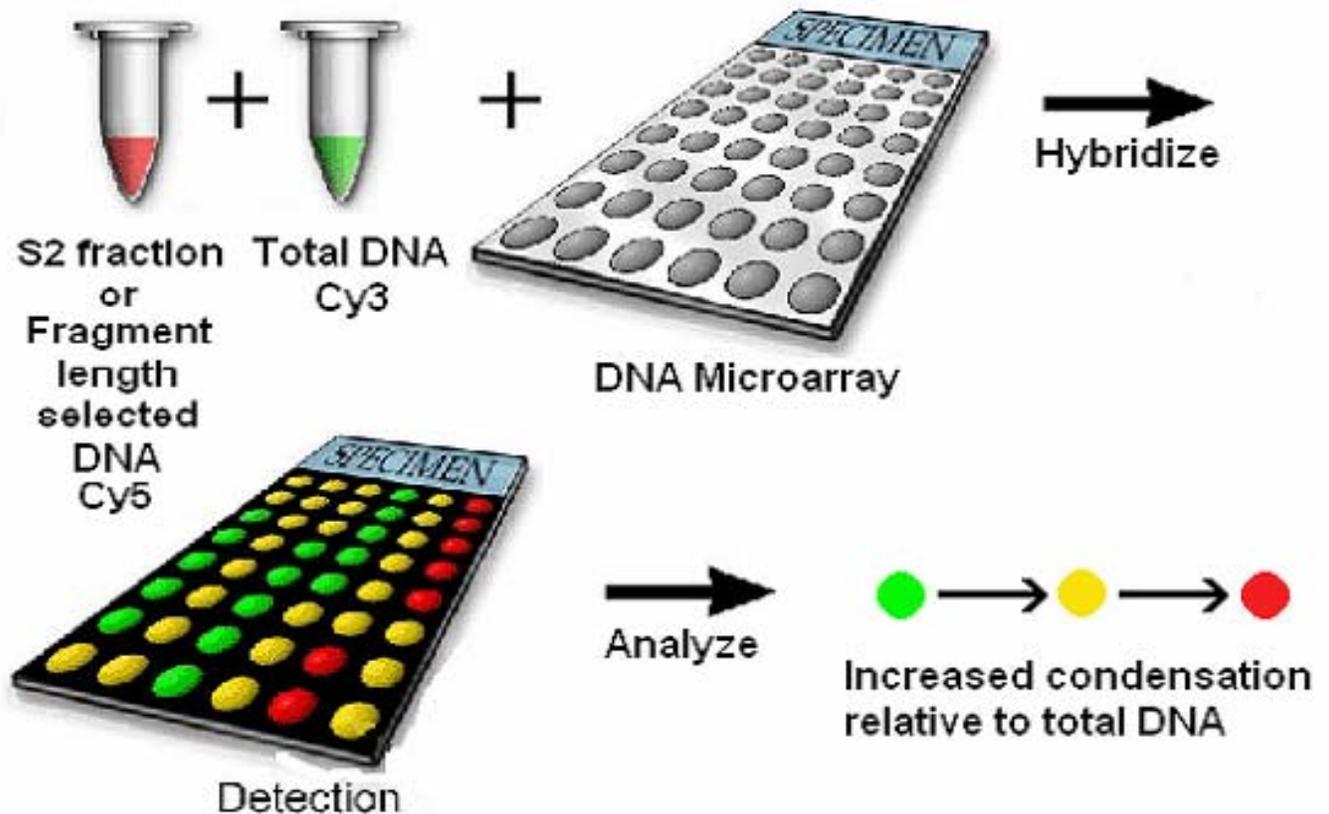


Figure 3.1: A diagram showing the chromatin array method. The condensed DNA is selectively extracted from “living” cells. The sample and the control DNA are labeled and hybridized to the microarray. After washing and scanning, the microarray data is normalized and the differential between the samples measured to determine enrichment or depletion from genomic levels. Differential probe intensity is then used in part to determine the condensation state of the gene. This figure was modified from <http://folk.uio.no/leonardm/ArrayCGHprinciple.htm>

3.1.1 Fragment Length Selection

Nuclease protection assays reveal information about local chromatin state by measuring susceptibility of the DNA to enzymatic digestion. For this assay the fragments would be separated by gel electrophoresis and then transferred to a membrane for probing (Weintraub and Groudine 1976). However, for microarray-based readout, the fragments must be separated by their chromatin state and left in solution for labeling and hybridization. Without a separation step, the only probes on the microarray that would

show any deviation in intensity from the normal genomic levels would be at hypersensitive sites. Because more complete digestion can occur when the chromatin state is fully open, there would be less target in these regions causing a reduction in signal (Weintraub and Groudine 1976). The problem of separating the chromatin state by fragment length was overcome using the same gel electrophoresis methods employed by the traditional blot-based chromatin state methods, but instead of transferring the DNA to a membrane, the selected fragment lengths (bands) were cut out of the gel and recovered for labeling (figure 3.2).

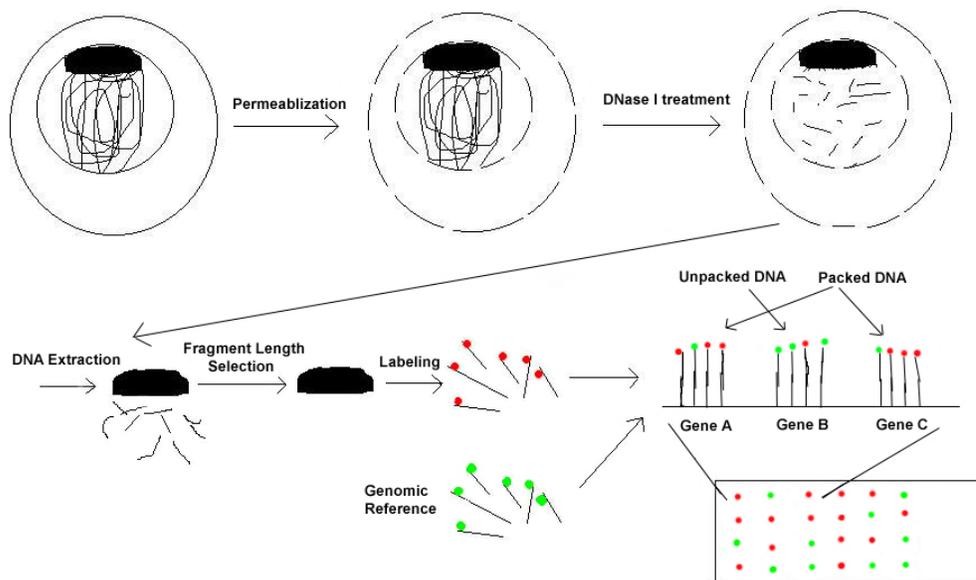


Figure 3.2: A diagram of the fragment length selection method and the chromatin array. The first steps of fragment length selection are similar to foot printing, however the results are not read out with a Southern Blot. The DNA is recovered and labeled so that it can be hybridized to a microarray. The relative intensity of the signal generated by the fragment length selected DNA as compared to the signal generated by the genomic reference for a given spot provides a measure of the chromatin condensation state.

The band selection is based on fragment length and completion and uniformity of digestion (figure 3.3). The region selected represents a compromise between enriching

for condensed chromatin and recovering enough DNA to label and hybridize. The top 40% of the band's intensity (which was the top ~33% of its length) was selected and excised. Of the 15 μ gs of DNA that are loaded per well, this region after extraction yielded on average ~3 μ g of DNA, sufficient for a single hybridization. While this method of separating chromatin by condensation state was successfully used for several microarray-based experiments, it proved cumbersome. In addition, the method proved difficult to scale up in order to interrogate the chromatin state of a large number of cell lines. Other fragment length selection methods utilizing salt or sugar gradients and ultracentrifugation to separate the DNA by its mass exist but are actually more troublesome than gel-based methods (Zardawi and Duncan 2003; Griesenbeck, Boeger et al. 2004; Ricke and Bielinsky 2005). Furthermore, all fragment length/mass methods of chromatin state measurement share the same flaw, that digestion efficiency is only correlated to chromatin condensation state (Rose and Garrard 1984).

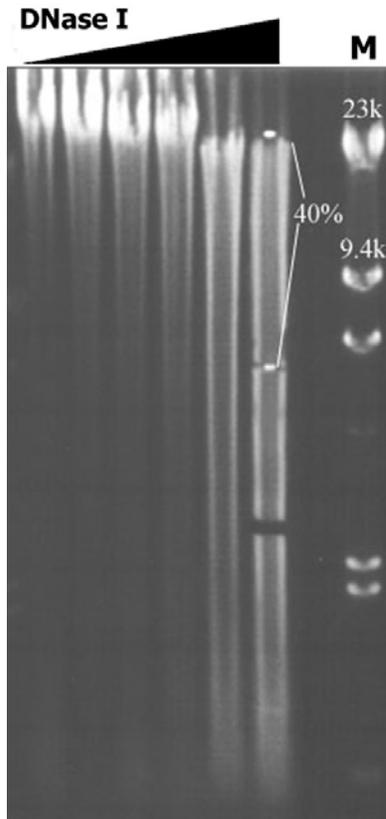


Figure 3.3: A fragment length selection gel showing the distribution of DNase I cleaved fragments with the concentration of DNase I increasing from left to right. M is the Lambda Hind III DNA mass standard. The lane digested with the highest concentration of DNase I (64U) was chosen for its uniform distribution of fragments around the 5 kb median. The fragments taken are the upper 40% of the total lane intensity (~6kb to 23kb), which because of the high mass of the fragments is enriched for condensed DNA. This section of the gel was excised and the DNA recovered for labeling.

3.1.2 Chromatin Fractionation

As discussed in Chapter 1, chromatin fractionation is another technique which can separate the chromatin by its condensation state (Rose and Garrard 1984). The chromatin fractionation method directly separates the chromatin into three fractions based on DNA/protein association. Given that the proteins associated with the chromatin define it as active or inactive, the separation has direct meaning in regards to the chromatin state (Huang, Barnard et al. 1986).

The chromatin fractionation method first utilizes micrococcal nuclease to generate mono- and oligonucleosomes which are separated into three fractions, designated S1, S2 and P (figure 3.4). Transcriptionally “active” DNA is found in the S1 and P fractions (Huang, Barnard et al. 1986). In MCF7, this comprises approximately 68% of the total DNA, though in mature B-cells it can be as low as 15%, since they produce a small subset of gene products (Huang, Barnard et al. 1986). The S1 fraction is depleted in histone H1 and enriched in high mobility group (HMG) proteins and heterogeneous ribonucleoproteins particles (HnRNPs), both of which are known to be associated with actively transcribed chromatin (Huang, Barnard et al. 1986). Likewise, the P fraction is highly enriched in non-histone proteins, and with further digestion can be partially converted to the S1 fraction (Rose and Garrard 1984; Huang, Barnard et al. 1986). The S2 fraction contains nucleosomes stoichiometrically associated with histone H1 and is highly deficient in non-histone proteins (Rose and Garrard 1984). The S2 fraction represents the most condensed chromatin fraction as indicated by previous studies, which have demonstrated that histone H1 is responsible for the formation of higher-order chromatin structures and is a general repressor of transcriptional activity (Allan, Cowling et al. 1981; Croston, Kerrigan et al. 1991). Figure 3.4 illustrates the differences between the S1, S2, and P fractions in terms of the protein compositions (A), DNA fractionation patterns (B), and as a molecular model of the digestion process (C). For the chromatin array studies, the S2 fraction was chosen because it is the least sensitive to digestion variations during fractionation and therefore the most reproducible.

Given that the separation of the chromatin states using the chromatin fractionation method is direct, far fewer cells are required to ensure a yield of classified DNA

sufficient to run the microarrays as compared to fragment length selection. It was clear that chromatin fractionation was likely the best method to separate chromatin, and consequently the chromatin microarray was largely developed based on this method. The fragment length selection-based method remained in use only to independently validate the results derived from the chromatin fractionation method.

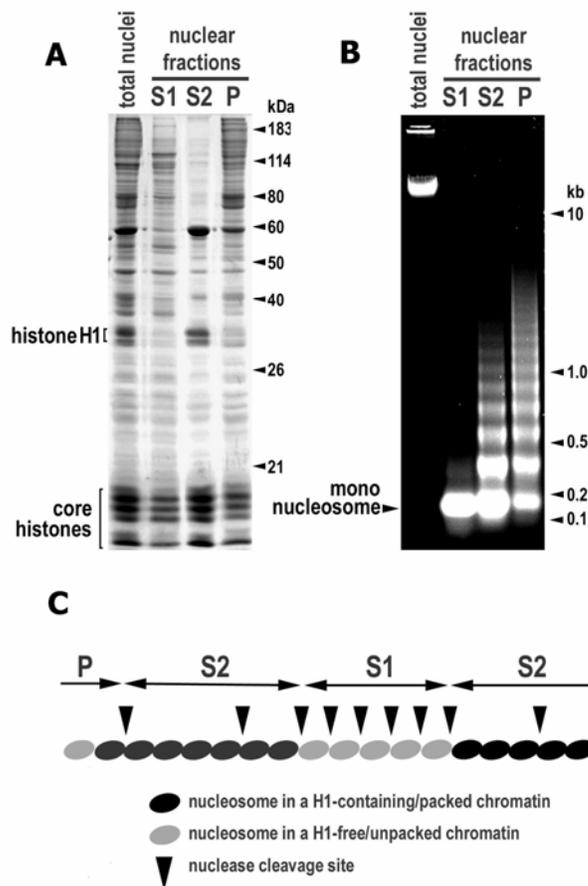


Figure 3.4: The chromatin solubility assay was used to extract the S1, S2, and P fractions. The result of that process is depicted as the patterns seen in the gels of the resulting protein and DNA. The protein gel (A) shows the depletion of the H1 histone and the enrichment of non-histone proteins in the S1 and P fractions. The DNA gel (B) shows the difference in the extent of the digestion between the fractions. The model (C) illustrates the basic concept of the chromatin fractionation assay.

3.2 Validation of the microarray-based method

3.2.1 Validation by traditional methods of chromatin state measurement

The initial validation was to determine if the chromatin fractionation had properly separated the chromatin by state. Because MCF7 had never been fractionated before, a test was performed utilizing the previously fractionated line BL60 (Huang, Barnard et al. 1986). The fractionated DNA was probed by Southern blotting for genes with known chromatin states to determine whether the fractionation was successful. The results of this work, which were done in collaboration with Dr. Widlak, are shown on table 3.1 (Widlak 2004). These results demonstrate that the fractionation was successful in dividing the chromatin by condensation state, with the actively transcribed Cκ and ribosomal DNA (rDNA) genes being excluded from the inactive S2 fraction and the inactive (unexpressed) β-globin gene being excluded from the active S1 fraction (Huang, Barnard et al. 1986; Widlak 2004).

Table 3.1: The validation of a selection of genes from the BL60 chromatin fractionation

Probe	Fraction			Total DNA
	S1	S2	P	
Total DNA	+++	+++	+++	+++
Cκ	+++	-	+++	+
β-globin	-	++	+	++/+
rDNA	-	-	+++	+

The level of signal measured on the Southern blot are indicated by the number of +s going from +++ which was the brightest to – which is no signal.

Once the chromatin fractionation method was validated using a well studied cell line, it was possible to extend the method to MCF7 and use the global parameters discussed in section 3.1.2 to validate the efficiency of the fractionation in that cell line. Comparison of published chromatin fractionation results for BL60 to the results of

MCF7, as seen in figure 3.4 (A and B), made it possible to determine if the MCF7 chromatin fractionation was successful (Huang, Barnard et al. 1986). In panel A, the S2 fraction was as expected; strongly enriched for histone H1, and by comparison to the other fractions strongly depleted for other chromatin associated proteins. In panel B, the distribution of DNA fragment lengths is also what would be expected from a successful chromatin fractionation. The S1 fraction, which is actively transcribed, is reduced to mononucleosomes, while the S2 fraction, which has an increased level of resistance to nuclease cleavage, shows much larger fragment lengths. The P fraction, which has the most resistance to cleavage, shows the largest fragments and the fewest mononucleosomes. Given how well these results match the published relationships on chromatin fractionation (Huang, Barnard et al. 1986), the chromatin fractionation of MCF7 was considered adequate to continue the validation.

3.2.2 Validation of the chromatin array method

The initial test of the results of any microarray-based technique is how reproducible are replicate experiments. In this case, as seen in figure 3.5, the reproducibility of the microarray data with minimal filtering is exceptional with an $R^2=0.791$, since studies have shown that R^2 s approaching 0.9 are the maximum for even high quality expression studies (Jarvinen, Hautaniemi et al. 2004; Draghici, Khatri et al. 2005). From a total of 19,437 unique genes on the microarrays, 7,955 genes passed the quality filters as discussed in the section 3.4 (see table 3.2). The average signal intensity of the S2 fraction in the experimental series after removing genes performing below a threshold defined by the cross-gene error model (Agilent 2006) was a 2.25 fold reduction from the average intensity of the total genomic DNA reference. The reason for this

change was likely a result of the fact that the microarrays used were designed for gene expression studies and therefore are biased toward genes known to be commonly expressed (Kane, Jatkoe et al. 2000).

Table 3.2: Measurements of the Chromatin Array method's reproducibility

	# Genes	Reproducibility	Concordance
Group 1: C/E >2	2,620	86%	84%
Group 2: C/E <0.5	1,493	76%	53%
Group 3: C/E <2 to >0.5	3,842	69%	64%
Genes passing all filters	7,955	--	--

The number of genes (19,437 possible) in each group that pass all data processing filters is shown. C/E is the ratio of condensation state over expression level. Reproducibility refers to the percentage of genes that yield similar results in an independent replicate experiment of chromatin solubility fractionation on a different array platform. Concordance refers to the percentage of genes between the merged fragment length selection data and the chromatin solubility data that are consistently classifiable to each group to be discussed in section 3.2.3.

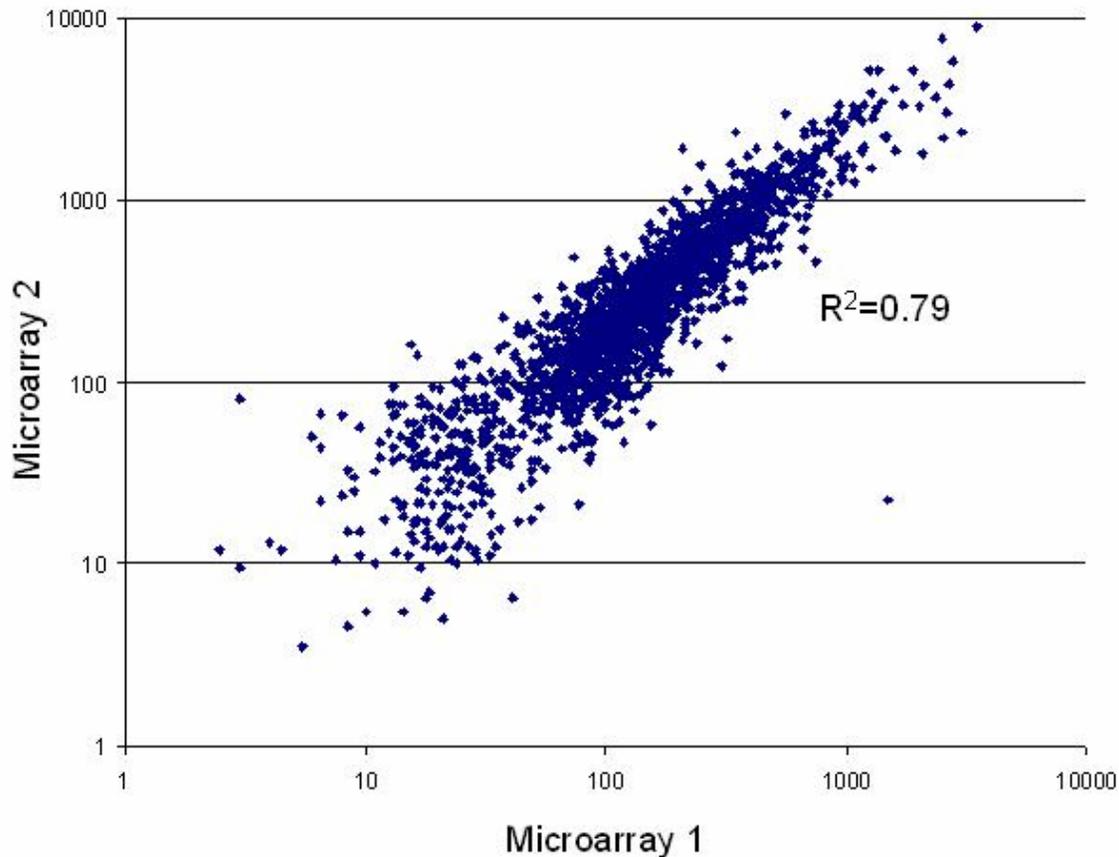


Figure 3.5: A comparison of replicate chromatin microarrays shows the reproducibility of the method. With the data filtered to only remove the low intensity data points the R^2 was 0.79, which given the type and quantity of data is exceptional.

The first test of the chromatin array was its ability to demonstrate the expected inverse relationship between chromatin condensation state and expression level (Foe 1978; Janicki, Tsukamoto et al. 2004). From figure 3.6, it can be seen that the condensation state of the genes (expressed as a ratio between the S2 fraction signal and total genomic DNA signal) shows the expected inverse correlation to the relative expression level with a linear regression line slope of -0.3 and a Pearson correlation of -0.16. While the inverse correlation was weaker than expected, the weakness can be traced to heterogeneity in the population of cells. This heterogeneity results from a subpopulation of cells that, in response to cell cycle or environmental effects, have

differing chromatin states and expression levels. The heterogeneity visible in the chromatin array data indicates a continuum of accessibility states, as opposed to a series of discrete steps. Because of the lack of inherent divisions in the chromatin array data, it is necessary to interpret the chromatin array results largely in the context of other data types, such as expression data.

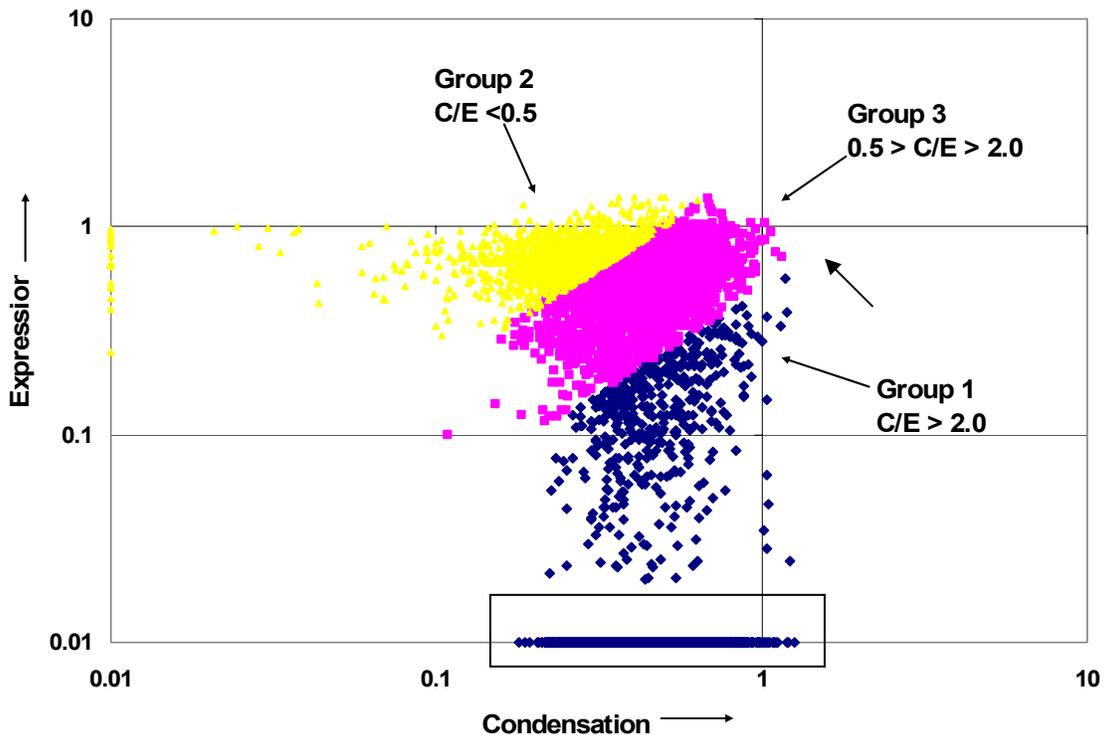


Figure 3.6: The log scale scatter plot of the co-normalized intensities (sequence abundances) illustrates the relationship between condensation (the ratio of the S2 fraction intensity to total genomic DNA intensity) and the absolute RNA expression. The yellow triangles are group 2 which has a condensation to expression (C/E) ratio of <0.5 . The pink squares represent group 3 which has indeterminate accessibility and a C/E ratio of <2.0 to >0.5 . The purple diamonds represent group 1 which has a C/E ratio of >2.0 . The box (bottom center) denotes the genes that exhibit differential chromatin compaction but are expressed at such low levels, expression data is not informative.

3.2.3 Cross validation of the two chromatin state fractionation techniques

A high throughput validation of the microarray results produced by the chromatin fractionation methods was not possible using traditional means. However, it was possible to evaluate the two chromatin fractionation techniques on the same microarray designs, thus allowing the reliability of the two techniques to be estimated. To accomplish this, both the chromatin fractionation and fragment length selection were carried out on MCF7 cells and the results compared. In order to compare the two methods the fragment length method was optimized to ensure that the fragment lengths selected for recovery would be equivalent to the S2 fraction (figure 3.3)

The microarray assays and data analysis of the fragment length-selected samples were performed as described in section 3.4. Genes that showed a 3-fold intensity change with respect to total genomic DNA reference were defined as the condensed (for genes with increased intensity relative to genomic levels) or relaxed (for genes with decreased spot intensity relative to genomic levels) sets, respectively. The gene lists from the fragment length selection process were merged with the chromatin solubility assay results for comparison (table 3.2.) 84% of the condensed genes correlated between the two chromatin state separation assays. The genes classified as relaxed by the two methods showed a 53% concordance based on raw intensity comparisons. It was anticipated that the fragment length selection method would be more error-prone for relaxed genes, since the recovery method used for the fragment selection process was not optimized for the separation of genes exhibiting an indeterminate accessibility from those that are most readily accessible (relaxed). The strong correlation (84%) between the two chromatin

state separation techniques demonstrates that the chromatin array concept based on chromatin fractionation is a robust method for chromatin condensation state analysis, given the data volume. As with any high throughput method, the results of the chromatin array method at the level of individual genes of interest should be validated by an independent method, before biological conclusions are drawn.

3.3 Insight into transcriptional regulation revealed by global high resolution chromatin studies

Since the chromatin condensation state results alone do not provide a means to delimit distinct chromatin states with certainty, the relationship between gene condensation and transcript abundance was used to further interpret the data and confirm the expected inverse relationship between chromatin compaction and gene expression (Foe 1978; Janicki, Tsukamoto et al. 2004). By using the ratio of the condensation state (as measured by relative intensity) to the RNA expression and then delimiting the groups by the standard 2-fold difference in intensity as the minimum statistically significant change, the data can be separated into three groups with a high degree of confidence in the assignment (figure 3.6). Group 1 (2,260 genes) has a condensation/expression (C/E) ratio of >2.0 making it the most condensed and expressed at low levels. Group 2 (1,493 genes) is the least condensed and most highly expressed with a C/E ratio <0.5 . Finally, Group 3 (3,842 genes) contains a C/E ratio between <2.0 to >0.5 with indeterminate accessibility (table 3.3). These groups represent the relationship between the chromatin state and gene expression, however because the number of genes in each group is so large, computer aided analysis and interpretation are required.

3.3.1 Combining chromatin array and expression data increases the robustness of the functional classification

Gene Ontology (GO) classifiers are an excellent means to identify and interpret, individual, statistically significant, biological relationships from large groups of data (Ashburner, Ball et al. 2000). Using the methods discussed in 3.4, the three groups were sub-classified so that functionalities that were over-represented within a group could be identified and interpreted. To demonstrate the synergistic effects of combining chromatin state data and expression data, the GO analysis was repeated using the groups (high and low expression) found in expression data alone. Table 3.3 shows that the addition of the chromatin array data substantially improved the significance scores for each identified term. The condensed group (2,620 genes, figure 3.6, purple diamonds) showed a condensation to expression ratio of >2.0 with 86% reproducibility between replicate fractionations on the different microarray types. Within the group with high condensation, the genes had a statistically significant ($p < 1.4E-2$) overlap with eight major GO terms (table 3.3) including “DNA repair”, a function associated with oncogenesis.

The relaxed group (1,493 genes, figure 3.6, yellow triangles) has a condensation to expression ratio of <0.5 . The reproducibility between replicate fractionations on different styles of arrays for this group was 76%. With the genes that have known pseudogenes excluded by the filtering, no GO term could be statistically associated with this group. The analysis was rerun with the 192 previously excluded genes having pseudogenic counterparts. This time the analysis found that several statistically significant GO terms were represented by this group (with $p < 1.8E-2$), most notably the

ribosomal genes and RNA binding genes (table 3.3). Both of these classes are constitutively expressed housekeeping genes and are known to be excluded from the S2 fraction (table 3.1) (Davis, Reudelhuber et al. 1983; Zhang, Harrison et al. 2002).

From the divergence seen in the GO terms that dominate the condensed and relaxed sets, clear differences in the gene functional types in each fraction can be identified. The division of the GO terms goes deeper, since while a number of signal transduction genes are in the relaxed fraction and are indeed transcribed (including *WNT7* which is one of primary oncogenes in MCF7), the WNT receptor signaling group is associated with the Group 1 and condensed genes (table 3.3) (Huguet, McMahon et al. 1994). It is these differences that may define the unique oncogenic signature of a cell line (Huguet, McMahon et al. 1994).

The last group of 3,842 genes (figure 3.6, pink squares) displays indeterminate accessibility with a condensation to expression ratio of <2.0 but >0.5 . This group showed 69% reproducibility between replicate fractionations on different styles of microarrays. The genes in this group cannot be identified from expression data alone, so in table 3.3 the “significance E” column is empty. However, the sample heterogeneity highlighted by the condensation data (section 3.2.2) is also apparent in the expression data as well. Because of heterogeneity, the genes represented by group 3 are likely a source of a large amount of known and unknown error in expression analysis. Based on closer inspection of the relationship between condensation and RNA expression, this group was determined to consist of three subgroups. The first of these subgroups included 1,166 genes that showed poor signal reliability, since the signal was marginally above threshold in one sample and below threshold in the other. The second subgroup of 1,412 genes

showed a significant change in raw signal, but intensity normalization converted the change to less than two-fold so they were excluded from further analysis. The last subgroup was the truly indeterminate set and had 1,264 genes that were more condensed more than their expression state indicated. 458 of these genes were involved in “cellular maintenance” (which is associated with genes such the *ABCA8* transporter) with a statistically significant ($p < 1.0E-3$) association to several GO terms (table 3.3.) Of the remaining 806 genes, 428 genes had no annotated function so they could not be considered further. The remaining 378 showed a statistically significant ($p < 8.5E-4$) overlap with the GO term “nucleic acid binding”. From these results it was confirmed that this subgroup of genes for which accessibility is indeterminate were not errors in the method but represent the genes that were heterogeneously expressed in the population due to cell cycle timing, environmental, or genetic factors like subclones in the population or variation in plating density.

Table 3.3: The functional signatures of the primary groups

GO Term Name	# Overlapping	Genes in Term	Significance C/E	Significance E
Group 1 C/E >2 (2,620)				
structural proteins	316	1576	1.21E-11	3.94E-10
plasma membrane	247	1192	3.15E-10	1.56E-08
transporter	268	1462	2.04E-05	1.13E-05
metabolism	168	856	2.32E-05	4.84E-05
signal transduction	525	1142	2.06E-04	7.00E-03
transcription factor	154	856	1.44E-02	1.56E-02
DNA repair	13	106	1.90E-02	-
Wnt receptor signaling*	5	12	6.72E-03	-
Group 2 C/E <0.5 (1,493)				
ribosome	59	139	6.11E-24	4.88E-19
RNA binding	72	374	5.53E-08	1.02E-02
cytosol	51	226	9.44E-08	1.02E-03
mitochondrial	65	459	1.78E-02	-
cell motility*	10	108	9.56E-03	-
Group 3 C/E <2 to >0.5 (3,842)				
signal transduction	252	1142	2.54E-82	-
serine threonine kinase	29	245	4.93E-10	-
proliferation	32	426	4.17E-06	-
cell adhesion	31	445	4.03E-05	-
cell cycle regulators	22	365	4.20E-03	-
nucleic acid binding	67	1925	8.50E-04	-

The Gene Ontology term names of the primary functional groups represented by the three major groups are shown. The number of genes that overlap with the term and the total number of genes in the term are shown for comparison. The significance value of the calculated relationship for each term from the combined dataset and from the expression data alone are given next. Terms with ‘-’ in the last column could not be found by expression alone. Significance was calculated from the overlap of the master list to the test list, and tested using a Bonferroni corrected hypergeometric distribution. Terms with asterisks were not identified as dominant terms in the initial search, but were found to be dominate sub-terms in that set.

3.3.2 Novel information provided by high-throughput global chromatin state measurements of synchronized cell culture to resolve heterogeneity issues

Genes with indeterminate accessibility (table 3.3 group 3), meaning they show a direct correlation between accessibility and expression, represent a significant portion of the genes. At first glance, the genes in this set are simply those that cannot be assigned to

a condensation group with high confidence. By filtering to remove the data from probes that cross-hybridize to non-functional elements like pseudogenes from the S2 fraction DNA and probes that are possible not specific in a CGH application, the remaining genes are likely to be heterogeneously expressed in a population of cells (Cho, Campbell et al. 1998; Su, Cooke et al. 2002). This analysis combined with directed experimentation has shown that this group has the potential to be a very rich source for new discoveries about chromatin's role in transcriptional control.

To test the assertion that heterogeneity in the sample is largely responsible for the genes with indeterminate accessibility (Group 3), the cell cycle state was synchronized using serum starvation. Serum starvation permits the capture of a synchronized (homogenous) population of cells for chromatin fractionation. Flow cytometry is employed to determine the fraction of cells in each phase of the cell cycle and to estimate the expected heterogeneity of the population (Darzynkiewicz and Juan 1997). 65% of the cells from the unsynchronized culture were in G1 and the subsequent chromatin fractionation experiment yielded approximately 32% of the DNA in the S2 fraction. By serum starving a culture for 24 hours and subsequently allowing six hours of recovery with 10% FBS before fractionation, over 80% of the population was in the G1 phase resulting in the S2 fraction being 42% of the recovered DNA. This change can be attributed to the fact that chromatin is in its most condensed form in the early G1 phase (Warters and Lyons 1992). The increase in the S2 fraction DNA caused a 16% increase in overall intensity after normalization of the synchronized chromatin array compared to the intensity of the chromatin array done on the unsynchronized sample. Of the genes that differ by at least two fold between the synchronized and unsynchronized samples, the GO

term “cell cycle regulators” (which includes the possible tumor suppressor *Fus1*) is significantly ($p < 6.0E-3$) overrepresented. The majority of these 61 cell cycle genes are measured as having more relaxed chromatin in synchronized cells, which correlates with the high expression levels measured in the unsynchronized culture. However, a small subset of genes go against this trend and show an increase in condensation, such as *CDKN2B*, a tumor suppressor involved in preventing cells from entering G1. Also of note is a strong increase in condensation of the mitotic motor *KNSL4*, which is only expressed in proliferating cells, indicating that the synchronized cells are not ready to divide. The results of the synchronization experiment indicate that by reducing a source of heterogeneity, the condensation state of the affected genes can be better correlated to the expression state. While the synchronization experiment showed the expected changes gene expression, in a manner that was predictable, 69% of genes that changed were not reported to be involved in cell cycle regulation. This means that cell cycle synchronization even by means as mild as serum starvation can alter the native chromatin state, and possibly introduce additional heterogeneity (Cooper 2003). For this reason cell cycle synchronization will not be used in any further chromatin array surveys.

3.3.3 Novel information resulting from chromatin state measurements of genes with expression state too low for microarray-based measurement

A majority of the chromatin array data can be interpreted in the context of other types of data; however, there is a subset of data that is novel to chromatin array data alone. The genes in the box in figure 3.6 represent information that can only be obtained using chromatin array data. 152 genes can be monitored by their chromatin state in spite of the fact that their expression level is below what can be measured using traditional

expression microarray technology. However, this number is artificially low because the data quality filters remove a significant portion of the genes whose expression is below detection. By turning off only the data filters for the expression measurements, the number of genes for which the chromatin state is measurable and reproducible, but show no expression above the calculated noise threshold, is increased to 1079. This figure represents an almost 14% increase in the amount of information that can be extracted from the chromatin array data, however even from the 152, significant discoveries are possible.

Examination of the 152 genes provides further evidence that inclusion of a gene on the list may indicate biological significance, since many of the genes are tissue-specific transcripts, including the neural specific *MAOB* and the hair specific keratin *KRTHA3B*. However, many of the genes on the list represent a significant opportunity for discovery since the telomerase gene *hTert* is included, as is the gene responsible for Ellis van Creveld syndrome, *EVC* (Atasu and Biren 2000). In MCF7, it is known that *hTert* is expressed at levels too low to be measured by expression microarray (Szutorisz, Lingner et al. 2003), which is confirmed by this experiment. From the chromatin array data it is clear that the gene maybe potentially expressible, with a chromatin state that is neither highly condensed nor highly relaxed. This data can then be compared to other chromatin array data, and the “express-ability” of *hTert* can be monitored without the transcript ever being detectible on expression microarrays.

In addition, because expression state is not needed to compare between chromatin states, promoters and other elements that are outside the transcribed region can be studied with the chromatin array. Using this technology, it should be possible to study the

function of insulator and other types of chromatin structural elements that are important in transcriptional regulation but, well outside the transcribed region (Gombert, Farris et al. 2003).

3.3.4 The chromatin state data provides significant insight into the relationship between chromosomal alteration and transcription

Using the publicly available aCGH data for MCF7 from the Breast aCGH project (Pollack, Sorlie et al. 2002), a map of gene amplifications and deletions in MCF7 was built. This dataset was chosen to be merged with our data, due to its wide availability and rigorous review. By overlaying this data on the chromatin array data, it was possible to evaluate the impact of these altered regions on chromatin condensation. In regions that were amplified, the genes were found to fall into two groups. The first group, containing genes that were highly condensed with low expression, was characterized by the GO terms in table 3.4 with $p < 2.6E-4$. Conversely, the second group had genes that were relaxed and showed high expression. This group was dominated by the GO terms “ribosomal” and “molecular chaperone,” which are broad groups of “housekeeping” genes (table 3.4.)

Table 3.4: The functional signatures of chromosomal alterations measured by accessibility state

GO Term Name	# Overlapping	Genes in Term	Significance
Amplified with increased condensation			
cell communication	147	3844	4.68E-09
cell growth and maintenance	131	3385	5.11E-08
transcription factors	46	926	2.57E-04
Amplified with decreased condensation			
ribosome	6	93	4.39E-03
molecular chaperone	5	96	1.02E-02
Copy number loss with increased condensation			
cell growth and maintenance	49	3385	6.64E-05
cell communication	52	3844	2.04E-04
tumor suppressor	5	177	2.87E-03

The Gene Ontology term names of the functional signatures represented in the CGH data as measured by accessibility state. Also shown is the number overlapping with the term and the total genes in the term with the significance values. Significance was calculated from the overlap of the master list to the test list, and tested using a Bonferroni corrected hypergeometric distribution.

Genes having aCGH measured copy number loss along with high condensation and weak expression, map to the GO terms “cell growth and maintenance”, “cell communication” and “tumor suppressors” with p values less than 2.9E-3 (table 3.4).

These terms include the gene functions traditionally associated with transformation, immortalization, and general oncogenesis, and possibly represent a substantial portion of the tumor-specific signature of MCF7 (Pollack, Sorlie et al. 2002). The genes that both showed a copy number loss as seen by aCGH and had reduced condensation (but strong expression) did not map to any GO terms with meaningful significance.

The aCGH data also yielded one additional relationship that was not expected. Genes found in the indeterminate accessibility group (table 3.3, group 3) showed strong ties to aCGH copy number changes including both copy number loss (p value =3.0E-8) and gain (p value = 9.5E-8). The 1,166 genes from Group 3 that were marginally above

threshold strongly overlapped (p value = $4.5E-25$) with amplified regions. This indicates that chromosomal abnormalities are another potential source of heterogeneity in the population. These results, determined from the chromatin array data, might represent a means to identify regions where chromosomal changes are not clonal in the population.

It is known that in cancer the alteration of the gene copy number can modify transcriptional activity to provide a selective advantage (Albertson, Collins et al. 2003). However, the regions that are altered are often large and contain genes that could be detrimental if deregulated, so additional regulatory mechanisms are required to provide fine control of these regions (Pollack, Sorlie et al. 2002; Albertson, Collins et al. 2003). During the aCGH and accessibility analysis, numerous regions of copy number change that shared a similar condensation state were identified. By overlaying the chromatin array data on the aCGH data it was possible to build high resolution maps of the differential chromatin states, and therefore regulatory regions. These results show that amplified and condensed regions, sometimes only a few hundred Kb wide, exhibiting low expression were interspersed at chromosomal positions 11q13, 7q21.11 to 7q22, 5q31 to 5q32 and 1q31.1 to 1q32. These regions contain a number of tumor suppressions including *DNAJC4* in 11q13 and *PPP2A* in 5q31. Amplified regions that primarily exhibit low condensation and increased expression were also seen. The 5q11 to 5q12 region is the most tantalizing example of this relationship, with the gene of greatest interest in that region being *DHFR* (Sullivan and Bickmore 2000), which by amplification can confer resistance to the chemotherapeutic compound methotrexate (Banerjee, Mayer-Kuckuk et al. 2002). Instances of low condensation and high expression from regions with copy number loss were also seen and include the 22q13.1

region, which is the location of the gene *H1FO*, a histone 1 variant known to be involved in maintaining basal gene repression and differentiation. The regions with copy number loss, high condensation, and low expression were also numerous and include the 3p21 region, which contains numerous tumor suppressor genes including *IHPK2*, and in the 22q13.2 region, the location of the tumor suppressor *RBX1*.

In conclusion, the use of chromatin condensation information to link the chromosomal state to gene expression level could provide a better understanding of diseases like cancer where large, enigmatic chromosomal alterations occur long before the selection that would make such changes oncogenic (Albertson, Collins et al. 2003). It is this ability to make transcriptional discoveries by mapping accessibility changes to transcriptional control of genes or regions that makes the chromatin array method powerful.

3.4 Materials and methods

3.4.1 Cell culture

MCF7 (American Type Culture Collection number HTB-22), which is a female breast cancer model line, was selected because it has been well studied. The cells were grown in RPMI 1640 media with 10% fetal bovine serum (FBS) to 90% confluence. Harvesting was done using 0.05% trypsin with two washes in sterile 4° C phosphate buffered saline. The harvested cells were counted and aliquoted for nuclease digestion. The remaining cells were pelleted and flash frozen for DNA or RNA extraction.

3.4.2 Cell cycle synchronization by serum starvation and population analysis

For the synchronized cultures, the cells were allowed to grow to 80% confluence under normal conditions. Then the media was removed and the cells washed twice in RPMI without FBS. The washed cells were grown in RPMI alone for 24 hours to serum starve and synchronize. The media was then replaced with RPMI supplemented with 10% FBS and the cells allowed to recover for 6 hours before harvesting for nuclease digestion. An aliquot was ethanol fixed according to the protocol developed by Darzynkiewicz, and stained with propidium iodide (Darzynkiewicz, Bruno et al. 1992). The stained cells were analyzed by flow cytometry to determine the proportion of cells in each phase of the cell cycle.

3.4.3 DNase I digestion and fragment length selection

The DNase I digestion protocol was based on Roque, et al, (Roque, Smith et al. 1996). Cells were permeabilized by 0.11% lysolecithin and incubated for 10 minutes in the presence of DNase I at concentrations up to 64U/2.0x10⁷ cells. The DNA was purified as described above and 15 µg of each sample was separated electrophoretically on 0.8% Seakem GTG agarose (BMA, Rockland, ME) gels in 1X TAE for 10 hours at 60V and then stained with ethidium bromide and visualized under UV light. The band that represents the sample with the uniform distribution of fragments from 23 kb to 1 kb with a median fragment length around 5 kb was selected from the DNase I concentration series (figure3.3). The upper portion of the lane containing 40% of the total DNA by intensity was then excised. Next, the sample was cut into pieces before being frozen at -20° C, snap thawed at 37° C, and placed in tubes with 0.22 micron filters (Gelman). The

samples were eluted by centrifugation at 13,000 x g for 10 minutes. The elutant was concentrated using a 30,000 Dalton centrifuge filter (Millipore) and resuspended in 20 μ L nuclease free water.

3.4.4 Micrococcal nuclease digestion and the chromatin fractionation assay

The chromatin fractionation assay was preformed as described by Rose and Garrard (Rose and Garrard 1984). Cell were lysed by incubation with a hypotonic buffer supplemented with non-ionic detergent NP-40, the nuclei were washed, resuspended in digestion buffer, and incubated with micrococcal nuclease to about 10% acid solubility. The nuclei were then centrifuged to obtain the S1 fraction, which represents the chromatin that was soluble in a buffer containing 5 mM MgCl₂ and 100 mM KCl. The nuclei were then resuspended in 2 mM EDTA for lysis and the S2 fraction was recovered in the supernatant following centrifugation. The pellet that remains after removal of the S2 fraction contained the P fraction. The DNA was purified from the samples by digestion with Proteinase K, extraction with phenol/chloroform, and ethanol-precipitation. DNA was re-suspended in Tris/EDTA pH 8.0, treated with RNase and re-purified as above.

3.4.5 Microarray characteristics

The microarrays initially used were produced by the UTSW microarray core and were spotted 70mer oligonucleotides from Operon (Qiagen). The arrays were printed on poly-L-Lysine coated slides and consisted of ~22,000 Unigene clusters (Schuler, Boguski et al. 1996) and ~2,000 blanks (negative controls), in addition to other control spots

including organism specific (*Arabidopsis*) probes. The hybridization was done at 62°C using a recirculating waterbath, in single slide hybridization chambers (Telechem, Sunnyvale, CA), under 24x60 mm coverslips. The chromatin solubility assay experiments were repeated on the MWG Human 30K oligonucleotide microarrays (MWG Biotech AG, High Point, NC) using only the A and B arrays for a total unique gene count of 19,473.

3.4.6 DNA labeling

Fluorescent labeling was achieved by random priming according to the protocol devised by Pollock et al. for aCGH and optimized for use as a genomic standard (Pollack, Perou et al. 1999; Weil, Macatee et al. 2002). Direct labeling of the sample was accomplished by the incorporation of Cy5 fluorophore (Amersham Biosciences) into ~2 µg of sample. An undigested control sample of total genomic DNA from the same cell line and at the same concentration was labeled with the Cy3 fluorophore (Amersham Biosciences, Piscataway, NJ) for use as the genomic reference. Sample purification and blocking was performed as specified in the protocol developed by Weil et al. (Weil, Macatee et al. 2002).

3.4.7 Chromatin hybridization

Sample hybridization was performed using the optimized protocol (Weil, Macatee et al. 2002) with modifications described here. The hybridization temperature was adjusted according to the MWG protocols to 42°C. Both types of microarrays were hybridized for 16-20 hours prior to washing. The three-buffer washing process is

standard, with the first buffer (0.2% SDS and 2X SSC) warmed to 30°C. The second buffer was 0.2X SSC and the third buffer was 0.1X SSC. The microarray was then dried by centrifugation. The dry microarray was scanned in an Axon 4000B scanner at a resolution of 10 microns/pixel with adjustments to the photomultiplier tube voltage used to balance the signal in the two channels.

3.4.8 RNA extraction and hybridization

RNA was extracted from the reserved snap frozen cells using the RNeasy kit (Qiagen). The resulting RNA was quantified and quality checked by the Bioanalyzer (Agilent). The RNA (20 µg each) was labeled with the Cyscribe kit (Amersham) using Cy3 or Cy5 fluorophores. By replicating the same RNA in both channels, the dye dependent variables were removed and each array produced two data points for each gene. The RNA was then degraded by the addition of a sodium hydroxide solution at 70° C for 10 minutes and the samples were subsequently neutralized with an equal amount of hydrochloric acid. The samples were washed as before and the appropriate blocking agents and buffers were added. Hybridization and washing conditions were similar to the conditions used for the chromatin microarrays (Weil, Macatee et al. 2002).

3.4.9 Data extraction, normalization and analysis

Using the GenePix 4.0 analysis package (Axon), the resulting images were overlaid with a grid to define the spots and the data was manually flagged to remove poor quality spots (Fielden, Halgren et al. 2002). The data was extracted from the seven chromatin microarrays and the six expression repeats and imported into GeneSpring 6.0

(Silicon Genetics) for normalization and analysis. The normalization of the chromatin microarrays consisted of dividing the S2 fraction signal by the genomic DNA control signal. The expression data was normalized to the median of the genes across the experimental series. The cross-gene error model was used to remove any genes that do not have sufficient signal to exceed the noise threshold of 125 intensity units calculated by the cross gene error model in the experimental series (Agilent 2006). To minimize the inaccuracies introduced by pseudogene cross hybridization, all genes that are known to have a pseudogene (Zhang, Harrison et al. 2003) were removed from the primary analysis.

3.4.10 Data interpretation

The GO annotation was done using the “find similar lists” function in GeneSpring. The Simplified Gene Ontology lists used represent only the subset of genes in each term that is actually probed for on the microarray because the lists are generated by parsing the gene annotation file. The similarity was determined using a Bonferroni corrected hypergeometric distribution test to find the significance of the overlap of the user generated (or master list) to all the possible Simplified Gene Ontology lists (or test lists). Only statistically significant lists with p values of approximately 0.01 were selected for a manual review of the content before being included as an overlapping term.

The expression results were verified for a subset of genes in each group using the SOURCE database (Diehn, Sherlock et al. 2003) to annotate the experimental data with data from published array experiments (Sherlock, Hernandez-Boussard et al. 2001), serial analysis of gene expression (SAGE) data (Lash, Tolstoshev et al. 2000), and sequence

data from Ensembl (Hubbard, Barker et al. 2002). The accessibility and expression data for this experimental series is stored in a MIAME-compliant database (Brazma, Hingamp et al. 2001) that can be accessed from the web site at <http://famine.swmed.edu>.

3.5 References

- Agilent (2006) "Cross-Gene error model " **Volume**, DOI:
- Albertson, D. G., C. Collins, et al. (2003). "Chromosome aberrations in solid tumors." Nat Genet **34**(4): 369-76.
- Allan, J., G. J. Cowling, et al. (1981). "Regulation of the higher-order structure of chromatin by histones H1 and H5." J Cell Biol **90**(2): 279-88.
- Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-9.
- Atasu, M. and S. Biren (2000). "Ellis-van Creveld syndrome: dental, clinical, genetic and dermatoglyphic findings of a case." J Clin Pediatr Dent **24**(2): 141-5.
- Banerjee, D., P. Mayer-Kuckuk, et al. (2002). "Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase." Biochim Biophys Acta **1587**(2-3): 164-73.
- Brazma, A., P. Hingamp, et al. (2001). "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data." Nat Genet **29**(4): 365-71.
- Cho, R. J., M. J. Campbell, et al. (1998). "A genome-wide transcriptional analysis of the mitotic cell cycle." Mol Cell **2**(1): 65-73.
- Cooper, S. (2003). "Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points." Faseb J **17**(3): 333-40.
- Croston, G. E., L. A. Kerrigan, et al. (1991). "Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription." Science **251**(4994): 643-9.
- Darzynkiewicz, Z., S. Bruno, et al. (1992). "Features of apoptotic cells measured by flow cytometry." Cytometry **13**(8): 795-808.
- Darzynkiewicz, Z. and G. Juan (1997). *Current Protocols in Cytometry*. J. P. Robinson. NY, J. Wiley & Sons: 7.5.1-7.5.24.
- Davis, A. H., T. L. Reudelhuber, et al. (1983). "Varigated chromatin structures of mouse ribosomal RNA genes." J Mol Biol **167**(1): 133-55.
- Diehn, M., G. Sherlock, et al. (2003). "SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data." Nucleic Acids Res **31**(1): 219-23.

- Draghici, S., P. Khatri, et al. (2005). "Reliability and reproducibility issues in DNA microarray measurements." Trends Genet.
- El-Osta, A. and A. P. Wolffe (2000). "DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease." Gene Expr **9**(1-2): 63-75.
- Esteller, M. and J. G. Herman (2002). "Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours." J Pathol **196**(1): 1-7.
- Fielden, M. R., R. G. Halgren, et al. (2002). "GP3: GenePix post-processing program for automated analysis of raw microarray data." Bioinformatics **18**(5): 771-3.
- Foe, V. E. (1978). "Modulation of ribosomal RNA synthesis in *Oncopeltus fasciatus*: an electron microscopic study of the relationship between changes in chromatin structure and transcriptional activity." Cold Spring Harb Symp Quant Biol **42 Pt 2**: 723-40.
- Gombert, W. M., S. D. Farris, et al. (2003). "The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain." Mol Cell Biol **23**(24): 9338-48.
- Griesenbeck, J., H. Boeger, et al. (2004). "Purification of defined chromosomal domains." Methods Enzymol **375**: 170-8.
- Huang, S. Y., M. B. Barnard, et al. (1986). "The active immunoglobulin kappa chain gene is packaged by non-ubiquitin-conjugated nucleosomes." Proc Natl Acad Sci U S A **83**(11): 3738-42.
- Hubbard, T., D. Barker, et al. (2002). "The Ensembl genome database project." Nucleic Acids Res **30**(1): 38-41.
- Huguet, E. L., J. A. McMahon, et al. (1994). "Differential expression of human Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue." Cancer Res **54**(10): 2615-21.
- Janicki, S. M., T. Tsukamoto, et al. (2004). "From silencing to gene expression: real-time analysis in single cells." Cell **116**(5): 683-98.
- Jarvinen, A. K., S. Hautaniemi, et al. (2004). "Are data from different gene expression microarray platforms comparable?" Genomics **83**(6): 1164-8.
- Jones, P. A. and R. Martienssen (2005). "A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop." Cancer Res **65**(24): 11241-6.
- Kane, M. D., T. A. Jatkoe, et al. (2000). "Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays." Nucleic Acids Res **28**(22): 4552-7.
- Lash, A. E., C. M. Tolstoshev, et al. (2000). "SAGEmap: a public gene expression resource." Genome Res **10**(7): 1051-60.
- Pollack, J. R., C. M. Perou, et al. (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet **23**(1): 41-6.
- Pollack, J. R., T. Sorlie, et al. (2002). "Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors." Proc Natl Acad Sci U S A **99**(20): 12963-8.
- Rauscher, F. J., 3rd (2005). "It is time for a Human Epigenome Project." Cancer Res **65**(24): 11229.
- Ricke, R. M. and A. K. Bielinsky (2005). "Easy detection of chromatin binding proteins by the Histone Association Assay." Biol Proced Online **7**: 60-9.

- Robertson, K. D. (2005). "DNA methylation and human disease." Nat Rev Genet **6**(8): 597-610.
- Roque, M. C., P. A. Smith, et al. (1996). "A developmentally modulated chromatin structure at the mouse immunoglobulin kappa 3' enhancer." Mol Cell Biol **16**(6): 3138-55.
- Rose, S. M. and W. T. Garrard (1984). "Differentiation-dependent chromatin alterations precede and accompany transcription of immunoglobulin light chain genes." J Biol Chem **259**(13): 8534-44.
- Scarano, M. I., M. Strazzullo, et al. (2005). "DNA methylation 40 years later: Its role in human health and disease." J Cell Physiol **204**(1): 21-35.
- Schuler, G. D., M. S. Boguski, et al. (1996). "A gene map of the human genome." Science **274**(5287): 540-6.
- Sherlock, G., T. Hernandez-Boussard, et al. (2001). "The Stanford Microarray Database." Nucleic Acids Res **29**(1): 152-5.
- Su, A. I., M. P. Cooke, et al. (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-70.
- Sullivan, B. A. and W. A. Bickmore (2000). "Unusual chromosome architecture and behaviour at an HSR." Chromosoma **109**(3): 181-9.
- Szutorisz, H., J. Lingner, et al. (2003). "A chromosome 3-encoded repressor of the human telomerase reverse transcriptase (hTERT) gene controls the state of hTERT chromatin." Cancer Res **63**(3): 689-95.
- Warters, R. L. and B. W. Lyons (1992). "Variation in radiation-induced formation of DNA double-strand breaks as a function of chromatin structure." Radiat Res **130**(3): 309-18.
- Weil, M. R., T. Macatee, et al. (2002). "Toward a universal standard: comparing two methods for standardizing spotted microarray data." Biotechniques **32**(6): 1310-4.
- Weintraub, H. and M. Groudine (1976). "Chromosomal subunits in active genes have an altered conformation." Science **193**(4256): 848-56.
- Widlak, P. (2004). "DNA microarrays, a novel approach in studies of chromatin structure." Acta Biochim Pol **51**(1): 1-8.
- Zardawi, I. M. and J. Duncan (2003). "Evaluation of a centrifuge method and thin-layer preparation in urine cytology." Acta Cytol **47**(6): 1038-42.
- Zhang, Z., P. Harrison, et al. (2002). "Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome." Genome Res **12**(10): 1466-82.
- Zhang, Z., P. M. Harrison, et al. (2003). "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." Genome Res **13**(12): 2541-58.

Chapter 4

Studies of alterations in the chromatin state of the MCF7 cancer model cell line induced by the drug 5-Aza-2'-Deoxycytidine

4.1 Selection of an epigenetic modifying drug

In the previous chapter most of the chromatin state data was reported as the ratio of condensation to expression. Expression data must be used because the chromatin state data has to be referenced against something in order to interpret the results. The cell cycle synchronization experiments discussed in 3.3.2 demonstrated that chromatin state to chromatin state comparisons could be interpreted without additional data. These experiments were just the first step, since the chromatin state differences represented the minority of data. In order to really determine the value of direct chromatin state comparisons, and then exploit them, the chromatin states being compared would need to be globally different and able to be interpreted in a biological context.

In order to accomplish this comparison, either different cell lines can be compared or a single cell line can be compared to itself after a chromatin state modifying treatment. When comparing different cell lines there is no “normal” condition to use as a reference, so the differences may be difficult to place in context. Conversely, with treatment-induced changes, the untreated chromatin state can be used as a reference state for comparison. Therefore, any differences measured between the chromatin states of each sample can be attributed to the treatment, which greatly facilitates interpretation. Given the difficulties with inter-cell line comparisons it was clear that treatment induced chromatin state change was the superior choice for validating direct chromatin state comparison on a large scale. However, the selection of a treatment method to systemically modify the chromatin state still required significant consideration.

There are numerous ways to globally modify chromatin state, ranging from interfering RNA-based suppression of the DNA methyltransferases (DNMTs) to exposure to ionizing radiation, but the best characterized are epigenetic modifying drugs (Warters and Lyons 1992; Gius, Cui et al. 2004). The two most common classes of epigenetic modifying drugs are DNA methylation inhibitors (i.e. 5-Aza-2'-deoxyCytidine (5 Aza dC) and zebularine) and histone deacetylase (HDAC) inhibitors (i.e. Sodium Butyrate and Trichostatin A (TSA) (Cameron, Bachman et al. 1999; Cheng, Matsen et al. 2003; Momparler 2005).

As discussed in section 1.2.2, histone acetylation is part of a key regulatory process that is involved in chromatin relaxation, therefore making the DNA available for transcription. HDAC inhibitors prevent the condensation of relaxed chromatin and therefore can globally alter the chromatin state (Cameron, Bachman et al. 1999). Of the two primary histone deacetylation inhibitors, toxicity concerns excluded the use of Sodium Butyrate and it had been shown that Trichostatin A was unlikely to produce significant change alone, because it cannot relax condensed chromatin (Shi, Wei et al. 2003). As a result, a treatment from the class of DNA methylation inhibitors was chosen for the chromatin state study.

In mammals, the most commonly methylated DNA base motif is the C in a CpG dinucleotide (Bird 1980). The DNMTs use S-adenosyl methionine as a methyl donor to transfer the methyl group to the 5 position of the cytidine ring, in a reaction that covalently bonds the enzyme to the substrate (figure 4.1a) (Gama-Sosa, Slagel et al. 1983). In most of the genome the CpG dinucleotide is actually found at levels far below expected if the distribution was random, but in certain regions called CpG islands it is

found at levels far above what is expected (McClelland and Ivarie 1982; Gardiner-Garden and Frommer 1987). CpG islands have long been used in gene finding, but it is the regulatory role that CpG islands play in transcriptional regulation that makes them interesting (Bird, Taggart et al. 1987; Larsen, Gundersen et al. 1992). CpG island methylation is mostly repressive on transcription, since methylated CpG islands have a condensed chromatin state, which denies the transcription factors access to the DNA (Bird 2002). CpG islands and CpG island methylation play important roles in transcriptional regulatory processes like imprinting, and aberrant methylation is associated with a number of diseases including cancer (Esteller and Herman 2002; Jones 2002; Robertson 2005). The rationale behind the use of DNA methylation inhibitors, for cancer treatment is that by reversing the aberrant methylation, the normal state can be restored or apoptosis induced. Furthermore unlike with traditional chemotherapeutics which are designed to be directly cytotoxic, epigenetic therapies are designed to have targeted cytotoxicity (Bird 2002; Goffin and Eisenhauer 2002; de Vos 2005).

Many DNA methylation inhibitors are base analogs that cannot be methylated by the DNMTs, so they reduce methylation stoichiometrically with incorporation into the genome (Goffin and Eisenhauer 2002). The “Aza” class of DNA methylation inhibitors has nitrogen substituted for the 5 carbon of the pyrimidine ring (figure 4.1b) while, zebularine is missing the amine from the 4 position, and other classes have modifications at the 6 position or use a non-ribose sugar.

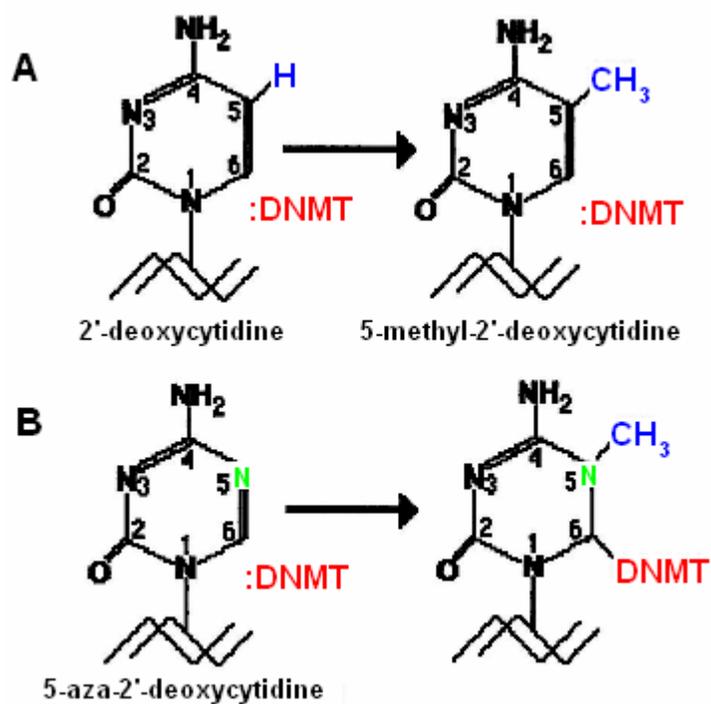


Figure 4.1: (A) A diagram showing the methylation of normal 2'-deoxycytidine with the methyl group added to the 5 position in presence of DNA methyl transferase (DNMT). (B) The structure of 5-aza-2'-deoxycytidine. Note that after methylation the transferase enzyme remains bound to the substrate. This sequesters the population of free enzyme and thus prevents DNA methylation at other sites in the genome.

While zebularine has shown great promise as a cancer therapeutic, because it has far fewer side effects compared to the 5 Aza drugs, it is expensive and difficult to obtain, as were many of the other non-Aza based inhibitors (Cheng, Matsen et al. 2003). Because the 5-Aza drugs have medical applications in the treatment of certain leukemia's and blood diseases, they are available cheaply and in large quantities (Lubbert and Minden 2005; Momparler 2005). The first 5 Aza drug was 5-Aza-cytidine (5 Aza C), but because it could be incorporated into RNA as well as DNA it was profoundly cytotoxic (Goffin and Eisenhauer 2002). To deal with this, the 2' deoxy-ribose form was chosen for most

medical applications, since its incorporation was restricted to DNA and 5 Aza dC proved less toxic (de Vos 2005). In addition, the drugs in the 5 Aza family have more than one mechanism of action, since they are not just base analogs, they are also “suicide substrates” for the DNMTs (figure 4.1b) (Oka, Meacham et al. 2005). This means that treatment with 5 Aza dC results not only in a reduction in the global CpG methylation level, but sequestration of the DNMTs and induction of double stranded breaks in the DNA (Momparler 2005). While the complex mechanism would seem to detract from the desirability of 5 Aza dC’s for use in this study, the drug has been used for over 20 years and there is a great deal of literature on its effects. By using the prior work as a guide, it was possible to correlate the results from the chromatin array to the results from previous studies and then make new observations about the model system and the effects of the drug.

4.2 Experimental details and validation

The goal of this experiment is to use the chromatin array and other complimentary techniques to determine the effects of 5 Aza dC treatment on the MCF7 cell line. While a full description of the treatment is provided in section 4.5, a brief description and discussion is warranted. The treatment of the MCF7 cells was carried out according to standard protocols but the size and number of plates was increased to produce the quantity of cells required (Bender, Pao et al. 1998). The MCF7 cells were seeded on to 150mm culture dishes at 500,000 cells per plate in RPMI containing 10% Fetal Bovine Serum. The cells were allowed to recover for 24 hours, before treatment. Seven plates received three 500nM treatments of 5 Aza dC, every other day. The control plates

received mock treatments. Twelve hours after final treatment the cells both treated and control were harvested and the cells within the groups were pooled. Several million cells from both the treated and control were flash frozen, so that RNA could be extracted for use in transcriptional studies. The remaining cells were fractionated to recover the chromatin for the chromatin state analysis, before and after treatment with 5 Aza dC.

In order to minimize the treatment-induced heterogeneity of the cellular population, high concentrations of 5 Aza dC were used to ensure that all cells were exposed to enough of the drug to induce global hypomethylation in MCF7, even though the drug has a very short half life (<15 minutes) in the cells (Ferguson, Vertino et al. 1997). The first concentration tested was 1,000nM 5 Aza dC; however, the nuclear membranes were so unstable that the chromatin fractionation failed. The concentration of 5 Aza dC used to treat the cells was reduced to 500nM, and number of cells used for the fractionation increased to account for the still diminished stability of the nuclear membrane, which allowed the treated cells to be fractionated at $\frac{3}{4}$ of the normal scale. The S1 fraction of the 5 Aza dC treated sample cleaved with 0.3 U of micrococcal nuclease contained 7% of the total recovered DNA (optimum = 10%) (Huang and Garrard 1986), so this fractionation was considered successful. 0.6U of micrococcal nuclease over-digested the chromatin and resulted in the S1 fraction containing only ~4% of total recovered DNA. Because the adequate recovery of the S2 fraction is largely independent of the amount of digestion, minor variations in the recovery of the S1 and P fractions do not bias the chromatin array results (Huang, Barnard et al. 1986).

The hybridization, data extraction and data analysis followed the methods developed and published by Weil et. al., and have been extensively discussed in the previous chapter (Weil, Widlak et al. 2004). However, the details of the arrays and experiment design are worth some additional discussion. The MWG Human 40K A microarray (MWG Biotech HighPoint, NC) was used for the chromatin arrays and some of expression array measurements. Additionally the Affymetrix Human 133 2+ GeneChip® (Mountain View, CA) was used to validate the expression measurements. By using two separate platforms for the expression measurement and using only data that reproduced across the platforms, the significance of the measurement is greatly increased (Woo, Affourtit et al. 2004). In order to measure alternative splicing in certain cases the MWG 40K array has more than one probe per gene. Using these probes it was also possible to measure chromatin state at a sub-gene level, in order to measure intragenic chromatin variation.

4.2.1 Examination of the microarray data variance

The first test of a microarray-based experiment is replicate reproducibility. Figure 4.2 is an example of the reproducibility of the chromatin array data. Even the raw data from replicate chromatin arrays hybridized with the 5 Aza dC treated samples show an excellent correlation with an $R^2=0.71$ (figure 4.2) (Hardiman 2004; Jarvinen, Hautaniemi et al. 2004; Engler, Mohapatra et al. 2006). To further ensure the results of this comparison were not influenced by external factors, the chromatin array data from the untreated MCF7 controls was compared with data from the chromatin arrays discussed in the section 3.2 (Weil, Widlak et al. 2004). The inter-platform and inter-fractionation

comparison ensures that the chromatin state measurements are reproducible. Because the microarrays were of different designs and therefore didn't have the same probe set, a direct comparison was not possible. However, indirect comparison is possible, if a conservative level of variance is allowed between the datasets (Jarvinen, Hautaniemi et al. 2004). By filtering for the common probes these arrays showed a good correlation, with more than 70% of the features showing a less than 35% variation in intensity between the data sets.

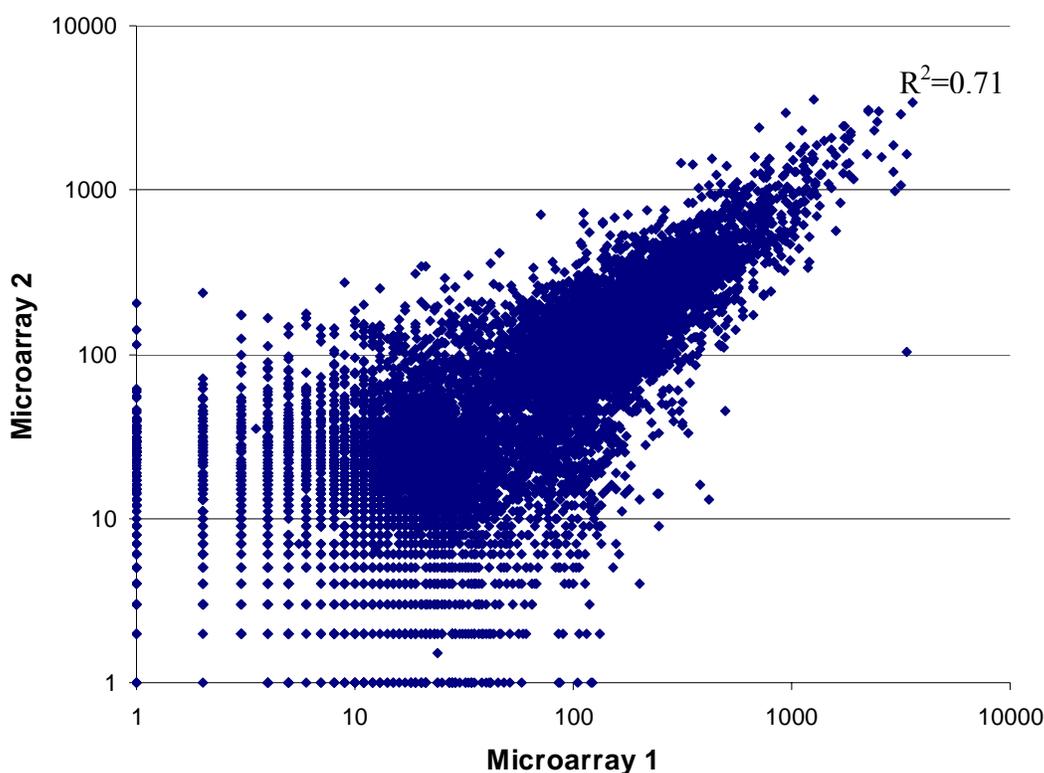


Figure 4.2: A scatter plot of the unfiltered raw intensity data (~20,000 spots) from the replicate chromatin arrays of the 5 Aza dC treated MCF7 cells. Replicate arrays are reproducible with an $R^2=0.71$.

Comparison of the expression data between the different MWG array types showed that greater than 80% of the genes had less than 35% variance in intensity among the probes common to the microarrays and internally reproducing. The final measure of

microarray reproducibility was the comparison between the MWG microarray data (expression) and the Affymetrix GeneChip® data. Because this comparison was across such radically different platforms, the spotted microarray data and the Affymetrix GeneChip® data were normalized and filtered to remove low-intensity features separately (Woo, Affourtit et al. 2004). Then the non-common features were excluded by filtering. Comparison of the remaining features showed that 77% showed a less than 35% variance in intensity across platforms. Once it was clear that the microarray portion of the experiment had yielded high quality data, it was still necessary to validate that the chromatin state changes induced by 5 Aza dC treatment could be measured using the chromatin array method.

4.2.2 Examination of genes known to be altered by 5 Aza dC treatment

Much known about the effects of 5 Aza dC treatment (including expression changes) on several of cell lines, including MCF7 (Gius, Cui et al. 2004; Shames 2005). By examining a subset of the genes known to be altered by 5 Aza dC treatment, and comparing the expected response to the observed response, it was possible to determine if the treatment was having the anticipated effect (figure 4.3) (Sherlock, Hernandez-Boussard et al. 2001; Diehn, Sherlock et al. 2003; Shames 2005). Furthermore, by looking at the treatment-induced chromatin condensation state changes for the same set of genes, it was possible to determine whether the effects of the treatment were manifested via changes in the chromatin state. Several cytidine kinases (including AK5), which increase the susceptibility of the cells to 5 Aza dC, show a decrease in expression (figure 4.3b) and a corresponding increase in condensation with treatment (figure 4.3a) (Momparler 2005). Additionally, several genes linked to increased 5 Aza dC resistance,

such as cytosine deaminases and several ABC transporters, showed treatment-induced chromatin relaxation (figure 4.3a), with reciprocal changes in expression (figure 4.3b). These results indicate the treatment induced the expected protective responses from the cells (Momparler 2005). In the set of genes where 5 Aza dC treatment induces change as a result of its direct epigenetic effects, the expected response was also seen, for example, in the increased relaxation of the *AZII* gene and the *MAGE* family of genes, together with increases in expression (figure 4.3) (Weber, Salgaller et al. 1994; Fujie, Mori et al. 1997; Furuta, Umebayashi et al. 2004).

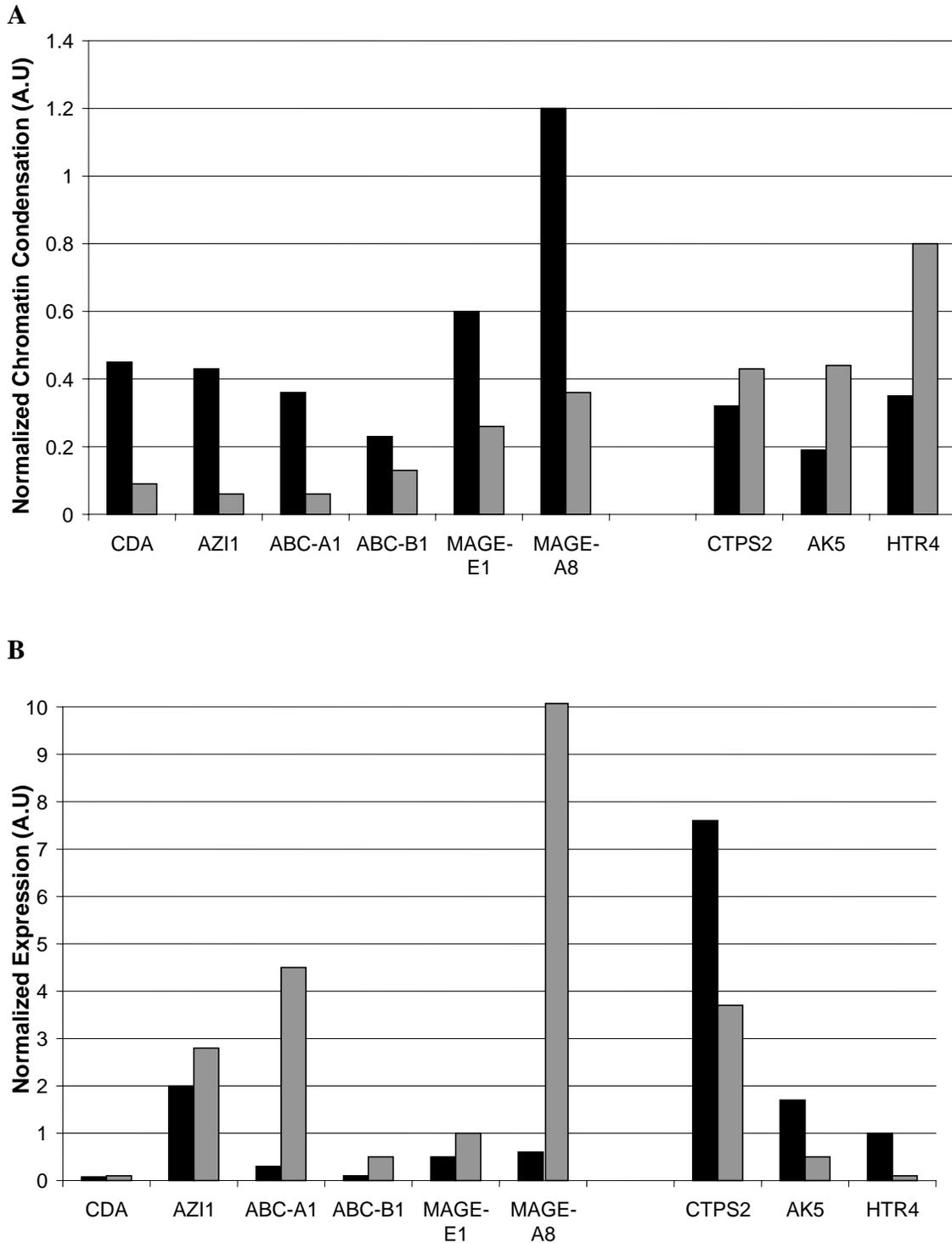


Figure 4.3 **A**: 5 Aza dC responsive genes exhibit chromatin state changes as measured by the chromatin array. **B**: The expression state of the 5 Aza dC responsive genes shows that there is a concordant change in expression, which is anticipated given the chromatin change. In both graphs the treated normalized intensity is shown in gray and the untreated normalized intensity is shown as black

4.3 The effects of 5 Aza dC treatment on the chromatin state of MCF7

4.3.1 Discoveries emerging from the experiment: Functional classification view

3,010 genes showed a greater than 2-fold increase in expression with 5 Aza dC (table 4.1a), with numerous Gene Ontologies (GO) (Ashburner, Ball et al. 2000) categories being over-represented ($p < 8.0E-3$). Inclusion of gene sets in categories such as ‘protein kinase cascade’, ‘cytoskeleton’, and ‘DNA binding’ (table 4.1a) was expected, and further suggested that the treatment altered the cells (Anderson and Gemmell 1984). However, these categories, and most of the remaining categories, were high-level ontological classifications. Because GO is a hierarchical classification system, high-level classifications are the least specific in regards to the nature of the function. For example the GO term ‘DNA binding’ contains all genes that are known to bind DNA regardless of the function or nature of the binding event.

In addition, 478 genes showed a greater than 2-fold decrease in expression after 5 Aza dC treatment but no GO categories were over-represented in the list (table 4.1b). The failure to find significant overlap suggests that this gene list contains a wide distribution of gene types.

Table 4.1: Functional classes based on expression change after 5 Aza dC treatment

A: Expression increased 2 fold (3010 genes)		
Gene Ontology	Significance	Genes overlapping %
protein kinase cascade	1.80E-27	31%
STAT cascade	3.10E-26	31%
cytoskeleton	8.50E-14	10%
golgi apparatus	3.70E-06	3%
cell cycle regulator	3.70E-05	5%
DNA binding	2.20E-05	3%
heat shock response	8.00E-04	1%
nucleosome	8.00E-03	1%
B: Expression decreased 2 fold (478 genes)		
Gene Ontology	Significance	Genes overlapping %
None	-	-

Using just the chromatin array results, 1,489 genes were more relaxed after 5 Aza dC treatment, and the GO categories associated with this list ($p < 5.3E-3$) included: 'oncogene', 'tumor suppressor', and 'cell proliferation' (table 4.2b). Among the tumor suppressors relaxed by treatment is the *TP53* binding partner *HIPK2*. These categories are more specific than the ones found with expression, but still highlight how the treatment globally affected chromatin state in the context of the cancer model system. 378 genes were condensed more than two-fold after treatment. Among the GO categories over-represented ($p < 4.8E-3$) in this group of genes were 'STAT cascade', 'metalloendopeptidase inhibitor', and 'caspase activation' (table 4.2a) (included in the genes with treatment induced condensation is the pro-proliferation gene *STAT1*). Because these categories represent lower-level classifications, they may be informative with regards to treatment effects. However, the discovery of a large set of genes that were condensed by the treatment was far more provocative, because 5 Aza dC treatment is almost synonymous with chromatin relaxation (Goffin and Eisenhauer 2002). Therefore

finding regions where chromatin condensation was increased by treatment was unanticipated and suggests that epigenetic regulation is even more complex than predicted by current models.

Table 4.2: Functional classes based on chromatin change after 5 Aza dC treatment

A: Chromatin condensation increased 2 fold (378 genes)

Gene Ontology	Significance	Genes overlapping %
STAT cascade	4.90E-10	45%
cell adhesion	1.90E-08	13%
metalloendopeptidase inhibitor	4.15E-08	3%
caspase activation	3.20E-07	3%
tyrosine phosphatase	5.00E-06	5%
Cathepsin S activity	1.70E-04	1%
extracellular matrix	2.40E-04	8%
DNA binding	4.80E-03	22%

B: Chromatin condensation decreased 2 fold (1489 genes)

Gene Ontology	Significance	Genes overlapping %
DNA binding	4.10E-21	19%
cancer	2.49E-17	6%
oncogene	2.90E-12	4%
tumor suppressor	3.10E-07	2%
Ca ⁺⁺ dependent receptor signaling	8.10E-06	4%
receptor antagonist activity	4.60E-05	<1%
morphogenesis	2.50E-04	1%
cell proliferation	1.60E-03	2%
estrogen receptor activity	5.30E-03	<1%

By combining data that was reproducible in both datasets (expression and chromatin condensation) and further sub-classifying expression changes by their chromatin state, the interpretation is greatly improved. From the list of genes with two-fold increased expression after 5 Aza dC treatment, 337 genes showed a two-fold decrease in condensation as expected. Several GO categories were statistically over-represented ($p < 8.3E-4$) in this group of genes that were not previously identified;

including 'microtubule motor', 'viral life cycle', and PP2A complex (table 4.3a). Also found was the gene *KIF2*, a microtubule motor protein associated with apoptosis and a possible tumor suppressor (Debernardi, Fontanella et al. 1997). In addition, 184 genes showed two-fold increases in both expression and chromatin condensation after treatment with 5 Aza dC. This type of change is indicative of the complex nature of chromatin regulation, especially given the potential for a gene to contain multiple chromatin domains that are associated with different isoforms (Levitsky, Podkolodnaya et al. 2001; de Leon, Montanez et al. 2005). The GO categories over-represented ($p < 7.4E-3$) upon sub-classification included 'protein deubiquitination', 'negative regulatory of anti-apoptosis', and "tubulin-tyrosine ligase activity" (table 4.3a). The stress induced gene *RTN4*, which has numerous isoforms resulting from alternative splicing events, was a member of this group as well (Acevedo, Yu et al. 2004).

The GO categories identified by the sub-classifications are potentially relevant to both the cancer system and the drug treatment, since it divided the previously identified classifications into the functionalities that were affected. For example, with this data it is possible to identify some of the cytoskeletal processes that were directly affected by 5 Aza dC treatment. The subdivision reduced the heterogeneity of the lists, and previously suppressed categories like 'viral life cycle' become significant ($p < 8.1E-5$). In this case, prior evidence indicated that the identification of this category was being suppressed, since decreased condensation of endogenous retrovirus DNA and even some viral gene re-expression (including *HCF2*) are known consequences of 5 Aza dC treatment (Ogasawara, Okada et al. 2003).

Table 4.3: Functional classifications of genes with 2 fold increased expression after treatment further sub-divided by chromatin state

A: Chromatin condensation decreased 2 fold (337 genes)

Gene Ontology	Significance	Genes overlapping %
cytoskeleton	2.40E-09	16%
STAT cascade	3.80E-07	37%
cell adhesion	4.58E-05	8%
microtubule motor	6.50E-05	3%
viral life cycle	8.10E-05	6%
morphogenesis	7.90E-04	4%
PP2A complex	8.30E-04	2%

B: Chromatin condensation increased 2 fold (184 genes)

Gene Ontology	Significance	Genes overlapping %
cytoskeleton	3.30E-09	21%
chaperone	2.30E-06	12%
protein deubiquitination	1.10E-04	2%
neg. regulation of anti-apoptosis	1.10E-04	2%
tubulin-tryosine ligase activity	5.10E-04	2%
anti-apoptosis	7.40E-03	3%

Numerous GO categories were over-represented by the 77 genes that showed a 2-fold decrease in expression and a 2-fold decrease in condensation after treatment with 5 Aza dC. The groups over represented ($p < 1.1E-3$) included ‘negative regulation of mitosis’, ‘estrogen receptor activity’, and ‘pyridine nucleotide biosynthesis’ (table 4.4a). These classifications were consistent with the inhibitory effects of 5 Aza dC treatment on cell division and estrogen-signaling pathways (including the *ESR1* gene which has multiple isoforms, all of which are not affected equally by treatment), which are critical for MCF7 survival (Olivier, Bautista et al. 1998; Carling, Kim et al. 2004; Parkinson, Sarkans et al. 2005).

Conversely, only 14 genes showed a 2-fold decrease in expression and 2-fold increase in condensation after treatment with 5 Aza dC. This group was over-represented

($p < 9.5E-3$) by the GO category Cathepsin S Activity and the parent category Lysosome (table 4.4b). The inclusion of these categories was done only for completeness, the number of genes (2) used to identify these categories makes it unlikely that the classification is biologically informative.

Of the genes listed on table 4.4, 61% have multiple probes on the microarrays used for the chromatin array, and therefore are known to have alternatively spliced forms. Of the remaining 39%, all are unannotated or minimally annotated, but of the genes with some annotation most have known alternative splice forms (Maglott, Ostell et al. 2005). Further analysis of the alternately spliced genes demonstrates that the probe that showed condensation after treatment was not the probe for the 3' end of the transcript (the one typically used for expression measurements), but was instead a probe for an internal region. This analysis suggests the possibility that these genes do not violate the expected inverse relationship between chromatin and expression nor are they an exception to the expected results of 5 Aza dC treatment with regard to chromatin state (Batsche, Yaniv et al. 2006). Since chromatin condensation occurs within the transcription unit at the level of individual exons/regions, and only certain isoforms are affected by the drug treatment. This indicates that the chromatin state of these regions is regulated separately (Nogues, Kadener et al. 2002).

From the results shown on tables 4.3 and 4.4, the combined expression and chromatin state datasets could enhance the sub-classification of the treatment-induced changes. With the more specific lower level GO classifications, the number of candidate genes is greatly reduced, which facilitates interpretation and therefore discovery.

Table 4.4: Functional classifications of genes with 2 fold decreased expression after treatment further sub-divided by chromatin state

A: Chromatin condensation decreased 2 fold (77 genes)

Gene Ontology	Significance	Genes overlapping %
neg. regulation of mitosis	1.10E-08	5%
chromatin remodeling complex	5.70E-08	5%
nitric-oxide synthase regulator	5.70E-08	5%
estrogen receptor activity	3.40E-05	5%
pyridine nucleotide biosynthesis	8.10E-05	4%
developmental processes	3.00E-04	28%
cell growth	1.10E-03	5%

B: Chromatin condensation increased 2 fold (14 genes)

Gene Ontology	Significance	Genes overlapping %
Cathepsin S activity	4.20E-08	21%
lysosome	9.50E-03	21%

4.3.2 Discoveries emerging from the experiment: gene view

Given the more than 3,000 genes that showed transcriptional or chromatin state changes in response to treatment, interpretation up to this point has largely been based on GO classification. While GO-based analysis has proven quite powerful, in revealing the “big” picture, the most common and direct method for examining microarray data is to examine the genes that show the greatest fold-change in response to the treatment. Since thousands of genes show a significant fold change (>2 fold) as a result of treatment, only the twenty with the biggest change from each group will be examined. A complete list of all the genes that have a greater than 2 fold change as a result of treatment both in expression and chromatin state is included in Appendix B.

The first group of twenty genes (table 4.5A) are those that show the greatest fold increase in expression following treatment. As can be seen from the table, with the

exception of three genes, all the genes show significant increases in expression but are not detectable reproducibly in the S2 fraction. The reason is that the S2 fraction contains the inactive chromatin and the genes on this list are so highly expressed, in at least one condition, that if their chromatin state is measurable in the S2 fraction it is only before transcriptional induction, if at all. This results in the fold changes being artificially large, denoted by the <0.001 . For three genes (*SLC6A8*, *GIP2*, and *ATP2A3*) it was possible to reproducibly measure their chromatin condensation states in the S2 fraction even after 5 Aza dC treatment induced expression. Examination of the functions of these genes shows that they are not related. However, further examination of the annotation shows that all of these genes are known to be expressed in normal mammary tissue and repressed in MCF7. In addition, each of these genes has a CpG island in its promoter (Larsen, Gundersen et al. 1992; Lash, Tolstoshev et al. 2000). As expected these findings indicate that 5 Aza dC treatment is able to induce the expression of genes normally expressed in mammary tissue that are presumably silenced by aberrant methylation of the gene's promoter CpG island.

The remaining genes on this list also have a CpG island in their promoter regions, which indicates that the 5 Aza dC treatment is also potentially relieving the repression of these genes. Many of these genes' functions are intriguing with regard to the cell's response to treatment including the putative chromatin structural protein *ACRC*, several chromatin remodeling proteins (*CITED2*, *EPAS1* and *PCAF*), and two proteins involved with sterol synthesis *CYP24A1* and *DHRS8*. Most of these genes are also known to be expressed in normal mammary tissue and repressed in MCF7, so induction by 5 Aza dC is not unexpected (Lash, Tolstoshev et al. 2000). Of greater potential interest are the

genes like *PCAF* and *RP2*, which are transcripts not known to be associated with normal mammary tissue, and therefore might represent the activation of novel pathways following treatment.

The twenty genes with the greatest fold decrease in expression following 5 Aza dC treatment are shown in Table 4.5B. Most of these genes are highly expressed and remain expressed even after treatment. The corresponding chromatin state changes are substantial, having gone from being almost completely open to partially closed, indicated by the artificially high fold changes denoted by >10 . The exception is the unannotated gene *KIAA1586*, which has a low level of expression that is silenced by treatment. While the gene is unannotated, expression studies show it to be nuclear localized and homology mapping indicates it has a zinc finger domain possibly indicating it is DNA binding protein (Hubbard, Barker et al. 2002). More interestingly, its promoter also contains a CpG island, but *KIAA1586* is not a transcript found in normal mammary tissue (Safran, Solomon et al. 2002).

The remaining genes on the list show a strong association ($p < 0.0001$) with the GO term metal binding, since *MTIX*, *PSPH*, *HFE*, *STC2*, and *CA12* all bind either iron and/or magnesium ions. Examination of these genes' annotations reveals two common features: they are not expressed in normal mammary tissue and their promoter regions contain CpG islands (Larsen, Gundersen et al. 1992; Lash, Tolstoshev et al. 2000). While this relationship seems to be in conflict with the normal model of how CpG islands and 5 Aza dC treatment interact, there are published models that potentially explain this. CpG island methylation is repressive on transcription, consequently, if the DNA element being repressed is a transcription repressor binding site then, the repressor is repressed and

transcription is activated. Examples of this type of regulation include the possible oncogenes *NTRK1* and *hTert* (Guilleret, Yan et al. 2002; Guilleret and Benhattar 2003; Fujimoto, Kitazawa et al. 2005). With this model it is possible that 5 Aza dC treatment is silencing genes induced by aberrant CpG island methylation of repressive elements.

Table 4.5: A selection of genes with the largest treatment induced expression changes**A: Expression induced by treatment**

Name	RefSeq#	Expression fold change	Chromatin fold change*
ACRC	AJ311392	17.23	<0.001
SGPP1	AJ293294	13.13	<0.001
CITED2	BC004377	9.32	<0.001
CYP24A1	L13286	8.39	<0.001
PCAF	NM_003884	8.39	<0.001
SLC6A8	U17986	7.20	0.26
MAP2K3	NM_145110	6.58	<0.001
INSR	S76825	6.40	<0.001
DHRS8	AF126780	6.05	<0.001
SYNGR3	BC009568	6.02	<0.001
RP2	AJ007590	6.00	<0.001
SLC38A1	AK074797	5.86	<0.001
SF3B1	AF054284	5.80	<0.001
NFYB	BC007035	5.78	<0.001
S100A12	D49549	5.63	<0.001
EPAS1	NM_001430	5.47	<0.001
LACTB2	AF151841	5.28	<0.001
G1P2	M13755	5.25	0.3
ATP2A3	NM_005173	5.15	0.27
DKFZP434B044	AK027395	5.11	<0.001

B: Expression repressed by treatment

Name	RefSeq#	Expression fold change	Chromatin fold change**
MT1X	BC032338	0.20	>10
KIAA1586	AB046806	0.20	3.7
NDP52	BC004130	0.24	>10
BTBD9	AK057507	0.25	>10
PA2G4	NM_006191	0.26	>10
RGNEF	AK025470	0.27	>10
PSPH	NM_004577	0.27	>10
PIPOX	AF134593	0.28	>10
ASB13	NM_024701	0.29	>10
IVNS1ABP	AB020657	0.30	>10
SCN11A	AF109737	0.30	>10
FLJ12973	AK023035	0.30	>10
EIF4G3	AF012072	0.30	>10
PFKFB3	AJ272439	0.31	>10
STC2	AB012664	0.31	>10
BTC	NM_001729	0.32	>10
COX4I2	NM_032609	0.32	>10
SPDEF	AB031549	0.32	>10
HFE	AF079408	0.33	>10

CA12 BC001012 0.44 >10

*The fold change of most of the data in this column is listed as <0.001, the reason for this is that in at least one condition the chromatin state has a value approaching zero (the numerator in the ratio calculation) and therefore the change is artificially high.

**The fold change of most of the data in this column is listed as >10, the reason for this is that in at least one condition the chromatin state is approaching zero (the denominator in ration calculation) and therefore the change is artificially high.

The next group of twenty genes that were examined (Table 4.6A) are genes with the greatest increase in chromatin condensation following 5 Aza dC treatment. Chromatin compaction is accompanied by an expected and significant reduction in expression level (>1.36 fold). Even a 1.36 fold increase in chromatin condensation is significant, given that the overall average chromatin change caused by treatment with 5 Aza dC is a 2 fold reduction in the chromatin condensation. The gene of highest significance on this list is *AK5* (Momparler 2005). As discussed in section 4.2.2, this gene is known to be suppressed by 5 Aza dC treatment. The significant pattern ($p < 0.0001$) of annotated functions of these genes is that half of the genes make protein products that interact with the extra-cellular space.

As with the genes in table 4.5B where expression is suppressed by 5 Aza dC treatment, these genes are also not known to be expressed in normal mammary tissue and are associated with CpG rich regions. This list is more complex because many of the “genes” on the list are actually probes for alternative splice forms of the same gene (including *PTPN22*, *COL4A3*, and *HTR4*). Further scrutiny shows that these regions often have differing chromatin states from other probes for the same gene and there are multiple CpG rich regions in the gene (intergenic CpG islands). Given that numerous instances of intragenic chromatin state variation were also identified as part of the GO

classification analysis, it is a widespread phenomena requiring further analysis and discussion and will therefore be the focus of section 4.4.

The final twenty genes (table 4.6B) to be discussed are those that showed the most relaxation as a result of 5 Aza dC treatment. As expected for this group, significant chromatin relaxation is accompanied by increased expression. This list has several genes that are known tumor suppressors, such as *PML* and *BBC3*, making the biological context of these changes relatively simple to interpret. However, the reason for disease related genes, like *NPC1L1*, and *G3BP*, appearing on this list is not as clear and represents an opportunity for discovering genes and pathways not previously thought to be involved either with the disease or the drug.

As with the genes that showed the biggest changes in expression after treatment, the genes on this list (>70%) are mostly known to be expressed in normal mammary tissue, but repressed in MCF7. Additionally, these genes mostly (>90%) have CpG islands in their promoters, but all of them have at least a CpG rich region associated with their promoter.

Table 4.6: A selection of genes with the largest treatment induced chromatin changes**A: Chromatin condensed by treatment**

Name	RefSeq#	Expression fold change	Chromatin fold change
KLRD1	AB009597	0.26	6.33
RANBP2L1	NM_032260	0.31	5.97
PCDHGC3	AF152503	0.20	4.10
LGALS8	AF074000	0.34	3.40
ATR	U49844	0.17	2.92
PTPN22	AF077031	0.31	2.90
PTPN2	BC008244	0.43	2.78
ZC3HDC7	AK000889	0.22	2.65
KCNA2	L02752	0.24	2.64
PTPN22	NM_012411	0.41	2.64
PCDHGC3	AF152336	0.59	2.53
HS3ST1	AF019386	0.08	2.53
CTSS	BC002642	0.37	2.51
HTR4	AJ278982	0.13	2.27
AK5	AK090967	0.30	2.27
ITPR2	AB012610	0.45	2.16
PAI-RBP1	BC003049	0.40	2.09
STATH	M18078	0.41	2.09
COL4A3	NM_031362	0.08	2.01
RNF32	AF325690	0.47	1.36

B: Chromatin relaxed by treatment

Name	RefSeq#	Expression fold change	Chromatin fold change
SRPR	BC008077	2.02	0.08
FLJ20014	BC009261	11.02	0.08
DNAJA4	BC031044	11.64	0.10
NPC1L1	AF192523	8.06	0.10
PML	AF230408	4.07	0.11
USP49	AF116602	2.28	0.13
LAMB2	AK094050	2.05	0.18
RPS8	BC005678	2.64	0.18
DHRS10	BC006283	2.10	0.19
TBX3	NM_016569	3.21	0.19
MLL	AF231999	9.97	0.19
HLA-A	AF255717	3.31	0.19
MID1	AF041206	5.52	0.21
BBC3	AF354656	2.27	0.21
TRIM10	NM_006778	3.03	0.22
ANKRD10	AK026017	2.26	0.22
FLJ11235	AK002097	5.95	0.23
FSHB	AL358944	3.57	0.23
G3BP	NM_005754	3.37	0.23
DUX4	D38024	2.79	0.24

The relationship between chromatin state and expression is not always as clear cut as it is with the genes on tables 4.5 and 4.6. Sometimes the results from the chromatin condensation data when paired with the expression data seem contradictory. For example, the tumor suppressor *TP53* showed a 38% chromatin relaxation following treatment but the expression was also decreased 68% (however the gene is still strongly expressed). Given the role that *TP53* plays in protein complex formation required for the regulation of other genes and the pro-apoptotic function of the protein itself, it is possible that the region is relaxed by treatment since the gene is regulated by the methylation state of a CpG rich region in the promoter region (Hodge, Peng et al. 2005). The level of transcription, on the other hand, is suppressed, at least in a subset of the cell population (Nieto, Samper et al. 2004). This finding is supported by previous work that shows even though *TP53* is not mutated and is expressed in the MCF7 cell line, measurable translation occurs only in response to DNA damage (Olivier, Bautista et al. 1998; Karpf, Moore et al. 2001). While the resolution of this apparently contradictory relationship was simple to explain, there are a number of genes where more comprehensive analysis must be performed to understand the nature of the relationship between expression levels and chromatin state.

4.3.3 Novel information derived from genes with expression levels too low to be measured with microarray based methods.

In this experimental series there were 640 genes where the chromatin state was measurable in both conditions, but expression was only measurable in one. As further evidence of the reproducibility the chromatin array method, this list includes 95% of the 152 genes shown in the previous chapter to have a measurable chromatin state but having

no expression using microarray-based measurement methods. Again, many of the genes on this list were tissue-specific transcripts (like the largely neural specific *BEXL1*), and not expected to be expressed in a breast cell line, such as MCF7. Treatment with 5 Aza dC, was however able to induce the expression of many tissue specific transcripts (like the testis-specific protein *TSGA10*). These results seem to demonstrate how non-specific the effects of 5 Aza dC treatment are, since the expression of a testis-specific transcript in a breast cancer cell line would seem to indicate uncontrolled global gene expression. This is however, not true, since SAGE and microarray-based expression studies have shown that *TSGA10* and other “tissue specific” transcripts are actually expressed in normal breast tissue, but not MCF7 (Diehn, Sherlock et al. 2003).

Of the genes with expression in only one condition (+/- drug), expression of the pro-apoptotic gene, *CASP6*, was induced more than 2 fold by treatment and while expression of the oncogene *MYB* was suppressed more than 2 fold by treatment. Additionally, alternative splicing products were induced by treatment, included a novel isoform of the tumor suppressor *PRDM2*, several isoforms of *ESR1*, and *DMD* discussed in the next section.

There were also 125 genes with measurable chromatin state but no expression as measured by microarray-based methods in either condition (table 4.7). Among the genes included on this list were the genes *NTRK1* and *BAD*. *BAD* is a tumor suppressor gene in the 11q13 region, which as discussed in the previous chapter (section 3.3.4). While *BAD* is known to be amplified in MCF7 the chromatin array results show that it is condensed, and its transcription is suppressed (Pollack, Perou et al. 1999). Additionally despite the 5 Aza dC induced relaxation of the region (-1.5 fold), expression of the pro-apoptotic gene

BAD remains suppressed while genes that provide a selective advantage like the antioxidant enzyme *PRDX5* is amplified, relaxed and expressed.

The results discussed here are only a sampling of the possible discoveries that can be made using the chromatin state data to identify genes with a differential behavior even when the expression is not measurable and therefore uninformative using microarray-based methods.

Table 4.7 Genes with measurable changes in chromatin state but no measurable expression

Genbank	Name	Fold*	Genbank	Name	Fold*	Genbank	Name	Fold*
NM_002780	PSG4	2.07	NM_003040	SLC4A2	0.61	AB041926	MINK1	0.42
BC024000	ZNF638	1.66	AF045764	GPR32	0.61	NM_001830	CLCN4	0.42
BC003634	TFCP2	1.56	AY042225	CD209	0.61	L15006	CTLA4	0.42
AF535149	FGF5	1.37	AF021792	BAD	0.61	X02661	OAS1	0.41
AB042557	PDE4DIP	1.26	L33801	GSK3B	0.60	BC018929	PHLDA1	0.40
AB030176	PADI2	1.22	AF007162	CRYAB	0.59	XM_044178	KIAA1211	0.40
AK001907	TDRD4	1.21	AB040147	RND1	0.59	BC016173	CDKN2C	0.40
AB033053	ZNF295	1.18	AF019039	TNPO2	0.58	BC016680	SP2	0.39
BC036669	COL25A1	1.15	NM_030756	TCF7L2	0.58	AF182277	CYP2B6	0.38
AF401652	GCNT2	1.11	AF131810	FAM26B	0.58	AF043380	FLG	0.37
BC026309	LILRB4	1.11	AF131218	C16orf5	0.57	AF064243	ITSN1	0.37
AK057479	NYD-SP25	1.10	AF423190	CACNB2	0.55	NM_001352	DBP	0.37
AK026486	FLJ22833	1.07	M60502	FLG	0.54	AF294278	PRDM16	0.37
D86971	KIAA0217	1.06	AB016517	FGF5	0.53	M96943	FLG	0.37
M15788	PPY	1.05	U26914	RREB1	0.53	BC033814	SP2	0.36
AF161521	FARSLB	1.03	NM_021641	ADAM12	0.53	BC012151	NFX1	0.35
AF063606		1.02	M84747	IL9R	0.53	AF381029	LMNA	0.35
AK097416	TRIM42	1.01	BC010019	MED8	0.53	AB044946	NDRG4	0.34
U81237	VWF	0.99	NM_002783	PSG7	0.53	AF251294	CCNL2	0.34
D28118	ZNF161	0.97	M23102	NTRK1	0.53	BC009240	TADA3L	0.33
U66496	LEPR	0.93	NM_005572	LMNA	0.52	AF078165	AXIN2	0.32
XM_117117	FLJ13072	0.93	BC007872	TK1	0.52	AB018295	NY-REN-7	0.31
AF043909	MUC5AC	0.93	U91543	CHD3	0.51	AB011076	UTF1	0.29
BC026691	PPM1A	0.92	NM_018955	UBB	0.51	U16153	ID4	0.29
NM_017745	BCOR	0.89	BC012755	CHGA	0.51	M81768	SLC9A1	0.27
U36501	SP100	0.89	AK026061	ABHD9	0.50	AK000413	FLJ20406	0.26
NM_015385	SORBS1	0.89	AK000914	FLJ10052	0.50	BC007223	DNCL2A	0.26
AK056425	KREMEN1	0.88	AB028963	KIAA1040	0.50	BC003381	KIAA0217	0.25
XM_098320	IRX1	0.86	NM_012188	FOXI1	0.49	NM_004218	RAB11B	0.25
BC009344	DOM3Z	0.85	AJ291679	DPEP3	0.49	AF131760	LRP10	0.24
AF123658	LZTS1	0.82	NM_003999	OSMR	0.49	NM_080741	NEU4	0.22
AF289573	KCTD13	0.77	BC025290	PLA2G4B	0.49	AK000900	FLJ10038	0.15
AF147782	ETV7	0.76	AF498964	RAC1	0.49	AK024264	ATP11A	0.15
AF044774	WDR22	0.75	AK090930	BTBD9	0.49			
NM_002019	FLT1	0.75	M96843	RPL10A	0.48			
AB059569	WNT10A	0.74	AK027209	TCF7L2	0.48			
AF327367	CHRNA10	0.73	BC001291	LY6K	0.48			
BC000993	ALS2CR15	0.72	M34986	EPOR	0.47			
BC005386	PLA2G1B	0.71	AK022593	KIAA1509	0.47			
AB006867	SOX15	0.67	AF144308	GPR44	0.47			
AB027464	FZD10	0.66	AK001341	ZDHHC4	0.47			
AF339750	REERG	0.66	BC016826	HSH2D	0.45			
L34155	LAMA3	0.64	AB010637	UPK3A	0.45			
AF156777	ASB1	0.64	AB001467	EFS	0.44			
BC019352	ZNF342	0.64	AF100745	LOC51136	0.44			
AK022536	C7orf19	0.63	BC009356	CDC42EP1	0.43			

* The fold change is after treatment with 5 Aza dC

4.4 Chromatin shows intragenic variation

In section 4.2 it was discussed that there is a 5 Aza dC-dependent destabilization of the nuclear membrane resulting in the treated cell line being susceptible to lysis during the early stages of the chromatin fractionation procedure. Microscopic examination of the cells after treatment showed spreading of both the cellular and nuclear membranes (figure 4.4) and the GO results indicated that the cytoskeleton was altered. Since membrane spreading is a characteristic of several muscular dystrophies (Lapidos, Kakkar et al. 2004; D'Angelo and Hetzer 2006), the genes linked to these diseases were examined to determine if alterations in one of them could be responsible for these changes.

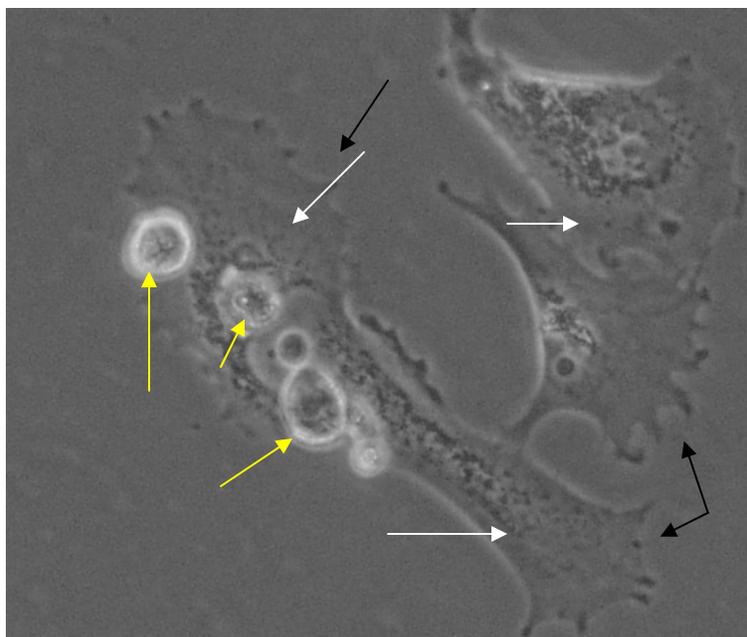


Figure 4.4: A micrograph of several MCF7 cells after treatment. Note the cytoplasmic spreading (denoted by the black arrows) and the enlarged nuclear area (denoted by the white arrows). The spheres (denoted by the yellow arrows) on top of the center cell are cells that detached from the substrate.

As discussed in section 4.3.3, it was observed that the 5 Aza dC treatment induced major changes in the chromatin state and the expression level of the various splice forms of *DMD* (figure 4.5.). These effects included the increase in expression (and a decrease in condensation) of the non-functional Dp427p2 variant. Conversely, the splice variants that encode potentially functional proteins (Dp260, Dp140 and Dp71) showed decreased expression and increased condensation after 5 Aza dC treatment. The suppression of functional *DMD* variants in conjunction with the hypomethylation of the structural heterochromatin might reduce the stability of the cells by weakening the integrity of the membranes (Lapidos, Kakkar et al. 2004; Taddei, Hediger et al. 2004; D'Angelo and Hetzer 2006). More interestingly, these results showed that the different regions of the

gene were in separate condensation states and consequently responded uniquely to the treatment. This indicates that multiple chromatin regulatory domains exist in the gene, possibly corresponding to the various promoters used by the alternate transcripts (de Leon, Montanez et al. 2005).

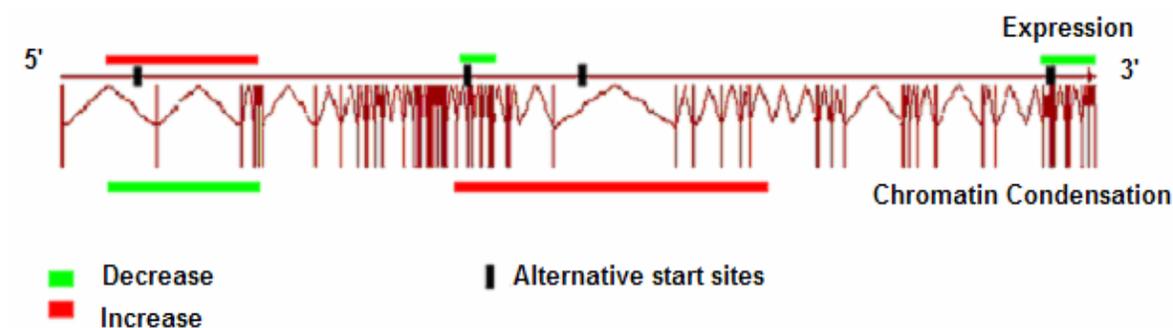


Figure 4.5: A graphic of the structure of DMD showing the areas interrogated by the chromatin and expression arrays and their response to treatment. The changes are displayed using the traditional microarray coloring scheme with red being increased intensity and green being decreased intensity. The tick marks represent the alternative start sites for the major short forms of the gene starting with Dp427p, Dp260, Dp140, Dp71, and Dp 40.

In an attempt to determine the sequence elements responsible for this behavior, computational methods were employed (section 4.5.7). Examination of the gene's structure revealed an absence of CpG islands and an overall GC content of less than 45%. Consequently, it is most probable that the alterations in start sites and splicing were indirect responses to the treatment and not a direct result of DNA hypo-methylation of a CpG island-linked regulatory mechanism. Further analysis of the region using DNA motif identification packages such as TRANSFAC also failed to specifically identify the

elements potentially responsible for the differential regulation (Wingender, Chen et al. 2000). The reason for the failure was that regions contained dozens of conserved protein binding DNA motifs. Additional attempts to use MEME (Bailey and Elkan 1994) along with other genes and gene regions that had a similar behavior to identify a novel motif that may be correlated to the differential regulation were unsuccessful. Again, the reason for this failure was that even after masking known repeats, the number of possible motifs found was still too great to identify a subset that could be correlated with the response.

4.4.1 Chromatin regulation of the hTert locus

The *DMD* example highlights the possibility that multiple chromatin regulatory domains can coexist in a single locus. As previously mentioned, there were numerous genes where the relationship between the chromatin condensation and expression state were in conflict at a sub-gene resolution, making it likely that this is a common occurrence. To address this possibility, the data set was searched for genes with multiple probes that showed differential states or responses. Of the 3,579 genes with multiple probes on the array, nearly a thousand genes had multiple probes and were identified as reproducible at the probe level on all the chromatin arrays (the full list is included in Appendix B.) This list contained a number of interesting genes with known links to cancer which are shown on table 4.8 (Widschwendter and Jones 2002; Furney, Higgins et al. 2006), but for brevity only two will be discussed in detail.

Table 4.8: Selected genes with intragenic chromatin variation

Name	Entrez Gene #	Variation type
<i>hTert</i>	7015	Short form probe condensed
<i>MBD1</i>	4152	Probe for long form condensed
<i>BRCA1</i>	672	Condensation variable across the gene
<i>CDKN1A</i>	1026	Condensation variable across the gene
<i>GNAS</i>	2778	Parental versus Maternal differences
<i>PRDM2</i>	7799	5' end open and 3' end closed
<i>RARB</i>	5915	Control protein is heterochromatin
<i>MLL</i>	4297	Fragile sites condensed
<i>PSEN</i>	5663	Condensation variable across the gene
<i>RUNX1</i>	861	Break sites show differential condensation
<i>HDAC3</i>	8841	Condensation across gene variable
<i>RB1</i>	5925	Short form probe condensed
<i>DNMT1</i>	1786	Short form probe condensed
<i>DNMT3B</i>	1789	Short form probe condensed
<i>DNMT3A</i>	1788	5' end open and 3' end closed
<i>Casp4</i>	837	Alternate start site condensed
<i>DMD</i>	1756	Condensation variable across the gene
<i>ESR1</i>	2099	Condensation variable across the gene
<i>ESR2</i>	2100	Condensation variable across the gene

The gene *hTert* was chosen since it represents one of the most enigmatic regulatory systems of any gene. Additionally, the gene was selected because it is difficult to study with expression microarray technologies, given that MCF7 has few copies of this transcript per cell, requiring signal strong amplification for reliable detection (Spiropoulou, Ferekidou et al. 2004). The reason *hTert* is considered enigmatic is that, besides having numerous alternative splice forms (Yi, White et al. 2000), promoter hypermethylation is required for transcriptional activity (Guilleret, Yan et al. 2002), which is contrary to the normal results of promoter hypermethylation.

For several years it has been suspected that active chromatin remodeling plays a role in the regulation of *hTert* (Ducrest, Szutorisz et al. 2002), but the proof came from a hypersensitivity site mapping study of the region in a telomerase positive breast cancer

cell line (Szutorisz, Lingner et al. 2003). In the published study, the authors found restriction hypersensitive sites in both the second exon and intron of the locus and the sensitivity of the sites was positively correlated to expression levels. Examination of the probe for this region (AB016767) in the chromatin array data showed that prior to treatment the chromatin state was relaxed. This finding was consistent with the published data, since DNase I hypersensitive sites are correlated with relaxed chromatin. After treatment, this region was more condensed, which was consistent with the model that 5 Aza dC treatment represses transcription of this gene by demethylating the promoter. New evidence by Renaud *et. al.* more directly confirms the chromatin array results by showing that there is a *CTCF* binding site in this region, which when bound to *CTCF* is sufficient to silence transcription downstream by inducing chromatin condensation (Renaud, Loukinov et al. 2005). However, CpG methylation in the region can prevent the binding of *CTCF*, and therefore relieve the repression of the full length transcript (Hark, Schoenherr et al. 2000). Given that the exon 2 and intron 2 regions are known to be methylated in MCF7 (Guilleret and Benhattar 2003), it is not at all surprising that treatment with the DNA methylation inhibitor 5 Aza dC induces a large increase in the chromatin condensation of the region (figure 4.6). Difficulties in mapping other regions prevented Szutorisz *et. al.* from examining them in detail, but these problems were not encountered with the chromatin array (Szutorisz, Lingner et al. 2003).

The probe for the $\alpha\beta\gamma$ deleted *hTert* splice form (AB086950) had a relaxed chromatin confirmation prior to treatment and was condensed by treatment with 5 Aza dC (figure 4.6). The probe for the full length transcript (AF015950) showed a heterogeneous condensation consistent with partially suppressed chromatin, and 5 Aza

dC treatment condensed it further (figure 4.6). Unlike the *DMD* example, in *hTert*, the analysis of the differential chromatin regulation via computational means was revealing since the correlated regulatory motifs (CpG islands) were known (Ulaner, Hu et al. 1998; Devereux, Horikawa et al. 1999).

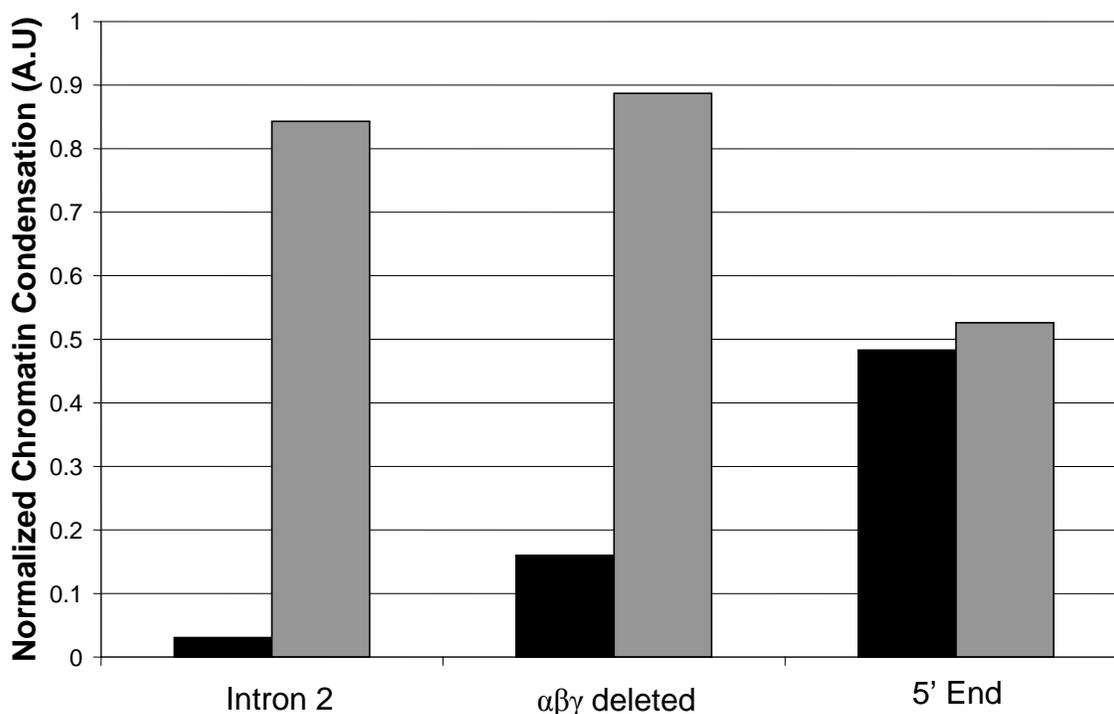


Figure 4.6: Intensity changes across probes specific to distinct *hTert* regions show that the chromatin state varies over the length of the loci. Black is untreated and gray is treated. The differences in native chromatin states of these gene slices, as well as their differing response to 5 Aza dC treatment, indicate that *hTert* regions undergo at least partial differential regulation.

The *hTert* locus promoter CpG island extends half-way through intron 2 (including the AB016767 probe) of the longest product and contains a great many protein-binding sites, even in regions that are outside what is traditionally considered the promoter region (Guilleret and Benhattar 2003; Renaud, Loukinov et al. 2005). Because

of this strong connection to CpG-mediated regulation, it can be surmised that additional CpG island elements could be involved in the regulatory processes of other functions, such as splice form control. Using the methods discussed in section 4.5, four distinct CpG islands were identified inside the locus in regions that encompass the exons spliced out in the α (Δ the first 36 bases of exon 6) and β (Δ exons 7 and 8) minus isoforms and is probed in part by AB086950 (figure 4.7). Sequence analysis revealed that the region was rich with putative splice-enhancing motifs and binding sites for other regulatory proteins (Hark, Schoenherr et al. 2000; Zhang and Chasin 2004). The combination of intragenic CpG islands and splice modulators can allow epigenetic processes to modulate splicing, without interfering with transcriptional start or elongation (Gonzalzo, Hayashida et al. 1998; Kleinjan, Seawright et al. 2004). The CpG islands that are associated with exons 6-8 encompass all the relevant splicing control regions necessary to modulate the splicing events (figure 4.7). The co-occurrence of intergenic CpG islands within each of the regions, but not within non-alternatively spliced exons (exons 3 or 5), strongly suggests a link between splice form control and epigenetic regulation.

methylation of exon 5's non-CpG island CpGs (table 4.9b), with most of the decrease resulting from strong demethylation of 5 specific CpGs. The results of this work clearly demonstrate that the chromatin state of *hTert* shows intragenic variation. Furthermore, it indicates these variations are potentially meaningful in the context of the transcriptional regulation of the gene, since the CpG island associated with exon 6 is resistant to demethylation by 5 Aza dC treatment, but this resistance is not coupled to the changes seen in exon 5 non-CpG island CpG dinucleotides (Hatada, Kato et al. 2002).

Table 4.9: *hTert* exons 5 and 6 bisulfite sequencing results.

A: Exon 6 CpG Island CpGs (19 CpGs averaged)

Cell Line	% Methylation before Aza	% Methylation after Aza
MCF7	95% +/-5%	90% +/-19%
H1299	93% +/-5%	0%

B: Exon 5 Non-CpG island CpGs (21 CpGs averaged)

Cell Line	% Methylation before Aza	% Methylation after Aza
MCF7	92% +/-7%	50% +/-40%*
H1299	88% +/-5%	76% +/-2%

*The standard deviation is due to the high heterogeneity of the clones, since 13 were nearly fully methylated (mean methylation state 85%) and the rest were nearly completely demethylated (mean methylation state 7%)

4.4.2 Chromatin regulation of the *PRDM2* locus

The gene *PRDM2* is a tumor suppressor with protein methyltransferase functionality. Furthermore, the gene encodes multiple isoforms, which have different binding partners, and therefore serve possibly non-interchangeable functions in the *Rb*, hemeoxygenase, or estrogen receptor controlled transcriptional pathways (Pruitt, Katz et al. 2000). On the microarray used for the chromatin array experiments were probes for

two of the putative *Rb* binding *PRDM2* isoforms, which have different 3' UTRs and COOH terminuses (figure 4.8).

The chromatin array results showed that the two regions of this gene had slightly different condensation states (the probe for the short form was 1.5 fold more condensed), even before treatment. However; treatment caused significant relaxation of the short form (-3 fold) but less relaxation in the long form (-1.5 fold), which was less than the average (-2 fold) relaxation measured in the dataset (table 4.10). The expression data confirm these results, with the expression of the transcripts that have the long form exon showing a -1.2 fold change after treatment, versus the average expression change seen in the dataset of +1.2 fold. Expression of transcripts containing the short form exon showed a 2.3 fold change after treatment. Expression of three other isoforms of *PRDM2* that are included in Affymetrix probe set but not in the probe set used for the chromatin array, were undetectable both before and after treatment, indicating the chromatin state change is isoform specific. Sequence analysis, showed that there is a CpG island in the promoter of the full length *PRDM2* transcript. Additionally, there is another CpG island in the region associated with several *PRDM2* transcripts that use an internal start site to produce transcripts that use the short form exon (figure 4.8).

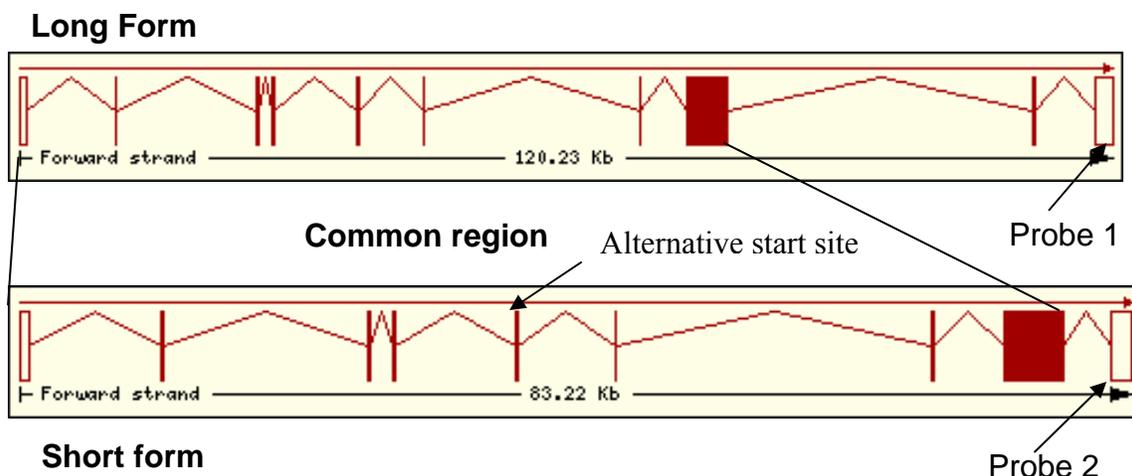


Figure 4.8: Isoforms of PRDM2 probed on the chromatin array. The common exons are noted, as are the unique exons with an approximate location of the probes on the chromatin array, and the regions the CQ-PCR primers and probes were designed from. The alternative start site is shown and all known transcripts from this promoter use the alternative 3' exon of the short form.

The chromatin array results were validated using a quantitative PCR (Q-PCR) technique called chromatin quantitative PCR (CQ-PCR). This method was developed as an extension Q-PCR methods originally used to validate aCGH results in order to now validate the chromatin array results at sub-kilobase resolution (Perry, Nobori et al. 1997). The TaqMan chemistry from ABI (Applied Biosystems Group, Foster City, CA) was chosen for QC-PCR given that it is the standard for expression microarray validation (Ginzinger 2002; Hembruff, Villeneuve et al. 2005). Also, the TaqMan chemistry ensures signal specificity in two different ways, since the primers and probe are specific to the region of interest; versus the SYBR green chemistry, which is gets all its specificity only from the primers during the amplification (Hembruff, Villeneuve et al. 2005). This enhanced specificity is important to ensure low noise, when quantitatively amplifying directly from low complexity genomic DNA, minor amplification products are possible, but are unlikely to also match the probe.

CQ-PCR is a novel application, so it was necessary to establish that the probes and the method worked as anticipated. The first test was to show that in total genomic DNA, the probes designed for the assay had equal amplification efficiency, and to establish the minimum quantity of DNA required for the amplification (figure 4.9). Concentrations covering between 100ng and 0.001ng of DNA per reaction were tested. For example, with 10 ng of total DNA per reaction, the average of the cycle threshold (Ct) for the short form assay was 28.4 and the long form assay had an averaged Ct of 28.3. With 1 ng of total DNA per reaction the average Ct for the short form assay was 31.7 and the long form assay had a Ct of 31.6. Controls with probe and primer, but with no DNA showed no amplification. These results demonstrate that the method and the probes are reproducible and that as little as 1 ng of DNA can be used for the assay.

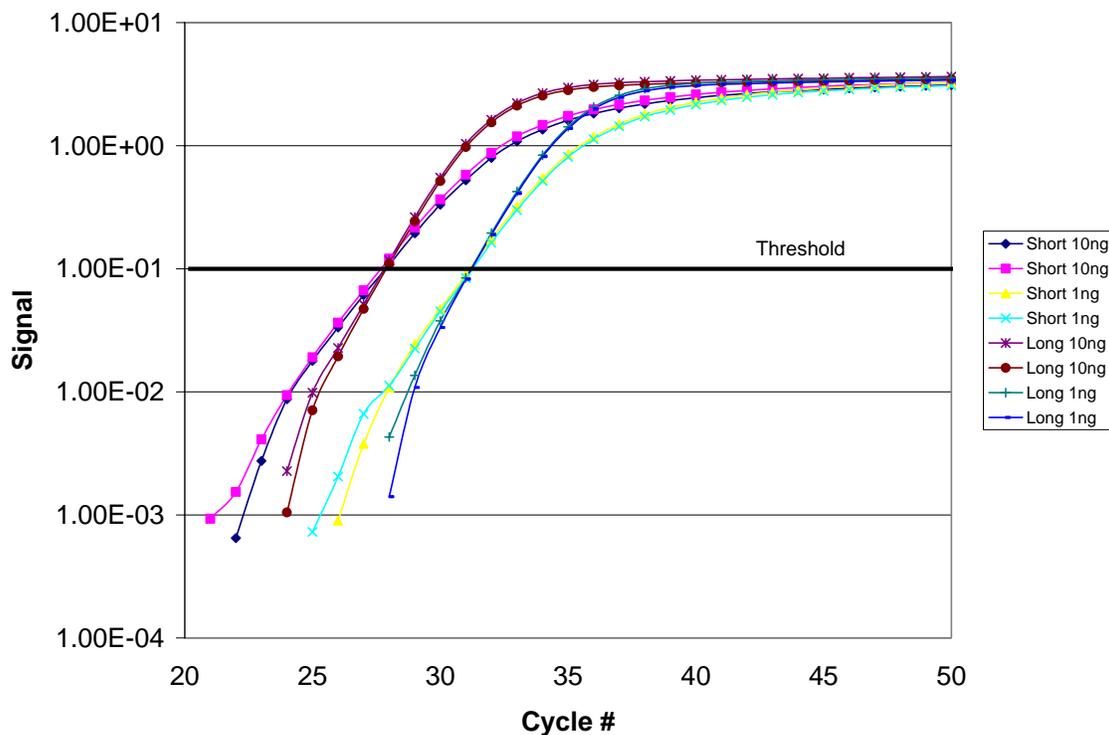


Figure 4.9: The amplification curves of the 1ng and 10 ngs of DNA per reaction tests of primers and probes for the two different *PRDM2* isoforms (called long and short). These results show that the primer/probe sets have equal amplification efficiency when the regions being amplified are at equal concentrations. Additionally the curve also shows that either 1 or 10 ng can be used for the experimental amplifications.

The CQ-PCR assay showed that after treatment, the long form had a -1.1 fold change from untreated levels, which is similar to the -1.2 level measured with the chromatin array (table 4.10). The CQ-PCR of the short form showed a -2 fold change in the treated sample versus the levels in the untreated. While this is not as high as the change measured on the chromatin array, it is significant and preserves the direction of the change (table 4.10). These results confirm the chromatin array data and show the existence of intragenic chromatin state variations, and that these changes are linked to alternative splicing. Using the chromatin array for high throughput screening and CQ-PCR for validation, it is now possible to rapidly and precisely measure the role that

chromatin state changes play in regulating transcriptional processes like alternative splicing and alternative product control.

Table 4.10: Fold changes in PRDM2 after treatment by platform

	Long Form	Short Form
Expression Array	-1.2	2.3
Chromatin Array	-1.5	-3
CQ-PCR	-1.1	-2

4.5 Materials and methods

4.5.1 5 Aza dC treatment

The MCF7 cells (ATCC number HTB-22, (ATCC Manassas, VA)) were plated at the rate of 500,000 cells per 150 mm dish and allowed to recover for 24 hours in RPMI with 10% Fetal Bovine Serum. Serum starvation was not used to synchronize the cell cycle of the population to G1, because it induced a broader than expected change in the chromatin state, so the synchronized chromatin state could not be considered native. The following day, 500 nM 5 Aza dC (Sigma-Aldrich, St. Louis, MO) dissolved in DMSO (500 nM is the IC 50 of 5 Aza dC in MCF7) was added by diluting it into fresh media (Ferguson, Vertino et al. 1997). Treatment with 5 Aza dC was repeated on the 3rd and 5th days, and on the 6th day the cells were harvested and the individual plates pooled. Control plates were seeded at the same rate and were treated similarly, but did not receive the 5 Aza dC treatment. The control plates were harvested at the same time as the treated plates and similarly pooled. For both treated and control, several million cells were removed, washed and flash frozen for later RNA extraction.

4.5.2 Chromatin fractionation

The chromatin fractionation assay was performed as described by Rose and Garrard (Rose and Garrard 1984). Treatment with 5 Aza dC made the nuclei unstable, decreasing the yield, but by increasing the number of cells used and the scale of nuclei prep, enough nuclei were recovered to complete the protocol at $\frac{3}{4}$ scale. The cells were lysed by incubation with a hypotonic buffer supplemented with non-ionic detergent NP-40, the nuclei were washed, resuspended in digestion buffer, and incubated with micrococcal nuclease to about 10% acid solubility. The nuclei were then centrifuged to obtain the S1 fraction, which represents the chromatin that was soluble in a buffer containing 5 mM MgCl₂ and 100 mM KCl. The nuclei were then resuspended in 2 mM EDTA for lysis and the S2 fraction was recovered as the supernatant following centrifugation. The pellet that remains after removal of the S2 fraction was the P fraction. The DNA was purified from the samples by digestion with Proteinase K, extraction with phenol/chloroform, and ethanol-precipitation. DNA was re-suspended in Tris/EDTA pH 8.0, treated with RNase and re-purified as above.

4.5.3 Microarray characteristics

As described in this chapter the larger MWG human 40K microarray (MWG-Biotech, High Point, NC) was used since it has a greater coverage of the genome and in some cases had multiple probes per gene. The expression measurements were reproduced using the Affymetrix Human 133 2.0 Plus chip and the Affymetrix protocols to validate the expression levels and increase the significance of the values.

4.5.4 Chromatin hybridization

The protocol for the chromatin array was similar to the original study (Weil, Widlak et al. 2004). A total of 5 arrays were run including biological replicates and technical replicate experiments to ensure differences in labeling and hybridization efficiencies were accounted for. The arrays were washed according to MWG protocol, and scanned on an Axon 4000B using GenePix 4.1 (Molecular Devices, Union City, CA).

4.5.5 RNA extraction and hybridization

The protocol for the expression array was similar to the original study (Weil, Widlak et al. 2004). The MWG human 40K microarray (MWG-Biotech, High Point, NC) was used, with each sample run against its self in the two channels, and in replicate to ensure differences in labeling and hybridization efficiencies were accounted for. The arrays were washed according to MWG protocol, and scanned on an Axon 4000B using GenePix 4.1 (Molecular Devices, Union City, CA).

4.5.6 Data extraction, normalization and analysis

The flagging of the spotted microarrays was done with GenePix 4.1 software package. The marginal flag was used to designate spots without defects but with low intensity, and the absent flag was used exclusively to denote unreliable data points. By using this flagging method, the ability to do a second pass analysis, specifically to look at features that are undetectable in one condition, was retained. The Affymetrix data was processed in their GCOS analysis package and exported as a pivot table. The data was imported into the GeneSpring 7.1 analysis package (Agilent, Redwood City, CA). The

chromatin data was normalized in a two-step process. The first consisted of normalizing total genomic DNA levels, and then subtracting the genomic intensity from the S2 fraction intensity to give a final intensity value. The data was filtered to remove genes with intensity below the calculated error model on all the arrays. The remaining genes were filtered again to remove data that showed >25% variability in the replicates. The replicates were averaged and genes with known pseudogenes filtered out (Zhang, Harrison et al. 2003). These results were compared with data from Weil et al, since those arrays provide a means to ensure the reproducibility of the chromatin array measurement (Weil, Widlak et al. 2004). While these arrays were different, thereby making direct comparison impossible, the common probes showed a good correlation with more than 70% of the intra-experiment reproducible features showing a less than 35% variation. The data was then divided into groups based on the difference between untreated and treated experiments with >2 fold differences in intensity used as the cut-off threshold.

The expression data from the spotted microarrays was filtered and normalized using the methods described in the original paper. The Affymetrix microarrays were normalized using the GCMS method and filtered to remove low-intensity features. The two types of expression data were compared, and once uncommon or internally irreproducible features were excluded, 77% of the remaining features showed a less than 35% variance in the normalized ratio across platforms. Comparison between the different MWG array types showed that greater than 80% of the genes had less than 35% variance among the common, internally reproducing features.

4.5.7 Functional classification of the data

The Gene Ontology (GO) classification was used to identify the functions in each list that are statistically over-represented within the list (Ashburner et al. 2000). The GO annotation was done using the “find similar lists” function in GeneSpring. The Simplified Gene Ontology terms do not contain all the genes in term as defined by GO. The Simplified Gene Ontology terms are generated by parsing the gene annotation file, so they only contain genes which are probed on the array. The similarity was determined using a Bonferroni corrected hypergeometric distribution test to find the significance of the overlap of the user generated (or master list) to all the possible Simplified Gene Ontology lists (or test lists). Only statistically significant lists with p values of approximately 0.01 were selected for a manual review of the content before being included as an overlapping term. The full tables are included in Appendix B.

As described in that work, the task of identifying significant GO categories from a test list is automated by the GeneSpring software. Given the hierarchical nature of the GO system, however, not all groups are informationally equal with regards to understanding the function of the genes it contains. Since high-level categories like DNA Binding contain all genes with a known DNA binding function, they contain a large number of genes that are otherwise unrelated. As the level of the ontology descends the hierarchy, the shared commonalities of the contained genes increase, and the number of genes in each category decreases. Since the significance of the overlap between the two lists is calculated by the probability of the overlap occurring randomly, to identify a low-level category requires significant noise reduction. This was accomplished by

reclassifying the test lists into subgroups, using multiple parameters identified in the previous round of analysis.

4.5.8 Selection and analysis of the genes with largest change

Most of the genes that showed the largest changes in expression or chromatin state were excluded from consideration since they were induced or repressed by treatment, so the fold changes were artificially large. Of the genes that remained the top twenty genes that had the largest, reproducible within the replicates, changes in response to 5 Aza dC treatment for the condition were selected. This generates four lists: largest increase and largest decrease in expression and largest increase and largest decrease in the chromatin condensation. To interpret the potential biological context of the genes included on the lists, the list was first examined against genes with known links to cancer (Ashburner, Ball et al. 2000; Widschwendter and Jones 2002). After the genes with known cancer relationships were identified, the significance of the remaining genes on the list was examined using various gene annotation databases, primarily SOURCE and EMSEMBL (Lash, Tolstoshev et al. 2000; Hubbard, Barker et al. 2002; Safran, Solomon et al. 2002; Diehn, Sherlock et al. 2003). Using the expression pattern and biological function, it was possible to gain a great deal of insight into the biological significance of these changes. CpG island analysis was necessary when a CpG island was not included as part of the annotation. This was done using the EMBOSS “newcpgisland” tool using the default parameters of CpG count Observed/Expected of >0.6 and a minimum length of 200 bases (Gardiner-Garden and Frommer 1987; Larsen, Gundersen et al. 1992; Rice, Longden et al. 2000).

4.5.9 *Alternative Splicing analysis*

The data was sorted by common name to identify all genes with multiple probes on the array, and all the genes where the signal for the different probes for each gene was reproducible between microarrays were selected for further analysis. The intra-gene analysis was carried out manually in Entrez Gene, with the longest functional transcript used as the reference sequence and additional probes assigned to locations in the genomic slice (Pruitt, Katz et al. 2000). The chromatin array data was overlaid on the transcript structure to visualize the chromatin data for each region. By looking at changes across the locus and between treatments, the regions of the gene that showed chromatin differences could be identified since they indicated the possibility of separate regulatory regions. The DMD locus was divided based on the isoforms, since the probes are for alternative start sites, and the isoforms of interest are Dp427C, Dp427p, Dp260, Dp140, Dp71, and Dp40. (In figure 4.5, they are shown left to right, starting with Dp427p.) For the *hTert* locus, the probe AB016767 was assigned to intron 2, while the probes AB086950 and AF015950 were located in the ABG-deleted region and the full-length transcript, respectively.

4.5.10 *Expanded analysis of the hTert locus*

The sequences of the introns bordering the alternatively-spliced exons were imported from Ensembl and searched for motifs linked to splice form modulation (Hertel, Lynch et al. 1997). The intragenic CpG island analysis was done using the “newcpgisland” tool from EMBOSS (Rice, Longden et al. 2000), with the default parameters (Gardiner-Garden and Frommer 1987) on the genomic slice recovered from Ensembl (Hertel, Lynch et al. 1997). The exon location was overlaid on the results.

Bisulfite sequencing was used to validate the methylation of the region from exon 5 to 6, spanning the region in three steps. 24 clones were sequenced for each sub region and condition, to give a composite of the methylation state before and after treatment with 5 Aza dC. Treated and untreated H1299 (ATCC # CRL-5803), which is a non-small cell lung cancer (NSCLC), were used for comparison since, unlike MCF7, the line is known to express *hTert* strongly and repress most alternative splice forms (Yi, Tesmer et al. 1999). In addition, after 5 Aza dC treatment the expression is strongly reduced (Shames 2005).

4.5.11 Chromatin quantitative PCR analysis

The CQ-PCR primers and probes were designed from the last exon of each of the *PRDM2* isoforms (shown in figure 4.8). The sequence of the regions were BLASTed and all non-unique sequences were masked (Altschul, Gish et al. 1990). The remaining sequence was sent to ABI's custom assays group and the TaqMan® primers and probe (with a FAM fluorophore) were designed for a 60° C extension (table 4.11) (Applied Biosystems Group, Foster City, CA). The amplification was performed on an ABI 7700 in UTSW's Genomics Core's, using the ABI Mastermix with UGM, according the standard amplification protocol with 2 minutes at 50° C hold for UGM activation, followed by holding at 95° C for 10 minutes, the cycle was 15 seconds at 95° C followed by 1 minute at 60° C for extension, which was repeated 50 times. The standard curve was run using total DNA concentrations ranging from 100 ng to 0.001 ng per reaction, with the 10 ng and 1 ng per reaction having the best amplification curves. The experimental

samples were run in triplicate with 1 ng of DNA per reaction, using the same protocol as the test.

Table 4.11: TaqMan® primers and probes for the PRDM2 CQ-PCR validation
Short form

Forward primer	AGAGATTTTCGGAGGCCATCCT
Reverse primer	GACTGACCGTCCTCCAGAGA
FAM probe	TCAGCCCCGTCCGCC
Long form	
Forward primer	CCACTTGGGCGCTGTTTT
Reverse primer	CTCTGGGTTTGATAGCTCACTCTT
FAM probe	CTCAGCTCCAATTC

4.6 References

- Acevedo, L., J. Yu, et al. (2004). "A new role for Nogo as a regulator of vascular remodeling." *Nat Med* **10**(4): 382-8.
- Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-10.
- Anderson, N. L. and M. A. Gemmell (1984). "Protein-pattern changes and morphological effects due to methionine starvation or treatment with 5-azacytidine of the phorbol-ester-sensitive cell lines HL-60, CCL-119, and U-937." *Clin Chem* **30**(12 Pt 1): 1956-64.
- Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nat Genet* **25**(1): 25-9.
- Bailey, T. L. and C. Elkan (1994). "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.
- Batsche, E., M. Yaniv, et al. (2006). "The human SWI/SNF subunit Brm is a regulator of alternative splicing." *Nat Struct Mol Biol* **13**(1): 22-9.
- Bender, C. M., M. M. Pao, et al. (1998). "Inhibition of DNA methylation by 5-aza-2'-deoxycytidine suppresses the growth of human tumor cell lines." *Cancer Res* **58**(1): 95-101.
- Bird, A. (2002). "DNA methylation patterns and epigenetic memory." *Genes Dev* **16**(1): 6-21.
- Bird, A. P. (1980). "DNA methylation and the frequency of CpG in animal DNA." *Nucleic Acids Res* **8**(7): 1499-504.

- Bird, A. P., M. H. Taggart, et al. (1987). "Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene." Embo J **6**(4): 999-1004.
- Cameron, E. E., K. E. Bachman, et al. (1999). "Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer." Nat Genet **21**(1): 103-7.
- Carling, T., K. C. Kim, et al. (2004). "A histone methyltransferase is required for maximal response to female sex hormones." Mol Cell Biol **24**(16): 7032-42.
- Cheng, J. C., C. B. Matsen, et al. (2003). "Inhibition of DNA methylation and reactivation of silenced genes by zebularine." J Natl Cancer Inst **95**(5): 399-409.
- D'Angelo, M. A. and M. W. Hetzer (2006). "The role of the nuclear envelope in cellular organization." Cell Mol Life Sci.
- de Leon, M. B., C. Montanez, et al. (2005). "Dystrophin Dp71 expression is down-regulated during myogenesis: role of Sp1 and Sp3 on the Dp71 promoter activity." J Biol Chem **280**(7): 5290-9.
- de Vos, D. (2005). "Epigenetic drugs: a longstanding story." Semin Oncol **32**(5): 437-42.
- Debernardi, S., E. Fontanella, et al. (1997). "Identification of a novel human kinesin-related gene (HK2) by the cDNA differential display technique." Genomics **42**(1): 67-73.
- Devereux, T. R., I. Horikawa, et al. (1999). "DNA methylation analysis of the promoter region of the human telomerase reverse transcriptase (hTERT) gene." Cancer Res **59**(24): 6087-90.
- Diehn, M., G. Sherlock, et al. (2003). "SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data." Nucleic Acids Res **31**(1): 219-23.
- Ducrest, A. L., H. Szutorisz, et al. (2002). "Regulation of the human telomerase reverse transcriptase gene." Oncogene **21**(4): 541-52.
- Engler, D. A., G. Mohapatra, et al. (2006). "A Pseudolikelihood Approach for Simultaneous Analysis of Array Comparative Genomic Hybridizations (aCGH)." Biostatistics.
- Esteller, M. and J. G. Herman (2002). "Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours." J Pathol **196**(1): 1-7.
- Ferguson, A. T., P. M. Vertino, et al. (1997). "Role of estrogen receptor gene demethylation and DNA methyltransferase.DNA adduct formation in 5-aza-2'deoxyctidine-induced cytotoxicity in human breast cancer cells." J Biol Chem **272**(51): 32260-6.
- Fujie, T., M. Mori, et al. (1997). "Expression of MAGE and BAGE genes in Japanese breast cancers." Ann Oncol **8**(4): 369-72.
- Fujimoto, M., R. Kitazawa, et al. (2005). "Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression." Oncogene **24**(32): 5108-18.
- Furney, S. J., D. G. Higgins, et al. (2006). "Structural and functional properties of genes involved in human cancer." BMC Genomics **7**(1): 3.
- Furuta, J., Y. Umebayashi, et al. (2004). "Promoter methylation profiling of 30 genes in human malignant melanoma." Cancer Sci **95**(12): 962-8.

- Gama-Sosa, M. A., V. A. Slagel, et al. (1983). "The 5-methylcytosine content of DNA from human tumors." Nucleic Acids Res **11**(19): 6883-94.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.
- Ginzinger, D. G. (2002). "Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream." Exp Hematol **30**(6): 503-12.
- Gius, D., H. Cui, 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**(4): 361-71.
- Goffin, J. and E. Eisenhauer (2002). "DNA methyltransferase inhibitors-state of the art." Ann Oncol **13**(11): 1699-716.
- Gonzalogo, M. L., T. Hayashida, et al. (1998). "The role of DNA methylation in expression of the p19/p16 locus in human bladder cancer cell lines." Cancer Res **58**(6): 1245-52.
- Guilleret, I. and J. Benhattar (2003). "Demethylation of the human telomerase catalytic subunit (hTERT) gene promoter reduced hTERT expression and telomerase activity and shortened telomeres." Exp Cell Res **289**(2): 326-34.
- Guilleret, I., P. Yan, et al. (2002). "Hypermethylation of the human telomerase catalytic subunit (hTERT) gene correlates with telomerase activity." Int J Cancer **101**(4): 335-41.
- Hardiman, G. (2004). "Microarray platforms--comparisons and contrasts." Pharmacogenomics **5**(5): 487-502.
- Hark, A. T., C. J. Schoenherr, et al. (2000). "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." Nature **405**(6785): 486-9.
- Hatada, I., A. Kato, et al. (2002). "A microarray-based method for detecting methylated loci." J Hum Genet **47**(8): 448-51.
- Hembruff, S. L., D. J. Villeneuve, et al. (2005). "The optimization of quantitative reverse transcription PCR for verification of cDNA microarray data." Anal Biochem **345**(2): 237-49.
- Hertel, K. J., K. W. Lynch, et al. (1997). "Common themes in the function of transcription and splicing enhancers." Curr Opin Cell Biol **9**(3): 350-7.
- Hodge, D. R., B. Peng, et al. (2005). "Interleukin 6 supports the maintenance of p53 tumor suppressor gene promoter methylation." Cancer Res **65**(11): 4673-82.
- Huang, S. Y., M. B. Barnard, et al. (1986). "The active immunoglobulin kappa chain gene is packaged by non-ubiquitin-conjugated nucleosomes." Proc Natl Acad Sci U S A **83**(11): 3738-42.
- Huang, S. Y. and W. T. Garrard (1986). "Temperature-dependent cleavage of chromatin by micrococcal nuclease near the nucleosome center." FEBS Lett **199**(1): 89-91.
- Hubbard, T., D. Barker, et al. (2002). "The Ensembl genome database project." Nucleic Acids Res **30**(1): 38-41.
- Jarvinen, A. K., S. Hautaniemi, et al. (2004). "Are data from different gene expression microarray platforms comparable?" Genomics **83**(6): 1164-8.
- Jones, P. A. (2002). "DNA methylation and cancer." Oncogene **21**(35): 5358-60.
- Karpf, A. R., B. C. Moore, et al. (2001). "Activation of the p53 DNA damage response pathway after inhibition of DNA methyltransferase by 5-aza-2'-deoxycytidine." Mol Pharmacol **59**(4): 751-7.

- Kleinjan, D. A., A. Seawright, et al. (2004). "Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription." Dev Biol **265**(2): 462-77.
- Lapidos, K. A., R. Kakkar, et al. (2004). "The dystrophin glycoprotein complex: signaling strength and integrity for the sarcolemma." Circ Res **94**(8): 1023-31.
- Larsen, F., G. Gundersen, et al. (1992). "CpG islands as gene markers in the human genome." Genomics **13**(4): 1095-107.
- Lash, A. E., C. M. Tolstoshev, et al. (2000). "SAGEmap: a public gene expression resource." Genome Res **10**(7): 1051-60.
- Levitsky, V. G., O. A. Podkolodnaya, et al. (2001). "Nucleosome formation potential of exons, introns, and Alu repeats." Bioinformatics **17**(11): 1062-4.
- Lubbert, M. and M. Minden (2005). "Decitabine in acute myeloid leukemia." Semin Hematol **42**(3 Suppl 2): S38-42.
- Maglott, D., J. Ostell, et al. (2005). "Entrez Gene: gene-centered information at NCBI." Nucleic Acids Res **33**(Database issue): D54-8.
- McClelland, M. and R. Ivarie (1982). "Asymmetrical distribution of CpG in an 'average' mammalian gene." Nucleic Acids Res **10**(23): 7865-77.
- Momparler, R. L. (2005). "Pharmacology of 5-Aza-2'-deoxycytidine (decitabine)." Semin Hematol **42**(2 Suppl 2): S9-S16.
- Nieto, M., E. Samper, et al. (2004). "The absence of p53 is critical for the induction of apoptosis by 5-aza-2'-deoxycytidine." Oncogene **23**(3): 735-43.
- Nogues, G., S. Kadener, et al. (2002). "Transcriptional activators differ in their abilities to control alternative splicing." J Biol Chem **277**(45): 43110-4.
- Ogasawara, H., M. Okada, et al. (2003). "Possible role of DNA hypomethylation in the induction of SLE: relationship to the transcription of human endogenous retroviruses." Clin Exp Rheumatol **21**(6): 733-8.
- Oka, M., A. M. Meacham, et al. (2005). "De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine." Oncogene **24**(19): 3091-9.
- Olivier, M., S. Bautista, et al. (1998). "Relaxed cell-cycle arrests and propagation of unrepaired chromosomal damage in cancer cell lines with wild-type p53." Mol Carcinog **23**(1): 1-12.
- Parkinson, H., U. Sarkans, et al. (2005). "ArrayExpress--a public repository for microarray gene expression data at the EBI." Nucleic Acids Res **33**(Database issue): D553-5.
- Perry, A., T. Nobori, et al. (1997). "Detection of p16 gene deletions in gliomas: a comparison of fluorescence in situ hybridization (FISH) versus quantitative PCR." J Neuropathol Exp Neurol **56**(9): 999-1008.
- Pollack, J. R., C. M. Perou, et al. (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet **23**(1): 41-6.
- Pruitt, K. D., K. S. Katz, et al. (2000). "Introducing RefSeq and LocusLink: curated human genome resources at the NCBI." Trends Genet **16**(1): 44-7.
- Renaud, S., D. Loukinov, et al. (2005). "CTCF binds the proximal exonic region of hTERT and inhibits its transcription." Nucleic Acids Res **33**(21): 6850-60.
- Rice, P., I. Longden, et al. (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet **16**(6): 276-7.

- Robertson, K. D. (2005). "DNA methylation and human disease." Nat Rev Genet **6**(8): 597-610.
- Rose, S. M. and W. T. Garrard (1984). "Differentiation-dependent chromatin alterations precede and accompany transcription of immunoglobulin light chain genes." J Biol Chem **259**(13): 8534-44.
- Safran, M., I. Solomon, et al. (2002). "GeneCards 2002: towards a complete, object-oriented, human gene compendium." Bioinformatics **18**(11): 1542-3.
- Shames, D. S., Girard, L., Gao, B., Sato, M., Xie, X., Zochbauer-Muller, S., Pollack, J., Fong, K., Ramirez, R., Shay, J.W., Gazdar, A.F., Minna, J.D. (2005). "A genome-wide search for acquired promoter methylation in human lung cancer." In preparation.
- Sherlock, G., T. Hernandez-Boussard, et al. (2001). "The Stanford Microarray Database." Nucleic Acids Res **29**(1): 152-5.
- Shi, H., S. H. Wei, et al. (2003). "Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation." Cancer Res **63**(9): 2164-71.
- Spiropoulou, T., L. Ferekidou, et al. (2004). "Effect of antineoplastic agents on the expression of human telomerase reverse transcriptase beta plus transcript in MCF-7 cells." Clin Biochem **37**(4): 299-304.
- Szutorisz, H., J. Lingner, et al. (2003). "A chromosome 3-encoded repressor of the human telomerase reverse transcriptase (hTERT) gene controls the state of hTERT chromatin." Cancer Res **63**(3): 689-95.
- Taddei, A., F. Hediger, et al. (2004). "The Function of Nuclear Architecture: A Genetic Approach." Annu Rev Genet.
- Ulaner, G. A., J. F. Hu, et al. (1998). "Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts." Cancer Res **58**(18): 4168-72.
- Warters, R. L. and B. W. Lyons (1992). "Variation in radiation-induced formation of DNA double-strand breaks as a function of chromatin structure." Radiat Res **130**(3): 309-18.
- Weber, J., M. Salgaller, et al. (1994). "Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine." Cancer Res **54**(7): 1766-71.
- Weil, M. R., P. Widlak, et al. (2004). "Global survey of chromatin accessibility using DNA microarrays." Genome Res **14**(7): 1374-81.
- Widschwendter, M. and P. A. Jones (2002). "DNA methylation and breast carcinogenesis." Oncogene **21**(35): 5462-82.
- Wingender, E., X. Chen, et al. (2000). "TRANSFAC: an integrated system for gene expression regulation." Nucleic Acids Res **28**(1): 316-9.
- Woo, Y., J. Affourtit, et al. (2004). "A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms." J Biomol Tech **15**(4): 276-84.
- Yi, X., V. M. Tesmer, et al. (1999). "Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels." Mol Cell Biol **19**(6): 3989-97.

- Yi, X., D. M. White, et al. (2000). "An alternate splicing variant of the human telomerase catalytic subunit inhibits telomerase activity." Neoplasia **2**(5): 433-40.
- Zhang, X. H. and L. A. Chasin (2004). "Computational definition of sequence motifs governing constitutive exon splicing." Genes Dev **18**(11): 1241-50.
- Zhang, Z., P. M. Harrison, et al. (2003). "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." Genome Res **13**(12): 2541-58.

Chapter 5

Improvements to the chromatin array method and its potential for expanded application

5.1 Development of an optimized microarray platform for the study of chromatin

In the previous chapters it was shown that the chromatin array can measure the chromatin state of genomic regions that do not express transcript, and that the chromatin state can vary greatly across a single locus. However, as discussed in Chapter 1 most commercial microarray platforms are designed to study only expressed sequences, and the probe for each gene is biased to the 3' end of the expressed sequence. Recently some microarrays, like those manufactured by MWG, have been designed to include multiple probes per gene so that processes like alternative splicing can be studied. While this improvement creates the opportunity to study intergenic chromatin variation on a small scale using the chromatin array method, even these new generation commercial microarrays still have too few probes per gene and the probes are still biased to include regions of the genome that are expressed. As with ChIP on chip, some of the most interesting chromatin array results may come from studying promoters or intronic regions, since they encode regulatory functions. Therefore, like ChIP on chip, the optimal chromatin array microarray platform uses an oligonucleotide probe set, to ensure high resolution that is tiled, to ensure complete coverage of the genome.

Since tiling the entire human genome without gaps is beyond current microarray production technology, an informatics approach must be used to optimize the design (Bertone, Gerstein et al. 2005). This approach would allow a probe set to be defined that has minimal gaps in regions of interest, and larger gaps in regions where genetic features such as repetitive elements will prevent accurate measurement of the chromatin state of

the region. The minimum tiling array for chromatin array use should include probes for the promoter, exon/intron boundaries, at least one unique region for each exon and intron, and the UTRs of each known transcript. The remaining probes should tile intragenic space, focusing on regions where sequence conservation between species is higher than expected. Using conservation between species the possibility of identifying non-translated or even non-coding transcripts which comprise the genetic “dark matter” is increased versus a shotgun approach (Johnson, Edwards et al. 2005). This approach would also make it possible to identify regions where the DNA’s interaction with proteins serves a regulatory function separate from transcription, such as distant promoter sites, chromatin loop domains or even replication start sites (Djeliova, Russev et al. 2001; Heng, Krawetz et al. 2001). By limiting the coverage to genomic regions that are likely to be of interest, this microarray could be constructed with current light-directed oligonucleotide synthesis technology (Bertone, Stolc et al. 2004). This microarray will have applications beyond chromatin array applications, since probes will also be applicable to aCGH, ChIP on chip, and expression (with an emphasis on alternative splicing and non-coding RNAs) experiments.

5.2 Development of methods to enable the analysis of the chromatin state of tissue samples

One of the most important advancements for the chromatin array method is its translation from being a research tool to a diagnostic tool. Chromatin state-based diagnostic tests would have wide application since cancer, heart disease, non-insulin dependent Diabetes, autoimmune and most aging related diseases have epigenetic components (Jiang, Bressler et al. 2004; Robertson 2005; Scarano, Strazzullo et al. 2005).

Given that the epigenetic state can be altered, diseases with an epigenetic causes should be much more treatable than diseases with a purely genetic or environmental cause (Rodenhiser and Mann 2006). In addition chromatin state based diagnostics could also be used to diagnose and treat imprinting disorders like Angelman's or map the changes caused by the loss of *MECP2* in Rett's syndrome (Jiang, Bressler et al. 2004; Lopez-Rangel and Lewis 2006). To create a chromatin-based diagnostic, several significant challenges must be overcome. The first is that biopsies and tissue samples are generally frozen or in some way preserved. However, the chromatin state is actively maintained, so even using snap freeze cryopreservation, after thawing the chromatin state is permanently altered (Van Blerkom 1989; Horowitz, Giannasca et al. 1990). Therefore, using current protocols, in order to study the chromatin state of cells obtained from tissue samples, a cell line must first be established. Given the amount of effort and time required to establish a cell line this is a difficult, if not impossible step, which precludes the use of the chromatin array for large-scale diagnostic purposes. Furthermore, even if it was possible to rapidly generate cell lines from patient samples, they do not perfectly mimic the natural cellular state, so again the chromatin state measurements would be perturbed (Perou, Sorlie et al. 2000). However, it may be possible that by using methods developed to freeze embryos, a "chromatin-safe" means of sample preservation can be created, which would permit the development of chromatin state-based diagnostic methods (Van Blerkom and Davis 1994).

To create a chromatin safe preservation technique, the normal chromatin state of a variety of cell lines must first be established. These cell lines can then be used to test various standard preservation methods (such as formalin fixation, or snap freezing) to

determine how well each can preserve the chromatin state. The best standard preservation method can then be used as a starting point to design a special method to minimize the damage to the chromatin state associated with sample preservation. Even if the best preservation method still causes changes to the chromatin state, as long as the changes are nonrandom, it is still possible to at least correlate the preserved chromatin state with normal state. This will allow the creation of a diagnostic chromatin test that can use preserved cells to measure at least a subset of genes.

Once a method to preserve the chromatin state is developed, the next step will be to test the effects of cellular heterogeneity on the chromatin array measurement. This step is a substantial challenge with translating most molecular biology techniques from samples derived from cell lines to samples derived from tissue, since tissues and tumors are not a homogenous mass of a single cell type. It will therefore be necessary to determine how dilute the cells of interest can be before the signal can no longer be detected. The initial work will be done using populations derived from mixing cells from a tumor line with cells from normal lines. This work will establish the maximum amount of cellular heterogeneity before the signal becomes undetectable. By correlation this number to the purity obtained by several tissue preparation methods it can be established if gross sectioning of the sample will suffice for a chromatin diagnostic or if microdissection or laser capture is necessary. If the purity required for a diagnostic based on changes in chromatin state is too high to be practical, then the diagnostic can be developed with uses the presence or absence of a particular sequence in a particular fraction similar to using DNA methylation as a biomarker (Ushijima 2005).

While the cell line-based validation will help establish lower limits of detection a great deal of testing using actual tissue samples will be required to optimize the method. However, tissue and tumor samples derived from patients are too precious; validation of the diagnostic will use tumors grown as xenografts in mice. Xenografts are a readily available source of tumor material and because the tumors can be derived from same cell lines used in the previous steps, the cross-validation will be greatly simplified. The greatest benefit of using xenograft material is that while the tumor mass will be very heterogeneous, the cancerous cell lines will be nearly clonal, and all the invasive tissues will be of host (mouse) origin. While this seems like host tissue contamination would complicate the analysis, it actually greater simplifies it. By adding probes to the microarray designed pick up regions of interest (both cancer associated regions and regions that will identify the contaminating tissue types) in the mouse DNA but not the human DNA, the signal can be deconvoluted computationally into the tumor and host components, and the amount of heterogeneity estimated. Information about the amount and type of host contamination of tumor will greatly improve the chances of creating a viable diagnostic-based on chromatin state since similar contamination is expected in human tissues as well. This data will also provide large amounts of unique data about host/tumor interaction.

The next area where improvements will need to be made is to reduce the number of cells required for the chromatin fractionation step. The current chromatin fractionation protocol calls for tens of millions of cells, but yields about a hundred micrograms of fractionated DNA. Since far less DNA is needed for a diagnostic test, the number of cells used for each fractionation can be reduced. Scaling back the number of cells that are

required should be a relatively simple matter, since $\frac{3}{4}$ scale experiments have already been successfully used. Reduced scale fractionations will likely not provide enough of a reduction in required sample size since the absolute minimum number of cells required is still based on how many cells are required to yield enough fractionated DNA (micrograms) to run even a single chromatin array. While chromatin array work requires micrograms of fractionated DNA per microarray and therefore a substantial amount of tissue, chromatin QPCR, which was discussed in Chapter 4 only requires 1 ng of DNA per reaction, so even biopsies might provide enough material. This means that once the sample preservation and chromatin fractionation scale issues are dealt with, a diagnostic test, based on the chromatin state measurements of a select group of “genes” is possible even from samples obtained, for example by needle biopsy.

5.3 Surveys to identify signatures of various cellular states

As discussed in Chapter 1, the major niche of expression microarray technology is for survey-type experiments where the hypothesis is: by testing enough variations of the variable of interest, discoveries can be made. This experimental model is quite applicable to chromatin array work as well and since the chromatin state is presumably more directly related to phenotype than expression, the potential for discovery is correspondingly greater (Rauscher 2005). By building on the work discussed in previous chapters using the chromatin array method and several additional breast cancer model cell lines, may be possible to make significant advances in the understanding of this disease. The following is two examples of how such a survey experiment might be conducted.

5.3.1 Breast cancer surveys

By focusing on a small number of carefully selected tumor model cell lines the goal of this experiment would be to identify the chromatin state changes associated with breast cancer and to use this knowledge to advance the treatment of the disease. The first step in this breast cancer survey experiment is to measure the chromatin state of “normal” mammary epithelial cells using cultured human mammary epithelial cell lines (Perou, Sorlie et al. 2000; Stampfer and Yaswen 2000). By knowing the chromatin state of the progenitor cells, and comparing the “normal” chromatin state to the chromatin state of the cancer cell lines, the cancer-associated chromatin changes can be identified. From this data alone, may be possible to gain substantial insight into the chromatin state changes and therefore transcriptional regulatory changes associated with breast cancer. In addition, the normal chromatin state can serve as a baseline, and the chromatin state of the other cell lines in the survey referenced to the normal condensation level. This will allow comparison of chromatin states from cell lines that are radically different, and possibly allow for chromatin array-based histological classification or staging (Subramaniam and Isaacs 2005).

The survey should then be expanded to include more breast cancer model cell lines, with a special emphasis on measuring the differences between estrogen receptor positive and negative cell lines. By studying the chromatin state differences and commonalities between these two disease sub-types, may be possible to better understand the significant differences in the response to treatment, and survival, between these two types of breast cancer (Gutierrez, Detre et al. 2005; Osborne and Schiff 2005). To

maximize the likelihood of discovery, the selection of the cell lines in the study will be guided by analyzing the results of published breast cancer expression microarray surveys (Perou, Sorlie et al. 2000; van 't Veer, Dai et al. 2002). The rationale for this is that expression is coupled to chromatin state, so by selecting cell lines where the expression patterns that are associated with the phenotypes of interest have been established, the chromatin array survey can build upon this knowledge. The chromatin array method is applied in a hypothesis driven experiment to elucidate the causes of the previously observed gene expression changes and identify critical regulatory pathways and genes that might be drug targets. While the primary experimental goals are hypothesis driven, chromatin state can be measured even if expression microarray is not capable to measuring a signal, so the chromatin array data will be a discovery tool as well. Given the number of possible target genes discussed in the previous two chapters, this type of approach has already been proved to be successful.

Following the completion of “normal” breast cancer chromatin survey, it would be useful to treat many of the cell lines used in the survey with 5 Aza dC or other epigenetic modifying agents like TSA. These treatments would determine which genes are susceptible to modulation by epigenetic therapies, and if the non-diseased chromatin state can be reestablished. The goal of epigenetic therapies is to create “gentle” chemotherapeutics, which would induce inactive genes and make the tumor cells kill themselves with no direct cytotoxic effects (Dowell and Minna 2004; Rosenfeld 2005). However, since 5 Aza dC and most other epigenetic modifying drugs have such global effects their use as chemotherapeutics is limited, but their role as discovery tools is just starting to become clear (Mompalmer 2005).

By combining epigenetic therapy and the chromatin array assay it is possible to identify the genomic regions that are altered by the epigenetic therapies. Then once the regions or genes of interest are identified, agents that mimic the desired epigenetic effects of the original drug can be developed. These mimetic agents, which can be as simple as interfering RNAs or as complex as novel small molecules that target the region and disrupt the chromatin state represent significant opportunity for creating new therapies (McLaughlin, Finn et al. 2003; Kawasaki and Taira 2005). These targeted epigenetic drugs would have a reduced risk of side effects, and might fulfill the promise of epigenetic therapy (Esteller 2005).

As previously discussed, drugs like 5 Aza dC, have limited use as chemotherapeutics, since 5 Aza dC also a known carcinogen, but 5 Aza dC is used to treat diseases like hairy cell leukemia and some solid tumors (Prasanna, Shack et al. 1995; Lubbert and Minden 2005; Momparler 2005). However, very little is known about the side effects of such treatment, especially over long periods of time. Therefore, in addition to the drug target discovery, treatment of these cell lines with epigenetic modifying drugs will allow a great deal of information about the long term effects of these drugs to be gathered. To accomplish this, the small fraction of cells that survive the 5 Aza dC treatment can be cultured post treatment until the population becomes semi-clonal. Then the chromatin states before, during and after treatment is compared to determine if the chromatin state prior to treatment is recapitulated, and if not, what changes from the original state are observed. This data would address how resilient the information stored in chromatin state is to perturbation, and possibly serve as a

accelerated model for longer term side effects following 5 Aza dC (or other epigenetic drug) treatment.

5.3.2. Development of a chromatin biosignature for other cancers

As for the breast cancer survey, other cancer or disease model cell lines can be surveyed to understand what role chromatin state alterations play in the disease process. The goal of these types of surveys is that by measuring the chromatin state across a large group of cell lines that have a histologically similar disease state, the molecular causes of the disease can be elucidated, in order to create a chromatin state-based molecular pathology for the disease. For example, by surveying a large number of Small Cell Lung Cancer (SCLC) cell lines, a SCLC chromatin state model can be built using the commonalities of all the SCLCs included in the survey, with the differences between the chromatin states and their frequencies mapped by region, as no two cell lines will be the exactly the same (Oshita, Ikehara et al. 2004). In addition, comparative expression studies of these same cell lines will provide context to interpret the chromatin state data. The SCLC model can be compared to the models derived from similar surveys of other lung cancers and normal cell lines, so that the SCLC-specific chromatin signature can, for example, be distinguished from a general lung cancer chromatin signature and the chromatin signature of normal lung. The use of subtractive analysis methods will reduce the data volume and increase the apparent complexity so that the variable of interest (like treatment response or metastatic potential) will be enriched thereby making it easier to identify its unique signature.

Developing a unified model of the epigenome would require extensive studies of the chromatin condensation state, since the chromatin state of every cell type and disease

process would have to be individually measured (Jones and Martienssen 2005). A tractable beginning to that process that would have immediate benefits would be a chromatin array study of the cell lines included in the NCI 60 (Ross, Scherf et al. 2000). This survey would allow for the building of a model of cancer chromatin, since the chromatin features that are generic to the cancer chromatin state can be separated from the tissue and cell-line specific chromatin features. Then, as technology improves, the cell lines used for other comparative expression surveys, like the Normal Tissue Atlas, could be included to understand the “normal” chromatin state (Su, Cooke et al. 2002). By studying a large number cell lines that approximate the normal chromatin state, they represent a significant opportunity to advance the understanding of how epigenetic processes contribute to transcriptional regulation and disease in a controlled experimental environment.

5.4 Correlating ChIP on chip results with chromatin array

ChIP on chip and chromatin array are both methods to study chromatin using a microarray-based platform (Das, Ramachandran et al. 2004; Weil, Widlak et al. 2004). While it would seem to be redundant to have two microarray-based platforms it is not. As discussed in Chapter 1, ChIP on chip is used to study the protein component of the chromatin, and chromatin array is used to measure the chromatin condensation state. Given the often direct relationship between the chromatin condensation state and the proteins bound to the chromatin, these two types of information are highly complementary if not synergistic (Jones and Martienssen 2005). The additional knowledge gained by actually correlating a protein binding or post-translational

modification event to the chromatin condensation state, would dramatically increase the understanding of chromatin regulation. From the structural genomics perspective combining ChIP on chip and chromatin array would greatly facilitate the discovery and characterization of new chromatin proteins and allow the function of chromatin remodeling complexes to be understood in context.

The difficulty is in selecting the chromatin proteins of interest since current GO classification lists more than a thousand proteins as chromatin or chromatin binding, and this doesn't even include post-translational modifications (Ashburner, Ball et al. 2000). Therefore any survey type approach would be severely limited since ChIP on chip is more hypotheses driven than most other types of microarray, since a protein of interest must be chosen in advance and everything is in the context of that choice (Das, Ramachandran et al. 2004). However, there are certain proteins and post-translational modifications (like the modification state of the lysines in the tails of the H3 and H4 histones or the location of CTCF) that stand out as obvious choices for a combined chromatin array and ChIP on chip study (Jenuwein and Allis 2001; Lewis and Murrell 2004).

A pilot project similar to the one discussed in Chapter 3, for developing the chromatin array is well suited, since many obstacles must be overcome to create a working combined method. Accordingly a single cell line (like MCF7) would be used, along with a carefully selected set of antibodies for the ChIP. The tiling microarray discussed in section 5.1 would be used for both the ChIP on chip and the chromatin array, because if chromatin state varies intragenically it is likely that the chromatin protein components do as well. Custom analysis and data display methods to link chromatin state

to protein state will also need to be created. While this is a massive amount of effort since the cell culture work alone is non-trivial, the amount of information about chromatin regulation that can be obtained from even this small scale experiment more than warrants effort.

Once a workflow for the wet work is in place and the data analysis schema is complete, scaling up is a simple matter since the antibodies and other reagents will have already been purchased and validated. As with the development of the chromatin array the second step should be comparing induced chromatin change in same cell line, since having a reference to understand the differences will greatly facilitate the analysis and interpretation of the data. Then once methods to compare cell lines are developed expanding the study becomes an incremental effort, and could become a component of the surveys discussed in section 5.3.

Of potentially greater utility is a hybrid of chromatin array and ChIP, which would replace the lysis step in traditional ChIP protocols with chromatin fractionation. As seen in figure 3.4 the chromatin fractionation is gentle enough for the chromatin associated proteins to remain bound to the chromatin and likely compatible with the crosslinking step used in ChIP. By separating the chromatin into fractions by its transcriptional state, then performing ChIP on each fraction separately the accuracy of the measurements can be increased since for most chromatin proteins fractionation by chromatin condensation state will represent an enrichment step (Cosgrove and Wolberger 2005). Additionally the context provided by the separation chromatin state is informative as to the nature and functions of the protein binding events, since many chromatin

proteins have roles both in transcriptional repression, and induction depending the other proteins in the complex (Roeder 2005).

5.4.1 The SouthWestern Array

An extremely difficult but potentially rewarding extension of the ChIP on chip and chromatin array correlation work is called the SouthWestern Array. The SouthWestern Array is an array-based hybrid of Southern and Western blots, or the chromatin array and ChIP on chip (Das, Ramachandran et al. 2004; Westermeier and Marouga 2005). The Southwestern array would allow the near simultaneous measurement of the chromatin state and the proteins that were associated with the region. This information is critical not only to building a model of transcriptional regulation, but also to understanding the epigenome (Rauscher 2005). The SouthWestern array would use covalent cross-linking to link the proteins to the DNA while the cell is still “alive”, much like ChIP on chip. The Protein/DNA hybrid is then recovered using chromatographic methods since the molecular weight of the DNA bound to the protein will make it straightforward to separate from the other proteins. The DNA is end labeled with a chemically modified base so a fluorescent label can be attached without amplification or synthesis of a new strand. This mixture is allowed to hybridize to a conventional microarray. The initial read-out is the quantity of DNA bound to protein, regardless of protein type, separated by the DNAs sequence. Antibodies to the proteins of interest are added, and the read out is protein level sorted by probe.

This makes the SouthWestern array unique since, unlike ChIP on chip, the proteins of interest need not be decided in advance. Furthermore, so long as the

fluorescence emission and excitation spectra do not overlap, there is no limit to the number of antibodies that can be used simultaneously (though the practical limit is six) (Huebschman, Schultz et al. 2002). Not only can the protein levels can be directly determined for each sub-gene region or possibly across the entire genome, but it also becomes feasible to have controls similar to Western blots, still.

The SouthWestern Array is conceptually very simple, but the reduction to practice has proven very difficult, since unbound DNA “out competes” the sterically hindered protein/DNA hybrid, and the sandwich-style second hybridization requires conditions that allow antibody binding but do not allow the hybridized DNA to melt off the probes. While the creation of the SouthWestern array would combine the best aspects of ChIP on chip with the chromatin array, there are a number of extremely complex technical challenges (like designing a standard protocol to efficient recover the DNA/Protein hybrid, and the creation of a method to attach enough fluors to the DNA that it in the sensitivity range of standard microarray scanners) that must be overcome before such a method is feasible.

5.5 References

- Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-9.
- Bertone, P., M. Gerstein, et al. (2005). "Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery." Chromosome Res **13**(3): 259-74.
- Bertone, P., V. Stolc, et al. (2004). "Global identification of human transcribed sequences with genome tiling arrays." Science **306**(5705): 2242-6.
- Cosgrove, M. S. and C. Wolberger (2005). "How does the histone code work?" Biochem Cell Biol **83**(4): 468-76.

- Das, P. M., K. Ramachandran, et al. (2004). "Chromatin immunoprecipitation assay." Biotechniques **37**(6): 961-9.
- Djeliova, V., G. Russev, et al. (2001). "Dynamics of association of origins of DNA replication with the nuclear matrix during the cell cycle." Nucleic Acids Res **29**(15): 3181-7.
- Dowell, J. E. and J. D. Minna (2004). "Cancer chemotherapy targeted at reactivating the expression of epigenetically inactivated genes." J Clin Oncol **22**(8): 1353-5.
- Esteller, M. (2005). "Dormant hypermethylated tumour suppressor genes: questions and answers." J Pathol **205**(2): 172-80.
- Gutierrez, M. C., S. Detre, et al. (2005). "Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase." J Clin Oncol **23**(11): 2469-76.
- Heng, H. H., S. A. Krawetz, et al. (2001). "Re-defining the chromatin loop domain." Cytogenet Cell Genet **93**(3-4): 155-61.
- Horowitz, R. A., P. J. Giannasca, et al. (1990). "Ultrastructural preservation of nuclei and chromatin: improvement with low-temperature methods." J Microsc **157** (Pt 2): 205-24.
- Huebschman, M. L., R. A. Schultz, et al. (2002). "Characteristics and capabilities of the hyperspectral imaging microscope." IEEE Eng Med Biol Mag **21**(4): 104-17.
- Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.
- Jiang, Y. H., J. Bressler, et al. (2004). "Epigenetics and human disease." Annu Rev Genomics Hum Genet **5**: 479-510.
- Johnson, J. M., S. Edwards, et al. (2005). "Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments." Trends Genet **21**(2): 93-102.
- Jones, P. A. and R. Martienssen (2005). "A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop." Cancer Res **65**(24): 11241-6.
- Kawasaki, H. and K. Taira (2005). "Transcriptional gene silencing by short interfering RNAs." Curr Opin Mol Ther **7**(2): 125-31.
- Lewis, A. and A. Murrell (2004). "Genomic imprinting: CTCF protects the boundaries." Curr Biol **14**(7): R284-6.
- Lopez-Rangel, E. and M. E. Lewis (2006). "Further evidence for epigenetic influence of MECP2 in Rett, autism and Angelman's syndromes." Clin Genet **69**(1): 23-5.
- Lubbert, M. and M. Minden (2005). "Decitabine in acute myeloid leukemia." Semin Hematol **42**(3 Suppl 2): S38-42.
- McLaughlin, F., P. Finn, et al. (2003). "The cell cycle, chromatin and cancer: mechanism-based therapeutics come of age." Drug Discov Today **8**(17): 793-802.
- Momparler, R. L. (2005). "Pharmacology of 5-Aza-2'-deoxycytidine (decitabine)." Semin Hematol **42**(2 Suppl 2): S9-S16.
- Osborne, C. K. and R. Schiff (2005). "Estrogen-receptor biology: continuing progress and therapeutic implications." J Clin Oncol **23**(8): 1616-22.
- Oshita, F., M. Ikehara, et al. (2004). "Genomic-wide cDNA microarray screening to correlate gene expression profile with chemoresistance in patients with advanced lung cancer." J Exp Ther Oncol **4**(2): 155-60.

- Perou, C. M., T. Sorlie, et al. (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-52.
- Prasanna, P., S. Shack, et al. (1995). "Phenylacetate in chemoprevention: in vitro and in vivo suppression of 5-aza-2'-deoxycytidine-induced carcinogenesis." Clin Cancer Res **1**(8): 865-71.
- Rauscher, F. J., 3rd (2005). "It is time for a Human Epigenome Project." Cancer Res **65**(24): 11229.
- Robertson, K. D. (2005). "DNA methylation and human disease." Nat Rev Genet **6**(8): 597-610.
- Rodenhiser, D. and M. Mann (2006). "Epigenetics and human disease: translating basic biology into clinical applications." Cmaj **174**(3): 341-8.
- Roeder, R. G. (2005). "Transcriptional regulation and the role of diverse coactivators in animal cells." FEBS Lett **579**(4): 909-15.
- Rosenfeld, C. S. (2005). "Clinical development of decitabine as a prototype for an epigenetic drug program." Semin Oncol **32**(5): 465-72.
- Ross, D. T., U. Scherf, et al. (2000). "Systematic variation in gene expression patterns in human cancer cell lines." Nat Genet **24**(3): 227-35.
- Scarano, M. I., M. Strazzullo, et al. (2005). "DNA methylation 40 years later: Its role in human health and disease." J Cell Physiol **204**(1): 21-35.
- Stampfer, M. R. and P. Yaswen (2000). "Culture models of human mammary epithelial cell transformation." J Mammary Gland Biol Neoplasia **5**(4): 365-78.
- Su, A. I., M. P. Cooke, et al. (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-70.
- Subramaniam, D. S. and C. Isaacs (2005). "Utilizing prognostic and predictive factors in breast cancer." Curr Treat Options Oncol **6**(2): 147-59.
- Ushijima, T. (2005). "Detection and interpretation of altered methylation patterns in cancer cells." Nat Rev Cancer **5**(3): 223-31.
- van 't Veer, L. J., H. Dai, et al. (2002). "Gene expression profiling predicts clinical outcome of breast cancer." Nature **415**(6871): 530-6.
- Van Blerkom, J. (1989). "Maturation at high frequency of germinal-vesicle-stage mouse oocytes after cryopreservation: alterations in cytoplasmic, nuclear, nucleolar and chromosomal structure and organization associated with vitrification." Hum Reprod **4**(8): 883-98.
- Van Blerkom, J. and P. W. Davis (1994). "Cytogenetic, cellular, and developmental consequences of cryopreservation of immature and mature mouse and human oocytes." Microsc Res Tech **27**(2): 165-93.
- Weil, M. R., P. Widlak, et al. (2004). "Global survey of chromatin accessibility using DNA microarrays." Genome Res **14**(7): 1374-81.
- Westermeyer, R. and R. Marouga (2005). "Protein detection methods in proteomics research." Biosci Rep **25**(1-2): 19-32.

APPENDIX A

Appendix A The MCF7 chromatin fractionation protocol

Buffer A

20mM Tris pH 7.5
3mM MgCl₂
10mM KCl
0.5mM EGTA
1mM DTT
0.25M Sucrose
0.5% NP-40

Buffer B

20mM Tris pH 7.5
3mM MgCl₂
10mM KCl
1mM DTT
0.25M Sucrose

Stop Buffer

1% SDS
80mM EDTA
50ug Proteinase K

Harvest:

Cells to be harvested should be ~80% confluent. The number of cells required varies by line but more is better since you can do multiple MNase concentrations with the extra.

(All volumes given are for 150 mm dishes.)

Suction off the media, and wash the dish with 3 mL warm PBS to remove the remaining media.

Use 2-3 mL warm trypsin 0.05% and incubate at RT or 37 C if the cells are resistant to trypsin.

Once the cells begin to come off (the trypsin will be come cloudy) neutralize the trypsin with an equal or greater volume of media. Use the pipette to loosen the cells remaining on dish by rinsing.

Pipette the suspended cell up and down repeatedly to break up any clumps.

Combine all the harvests into a conical vial, and if necessary rinse the plates again with 5 mL of media to recover the remaining cells.

Spin the cells at 1,000 RPM for 5 minutes at 4 C

Suction off the media, and resuspend the pellet in 5 mL of 4 C PBS, pipetting up and down to break up the clumps. Add an additional 15 mL of 4 C PBS

Spin the cells at 1000 RPM for 5 minutes at 4 C

Suction off the PBS, and resuspend the pellet in 5 mL of 4 C PBS, pipetting up and down to break up the clumps. Add an additional 15 mL of 4 C PBS.

Spin the cells at 1000 RPM for 5 minutes at 4 C

Lysis:

Suction off the PBS, and resuspend the pellet in 5 mL of 4 C Buffer A (made up fresh), pipetting up and down to break up the clumps. Add an additional 15 mL of 4 C Buffer A.

Incubate for 10 minutes on ice

Spin the nuclei at 1000 RPM for 5 minutes at 4 C, (if the buffer is flocculent increase the speed of the centrifuge to 1200 RPM and spin for 2 additional minutes).

Suction off the Buffer A, and resuspend the pellet in 5 mL of 4 C Buffer B (made up fresh), pipetting up and down to break up the clumps.

If the cell line produces stable nuclei add an additional 15 mL of 4 C Buffer B and repeat the wash after spinning with an additional 20mL of Buffer B, if the cell line is known to be difficult to work with do a single wash with 40mL of Buffer B.

Spin the nuclei at 1000 RPM for 5 minutes at 4 C, (if the buffer is flocculent increase the speed of the centrifuge (up to 1800 RPM) and spin for 2 additional minutes).

Resuspend the nuclei in Buffer B (0.5mg/mL is optimal.) Use the 260 absorbance to determine the conc of chromatin since the nuclei are mostly DNA.

Add 4 volumes Nuclei to 1 volume Buffer B, with 5mM CaCl₂, and Mircococcal Nuclease (MNase)

The optimum concentrations of MNase are between 0.3 and 0.8 U/uL, but the required amount must be determined experimentally. 0.4U/uL is a good starting place for MCF7.

Incubate at 33 C for 10 min, invert to the tube to mix every 3 minutes, but do not vortex. (The remaining volume of the nuclei is mixed with 400 uL of buffer B and 100 uL of Stop buffer, and incubated at 52 C.)

The digestion is stopped by adding EGTA, and incubated for 10 minutes on ice.

The tube is centrifuged at 3000 RPM for 3 minutes and 4 C.

S1

The supernatant is carefully removed (the pellet must not be left wet) and saved in a new tube.

The supernatant is centrifuged at 14,000 RPM at 4 C for 5 minutes.

The supernatant is removed to a new tube (labeled S1) after the addition of 100 uL of Stop buffer. (The pellet is discarded)

S2

The S1 pellet is resuspended in 500 uL of 2 mM EDTA, and incubated on ice for 10 minutes. Centrifuge at 14,000 RPM for 5 minutes at 4 C

Remove the supernatant, and save in a new tube labeled S2, with 100 uL of Stop buffer.

P

The S2 pellet is resuspended in 500 uL of Buffer B and 100 uL of Stop buffer and the tube labeled P.

The tubes are incubated over night at 52 C, before phenol prepping.

APPENDIX B

Appendix B Supplemental data from: 5-AZA-2'-DEOXYCYTIDINE INDUCED CHROMATIN MODULATION SHOWS DIFFERENTIAL REGULATION AT THE SUB-GENE LEVEL

All supplemental data tables are included electronically

Supplementary Table 1: Expression increased by Aza.....	181
Supplementary Table 2: Expression down by Aza.....	248
Supplementary Table 3 Chromatin down by Aza.....	259
Supplementary Table 4: Chromatin condensed by treatment	290
Supplementary table 5: Genes with multiple probes.....	298

APPENDIX B
**Appendix B Supplemental data from: 5-AZA-2'-DEOXYCYTIDINE
INDUCED CHROMATIN MODULATION SHOWS DIFFERENTIAL
REGULATION AT THE SUB-GENE LEVEL**

Supplementary Table 1: Expression increased by Aza

Genbank	Description
AA001203	cancer susceptibility candidate 3
AA001414	serine/threonine kinase 35
AA031528	hypothetical protein FLJ20280
AA044835	solute carrier family 35, member F5
AA046424	peroxisomal acyl-CoA thioesterase 2B
AA046650	Tara-like protein
AA053830	C-terminal binding protein 1
AA058770	glucocorticoid induced transcript 1
	Transcribed sequence with strong similarity to protein ref:NP_002745.1 (H.sapiens) mitogen-activated protein kinase 13; mitogen-activated protein kinase p38 delta; stress-activated protein kinase 4 [Homo sapiens]
AA088543	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme A
AA115117	z117c11.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
AA133277	IMAGE:502196 3', mRNA sequence.
AA135522	KIAA0089 protein
AA143793	Rab coupling protein
AA148301	COMM domain containing 7
AA149594	TGFB inducible early growth response 2
AA150460	KIAA1404 protein
	zo02d03.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
AA151917	IMAGE:566501 3', mRNA sequence.
AA156240	retinoic acid induced 3
AA156777	KIAA1219 protein
AA156961	helicase, ATP binding 1
AA160181	HDCMA18P protein
AA161130	hypothetical protein MGC3121
AA164751	tumor necrosis factor receptor superfamily, member 6
AA176780	tripartite motif-containing 44
AA187963	KIAA1228 protein
AA194149	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1
AA195017	KIAA1609 protein
AA195124	KIAA1609 protein
	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1
AA195259	mitogen-activated protein kinase 1
AA195999	MRNA; cDNA DKFZp586K1123 (from clone DKFZp586K1123)
AA205593	zq55f01.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone
AA206161	IMAGE:645529 3', mRNA sequence.
AA284075	kinesin 2 60/70kDa
AA284075	kinesin 2 60/70kDa
AA291203	Notch homolog 2 (Drosophila)

AA292373 collagen, type VI, alpha 1
 AA343057 EST48728 Fetal spleen Homo sapiens cDNA 3' end, mRNA sequence.
 AA362254 Transcribed sequences
 AA406435 hypothetical protein FLJ33761
 Transcribed sequence with weak similarity to protein sp:P39195 (H.sapiens)
 AA417878 ALU8_HUMAN Alu subfamily SX sequence contamination warning entry
 AA428286 hypothetical protein FLJ90492
 AA432267 adenylate kinase 3 like 1
 zx11c09.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
 IMAGE:786160 3' similar to contains Alu repetitive element;contains element
 AA448858 PTR5 repetitive element ;, mRNA sequence.
 Similar to RIKEN cDNA 5033406L14 gene, clone IMAGE:5591379, mRNA,
 partial cds
 AA454190 activated RNA polymerase II transcription cofactor 4
 AA456973 mannosidase, alpha, class 2C, member 1
 AA460970 H3 histone, family 3A
 AA477655 aa34e11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:815180 3',
 mRNA sequence.
 AA481141 insulin receptor
 AA485908 CDNA FLJ34866 fis, clone NT2NE2014113
 AA489100 CDNA: FLJ22994 fis, clone KAT11918
 AA490685 insulin receptor tyrosine kinase substrate
 AA496034 ne43d10.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:900115 3'
 similar to contains Alu repetitive element;, mRNA sequence.
 AA502936 zinc finger protein 262
 AA521508 Splicing factor, arginine/serine-rich, 46kD
 AA524274 Transcribed sequence with moderate similarity to protein sp:P39192
 (H.sapiens) ALU5_HUMAN Alu subfamily SC sequence contamination warning
 entry
 AA524299 sorting nexin 4
 AA524345 jumonji domain containing 1
 AA524505 phosphatidylinositol-3-phosphate associated protein
 AA524700 KIAA0924 protein
 AA526904 hypothetical protein MGC12538
 AA526907 nucleoporin 98kDa
 AA527238 transportin 1
 AA527296 MRNA; cDNA DKFZp761D1624 (from clone DKFZp761D1624)
 AA528080 ubiquitin specific protease 30
 AA528138 FLJ39739 protein
 AA532655 phosphoinositol 3-phosphate-binding protein-3
 AA535361 H.sapiens (xs157) mRNA, 315bp
 AA554945 major histocompatibility complex, class I, A
 AA573862 hypothetical protein BC004337
 AA574240 eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa
 AA577698 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2
 AA587884 transforming growth factor, beta receptor I (activin A receptor type II-like
 kinase, 53kDa)
 AA604375 glycerol-3-phosphate dehydrogenase 2 (mitochondrial)
 AA613031 myosin ID
 AA621962 Transcribed sequences
 AA625847 Clone IMAGE:5289004, mRNA
 AA628423

AA628539 eukaryotic translation initiation factor 3, subunit 9 eta, 116kDa
 AA631103 CDNA clone IMAGE:6186815, partial cds
 AA632295 stonin 2
 AA634272 signal transducer and activator of transcription 3 (acute-phase response factor)
 AA639752 DnaJ (Hsp40) homolog, subfamily C, member 3
 AA643304 N-terminal asparagine amidase
 AA649070 hypothetical protein DKFZp667E0512
 AA678047 methylmalonic aciduria (cobalamin deficiency) type A
 AA700167 son of sevenless homolog 1 (Drosophila)
 AA700485 adaptor-related protein complex 3, mu 1 subunit
 Transcribed sequence with weak similarity to protein pir:A32422 (H.sapiens)
 AA701676 A32422 dihydrolipoamide S-
 AA702248 CDNA clone IMAGE:4047449, partial cds
 zj11c05.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone
 IMAGE:449960 3' similar to contains MER17.t1 MER17 repetitive element ;,
 mRNA sequence.
 AA703326
 AA705628 solute carrier family 16 (monocarboxylic acid transporters), member 5
 AA719797 CDNA clone IMAGE:3865861, partial cds
 AA722799 endothelial and smooth muscle cell-derived neuropilin-like protein
 AA723057 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa
 ai08h05.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
 1342233 3' similar to gb:D38081 THROMBOXANE A2 RECEPTOR (HUMAN);,
 mRNA sequence.
 AA725102
 ai17c10.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
 1343058 3', mRNA sequence.
 AA725362
 Transcribed sequence with strong similarity to protein ref:NP_073584.1
 AA728758 (H.sapiens) hypothetical protein FLJ22558 [Homo sapiens]
 nz03g08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1286750 3',
 mRNA sequence.
 AA740875
 AA741074 Wolf-Hirschhorn syndrome candidate 1-like 1
 AA742659 zinc finger protein 567
 AA747303 exosome component Rrp40
 AA747309 zinc finger protein 90 homolog (mouse)
 AA760738 CDNA FLJ42331 fis, clone TSTOM2000588
 AA761259 Rap guanine nucleotide exchange factor (GEF) 1
 AA772172 hypothetical protein LOC196264
 AA775177 protein tyrosine phosphatase, receptor type, E
 AA775408 hypothetical protein LOC283241
 AA780381 mitogen-activated protein kinase kinase 3
 nz41a04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1290318 3',
 mRNA sequence.
 AA805651
 AA806142 CAP, adenylate cyclase-associated protein 1 (yeast)
 AA808018 cyclin I
 AA810864 Chromosome 2, 10 repeat region, complete sequence
 AA811466 WD repeat domain 20
 ob72f05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336929 3',
 mRNA sequence.
 AA811923
 AA813018 tight junction protein 1 (zona occludens 1)
 AA813332 CASK-interacting protein CIP98

od55c12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371862 3', mRNA sequence.
 AA825563 Transcribed sequences
 AA825721 UL16 binding protein 2
 AA831769 solute carrier family 30 (zinc transporter), member 5
 AA836340 Transcribed sequence with moderate similarity to protein sp:P39195 (H.sapiens) ALU8_HUMAN Alu subfamily SX sequence contamination warning entry
 AA865357 Clone IMAGE:5299888, mRNA
 AA868896 ubiquitin-conjugating enzyme E2B (RAD6 homolog)
 AA877765 hypothetical protein FLJ32312
 AA878516 nuclear receptor binding factor 2
 AA883074 NADPH oxidase, EF hand calcium-binding domain 5
 AA885360 Transcribed sequence with strong similarity to protein sp:P42696 (H.sapiens) Y117_HUMAN Hypothetical RNA-binding protein KIAA0117
 AA887480 ok92b01.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1521385 3', mRNA sequence.
 AA902326 Transcribed sequence
 AA902480 ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)
 AA910614 solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
 AA916851 RAB14, member RAS oncogene family
 AA919115 trafficking protein particle complex 6B
 AA936745 thyroid hormone receptor interactor 12
 AA973551 hypothetical protein LOC51321
 AA977218 dynein, cytoplasmic, heavy polypeptide 1
 AB002323 chromosome 6 open reading frame 133
 AB002347 lysosomal apyrase-like 1
 AB002390 D15F37 (pseudogene)
 AB002391 suppressor of cytokine signaling 1
 AB005043 KIAA0286 protein
 AB006624 zinc finger, ZZ-type with EF hand domain 1
 AB007859 KIAA0415 gene product
 AB007875 putative L-type neutral amino acid transporter
 AB007896 immunoglobulin superfamily, member 3
 AB007935 KIAA0483 protein
 AB007952 sperm associated antigen 9
 AB011088 KIAA0528 gene product
 AB011100 PRP4 pre-mRNA processing factor 4 homolog B (yeast)
 AB011108 hypothetical protein MGC23401
 AB011118 formin binding protein 1
 AB011126 BDG-29 proten
 AB011151 huntingtin interacting protein-1-related
 AB013384 ATPase, Class II, type 9A
 AB014511 synovial sarcoma translocation gene on chromosome 18-like 1
 AB014593 atrophin-1 interacting protein 1
 AB014605 ubiquitin-conjugating enzyme E2E 3 (UBC4/5 homolog, yeast)
 AB017644 ubiquitin specific protease 34
 AB018272 Rho-related BTB domain containing 1
 AB018283 KIAA0802 protein
 AB018345 pecanex homolog (Drosophila)

AB019691 A kinase (PRKA) anchor protein (yotiao) 9
AB020631 pre-mRNA cleavage complex II protein Pcf11
AB020657 influenza virus NS1A binding protein
AB020663 rabconnectin-3
AB020682 F-box only protein 21
AB020712 SEC31-like 1 (*S. cerevisiae*)
AB020724 sec1 family domain containing 1
AB020980 putative membrane protein
AB023231 formin binding protein 4
AB028628 ASF1 anti-silencing function 1 homolog A (*S. cerevisiae*)
AB028839 ubiquitination factor E4B (UFD2 homolog, yeast)
AB028973 Homo sapiens mRNA for KIAA1050 protein, partial cds.
AB029001 KIAA1078 protein
AB029006 spastic paraplegia 4 (autosomal dominant; spastin)
AB029020 ubiquitin specific protease 33
AB029032 hypothetical protein KIAA1109
AB029156 hepatoma-derived growth factor, related protein 3
AB029551 RING1 and YY1 binding protein
AB030710 GABA(A) receptor-associated protein-like 2
AB032179 erythrocyte membrane protein band 4.1 like 4B
AB032951 protein kinase C binding protein 1
AB032983 ras homolog gene family, member C like 1
AB033054 KIAA1228 protein
AB033066 KIAA1240 protein
AB033076 likely homolog of rat kinase D-interacting substance of 220 kDa
AB033078 sphingosine-1-phosphate lyase 1
AB034951 heat shock 70kDa protein 8
AB035482 chromosome 1 open reading frame 38
AB035745 Down syndrome critical region gene 5
AB036063 ribonucleotide reductase M2 B (TP53 inducible)
AB037703 survival of motor neuron protein interacting protein 1
AB037732 KIAA1311 protein
AB037743 TBC1 domain family, member 14
AB037791 hypothetical protein FLJ10980
AB037797 arrestin domain containing 3
AB037810 signal-induced proliferation-associated 1 like 2
AB037845 Rho GTPase activating protein 21
AB037901 jumonji domain containing 2C
AB040896 KIAA1463 protein
AB040903 RCC1-like
AB040914 Shroom-related protein
AB040927 SH3 multiple domains 2
AB044807 InaD-like protein
AB045223 cisplatin resistance related protein CRR9p
AB046809 zinc finger, FYVE domain containing 1
AB046817 synaptotagmin-like 2
AB046821 sarcolemma associated protein
AB046857 nuclear receptor coactivator 5
AB050049 methylcrotonoyl-Coenzyme A carboxylase 2 (beta)
AB051450 transducer of ERBB2, 2

AB051487 dual specificity phosphatase 16
 AB051495 kinesin family member 21A
 AB051499 KIAA1712
 AB051515 KIAA1728 protein
 AB066484 heterogeneous nuclear ribonucleoprotein D-like
 AC004472
 AC005034 Homo sapiens BAC clone RP11-342K6 from 2, complete sequence.
 AC005600
 AD001527
 AF000974 thyroid hormone receptor interactor 6
 AF001212 proteasome (prosome, macropain) 26S subunit, non-ATPase, 11
 AF003934 growth differentiation factor 15
 AF006010 progesterin induced protein
 AF007217 thyroid hormone receptor interactor 11
 AF010227 nuclear receptor coactivator 3
 AF013168 tuberous sclerosis 1
 AF014403 phosphatidic acid phosphatase type 2A
 AF015043 SH3-domain binding protein 4
 AF015186 hypothetical protein ET
 AF015593 ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)
 AF020043 chondroitin sulfate proteoglycan 6 (bamacan)
 AF021834 tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)
 phosphatase and tensin homolog (mutated in multiple advanced cancers 1),
 pseudogene 1
 AF023139 serine protease inhibitor, Kunitz type, 2
 AF033026 3'-phosphoadenosine 5'-phosphosulfate synthase 1
 AF035620 BRCA1 associated protein
 AF042386 peptidylprolyl isomerase E (cyclophilin E)
 AF043453 Homo sapiens sorting nexin 2 (SNX2) mRNA, complete cds.
 AF044221 ubiquitin-like 3
 AF047022 activating transcription factor 1
 AF047598 origin recognition complex, subunit 4-like (yeast)
 AF049103 huntingtin interacting protein B
 AF052094 endothelial PAS domain protein 1
 AF052179 ADP-ribosylation factor 1
 AF056085 G protein-coupled receptor 51
 AF060922 BCL2/adenovirus E1B 19kDa interacting protein 3-like
 AF061730 CGI-150 protein
 AF061731 chromosome 14 open reading frame 119
 AF061832 heterogeneous nuclear ribonucleoprotein M
 AF062347 zinc finger protein 216
 AF063592 Brain my034 protein mRNA, complete cds
 AF064801 ring finger 139
 AF065391 zinc finger protein 265
 AF067173 mago-nashi homolog, proliferation-associated (Drosophila)
 AF069506 RAS, dexamethasone-induced 1
 AF070448 cathepsin L2
 AF070536 Clone 24566 mRNA sequence
 AF070560 O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)

AF070569 hypothetical protein MGC14376
 AF072242 methyl-CpG binding domain protein 2
 AF077973 ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeier-Vogt disease)
 AF080157 conserved helix-loop-helix ubiquitous kinase
 AF082283 B-cell CLL/lymphoma 10
 AF083105 SRY (sex determining region Y)-box 13
 AF083441 putative translation initiation factor
 AF084462 Ras-like without CAAX 1
 AF087481 Jumonji, AT rich interactive domain 1B (RBP2-like)
 AF087980 immunoglobulin superfamily, member 10
 AF090900 lin-7 homolog C (*C. elegans*)
 AF091085 serologically defined breast cancer antigen 84
 AF092128 integral membrane protein 2B
 AF098865 squalene epoxidase
 AF100752 valosin-containing protein
 AF100763 protein kinase, AMP-activated, alpha 1 catalytic subunit
 AF101051 claudin 1
 AF106037 type 1 tumor necrosis factor receptor shedding aminopeptidase regulator
 Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal
 domain, 2
 AF109161 CREBBP/EP300 inhibitory protein 1
 AF109873 ATPase, H⁺ transporting, lysosomal 50/57kDa, V1 subunit H
 AF112204 UMP-CMP kinase
 AF112216 GK001 protein
 AF114784 methyl-CpG binding domain protein 4
 AF115403 E74-like factor 5 (ets domain transcription factor)
 AF115512 DnaJ (Hsp40) homolog, subfamily B, member 9
 AF116273 BCL2-associated athanogene
 AF116702 predicted protein of HQ2446; Homo sapiens PRO2446 mRNA, complete cds.
 AF116827 component of oligomeric golgi complex 6
 AF118652 thioredoxin-like 2
 AF119814 zinc finger protein 291
 AF121855 sorting nexin 5
 AF126782 dehydrogenase/reductase (SDR family) member 7
 AF129536 F-box only protein 6
 AF130099 hypothetical protein FLJ14753
 AF131743 XTP3-transactivated protein B
 AF131748 succinate-CoA ligase, GDP-forming, beta subunit
 AF131780 hypothetical protein LOC51061
 AF131793 hypothetical protein FLJ37306
 AF131831 non-kinase Cdc42 effector protein SPEC2
 AF131854 hypothetical protein MGC3067
 intelligence reducing insertion protein INGRIN; Homo sapiens OPA-containing
 protein (HOPA) gene, complete cds.
 AF132033 Kruppel-like factor 5 (intestinal)
 AF132818 heat shock 27kDa protein 8
 AF133207 cofilin 2 (muscle)
 AF136972 protein phosphatase 1B (formerly 2C), magnesium-dependent, beta isoform
 AF139131 beclin 1 (coiled-coil, myosin-like BCL2 interacting protein)
 AF142408 MLL septin-like fusion
 AF142482 TEF-5; Homo sapiens transcription enhancer factor-5 mRNA, complete cds.

AF145020 phospholipase A2-activating protein
 AF148949 Clone IMAGE:5242625, mRNA
 AF151047 chromosome 14 open reading frame 100
 AF151079 Homo sapiens HSPC245 mRNA, complete cds.
 AF151842 DKFZP564O123 protein
 AF151853 preimplantation protein 3
 AF153330 solute carrier family 19 (thiamine transporter), member 2
 AF154848 UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 3
 AF158555 glutaminase
 AF164794 tumor differentially expressed 2
 AF176699 F-box and leucine-rich repeat protein 4
 AF176704 F-box only protein 9
 AF177171 tropomodulin 3 (ubiquitous)
 AF180519 GABA(A) receptors associated protein like 3
 AF182645 IK cytokine, down-regulator of HLA II
 AF183417 microtubule-associated protein 1 light chain 3 beta
 AF184213 pre-mRNA branch site protein p14
 AF191495 F11 receptor
 AF201932 F-box only protein 8
 AF202092 autophagy Apg3p/Aut1p-like
 AF204231 golgin-67
 AF208694 hypothetical protein IMPACT
 AF210057 chromosome 3 open reading frame 1
 AF212221 myosin regulatory light chain interacting protein
 AF212230 hypothetical protein GL009
 AF216292 heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)
 AF216644 transmembrane protein vezatin
 AF217190 DEAH (Asp-Glu-Ala-His) box polypeptide 36
 AF217514 chromosome 20 open reading frame 111
 AF217963 melanoma antigen, family D, 1
 AF217990 homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1
 AF220137 tripartite motif-containing 33
 AF225417 transmembrane protein vezatin
 AF225423 mesenchymal stem cell protein DSC92
 AF226044 Homo sapiens HSNFRK (HSNFRK) mRNA, complete cds.
 AF226732 NPD007 protein
 AF227192 tudor and KH domain containing
 AF228422 normal mucosa of esophagus specific 1
 AF229179 kidney-specific membrane protein
 AF230929 annexin A9
 AF232674 downstream neighbor of SON
 AF233225 F-box only protein 7
 AF246144 cyclin D-type binding-protein 1
 AF251050 tigger transposable element derived 7
 AF251052 RALBP1 associated Eps domain containing 1
 AF251053 X-linked protein
 AF262027 member of eIF-5A family; Homo sapiens eIF-5A2 mRNA, complete cds.
 AF263293 SH3-domain GRB2-like endophilin B1
 AF267856 hypothetical protein dJ465N24.2.1

AF267863 RAB5B, member RAS oncogene family
 AF267864 likely ortholog of mouse IRA1 protein
 AF268193 likely ortholog of mouse IRA1 protein
 AF269167 hypothetical protein FLJ20364
 AF274950 hypothetical protein FLJ10637
 AF275260 chemokine (C-X-C motif) ligand 16
 AF275719 heat shock 90kDa protein 1, beta
 AF275945 epithelial V-like antigen 1
 AF278532 netrin 4
 AF279891 DEAH (Asp-Glu-Ala-His) box polypeptide 15
 AF279893 actin related protein 2/3 complex, subunit 2, 34kDa
 AF279903 ribosomal protein L15
 AF285758 lysyl-tRNA synthetase
 AF288208 UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1
 AF290195 hypertension-related calcium-regulated gene CaRG; Homo sapiens
 AF291676 hypertension-related calcium-regulated gene mRNA, complete cds.
 AF293841 MIP-2A; Homo sapiens MBP-1 interacting protein-2A mRNA, complete cds.
 AF305057 APG5 autophagy 5-like (*S. cerevisiae*)
 AF306508 SUMO1/sentrin specific protease 6
 AF307338 B aggressive lymphoma gene
 AF309553 recombination protein REC14
 AF311312 sperm associated antigen 1
 AF318575 UPF3 regulator of nonsense transcripts homolog A (yeast)
 AF320070 Homo sapiens hepatocellular carcinoma-associated protein HCA10 mRNA,
 complete cds.
 AF320999 reticulon 4
 AF323665 ovarian carcinoma immunoreactive antigen
 AF327443 calpastatin
 AF328864 selenoprotein S
 AF332197 sine oculis homeobox homolog 2 (*Drosophila*)
 AF333336 reticulon 4
 AF338193 similar to other members of this gene family: FKSG48 to FKSG70; Homo
 sapiens FKSG65 (FKSG65) mRNA, complete cds.
 AF339824 hypothetical protein LOC283476
 AF352728 stimulated by retinoic acid gene 6
 AF353991 BBP-like protein 1
 AF353992 BBP-like protein 2
 AF362887 tropomyosin 4
 AF364036 RNA binding motif protein 15
 AF479418 KIAA0261
 AF493921 Ras homolog enriched in brain
 AF523265 Hypothetical protein LOC286286 (LOC286286), mRNA
 AFFX-HSAC07/X00351_5
 AI003508 Transcribed sequences
 AI005245 hypothetical protein FLJ35036
 AI016620 signal sequence receptor, alpha (translocon-associated protein alpha)
 AI026938 thyroid hormone receptor associated protein 6
 AI028528 ow44c09.x1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
 IMAGE:1649680 3', mRNA sequence.

AI038402 Transcribed sequence with strong similarity to protein pir:A57377 (H.sapiens)
 A57377 transcription factor NFATx - human
 AI039874 NAD(P)H dehydrogenase, quinone 1
 AI051248 hypothetical protein FLJ32115
 AI052736 hypothetical protein DKFZp564A022
 AI056815 paired basic amino acid cleaving system 4
 AI073822 casein kinase 1, gamma 3
 AI087937 hypothetical protein FLJ11036
 AI089325 epithelial membrane protein 2
 AI090268 Transcribed sequence with moderate similarity to protein pir:B34087
 (H.sapiens) B34087 hypothetical protein
 AI091533 acidic repeat containing
 AI092770 Mesenchymal stem cell protein DSC96 mRNA, partial cds
 AI093221 occludin
 AI093579 integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)
 AI096375 KIAA1750 protein
 AI096477 zinc finger protein 363
 AI122754 steroid sulfatase (microsomal), arylsulfatase C, isozyme S
 AI122756 similar to hypothetical protein FLJ13659
 AI122787 qa48h04.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1690039
 3' similar to contains Alu repetitive element;, mRNA sequence.
 AI123320 nanos homolog 1 (Drosophila)
 AI123518 zinc fingers and homeoboxes 1
 AI123815 FLJ21963 protein
 AI125646 zinc finger protein 207
 AI125859 qe01g02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1737746
 3', mRNA sequence.
 AI129346 NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa
 AI130969 collagen, type V, alpha 1
 AI131008 thyroid hormone receptor interactor 3
 AI139569 SWAP-70 protein
 AI139990 Transcribed sequences
 AI140607 LOC389629 (LOC389629), mRNA
 AI141556 BAI1-associated protein 1
 AI141784 FAD104
 AI143307 zinc finger protein 598
 AI147621 chromosome 10 open reading frame 32
 AI149508 origin recognition complex, subunit 4-like (yeast)
 AI150000 CDNA clone IMAGE:4797120, partial cds
 AI151434 hypothetical protein DKFZp762N1910
 AI167164 myotubularin related protein 1
 AI168350 exocyst complex component 8
 AI174548 methyl-CpG binding domain protein 6
 AI184562 U2-associated SR140 protein
 AI184802 PRP4 pre-mRNA processing factor 4 homolog (yeast)
 AI193090 hypothetical protein FLJ22875
 AI201534 sec13-like protein
 AI202327 cytoplasmic polyadenylation element binding protein 2
 AI206888 potassium channel, subfamily V, member 2
 AI214061 tropomyosin 4

AI215106 insulin receptor
 AI217906 chromosome 14 open reading frame 103
 AI218219 heat shock 90kDa protein 1, beta
 AI219073 EPS8-like 1
 AI221300 potassium channel modulatory factor 1
 qk25b10.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1869979 3', mRNA sequence.
 AI245517
 AI251890 CDC-like kinase 1
 AI261542 chromosome 14 open reading frame 32
 AI268315 glutamine-fructose-6-phosphate transaminase 1
 AI275597 glucocorticoid receptor DNA binding factor 1
 qm48h03.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1892021 3', mRNA sequence.
 AI277336
 AI277652 Full length insert cDNA clone ZD42C02
 AI278616 SET translocation (myeloid leukemia-associated)
 AI279868 activating transcription factor 7
 AI280328 Transcribed sequences
 AI288759 AT-binding transcription factor 1
 AI289311 jub, ajuba homolog (Xenopus laevis)
 AI289774 Transcribed sequences
 AI300168 hypothetical protein FLJ31413
 AI302244 KIAA1915 protein
 AI307763 arginase, type II
 AI307808 ring finger protein 141
 AI309784 CTF8, chromosome transmission fidelity factor 8 homolog (S. cerevisiae)
 AI310001 hypothetical protein FLJ22789
 Transcribed sequence with weak similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 AI312083 ta77f02.x2 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2050107 3' similar to gb:L19779 HISTONE H2A.1 (HUMAN);, mRNA sequence.
 AI313324
 AI333232 RAB18, member RAS oncogene family
 AI335267 mitogen-activated protein kinase kinase 1 interacting protein 1
 AI336848 MSTP105 (MST105) mRNA, complete cds
 AI337901 ocular development-associated gene
 AI340270 ariadne homolog 2 (Drosophila)
 AI341146 E2F transcription factor 7
 AI341537 NADPH oxidase-related, C2 domain-containing protein
 pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1
 AI346026
 AI346504 chromosome 14 open reading frame 109
 AI347136 telomeric repeat binding factor (NIMA-interacting) 1
 AI348001 similar to hypothetical protein MGC17347
 qp61g12.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1927558 3', mRNA sequence.
 AI348094
 AI352424 PWWP domain containing 1
 AI356283 hypothetical protein FLJ34443
 AI356412 v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
 Transcribed sequence with moderate similarity to protein sp:P39188 (H.sapiens) ALU1_HUMAN Alu subfamily J sequence contamination warning entry
 AI360054
 AI372850 ubiquitin-conjugating enzyme E2-like

AI376549 Transcribed sequences
 AI377497 mob protein
 AI379517 ral guanine nucleotide dissociation stimulator-like 3
 AI380514 Transcribed sequences
 AI382195 LOC388727 (LOC388727), mRNA
 AI393309 Similar to RIKEN cDNA 1110033O09 gene
 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 AI393725 ubiquitin specific protease 8
 AI393759 CDNA: FLJ21778 fis, clone HEP00201
 AI394438 phosphoinositide-3-kinase, class 2, alpha polypeptide
 AI401379 hypothetical protein FLJ11193
 AI422099 chromodomain helicase DNA binding protein 1-like
 AI422414 Transcribed sequences
 AI423056 hypothetical protein DKFZp547A023
 AI431788 glycogen synthase kinase 3 beta
 AI432196 nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
 AI434509 hypothetical protein FLJ38499
 AI435036 ubiquitin specific protease 16
 Transcribed sequence with weak similarity to protein ref:NP_071431.1
 (H.sapiens) cytokine receptor-like factor 2; cytokine receptor CRL2 precursor
 [Homo sapiens]
 AI436356 nuclear receptor coactivator 3
 AI438999 retinal short chain dehydrogenase reductase
 AI440266 PRO1073 protein
 AI446756 hypothetical protein LOC162073
 tm08f07.x1 NCI_CGAP_Co14 Homo sapiens cDNA clone IMAGE:2156005 3'
 AI469425 similar to contains MSR1.t2 MSR1 repetitive element ;, mRNA sequence.
 AI472139 CUG triplet repeat, RNA binding protein 1
 tm03d11.x1 NCI_CGAP_Co14 Homo sapiens cDNA clone IMAGE:2155509 3'
 AI473796 similar to gb:S41458 ROD CGMP-SPECIFIC 3',5'-CYCLIC
 PHOSPHODIESTERASE BETA-SUBUNIT (HUMAN);, mRNA sequence.
 tm39e01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2160504 3'
 AI478300 similar to contains Alu repetitive element;, mRNA sequence.
 AI478879 KIAA1128 protein
 AI492902 topoisomerase-related function protein 4-2
 AI493245 CD44 antigen (homing function and Indian blood group system)
 qz51f12.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2030447 3',
 AI493853 mRNA sequence.
 AI498144 chromosome 20 open reading frame 194
 tm85a05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2164880 3',
 AI498207 mRNA sequence.
 tm85a05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2164880 3',
 AI498207 mRNA sequence.
 AI521883 Transcribed sequences
 AI535735 LATS, large tumor suppressor, homolog 2 (Drosophila)
 tp32g06.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2189530 3'
 AI537887 similar to gb:M81635 ERYTHROCYTE BAND 7 INTEGRAL MEMBRANE
 PROTEIN (HUMAN);, mRNA sequence.
 AI538172 retinoblastoma binding protein 6
 AI559300 LOC132671

AI559696 likely ortholog of mouse acyl-Coenzyme A thioesterase 2, mitochondrial
 AI559701 hypothetical protein BC008217
 AI561070 MRNA; cDNA DKFZp313O038 (from clone DKFZp313O038)
 AI564840 Transcribed sequences
 AI569503 CTL2 gene
 AI569932 Similar to RIKEN cDNA 5830415L20 (LOC401015), mRNA
 AI570450 Clone IMAGE:5271371, mRNA
 AI571419 mitogen-activated protein kinase kinase 1
 AI571996 signal transducing adaptor molecule (SH3 domain and ITAM motif) 2
 AI580100 MRNA; cDNA DKFZp564H0764 (from clone DKFZp564H0764)
 AI583393 ts09b07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2228053 3'
 AI587307 similar to gb:M77016 TROPOMODULIN (HUMAN);, mRNA sequence.
 AI590053 mannosidase, endo-alpha
 AI590088 tr75d05.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224137 3'
 AI609256 similar to contains Alu repetitive element;, mRNA sequence.
 AI610112 TEA domain family member 1 (SV40 transcriptional enhancer factor)
 AI610347 solute carrier family 30 (zinc transporter), member 9
 AI611074 tigger transposable element derived 2
 AI613010 Transcribed sequences
 AI621223 Transcribed sequences
 AI621225 tx57h11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2273733 3',
 AI625550 mRNA sequence.
 AI627666 huntingtin interacting protein 14
 AI628605 KIAA1737
 AI631846 filamin A, alpha (actin binding protein 280)
 AI632214 hypothetical protein BC014311
 AI632728 chromosome 14 open reading frame 138
 AI633566 hypothetical protein BC009980
 AI634866 Clone 24987 mRNA sequence
 AI638714 CDNA FLJ31683 fis, clone NT2RI2005353
 AI650251 microtubule-associated protein, RP/EB family, member 1
 AI650819 CDW92 antigen
 AI652586 GLE1 RNA export mediator-like (yeast)
 AI652899 hypothetical protein MGC33214
 AI653608 cullin 4B
 AI653730 SWI/SNF related, matrix associated, actin dependent regulator of chromatin,
 AI653730 subfamily a, member 5
 AI656706 Transcribed sequence with weak similarity to protein ref:NP_060265.1
 AI658548 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 AI659020 leucine-rich PPR-motif containing
 AI659219 calmodulin 1 (phosphorylase kinase, delta)
 AI659800 calmodulin 1 (phosphorylase kinase, delta)
 AI668786 zinc finger and BTB domain containing 1
 AI669535 zinc finger, DHHC domain containing 21
 CDNA FLJ36689 fis, clone UTERU2008653, highly similar to GLYCODELIN
 PRECURSOR
 hypothetical protein FLJ38725
 mesoderm induction early response 1
 interleukin 17D

AI669749 ribonuclease L (2',5'-oligoadenylate synthetase-dependent)
 AI669947 hypothetical protein LOC286148
 AI669957 eukaryotic translation initiation factor 2C, 4
 AI670903 YY1 transcription factor
 AI672356 heterogeneous nuclear ribonucleoprotein A0
 AI674647 putative intramembrane cleaving protease
 AI674731 hypothetical protein MGC35274
 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila);
 translocated to, 4
 AI675354
 AI675780 transcription elongation factor A (SII), 3
 AI675844 CDNA FLJ31553 fis, clone NT2RI2001178
 AI676103 protein x 0001
 Transcribed sequence with weak similarity to protein sp:P39188 (H.sapiens)
 AI678013 ALU1_HUMAN Alu subfamily J sequence contamination warning entry
 AI678717 sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)
 AI681120 RAN binding protein 2
 AI684626 trinucleotide repeat containing 6
 AI686303 KIAA1935 protein
 wd40d12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 AI688663 IMAGE:2330615 3', mRNA sequence.
 AI690465 hypothetical protein LOC283278
 AI692203 ring finger protein 3
 AI692595 zinc finger, SWIM domain containing 6
 AI693689 Transcribed sequences
 AI694023 thyroid hormone receptor interactor 8
 Transcribed sequence with weak similarity to protein sp:P39194 (H.sapiens)
 AI694544 ALU7_HUMAN Alu subfamily SQ sequence contamination warning entry
 syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component
 2)
 AI695684
 AI697976 hypothetical protein LOC220213
 AI700633 CDNA: FLJ22642 fis, clone HSI06970
 AI700962 protein phosphatase 1, catalytic subunit, beta isoform
 AI701055 transforming growth factor beta regulator 1
 AI701949 heterogeneous nuclear ribonucleoprotein K
 AI703476 G protein-coupled receptor 27
 AI709335 solute carrier family 18 (vesicular monoamine), member 2
 AI718385 solute carrier family 26 (sulfate transporter), member 2
 AI718937 potassium channel tetramerisation domain containing 12
 AI720923 DEAH (Asp-Glu-Ala-His) box polypeptide 33
 AI732381 keratin 20
 Transcribed sequence with weak similarity to protein ref:NP_036553.1
 AI732587 (H.sapiens) rearranged L-myc fusion sequence; Zn-15 related [Homo sapiens]
 Transcribed sequence with moderate similarity to protein sp:P12947
 AI733287 (H.sapiens) RL31_HUMAN 60S ribosomal protein L31
 AI733465 collagen, type IX, alpha 2
 AI734929 poly(A) binding protein, cytoplasmic 1
 AI735261 hypothetical protein MGC54289
 AI735576 ROD1 regulator of differentiation 1 (S. pombe)
 AI738919 ligand of numb-protein X
 AI741411 DDHD domain containing 1
 AI741415 hypothetical protein MGC5509

AI741439 solute carrier family 8 (sodium/calcium exchanger), member 1
 AI742057 hypothetical protein LOC129607
 wg39c02.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 AI742210 IMAGE:2367458 3', mRNA sequence.
 AI742553 protein kinase, lysine deficient 1
 AI742668 CGI-72 protein
 AI742838 dedicator of cytokinesis 11
 AI742868 regulatory factor X-associated protein
 TAF4 RNA polymerase II, TATA box binding protein (TBP)-associated factor,
 135kDa
 AI744029
 AI744148 KIAA0431 protein
 AI744726 KIAA1327 protein
 AI749193 ubiquitin protein ligase E3B
 AI753792 related RAS viral (r-ras) oncogene homolog 2
 AI754871 Transcribed sequences
 AI758191 chromosome 9 open reading frame 55
 AI758763 transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)
 AI760629 chloride channel 3
 AI760812 SPT3-associated factor 42
 AI761436 Hypothetical protein LOC149692 (LOC149692), mRNA
 AI761621 nuclear receptor subfamily 1, group D, member 2
 AI761748 nuclear receptor coactivator 3
 AI761759 calnexin
 AI761989 glucocorticoid induced transcript 1
 AI762552 heterogeneous nuclear ribonucleoprotein D-like
 AI762874 translocase of inner mitochondrial membrane 22 homolog (yeast)
 AI762876 myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila)
 AI765607 Transcribed sequences
 AI767447 hypothetical protein FLJ13456
 AI767750 TATA element modulatory factor 1
 wh39d09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2383121
 AI767756 3', mRNA sequence.
 Transcribed sequence with weak similarity to protein ref:NP_112159.1
 (H.sapiens) hypothetical protein FLJ21617; erythroid differentiation-related
 factor 1 [Homo sapiens]
 AI769269
 AI769569 mastermind-like 2 (Drosophila)
 AI769587 similar to Rho GTPase activating protein 12
 AI769794 hypothetical protein FLJ14681
 AI791303 Transcribed sequences
 wh81c10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387154 3',
 mRNA sequence.
 AI798098
 AI798790 bullous pemphigoid antigen 1, 230/240kDa
 AI798924 Transcribed sequences
 AI799061 fem-1 homolog b (C. elegans)
 AI801666 Transcribed sequences
 AI802955 chronic myelogenous leukemia tumor antigen 66
 AI803010 Clone IMAGE:5301910, mRNA
 wf14b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 AI806583 IMAGE:2350559 3', mRNA sequence.
 AI806781 Transcribed sequences

AI807036 hypothetical protein FLJ14803
 AI807073 hypothetical protein FLJ10876
 AI808746 KIAA1826 protein
 AI809483 mitochondrial ribosomal protein L39
 AI809961 phosphonoformate immuno-associated protein 5
 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 AI810266 splicing factor, arginine/serine-rich 12
 AI810380 CDNA FLJ26764 fis, clone PRS02668
 AI810669 Similar to diaphanous homolog 3 (Drosophila), clone IMAGE:5277415, mRNA
 AI813331 eukaryotic translation initiation factor 2C, 2
 AI813489 Bardet-Biedl syndrome 4
 AI813772 zinc finger, DHHC domain containing 2
 AI814257 Transcribed sequences
 AI814925 HGS_RE408
 AI816071 protein phosphatase 1, regulatory (inhibitor) subunit 12A
 AI817061 Sp6 transcription factor
 AI817264 wk41h05.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2418009 3',
 mRNA sequence.
 AI818048 DKFZP564G092 protein
 AI819938 MRNA similar to joined to JAZF1 (cDNA clone MGC:52103 IMAGE:5736798),
 complete cds
 AI820796 secretory carrier membrane protein 4
 AI820875 Transcribed sequences
 AI821780 transmembrane, prostate androgen induced RNA
 AI821781 phosphodiesterase 4D interacting protein (myomegalin)
 AI821791 Transcribed sequence with moderate similarity to protein sp:P39195
 (H.sapiens) ALU8_HUMAN Alu subfamily SX sequence contamination warning
 entry
 AI821995 wh53b02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2384427
 3', mRNA sequence.
 AI823360 MYC induced nuclear antigen
 AI823896 phospholipid scramblase 1
 AI825926 syntaxin binding protein 5 (tomosyn)
 AI825998 isocitrate dehydrogenase 3 (NAD+) alpha
 AI826060 solute carrier family 25, member 29
 AI826268 EGF-containing fibulin-like extracellular matrix protein 1
 AI826799 catenin (cadherin-associated protein), alpha 1, 102kDa
 AI826881 myosin, heavy polypeptide 9, non-muscle
 AI827941 hypothetical protein LOC203522
 AI828015 SNF2 histone linker PHD RING helicase
 AI828221 ligase IV, DNA, ATP-dependent
 AI829314 occludin
 AI829721 hypothetical protein LOC285148
 AI829927 flightless I homolog (Drosophila)
 AI830227 hypothetical protein DKFZp564B1023
 AI831738 hypothetical protein DKFZp564B1023
 Transcribed sequence with weak similarity to protein ref:NP_112159.1
 (H.sapiens) hypothetical protein FLJ21617; erythroid differentiation-related
 factor 1 [Homo sapiens]
 AI833186 nudE nuclear distribution gene E homolog 1 (A. nidulans)
 AI857685

AI862255 ATPase, H⁺ transporting, lysosomal 9kDa, V0 subunit e
 AI862477 hypothetical protein FLJ11526
 wI98g02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2432978 3',
 mRNA sequence.
 AI869704
 AI870473 CDNA FLJ26369 fis, clone HRT06001
 AI871160 Transcribed sequences
 AI871408 Transcribed sequences
 AI871620 phosphoribosyl transferase domain containing 1
 AI871745 hypothetical protein FLJ33761
 AI873425 decapping enzyme hDcp2
 AI884858 tumor suppressor candidate 3
 AI885109 bromodomain containing 7
 AI885815 Similar to cleavage and polyadenylation specific factor 6, 68kD subunit, clone
 IMAGE:4819488, mRNA
 AI888099 Machado-Joseph disease (spinocerebellar ataxia 3, olivopontocerebellar ataxia
 3, autosomal dominant, ataxin 3)
 AI888503 CDNA: FLJ21652 fis, clone COL08582
 Transcribed sequence with moderate similarity to protein sp:P39192
 (H.sapiens) ALU5_HUMAN Alu subfamily SC sequence contamination warning
 entry
 AI890133
 AI890347 TUWD12
 AI890947 RNA-binding region (RNP1, RRM) containing 2
 we11g05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2340824 3',
 mRNA sequence.
 AI912618
 AI913365 methyl-CpG binding domain protein 4
 AI913749 similar to PH (pleckstrin homology) domain
 AI915947 hypothetical protein LOC339448
 AI916261 FLJ37099 protein
 AI916887 Transcribed sequences
 AI919519 potassium channel tetramerisation domain containing 1
 AI921002 hypothetical protein FLJ32702
 AI922509 retinoblastoma-associated protein 140
 AI922968 KIAA0303 protein
 AI923492 major histocompatibility complex, class I, A
 AI924685 hypothetical protein FLJ22557
 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 AI925316 NP220 nuclear protein
 AI927329
 AI927382 chromosome 6 open reading frame 113
 AI927770 sel-1 suppressor of lin-12-like (C. elegans)
 AI928344 RNA binding motif protein, X-linked
 wo67g04.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2460438 3',
 mRNA sequence.
 AI928428
 AI928526 heme-regulated initiation factor 2-alpha kinase
 AI932370 spastic ataxia of Charlevoix-Saguenay (sacsin)
 AI934469 KIAA0779 protein
 AI935115 hypothetical protein FLJ38101
 AI935123 chromosome 14 open reading frame 78
 AI935180 DnaJ (Hsp40) homolog, subfamily C, member 5
 AI935246 karyopherin alpha 4 (importin alpha 3)
 AI935657 BTB (POZ) domain containing 7

AI939336 transcription factor MLR1
 AI949179 BCL2-like 11 (apoptosis facilitator)
 AI950314 HIV-1 rev binding protein 2
 AI951998 DAZ associated protein 1
 AI952986 chromosome 14 open reading frame 170
 AI953847 IBR domain containing 2
 wx71b04.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2549071
 AI953914 3', mRNA sequence.
 AI954041 fatty acid synthase
 AI962933 heat shock 90kDa protein 1, alpha
 AI963008 heterogeneous nuclear ribonucleoprotein A2/B1
 AI967961 MRNA; cDNA DKFZp586K1123 (from clone DKFZp586K1123)
 AI970823 coiled-coil domain containing 8
 AI978623 KIAA0657 protein
 AI982758 CDNA FLJ33653 fis, clone BRAMY2024715
 wt46h06.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2510555 3',
 mRNA sequence.
 AI983021 pleckstrin homology domain containing, family B (evectins) member 2
 AI983043 pleckstrin homology domain containing, family B (evectins) member 2
 AI984051 thyroid hormone receptor associated protein 1
 AI984061 hypothetical protein LOC90637
 AI984479 poly(A) polymerase alpha
 AI986295 ankyrin repeat domain 17
 AI989530 DKFZP434D146 protein
 AI990523 ankyrin repeat and sterile alpha motif domain containing 1
 AI992283 TNF receptor-associated factor 4
 AJ001306 InaD-like protein
 AJ002572 chromosome 21 open reading frame 107
 AJ005866 solute carrier family 35, member D2
 MCM3 minichromosome maintenance deficient 3 (*S. cerevisiae*) associated
 protein
 AJ010089 aminopeptidase puromycin sensitive
 AJ132583 Homo sapiens CXCR4 gene encoding receptor CXCR4.
 AJ224869 Homo sapiens mRNA for putative protein TH1, partial, clone IMAGE ID
 785447.
 AJ238374 TH1-like (*Drosophila*)
 AJ238379 TH1-like (*Drosophila*)
 AJ250229 chromosome 11 open reading frame 1
 AJ251830 PERP, TP53 apoptosis effector
 AJ277276 transcriptional regulating factor 1
 AJ297586 major histocompatibility complex, class II, DR beta 3
 AK000004 FGD1 family, member 3
 AK000095 CDNA FLJ20088 fis, clone COL03869
 AK000116 debranching enzyme homolog 1 (*S. cerevisiae*)
 AK000293 Homo sapiens cDNA FLJ20286 fis, clone HEP04358.
 AK000478 GTPase activating RANGAP domain-like 4
 AK000617 hypothetical protein LOC92912
 unnamed protein product; Homo sapiens cDNA FLJ20819 fis, clone
 ADSE00511.
 AK000826 chemokine-like factor super family 4
 AK000855 chemokine-like factor super family 4
 AK001017 Nijmegen breakage syndrome 1 (nibrin)
 AK001105 LAG1 longevity assurance homolog 2 (*S. cerevisiae*)
 AK001135 SEC23 interacting protein

AK001289 bridging integrator 3
 AK001389 ankyrin repeat and MYND domain containing 2
 AK001513 KIAA1414 protein
 AK001574 golgi reassembly stacking protein 1, 65kDa
 AK001618 TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa
 AK001731 hypothetical protein MGC17943
 AK001750 hypothetical protein FLJ11046
 AK001899 APG5 autophagy 5-like (*S. cerevisiae*)
 AK002064 DNA polymerase-transactivated protein 6
 AK002075 myelin expression factor 2
 AK002111 karyopherin alpha 6 (importin alpha 7)
 AK002165 DKFZP564O123 protein
 AK002174 kelch-like 5 (*Drosophila*)
 AK021433 chromosome 6 open reading frame 109
 AK021464 enoyl Coenzyme A hydratase domain containing 1
 AK021761 RAB24, member RAS oncogene family
 AK021879 Homo sapiens cDNA FLJ11817 fis, clone HEMBA1006421.
 AK021890 dishevelled associated activator of morphogenesis 1
 AK022130 hypothetical protein FLJ10826
 AK022142 GRB2-associated binding protein 1
 AK022181 CDNA FLJ12119 fis, clone MAMMA1000092
 AK022248 actin-related protein 10 homolog (*S. cerevisiae*)
 AK022330 T-box 6
 AK022815 hypothetical protein FLJ13910
 AK022838 GRIP and coiled-coil domain containing 2
 AK022852 signal-induced proliferation-associated 1 like 2
 AK022888 WD repeat and FYVE domain containing 1
 AK022894 dendritic cell-derived ubiquitin-like protein
 AK023092 ubiquitin specific protease 49
 AK023140 unnamed protein product; Homo sapiens cDNA FLJ13078 fis, clone NT2RP3002002.
 AK023183 hypothetical protein FLJ11171
 AK023204 lamina-associated polypeptide 1B
 AK023354 unnamed protein product; Homo sapiens cDNA FLJ13292 fis, clone OVARC1001180, weakly similar to UBIQUITIN-LIKE PROTEIN DSK2.
 AK023596 KIAA1221 protein
 AK023754 Similar to Transcription factor HES-2 (Hairy and enhancer of split 2) (LOC388592), mRNA
 AK023841 DKFZP434C212 protein
 AK023950 chromosome 11 open reading frame 23
 AK023981 hypothetical protein LOC119504
 AK024029 modulator of apoptosis 1
 AK024274 hypothetical protein FLJ10074
 AK024379 Homo sapiens cDNA FLJ14317 fis, clone PLACE3000401.
 AK024412 kelch-like 12 (*Drosophila*)
 AK024828 chromosome 6 open reading frame 145
 AK024836 major histocompatibility complex, class I, C
 AK024846 SET domain-containing protein 7
 AK024896 mitochondrial ribosomal protein S6

AK024913 CDNA: FLJ21260 fis, clone COL01441
 AK024927 CDNA: FLJ21274 fis, clone COL01781
 AK024967 male sterility domain containing 2
 AK025007 amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 13
 AK025344 CDNA: FLJ21691 fis, clone COL09555
 AK025444 pleckstrin homology-like domain, family B, member 2
 AK025504 KIAA0251 protein
 AK025567 jub, ajuba homolog (*Xenopus laevis*)
 AK025603 HSPC159 protein
 AK025633 component of oligomeric golgi complex 1
 AK025731 YY1 transcription factor
 AK025794 PI-3-kinase-related kinase SMG-1
 AK025831 solute carrier family 39 (zinc transporter), member 9
 AK025925 WD-repeat protein
 AK025933 KIAA0692 protein
 AK026025 KIAA0515
 AK026133 sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4B
 AK026317 hypothetical protein BC001584
 AK026420 desmuslin
 AK026437 KIAA1387 protein
 AK026630 hypothetical protein FLJ13188
 AK026646 surfeit 4
 AK026678 stromal antigen 2
 AK026736 integrin, beta 6
 AK026921 solute carrier family 17 (anion/sugar transporter), member 5
 AK026926 Homo sapiens cDNA: FLJ23273 fis, clone HEP02611, highly similar to HSU79278 Human protein disulfide isomerase-related protein P5 mRNA.
 AK026933 eukaryotic translation initiation factor 5
 AK026954 retinoblastoma binding protein 6
 AK027164 FLJ10378 protein
 AK027217 LIM protein (similar to rat protein kinase C-binding enigma)
 AK054976 histidine triad nucleotide binding protein 1
 AK090649 Clone IMAGE:3606519, mRNA, partial cds
 AK091107 hypothetical protein LOC283687
 AK094254 similar to solute carrier family 25 , member 16
 AK094613 ribosomal protein S24
 AK095315 F-box only protein 9
 AK096484 Homo sapiens cDNA FLJ39165 fis, clone OCBBF2002663.
 AK096921 CDNA FLJ39602 fis, clone SKNSH2005061
 AK097649 MRNA; cDNA DKFZp686E1246 (from clone DKFZp686E1246)
 AL008582
 AL021707
 AL031282
 AL031427 Human DNA sequence from clone RP1-167A19 on chromosome 1p32.1-33, complete sequence.
 AL031588
 AL031685
 AL033377 Human DNA sequence from clone RP1-287G14 on chromosome 6q23.1-24.3, complete sequence.

AL034417 Human DNA sequence from clone CTA-215D11 on chromosome 1p36.12-36.33, complete sequence.

AL035086 Human DNA sequence from clone RP1-44A20 on chromosome 6q23.1-24.3, complete sequence.

AL035541

AL035604 Human DNA sequence from clone RP1-38C16 on chromosome 6q22.33-24.1, complete sequence.
continues in dJ1041C10 (AL162615) match: proteins: Tr:Q9WVK5 Tr:O43286 Tr:O88419 Tr:O60514; Human DNA sequence from clone RP5-1063B2 on chromosome 20q13.1-13.2 Contains the 3' part of the B4GALT5 gene for beta 1,4-galactosyltransferase, polypeptide 5, ESTs, STSs and GSSs, complete sequence.

AL035683 continued from bA177G23.1 in Em:AL451064 match: proteins: Tr:O75368 Sw:P55822 Tr:Q9BPY5 Tr:Q9BRB8 Sw:Q9WUZ7; Human DNA sequence from clone RP1-75K24 on chromosome 6q13-15 Contains the the 3' end of the SH3BGR2 gene for SH3 domain binding glutamic acid-rich protein-like 2, complete sequence.

AL035700

AL037339 PTK2 protein tyrosine kinase 2

AL038450 Transcribed sequences

AL039831 hypothetical protein FLJ10770

AL039862 breast cancer membrane protein 101

AL040394 3-phosphoinositide dependent protein kinase-1
ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)

AL040708

AL041280 hypothetical protein DKFZp313M0720

AL043571 RAN binding protein 2-like 1

AL043967 F-box and WD-40 domain protein 2

AL044092 hypothetical protein MGC18216

AL044651 ADP-ribosylation factor interacting protein 1 (arfaptin 1)

AL045882 hypothetical protein MGC16202

AL046054 prostate tumor overexpressed gene 1

AL049261 FGF receptor activating protein 1

AL049265 MRNA; cDNA DKFZp564F053 (from clone DKFZp564F053)

AL049305 FLJ21940 protein

AL049385 Homo sapiens mRNA; cDNA DKFZp586M1418 (from clone DKFZp586M1418).

AL049650

AL049942 zinc finger protein 337

AL049987 MRNA; cDNA DKFZp564F112 (from clone DKFZp564F112)

AL050050 GTPase activating RANGAP domain-like 2 pseudogene

AL050102 erythroid differentiation-related factor 1

AL050205 c-Mpl binding protein

AL050226 succinate-CoA ligase, GDP-forming, beta subunit

AL050348

AL050374 ankyrin repeat and BTB (POZ) domain containing 2

AL050376 AHA1, activator of heat shock 90kDa protein ATPase homolog 2 (yeast)

AL078633

AL079292 DEAH (Asp-Glu-Ala-His) box polypeptide 29

AL079310 high-mobility group protein 2-like 1

AL080081 DnaJ (Hsp40) homolog, subfamily B, member 9

AL080111 NIMA (never in mitosis gene a)-related kinase 7

AL080149 bromodomain containing 1

AL080183 hypothetical protein FLJ10824
 AL096741 ASC-1 complex subunit P100
 AL096776
 AL096842 mitochondrial tumor suppressor gene 1
 AL109688 MRNA full length insert cDNA clone EUROIMAGE 131775
 AL109714 hypothetical protein LOC283687
 AL109824
 AL109923
 AL109928
 AL110191 delta sleep inducing peptide, immunoreactor
 AL110236 HCF-binding transcription factor Zhangfei
 AL110238 RNA polymerase I transcription factor RRN3
 strong similarity to mouse Magi-1; Homo sapiens mRNA; cDNA
 AL110296 DKFZp434B203 (from clone DKFZp434B203); partial cds.
 Human DNA sequence from clone RP5-876B10 on chromosome 1q42.12-43,
 complete sequence.
 AL117352
 AL117354
 AL117499 chromosome 19 open reading frame 13
 AL117516 strand-exchange protein 1
 AL117553 MRNA; cDNA DKFZp564F1171 (from clone DKFZp564F1171)
 AL117600 DKFZP564J0863 protein
 AL117607 MRNA; cDNA DKFZp564N0763 (from clone DKFZp564N0763)
 AL118502
 AL118506
 AL118798 CD47 antigen (Rh-related antigen, integrin-associated signal transducer)
 AL120704 karyopherin alpha 3 (importin alpha 4)
 DKFZp762N103_s1 762 (synonym: hmel2) Homo sapiens cDNA clone
 DKFZp762N103 3', mRNA sequence.
 AL121021
 AL121585
 AL122091 hypothetical protein from EUROIMAGE 1977056
 AL132665 BCL2/adenovirus E1B 19kDa interacting protein 3-like
 AL132773
 Human DNA sequence from clone RP3-393D12 on chromosome 6q16.1-16.3,
 complete sequence.
 AL132776
 AL133577 MRNA; cDNA DKFZp434G0972 (from clone DKFZp434G0972)
 AL133580 short coiled-coil protein
 AL133653 testis expressed sequence 15
 DKFZp762M0710_s1 762 (synonym: hmel2) Homo sapiens cDNA clone
 AL134904 DKFZp762M0710 3', mRNA sequence.
 MRNA similar to RIKEN cDNA 2700049P18 gene (cDNA clone MGC:57827
 IMAGE:6064384), complete cds
 AL135396 phosphatidylinositol binding clathrin assembly protein
 AL135735
 AL135787
 AL136179
 AL136560 hypothetical protein DKFZp761C121
 AL136583 solute carrier family 37 (glycerol-3-phosphate transporter), member 3
 AL136592 nudix (nucleoside diphosphate linked moiety X)-type motif 12
 AL136597 SBBI26 protein
 AL136598 protein associated with PRK1
 AL136599 SUMO1/sentrin specific protease 7

AL136601 staufen, RNA binding protein (Drosophila)
 AL136629 TSPY-like
 AL136667 HSPC039 protein
 AL136705 phosphoglucomutase 2
 AL136719 spindlin
 AL136736 KIAA1549 protein
 AL136784 ankyrin repeat domain 27 (VPS9 domain)
 AL136800 KIAA1287 protein
 AL136827 KIAA0982 protein
 AL136829 ring finger protein 146
 AL136842 CDC42 effector protein (Rho GTPase binding) 3
 AL136861 hypothetical protein DKFZp434B044
 AL136883 putative homeodomain transcription factor 2
 AL136885 chromosome 10 open reading frame 45
 AL136920 poly(A) binding protein interacting protein 1
 AL136924 Ras and Rab interactor 2

 AL136932 KIAA0922, elongation; Homo sapiens mRNA; cDNA DKFZp586H1322 (from clone DKFZp586H1322); complete cds.
 AL137317 hypothetical protein LOC284058
 AL137398 Similar to hypothetical protein (LOC388094), mRNA
 AL137438 SEC15-like 1 (S. cerevisiae)
 AL137616 LOC388443 (LOC388443), mRNA
 AL137679 3' exoribonuclease
 AL137725 epiplakin 1
 AL137725 epiplakin 1

 AL137751 radixin (Homo sapiens); Homo sapiens mRNA; cDNA DKFZp434I0812 (from clone DKFZp434I0812); partial cds.
 AL137753 KIAA1033 protein
 AL138444 mitochondrial ribosomal protein S25

 match: proteins: Tr:O02330 Tr:Q9Y2J5 Tr:Q9V5I7; Human DNA sequence from clone RP1-19N1 on chromosome Xq21.33-22.3 Contains a gene for a novel protein. Contains ESTs, STSs and GSSs, complete sequence.
 AL139812 G protein pathway suppressor 2
 AL157493
 AL161725
 AL162039 MOB1, Mps One Binder kinase activator-like 1A (yeast)
 AL162047 nuclear receptor coactivator 4
 AL162069 hypothetical protein LOC144501
 AL162074 CDC42 effector protein (Rho GTPase binding) 4
 AL163248

 AL353950 protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform (calcineurin A alpha)
 AL359338 cardiomyopathy associated 5
 AL359567 MRNA; cDNA DKFZp547D023 (from clone DKFZp547D023)
 AL359571 ninein (GSK3B interacting protein)
 AL359577 kelch repeat and BTB (POZ) domain containing 6
 AL359652 EST from clone DKFZp434A0411, full insert
 AL359939 vacuolar protein sorting 54 (yeast)
 AL360145 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 839551.
 AL365404 G protein-coupled receptor 108
 AL450314 kraken-like

AL512694 NAD synthetase 1
 AL513583 AL513583 Homo sapiens PLACENTA Homo sapiens cDNA clone
 XCL0BA001ZA05 3-PRIME, mRNA sequence.
 AL514271 cytochrome b-561
 AL514547 AL514547 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
 CLOBB004ZC07 3-PRIME, mRNA sequence.
 AL515061 5'-nucleotidase, cytosolic II-like 1
 AL515270 RNA (guanine-9-) methyltransferase domain containing 2
 AL515918 voltage-dependent anion channel 1
 AL516202 chromosome 2 open reading frame 22
 AL516350 actin related protein 2/3 complex, subunit 5, 16kDa
 AL517946 RNA binding motif, single stranded interacting protein 1
 AL518328 Similar to basic leucine zipper and W2 domains 1 (LOC151579), mRNA
 AL519376 AL519376 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo
 sapiens cDNA clone CS0DB004YH18 3-PRIME, mRNA sequence.
 AL519710 immunoglobulin superfamily, member 4
 AL520657 hypothetical protein MGC14327
 AL521101 WD repeat domain 5
 AL522406 hypothetical protein FLJ12661
 AL522667 putative nuclear protein ORF1-FL49
 AL523776 OTU domain, ubiquitin aldehyde binding 1
 AL524045 hypothetical protein MGC12103
 AL524175 hypothetical protein LOC116064
 AL524467 Clone IMAGE:4797078, mRNA
 AL525206 translokin
 AL525780 solute carrier family 39 (zinc transporter), member 13
 AL525798 acyl-CoA synthetase long-chain family member 3
 AL527365 RAD23 homolog B (*S. cerevisiae*)
 AL530462 zinc finger protein 364
 AL534702 AL534702 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
 CS0DF006YN17 3-PRIME, mRNA sequence.
 AL537042 ADP-ribosylation factor 4
 AL537707 putative translation initiation factor
 AL539253 AL539253 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
 CS0DF034YK07 5-PRIME, mRNA sequence.
 AL541302 AL541302 Homo sapiens PLACENTA Homo sapiens cDNA clone
 CS0DE006YI10 5-PRIME, mRNA sequence.
 AL548941 AL548941 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DI042YI15 3-PRIME, mRNA sequence.
 AL550722 small membrane protein 1
 AL550977 rabaptin, RAB GTPase binding effector protein 1
 AL551823 mitochondrial ribosomal protein L32
 AL553942 hypothetical protein FLJ31951
 AL555107 KIAA1892
 AL558987 ring finger protein 149
 AL559283 sorting nexin 12
 AL561281 hypothetical protein FLJ20373
 AL562031 AL562031 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo
 sapiens cDNA clone CS0DB003YG22 3-PRIME, mRNA sequence.

AL562398 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC008YD20 3-PRIME, mRNA sequence.
 AL562528 ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast)
 AL563460 GATA binding protein 2
 AL564683 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CS0DM007YK12 3-PRIME, mRNA sequence.
 AL565516 pantothenate kinase 3
 AL565621 coactosin-like 1 (Dictyostelium)
 AL565741 hypothetical gene supported by AF038182; BC009203
 AL565749 tubulin, beta, 4
 AL566034 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF016YJ06 3-PRIME, mRNA sequence.
 AL566172 ATPase, H⁺ transporting, lysosomal 38kDa, V0 subunit d isoform 1
 AL567411 cyclin-dependent kinase 5, regulatory subunit 1 (p35)
 AL567431 chromosome 9 open reading frame 5
 AL567808 zinc finger protein 23 (KOX 16)
 AL567820 actin, gamma 1
 AL569476 ankyrin repeat domain 13
 AL569804 likely ortholog of mouse semaF cytoplasmic domain associated protein 3 membrane cofactor protein (CD46, trophoblast-lymphocyte cross-reactive antigen)
 AL570661 hypothetical protein MGC14156
 AL570697
 AL571375 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI009YH10 3-PRIME, mRNA sequence.
 AL572015 hypothetical protein FLJ20313
 AL573637 UDP-glucuronate decarboxylase 1
 AL573951 butyrate-induced transcript 1
 AL574096 tissue factor pathway inhibitor 2
 AL575306 H19, imprinted maternally expressed untranslated mRNA
 AL575747 Transcribed sequence with weak similarity to protein ref:NP_055131.1 (H.sapiens) calcium-regulated heat-stable protein
 AL576253 dedicator of cytokinesis 9
 AL577809 MKI67 (FHA domain) interacting nucleolar phosphoprotein
 AL577823 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK007YN19 3-PRIME, mRNA sequence.
 AL578310 protein tyrosine phosphatase type IVA, member 1
 AL578336 cirrhosis, autosomal recessive 1A (cirhin)
 AL578668 synaptotagmin binding, cytoplasmic RNA interacting protein
 AL582889 exosomal core protein CSL4
 AL589593 MRNA; cDNA DKFZp686D04119 (from clone DKFZp686D04119)
 AL832312 zinc finger CCCH type domain containing 5
 AL832653 MRNA; cDNA DKFZp313A1412 (from clone DKFZp313A1412)
 AL834286 hypothetical protein FLJ13815
 AP000693 Homo sapiens genomic DNA, chromosome 21q22.2, BAC clone:KB739C11, CBR1-HLCS region.
 AP001745
 AP001745
 AU134977 MRNA; cDNA DKFZp686L01105 (from clone DKFZp686L01105)
 AU144066 zinc finger protein 24 (KOX 17)
 AU144104 general transcription factor IIA, 1, 19/37kDa

AU144503 Transcribed sequence with weak similarity to protein pir:A36353 (H.sapiens)
 AU145019 A36353 DNA repair protein XRCC1 - human
 AU145192 GRP1-binding protein GRSP1
 AU145309 actin, gamma 1
 AU145351 papilin, proteoglycan-like sulfated glycoprotein
 AU146275 protein tyrosine phosphatase, receptor type, F
 AU146275 HEMBB1 Homo sapiens cDNA clone HEMBB1000004 3', mRNA
 sequence.
 AU146850 KIAA1340 protein
 AU146924 CDNA FLJ12033 fis, clone HEMBB1001899
 AU147416 Wilms tumor 1 associated protein
 AU147591 MRNA fragment.
 AU148466 karyopherin (importin) beta 3
 AU149385 ELL associated factor 1
 AU149908 UBX domain containing 2
 AU150691 CDNA FLJ43745 fis, clone TESTI2019648
 AU150728 zinc finger protein 267
 AU151331 ligand of numb-protein X 2
 AU151342 CDNA FLJ12935 fis, clone NT2RP2004982
 AU151801 complement component 1, q subcomponent binding protein
 AU152505 mitogen-activated protein kinase 8
 AU153138 KIAA0033 protein
 AU154321 karyopherin alpha 6 (importin alpha 7)
 AU154691 desmocollin 2
 AU154740 SPTF-associated factor 65 gamma
 AU155298 chromodomain helicase DNA binding protein 1
 AU156837 CDNA FLJ13747 fis, clone PLACE3000276
 AU157224 hypothetical protein DKFZp434D2328
 AU157541 hypothetical protein FLJ22833
 AU157543 CDNA FLJ14294 fis, clone PLACE1008181
 AU157716 CDNA FLJ13585 fis, clone PLACE1009150
 AU157881 PLACE1 Homo sapiens cDNA clone PLACE1009921 3', mRNA
 sequence.
 AU157915 hypothetical protein FLJ31657
 AU158062 RAB2, member RAS oncogene family
 AU158463 hypothetical protein LOC196996
 AU158495 Notch homolog 2 (Drosophila)
 AV681807 v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
 AV682285 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 AV682436 phosphoinositide-3-kinase, class 2, alpha polypeptide
 AV686514 epithelial membrane protein 2
 AV691491 hypothetical protein LOC161291
 AV692425 Clone IMAGE:4798349, mRNA
 AV693403 AV693403 GKC Homo sapiens cDNA clone GKCFCB09 5', mRNA sequence.
 AV699347 Transcribed sequences
 AV699389 AV699389 GKC Homo sapiens cDNA clone GKCETA09 3', mRNA sequence.
 AV699744 Transcribed sequence with strong similarity to protein pir:T12523 (H.sapiens)
 T12523 hypothetical protein DKFZp434K063.1 - human
 AV700003 ADP-ribosylation factor-like 6 interacting protein 2

AV700323 mannosidase, alpha, class 2A, member 1
 AV700415 Transcribed sequences
 AV700849 AV700849 GKC Homo sapiens cDNA clone GKCEEG04 3', mRNA sequence.
 AV701177 arrestin domain containing 4
 AV701318 CDNA clone IMAGE:4446165, partial cds
 AV701987 chromosome 10 open reading frame 9
 AV702197 CDNA FLJ45369 fis, clone BRHIP3017325
 AV702506 AV702506 ADB Homo sapiens cDNA clone ADBCBG02 5', mRNA sequence.
 AV703054 insulin receptor
 AV703731 AV703731 ADB Homo sapiens cDNA clone ADBCWA12 5', mRNA sequence.
 AV704551 COMM domain containing 6
 splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)
 AV705803
 AV705934 AV705934 ADB Homo sapiens cDNA clone ADBDCE04 5', mRNA sequence.
 AV706343 Ras-associated protein Rap1
 AV706522 AV706522 ADB Homo sapiens cDNA clone ADBDAF07 5', mRNA sequence.
 AV709094 CDNA FLJ33107 fis, clone TRACH2000959
 AV710542 AV710542 Cu Homo sapiens cDNA clone CuAAJF07 5', mRNA sequence.
 protein tyrosine phosphatase-like (proline instead of catalytic arginine), member b
 AV712602
 AV712733 filamin B, beta (actin binding protein 278)
 AV712912 protein x 013
 AV714014 splicing factor, arginine/serine-rich 2, interacting protein
 AV715578 decapping enzyme hDcp2
 AV715767 LIM protein (similar to rat protein kinase C-binding enigma)
 AV718192 triple functional domain (PTPRF interacting)
 AV721987 cancer susceptibility candidate 3
 AV724107 chromosome 6 open reading frame 56
 AV724183 Clone IMAGE:5294815, mRNA
 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
 AV725365
 AV725664 phosphatidic acid phosphatase type 2B
 AV727101 AV727101 HTC Homo sapiens cDNA clone HTCBGF05 5', mRNA sequence.
 AV727449 p300/CBP-associated factor
 AV728268 sortilin-related receptor, L(DLR class) A repeats-containing
 AV728526 macrophage expressed gene 1
 AV728658 hypothetical protein FLJ10824
 AV729462 similar to FRG1 protein (FSHD region gene 1 protein)
 AV729634 DnaJ (Hsp40) homolog, subfamily C, member 6
 Transcribed sequence with weak similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 AV734194
 AV734646 chromosome 6 open reading frame 187
 AV734843 AV734843 cdA Homo sapiens cDNA clone cdAAHD10 5', mRNA sequence.
 AV735100 AV735100 cdA Homo sapiens cDNA clone cdAAID09 5', mRNA sequence.
 AV740426 ankyrin repeat domain 28
 AV740879 chromosome 14 open reading frame 43
 AV756026 chromodomain helicase DNA binding protein 2
 AV757384 RNA-binding region (RNP1, RRM) containing 7
 AV757675 AV757675 BM Homo sapiens cDNA clone BMFAVB12 5', mRNA sequence.
 AV760596 CGI-125 protein
 AV762065 casein kinase 1, epsilon

AW001026 chromosome 14 open reading frame 58
 AW001101 KIAA0368
 AW001847 amyloid beta (A4) precursor-like protein 2
 AW002970 chromosome 18 open reading frame 9
 AW003222 ankyrin repeat and SOCS box-containing 7
 AW005535 RAP2B, member of RAS oncogene family
 AW005866 Transcribed sequences
 AW006750 DRE1 protein
 AW006938 neurolysin (metallopeptidase M3 family)
 AW007319 Transcribed sequences
 ws52h07.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2500861
 AW007532 3', mRNA sequence.
 AW007710 MRNA for FLJ00016 protein, partial cds
 AW008976 hypothetical protein DKFZp761O0113
 Transcribed sequence with moderate similarity to protein pdb:2RMA
 (H.sapiens) S Chain S, Cyclophilin A
 AW009630 Full length insert cDNA YI37C01
 AW014647 protein containing single MORN motif in testis
 AW015683 leptin receptor
 AW026535 MRNA, chromosome 1 specific transcript KIAA0500.
 AW028110 KIAA1946 protein
 AW043602 cyclin-dependent kinase 6
 AW051349 CDNA FLJ36689 fis, clone UTERU2008653, highly similar to GLYCODELIN
 AW051365 PRECURSOR
 wz04a05.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2557040 3'
 similar to gb:X53416 ENDOTHELIAL ACTIN-BINDING PROTEIN (HUMAN);,
 AW051856 mRNA sequence.
 AW055008 aminopeptidase puromycin sensitive
 AW080618 hypothetical protein LOC116068
 AW081685 Hypothetical LOC339226 (LOC339226), mRNA
 AW084511 mucin 20
 AW117765 peroxisome biogenesis factor 13
 AW118862 ras responsive element binding protein 1
 AW129145 Transcribed sequences
 AW129593 tudor repeat associator with PCTAIRE 2
 AW134535 cyclin G2
 AW135306 Transcribed sequences
 UI-H-BI1-acp-b-03-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
 AW137073 IMAGE:2714837 3', mRNA sequence.
 AW138157 hypothetical protein MGC24665
 AW138767 hypothetical protein FLJ23563
 AW138794 Clone 24841 mRNA sequence
 AW139179 fem-1 homolog b (C. elegans)
 AW149498 BTB (POZ) domain containing 6
 xg42b08.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2630199 3'
 AW150923 similar to contains Alu repetitive element;,, mRNA sequence.
 AW152589 B-cell receptor-associated protein 29
 AW161626 transportin 1
 v-abl Abelson murine leukemia viral oncogene homolog 2 (arg, Abelson-related
 AW173164 gene)

AW173623 tumor differentially expressed 1
 AW188940 beta-2-microglobulin
 AW190070 ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
 AW190873 ROD1 regulator of differentiation 1 (S. pombe)
 AW193511 HMBA-inducible
 AW194730 serine/threonine kinase 17a (apoptosis-inducing)
 xn38a03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2695948 3',
 AW195407 mRNA sequence.
 xn84c09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:2701168 3' similar to contains L1.t1 L1 repetitive element ;, mRNA
 sequence.
 AW195572 Clone IMAGE:5312086, mRNA
 AW205474 Transcribed sequences
 AW206286 DEAD (Asp-Glu-Ala-Asp) box polypeptide 43
 AW206656 methyl-CpG binding domain protein 6
 AW207668 xn20d09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2694257 3',
 mRNA sequence.
 AW235608 zinc finger protein 363
 AW236209 HBxAg transactivated protein 2
 AW238632 RNA-binding protein
 AW241742 Transcribed sequences
 AW242920 xm91g11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2691620 3'
 AW242981 similar to TR:O95620 O95620 PP35. [1] ;, mRNA sequence.
 xm92d07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2691661
 AW243097 3', mRNA sequence.
 AW246673 chromosome 6 open reading frame 37
 AW250952 dipeptidylpeptidase 3
 AW268365 ERO1-like (S. cerevisiae)
 AW268585 casein kinase 1, alpha 1
 AW269397 spermatogenesis associated 13
 AW269834 phosphodiesterase 7A
 AW270932 transducer of Cdc42-dependent actin assembly 1
 xs16c07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2769804 3',
 AW271609 mRNA sequence.
 AW272611 thymopoietin
 xv24c03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 AW273811 IMAGE:2814052 3', mRNA sequence.
 Transcribed sequence with moderate similarity to protein sp:O15320
 AW274747 (H.sapiens) MEA6_HUMAN Meningioma-expressed antigen 6/11
 AW291264 ADP-ribosylation factor-like 5
 AW292751 SAM, WWE and DDHD domain containing 1
 AW292882 synovial sarcoma translocation, chromosome 18
 AW293296 Transcribed sequences
 UI-H-BI2-ahj-d-03-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
 AW293341 IMAGE:2726837 3', mRNA sequence.
 neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson
 AW293356 disease)
 AW294022 KIAA1718 protein
 AW294630 SEC15-like 2 (S. cerevisiae)
 AW294869 SKI-like
 AW295395 CDNA FLJ41107 fis, clone BLADE2007923

AW295549 additional sex combs like 2 (Drosophila)
 AW296050 hypothetical protein FLJ10808
 AW298119 sec1 family domain containing 1
 AW298170 mitogen-activated protein kinase kinase kinase 5
 AW299555 ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, *C. elegans*)
 xs43f10.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2772427 3'
 AW299815 similar to contains L1.t2 L1 repetitive element ;, mRNA sequence.
 AW299828 hypothetical protein LOC255189
 AW299958 3'-phosphoadenosine 5'-phosphosulfate synthase 2
 AW300204 solute carrier family 30 (zinc transporter), member 8
 AW300965 family with sequence similarity 11, member A
 AW301218 THAP domain containing 9
 AW303865 chromosome 20 open reading frame 22
 AW339310 dystrobrevin, alpha
 AW341649 tumor protein p53 inducible nuclear protein 1
 Transcribed sequence with weak similarity to protein ref:NP_055301.1
 (H.sapiens) neuronal thread protein [Homo sapiens]
 AW402635 MRNA; cDNA DKFZp686J18198 (from clone DKFZp686J18198)
 AW409599 secretory carrier membrane protein 2
 AW438674 melanoma antigen, family A, 4
 AW444761 cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
 AW450329 CDNA FLJ36584 fis, clone TRACH2013450
 AW451785 Transcribed sequences
 AW451954 AP2 associated kinase 1
 pleckstrin homology domain containing, family C (with FERM domain) member
 1
 AW469573
 AW473802 golgi membrane protein SB140
 Similar to bA182L21.1 (novel protein similar to hypothetical proteins)
 (LOC399788), mRNA
 AW501360
 AW511135 nudix (nucleoside diphosphate linked moiety X)-type motif 4
 AW511319 Mesenchymal stem cell protein DSC96 mRNA, partial cds
 xx76g11.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2849636
 AW512586 3', mRNA sequence.
 AW513835 development and differentiation enhancing factor 1
 AW517686 ATPase, Ca⁺⁺ transporting, plasma membrane 4
 TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor,
 250kDa
 AW575233
 AW575379 protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
 AW611550 Full length insert cDNA clone YW29A12
 AW611917 MRNA; cDNA DKFZp686F2198 (from clone DKFZp686F2198)
 AW612311 CDNA FLJ38130 fis, clone D6OST2000464
 AW628168 Transcribed sequences
 AW664850 nuclear protein double minute 1
 AW665538 Transcribed sequences
 AW731710 histidine triad nucleotide binding protein 3
 AW771590 ras homolog gene family, member Q
 AW771935 TEA domain family member 1 (SV40 transcriptional enhancer factor)
 Transcribed sequence with weak similarity to protein ref:NP_060265.1
 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 AW796364 RC3-LT0023-200100-012-e08 LT0023 Homo sapiens cDNA, mRNA sequence.
 AW836210
 AW873330 B-cell receptor-associated protein 29

AW873564 Transcribed sequences
 AW902062 Transcribed sequences
 EST372778 MAGE resequences, MAGF Homo sapiens cDNA, mRNA
 sequence.
 AW960707
 AW962511 hypothetical protein FLJ22531
 EST375401 MAGE resequences, MAGH Homo sapiens cDNA, mRNA
 sequence.
 AW963328
 AW970089 nanos homolog 1 (Drosophila)
 AW974642 CDNA FLJ36478 fis, clone THYMU2017362
 AW975057 CDC-like kinase 4
 AW976269 chromosome 14 open reading frame 101
 EST388456 MAGE resequences, MAGN Homo sapiens cDNA, mRNA
 sequence.
 AW976347
 AW977401 ubiquitin specific protease 38
 AY005111 LUC7-like (S. cerevisiae)
 AY007126 hypothetical protein LOC339229
 AY007128 CDNA FLJ26765 fis, clone PRS02774
 AY099328 GABA(A) receptor-associated protein
 BC000018 actin related protein 2/3 complex, subunit 5-like
 BC000103 NCK adaptor protein 2
 BC000182 annexin A4
 BC000185 carnitine palmitoyltransferase 1A (liver)
 BC000296 oxysterol binding protein-like 2
 BC000324 granulins
 BC000353
 BC000373 amyloid beta (A4) precursor-like protein 2
 BC000398 platelet-activating factor acetylhydrolase, isoform Ib, beta subunit 30kDa
 BC000400 protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform
 BC000433 mitogen-activated protein kinase 13
 BC000464 JM5 protein
 BC000474 tumor protein p53 inducible protein 3
 BC000478 heat shock 70kDa protein 9B (mortalin-2)
 BC000527 Ewing sarcoma breakpoint region 1
 BC000580 hypoxia-inducible factor prolyl 4-hydroxylase
 BC000629 aspartyl-tRNA synthetase
 BC000665 t-complex 1
 BC000687 translocation associated membrane protein 1
 BC000704 transmembrane 4 superfamily member 8
 BC000723 carnitine acetyltransferase
 BC000738 emerin (Emery-Dreifuss muscular dystrophy)
 BC000756 mitochondrial ribosomal protein L4
 BC000758 chromosome 6 open reading frame 80
 BC000764 chromosome 6 open reading frame 166
 BC000771 tropomyosin 3
 BC000832 zinc finger protein 183 (RING finger, C3HC4 type)
 BC000836 yippee protein
 BC000878 lactamase, beta 2
 BC000905 RAB1A, member RAS oncogene family
 BC000915 PDZ and LIM domain 1 (elfin)
 BC000948 hypothetical protein FLJ10525
 BC000977 aminolevulinic acid, delta-, dehydratase

BC001004
 BC001033 transmembrane protein 14B
 BC001081 anaphase promoting complex subunit 5
 BC001131 Homo sapiens cDNA clone IMAGE:2989839, with apparent retained intron.
 BC001169 esterase D/formylglutathione hydrolase
 BC001193 Homo sapiens histone 3, H2a, mRNA (cDNA clone IMAGE:3355200).
 BC001196 heparan sulfate 6-O-sulfotransferase 1
 BC001220 hypothetical protein MGC874
 BC001228 COMM domain containing 2
 BC001288 decay accelerating factor for complement (CD55, Cromer blood group system)
 BC001312 thioredoxin domain containing 7 (protein disulfide isomerase)
 tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein,
 beta polypeptide
 BC001359
 BC001364 SEC22 vesicle trafficking protein-like 1 (*S. cerevisiae*)
 BC001373 mesoderm development candidate 1
 BC001404 KIAA0310
 BC001450 nuclear cap binding protein subunit 1, 80kDa
 BC001659 RNA binding protein S1, serine-rich domain
 BC001709 NAD kinase
 BC001716 poly(A) binding protein interacting protein 2
 BC001771 general transcription factor IIF, polypeptide 2, 30kDa
 BC001903 interleukin 10 receptor, beta
 BC002356 nucleobindin 1
 BC002374 karyopherin alpha 1 (importin alpha 5)
 BC002446 DnaJ (Hsp40) homolog, subfamily B, member 6
 BC002449 EF hand domain containing 1
 BC002506 programmed cell death 10
 BC002515 aldehyde dehydrogenase 7 family, member A1
 BC002548 Mov10, Moloney leukemia virus 10, homolog (mouse)
 BC002556
 BC002559 high-glucose-regulated protein 8
 BC002640 intraflagellar transport protein IFT20
 BC002649 Homo sapiens histone 1, H1c, mRNA (cDNA clone IMAGE:3608862).
 BC002675 HESB like domain containing 2
 BC002682 dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)
 BC002700 keratin 7
 BC002704 signal transducer and activator of transcription 1, 91kDa
 BC002719 eukaryotic translation initiation factor 3, subunit 1 alpha, 35kDa
 BC002729 protein arginine N-methyltransferase 6
 BC002842 histone 1, H2bd
 BC002906 uridine monophosphate kinase
 BC003005 unactive progesterone receptor, 23 kD
 BC003049 PAI-1 mRNA-binding protein
 BC003060 fucosidase, alpha-L- 2, plasma
 BC003090 COP9 constitutive photomorphogenic homolog subunit 8 (*Arabidopsis*)
 BC003376 ELAV (embryonic lethal, abnormal vision, *Drosophila*)-like 1 (Hu antigen R)
 BC003390 chromosome 13 open reading frame 12
 BC003418 cyclic AMP phosphoprotein, 19 kD
 BC003560 ribophorin II
 BC003564 ATPase, H⁺ transporting, lysosomal 13kDa, V1 subunit G isoform 1

BC003572	karyopherin (importin) beta 1
BC003573	farnesyl-diphosphate farnesyltransferase 1
BC003600	LIM domain only 4
BC003623	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
BC003682	cell division cycle 42 (GTP binding protein, 25kDa)
BC003701	DKFZP434I116 protein
BC004130	nuclear domain 10 protein
BC004169	hypothetical protein MGC2714
BC004227	metastasis associated family, member 3
BC004247	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
BC004331	chromosome 9 open reading frame 99
BC004354	
BC004372	CD44 antigen (homing function and Indian blood group system)
BC004395	apolipoprotein L, 2
BC004419	vacuolar protein sorting 24 (yeast)
BC004421	zinc finger protein 330
BC004443	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E isoform 1
BC004446	chromosome 20 open reading frame 24
BC004489	major histocompatibility complex, class I, C
BC004523	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 4
BC004542	plexin B2
BC004548	hypothetical protein PRO1853
BC004815	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain
BC004818	chromosome 14 open reading frame 129
BC004911	Bernardinelli-Seip congenital lipodystrophy 2 (seipin)
BC004995	MARVEL (membrane-associating) domain containing 1
BC004998	likely ortholog of mouse membrane bound C2 domain containing protein
BC005009	hypothetical protein MGC10500
BC005064	hypothetical protein MGC12972
BC005084	ring finger protein 135
BC005097	cytoplasmic FMR1 interacting protein 1
BC005123	serine palmitoyltransferase, long chain base subunit 2
BC005127	adipose differentiation-related protein
BC005169	Ser/Thr-like kinase
BC005181	N-acetyltransferase 5 (ARD1 homolog, <i>S. cerevisiae</i>)
BC005295	poly(A) binding protein interacting protein 1
BC005336	uncharacterized hematopoietic stem/progenitor cells protein MDS031
BC005345	general transcription factor IIH, polypeptide 2, 44kDa
BC005352	tumor necrosis factor, alpha-induced protein 8
BC005372	Homo sapiens, clone IMAGE:3932794, mRNA.
BC005808	mitochondrial translation optimization 1 homolog (<i>S. cerevisiae</i>)
BC005821	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)
BC005871	hypothetical protein MGC4248
BC005932	proteasome (prosome, macropain) subunit, alpha type, 1
BC005978	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)
BC005999	zinc finger, FYVE domain containing 21
BC006088	tumor differentially expressed 1
BC006110	hypothetical protein MGC12966

BC006177 metastasis associated 1
 BC006230 monoglyceride lipase
 BC006237 HECT domain containing 1
 BC006279 similar to Zinc finger protein 136
 BC006316 KIAA1191 protein
 BC006332 clathrin, light polypeptide (Lcb)
 BC006471 ALL1-fused gene from chromosome 1q
 BC007934
 BC009506 putative MAPK activating protein
 Homo sapiens, Similar to neuronal thread protein, clone IMAGE:4102657,
 mRNA.
 BC009590
 BC009712 ATP-binding cassette, sub-family D (ALD), member 3
 BC011002 hypothetical protein LOC113230
 BC012117 serologically defined colon cancer antigen 10
 BC012484 Homo sapiens cDNA clone IMAGE:4472100, partial cds.
 BC012846 isocitrate dehydrogenase 1 (NADP+), soluble
 BC016017 hypothetical protein LOC157697
 BC016183 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
 BC016339 cysteine-rich motor neuron 1
 BC020552 programmed cell death 6
 BC022079 CG10806-like
 BC028081 chromosome 2 open reading frame 18
 BC028974 hypothetical protein LOC126295
 BC029352 retinoblastoma binding protein 6
 BC029890 annexin A8
 BC032406 hypothetical protein FLJ23129
 BC040456 F-box only protein 8
 BC040628 Clone IMAGE:5745274, mRNA
 BC040884 Homo sapiens, clone IMAGE:5735420, mRNA.
 BC042510 carboxyl ester lipase (bile salt-stimulated lipase)
 BC043601 Homo sapiens, clone IMAGE:5228040, mRNA.
 BE048571 hypothetical protein MGC16121
 BE217882 KIAA1718 protein
 BE218028 dynactin 4 (p62)
 BE220265 wingless-type MMTV integration site family, member 9A
 hu22a05.x1 NCI_CGAP_Mel15 Homo sapiens cDNA clone IMAGE:3170768 3'
 similar to gb:K01228 PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR
 (HUMAN);, mRNA sequence.
 BE221212
 hu51g06.x1 NCI_CGAP_Brn41 Homo sapiens cDNA clone IMAGE:3173626 3',
 mRNA sequence.
 BE222709
 BE250417 adenovirus 5 E1A binding protein
 BE256900 jumonji domain containing 2B
 601112031F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3352879 5',
 mRNA sequence.
 BE259137
 600944342T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960218 3',
 mRNA sequence.
 BE299671
 BE302305 transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)
 BE311760 high-mobility group box 1
 BE326738 hypothetical protein MGC24180
 hs98f09.x1 NCI_CGAP_Kid13 Homo sapiens cDNA clone IMAGE:3145289 3',
 mRNA sequence.
 BE328496

BE348688 amine oxidase (flavin containing) domain 1
 BE350312 Transcribed sequences
 BE385892 KIAA1228 protein
 BE396879 similar to splicing factor, arginine/serine-rich 4
 BE464077 BRCA2 and CDKN1A interacting protein
 hz75g08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3213854 3',
 mRNA sequence.
 BE467916 tripartite motif-containing 4
 BE501464 Transcribed sequences
 BE501976 MAX gene associated
 BE502432 hypothetical protein FLJ30525
 BE502436 yippee-like 2 (Drosophila)
 BE502982 CDNA FLJ26687 fis, clone MPG06534
 BE504098 601064390F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3450599 5',
 mRNA sequence.
 BE541042 vestigial like 1 (Drosophila)
 BE542323 601068994F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3455374 5',
 mRNA sequence.
 BE543064 adducin 3 (gamma)
 BE545756 ras homolog gene family, member V
 BE547917 Transcribed sequences
 BE550501 hypothetical protein BC007882
 BE563042 transcription factor CP2-like 2
 BE566136 601339864F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3682406 5',
 mRNA sequence.
 BE566290 hypothetical protein MGC10198
 BE613081 class II bHLH protein MIST1
 BE616972 601441509T1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845643 3',
 mRNA sequence.
 BE617588 cell division cycle and apoptosis regulator 1
 BE617899 neuropilin 1
 BE620457 601483124T1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885917 3',
 mRNA sequence.
 BE620598 cytoplasmic polyadenylation element binding protein 4
 BE620832 Clone IMAGE:5285100, mRNA
 BE621082 ubiquitin-conjugating enzyme E2D 2 (UBC4/5 homolog, yeast)
 BE621259 601440792T1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915695 3',
 mRNA sequence.
 BE622627 CCAAT/enhancer binding protein (C/EBP), gamma
 BE622659 601441191T1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3916031 3',
 mRNA sequence.
 BE622841 Rho GTPase activating protein 18
 BE644830 SEC24 related gene family, member A (S. cerevisiae)
 BE645231 7a46g12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3221830 3'
 BE671949 similar to contains MER15.b2 MER15 repetitive element ;, mRNA sequence.
 Transcribed sequence with weak similarity to protein ref:NP_071385.1
 BE672408 (H.sapiens) hypothetical protein FLJ20958 [Homo sapiens]
 BE674006 Transcribed sequences
 BE674143 7d75g01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278832 3'
 similar to contains element MER1 repetitive element ;, mRNA sequence.

BE674964 Transcribed sequences
 BE675139 cAMP responsive element binding protein 3-like 2
 BE675800 retinoid X receptor, alpha
 BE675995 non-kinase Cdc42 effector protein SPEC2
 BE708432 PRO1073 protein
 BE738425 chromosome 14 open reading frame 147
 BE740137 WD repeat domain 5
 BE740761 histone 1, H4h
 BE741869 hypothetical protein BC013767
 BE744389 hypothetical protein MGC49942
 BE778706 CDNA clone IMAGE:5092935, partial cds
 BE780075 transmembrane trafficking protein
 BE782754 syntaxin 16
 601480343F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3882995 5', mRNA sequence.
 BE789346
 BE855765 RGM domain family, member B
 BE856541 hypothetical protein BC016683
 BE866412 suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)
 BE872974 acid phosphatase 1, soluble
 BE876628 GA binding protein transcription factor, alpha subunit 60kDa
 BE877775 CDNA: FLJ22165 fis, clone HRC00470
 BE877775 CDNA: FLJ22165 fis, clone HRC00470
 BE879873 progesterone receptor membrane component 2
 BE880703 sphingosine-1-phosphate phosphatase 1
 BE883841 sestrin 3
 BE891920 actin related protein 2/3 complex, subunit 4, 20kDa
 Transcribed sequence with strong similarity to protein ref:NP_073729.1 (H.sapiens) microtubule-associated proteins 1A/1B light chain 3 [Homo sapiens]
 BE893893
 BE895685 KIAA0853
 BE898861 heterogeneous nuclear ribonucleoprotein C (C1/C2)
 BE903880 CD44 antigen (homing function and Indian blood group system)
 601498628F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900702 5', mRNA sequence.
 BE904551
 UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNAc-T10)
 601498152F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3899964 5', mRNA sequence.
 BE908995
 BE927766 Transcribed sequences
 BE930512 hypothetical protein MGC5370
 BE961916 F-box only protein, helicase, 18
 BE962299 hypothetical protein FLJ10099
 601656052R1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3855805 3', mRNA sequence.
 BE962679
 601656874R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3865699 3', mRNA sequence.
 BE963245
 601658192R1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876166 3', mRNA sequence.
 BE964484
 601658524R1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885940 3', mRNA sequence.
 BE964598

BE964704 601658243R1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885682 3', mRNA sequence.
 BE965369 601659282R1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3895653 3', mRNA sequence.
 BE966247 F-box only protein 22
 BE966267 601660507R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906240 3', mRNA sequence.
 BE966922 601660942R1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915610 3', mRNA sequence.
 BE967331 601649324F1 NIH_MGC_73 Homo sapiens cDNA clone IMAGE:3933172 5', mRNA sequence.
 BE967532 midline 1 (Opitz/BBB syndrome)
 BE968786 tissue inhibitor of metalloproteinase 2
 BE972723 Transcribed sequence with weak similarity to protein sp:P39189 (H.sapiens)
 BE972774 ALU2_HUMAN Alu subfamily SB sequence contamination warning entry
 BE974098 unc-84 homolog A (C. elegans)
 BF000155 tumor protein D52
 BF000409 TNF receptor-associated factor 4
 BF001267 similar to S. cerevisiae SSM4
 BF001614 dedicator of cytokinesis 7
 BF001665 zinc finger protein ANC_2H01
 BF001670 O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)
 BF001714 ephrin-B2
 BF001919 carnitine palmitoyltransferase 1A (liver)
 BF002296 hypothetical protein LOC253512
 BF031829 thyroid hormone receptor interactor 8
 BF032213 desmoglein 2
 BF034561 chromosome 20 open reading frame 45
 BF05107 G-rich RNA sequence binding factor 1
 BF055201 RNA-binding region (RNP1, RRM) containing 7
 BF055289 7j76d08.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3392367 3', mRNA sequence.
 BF057649 hypothetical protein BC018453
 BF057717 peroxisomal membrane protein 4, 24kDa
 BF057855 hypothetical protein LOC147991
 BF059136 TATA element modulatory factor 1
 BF060747 ROD1 regulator of differentiation 1 (S. pombe)
 BF060776 hypothetical protein LOC130576
 BF061845 7j60e11.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3390860 3', mRNA sequence.
 BF062037 trans-golgi network protein 2
 BF062203 chromosome 7 open reading frame 30
 BF062244 KIAA2026
 BF064162 lin-7 homolog A (C. elegans)
 BF107565 endothelial-derived gene 1
 BF109251 601823915F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043714 5', mRNA sequence.
 BF109906 baculoviral IAP repeat-containing 4
 CDNA FLJ41747 fis, clone HSYRA2006873

BF110053 claudin 23
 BF111819 Nance-Horan syndrome (congenital cataracts and dental anomalies)
 mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase,
 isoenzyme A
 BF195431 HSPC056 protein
 BF195973 zinc finger protein 481
 BF214688 myosin IB
 BF215996 homeodomain interacting protein kinase 2
 BF218115 ATPase family homolog up-regulated in senescence cells
 BF218804 vav-1 interacting Kruppel-like protein
 BF219240 F-box and WD-40 domain protein 7 (archipelago homolog, Drosophila)
 BF222826 Clone IMAGE:5289004, mRNA
 BF222895 kinesin family member 5B
 BF223224 chromosome 6 open reading frame 136
 BF224092 transmembrane protein 5
 BF224146 Similar to RIKEN cDNA 1810037117 (LOC401152), mRNA
 BF244081 601854085F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4073973 5',
 mRNA sequence.
 BF246131 eukaryotic translation initiation factor 4B
 BF247371 solute carrier family 38, member 1
 BF247552 Yes-associated protein 1, 65kDa
 BF247906 Similar to poly(A) binding protein interacting protein 1 isoform 1; polyadenylate
 binding protein-interacting protein 1; PABC1-interacting protein 1
 (LOC388345), mRNA
 BF248165 CDNA FLJ38130 fis, clone D6OST2000464
 BF304695 regulator of G-protein signalling 16
 BF304996 phosphatidylinositol 3,4,5-trisphosphate-dependent RAC exchanger 1
 BF308645 601897391F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126486 5',
 mRNA sequence.
 BF311866 TIGA1
 BF314746 N-acetylglucosamine-1-phosphotransferase, gamma subunit
 BF338332 riboflavin kinase
 BF340123 MRNA; cDNA DKFZp779B1535 (from clone DKFZp779B1535)
 BF341845 transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
 BF343007 MRNA; cDNA DKFZp434D0818 (from clone DKFZp434D0818)
 BF346665 tripartite motif-containing 44
 BF431488 Hypothetical protein FLJ25157 (FLJ25157), mRNA
 BF431962 myosin IB
 BF432550 tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)
 BF433902 general transcription factor IIIc, polypeptide 4, 90kDa
 BF434224 musashi homolog 2 (Drosophila)
 BF435123 ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)
 BF435617 Transcribed sequence with moderate similarity to protein pdb:1LBG (E. coli) B
 Chain B, Lactose Operon Repressor Bound To 21-Base Pair Symmetric
 Operator Dna, Alpha Carbons Only
 BF435769 Clone IMAGE:5311370, mRNA
 BF437161 CDNA FLJ39656 fis, clone SMINT2005956
 BF437260 retinitis pigmentosa 9 (autosomal dominant)
 BF438270 RAB27B, member RAS oncogene family
 BF438386 pleckstrin homology domain containing, family F (with FYVE domain) member
 BF439250 2

BF444916 FAD104
 BF446578 RasGEF domain family, member 1A
 BF446673 hemicentin
 BF447037 junction-mediating and regulatory protein
 BF447954 CDNA FLJ26063 fis, clone PRS04788
 nad16a01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3365640
 BF448159 3', mRNA sequence.
 BF475862 ATPase, Class VI, type 11C
 7r08f01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3707280 3'
 BF477980 similar to contains Alu repetitive element;, mRNA sequence.
 BF478120 RecQ protein-like 5
 BF507862 CDNA FLJ46512 fis, clone THYMU3031878
 UI-H-BI4-aow-g-12-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
 BF507952 IMAGE:3086422 3', mRNA sequence.
 Transcribed sequence with strong similarity to protein ref:NP_061145.1
 (H.sapiens) ZNF135-like protein; zinc finger ZNF140-like protein [Homo
 sapiens]
 BF508739 UI-H-BI4-aor-e-06-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
 BF508848 IMAGE:3085907 3', mRNA sequence.
 UI-H-BI4-aoy-b-12-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
 BF509391 IMAGE:3086542 3', mRNA sequence.
 BF511231 tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)
 BF512188 Spir-1 protein
 BF513060 hypothetical protein FLJ11273
 BF513121 Clone IMAGE:4794726, mRNA
 BF514079 Kruppel-like factor 4 (gut)
 BF514723 helicase with zinc finger domain
 BF514975 hypothetical protein BM-002
 BF515750 hypothetical protein DKFZp547B1713
 BF515889 peptidylprolyl isomerase (cyclophilin)-like 4
 BF525395 zinc finger protein 385
 BF528646 SEC6-like 1 (*S. cerevisiae*)
 BF572029 chromodomain helicase DNA binding protein 6
 BF576458 nuclear receptor coactivator 1
 BF589024 kinectin 1 (kinesin receptor)
 BF589207 chromosome 10 open reading frame 47
 BF590861 chromosome 9 open reading frame 40
 BF591360 postsynaptic protein CRIPT
 BF591556 ras responsive element binding protein 1
 BF591611 hypothetical protein MGC45594
 BF593382 protein kinase, AMP-activated, beta 2 non-catalytic subunit
 BF593817 sideroflexin 1
 BF593914 TAK1-binding protein 3
 BF593940 succinate-CoA ligase, GDP-forming, beta subunit
 BF594695 Transcribed sequences
 BF668950 hypothetical protein FLJ12178
 BF669455 CD164 antigen, sialomucin
 602137554F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4274077 5',
 BF673888 mRNA sequence.
 BF675218 LOC388789 (LOC388789), mRNA
 BF675754 hypothetical protein MGC22014

BF676081 CDNA FLJ11174 fis, clone PLACE1007367
 BF679700 ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)
 BF680438 hypothetical protein FLJ23749
 Transcribed sequence with strong similarity to protein pir:S65784 (H.sapiens)
 BF683426 S65784 ribosomal protein L29, cytosolic - human
 602184834T1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299201 3',
 mRNA sequence.
 BF689173 UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5
 BF691447 ATP-binding cassette, sub-family A (ABC1), member 5
 BF693921 low density lipoprotein receptor-related protein 11
 BF696304 hypothetical protein FLJ90709
 BF696757 destrin (actin depolymerizing factor)
 BF697964 cyclin-dependent kinase 8
 BF700678 similar to RIKEN cDNA 2700047N05
 BF735901 hypothetical protein DKFZp761P0423
 BF739767 transcription factor (p38 interacting protein)
 BF739930 purine-rich element binding protein A
 BF739943 CDNA FLJ16053 fis, clone PLACE6019385, moderately similar to MITOGEN-
 ACTIVATED PROTEIN KINASE KINASE KINASE 5 (EC 2.7.1.-)
 BF739979 acheron
 BF792126 ornithine decarboxylase antizyme inhibitor
 BF793951 protein phosphatase 1, regulatory (inhibitor) subunit 15B
 BF796046 p21 (CDKN1A)-activated kinase 2
 BF796470 Transcribed sequences
 BF801735 transmembrane gamma-carboxyglutamic acid protein 4
 BF905445 sorting nexin family member 27
 BF939727 synaptonemal complex protein 3
 BF940025 hypothetical protein MGC39830
 BF940270 zinc finger-like protein 9
 BF940944 chromosome 3 open reading frame 6
 BF941088 zinc finger protein 406
 BF941325 hypothetical protein MGC10067
 BF965546 septin 10
 BF966021 602287009T1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4375586 3',
 mRNA sequence.
 BF966540 CDNA clone IMAGE:4346793, partial cds
 BF968134 ring finger protein 4
 BF968633 hypothetical protein
 BF969397 nuclear protein UKp68
 BF970253 hypothetical protein LOC283464
 BF977231 CGG triplet repeat binding protein 1
 BF979809 CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5,
 EJ16, EJ30, EL32 and G344)
 BF983379 602275807F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4363945 5',
 mRNA sequence.
 BG024886 KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2
 BG026159 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 BG027926 guanine nucleotide binding protein (G protein) alpha 12
 BG028884 ras homolog gene family, member E
 BG054844

BG054966 chromosome 19 open reading frame 13
 BG107845 membrane component, chromosome 11, surface marker 1
 BG109230 CDNA FLJ25573 fis, clone JTH06531
 602280715F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4368554 5',
 mRNA sequence.
 BG109597
 BG111761 guanine nucleotide binding protein (G protein), gamma 12
 BG112118 hypothetical protein LOC134218
 BG149218 phosphorylase kinase, beta
 BG150085 chromosome 11 open reading frame 1
 BG150636 hypothetical protein FLJ20519
 Transcribed sequence with weak similarity to protein ref:NP_060219.1
 (H.sapiens) hypothetical protein FLJ20294 [Homo sapiens]
 BG163591
 BG165333 membrane associated guanylate kinase interacting protein-like 1
 BG167675 hypothetical protein BC017881
 BG167841 MOB1, Mps One Binder kinase activator-like 2C (yeast)
 v-ral simian leukemia viral oncogene homolog B (ras related; GTP binding
 protein)
 BG169673
 602322848F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4426211 5',
 mRNA sequence.
 BG170478
 BG170541 met proto-oncogene (hepatocyte growth factor receptor)
 602331329F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4432615 5',
 mRNA sequence.
 BG180437
 BG231494 Full length insert cDNA clone ZB42D04
 BG231691 hypothetical protein FLJ32205
 BG231979 protein kinase, lysine deficient 2
 BG249221 hypothetical protein FLJ25952
 602362443F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4470898 5',
 mRNA sequence.
 BG250310
 602363024F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4471541 5',
 mRNA sequence.
 BG250721
 BG253884 HSPC063 protein
 BG258784 membrane component, chromosome 11, surface marker 1
 BG260519 likely ortholog of mouse aquarius
 BG285011 AT rich interactive domain 5B (MRF1-like)
 602382536F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4500129 5',
 mRNA sequence.
 BG286537
 BG287153 mannosidase, alpha, class 1A, member 1
 BG287862 hypothetical protein MGC5395
 BG289456 ubiquitin-specific proteinase 31
 BG290577 sperm associated antigen 9
 BG290908 ATPase, Class I, type 8B, member 1
 602386841F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4515730 5',
 mRNA sequence.
 BG291039
 BG292065 protein kinase C-like 2
 BG292405 implantation-associated protein
 BG338251 RAB7, member RAS oncogene family-like 1
 BG340967 trafficking protein particle complex 1
 602455521F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583885 5',
 mRNA sequence.
 BG386322
 spinocerebellar ataxia 7 (olivopontocerebellar atrophy with retinal
 degeneration)
 BG390306

BG398414 replication protein A1, 70kDa
 BG402105 RB1-inducible coiled-coil 1
 BG402553 chromosome 9 open reading frame 5
 BG403671 THO complex 2
 BG413366 MRNA; cDNA DKFZp686N0886 (from clone DKFZp686N0886)
 BG420237 heat shock 90kDa protein 1, alpha
 BG432350 chromosome 20 open reading frame 108
 BG434174 stoned B-like factor
 BG434703 dishevelled associated activator of morphogenesis 1
 BG480592 chromosome 20 open reading frame 11
 BG481972 eukaryotic translation initiation factor 5
 BG484552 chromosome 12 open reading frame 2
 BG485129 602503750F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4617157 5', mRNA sequence.
 BG495327 PRO0149 protein
 BG499947 mitogen-activated protein kinase associated protein 1
 BG500301 602546969F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669168 5', mRNA sequence.
 BG501219 hypothetical protein MGC23909
 BG528818 602579703F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4713600 5', mRNA sequence.
 BG530368 cyclin I
 BG534245 casein kinase 1, alpha 1
 BG534952 PRO1073 protein
 BG537190 ferritin, light polypeptide
 BG537255 602567289F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4691639 5', mRNA sequence.
 BG538564 KIAA1387 protein
 BG540048 protein phosphatase 2C, magnesium-dependent, catalytic subunit
 BG542521 KIAA0877 protein
 BI461155 CDNA FLJ33518 fis, clone BRAMY2005711
 BM354219 UI-E-EO1-aid-f-10-0-UI.s1 UI-E-EO1 Homo sapiens cDNA clone UI-E-EO1-aid-f-10-0-UI 3', mRNA sequence.
 BM677635 UI-CF-EC1-aec-p-21-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone UI-CF-EC1-aec-p-21-0-UI 3', mRNA sequence.
 BM978026 CDNA FLJ43258 fis, clone HHDPC1000001
 BU147359 MRNA; cDNA DKFZp564C2063 (from clone DKFZp564C2063)
 BU683415 hypothetical protein BC001610
 BU686397 hypothetical protein BC001610
 BU686397 leucine-rich PPR-motif containing
 CA430402 UI-H-EI0-ayo-p-15-0-UI.s1 NCI_CGAP_EI0 Homo sapiens cDNA clone UI-H-EI0-ayo-p-15-0-UI 3', mRNA sequence.
 CA448665 inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
 D13889 GDP dissociation inhibitor 2
 D13988 xeroderma pigmentosum, complementation group C
 D21089 heme oxygenase (decycling) 2
 D21243 protein-L-isoaspartate (D-aspartate) O-methyltransferase
 D25547 CD58 antigen, (lymphocyte function-associated antigen 3)
 D28586 IQ motif containing GTPase activating protein 1
 D29640

D29641	KIAA0052 protein
D31885	ADP-ribosylation factor-like 6 interacting protein
D31888	REST corepressor
D42043	raft-linking protein
D42055	neural precursor cell expressed, developmentally down-regulated 4
D42063	RAN binding protein 2
D43949	KIAA0082
D63480	KIAA0146 protein
D63879	squamous cell carcinoma antigen recognised by T cells 3
D79994	ankyrin repeat domain 15
D80004	KIAA0182 protein
D80010	lipin 1
D81004	Transcribed sequences
D83485	glucose regulated protein, 58kDa
D83702	cryptochrome 1 (photolyase-like)
D83782	SREBP CLEAVAGE-ACTIVATING PROTEIN
D84109	RNA binding protein with multiple splicing
D84454	solute carrier family 35 (UDP-galactose transporter), member A2
D85181	sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like
D86957	septin 8
D86985	KIAA0232 gene product
D87078	pumilio homolog 2 (Drosophila)
D87440	KIAA0252
D87446	RW1 protein
D87450	KIAA0261
D87682	KIAA0241 protein
D89729	exportin 1 (CRM1 homolog, yeast)
H05010	Nedd4 family interacting protein 1
H28915	Transcribed sequences
	yp64g06.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
H41167	IMAGE:192250 3', mRNA sequence.
H73636	hypothetical protein MGC10067
H93013	nuclear factor (erythroid-derived 2)-like 1
H98994	MRNA; cDNA DKFZp313M1034 (from clone DKFZp313M1034)
	guanine nucleotide binding protein (G protein), alpha inhibiting activity
J03198	polypeptide 3
J03250	topoisomerase (DNA) I
	dihydrolipoamide dehydrogenase (E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex, branched chain keto acid dehydrogenase complex)
J03620	Homo sapiens M2 mitochondrial autoantigen dihydrolipoamide
J03866	acetyltransferase mRNA, complete cds.
J04183	lysosomal-associated membrane protein 2
J05008	Homo sapiens endothelin-1 (EDN1) gene, complete cds.
J05021	villin 2 (ezrin)
	alpha-tubulin isotype H2-alpha; Human alpha-tubulin isotype H2-alpha gene, last exon.
K03460	
L04636	complement component 1, q subcomponent binding protein
L12387	sorcin
L12723	heat shock 70kDa protein 4
L14561	

L18964	protein kinase C, iota
L19184	peroxiredoxin 1
L20320	cyclin-dependent kinase 7 (MO15 homolog, <i>Xenopus laevis</i> , cdk-activating kinase)
L22453	ribosomal protein L3
L24521	hepatoma-derived growth factor (high-mobility group protein 1-like)
L48722	<i>Homo sapiens</i> (clone hh18) protein tyrosine phosphatase (ptp-IV1r) gene, 5' end of cds.
L76416	SMT3 suppressor of mif two 3 homolog 2 (yeast)
M12679	major histocompatibility complex, class I, C
M13981	inhibin, alpha
M18468	protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1)
M19154	
M19156	keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)
M23114	ATPase, Ca ⁺⁺ transporting, cardiac muscle, slow twitch 2
M23254	calpain 2, (m/II) large subunit
M23612	RAS p21 protein activator (GTPase activating protein) 1
M24915	CD44 antigen (homing function and Indian blood group system)
M25915	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)
M25915	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)
M27877	zinc finger protein 83 (HPF1)
M28213	RAB2, member RAS oncogene family
M28882	melanoma cell adhesion molecule
M31659	GT mitochondrial solute carrier protein homologue; putative; Human GT mitochondrial solute carrier protein homologue mRNA, complete cds.
M34715	pregnancy specific beta-1-glycoprotein 1
M55580	spermidine/spermine N1-acetyltransferase
M60316	bone morphogenetic protein 7 (osteogenic protein 1)
M63310	annexin A3
M64992	proteasome (prosome, macropain) subunit, alpha type, 1
M65217	heat shock transcription factor 2
M72709	
M74089	hypothetical protein BC017169
M80261	APEX nuclease (multifunctional DNA repair enzyme) 1
M81635	The gene codes for an erythrocyte membrane protein which is deficient in patients with hereditary stomatocytosis. These membranes show very high permeability to Na ⁺ and K ⁺ .; putative; <i>Homo sapiens</i> erythrocyte membrane protein mRNA, complete cds.
M85289	heparan sulfate proteoglycan 2 (perlecan)
M86849	gap junction protein; <i>Homo sapiens</i> connexin 26 (GJB2) mRNA, complete cds.
M87771	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)
M93651	SET translocation (myeloid leukemia-associated)
M97655	6-pyruvoyltetrahydropterin synthase
N21390	src family associated phosphoprotein 2
N21426	yx58d01.s1 Soares melanocyte 2NbHM <i>Homo sapiens</i> cDNA clone IMAGE:265921 3', mRNA sequence.

N21442 protein tyrosine phosphatase, non-receptor type 21
 N22548 Rho-associated, coiled-coil containing protein kinase 1
 N22859 KIAA0934 protein
 N23018 C-terminal binding protein 2
 N25727 serine hydroxymethyltransferase 1 (soluble)
 yx83c03.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone
 N25732 IMAGE:268324 3', mRNA sequence.
 N29918 zinc finger and BTB domain containing 10
 N30878 Trichohyalin (THH), mRNA
 N32035 Clone 24889 mRNA sequence
 N32526 serologically defined colon cancer antigen 8
 N34514 CDNA FLJ41867 fis, clone OCBBF2005546
 N36085 CDNA FLJ31683 fis, clone NT2RI2005353
 N37081 hypothetical protein FLJ20618
 yz17a12.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone
 N45308 IMAGE:283294 3', mRNA sequence.
 N48361 MRNA; cDNA DKFZp564O1016 (from clone DKFZp564O1016)
 N51370 a disintegrin and metalloproteinase domain 10
 N51514 PTD016 protein
 Transcribed sequence with weak similarity to protein ref:NP_060265.1
 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 N51961 translational inhibitor protein p14.5
 N54448 Transcribed sequences
 N63384 Transcribed sequences
 N63936 similar to ubiquitin binding protein
 yz86a05.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone
 N64593 IMAGE:289904 3', mRNA sequence.
 N64643 senataxin
 N64802 nuclear protein UKp68
 yz37h06.s1 Morton Fetal Cochlea Homo sapiens cDNA clone IMAGE:285275
 N66307 3', mRNA sequence.
 yz79b06.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone
 N73682 IMAGE:289235 3', mRNA sequence.
 N74222 zinc finger protein 75a
 N74607 aquaporin 3
 N90779 tumor suppressor TSBF1
 N95414 integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)
 N95443 RAB22A, member RAS oncogene family
 N99438 signal peptidase complex (18kD)
 NM_000019 acetyl-Coenzyme A acetyltransferase 1 (acetoacetyl Coenzyme A thiolase)
 NM_000053 ATPase, Cu⁺⁺ transporting, beta polypeptide (Wilson disease)
 NM_000100 cystatin B (stefin B)
 NM_000107 damage-specific DNA binding protein 2, 48kDa
 NM_000126 electron-transfer-flavoprotein, alpha polypeptide (glutaric aciduria II)
 NM_000127 exostoses (multiple) 1
 NM_000142 fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)
 NM_000152 glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II)
 NM_000153 galactosylceramidase (Krabbe disease)
 NM_000161 GTP cyclohydrolase 1 (dopa-responsive dystonia)
 NM_000165 gap junction protein, alpha 1, 43kDa (connexin 43)
 NM_000177 gelsolin (amyloidosis, Finnish type)

NM_000213 integrin, beta 4
 NM_000259 myosin VA (heavy polypeptide 12, myosin)
 NM_000270 nucleoside phosphorylase
 NM_000271 Niemann-Pick disease, type C1
 NM_000274 ornithine aminotransferase (gyrate atrophy)
 NM_000276 oculocerebrorenal syndrome of Lowe
 NM_000285 peptidase D
 NM_000286 peroxisomal biogenesis factor 12
 NM_000291 phosphoglycerate kinase 1
 NM_000297 polycystic kidney disease 2 (autosomal dominant)
 NM_000308 protective protein for beta-galactosidase (galactosialidosis)
 prion protein (p27-30) (Creutzfeld-Jakob disease, Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia)
 NM_000311
 NM_000328 retinitis pigmentosa GTPase regulator
 tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)
 NM_000362
 NM_000389 cyclin-dependent kinase inhibitor 1A (p21, Cip1)
 NM_000403 galactose-4-epimerase, UDP-
 NM_000404 galactosidase, beta 1
 NM_000416 interferon gamma receptor 1
 NM_000435 Notch homolog 3 (Drosophila)
 galactosamine (N-acetyl)-6-sulfate sulfatase (Morquio syndrome, mucopolysaccharidosis type IVA)
 NM_000512
 NM_000532 propionyl Coenzyme A carboxylase, beta polypeptide
 NM_000574 decay accelerating factor for complement (CD55, Cromer blood group system)
 NM_000593 transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
 CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)
 NM_000611
 solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
 NM_000617
 NM_000627 latent transforming growth factor beta binding protein 1
 amylo-1, 6-glucosidase, 4-alpha-glucanotransferase (glycogen debranching enzyme, glycogen storage disease type III)
 NM_000645
 synonym: AAC1; arylamine N-acetyltransferase-1; arylamide acetylase 1 (N-acetyltransferase 1); go_function: arylamine N-acetyltransferase activity [go:0004060] [evidence E]; Homo sapiens N-acetyltransferase 1 (arylamine N-acetyltransferase) (NAT1), mRNA.
 NM_000662
 NM_000690 aldehyde dehydrogenase 2 family (mitochondrial)
 NM_000693 aldehyde dehydrogenase 1 family, member A3
 NM_000694 aldehyde dehydrogenase 3 family, member B1
 NM_000695 aldehyde dehydrogenase 3 family, member B2
 NM_000696 aldehyde dehydrogenase 9 family, member A1
 NM_000782 cytochrome P450, family 24, subfamily A, polypeptide 1
 NM_000852 glutathione S-transferase pi
 NM_000859 3-hydroxy-3-methylglutaryl-Coenzyme A reductase
 NM_000876 insulin-like growth factor 2 receptor
 NM_000884 IMP (inosine monophosphate) dehydrogenase 2
 NM_000916 oxytocin receptor
 NM_000938 polymerase (RNA) II (DNA directed) polypeptide B, 140kDa
 NM_000942 peptidylprolyl isomerase B (cyclophilin B)
 NM_001068 topoisomerase (DNA) II beta 180kDa

NM_001110 a disintegrin and metalloproteinase domain 10
NM_001124 adrenomedullin
NM_001153 annexin A4
NM_001157 annexin A11
NM_001166 baculoviral IAP repeat-containing 2
NM_001187 B melanoma antigen
NM_001188 BCL2-antagonist/killer 1
NM_001207 basic transcription factor 3
NM_001219 calumenin
NM_001222 calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma
NM_001227 caspase 7, apoptosis-related cysteine protease
NM_001241 cyclin T2
NM_001253 CDC5 cell division cycle 5-like (S. pombe)
NM_001305 claudin 4
NM_001306 claudin 3
NM_001315 mitogen-activated protein kinase 14
NM_001320 casein kinase 2, beta polypeptide
NM_001326 cleavage stimulation factor, 3' pre-RNA, subunit 3, 77kDa
NM_001349 aspartyl-tRNA synthetase
NM_001357 DEAH (Asp-Glu-Ala-His) box polypeptide 9
NM_001418 eukaryotic translation initiation factor 4 gamma, 2
NM_001421 E74-like factor 4 (ets domain transcription factor)
NM_001424 epithelial membrane protein 2
NM_001449 four and a half LIM domains 1
NM_001456 filamin A, alpha (actin binding protein 280)
NM_001483 glioblastoma amplified sequence
NM_001514 general transcription factor IIB
NM_001527 histone deacetylase 2
NM_001539 DnaJ (Hsp40) homolog, subfamily A, member 1
NM_001613 actin, alpha 2, smooth muscle, aorta
NM_001630 annexin A8
NM_001634 adenosylmethionine decarboxylase 1
NM_001642 amyloid beta (A4) precursor-like protein 2
NM_001647 apolipoprotein D
NM_001655 archain 1
NM_001659 ADP-ribosylation factor 3
NM_001660 ADP-ribosylation factor 4
NM_001674 activating transcription factor 3
NM_001695 ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C, isoform 1
NM_001724 2,3-bisphosphoglycerate mutase
NM_001731 B-cell translocation gene 1, anti-proliferative
NM_001743 calmodulin 2 (phosphorylase kinase, delta)
NM_001752 catalase
NM_001769 CD9 antigen (p24)
NM_001786 cell division cycle 2, G1 to S and G2 to M
NM_001823 creatine kinase, brain
NM_001829 chloride channel 3
NM_001833 clathrin, light polypeptide (Lca)
NM_001834 clathrin, light polypeptide (Lcb)
NM_001873 carboxypeptidase E

NM_001894 casein kinase 1, epsilon
 NM_001899 cystatin S
 NM_001908 cathepsin B
 NM_001912 cathepsin L
 NM_001913 cut-like 1, CCAAT displacement protein (Drosophila)
 NM_001924 growth arrest and DNA-damage-inducible, alpha
 NM_001938 down-regulator of transcription 1, TBP-binding (negative cofactor 2)
 NM_001969 eukaryotic translation initiation factor 5
 NM_001981 epidermal growth factor receptor pathway substrate 15
 NM_002041 GA binding protein transcription factor, beta subunit 2, 47kDa
 NM_002056 glutamine-fructose-6-phosphate transaminase 1
 NM_002064 glutaredoxin (thioltransferase)
 NM_002077 golgi autoantigen, golgin subfamily a, 1
 NM_002078 golgi autoantigen, golgin subfamily a, 4
 NM_002087 granulin
 NM_002094 G1 to S phase transition 1
 NM_002095 general transcription factor IIE, polypeptide 2, beta 34kDa
 NM_002106 H2A histone family, member Z
 NM_002114 human immunodeficiency virus type I enhancer binding protein 1
 NM_002194 inositol polyphosphate-1-phosphatase
 NM_002205 integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
 NM_002229 jun B proto-oncogene
 NM_002245 potassium channel, subfamily K, member 1
 potassium intermediate/small conductance calcium-activated channel,
 subfamily N, member 4
 NM_002250
 NM_002281 keratin, hair, basic, 1
 NM_002291 laminin, beta 1
 NM_002318 lysyl oxidase-like 2
 NM_002356 myristoylated alanine-rich protein kinase C substrate
 NM_002364 melanoma antigen, family B, 2
 NM_002372 mannosidase, alpha, class 2A, member 1
 NM_002376 MAP/microtubule affinity-regulating kinase 3
 NM_002380 matrilin 2
 NM_002381 matrilin 3
 NM_002388 MCM3 minichromosome maintenance deficient 3 (S. cerevisiae)
 NM_002408 mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
 NM_002434 N-methylpurine-DNA glycosylase
 NM_002443 microseminoprotein, beta-
 NM_002463 myxovirus (influenza virus) resistance 2 (mouse)
 NM_002526 5'-nucleotidase, ecto (CD73)
 NM_002539 ornithine decarboxylase 1
 NM_002641 phosphatidylinositol glycan, class A (paroxysmal nocturnal hemoglobinuria)
 NM_002647 phosphoinositide-3-kinase, class 3
 NM_002669 pleiotropic regulator 1 (PRL1 homolog, Arabidopsis)
 NM_002710 protein phosphatase 1, catalytic subunit, gamma isoform
 protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta
 isoform
 NM_002716
 protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific
 extinguisher 1)
 NM_002734
 NM_002748 mitogen-activated protein kinase 6
 NM_002773 protease, serine, 8 (prostatic)

NM_002784 pregnancy specific beta-1-glycoprotein 9
 NM_002786 proteasome (prosome, macropain) subunit, alpha type, 1
 NM_002787 proteasome (prosome, macropain) subunit, alpha type, 2
 NM_002789 proteasome (prosome, macropain) subunit, alpha type, 4
 NM_002790 proteasome (prosome, macropain) subunit, alpha type, 5
 NM_002793 proteasome (prosome, macropain) subunit, beta type, 1
 NM_002794 proteasome (prosome, macropain) subunit, beta type, 2
 NM_002796 proteasome (prosome, macropain) subunit, beta type, 4
 proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional
 protease 2)
 NM_002800
 NM_002829 protein tyrosine phosphatase, non-receptor type 3
 NM_002835 protein tyrosine phosphatase, non-receptor type 12
 NM_002858 ATP-binding cassette, sub-family D (ALD), member 3
 NM_002859 Paxillin (PXN), mRNA
 phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type
 VI)
 NM_002863
 NM_002887 arginyl-tRNA synthetase
 NM_002896 RNA binding motif protein 4
 NM_002902 reticulocalbin 2, EF-hand calcium binding domain
 NM_002925 regulator of G-protein signalling 10
 restin (Reed-Steinberg cell-expressed intermediate filament-associated
 protein)
 NM_002956
 NM_002958 RYK receptor-like tyrosine kinase
 NM_002964 S100 calcium binding protein A8 (calgranulin A)
 NM_002965 S100 calcium binding protein A9 (calgranulin B)
 NM_002967 scaffold attachment factor B
 spinocerebellar ataxia 2 (olivopontocerebellar ataxia 2, autosomal dominant,
 ataxin 2)
 NM_002973
 NM_002997 syndecan 1
 NM_003009 selenoprotein W, 1
 NM_003010 mitogen-activated protein kinase kinase 4
 NM_003011 SET translocation (myeloid leukemia-associated)
 NM_003022 SH3 domain binding glutamic acid-rich protein like
 SWI/SNF related, matrix associated, actin dependent regulator of chromatin,
 subfamily a, member 1
 NM_003069
 SWI/SNF related, matrix associated, actin dependent regulator of chromatin,
 subfamily a, member 1
 NM_003069
 NM_003092 small nuclear ribonucleoprotein polypeptide B"
 NM_003097 small nuclear ribonucleoprotein polypeptide N
 NM_003103 SON DNA binding protein
 NM_003105 sortilin-related receptor, L(DLR class) A repeats-containing
 NM_003119 spastic paraplegia 7, paraplegin (pure and complicated autosomal recessive)
 NM_003135 signal recognition particle 19kDa
 NM_003136 signal recognition particle 54kDa
 NM_003138 synonym: SFRSK2; isoform b is encoded by transcript variant 2;
 NM_003139 signal recognition particle receptor ('docking protein')
 NM_003142 Sjogren syndrome antigen B (autoantigen La)
 NM_003161 ribosomal protein S6 kinase, 70kDa, polypeptide 1
 NM_003169 suppressor of Ty 5 homolog (*S. cerevisiae*)
 NM_003191 synonyms: ThrRS, MGC9344; threonine--tRNA ligase;.

NM_003194 TATA box binding protein
 NM_003201 transcription factor A, mitochondrial
 NM_003217 testis enhanced gene transcript (BAX inhibitor 1)
 NM_003234 transferrin receptor (p90, CD71)
 tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)
 NM_003254 tissue inhibitor of metalloproteinase 2
 NM_003257 tight junction protein 1 (zona occludens 1)
 NM_003288 tumor protein D52-like 2
 NM_003297 nuclear receptor subfamily 2, group C, member 1
 NM_003299 tumor rejection antigen (gp96) 1
 NM_003304 transient receptor potential cation channel, subfamily C, member 1
 NM_003330 thioredoxin reductase 1
 ubiquitin-activating enzyme E1 (A1S9T and BN75 temperature sensitivity complementing)
 NM_003334
 NM_003336 ubiquitin-conjugating enzyme E2A (RAD6 homolog)
 NM_003366 ubiquinol-cytochrome c reductase core protein II
 tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
 NM_003406
 NM_003440 zinc finger protein 140 (clone pHZ-39)
 NM_003452 zinc finger protein 189
 NM_003469 secretogranin II (chromogranin C)
 NM_003470 ubiquitin specific protease 7 (herpes virus-associated)
 NM_003473 signal transducing adaptor molecule (SH3 domain and ITAM motif) 1
 NM_003489 nuclear receptor interacting protein 1
 NM_003496 transformation/transcription domain-associated protein
 NM_003516 histone 2, H2aa
 NM_003522 histone 1, H2bg
 NM_003526 histone 1, H2bc
 NM_003528 histone 2, H2be
 NM_003536 histone 1, H3h
 NM_003567 breast cancer anti-estrogen resistance 3
 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5
 NM_003601
 NM_003618 mitogen-activated protein kinase kinase kinase kinase 3
 NM_003620 protein phosphatase 1D magnesium-dependent, delta isoform
 NM_003628 plakophilin 4
 NM_003645 solute carrier family 27 (fatty acid transporter), member 2
 NM_003645 solute carrier family 27 (fatty acid transporter), member 2
 NM_003648 diacylglycerol kinase, delta 130kDa
 NM_003664 adaptor-related protein complex 3, beta 1 subunit
 NM_003666 basic leucine zipper nuclear factor 1 (JEM-1)
 NM_003668 mitogen-activated protein kinase-activated protein kinase 5
 NM_003715 vesicle docking protein p115
 NM_003729 RNA terminal phosphate cyclase domain 1
 NM_003744 numb homolog (Drosophila)
 NM_003746 dynein, cytoplasmic, light polypeptide 1
 NM_003750 eukaryotic translation initiation factor 3, subunit 10 theta, 150/170kDa
 NM_003761 vesicle-associated membrane protein 5 (myobrevin)
 NM_003791 membrane-bound transcription factor protease, site 1

NM_003798 catenin (cadherin-associated protein), alpha-like 1
 NM_003816 a disintegrin and metalloproteinase domain 9 (meltrin gamma)
 NM_003819 poly(A) binding protein, cytoplasmic 4 (inducible form)
 NM_003870 IQ motif containing GTPase activating protein 1
 NM_003887 development and differentiation enhancing factor 2
 NM_003899 Rho guanine nucleotide exchange factor (GEF) 7
 NM_003900 sequestosome 1
 NM_003903 CDC16 cell division cycle 16 homolog (*S. cerevisiae*)
 NM_003909 copine III
 NM_003915 copine I
 NM_003922 hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1
 (CHC1)-like domain (RLD) 1
 NM_003928 CAAX box 1
 NM_003937 kynureninase (L-kynurenine hydrolase)
 NM_003942 ribosomal protein S6 kinase, 90kDa, polypeptide 4
 NM_003979 retinoic acid induced 3
 NM_004068 adaptor-related protein complex 2, mu 1 subunit
 NM_004078 cysteine and glycine-rich protein 1
 NM_004105 EGF-containing fibulin-like extracellular matrix protein 1
 NM_004109 ferredoxin 1
 NM_004116 FK506 binding protein 1B, 12.6 kDa
 NM_004124 glia maturation factor, beta
 NM_004156 protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform
 NM_004161 RAB1A, member RAS oncogene family
 NM_004163 RAB27B, member RAS oncogene family
 NM_004175 small nuclear ribonucleoprotein D3 polypeptide 18kDa
 NM_004199 procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase),
 alpha polypeptide II
 NM_004209 synaptogyrin 3
 NM_004215 estrogen receptor binding site associated, antigen, 9
 NM_004236 thyroid receptor interacting protein 15
 NM_004238
 NM_004251 RAB9A, member RAS oncogene family
 NM_004289 nuclear factor (erythroid-derived 2)-like 3
 NM_004300 acid phosphatase 1, soluble
 NM_004317 arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)
 NM_004339 pituitary tumor-transforming 1 interacting protein
 NM_004356 CD81 antigen (target of antiproliferative antibody 1)
 NM_004360 cadherin 1, type 1, E-cadherin (epithelial)
 NM_004362 calmegin
 NM_004363 carcinoembryonic antigen-related cell adhesion molecule 5
 NM_004414 Down syndrome critical region gene 1
 NM_004415 desmoplakin
 NM_004423 dishevelled, dsh homolog 3 (*Drosophila*)
 NM_004427 polyhomeotic-like 2 (*Drosophila*)
 NM_004428 ephrin-A1
 NM_004446 glutamyl-prolyl-tRNA synthetase
 NM_004453 electron-transferring-flavoprotein dehydrogenase
 NM_004457 acyl-CoA synthetase long-chain family member 3

NM_004481 UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)
 NM_004492 general transcription factor IIA, 2, 12kDa
 NM_004496 forkhead box A1
 NM_004504 HIV-1 Rev binding protein
 NM_004515 interleukin enhancer binding factor 2, 45kDa
 NM_004516 interleukin enhancer binding factor 3, 90kDa
 NM_004517 integrin-linked kinase
 NM_004520 kinesin heavy chain member 2
 NM_004524 lethal giant larvae homolog 2 (Drosophila)
 NM_004544 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa
 NM_004595 spermine synthase
 NM_004630 splicing factor 1
 NM_004639 HLA-B associated transcript 3
 NM_004642 CDK2-associated protein 1
 NM_004663 RAB11A, member RAS oncogene family
 NM_004669 chloride intracellular channel 3
 NM_004688 N-myc (and STAT) interactor
 NM_004710 synaptogyrin 2
 NM_004725 BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)
 NM_004730 eukaryotic translation termination factor 1
 NM_004733 solute carrier family 33 (acetyl-CoA transporter), member 1
 NM_004735 leucine rich repeat (in FLII) interacting protein 1
 NM_004755 ribosomal protein S6 kinase, 90kDa, polypeptide 5
 NM_004757 small inducible cytokine subfamily E, member 1 (endothelial monocyte-activating)
 NM_004763 integrin beta 1 binding protein 1
 NM_004768 splicing factor, arginine/serine-rich 11
 NM_004782 synaptosomal-associated protein, 29kDa
 NM_004789 LIM homeobox 2
 NM_004798 kinesin family member 3B
 NM_004800 transmembrane 9 superfamily member 2
 NM_004815 PTPL1-associated RhoGAP 1
 NM_004817 tight junction protein 2 (zona occludens 2)
 NM_004830 cofactor required for Sp1 transcriptional activation, subunit 3, 130kDa
 NM_004834 mitogen-activated protein kinase kinase kinase 4
 NM_004836 eukaryotic translation initiation factor 2-alpha kinase 3
 NM_004856 kinesin family member 23
 NM_004872 chromosome 1 open reading frame 8
 NM_004893 H2A histone family, member Y
 NM_004896 vacuolar protein sorting 26 (yeast)
 NM_004915 ATP-binding cassette, sub-family G (WHITE), member 1
 NM_004949 desmocollin 2
 NM_004993 Machado-Joseph disease (spinocerebellar ataxia 3, olivopontocerebellar ataxia 3, autosomal dominant, ataxin 3)
 NM_004999
 NM_005008 NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)
 NM_005010 neuronal cell adhesion molecule
 NM_005022 profilin 1
 NM_005032 plastin 3 (T isoform)
 NM_005034 polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa

NM_005051 glutaminyl-tRNA synthetase
 NM_005056 Jumonji, AT rich interactive domain 1A (RBBP2-like)
 splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)
 NM_005066 synonyms: XE7, XE7Y, MGC39904; isoform 1 is encoded by transcript variant 1; Pseudoautosomal gene XE7 (Y chromosome);
 NM_005088 interferon, alpha-inducible protein (clone IFI-15K)
 NM_005101 golgi autoantigen, golgin subfamily a, 5
 NM_005113 golgi autoantigen, golgin subfamily a, 5
 NM_005132 REC8-like 1 (yeast)
 NM_005154 ubiquitin specific protease 8
 NM_005163 v-akt murine thymoma viral oncogene homolog 1
 NM_005173 ATPase, Ca⁺⁺ transporting, ubiquitous
 NM_005180 COMM domain containing 3
 NM_005204 mitogen-activated protein kinase kinase kinase 8
 NM_005224 AT rich interactive domain 3A (BRIGHT- like)
 ems1 sequence (mammary tumor and squamous cell carcinoma-associated (p80/85 src substrate)
 NM_005231 GA binding protein transcription factor, beta subunit 2, 47kDa
 NM_005254 guanine nucleotide binding protein (G protein), beta polypeptide 2
 NM_005273 H3 histone, family 3B (H3.3B)
 NM_005324 H3 histone, family 3B (H3.3B)
 NM_005333 holocytochrome c synthase (cytochrome c heme-lyase)
 NM_005336 high density lipoprotein binding protein (vigilin)
 NM_005342 high-mobility group box 3
 NM_005345 heat shock 70kDa protein 1A
 NM_005397 podocalyxin-like
 NM_005426 tumor protein p53 binding protein, 2
 NM_005433 v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1
 NM_005471 glucosamine-6-phosphate deaminase 1
 NM_005475 lymphocyte adaptor protein
 NM_005496 SMC4 structural maintenance of chromosomes 4-like 1 (yeast)
 NM_005513 general transcription factor IIE, polypeptide 1, alpha 56kDa
 NM_005516 major histocompatibility complex, class I, E
 NM_005561 lysosomal-associated membrane protein 1
 NM_005567 lectin, galactoside-binding, soluble, 3 binding protein
 NM_005614 Ras homolog enriched in brain
 NM_005620 S100 calcium binding protein A11 (calgizzarin)
 NM_005621 S100 calcium binding protein A12 (calgranulin C)
 NM_005626 splicing factor, arginine/serine-rich 4
 NM_005629 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
 NM_005638 synaptobrevin-like 1
 TAF7 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 55kDa
 NM_005642 zinc finger protein 354A
 NM_005649 ring finger protein 103
 TATA box binding protein (TBP)-associated factor, RNA polymerase I, A, 48kDa
 NM_005681 ATP-binding cassette, sub-family B (MDR/TAP), member 6
 NM_005703 polyglutamine binding protein 1
 NM_005710 thioredoxin domain containing 7 (protein disulfide isomerase)
 NM_005742 thioredoxin domain containing 7 (protein disulfide isomerase)
 NM_005744 ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1

(Drosophila)

NM_005745 B-cell receptor-associated protein 31

NM_005765 ATPase, H⁺ transporting, lysosomal accessory protein 2

NM_005778 RNA binding motif protein 5

NM_005794 dehydrogenase/reductase (SDR family) member 2

NM_005827 solute carrier family 35, member B1

NM_005839 serine/arginine repetitive matrix 1

NM_005851 tumor suppressor deleted in oral cancer-related 1

NM_005875 translation factor sui1 homolog

NM_005877 splicing factor 3a, subunit 1, 120kDa

NM_005882 macrophage erythroblast attacher

NM_005895 golgi autoantigen, golgin subfamily a, 3

NM_005896 isocitrate dehydrogenase 1 (NADP⁺), soluble
membrane component, chromosome 17, surface marker 2 (ovarian carcinoma antigen CA125)

NM_005899 antigen CA125)

NM_005908 mannosidase, beta A, lysosomal

NM_005917 malate dehydrogenase 1, NAD (soluble)
myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila);
translocated to, 2

NM_005935

NM_005977 ring finger protein (C3H2C3 type) 6

NM_006002 ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)

NM_006016 CD164 antigen, sialomucin

NM_006024 Tax1 (human T-cell leukemia virus type I) binding protein 1

NM_006029 paraneoplastic antigen MA1

NM_006035 CDC42 binding protein kinase beta (DMPK-like)

NM_006044 histone deacetylase 6

NM_006070 TRK-fused gene
Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal
domain, 2

NM_006079

NM_006096 N-myc downstream regulated gene 1

NM_006103 WAP four-disulfide core domain 2

NM_006117 peroxisomal D3,D2-enoyl-CoA isomerase

NM_006122 mannosidase, alpha, class 2A, member 2

NM_006134 chromosome 21 open reading frame 4

NM_006145 DnaJ (Hsp40) homolog, subfamily B, member 1

NM_006148 LIM and SH3 protein 1

NM_006156 neural precursor cell expressed, developmentally down-regulated 8

NM_006166 nuclear transcription factor Y, beta

NM_006197 pericentriolar material 1

NM_006224 phosphatidylinositol transfer protein

NM_006241 protein phosphatase 1, regulatory (inhibitor) subunit 2

NM_006246 protein phosphatase 2, regulatory subunit B (B56), epsilon isoform

NM_006263 proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)

NM_006265 RAD21 homolog (S. pombe)

NM_006276 splicing factor, arginine/serine-rich 7, 35kDa

NM_006278 sialyltransferase 4C (beta-galactoside alpha-2,3-sialyltransferase)

NM_006283 transforming, acidic coiled-coil containing protein 1

NM_006291 tumor necrosis factor, alpha-induced protein 2

NM_006292 tumor susceptibility gene 101

NM_006303 JTV1 gene

NM_006309 leucine rich repeat (in FLII) interacting protein 2

NM_006311 nuclear receptor co-repressor 1
NM_006317 brain abundant, membrane attached signal protein 1
NM_006323 SEC24 related gene family, member B (*S. cerevisiae*)
NM_006324 craniofacial development protein 1
NM_006327 translocase of inner mitochondrial membrane 23 homolog (yeast)
NM_006332 interferon, gamma-inducible protein 30
NM_006333 nuclear DNA-binding protein
NM_006349 zinc finger, HIT domain containing 1
NM_006353 high mobility group nucleosomal binding domain 4
NM_006359 solute carrier family 9 (sodium/hydrogen exchanger), isoform 6
NM_006363 Sec23 homolog B (*S. cerevisiae*)
NM_006365 transcriptional activator of the c-fos promoter
NM_006367 CAP, adenylate cyclase-associated protein 1 (yeast)
NM_006380 amyloid beta precursor protein (cytoplasmic tail) binding protein 2
NM_006393 nebullette
NM_006402 hepatitis B virus x interacting protein
NM_006404 protein C receptor, endothelial (EPCR)
NM_006409 actin related protein 2/3 complex, subunit 1A, 41kDa
NM_006416 chromosome 6 open reading frame 165
NM_006421 brefeldin A-inhibited guanine nucleotide-exchange protein 1
NM_006429 chaperonin containing TCP1, subunit 7 (eta)
NM_006432 Niemann-Pick disease, type C2
NM_006437 ADP-ribosyltransferase (NAD⁺; poly (ADP-ribose) polymerase)-like 1
NM_006447 ubiquitin specific protease 16
NM_006462 chromosome 20 open reading frame 18
NM_006466 polymerase (RNA) III (DNA directed) polypeptide F, 39 kDa
NM_006471 myosin regulatory light chain MRCL3
NM_006472 thioredoxin interacting protein
NM_006493 ceroid-lipofuscinosis, neuronal 5
NM_006526 zinc finger protein 217
NM_006527 stem-loop (histone) binding protein
NM_006531 Probe hTg737 (polycystic kidney disease, autosomal recessive)
NM_006554 metaxin 2
NM_006570 Ras-related GTP binding A
NM_006575 mitogen-activated protein kinase kinase kinase kinase 5
NM_006595 apoptosis inhibitor 5
NM_006598 solute carrier family 12 (potassium/chloride transporters), member 7
NM_006608 putative homeodomain transcription factor 1
NM_006618 Jumonji, AT rich interactive domain 1B (RBP2-like)
NM_006621 S-adenosylhomocysteine hydrolase-like 1
NM_006622 polo-like kinase 2 (*Drosophila*)
NM_006625 FUS interacting protein (serine-arginine rich) 1
NM_006628 cyclic AMP phosphoprotein, 19 kD
NM_006667 progesterone receptor membrane component 1
NM_006698 bladder cancer associated protein
NM_006706 transcription elongation regulator 1 (CA150)
NM_006720 actin binding LIM protein 1
NM_006743 RNA binding motif protein 3
NM_006759 UDP-glucose pyrophosphorylase 2
NM_006763 BTG family, member 2

NM_006788 ralA binding protein 1
 NM_006791 mortality factor 4 like 1
 NM_006805 heterogeneous nuclear ribonucleoprotein A0
 NM_006808 Sec61 beta subunit
 NM_006823 protein kinase (cAMP-dependent, catalytic) inhibitor alpha
 NM_006825 cytoskeleton-associated protein 4
 NM_006827 transmembrane trafficking protein
 NM_006829 adipose specific 2
 NM_006855 KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 3
 NM_006915 retinitis pigmentosa 2 (X-linked recessive)
 NM_006930 S-phase kinase-associated protein 1A (p19A)
 NM_006952 uroplakin 1B
 NM_006999 polymerase (DNA directed) sigma
 NM_007019 ubiquitin-conjugating enzyme E2C
 NM_007063 TBC1 domain family, member 8 (with GRAM domain)
 NM_007081 RAB, member of RAS oncogene family-like 2B
 NM_007085 follistatin-like 1
 NM_007097 clathrin, light polypeptide (Lcb)
 NM_007158 NRAS-related gene
 NM_007192 suppressor of Ty 16 homolog (*S. cerevisiae*)
 NM_007214 SEC63-like (*S. cerevisiae*)
 NM_007216 Hermansky-Pudlak syndrome 5
 NM_007229 protein kinase C and casein kinase substrate in neurons 2
 NM_007282 ring finger protein 13
 NM_007286 synaptopodin
 NM_007318 presenilin 1 (Alzheimer disease 3)
 NM_007362 nuclear cap binding protein subunit 2, 20kDa
 NM_007372 DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
 NM_012062 dynamin 1-like
 NM_012120 CD2-associated protein
 NM_012121 CDC42 effector protein (Rho GTPase binding) 4
 NM_012124 cysteine and histidine-rich domain (CHORD)-containing, zinc binding protein 1
 NM_012155 echinoderm microtubule associated protein like 2
 NM_012164 F-box and WD-40 domain protein 2
 NM_012168 F-box only protein 2
 NM_012175 F-box only protein 3
 NM_012177 F-box only protein 5
 NM_012214 mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme A
 NM_012215 meningioma expressed antigen 5 (hyaluronidase)
 NM_012238 sirtuin (silent mating type information regulation 2 homolog) 1 (*S. cerevisiae*)
 NM_012249 ras homolog gene family, member Q
 NM_012290 tousled-like kinase 1
 NM_012325 microtubule-associated protein, RP/EB family, member 1
 NM_012341 GTP binding protein 4
 NM_012392 PEF protein with a long N-terminal hydrophobic domain (peflin)
 NM_012413 glutaminyl-peptide cyclotransferase (glutaminyl cyclase)
 NM_012415 RAD54 homolog B (*S. cerevisiae*)
 NM_012432 SET domain, bifurcated 1
 NM_012433 splicing factor 3b, subunit 1, 155kDa

NM_012446 single-stranded DNA binding protein 2
 NM_013229 apoptotic protease activating factor
 NM_013232 programmed cell death 6
 NM_013238 DnaJ (Hsp40) homolog, subfamily D, member 1
 NM_013254 TANK-binding kinase 1
 NM_013283 methionine adenosyltransferase II, beta
 NM_013314 B-cell linker
 NM_013315 transmembrane phosphatase with tensin homology
 NM_013352 squamous cell carcinoma antigen recognized by T cells 2
 NM_013360 zinc finger protein 222
 NM_013374 programmed cell death 6 interacting protein
 NM_013392 nuclear receptor binding protein
 NM_013438 ubiquitin 1
 NM_013446 makorin, ring finger protein, 1
 NM_013448 bromodomain adjacent to zinc finger domain, 1A
 NM_013451 fer-1-like 3, myoferlin (*C. elegans*)
 NM_013989 deiodinase, iodothyronine, type II
 NM_013995 lysosomal-associated membrane protein 2
 NM_014000 vinculin
 NM_014017 mitogen-activated protein-binding protein-interacting protein
 NM_014028 osteopetrosis associated transmembrane protein 1
 NM_014031 solute carrier family 27 (fatty acid transporter), member 6
 NM_014038 basic leucine zipper and W2 domains 2
 NM_014041 signal peptidase 12kDa
 NM_014044 unc-50 homolog (*C. elegans*)
 NM_014048 myocardin-related transcription factor B
 NM_014153 zinc-finger protein AY163807
 NM_014161 mitochondrial ribosomal protein L18
 NM_014165 chromosome 6 open reading frame 66
 NM_014175 mitochondrial ribosomal protein L15
 NM_014183 dynein, cytoplasmic, light polypeptide 2A
 NM_014239 eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa
 protein tyrosine phosphatase-like (proline instead of catalytic arginine),
 member a
 NM_014241
 NM_014247
 NM_014254 transmembrane protein 5
 NM_014258 synaptonemal complex protein 2
 NM_014294 translocation associated membrane protein 1
 quinolinate phosphoribosyltransferase (nicotinate-nucleotide
 pyrophosphorylase (carboxylating))
 NM_014298
 NM_014315 kelch domain containing 2
 NM_014322 opsin 3 (encephalopsin, panopsin)
 NM_014333 immunoglobulin superfamily, member 4
 NM_014345 zinc finger protein 318
 NM_014350 tumor necrosis factor, alpha-induced protein 8
 NM_014366 nucleostemin
 NM_014392 DNA segment on chromosome 4 (unique) 234 expressed sequence
 NM_014400 GPI-anchored metastasis-associated protein homolog
 NM_014413 heme-regulated initiation factor 2-alpha kinase
 NM_014415 zinc finger protein
 NM_014453 putative breast adenocarcinoma marker (32kD)

NM_014574 striatin, calmodulin binding protein 3
 NM_014577 bromodomain containing 1
 NM_014635 GRIP and coiled-coil domain containing 2
 NM_014639 KIAA0372
 NM_014666 enthoprotin
 NM_014670 basic leucine zipper and W2 domains 1
 NM_014676 pumilio homolog 1 (Drosophila)
 NM_014704
 NM_014708 kinetochore associated 1
 NM_014732
 NM_014734 KIAA0247
 NM_014735 PHD finger protein 16
 NM_014745 KIAA0233 gene product
 NM_014751 metastasis suppressor 1
 NM_014757 mastermind-like 1 (Drosophila)
 NM_014762 24-dehydrocholesterol reductase
 NM_014764 DAZ associated protein 2
 NM_014766 secernin 1
 NM_014767 synonym: testican-2;
 NM_014774
 go_function: transcription factor activity [goid 0003700] [evidence IEA];
 go_process: regulation of transcription, DNA-dependent [goid 0006355]
 NM_014779 [evidence IEA]; Homo sapiens KIAA0669 gene product (KIAA0669), mRNA.
 NM_014789
 NM_014803
 NM_014805 EPM2A (laforin) interacting protein 1
 NM_014822 SEC24 related gene family, member D (S. cerevisiae)
 go_function: nucleic acid binding [goid 0003676] [evidence IEA]; Homo sapiens
 NM_014827 KIAA0663 gene product (KIAA0663), mRNA.
 NM_014837 chromosome 1 open reading frame 16
 NM_014840
 NM_014846 KIAA0196 gene product
 NM_014864 family with sequence similarity 20, member B
 NM_014868 ring finger protein 10
 NM_014873
 NM_014885 anaphase-promoting complex subunit 10
 NM_014888 family with sequence similarity 3, member C
 NM_014890 downregulated in ovarian cancer 1
 NM_014904 KIAA0941 protein
 NM_014905 glutaminase
 NM_014914 centaurin, gamma 2
 NM_014921 latrophilin 1
 NM_014923 KIAA0970 protein
 NM_014925 synonym: PR01365; Homo sapiens KIAA1002 protein (KIAA1002), mRNA.
 NM_014933 SEC31-like 1 (S. cerevisiae)
 NM_014937 inositol polyphosphate-5-phosphatase F
 NM_014938 Mlx interactor
 NM_014939 KIAA1012
 NM_014947
 NM_014949

NM_014962 BTB (POZ) domain containing 3
NM_014968
NM_014969 KIAA0893 protein
NM_014977 apoptotic chromatin condensation inducer in the nucleus
NM_015072 KIAA0998
NM_015251
NM_015271 tripartite motif-containing 2
NM_015310 ADP-ribosylation factor guanine nucleotide factor 6
NM_015339 activity-dependent neuroprotector
NM_015362 S-phase 2 protein
NM_015368 pannexin 1
NM_015369 TP53TG3 protein
NM_015384 IDN3 protein
NM_015400 DKFZP586N0721 protein
NM_015485 RWD domain containing 3
NM_015523 small fragment nuclease
NM_015542 UPF2 regulator of nonsense transcripts homolog (yeast)
NM_015610 DKFZP434J154 protein
NM_015626 WD repeat and SOCS box-containing 1
NM_015874 recombining binding protein suppressor of hairless (Drosophila)
NM_015878 ornithine decarboxylase antizyme inhibitor
NM_015904 eukaryotic translation initiation factor 5B
NM_015907 leucine aminopeptidase 3
NM_015925 liver-specific bHLH-Zip transcription factor
NM_015928 androgen-induced proliferation inhibitor
NM_015946 pelota homolog (Drosophila)
NM_015959 thioredoxin-related transmembrane protein 2
NM_015976 sorting nexin 7
NM_015987 heme binding protein 1
NM_015993 transmembrane 4 superfamily member 11 (plasmolipin)
NM_016002 CGI-49 protein
NM_016003 DKFZP434J154 protein
NM_016008 dynein 2 light intermediate chain
NM_016056 CGI-119 protein
NM_016077 CGI-147 protein
NM_016078 family with sequence similarity 18, member B
NM_016096 HSPC038 protein
NM_016099 golgi autoantigen, golgin subfamily a, 7
NM_016121 potassium channel tetramerisation domain containing 3
NM_016126 HSPCO34 protein
NM_016141 dynein, cytoplasmic, light intermediate polypeptide 1
NM_016185 hematological and neurological expressed 1
NM_016201 angiomin like 2
NM_016217 headcase homolog (Drosophila)
NM_016227 chromosome 1 open reading frame 9
NM_016245 dehydrogenase/reductase (SDR family) member 8
NM_016261 likely ortholog of mouse tubulin, delta 1
NM_016271 ring finger protein 138
NM_016281 STE20-like kinase
NM_016289 MO25 protein

NM_016290 receptor associated protein 80
NM_016291 inositol hexaphosphate kinase 2
NM_016303 pp21 homolog
NM_016304 chromosome 15 open reading frame 15
NM_016308 UMP-CMP kinase
NM_016315 GULP, engulfment adaptor PTB domain containing 1
NM_016316 REV1-like (yeast)
NM_016357 epithelial protein lost in neoplasm beta
NM_016365
NM_016374 AT rich interactive domain 4B (RBP1- like)
NM_016397 TH1-like (Drosophila)
NM_016401 hypothetical protein HSPC138
NM_016406 hypothetical protein HSPC155
NM_016410 chromosome 9 open reading frame 83
NM_016441 cysteine-rich motor neuron 1
NM_016451 coatomer protein complex, subunit beta
NM_016570 PTX1 protein
NM_016598 zinc finger, DHHC domain containing 3
NM_016603 chromosome 5 open reading frame 5
NM_016614 TRAF and TNF receptor associated protein
NM_016617 hypothetical protein BM-002
NM_016627 hypothetical protein LOC51321
NM_016641 membrane interacting protein of RGS16
NM_017455 stromal cell derived factor receptor 1
NM_017491 WD repeat domain 1
NM_017512 rTS beta protein
NM_017515 solute carrier family 35, member F2
NM_017528 Williams Beuren syndrome chromosome region 22
NM_017540 UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNAc-T10)
NM_017572 MAP kinase-interacting serine/threonine kinase 2
NM_017582 ubiquitin-conjugating enzyme E2Q (putative)
NM_017612 zinc finger, CCHC domain containing 8
NM_017617 Notch homolog 1, translocation-associated (Drosophila)
NM_017623 cyclin M3
NM_017634 potassium channel tetramerisation domain containing 9
NM_017657 aftiphilin protein
NM_017660 p66 alpha
NM_017696
NM_017697 hypothetical protein FLJ20171
NM_017724 leucine rich repeat (in FLII) interacting protein 2
NM_017733 hypothetical protein FLJ20265
NM_017740 zinc finger, DHHC domain containing 7
NM_017742 zinc finger, CCHC domain containing 2
NM_017743 dipeptidylpeptidase 8
NM_017745 BCL6 co-repressor
NM_017750 hypothetical protein FLJ20296
NM_017775 hypothetical protein FLJ20343
NM_017780
NM_017782

NM_017792 hypothetical protein FLJ20373
NM_017832 hypothetical protein FLJ20457
NM_017835 chromosome 21 open reading frame 59
NM_017859 uridine kinase-like 1
NM_017866 hypothetical protein FLJ20533
NM_017892
NM_017897 hypothetical protein FLJ20604
NM_017925 chromosome 9 open reading frame 55
NM_017927 mitofusin 1
NM_017939 hypothetical protein FLJ20718
NM_017945 solute carrier family 35, member A5
NM_017966
NM_017983 hypothetical protein FLJ10055
NM_017994 hypothetical protein FLJ10099
NM_018009 TAP binding protein related
NM_018011 hypothetical protein FLJ10154
NM_018036 chromosome 14 open reading frame 103
NM_018045 hypothetical protein FLJ10276
NM_018051 hypothetical protein FLJ10300
NM_018069
NM_018072
NM_018087 hypothetical protein FLJ10407
NM_018095 kelch repeat and BTB (POZ) domain containing 4
NM_018107 RNA-binding region (RNP1, RRM) containing 4
NM_018108 chromosome 14 open reading frame 130
NM_018115 hypothetical protein FLJ10498
NM_018117 WD repeat domain 11
NM_018128 hypothetical protein FLJ10534
NM_018144 Sec61 alpha 2 subunit (*S. cerevisiae*)
NM_018147 Fas apoptotic inhibitory molecule
NM_018149 hypothetical protein FLJ10587
NM_018171 DIP13 beta
NM_018179 activating transcription factor 7 interacting protein
NM_018184 ADP-ribosylation factor-like 10C
NM_018200 high-mobility group 20A
NM_018214 leucine rich repeat containing 1
NM_018265 hypothetical protein FLJ10901
NM_018281 hypothetical protein FLJ10948
NM_018287 Rho GTPase activating protein 12
NM_018288 PHD finger protein 10
NM_018339 riboflavin kinase
NM_018362 lin-7 homolog C (*C. elegans*)
NM_018368 chromosome 6 open reading frame 209
NM_018370 hypothetical protein FLJ11259
NM_018372 receptor-interacting factor 1
NM_018390 hypothetical protein FLJ11323
NM_018407 lysosomal associated protein transmembrane 4 beta
NM_018428 hepatocellular carcinoma-associated antigen 66
NM_018434 ring finger protein 130
NM_018439 hypothetical protein IMPACT

NM_018443 zinc finger protein 302
 NM_018446 glycosyltransferase AD-017
 NM_018448 TBP-interacting protein
 NM_018458 synonym: BM042; Homo sapiens KIAA1280 protein (KIAA1280), mRNA.
 NM_018459 synonyms: BM045, HIBDL; 3-hydroxyisobutyrate dehydrogenase-like;
 NM_018466 uncharacterized hematopoietic stem/progenitor cells protein MDS031
 synonyms: ATA2, SAT2, SNAT2, PRO1068, KIAA1382; amino acid transporter
 2; system A amino acid transporter;
 NM_018573 2; system A amino acid transporter;
 NM_018584 calcium/calmodulin-dependent protein kinase II
 synonyms: PRO2577, FLJ13784; Homo sapiens hypothetical protein
 NM_018630 MGC3067 (MGC3067), mRNA.
 NM_018686 cytidine monophosphate N-acetylneuraminic acid synthetase
 NM_018695 erbb2 interacting protein
 NM_018834 matrin 3
 NM_018835 membrane-associated nucleic acid binding protein
 NM_018844 B-cell receptor-associated protein 29
 NM_018846 SBB126 protein
 NM_018848 McKusick-Kaufman syndrome
 NM_018963 chromosome 21 open reading frame 107
 NM_018963 chromosome 21 open reading frame 107
 NM_018975 telomeric repeat binding factor 2, interacting protein
 NM_018976 solute carrier family 38, member 2
 NM_019005 hypothetical protein FLJ20323
 NM_019027 RNA-binding protein
 NM_019048 HCV NS3-transactivated protein 1
 NM_019057 hypothetical protein FLJ10404
 NM_019058 DNA-damage-inducible transcript 4
 NM_019067 hypothetical protein FLJ10613
 NM_019556 motile sperm domain containing 1
 NM_019887 diablo homolog (Drosophila)
 NM_019896 polymerase (DNA-directed), epsilon 4 (p12 subunit)
 NM_020127 tuftelin 1
 NM_020166 methylcrotonoyl-Coenzyme A carboxylase 1 (alpha)
 NM_020182 transmembrane, prostate androgen induced RNA
 NM_020184 cyclin M4
 NM_020193 chromosome 11 open reading frame 30
 NM_020216 arginyl aminopeptidase (aminopeptidase B)
 NM_020228 PR domain containing 10
 NM_020232 tumor necrosis factor superfamily, member 5-induced protein 1
 NM_020307 cyclin L1
 NM_020375 chromosome 12 open reading frame 5
 NM_020387 RAB25, member RAS oncogene family
 NM_020632 ATPase, H⁺ transporting, lysosomal V0 subunit a isoform 4
 NM_020657 zinc finger protein 304
 NM_020662 MRS2-like, magnesium homeostasis factor (S. cerevisiae)
 NM_020673 RAB22A, member RAS oncogene family
 NM_020987 ankyrin 3, node of Ranvier (ankyrin G)
 NM_021003 protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform
 NM_021039
 NM_021052 histone 1, H2ae
 NM_021074 NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa

NM_021101 claudin 1
 NM_021129 pyrophosphatase (inorganic)
 protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)
 NM_021132
 NM_021137 tumor necrosis factor, alpha-induced protein 1 (endothelial)
 NM_021139 UDP glycosyltransferase 2 family, polypeptide B4
 NM_021145 cyclin D binding myb-like transcription factor 1
 NM_021188 zinc finger protein 410
 NM_021213 phosphatidylcholine transfer protein
 NM_021238 chromosome 12 open reading frame 14
 NM_021615 carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6
 NM_021626 likely homolog of rat and mouse retinoid-inducible serine carboxypeptidase
 NM_021632 zinc finger protein 350
 NM_021727 fatty acid desaturase 3
 NM_021824 NIF3 NGG1 interacting factor 3-like 1 (S. pombe)
 NM_021873 cell division cycle 25B
 NM_021930 Rad50-interacting protein 1
 NM_021934 hypothetical protein FLJ11773
 NM_021941 chromosome 21 open reading frame 97
 NM_021942 FLJ12716 protein
 NM_021960 myeloid cell leukemia sequence 1 (BCL2-related)
 NM_021970 mitogen-activated protein kinase kinase 1 interacting protein 1
 NM_021972 sphingosine kinase 1
 v-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor of kappa
 NM_021975 light polypeptide gene enhancer in B-cells 3, p65 (avian)
 NM_021991 junction plakoglobin
 NM_021993 synonyms: NSSR, TASR
 NM_021994 zinc finger protein (C2H2 type) 277
 NM_022041 giant axonal neuropathy (gigaxonin)
 UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-
 NM_022087 acetylgalactosaminyltransferase 11 (GalNAc-T11)
 NM_022121 PERP, TP53 apoptosis effector
 NM_022129 MAWD binding protein
 NM_022130 golgi phosphoprotein 3 (coat-protein)
 NM_022168 melanoma differentiation associated protein-5
 NM_022170 Williams-Beuren syndrome chromosome region 1
 NM_022365 DnaJ (Hsp40) homolog, subfamily C, member 1
 NM_022459 exportin 4
 NM_022476 fused toes homolog (mouse)
 NM_022496 actin-related protein 6
 NM_022744 hypothetical protein FLJ13868
 NM_022766 ceramide kinase
 NM_022776 oxysterol binding protein-like 11
 NM_022781 ring finger protein 38
 NM_022838 hypothetical protein FLJ12969
 NM_023005 bromodomain adjacent to zinc finger domain, 1B
 NM_023009 MARCKS-like protein
 NM_023012 similar to splicing factor, arginine/serine-rich 4
 NM_023039 ankyrin repeat, family A (RFXANK-like), 2
 NM_023915 G protein-coupled receptor 87

NM_023925 C1q domain containing 1
NM_023926 hypothetical protein FLJ12895
NM_023941
NM_024047 nudix (nucleoside diphosphate linked moiety X)-type motif 9
NM_024057 nucleoporin Nup37
NM_024064 synonyms: PKCL, PKC-L, PRKCL, MGC5363;
NM_024065 phosducin-like 3
NM_024077 SECIS binding protein 2
NM_024085 hypothetical protein FLJ22169
NM_024091 hypothetical protein MGC5297
NM_024095 ankyrin repeat and SOCS box-containing 8
NM_024107 hypothetical protein MGC3123
NM_024315 chromosome 7 open reading frame 23
synonym: MGC4342; go_function: calcium ion binding [goid 0005509]
NM_024329 [evidence IEA]; Homo sapiens EF hand domain containing 2 (EFHD2), mRNA.
NM_024430 proline-serine-threonine phosphatase interacting protein 2
NM_024586 oxysterol binding protein-like 9
NM_024604 hypothetical protein FLJ21908
NM_024609
NM_024610 HSPB (heat shock 27kDa) associated protein 1
NM_024611 NMDA receptor-regulated gene 2
pleckstrin homology domain containing, family F (with FYVE domain) member
2
NM_024613
NM_024617
NM_024635 corneal wound healing-related protein
NM_024639 zinc finger protein 322A
NM_024653 PRKR interacting protein 1 (IL11 inducible)
NM_024666 hypothetical protein FLJ11506
NM_024673
NM_024675 hypothetical protein FLJ21816
NM_024680 FLJ23311 protein
NM_024712 engulfment and cell motility 3 (ced-12 homolog, C. elegans)
NM_024723 MICAL-like 2
NM_024756 multimerin 2
NM_024810 hypothetical protein FLJ23018
NM_024814 Cas-Br-M (murine) ecotropic retroviral transforming sequence-like 1
NM_024824 nuclear protein UKp68
NM_024829 hypothetical protein FLJ22662
NM_024830 hypothetical protein FLJ12443
NM_024852 eukaryotic translation initiation factor 2C, 3
NM_024897 progesterin and adipoQ receptor family member VI
NM_024901 hypothetical protein FLJ22457
NM_024918 chromosome 20 open reading frame 172
NM_024920 hypothetical protein FLJ14281
NM_024948 hypothetical protein FLJ13397
NM_025027 hypothetical protein FLJ14260
NM_025065 RNA processing factor 1
NM_025076 UDP-glucuronate decarboxylase 1
NM_025125 hypothetical protein FLJ13263
NM_025126 ring finger protein 34

NM_025137 hypothetical protein FLJ21439
 NM_025138 hypothetical protein FLJ12661
 NM_025140 limkain beta 2
 NM_025151 Rab coupling protein
 NM_025160 WD repeat domain 26
 NM_025176 KIAA0980 protein
 NM_025195 phosphoprotein regulated by mitogenic pathways
 NM_025238 BTB (POZ) domain containing 1
 NM_025264 chromosome 2 open reading frame 8
 NM_030570 uroplakin 3B
 NM_030579 cytochrome b5 outer mitochondrial membrane precursor
 NM_030757 makorin, ring finger protein, 4
 NM_030777 solute carrier family 2 (facilitated glucose transporter), member 10
 NM_030790 T-cell immunomodulatory protein
 NM_030804
 NM_030917 FIP1 like 1 (*S. cerevisiae*)
 NM_030940 HESB like domain containing 2
 NM_030952 likely ortholog of rat SNF1/AMP-activated protein kinase
 NM_030969 transmembrane protein 14B
 NM_030979 poly(A) binding protein, cytoplasmic 3
 NM_031213 hypothetical protein MGC5244
 NM_031284 ATP-dependent glucokinase
 NM_031286 SH3 domain binding glutamic acid-rich protein like 3
 NM_031298 hypothetical protein MGC2963
 NM_032876 jub, ajuba homolog (*Xenopus laevis*)
 integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes
 NM_133376 MDF2, MSK12)
 NM_144707 prominin 2
 NM_147174 heparan sulfate 6-O-sulfotransferase 2
 NM_172193 kelch domain containing 1
 R38084 yc92d06.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:23525
 3', mRNA sequence.
 R38389 yc96b12.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:23744
 3', mRNA sequence.
 R39094 yf50c09.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:25528
 3', mRNA sequence.
 R45298 CDNA FLJ25042 fis, clone CBL03351
 R49102 yg69b04.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
 IMAGE:38588 3', mRNA sequence.
 R60068 DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
 R66534 hypothetical protein FLJ35954
 R66713 yi33f06.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:141059
 3' similar to contains Alu repetitive element;contains L1 repetitive element ;,
 mRNA sequence.
 TATAA binding protein; This sequence comes from 3; conceptual translation
 S49765 presented here differs from translation in publication; homeobox gene [human,
 mRNA Partial, 769 nt].
 S69189 acyl-Coenzyme A oxidase 1, palmitoyl
 S73498 UDP-N-acteylglucosamine pyrophosphorylase 1

T32429 EST48411 Human Spleen Homo sapiens cDNA 3' end similar to None, mRNA
 sequence.
 T52999 zinc finger protein 92 (HTF12)
 ya86b01.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA clone
 T53175 IMAGE:68521 5' similar to contains Alu repetitive element, mRNA sequence.
 T59859 hypothetical protein MGC11102
 yc39d08.s1 Stratagene liver (#937224) Homo sapiens cDNA clone
 T67821 IMAGE:83055 3', mRNA sequence.
 T81452 hypothetical protein FLJ12903
 T89044 kelch-like 11 (Drosophila)
 U01351
 U04627 hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A
 thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit
 U07802 Human Tis11d gene, complete cds.
 U13261 methionyl aminopeptidase 2
 U15174 Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 3 (BNIP3)
 mRNA, complete cds.
 U21915 Human chromosome 5q13.1 clone 5G8 mRNA.
 U24223 poly(rC) binding protein 1
 U26710 Cas-Br-M (murine) ecotropic retroviral transforming sequence b
 tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein,
 zeta polypeptide
 U28964
 U32645 E74-like factor 4 (ets domain transcription factor)
 U38545 phospholipase D1, phosphatidylcholine-specific
 U39361 ubiquitin-conjugating enzyme E2 variant 1
 U40490 nicotinamide nucleotide transhydrogenase
 U42349 tumor suppressor candidate 3
 U46006 cysteine and glycine-rich protein 2
 U46024 myotubular myopathy 1
 U47077 protein kinase, DNA-activated, catalytic polypeptide
 U47635 myotubularin related protein 6
 U49188 tumor differentially expressed 1
 U49245 Rab geranylgeranyltransferase, beta subunit
 U50529 CG016
 U53823 occludin
 alternatively spliced mRNA corresponding to the rodent testis enriched heat
 shock protein 70 kDa family member; Human heat shock protein mRNA,
 complete cds.
 U56725
 U58766 tissue specific transplantation antigen P35B
 U59863 TRAF family member-associated NFkB activator
 U60521 caspase 9, apoptosis-related cysteine protease
 U61167 intersectin 2
 U61500 transmembrane protein 1
 U64661 Human poly(A)-binding protein processed pseudogene3.
 U64675 RAN binding protein 2-like 1
 U64898 nardilysin (N-arginine dibasic convertase)
 U65019 MAD, mothers against decapentaplegic homolog 2 (Drosophila)
 U67122 ubiquitin-like 1 (sentrin)
 U70451 myeloid differentiation primary response gene (88)
 U73844 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
 U76542 pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde

	synthetase)
U78303	sin3-associated polypeptide, 18kDa
U79283	D site of albumin promoter (albumin D-box) binding protein
U79297	hypothetical protein LOC157567
U79751	basic leucine zipper nuclear factor 1 (JEM-1)
	see GenBank Accession Number U01184 for cDNA; similar to <i>Drosophila melanogaster</i> flil in GenBank Accession Number U01182 and <i>Caenorhabditis elegans</i> flil homolog in GenBank Accession Number U01183; Homo sapiens FLII gene, complete cds.
U80184	
U83867	spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)
U84487	chemokine (C-X3-C motif) ligand 1
	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)
U85658	
	41 kDa protein; Homo sapiens mRNA-associated protein mrnp41 mRNA, complete cds.
U85943	
U87408	parathyroid hormone-responsive B1 gene
U88964	interferon stimulated gene 20kDa
U89281	3-hydroxysteroid epimerase
U92014	Human clone 121711 defective mariner transposon Hsmar2 mRNA sequence.
U94831	transmembrane 9 superfamily member 1
U96180	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)
	za51e06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
W02593	IMAGE:296098 5', mRNA sequence.
W05463	apoptosis-related protein PNAS-1
W19873	THAP domain containing 11
	zb58e03.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
W21283	IMAGE:307804 5', mRNA sequence.
	65F1 Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional, mRNA sequence.
W22165	
W56760	CDNA: FLJ20892 fis, clone ADKA03430
W63676	hypothetical protein BC016005
W67511	G protein-coupled receptor 115
W67644	putative translation initiation factor
W67887	protein phosphatase 1, catalytic subunit, beta isoform
W68084	EGF-like-domain, multiple 5
W69265	SEC6-like 1 (<i>S. cerevisiae</i>)
W69365	TcD37 homolog
W72053	trans-golgi network protein 2
W72338	UDP-glucose ceramide glucosyltransferase
W73272	phosphodiesterase 8A
	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)
W74620	
W80623	hypothetical protein FLJ36175
W80678	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog
W81119	serine palmitoyltransferase, long chain base subunit 2
W87466	hypothetical protein BC001096
	zd92g07.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
W92744	IMAGE:356988 3', mRNA sequence.
W93695	CDNA FLJ39417 fis, clone PLACE6016942
W93787	golgi reassembly stacking protein 2, 55kDa

X03363	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
X06989	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
X57198	transcription elongation factor A (SII), 1
X62048	WEE1 homolog (S. pombe)
X63381	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
X69397	CD24 antigen (small cell lung carcinoma cluster 4 antigen)
X76061	retinoblastoma-like 2 (p130)
X79510	protein tyrosine phosphatase, non-receptor type 21
X79683	H.sapiens LAMB2 mRNA for beta2 laminin.
X98743	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18
X99142	keratin, hair, basic, 6 (monilethrix)
Y15724	Homo sapiens SERCA3 gene, exons 1-7 (and joined CDS).
Z22551	kinectin 1 (kinesin receptor)
Z24725	pleckstrin homology domain containing, family C (with FERM domain) member 1
Z25521	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)
Z48199	H.sapiens syndecan-1 gene (exons 2-5).
Z48950	H.sapiens hH3.3B gene for histone H3.3.
Z78330	ARP3 actin-related protein 3 homolog (yeast)
Z92546	

Supplementary Table 2: Expression down by Aza

Genbank	Description
AA001375	Transcribed sequences
AA017429	Transcribed sequences
AA054734	CDKN1A interacting zinc finger protein 1
AA115105	zI03a11.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:491228 3', mRNA sequence.
AA148929	CDNA FLJ45363 fis, clone BRHIP3015854
AA161476	zq49g09.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:633088 3' similar to contains Alu repetitive element;contains element MER28 repetitive element ;, mRNA sequence.
AA173223	zp31b04.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone IMAGE:611023 3' similar to contains Alu repetitive element;, mRNA sequence.
AA176313	CDNA: FLJ23262 fis, clone COL05922
AA292789	hypothetical protein LOC146909
AA398043	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
AA398590	hypothetical protein MGC29898
AA459699	hypothetical protein FLJ11795
AA461242	ras homolog gene family, member J
AA502768	FLJ32363 protein
AA514634	hypothetical protein FLJ10979
AA515560	high density lipoprotein binding protein (vigilin)
AA552969	sialyltransferase 8D (alpha-2, 8-polysialyltransferase)
AA603344	np29d10.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1117747 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;, mRNA sequence.
AA669799	acetylserotonin O-methyltransferase-like

AA700440 CDNA FLJ30779 fis, clone FEBRA2000815
 AA721240 Transcribed sequences
 AA731688 Transcribed sequences
 AA761055 Transcribed sequences
 MCM5 minichromosome maintenance deficient 5, cell division cycle 46 (S. cerevisiae)
 AA807529
 AA812086 hypothetical protein MGC40574
 AA846789 LOC399722 (LOC387636), mRNA
 AA853175 solute carrier family 16 (monocarboxylic acid transporters), member 3
 aj75a10.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone IMAGE:1402266 3' similar to SW:NIDM_BOVIN Q02373 NADH-UBIQUINONE OXIDOREDUCTASE PDSW SUBUNIT ;, mRNA sequence.
 AA854479
 AA861435 hypothetical protein MGC47869
 AA872187 CDNA FLJ44231 fis, clone THYMU3006767
 AA994334 B-cell CLL/lymphoma 10
 AB004064 TR; Homo sapiens mRNA for tomoregulin, complete cds.
 AB012143 RNA guanylyltransferase and 5'-phosphatase
 AB014583 KIAA0683 gene product
 AB017546 peroxisomal biogenesis factor 14
 AB020648 KIAA0841
 AB037804 KIAA1383 protein
 AB046828 KIAA1608
 AB049117 actin related protein M1
 AB051452 RNA polymerase III subunit RPC8
 AF008442 polymerase (RNA) I polypeptide C, 30kDa
 AF032897 potassium voltage-gated channel, subfamily H (eag-related), member 7
 AF034956 RAD51-like 3 (S. cerevisiae)
 AF052100 vesicle-associated membrane protein 4
 AF060924 hypothetical protein HSPC242
 AF070565 Homo sapiens clone 24425 mRNA sequence.
 C-terminus similar to human metallothionein-IF: Swiss-Prot Accession Number P04733; Homo sapiens hqp0376 protein mRNA, complete cds.
 AF078844
 AF085965 nuclear domain 10 protein
 AF086279 Full length insert cDNA clone ZD45F06
 AF086328 Full length insert cDNA clone ZD54C03
 protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), gamma isoform
 AF086924
 AF094508 Homo sapiens dentin phosphoryn mRNA, complete cds.
 AF100751 FK506 binding protein 7
 AF112345 integrin, alpha 10
 AF119873 PRO2275 mRNA, complete cds
 AF119875 predicted protein of HQ2309; Homo sapiens PRO2309 mRNA, complete cds.
 AF125393 RAB27A, member RAS oncogene family
 AF131799 sidekick homolog 1 (chicken)
 AF136970 pipecolic acid oxidase
 AF139625 centromere protein J
 AF161347 trinucleotide repeat containing 5
 AF181660 myelin protein zero-like 1
 AF212236 ADP-ribosylhydrolase like 2
 AF222345 suppressor of fused homolog (Drosophila)
 AF225416 kinetochore protein Spc25

AF258562 deoxythymidylate kinase (thymidylate kinase)
 AF265439 mitochondrial ribosomal protein S15
 AF295728 chromosome 18 open reading frame 2
 AF316855 hypothetical protein FLJ22795
 AFFX-BioC-3
 AI003763 hypothetical protein FLJ12681
 ox05b02.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone
 AI021939 IMAGE:1655403 3', mRNA sequence.
 ox08b03.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone
 AI023774 IMAGE:1655693 3' similar to TR:O04471 O04471 SIMILAR TO
 SACCHAROMYCES HYPOTHETICAL PROTEIN P9642.2. ;contains Alu
 repetitive element;; mRNA sequence.
 ov84h08.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1644063
 AI028310 3', mRNA sequence.
 ow70h07.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone
 AI032585 IMAGE:1652221 3' similar to contains Alu repetitive element;contains element
 PTR5 repetitive element ;, mRNA sequence.
 AI073549 estrogen receptor 1
 AI074450 Clone IMAGE:121662 mRNA sequence
 AI078196 Transcribed sequences
 AI128170 Transcribed sequences
 Transcribed sequence with weak similarity to protein ref:NP_060265.1
 AI140917 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
 AI143752 Tu translation elongation factor, mitochondrial
 AI144156 CDNA FLJ41972 fis, clone SKNMC2003987
 AI188346 similar to RIKEN cDNA 2610307121
 Hypothetical protein LOC283507, mRNA (cDNA clone IMAGE:5314144), partial
 AI203405 cds
 AI220427 Transcribed sequences
 qg43e04.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1837950
 AI220472 3', mRNA sequence.
 AI245026 spermatogenesis associated 11
 AI249173 Clone IMAGE:4151011, mRNA
 AI252582 ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e
 AI271418 hypothetical protein LOC150763
 Transcribed sequence with weak similarity to protein ref:NP_062553.1
 AI276956 (H.sapiens) hypothetical protein FLJ11267 [Homo sapiens]
 AI278204 MRNA; cDNA DKFZp564O0862 (from clone DKFZp564O0862)
 AI300126 CDNA FLJ41470 fis, clone BRSTN2019079
 AI307915 SAM pointed domain containing ets transcription factor
 AI340264 ropporin, rhopilin associated protein 1
 AI341602 Clone IMAGE:5200861, mRNA
 qt26c08.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
 AI342132 IMAGE:1949102 3', mRNA sequence.
 AI346350 polymyositis/scleroderma autoantigen 1, 75kDa
 AI357042 v-myb myeloblastosis viral oncogene homolog (avian)
 AI357143 Transcribed sequences
 AI421812 hypothetical protein FLJ13912
 AI435828 stanniocalcin 2
 AI493276 Transcribed sequences
 AI560951 mitochondrial ribosomal protein L52

AI589978 chromosome 14 open reading frame 116
 AI623202 PR domain containing 16
 AI627636 Hypothetical gene supported by BC040598 (LOC400960), mRNA
 AI640483 CDNA FLJ44273 fis, clone TOVAR2001281
 AI658662 synaptopodin 2
 AI659180 translin
 AI671257 hypothetical protein DKFZp547K1113
 AI672363 vacuolar protein sorting 33B (yeast)
 AI677701 RNA-binding region (RNP1, RRM) containing 6
 AI690169 Clone IMAGE:5315196, mRNA
 AI692169 wd37e07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:2330340 3', mRNA sequence.
 AI693543 CDNA FLJ31443 fis, clone NT2NE2000808
 solute carrier family 13 (sodium-dependent dicarboxylate transporter), member
 3
 AI700882
 AI703265 hypothetical protein FLJ90119
 AI733356 SMC1 structural maintenance of chromosomes 1-like 1 (yeast)
 AI739231 MRNA; cDNA DKFZp686C1238 (from clone DKFZp686C1238)
 AI742383 Transcribed sequences
 AI743416 hypothetical protein FLJ33718
 AI744580 wg04b09.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:2364089 3' similar to contains Alu repetitive element;contains L1.t2 L1
 repetitive element ;, mRNA sequence.
 AI753488 glutamate-cysteine ligase, modifier subunit
 AI762113 GDP-mannose 4,6-dehydratase
 AI809328 Similar to hypothetical protein A130030D10 (LOC344287), mRNA
 AI825212 Alstrom syndrome 1
 AI872645 dynein, axonemal, heavy polypeptide 5
 AI885411 hypothetical protein FLJ22222
 AI911434 myeloid leukemia factor 1
 AI922198 Hermansky-Pudlak syndrome 3
 AI926924 Transcribed sequences
 AI961235 hypothetical protein FLJ12505
 AI961429 ubiquinol-cytochrome c reductase core protein II
 AI971519 x 009 protein
 AI983115 wu18b02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2517291 3'
 similar to contains Alu repetitive element;, mRNA sequence.
 AI990178 Transcribed sequences
 AJ003030 syntrophin, gamma 1
 AJ224170 matrin 3
 AK000016 Homo sapiens cDNA FLJ20009 fis, clone ADKA03183.
 AK000721 timeless homolog (Drosophila)
 AK000818 hypothetical protein FLJ20811
 AK001261 RA-regulated nuclear matrix-associated protein
 AK021607 essential meiotic endonuclease 1 homolog 1 (S. pombe)
 AK022442 Homo sapiens cDNA FLJ12380 fis, clone MAMMA1002556.
 AK022955 transcription factor RAM2
 AK024138 Homo sapiens cDNA FLJ14076 fis, clone HEMBB1001925.
 AK024273 COP9 constitutive photomorphogenic homolog subunit 7B (Arabidopsis)
 AK024855 Homo sapiens cDNA: FLJ21202 fis, clone COL00293.
 AK024995 Homo sapiens cDNA: FLJ21342 fis, clone COL02673.

AK025101 CDNA: FLJ21448 fis, clone COL04473
 AK025654 CTP synthase II
 AK026825 LOC389177 (LOC389177), mRNA
 AK027107 CDNA FLJ41259 fis, clone BRAMY2034305
 AK027184 fetal Alzheimer antigen
 AK055438 transmembrane 6 superfamily member 1
 AK056519 ATP-binding cassette, sub-family C (CFTR/MRP), member 9
 AK091836 MRNA; cDNA DKFZp686I05132 (from clone DKFZp686I05132)
 AK092544 CDNA FLJ35225 fis, clone PROST2001116
 unnameD protein product; Homo sapiens cDNA FLJ38183 fis, clone
 FCBBF1000155.
 AK095502
 AK096657 hypothetical protein LOC145783
 AK097492 CDNA FLJ40173 fis, clone TESTI2016922
 AK097497 CDNA FLJ40178 fis, clone TESTI2017932
 AL041761 Clone IMAGE:4753714, mRNA
 AL049390 MRNA; cDNA DKFZp586O1318 (from clone DKFZp586O1318)
 known gamma-adaptin / chimeric ?; Homo sapiens mRNA; cDNA
 DKFZp564D066 (from clone DKFZp564D066); partial cds.
 AL050025 MRNA; cDNA DKFZp586E121 (from clone DKFZp586E121)
 AL050122
 AL050353 Opa-interacting protein 2
 AL079341
 AL080280 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 85905.
 AL096779 chromosome 22 open reading frame 4
 Human DNA sequence from clone RP3-526F5 on chromosome Xq26.3-28,
 complete sequence.
 AL109622
 AL117381
 AL136840 MCM10 minichromosome maintenance deficient 10 (*S. cerevisiae*)
 AL137067
 AL138240 DEAD (Asp-Glu-Ala-Asp) box polypeptide 56
 AL138349 KIAA0367
 AL390738
 AL571557 MRNA; cDNA DKFZp686G03142 (from clone DKFZp686G03142)
 AL571598 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DI024YM24 3-PRIME, mRNA sequence.
 AL571598
 AL572471 centromere protein H
 AL700385 Clone IMAGE:4042121, mRNA, partial cds
 AL707919 Clone IMAGE:4290175, mRNA
 AL713632 Homo sapiens mRNA; cDNA DKFZp761B0221 (from clone DKFZp761B0221).
 AL831814 Rho guanine nucleotide exchange factor (GEF) 7
 AL831879 Homo sapiens mRNA; cDNA DKFZp547I1410 (from clone DKFZp547I1410).
 AL832255 MRNA; cDNA DKFZp667D1717 (from clone DKFZp667D1717)
 AL832400 NACHT, leucine rich repeat and PYD containing 1
 AL832710 plexin D1
 AL833769 serologically defined colon cancer antigen 10
 AU144567 ankyrin repeat and SOCS box-containing 3
 AU145100 solute carrier family 24 (sodium/potassium/calcium exchanger), member 5
 AU146391 HEMBB1 Homo sapiens cDNA clone HEMBB1000343 3', mRNA
 sequence.
 AU146391 CDNA FLJ12004 fis, clone HEMBB1001564, moderately similar to VACUOLAR
 AU146809 ATP SYNTHASE SUBUNIT H (EC 3.6.1.34)

AU147017 HEMBB1 Homo sapiens cDNA clone HEMBB1002152 3', mRNA sequence.
 AU157303 histamine N-methyltransferase
 Transcribed sequence with moderate similarity to protein pir:T02670
 AV648418 (H.sapiens) T02670 probable thromboxane A2 receptor isoform beta - human
 AV696977 GKC Homo sapiens cDNA clone GKCEEG06 5', mRNA sequence.
 AV700969 MRNA; cDNA DKFZp686K14148 (from clone DKFZp686K14148)
 AV738806 AV738806 CB Homo sapiens cDNA clone CBCAWB04 5', mRNA sequence.
 AW001030 chromosome 9 open reading frame 52
 wq63e07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2475972 3', mRNA sequence.
 AW003173
 AW003893 quaking homolog, KH domain RNA binding (mouse)
 AW004065 hypothetical protein FLJ35843
 AW024383 ribosomal protein S21
 AW024741 MRNA; cDNA DKFZp586B1024 (from clone DKFZp586B1024)
 AW025529 serologically defined breast cancer antigen NY-BR-20
 AW051712 Transcribed sequences
 AW057819 pyruvate dehydrogenase (lipoamide) alpha 1
 AW075105 DNA replication factor
 AW080025 Transcribed sequences
 AW081561 hypothetical protein FLJ25006
 AW134976 KIAA0984 protein
 AW183692 chromosome 21 open reading frame 107
 AW189966 KIAA0841
 AW190316 NADH:ubiquinone oxidoreductase MLRQ subunit homolog
 AW190316 NADH:ubiquinone oxidoreductase MLRQ subunit homolog
 AW195928 E3 ubiquitin ligase SMURF2
 AW237225 syntaxin binding protein 4
 AW242720 MRNA; cDNA DKFZp686J0156 (from clone DKFZp686J0156)
 AW263087 GDNF family receptor alpha 1
 AW263497 synaptotagmin-like 5
 xr49a07.x1 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:2763444 3', mRNA sequence.
 AW271106
 AW273860 influenza virus NS1A binding protein
 AW298438 nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3
 AW302047 ribosomal protein S10
 AW303136 ribosomal protein L38
 AW451271 developmentally regulated RNA-binding protein 1
 AW451961 adenylate cyclase activating polypeptide 1 (pituitary) receptor type I
 AW468382 ATPase, H⁺ transporting, lysosomal 50/57kDa, V1 subunit H
 AW512196 serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 1
 AW572778 Transcribed sequences
 Transcribed sequence with weak similarity to protein ref:NP_060312.1
 AW664964 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
 AW665177 Clone IMAGE:4830853, mRNA
 AW665375 Transcribed sequences
 AW665713 DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26
 AW969913 Transcribed sequences
 AW972824 Transcribed sequences
 AW974997 tousled-like kinase 1

AY042224 CD209 antigen
 BC000535 peter pan homolog (Drosophila)
 BC000658 stanniocalcin 2
 BC000747 homolog of Yeast RRP4 (ribosomal RNA processing 4), 3'-5'-exoribonuclease
 BC001560 PTPRF interacting protein, binding protein 1 (liprin beta 1)
 BC001867 Homo sapiens cDNA clone MGC:1167 IMAGE:3536204, complete cds.
 BC002836
 BC003068 solute carrier family 19 (folate transporter), member 1
 BC003186 DNA replication complex GINS protein PSF2
 BC003621 heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)
 BC003679 ATP synthase, H⁺ transporting, mitochondrial F0 complex, subunit e
 BC004123 Homo sapiens cDNA clone IMAGE:3929520, partial cds.
 BC004277 hypothetical protein FLJ10719
 BC005102 hypothetical protein MGC13186
 BC005369 egl nine homolog 1 (C. elegans)
 BC005810 stem cell growth factor; lymphocyte secreted C-type lectin
 BC011119 Homo sapiens, clone IMAGE:3047997, mRNA.
 BC011549 ATP synthase, H⁺ transporting, mitochondrial F0 complex, subunit s (factor B)
 BC012946 Rho-guanine nucleotide exchange factor
 BC013582 Fanconi anemia, complementation group D2
 BC014936 Clone IMAGE:3921647, mRNA
 BC015940 5'-nucleotidase, ecto (CD73)
 BC019842 Clone IMAGE:4516253, mRNA
 BC019880 Homo sapiens hypothetical protein MGC21881, mRNA (cDNA clone IMAGE:3867537), partial cds.
 BC019880 Homo sapiens hypothetical protein MGC21881, mRNA (cDNA clone IMAGE:3867537), partial cds.
 BC020640 CGI-146 protein
 BC021121 protein kinase, lysine deficient 1
 BC022234 Homo sapiens, clone IMAGE:4401081, mRNA.
 BC022297 diacylglycerol kinase, epsilon 64kDa
 BC022406 Homo sapiens, clone IMAGE:4716286, mRNA.
 BC024000 NP220 nuclear protein
 BC027254
 BC027983 CDNA clone IMAGE:3895112, containing frame-shift errors
 BC028845 Homo sapiens cDNA clone IMAGE:4796102, partial cds.
 BC029470 Clone IMAGE:4723407, mRNA, partial cds
 BC029803 Homo sapiens hypothetical protein MGC39900, mRNA (cDNA clone IMAGE:5187354), containing frame-shift errors.
 BC030524 claudin 19
 BC030552 Clone IMAGE:5223566, mRNA
 BC030578 eukaryotic translation initiation factor 4 gamma, 3
 BC031239 Clone IMAGE:5259303, mRNA
 BC033321 Clone IMAGE:4828738, mRNA
 BC033718 chromosome 21 open reading frame 106
 BC036731 zinc finger protein 450
 BC037412 collagen, type V, alpha 3
 BC038774 Hypothetical protein LOC283914, mRNA (cDNA clone IMAGE:4837572), partial cds
 BC038780 CDNA clone IMAGE:5271447, partial cds
 BC039122 DKFZP566B183 protein

BC039241 LOC126731
 BC039507 Hypothetical gene supported by BC039507 (LOC400656), mRNA
 BC040572 CDNA FLJ35761 fis, clone TESTI2004791
 BC041818 Similar to RIKEN cDNA 4933434I20 gene, clone IMAGE:5267198, mRNA
 BC042071 Clone IMAGE:5742522, mRNA
 BC042076 Keratin 7, mRNA (cDNA clone IMAGE:5743968), partial cds
 BC042986 CDNA clone IMAGE:5296106, partial cds
 BC042988 MUF1 protein
 BC043282 CDNA clone IMAGE:5297486, partial cds
 BE092211 IL2-BT0734-240400-071-D01 BT0734 Homo sapiens cDNA, mRNA sequence.
 BE207758 MRNA; cDNA DKFZp762M127 (from clone DKFZp762M127)
 BE220224 Transcribed sequences
 BE222220 hypothetical protein LOC93273
 601145652F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161124 5',
 BE312027 mRNA sequence.
 601159345F2 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510995 5',
 BE380031 mRNA sequence.
 BE464269 U2-associated SR140 protein
 BE501980 CDNA clone IMAGE:4800077, partial cds
 BE502030 RAB27A, member RAS oncogene family
 BE544837 PHD finger protein 19
 BE560461 solute carrier family 2 (facilitated glucose/fructose transporter), member 5
 7d74a04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3278670 3'
 BE674061 similar to WP:Y48C3A.S CE22150 ;, mRNA sequence.
 BE873420 UDP glycosyltransferase 1 family, polypeptide A10
 601437912F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922971 5',
 BE895437 mRNA sequence.
 601499337F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3901160 5',
 BE905157 mRNA sequence.
 BE963444 hypothetical protein A-211C6.1
 601658585R1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885812 3',
 BE964655 mRNA sequence.
 BF001806 antigen identified by monoclonal antibody Ki-67
 BF055200 homeo box A13
 BF062364 olfactory receptor, family 7, subfamily E, member 24
 BF109140 ubiquitin specific protease 13 (isopeptidase T-3)
 BF115777 hypothetical protein FLJ10979
 BF217471 adenylate cyclase 3
 BF224073 acetyl-Coenzyme A acetyltransferase 2 (acetoacetyl Coenzyme A thiolase)
 BF246115 metallothionein 1F (functional)
 BF308250 TRAF6-inhibitory zinc finger protein
 BF339845 KIAA1586
 Similar to dJ132F21.2 (Contains a novel protein similar to the L82E from
 BF437747 Drosophila) (LOC343574), mRNA
 Transcribed sequence with moderate similarity to protein sp:P39188
 (H.sapiens) ALU1_HUMAN Alu subfamily J sequence contamination warning
 entry
 BF438028
 BF448693 Transcribed sequences
 BF507371 Transcribed sequences
 BF510789 ADP-ribosylarginine hydrolase
 BF592982 gamma tubulin ring complex protein (76p gene)

BF669264 602120534F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4277696 5', mRNA sequence.
 BF672306 CDNA FLJ90295 fis, clone NT2RP2000240.
 BF678148 CDNA FLJ46495 fis, clone THYMU3028461
 BF692729 602079807F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4244520 5', mRNA sequence.
 BF732480 nae10f02.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3435003 3', mRNA sequence.
 BF732484 Transcribed sequences
 BF752277 carbonic anhydrase XII
 BF973374 MRNA; cDNA DKFZp686P09209 (from clone DKFZp686P09209)
 BF973387 ELOVL family member 5, elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3-like, yeast)
 BF981643 septin 10
 BG105700 602311907F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4421928 5', mRNA sequence.
 BG107203 expressed in hematopoietic cells, heart, liver
 BG180941 heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa)
 BG472587 602514350F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4645981 5', mRNA sequence.
 BG535396 hypothetical protein DKFZp762C1112
 BG704082 Clone IMAGE:4819737, mRNA
 BG819064 Hypothetical protein LOC284570, mRNA (cDNA clone IMAGE:4943937), containing frame-shift errors
 BI548390 CDNA clone IMAGE:5260914, partial cds
 BM682057 UI-E-EO1-aiw-m-22-0-UI.s1 UI-E-EO1 Homo sapiens cDNA clone UI-E-EO1-aiw-m-22-0-UI 3', mRNA sequence.
 BM786513 CDNA clone MGC:61706 IMAGE:6162269, complete cds
 BQ017399 chromosome 21 open reading frame 106
 BQ671034 nicotinamide nucleotide adenylyltransferase 1
 BU952437 CDNA FLJ42571 fis, clone BRACE3008036
 D21851 leucyl-tRNA synthetase 2, mitochondrial
 D81792 PBX/knotted 1 homeobox 2
 H29479 Transcribed sequences
 H78106 CDNA FLJ41318 fis, clone BRAMY2044246
 H90656 nicotinamide nucleotide adenylyltransferase 2
 M73554 cyclin D1 (PRAD1: parathyroid adenomatosis 1)
 N33616 Full length insert cDNA YU76C01
 N34846 DAZ associated protein 2
 N79004 sine oculis homeobox homolog 1 (Drosophila)
 NM_000125 estrogen receptor 1
 NM_000376 vitamin D (1,25- dihydroxyvitamin D3) receptor
 NM_000410 hemochromatosis
 NM_000709 branched chain keto acid dehydrogenase E1, alpha polypeptide (maple syrup urine disease)
 NM_000805 gastrin
 NM_001123 adenosine kinase
 NM_001172 arginase, type II
 NM_001218 carbonic anhydrase XII

NM_001729 betacellulin
 NM_001790 cell division cycle 25C
 NM_001881 cAMP responsive element modulator
 NM_002073 guanine nucleotide binding protein (G protein), alpha z polypeptide
 NM_002228 v-jun sarcoma virus 17 oncogene homolog (avian)
 NM_002412 O-6-methylguanine-DNA methyltransferase
 NM_002439 mutS homolog 3 (E. coli)
 NM_002450
 NM_002490 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa
 NM_002515 neuro-oncological ventral antigen 1
 NM_002757 mitogen-activated protein kinase kinase 5
 NM_002875 RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)
 NM_003358 UDP-glucose ceramide glucosyltransferase
 NM_003412 Zic family member 1 (odd-paired homolog, Drosophila)
 NM_003608 G protein-coupled receptor 65
 NM_003897 immediate early response 3
 NM_004260 RecQ protein-like 4
 NM_004350 runt-related transcription factor 3
 NM_004566 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
 NM_004577 phosphoserine phosphatase
 NM_004583 RAB5C, member RAS oncogene family
 NM_004853 syntaxin 8
 NM_005187 core-binding factor, runt domain, alpha subunit 2; translocated to, 3
 NM_005213 cystatin A (stefin A)
 NM_005264 GDNF family receptor alpha 1
 NM_005375 v-myb myeloblastosis viral oncogene homolog (avian)
 NM_005391 pyruvate dehydrogenase kinase, isoenzyme 3
 NM_005887 deleted in lymphocytic leukemia, 1
 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila);
 translocated to, 7
 NM_005938
 NM_005950 metallothionein 1G
 NM_005952 metallothionein 1X
 NM_005953 metallothionein 2A
 NM_006037 histone deacetylase 4
 NM_006221 ubiquitin-like 5
 NM_006259 protein kinase, cGMP-dependent, type II
 NM_006392 nucleolar protein 5A (56kDa with KKE/D repeat)
 NM_006491 neuro-oncological ventral antigen 1
 NM_006501 myelin-associated oligodendrocyte basic protein
 NM_006512 serum amyloid A4, constitutive
 NM_006567 phenylalanine-tRNA synthetase 1 (mitochondrial)
 NM_006579 emopamil binding protein (sterol isomerase)
 NM_006824 EBNA1 binding protein 2
 NM_006867 RNA binding protein with multiple splicing
 NM_007100 ATP synthase, H⁺ transporting, mitochondrial F₀ complex, subunit e
 synonym: MACAM1; isoform a precursor is encoded by transcript variant 1;
 NM_007164 mucosal addressin cell adhesion molecule-1;
 NM_012225 nucleotide binding protein 2 (MinD homolog, E. coli)
 NM_014139 sodium channel, voltage-gated, type XI, alpha
 NM_014484 molybdenum cofactor synthesis 3

NM_014502 PRP19/PSO4 homolog (*S. cerevisiae*)
 NM_014569 zinc finger protein 95 homolog (mouse)
 NM_014641
 NM_014668 GREB1 protein
 NM_014956
 NM_015380 CGI-51 protein
 NM_015936 CGI-04 protein
 NM_015950 mitochondrial ribosomal protein L2
 nemo-like kinase; go_component: nucleus [goid 0005634] [evidence TAS]
 [pmid 9448268];
 NM_016231
 NM_016373 WW domain containing oxidoreductase
 NM_016653 sterile alpha motif and leucine zipper containing kinase AZK
 NM_017566 hypothetical protein DKFZp434G0522
 NM_017688 B-box and SPRY domain containing
 NM_017739 O-linked mannose beta1,2-N-acetylglucosaminyltransferase
 NM_018076 armadillo repeat containing 4
 NM_018603
 NM_018693 F-box only protein 11
 NM_020667 Homo sapiens COBW-like protein (LOC55871), mRNA.
 NM_021922 Fanconi anemia, complementation group E
 NM_022063 hypothetical protein FLJ13188
 NM_022346 chromosome condensation protein G
 NM_022782 M-phase phosphoprotein 9
 NM_023077 hypothetical protein FLJ12439
 NM_024701 ankyrin repeat and SOCS box-containing 13
 NM_024744 amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 8
 NM_024745 likely ortholog of mouse Shc SH2-domain binding protein 1
 NM_024784 zinc finger and BTB domain containing 3
 NM_025059 chromosome 6 open reading frame 97
 NM_025208 platelet derived growth factor D
 NM_030941 exonuclease NEF-sp
 NM_052931 SLAM family member 6
 NM_058173 small breast epithelial mucin
 NM_139173 CG10806-like
 NM_145702 eukaryotic translation initiation factor 4E-like 3
 NM_152633 hypothetical protein FLJ34064
 NM_153224 hypothetical protein MGC34034
 NM_171846 lactamase, beta
 NM_173580 hypothetical protein FLJ39058
 NM_173596 solute carrier family 39 (metal ion transporter), member 5
 NM_173672 peptidylprolyl isomerase (cyclophilin)-like 6
 NM_173714 NADH dehydrogenase, subunit 6 (complex I);
 yf76f07.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:28365
 R40892 3', mRNA sequence.
 R42449 hypothetical protein FLJ10357
 R52640 ubiquitin protein ligase E3A (human papilloma virus E6-associated protein,
 Angelman syndrome)
 R54026 Transcribed sequences
 Hypothetical protein LOC255512, mRNA (cDNA clone IMAGE:5274144), partial
 R67325 cds

R78604	Transcribed sequence with moderate similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
S75174	E2F transcription factor 4, p107/p130-binding
T75480	potassium channel tetramerisation domain containing 6
T82467	Transcribed sequences
T90703	Transcribed sequences
U38276	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3F
U38372	Human huntingtin associated protein (hHAP1) mRNA, partial cds. Homo sapiens chromosome X clone Qc-7G6, QLL-F1720, QLL-C1335, Qc-8B7, Qc-11H12, Qc-7F6, QLL-E153, Qc-10E8, Qc-10B7 map q28, complete sequence.
U52111	
U69199	KIAA1977 protein
U78181	amiloride-sensitive cation channel 2, neuronal
W58255	G protein-coupled receptor 61
W65369	SET domain, bifurcated 2
W74640	TBC1 domain family, member 16
W81117	selenoprotein K
X98258	M-phase phosphoprotein 9
Z84723	Human DNA sequence from phage LAW2 from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains Interleukin 9 receptor pseudogene.

Supplementary table 3 Chromatin down by Aza

Genbank	Description
AA005137	pyridoxine-5'-phosphate oxidase
AA015606	KCCR13L
AA026297	ze91g12.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA
AA044730	ubiquitin specific protease 31
AA045183	glucocorticoid modulatory element binding protein 2
AA053830	C-terminal binding protein 1
AA070330	dihydroorotate dehydrogenase
AA127686	hypothetical protein FLJ14775
AA143060	melanoma associated antigen (mutated) 1
AA149639	quaking homolog, KH domain RNA binding (mouse)
AA161026	zo62a12.s1 Stratagene pancreas (#937208)
AA161140	huntingtin interacting protein 2
AA195017	KIAA1609 protein
AA195124	KIAA1609 protein
AA195408	zr36b11.s1 Soares_NhHMPu_S1 Homo sapiens
AA251505	hypothetical protein FLJ31528
AA284075	kinesin 2 60/70kDa
AA284075	kinesin 2 60/70kDa
AA398043	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
AA427737	protein tyrosine phosphatase type IVA, member 3
AA448208	calcium channel, voltage-dependent, L type, alpha 1B subunit
AA457033	peptidylprolyl isomerase A (cyclophilin A)
AA464753	actin related protein 2/3 complex, subunit 1B, 41kDa
AA470761	HMBA-inducible
AA481656	protocadherin 1 (cadherin-like 1)
AA489681	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2

AA523105 TRIAD3 protein
AA523537 sema domain, immunoglobulin domain (Ig)
AA524740 SPPL2b
AA526470 T-box 3 (ulnar mammary syndrome)
AA527180 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
AA527578 solute carrier family 16 (monocarboxylic acid transporters), member 3
AA530892 dual specificity phosphatase 1
AA552953 hypothetical protein FLJ31528
AA555113 Transcribed sequence with moderate similarity to protein sp:P05388
AA573862 major histocompatibility complex, class I, A
AA576961 pleckstrin homology-like domain, family A, member 1
AA581879 Hypothetical protein similar to topoisomerase (DNA) III beta (H. sapiens)
AA587390 mucin 5, subtypes A and C, tracheobronchial/gastric
AA608820 neurexin 2
AA629059 leishmanolysin-like (metallopeptidase M8 family)
AA632295 stonin 2
AA634220 neurofascin
AA639585 zinc finger protein 576
AA648913 baculoviral IAP repeat-containing 5 (survivin)
AA649851 UPF3 regulator of nonsense transcripts homolog A (yeast)
AA677438 tripartite motif-containing 10
AA704766 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
AA706316 hypothetical gene ZD52F10
AA715041 nx94c09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1269904
AA758795 guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O
AA761673 helicase with zinc finger domain
AA766646 oa38d07.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:130724
AA775396 leucine-rich repeats and immunoglobulin-like domains 1
AA777852 solute carrier family 22 (organic anion transporter), member 7
AA789278 ribosomal protein L13
AA806142 CAP, adenylate cyclase-associated protein 1 (yeast)
AA811509 hypothetical protein LOC285458
AA811923 ob72f05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336929
AA829860 cyclin M3
AA853175 solute carrier family 16 (monocarboxylic acid transporters), member 3
AA853175 solute carrier family 16 (monocarboxylic acid transporters), member 3
AA877910 nr12h12.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1161671 .
AA890272 hypothetical protein FLJ00058
AA912743 PTK2 protein tyrosine kinase 2
AA923372 hypothetical protein MGC46719
AA935633 quaking homolog, KH domain RNA binding (mouse)
AA947302 inhibitor of growth family, member 1
AA961748 ribosomal protein L13
AA976278 ribosomal protein S8
AB001467 embryonal Fyn-associated substrate
AB002304 capicua homolog (Drosophila)
AB002361 KIAA0363 protein
AB006589 estrogen receptor 2 (ER beta)
AB006590 estrogen receptor 2 (ER beta)
AB006622 KIAA0284

AB006627 astrotactin
 AB006628 FCH domain only 1
 AB006630 transcription factor 20 (AR1)
 AB007830 scavenger receptor class A, member 3
 AB007830 scavenger receptor class A, member 3
 AB011093 rho/rac guanine nucleotide exchange factor (GEF) 18
 AB016823 microtubule-associated protein, RP/EB family, member 2
 AB022433 sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B
 AB022435 huntingtin interacting protein 2
 AB028021 forkhead box A2
 AB028021 forkhead box A2
 AB028869 baculoviral IAP repeat-containing 5 (survivin)
 AB028953 KIAA1030 protein
 AB028953 KIAA1030 protein
 AB028989 mitogen-activated protein kinase 8 interacting protein 3
 AB028989 mitogen-activated protein kinase 8 interacting protein 3
 AB029009 KIAA1086
 AB030034 sterile alpha motif and leucine zipper containing kinase AZK
 AB032946 calcium channel, voltage-dependent, alpha 1I subunit
 AB033941 dickkopf homolog 2 (*Xenopus laevis*)
 AB035266 neurexin 2
 AB035482 chromosome 1 open reading frame 38
 AB040900 KIAA1467 protein
 AB040900 KIAA1467 protein
 AB042719 MCM10 minichromosome maintenance deficient 10 (*S. cerevisiae*)
 AB044548 eukaryotic translation initiation factor 4E binding protein 1
 AB047005 microtubule associated serine/threonine kinase 2
 AB048286 diacylglycerol O-acyltransferase homolog 2 (mouse)
 AB050468 leucine-rich repeats and immunoglobulin-like domains 1
 AB054576 neuronal PAS domain protein 3
 AB054985 calcium channel, voltage-dependent, beta 1 subunit
 AC004410 Hypothetical 43.4 kDa human protein; Homo sapiens chromosome 19, fosmid 39554
 AC004770
 AC004893
 AC005591
 AC006986 Homo sapiens BAC clone RP11-155J5 from Y, complete sequence.
 AF000424 leukocyte specific transcript 1
 AF000425 leukocyte specific transcript 1
 AF000426 leukocyte specific transcript 1
 AF003114 cysteine-rich, angiogenic inducer, 61
 AF010404 myeloid/lymphoid or mixed-lineage leukemia 2
 AF015730 glutamate receptor, ionotropic, N-methyl D-aspartate 1
 AF015731 glutamate receptor, ionotropic, N-methyl D-aspartate 1
 AF017307 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
 AF024540 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, *Drosophila*)
 AF026071 tumor necrosis factor receptor superfamily, member 25
 AF029778 jagged 2
 AF035278 supervillin
 AF035321 dynamin 1
 AF040723 huntingtin-associated protein 1 (neuroan 1)

AF041209 midline 1 (Opitz/BBB syndrome)
 AF044076 inhibitor of growth family, member 1
 AF047004 sarcosine dehydrogenase
 AF047190 sarcosine dehydrogenase
 AF051428 estrogen receptor 2 (ER beta)
 AF051851 supervillin
 AF059203 sterol O-acyltransferase 2
 AF060555 estrogen receptor 2 (ER beta)
 AF060925 hematological and neurological expressed 1
 AF063608 kinesin family member 5A
 AF064102 CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)
 AF064102 CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)
 AF064103 CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)
 AF064103 CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)
 AF067170 endosulfine alpha
 AF068220 ATPase, Ca⁺⁺ transporting, ubiquitous
 AF071594 Wolf-Hirschhorn syndrome candidate 1
 AF072506 Homo sapiens endogenous retrovirus W envelope protein precursor mRNA, complete
 AF073482 myotubularin related protein 7
 AF074382 inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma
 AF080216 B-cell CLL/lymphoma 11A (zinc finger protein)
 AF083389 Wolf-Hirschhorn syndrome candidate 1
 AF087138 ATP-binding cassette, sub-family C (CFTR/MRP), member 8
 AF087942 glycogenin
 AF090094 hypothetical protein FLJ14668
 AF091453 inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma
 AF096304 transmembrane 7 superfamily member 2
 AF101388 activating transcription factor 5
 AF106019 DEAD (Asp-Glu-Ala-Asp) box polypeptide 20
 AF112203 px19-like protein
 AF113007 DKFZP586A0522 protein
 AF113700 selenium binding protein 1
 AF113700 selenium binding protein 1
 AF115403 E74-like factor 5 (ets domain transcription factor)
 AF115765
 AF116624 predicted protein of HQ1146; Homo sapiens PRO1146 mRNA, complete cds.
 AF116639 chromosome 14 open reading frame 2
 AF116687 predicted protein of HQ2121; Homo sapiens PRO2121 mRNA, complete cds.
 AF118265 intronless; Homo sapiens orphan G protein-coupled receptor GPR44 (GPR44) gene
 AF119848 predicted protein of HQ1580; Homo sapiens PRO1580 mRNA, complete cds.
 AF119878 predicted protein of HQ2353; Homo sapiens PRO2353 mRNA, complete cds.
 AF119889 ARF protein
 AF120274 artemin
 AF120493 elastase 1, pancreatic
 AF124790 estrogen receptor 2 (ER beta)
 AF126008 A kinase (PRKA) anchor protein 13
 AF127481 normal LBC; an alternative 5' translation initiation start site exists upstream
 AF129266 sarcosine dehydrogenase
 AF129266 sarcosine dehydrogenase
 AF133207 heat shock 27kDa protein 8

AF133588 RAB36, member RAS oncogene family
AF142418 quaking homolog, KH domain RNA binding (mouse)
AF142421 quaking homolog, KH domain RNA binding (mouse)
AF142421 quaking homolog, KH domain RNA binding (mouse)
AF142573 CocoaCrisp
AF151022 brain protein 16
AF151045 melanoma associated antigen (mutated) 1
AF152318 protocadherin gamma subfamily C, 3
AF152524 protocadherin gamma subfamily C, 3
AF152524 protocadherin gamma subfamily C, 3
AF152528 protocadherin beta 18 pseudogene
AF153330 solute carrier family 19 (thiamine transporter), member 2
AF153607 px19-like protein
AF155097 NY-REN-7 antigen
AF157319 CGI-30 protein
AF157510 endosulfine alpha
AF161492 CGI-30 protein
AF162428 sarcosine dehydrogenase
AF163762 a disintegrin-like and metalloprotease (repolysin type) with thrombospondin
AF170911 solute carrier family 23 (nucleobase transporters), member 1
AF177473 transient receptor potential cation channel, subfamily M, member 5
AF190825 purinergic receptor P2X, ligand-gated ion channel, 2
AF192523 NPC1 (Niemann-Pick disease, type C1, gene)-like 1
AF192523 NPC1 (Niemann-Pick disease, type C1, gene)-like 1
AF198254 IGF-II mRNA-binding protein 1
AF201931 zinc finger, DHHC domain containing 4
AF208012 tumor protein D52-like 1
AF210455 solute carrier family 22 (organic anion transporter), member 7
AF210455 solute carrier family 22 (organic anion transporter), member 7
AF211189 calcium channel, voltage-dependent, alpha 11 subunit
AF214519 protein kinase, AMP-activated, gamma 3 non-catalytic subunit
AF216185 Ellis van Creveld syndrome
AF220122 tripartite motif-containing 10
AF220122 tripartite motif-containing 10
AF220123 tripartite motif-containing 10
AF222711 C-terminal binding protein 2
AF229061 NACHT, leucine rich repeat and PYD containing 1
AF229062 NACHT, leucine rich repeat and PYD containing 1
AF230401 promyelocytic leukemia
AF230401 promyelocytic leukemia
AF230409 promyelocytic leukemia
AF230410 promyelocytic leukemia
AF230411 promyelocytic leukemia
AF242768 mesenchymal stem cell protein DSC43
AF243433 emopamil binding related protein, delta8-delta7 sterol isomerase related protein
AF246238 growth factor receptor-bound protein 2
AF248965 CGI-30 protein
AF254868 nyctalopin
AF260426 purinergic receptor P2X, ligand-gated ion channel, 2
AF267865 nucleoporin 50kDa

AF272374 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AF272379 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AF272384 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AF277374 enhancer of polycomb homolog 1 (Drosophila)
 AF279893 actin related protein 2/3 complex, subunit 2, 34kDa
 AF279900 MCM7 minichromosome maintenance deficient 7 (*S. cerevisiae*)
 AF282167 EPS8-like 1
 AF285592 ferritin, heavy polypeptide-like 17
 AF287892 Siglec8 splice variant; splice variant of the sequence
 AF287892 Siglec8 splice variant; splice variant of the sequence
 AF293363 sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B
 AF294326 core-binding factor, beta subunit
 AF295039 calcium-binding tyrosine-(Y)-phosphorylation regulated (fibrousheathin 2)
 AF295773 ral guanine nucleotide dissociation stimulator
 AF300796 cytosolic sialic acid 9-O-acetylerase homolog
 AF303378 cytosolic sialic acid 9-O-acetylerase homolog
 AF305239 homeodomain interacting protein kinase 3
 AF308472 ATP-binding cassette, sub-family B (MDR/TAP), member 6
 AF310105 NACHT, leucine rich repeat and PYD containing 1
 AF311306 7TM; PSGR; Homo sapiens prostate specific G-protein coupled receptor gene
 AF312024 cadherin related 23
 AF312028 protein phosphatase 1, regulatory (inhibitor) subunit 12C
 AF318575 UPF3 regulator of nonsense transcripts homolog A (yeast)
 AF321125 DNA replication factor
 AF323729 oxysterol binding protein-like 7
 AF329403 cytoplasmic polyadenylation element binding protein 1
 AF329841 C1q and tumor necrosis factor related protein 5
 AF331844 sclerosteosis
 AF332009 nuclear transcription factor, X-box binding 1
 AF332192 regulatory factor X, 4 (influences HLA class II expression)
 AF332197 sine oculis homeobox homolog 2 (Drosophila)
 AF332558 BCL2 binding component 3
 AF336851 DEAD (Asp-Glu-Ala-Asp) box polypeptide 27
 AF342736 Bardet-Biedl syndrome 2
 AF422798 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AF465843 sterile alpha motif and leucine zipper containing kinase AZK
 AF486838 rho/rac guanine nucleotide exchange factor (GEF) 2
 AF493886 ARF protein
 AF513360 Homo sapiens enverin mRNA, complete cds.
 AFFX-hum_alu
 AI003763 hypothetical protein FLJ12681
 AI005043 Wiskott-Aldrich syndrome protein interacting protein
 AI016305 stonin 2
 AI022089 casein kinase 2, alpha prime polypeptide
 AI028474 hypothetical protein FLJ21438
 AI040329 neuro-oncological ventral antigen 2
 AI074983 a disintegrin-like and metalloprotease (repolysin type) with thrombospondin type 1
 AI082085 kinesin 2 60/70kDa
 AI088145 muscleblind-like 2 (Drosophila)
 AI091533 acidic repeat containing

AI114716 quaking homolog, KH domain RNA binding (mouse)
AI123426 CCR4-NOT transcription complex, subunit 2
AI129320 sterile alpha motif and leucine zipper containing kinase AZK
AI144394 zinc and ring finger protein 1
AI149639 fibrosin 1
AI160339 signal recognition particle receptor, B subunit
AI168198 C-terminal binding protein 1
AI186735 ribosomal protein L13
AI191118 peptidylprolyl isomerase A (cyclophilin A)
AI199541 zinc and ring finger protein 1
AI205309 methyl-CpG binding domain protein 1
AI219073 EPS8-like 1
AI243209 ADP-ribosylhydrolase like 1
AI246769 homeo box A9
AI247478 PTK2 protein tyrosine kinase 2
AI275938 tumor necrosis factor receptor superfamily, member 25
AI288424 single-minded homolog 2 (Drosophila)
AI291720 CGI-30 protein
AI343292 EPS8-like 1
AI343931 TAR (HIV) RNA binding protein 2
AI349089 homeo box A9
AI355279 nucleolar and coiled-body phosphoprotein 1
AI356398 zinc finger protein 36, C3H type-like 2
AI356774 leucine-rich repeats and immunoglobulin-like domains 1
AI369389 ribosomal protein L13
AI373205 nuclear receptor co-repressor 2
AI375694 postmeiotic segregation increased 2-like 6
AI377007 chromosome 22 open reading frame 1
AI379338 N-acylsphingosine amidohydrolase (acid ceramidase) 1
AI380156 hypothetical protein LOC285458
AI381472 transcription factor 20 (AR1)
AI400292 zinc finger, DHHC domain containing 4
AI401287 zinc finger protein-like 1
AI418253 postmeiotic segregation increased 2-like 5
AI421559 ral guanine nucleotide dissociation stimulator
AI421972 nuclear receptor binding SET domain protein 1
AI433419 Down syndrome cell adhesion molecule like 1
AI435747 th53f01.x1 NCI_CGAP_Br17 Homo sapiens cDNA clone IMAGE:2122009
AI459136 KIAA0685
AI475164 sterile alpha motif and leucine zipper containing kinase AZK
AI479073 egl nine homolog 2 (C. elegans)
AI479104 F-box and leucine-rich repeat protein 12
AI498144 chromosome 20 open reading frame 194
AI499095 nuclear receptor co-repressor 2
AI521646 to66a06.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2183218
AI523593 A kinase (PRKA) anchor protein 13
AI554467 ribosomal protein L13
AI560217 interleukin 17 receptor C
AI583181 hepatitis delta antigen-interacting protein A
AI613383 eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)

AI621079 peroxisomal farnesylated protein
AI623202 PR domain containing 16
AI635302 KIAA1061 protein
AI638208 stathmin-like 4
AI640492 KCCR13L
AI652861 CUG triplet repeat, RNA binding protein 2
AI653267 growth factor receptor-bound protein 2
AI655903 G protein-coupled receptor 124
AI659683 ankyrin 1, erythrocytic
AI660497 hypothetical protein FLJ20422
AI674926 A kinase (PRKA) anchor protein 13
AI676103 protein x 0001
AI680721 zinc and ring finger protein 1
AI681702 hypothetical protein FLJ12442
AI681702 hypothetical protein FLJ12442
AI692870 enabled homolog (Drosophila)
AI693985 forkhead box A2
AI694332 ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)
AI697108 mucin 5, subtypes A and C, tracheobronchial/gastric
AI701430 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
AI703476 G protein-coupled receptor 27
AI708767 peptidylprolyl isomerase A (cyclophilin A)
AI732599 RNA polymerase I associated factor 53
AI733480 orthopedia homolog (Drosophila)
AI735692 natural cytotoxicity triggering receptor 3
AI738987 hypothetical protein MGC3047
AI740763 protein-O-mannosyltransferase 2
AI742570 hypothetical protein FLJ13710
AI742586 dehydrogenase/reductase (SDR family) member 10
AI742931 protein phosphatase 1, regulatory (inhibitor) subunit 16A
AI743612 KIAA1238 protein
AI744658 D2-2 mRNA, 3'UTR
AI760919 KIAA1238 protein
AI767884 zinc finger protein-like 1
AI767884 zinc finger protein-like 1
AI770166 Wolf-Hirschhorn syndrome candidate 1
AI770171 nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
AI795908 pleckstrin homology-like domain, family A, member 1
AI798823 pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1)
AI805082 olfactory receptor, family 51, subfamily E, member 2
AI806338 T-box 3 (ulnar mammary syndrome)
AI820043 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
AI820043 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
AI821777 Transcribed sequences
AI824954 SRY (sex determining region Y)-box 3
AI828004 HLA-B associated transcript 8
AI858000 hypothetical protein FLJ10385
AI860326 melanoma associated antigen (mutated) 1
AI867445 forkhead box D3
AI889160 Cdk5 and Abl enzyme substrate 1

AI912275 B-cell CLL/lymphoma 11A (zinc finger protein)
 AI916555 coatomer protein complex, subunit zeta 2
 AI923492 major histocompatibility complex, class I, A
 AI923713 cyclin M3
 AI934125 bicaudal D homolog 2 (Drosophila)
 AI934569 N-acylsphingosine amidohydrolase (acid ceramidase) 1
 AI934753 tweety homolog 3 (Drosophila)
 AI935308 fucosidase, alpha-L- 1, tissue
 AI939402 homeo box (H6 family) 1
 AI953599 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 AI953822 ribosomal protein, large, P0
 AI954654 DiGeorge syndrome critical region gene 14
 AI955239 oxysterol binding protein-like 7
 AI963713 enabled homolog (Drosophila)
 AI972623 Williams-Beuren syndrome chromosome region 17
 AI990766 fragile X mental retardation, autosomal homolog 1
 AI991669 hypothetical protein MGC3047
 AI991703 hypothetical protein FLJ31528
 AI991996 ws43e12.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2499982
 AI992095 mesenchymal stem cell protein DSC43
 AI992271 hypothetical protein FLJ20014
 AJ010395 Homo sapiens DKC1 gene, exons 1 to 11.
 AJ011712
 AJ012502 ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
 AJ012502 ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
 AK000003 PC2 (positive cofactor 2, multiprotein complex) glutamine/Q-rich-associated protein
 AK000110 ubiquitin specific protease 31
 AK000582 hypothetical protein FLJ13710
 AK000586 chromosome 20 open reading frame 43
 AK002166 A kinase (PRKA) anchor protein 11
 AK021715 DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
 AK022014 A kinase (PRKA) anchor protein 13
 AK022014 A kinase (PRKA) anchor protein 13
 AK022023 KIAA0889 protein
 AK022115 hypothetical protein FLJ12057
 AK022669 KIAA0685
 AK023092 ubiquitin specific protease 49
 AK023804 protein-O-mannosyltransferase 2
 AK023850 protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF)
 AK023900 CMP-NeuAC:(beta)-N-acetylgalactosaminide (alpha)2,6-sialyltransferase member VI
 AK024117 Homo sapiens cDNA FLJ14055 fis, clone HEMBB1000264
 AK024132 hypothetical protein FLJ13710
 AK024386 glyoxylate reductase/hydroxypyruvate reductase
 AK024437 mitogen-activated protein kinase 8 interacting protein 3
 AK024467 hypothetical gene FLJ00060
 AK024467 hypothetical gene FLJ00060
 AK024649 leucine rich repeat containing 8
 AK024714 kinesin 2 60/70kDa
 AK024945 Homo sapiens cDNA: FLJ21292 fis, clone COL01969.
 AK024945 Homo sapiens cDNA: FLJ21292 fis, clone COL01969.

AK024968 breast carcinoma amplified sequence 1
 AK024968 breast carcinoma amplified sequence 1
 AK024982 transcriptional adaptor 3 (NGG1 homolog, yeast)-like
 AK025108 enabled homolog (Drosophila)
 AK025352 microtubule associated serine/threonine kinase 2
 AK025352 microtubule associated serine/threonine kinase 2
 AK025352 microtubule associated serine/threonine kinase 2
 AK025366 phospholipase C, epsilon 1
 AK025366 phospholipase C, epsilon 1
 AK025394 Ellis van Creveld syndrome
 AK025419 Spi-B transcription factor (Spi-1/PU.1 related)
 AK025573 Homo sapiens cDNA: FLJ21920 fis, clone HEP04049.
 AK025627 Cdk5 and Abl enzyme substrate 1
 AK025720 VAMP (vesicle-associated membrane protein)-associated protein B and C
 AK025898 low density lipoprotein receptor-related protein 10
 AK025940 Down syndrome cell adhesion molecule like 1
 AK026105 TBC1 domain family, member 2
 AK026118 chromosome 20 open reading frame 43
 AK026181 pleckstrin homology-like domain, family A, member 1
 AK026188 protocadherin gamma subfamily C, 3
 AK026340 ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)
 AK026666 hypothetical protein MGC3207
 AK026752 glyoxylate reductase/hydroxypyruvate reductase
 AK055684 KIAA1061 protein
 AK074510 pleckstrin homology-like domain, family A, member 1
 AK091796 hypothetical protein MGC13045
 AK093331 KIAA1061 protein
 AK095055 ribosomal protein S9
 AK097281 dynein, cytoplasmic, light polypeptide 2A
 AL008583
 AL021395
 AL031588
 AL031651
 AL031718
 AL031718
 AL031781
 AL031781
 AL035682
 AL036344 nucleoporin 50kDa
 AL037339 PTK2 protein tyrosine kinase 2
 AL037401 nuclear receptor subfamily 2, group F, member 2
 AL039795 TAR DNA binding protein
 AL039870 hypothetical protein FLJ00058
 AL040708 ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)
 AL041451 Hermansky-Pudlak syndrome 4
 AL049688 calcium/calmodulin-dependent protein kinase IG
 AL050008 breast cancer metastasis suppressor 1
 AL050358 solute carrier family 26, member 10
 AL050388 superoxide dismutase 2, mitochondrial
 AL079281 neuronal PAS domain protein 3

AL110138 KIAA1211 protein
 AL110249 chromosome 20 open reading frame 194
 AL117487 transcriptional adaptor 3 (NGG1 homolog, yeast)-like
 AL117499 chromosome 19 open reading frame 13
 AL117549 myotubularin related protein 7
 AL117549 myotubularin related protein 7
 AL117586 Homo sapiens mRNA; cDNA DKFZp434M148 (from clone DKFZp434M148).
 AL117630 NY-REN-7 antigen
 AL121873
 AL121994 Human DNA sequence from clone RP4-781L3 on chromosome 1p34.3-36.11
 AL122081 cadherin related 23
 AL122081 cadherin related 23
 AL133427 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 261172.
 AL133598 chimeric cDNA?; Homo sapiens mRNA; cDNA DKFZp434H052
 AL133646 glucocorticoid modulatory element binding protein 2
 AL136552 sema domain, immunoglobulin domain (Ig), transmembrane domain TM
 AL136562 coiled-coil domain containing 3
 AL136595 MYB binding protein (P160) 1a
 AL136597 SBBI26 protein
 AL136774 hypothetical protein DKFZp434P0316
 AL136811 F-box only protein 24
 AL136840 MCM10 minichromosome maintenance deficient 10 (*S. cerevisiae*)
 AL136903 zinc and ring finger protein 1
 AL136903 zinc and ring finger protein 1
 AL137316 KIAA1609 protein
 AL137653 C-terminal binding protein 1
 AL137674 CCR4-NOT transcription complex, subunit 2
 AL137713 DiGeorge syndrome critical region gene 14
 AL137713 DiGeorge syndrome critical region gene 14
 AL162057 neuronal pentraxin receptor
 AL359931 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 288936.
 AL360266 melanoma associated antigen (mutated) 1
 AL390183 LOC389632 (LOC389632), mRNA
 AL512686 guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O
 protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) 2; Homo sapiens mRNA;
 cDNA DKFZp667H046 (from clone DKFZp667H046).
 AL512703 Homo sapiens mRNA; cDNA DKFZp667N057 (from clone DKFZp667N057).
 AL513917 solute carrier family 16 (monocarboxylic acid transporters), member 3
 AL515916 hypothetical protein FLJ12681
 AL520719 CGI-07 protein
 AL526448 nuclear receptor binding SET domain protein 1
 AL529396 hypothetical protein MGC3262
 AL531750 collagen, type VI, alpha 2
 AL535113 phospholipase C, beta 4
 AL550657 basigin (OK blood group)
 AL554245 nuclear receptor subfamily 2, group F, member 2
 AL558164 hypothetical protein FLJ10922
 AL562298 hypothetical protein LOC285458
 AL563476 AL563476 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
 cDNA clone CS0DD006YJ02 3-PRIME, mRNA sequence.

AL565715 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF007YG22 3-PRIME, mRNA sequence.
 AL566294 muscleblind-like 2 (Drosophila)
 AL568804 AL568804 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YN15 3-PRIME, mRNA sequence.
 AL571375 AL571375 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI009YH10 3-PRIME, mRNA sequence.
 AL575337 RAB11B, member RAS oncogene family
 AL578668 synaptotagmin binding, cytoplasmic RNA interacting protein
 AL832400 NACHT, leucine rich repeat and PYD containing 1
 AL832400 NACHT, leucine rich repeat and PYD containing 1
 AL832902 Homo sapiens mRNA; cDNA DKFZp762C0813 (from clone DKFZp762C0813).
 AL833178 Similar to Serine/threonine protein kinase PRKX (Protein kinase PKX1) (LOC390644), mRNA
 AL833178 Similar to Serine/threonine protein kinase PRKX (Protein kinase PKX1) (LOC390644), mRNA
 AL833290 nuclear receptor co-repressor 2
 AL834350 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
 AU145723 transmembrane protein 2
 AU146440 Transcribed sequence with weak similarity to protein ref:NP_060190.1 (H.sapiens) hypothetical protein FLJ20234 [Homo sapiens]
 AU150065 mesenchymal stem cell protein DSC43
 AU154358 KIAA0217 protein
 AV686550 hypothetical protein FLJ12681
 AV693403 AV693403 GKC Homo sapiens cDNA clone GKCF09 5', mRNA sequence.
 AV700721 ADP-ribosylation factor 6
 AV703465 AV703465 ADB Homo sapiens cDNA clone ADBCHG08 5', mRNA sequence.
 AV713720 AV713720 DCB Homo sapiens cDNA clone DCBBJA12 5', mRNA sequence.
 AV714029 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AV724266 ankyrin repeat domain 10
 AV726322 endosulfine alpha
 AV726322 endosulfine alpha
 AW002079 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 AW003512 arachidonate lipoxygenase 3
 AW003893 quaking homolog, KH domain RNA binding (mouse)
 AW005237 chromosome 9 open reading frame 25
 AW006288 hypothetical protein MGC10986
 AW014373 mesenchymal stem cell protein DSC43
 AW015263 zinc finger, FYVE domain containing 19
 AW016039 calcium/calmodulin-dependent protein kinase IG
 AW016304 dishevelled, dsh homolog 2 (Drosophila)
 AW024420 dual specificity phosphatase 1
 AW024870 Wolf-Hirschhorn syndrome candidate 1
 AW025687 hypothetical protein FLJ39501
 AW044553 ankyrin 1, erythrocytic
 AW057545 egl nine homolog 2 (C. elegans)
 AW058622 Wiskott-Aldrich syndrome protein interacting protein
 AW069181 cr43e01.x1 Human bone marrow stromal cells Homo sapiens cDNA clone HBMSC_cr43e01 3', mRNA sequence.
 AW071829 eukaryotic translation initiation factor 2C, 1

AW075105 DNA replication factor
 AW080549 fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group included)
 AW084791 zinc finger, DHHC domain containing 12
 AW102570 KIAA1238 protein
 AW102637 neurofascin
 AW118878 supervillin
 AW118878 supervillin
 AW119113 thrombomodulin
 AW134523 fibrosin 1
 AW135012 KIAA1238 protein
 AW167553 ubiquitin-conjugating enzyme E2, J2 (UBC6 homolog, yeast)
 AW188009 ATPase, Ca⁺⁺ transporting, ubiquitous
 AW193511 HMBA-inducible
 AW193656 inhibitor of growth family, member 1
 AW195920 postmeiotic segregation increased 2-like 5
 AW204766 A kinase (PRKA) anchor protein 13
 AW269335 endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2
 AW269415 oxysterol binding protein-like 10
 AW269836 zinc finger, RAN-binding domain containing 1
 AW274658 inhibitor of growth family, member 1
 AW276522 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
 AW276914 muscleblind-like 2 (Drosophila)
 AW291083 VAMP (vesicle-associated membrane protein)-associated protein B and C
 AW291398 nuclear transcription factor, X-box binding 1
 AW291829 homeodomain interacting protein kinase 3
 AW293282 leucine-rich repeats and immunoglobulin-like domains 1
 AW299534 rho/rac guanine nucleotide exchange factor (GEF) 2
 AW340508 KIAA1061 protein
 AW375186 claudin 23
 AW379790 RC3-HT0253-181099-011-c03 HT0253 Homo sapiens cDNA, mRNA sequence.
 AW409794 KIAA1238 protein
 AW449022 pyridoxal (pyridoxine, vitamin B6) kinase
 AW450345 hypothetical protein MGC16664
 AW450386 WD repeat domain 34
 AW451036 chromosome 20 open reading frame 43
 Transcribed sequence with strong similarity to protein pir:A56126 (H.sapiens) A56126
 AW468717 peroxisomal targeting signal 1 receptor - human
 AW469523 diacylglycerol O-acyltransferase homolog 2 (mouse)
 AW500180 hypothetical protein MGC46719
 AW514174 KIAA0217 protein
 AW514267 NY-REN-7 antigen
 AW574664 ribosomal protein L13
 AW575379 protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
 AW575773 A kinase (PRKA) anchor protein 13
 AW589982 nucleoporin 50kDa
 AW593269 KIAA0685
 AW593467 oxysterol binding protein-like 10
 AW593859 chromosome 19 open reading frame 13
 AW628686 KIAA0217 protein
 AW629423 nucleoporin 50kDa

AW665096 HMBA-inducible
 AW771015 phospholipase C, epsilon 1
 AW772123 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
 hi91a07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3009300 3', mRNA
 sequence.
 AW872374 supervillin
 AW872377 supervillin
 AY029208 collagen, type VI, alpha 2
 AY151049 cytochrome P450, family 4, subfamily B, polypeptide 1
 BC000069 retinoic acid receptor responder (tazarotene induced) 2
 BC000080 promyelocytic leukemia
 BC000154 thyroid hormone receptor interactor 6
 BC000154 thyroid hormone receptor interactor 6
 BC000265
 BC000436 endosulfine alpha
 BC000496 peroxisomal farnesylated protein
 BC000767 THAP domain containing 4
 BC000787 TRIAD3 protein
 BC001007 protein kinase, AMP-activated, beta 1 non-catalytic subunit
 BC001186 protocadherin beta 5
 BC001703 hypothetical protein MGC3207
 BC001727 ankyrin repeat domain 10
 BC001788 SPPL2b
 BC001800 orthopedia homolog (Drosophila)
 BC001865
 BC002327
 BC002471 complexin 1
 BC002486 C-terminal binding protein 2
 BC002635 colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)

 BC002665 proteolipid protein 1 (Pelizaeus-Merzbacher disease, spastic paraplegia 2, uncomplicated)
 BC002671 dual specificity phosphatase 4
 BC002769 chromosome 20 open reading frame 43
 BC002844 nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
 BC002918 carbohydrate (chondroitin 4) sulfotransferase 12
 BC002983 likely ortholog of mouse gene rich cluster, C9 gene
 BC003080 HSPC171 protein
 BC003105 protein tyrosine phosphatase type IVA, member 3
 BC003179 BENE protein
 BC003381 KIAA0217 protein
 BC003551 transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)
 BC003655 ribosomal protein, large, P0
 BC004242 heterogeneous nuclear ribonucleoprotein U-like 1
 BC004247 ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
 BC004492 DKFZP586A0522 protein
 BC004865 cleft lip and palate associated transmembrane protein 1
 BC004877 uracil-DNA glycosylase 2
 BC004907 EPS8-like 1
 BC004936 stearoyl-CoA desaturase 4
 BC004936 stearoyl-CoA desaturase 4
 BC004954 ribosomal protein L13

BC005020	peptidylprolyl isomerase F (cyclophilin F)
BC005050	nicolin 1
BC005174	activating transcription factor 5
BC005863	ribosomal protein, large, P0
BC005876	ATPase, H ⁺ transporting, lysosomal 21kDa, V0 subunit c"
BC005982	peptidylprolyl isomerase A (cyclophilin A)
BC006104	RIO kinase 1 (yeast)
BC006129	hypothetical protein MGC13045
BC006134	hypothetical protein MGC13053
BC006252	solute carrier family 6 (neurotransmitter transporter, taurine), member 6
BC006283	dehydrogenase/reductase (SDR family) member 10
BC006294	dehydrogenase/reductase (SDR family) member 10
BC006314	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)
BC006411	interleukin 17 receptor C
	synonyms: LIG, HYPG; Homo sapiens huntingtin interacting protein 2, mRNA (cDNA clone IMAGE:3946309), complete cds.
BC006419	IMAGE:3946309), complete cds.
BC006439	protocadherin gamma subfamily C, 3
BC007085	serine palmitoyltransferase, long chain base subunit 1
BC007223	dynein, cytoplasmic, light polypeptide 2A
BC008000	nuclear receptor co-repressor 2
BC008840	hypothetical protein FLJ13710
BC008840	hypothetical protein FLJ13710
BC009036	chromosome 22 open reading frame 1
BC009757	RNA polymerase I associated factor 53
BC010094	phosphatidylinositol glycan, class Q
BC010943	oncostatin M receptor
BC012487	
BC014215	sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)
BC014215	sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)
BC015129	chromosome 22 open reading frame 1
BC015667	kelch repeat and BTB (POZ) domain containing 9
BC015667	kelch repeat and BTB (POZ) domain containing 9
BC016183	protein phosphatase 2, regulatory subunit B (B56), gamma isoform
BC016183	protein phosphatase 2, regulatory subunit B (B56), gamma isoform
BC016740	chromosome 22 open reading frame 1
BC016828	N-acylsphingosine amidohydrolase (acid ceramidase) 1
BC017893	chromosome 22 open reading frame 1
BC017963	solute carrier family 22 (organic anion transporter), member 7
BC020869	ARF protein
BC021928	chromosome 22 open reading frame 1
BC022013	Similar to FKBP6 protein (LOC392732), mRNA
	Homo sapiens KIAA0685, mRNA (cDNA clone IMAGE:4250008), with apparent retained intron.
BC022346	intron.
BC025707	potassium large conductance calcium-activated channel, subfamily M, beta member 1
BC025994	solute carrier family 4, sodium bicarbonate cotransporter, member 8
BC027465	RAB36, member RAS oncogene family
BC028064	RAB39, member RAS oncogene family
	Homo sapiens NY-REN-7 antigen, mRNA (cDNA clone IMAGE:5266126), with apparent retained intron.
BC028606	intron.
BC029446	translocase of inner mitochondrial membrane 17 homolog B (yeast)

BC031044 DnaJ (Hsp40) homolog, subfamily A, member 4
 BC031044 DnaJ (Hsp40) homolog, subfamily A, member 4
 BC031076 protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
 BC032124 bromodomain containing 3
 BC034024 chromosome 22 open reading frame 1
 BC034784 adenovirus 5 E1A binding protein
 BC034784 adenovirus 5 E1A binding protein
 BC036391 CUG triplet repeat, RNA binding protein 2
 BC036896 Ellis van Creveld syndrome
 BC037823 hypothetical protein DKFZp564O1664
 BC039181 hypothetical protein LOC158833
 BC040275 Ras protein-specific guanine nucleotide-releasing factor 1
 BC040472 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 BC040472 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 BC040966 cytosolic sialic acid 9-O-acetyesterase homolog
 BC040966 cytosolic sialic acid 9-O-acetyesterase homolog
 BC043202 PTK2 protein tyrosine kinase 2
 BE018269 solute carrier family 13 (sodium/sulfate symporters), member 4
 BE046521 cut-like 1, CCAAT displacement protein (Drosophila)
 BE138888 ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)
 BE218313 quaking homolog, KH domain RNA binding (mouse)
 BE220265 wingless-type MMTV integration site family, member 9A
 BE250417 adenovirus 5 E1A binding protein
 BE254644 aryl hydrocarbon receptor interacting protein-like 1
 BE273906 hypothetical protein FLJ31528
 BE299671 600944342T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960218 3', mRNA sequence.
 BE328496 hs98f09.x1 NCI_CGAP_Kid13 Homo sapiens cDNA clone IMAGE:3145289 3', mRNA sequence.
 BE348997 ribosomal protein S9
 BE350315 receptor tyrosine kinase-like orphan receptor 2
 BE379761 stonin 2
 BE463815 hypothetical protein MGC3047
 BE465475 kelch repeat and BTB (POZ) domain containing 9
 BE501428 protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 3
 BE504672 chromosome 13 open reading frame 7
 BE551998 homeobox protein GSH-2
 BE644818 SPPL2b
 BE645018 ubiquitin specific protease 31
 BE669782 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 BE670563 guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O
 BE671156 microtubule-associated protein, RP/EB family, member 2
 BE673925 Ras-GTPase-activating protein SH3-domain-binding protein
 BE676642 cleavage and polyadenylation specific factor 1, 160kDa
 BE731738 Transcribed sequence with strong similarity to protein pdb:2RMA (H.sapiens) S Chain S, Cyclophilin A
 BE778059 zinc finger protein 67 homolog (mouse)
 BE783723 malignant fibrous histiocytoma amplified sequence 1
 BE786598 microtubule associated serine/threonine kinase 2

601479952F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3882661 5', mRNA
 BE789211 sequence.
 BE793789 Wolf-Hirschhorn syndrome candidate 1
 BE795104 KCCR13L
 BE795104 KCCR13L
 BE855799 KIAA1211 protein
 BE856607 nischarin
 BE962920 ubiquitin-conjugating enzyme E2, J2 (UBC6 homolog, yeast)
 BE966372 hepatitis delta antigen-interacting protein A
 BE967532 midline 1 (Opitz/BBB syndrome)
 BE974210 KIAA0217 protein
 BF000697 supervillin
 BF030331 leucine rich repeat containing 8
 BF055366 endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2
 Transcribed sequence with strong similarity to protein ref:NP_149018.1 (H.sapiens)
 BF056991 leishmanolysin-2 [Homo sapiens]
 BF057352 myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)
 BF109740 Transcribed sequences
 BF110053 claudin 23
 BF110053 claudin 23
 BF111870 Wolf-Hirschhorn syndrome candidate 1
 BF112057 interleukin 17 receptor C
 BF115826 Williams-Beuren syndrome chromosome region 17
 a disintegrin and metalloproteinase domain 17 (tumor necrosis factor, alpha, converting
 enzyme)
 BF115870 DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
 BF129093 601843510F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4064187 5', mRNA
 sequence.
 BF185219 calpain 5
 BF195709 kinesin family member 5A
 BF196255 interleukin 17 receptor C
 BF196320 neuronal PAS domain protein 3
 BF196935 A kinase (PRKA) anchor protein 13
 BF222823 enabled homolog (Drosophila)
 BF223300 heat shock 27kDa protein family, member 7 (cardiovascular)
 BF306676 602034795F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4182742 5', mRNA
 sequence.
 BF337195 interferon induced transmembrane protein 3 (1-8U)
 BF338947 602038688F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4186629 5', mRNA
 sequence.
 BF339357 ATPase, H⁺ transporting, lysosomal 13kDa, V1 subunit G isoform 2
 BF340635 zinc and ring finger protein 1
 BF432625 tau tubulin kinase 1
 BF435053 mitogen-activated protein kinase 8 interacting protein 3
 BF445013 branched chain keto acid dehydrogenase E1, alpha polypeptide (maple syrup urine
 disease)
 BF446281 Wiskott-Aldrich syndrome protein interacting protein
 BF446719 kallikrein 3, (prostate specific antigen)
 BF476883 chromosome 14 open reading frame 2
 BF507638 G protein-coupled receptor 124
 BF511315

BF530348	coagulation factor XII (Hageman factor)
BF568780	bone marrow stromal cell-derived ubiquitin-like
BF589232	KIAA0217 protein
BF589368	zinc finger protein 36, C3H type-like 2
BF593594	kinesin 2 60/70kDa
BF593867	hypothetical protein FLJ37078
BF724216	heterogeneous nuclear ribonucleoprotein U-like 1
BF739959	malignant fibrous histiocytoma amplified sequence 1
BF940735	protein kinase C and casein kinase substrate in neurons 3
BF965437	KIAA0153 protein
BF977837	KIAA0527 protein
BF983948	signal recognition particle receptor, B subunit
BF984434	C-terminal binding protein 1
BG025078	fragile X mental retardation, autosomal homolog 1
BG034239	actin related protein 2/3 complex, subunit 2, 34kDa
BG054966	chromosome 19 open reading frame 13
BG055137	T-box 3 (ulnar mammary syndrome)
BG056174	dynein, cytoplasmic, light polypeptide 2A
BG105204	CCR4-NOT transcription complex, subunit 2
BG105204	CCR4-NOT transcription complex, subunit 2
BG122789	602351740F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4450040 5', mRNA sequence.
BG150494	RAB28, member RAS oncogene family
BG163478	hypothetical protein FLJ13710
BG164064	ubiquitin-conjugating enzyme E2 variant 1
BG171548	NEDD8-conjugating enzyme
BG251266	hepatitis delta antigen-interacting protein A
BG252318	602365906F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4474128 5', mRNA sequence.
BG290819	hypothetical protein MGC2615
BG291007	CGI-07 protein
BG292367	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
BG338983	WD repeat domain 34
BG396520	ribosomal protein L13
BG420865	zinc finger protein-like 1
BG474541	signal recognition particle receptor ('docking protein')
BG489075	zinc finger protein 524
BG500067	602545874F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4668234 5', mRNA sequence.
BG722779	sirtuin (silent mating type information regulation 2 homolog) 2 (S. cerevisiae)
BM193618	B-cell CLL/lymphoma 11A (zinc finger protein)
BQ021146	baculoviral IAP repeat-containing 5 (survivin)
BU189724	hypothetical protein FLJ13231
CA431092	UI-H-FL1-bge-p-03-0-UI.s1 NCI_CGAP_FL1 Homo sapiens cDNA clone UI-H-FL1-bge-p-03-0-UI 3', mRNA sequence.
CA431092	UI-H-FL1-bge-p-03-0-UI.s1 NCI_CGAP_FL1 Homo sapiens cDNA clone UI-H-FL1-bge-p-03-0-UI 3', mRNA sequence.
D13515	glutamate receptor, ionotropic, N-methyl D-aspartate 1
D13902	peptide YY
D21262	nucleolar and coiled-body phosphoprotein 1

D25216	leucine rich repeat containing 14
D28468	D site of albumin promoter (albumin D-box) binding protein ORF; Homo sapiens FSHD gene for facioscapulohumeral muscular dystrophy, complete cds, 4Z4 tandem repeat unit.
D38024	MCM7 minichromosome maintenance deficient 7 (<i>S. cerevisiae</i>)
D55716	aquaporin 4
D63412	aquaporin 4
D63412	aquaporin 4
D83703	peroxisomal biogenesis factor 6
D83782	SREBP CLEAVAGE-ACTIVATING PROTEIN
D86988	regulator of nonsense transcripts 1
D87012	putative; Homo sapiens immunoglobulin lambda gene locus DNA, clone:61D6.
D87914	ornithine decarboxylase antizyme 1
D88435	cyclin G associated kinase
D88435	cyclin G associated kinase
D89324	Homo sapiens DNA for alpha (1,3/1,4) fucosyltransferase, complete cds. yl74b03.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:43734 3', mRNA sequence.
H04790	sterile alpha motif and leucine zipper containing kinase AZK
H28667	yr33g06.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:207130 3' similar to gb:X58521 NUCLEAR PORE GLYCOPROTEIN P62 (HUMAN);, mRNA sequence.
H48515	fragile X mental retardation, autosomal homolog 1
H48840	ficolin (collagen/fibrinogen domain containing lectin) 2 (hucolin)
H62781	IGF-II mRNA-binding protein 1
H93038	IGF-II mRNA-binding protein 1
J02871	cytochrome P450, family 4, subfamily B, polypeptide 1
K03199	tumor protein p53 (Li-Fraumeni syndrome)
L05666	glutamate receptor, ionotropic, N-methyl D-aspartate 1
L08599	cadherin 1, type 1, E-cadherin (epithelial)
L13720	growth arrest-specific 6
L20469	dopamine receptor D3
L22179	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
L26584	Ras protein-specific guanine nucleotide-releasing factor 1
L29349	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)
L29511	growth factor receptor-bound protein 2
L35853	sarcoglycan, alpha (50kDa dystrophin-associated glycoprotein)
L42374	
L42375	protein phosphatase 2, regulatory subunit B (B56), gamma isoform
L77566	DiGeorge syndrome critical region gene 14
L78207	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
M15465	pyruvate kinase, liver and RBC
M17863	insulin-like growth factor 2 (somatomedin A)
M28880	ankyrin 1, erythrocytic
M32221	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)
M55683	matrilin 1, cartilage matrix protein
M57763	ADP-ribosylation factor 6
M64445	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)
M64497	nuclear receptor subfamily 2, group F, member 2
M79462	promyelocytic leukemia
M90360	A kinase (PRKA) anchor protein 13
M92301	calcium channel, voltage-dependent, beta 1 subunit

M94065	dihydroorotate dehydrogenase
M94893	testis specific protein, Y-linked
M94893	testis specific protein, Y-linked
M96739	nescient helix loop helix 1
M98478	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)
M98825	fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
N23018	C-terminal binding protein 2
N25546	Transcribed sequences
N29712	yw78e05.s1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:258368 3', mRNA sequence.
N29712	yw78e05.s1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:258368 3', mRNA sequence.
N36839	CUG triplet repeat, RNA binding protein 2
N36842	UPF3 regulator of nonsense transcripts homolog A (yeast)
N45111	bicaudal D homolog 2 (Drosophila)
N45231	DnaJ (Hsp40) homolog, subfamily A, member 4
N79572	zb13b12.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:301919 3', mRNA sequence.
N90866	CDW52 antigen (CAMPATH-1 antigen)
N91109	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)
N91109	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)
NM_000037	ankyrin 1, erythrocytic
NM_000050	argininosuccinate synthetase
NM_000053	ATPase, Cu ⁺⁺ transporting, beta polypeptide (Wilson disease)
NM_000079	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)
NM_000147	fucosidase, alpha-L- 1, tissue
NM_000273	G protein-coupled receptor 143
NM_000275	oculocutaneous albinism II (pink-eye dilution homolog, mouse)
NM_000287	peroxisomal biogenesis factor 6
NM_000298	pyruvate kinase, liver and RBC
NM_000319	peroxisome receptor 1
NM_000361	thrombomodulin
NM_000361	thrombomodulin
NM_000363	troponin I, cardiac
NM_000381	midline 1 (Opitz/BBB syndrome)
NM_000481	aminomethyltransferase (glycine cleavage system protein T)
NM_000482	apolipoprotein A-IV
NM_000492	cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)
NM_000505	coagulation factor XII (Hageman factor)
NM_000510	follicle stimulating hormone, beta polypeptide
NM_000546	tumor protein p53 (Li-Fraumeni syndrome)
NM_000612	insulin-like growth factor 2 (somatomedin A)
NM_000709	branched chain keto acid dehydrogenase E1, alpha polypeptide (maple syrup urine disease)
NM_000711	
NM_000718	calcium channel, voltage-dependent, L type, alpha 1B subunit
NM_000723	calcium channel, voltage-dependent, beta 1 subunit

NM_000727 calcium channel, voltage-dependent, gamma subunit 1
 NM_000796 dopamine receptor D3
 NM_000820 growth arrest-specific 6
 NM_000831 glutamate receptor, ionotropic, kainate 3
 NM_000883 IMP (inosine monophosphate) dehydrogenase 1
 NM_000933 phospholipase C, beta 4
 NM_000965 retinoic acid receptor, beta
 NM_000975 ribosomal protein L11
 NM_001002 ribosomal protein, large, P0
 NM_001012 ribosomal protein S8
 NM_001013 ribosomal protein S9
 NM_001037 sodium channel, voltage-gated, type I, beta
 NM_001168 baculoviral IAP repeat-containing 5 (survivin)
 NM_001262 cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)
 NM_001294 cleft lip and palate associated transmembrane protein 1
 NM_001299 calponin 1, basic, smooth muscle
 NM_001328 C-terminal binding protein 1
 NM_001329 C-terminal binding protein 2
 NM_001363 dyskeratosis congenita 1, dyskerin
 NM_001386 dihydropyrimidinase-like 2
 NM_001394 dual specificity phosphatase 4
 NM_001395 dual specificity phosphatase 9
 NM_001401 endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2
 NM_001422 E74-like factor 5 (ets domain transcription factor)
 BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (*S. cerevisiae*)
 NM_001519
 NM_001540 heat shock 27kDa protein 1
 NM_001554 cysteine-rich, angiogenic inducer, 61
 NM_001585 chromosome 22 open reading frame 1
 NM_001648 kallikrein 3, (prostate specific antigen)
 NM_001663 ADP-ribosylation factor 6
 NM_001701 bile acid Coenzyme A: amino acid N-acyltransferase (glycine N-choloyltransferase)
 NM_001702 brain-specific angiogenesis inhibitor 1
 NM_001755 core-binding factor, beta subunit
 NM_001803 CDW52 antigen (CAMPATH-1 antigen)
 NM_001842 ciliary neurotrophic factor receptor
 NM_001896 casein kinase 2, alpha prime polypeptide
 NM_001913 cut-like 1, CCAAT displacement protein (*Drosophila*)
 NM_001928 D component of complement (adipsin)
 NM_001960 eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
 NM_002103 glycogen synthase 1 (muscle)
 NM_002251 potassium voltage-gated channel, delayed-rectifier, subfamily S, member 1
 NM_002279 keratin, hair, acidic, 3B
 NM_002283 keratin, hair, basic, 5
 NM_002379 matrilin 1, cartilage matrix protein
 NM_002415 macrophage migration inhibitory factor (glycosylation-inhibiting factor)
 NM_002431 menage a trois 1 (CAK assembly factor)
 NM_002502 nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
 NM_002504 nuclear transcription factor, X-box binding 1
 NM_002516 neuro-oncological ventral antigen 2

NM_002563 purinergic receptor P2Y, G-protein coupled, 1
 NM_002587 protocadherin 1 (cadherin-like 1)
 NM_002588 protocadherin gamma subfamily C, 3
 NM_002613 3-phosphoinositide dependent protein kinase-1
 NM_002621 properdin P factor, complement
 NM_002675 promyelocytic leukemia
 NM_002686 phenylethanolamine N-methyltransferase
 NM_002707 protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform
 NM_002719 protein phosphatase 2, regulatory subunit B (B56), gamma isoform
 NM_002761 protamine 1
 NM_002778 prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)
 NM_002857
 NM_002872 ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)
 NM_002908 v-rel reticuloendotheliosis viral oncogene homolog (avian)
 NM_002908 v-rel reticuloendotheliosis viral oncogene homolog (avian)
 NM_002920 regulatory factor X, 4 (influences HLA class II expression)
 NM_002920 regulatory factor X, 4 (influences HLA class II expression)
 NM_002993 chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)
 NM_003033 sialyltransferase 4A (beta-galactoside alpha-2,3-sialyltransferase)
 NM_003043 solute carrier family 6 (neurotransmitter transporter, taurine), member 6
 NM_003116 sperm associated antigen 4
 NM_003121 Spi-B transcription factor (Spi-1/PU.1 related)
 NM_003139 signal recognition particle receptor ('docking protein')
 NM_003174 supervillin
 NM_003183 a disintegrin and metalloproteinase domain 17 (tumor necrosis factor, alpha, converting enzyme)
 NM_003192 tubulin-specific chaperone c
 NM_003223 transcription factor AP-4 (activating enhancer binding protein 4)
 NM_003282 troponin I, skeletal, fast
 NM_003287 tumor protein D52-like 1
 NM_003308 testis specific protein, Y-linked
 NM_003387 Wiskott-Aldrich syndrome protein interacting protein
 NM_003395 wingless-type MMTV integration site family, member 9A
 NM_003456 zinc finger protein 205
 NM_003465 chitinase 1 (chitotriosidase)
 NM_003496 transformation/transcription domain-associated protein
 NM_003550 MAD1 mitotic arrest deficient-like 1 (yeast)
 NM_003579 RAD54-like (*S. cerevisiae*)
 NM_003602 FK506 binding protein 6, 36kDa
 NM_003657 breast carcinoma amplified sequence 1
 NM_003672 CDC14 cell division cycle 14 homolog A (*S. cerevisiae*)
 NM_003681 pyridoxal (pyridoxine, vitamin B6) kinase
 NM_003695 lymphocyte antigen 6 complex, locus D
 NM_003775 endothelial differentiation, G-protein-coupled receptor 6
 NM_003790 tumor necrosis factor receptor superfamily, member 25
 NM_003802 myosin, heavy polypeptide 13, skeletal muscle
 NM_003807 tumor necrosis factor (ligand) superfamily, member 14
 NM_003833 matrilin 4
 NM_003881 WNT1 inducible signaling pathway protein 2
 NM_003897 immediate early response 3

NM_003944 selenium binding protein 1
 NM_003963 transmembrane 4 superfamily member 5
 NM_003976 artemin
 NM_003999 oncostatin M receptor
 NM_004055 calpain 5
 NM_004108 ficolin (collagen/fibrinogen domain containing lectin) 2 (hucolin)
 NM_004130 glycogenin
 NM_004160 peptide YY
 NM_004178 TAR (HIV) RNA binding protein 2
 NM_004204 phosphatidylinositol glycan, class Q
 NM_004207 solute carrier family 16 (monocarboxylic acid transporters), member 3
 NM_004208 programmed cell death 8 (apoptosis-inducing factor)
 NM_004237 thyroid hormone receptor interactor 13
 NM_004246 glucagon-like peptide 2 receptor
 NM_004249 RAB28, member RAS oncogene family
 NM_004263 sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4F
 NM_004275 ubiquitin specific protease 49
 NM_004283 MBC3205
 NM_004327 breakpoint cluster region
 NM_004360 cadherin 1, type 1, E-cadherin (epithelial)
 NM_004369 collagen, type VI, alpha 3
 NM_004397 DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
 NM_004413 dipeptidase 1 (renal)
 NM_004417 dual specificity phosphatase 1
 NM_004420 dual specificity phosphatase 8
 NM_004422 dishevelled, dsh homolog 2 (Drosophila)
 NM_004431 EphA2
 NM_004560 receptor tyrosine kinase-like orphan receptor 2
 NM_004584 RAD9 homolog A (S. pombe)
 NM_004636 sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B
 NM_004636 sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B
 NM_004668 maltase-glucoamylase (alpha-glucosidase)
 NM_004723 rho/rac guanine nucleotide exchange factor (GEF) 2
 NM_004738 VAMP (vesicle-associated membrane protein)-associated protein B and C
 NM_004741 nucleolar and coiled-body phosphoprotein 1
 NM_004762 pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1)
 NM_004778 G protein-coupled receptor 44
 NM_004789 LIM homeobox 2
 NM_004826 endothelin converting enzyme-like 1
 NM_004848 chromosome 1 open reading frame 38
 NM_004858 solute carrier family 4, sodium bicarbonate cotransporter, member 8
 NM_004894 chromosome 14 open reading frame 2
 NM_004907 immediate early response 2
 NM_004928 chromosome 21 open reading frame 2
 NM_004928 chromosome 21 open reading frame 2
 NM_004957 folylpolyglutamate synthase
 NM_004984 kinesin family member 5A

NM_005029 paired-like homeodomain transcription factor 3
 NM_005044 protein kinase, X-linked
 NM_005069 single-minded homolog 2 (Drosophila)
 NM_005070 solute carrier family 4, anion exchanger, member 3
 NM_005087 fragile X mental retardation, autosomal homolog 1
 NM_005125 copper chaperone for superoxide dismutase
 NM_005145 hypothetical protein FLJ00058
 NM_005173 ATPase, Ca⁺⁺ transporting, ubiquitous
 NM_005235 v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
 NM_005245 FAT tumor suppressor homolog 1 (Drosophila)
 NM_005255 cyclin G associated kinase
 NM_005283 chemokine (C motif) receptor 1
 NM_005291 G protein-coupled receptor 17
 NM_005323
 NM_005323
 NM_005339 huntingtin interacting protein 2
 NM_005355 kinesin family member 25
 NM_005372 v-mos Moloney murine sarcoma viral oncogene homolog
 NM_005393 plexin B3
 NM_005394
 NM_005537 inhibitor of growth family, member 1
 NM_005567 lectin, galactoside-binding, soluble, 3 binding protein
 NM_005607 PTK2 protein tyrosine kinase 2
 NM_005609 phosphorylase, glycogen; muscle (McArdle syndrome, glycogen storage disease type V)
 NM_005629 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
 NM_005630 solute carrier organic anion transporter family, member 2A1
 NM_005634 SRY (sex determining region Y)-box 3
 NM_005689 ATP-binding cassette, sub-family B (MDR/TAP), member 6
 NM_005729 peptidylprolyl isomerase F (cyclophilin F)
 NM_005731 actin related protein 2/3 complex, subunit 2, 34kDa
 NM_005734 homeodomain interacting protein kinase 3
 NM_005744 ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (Drosophila)
 NM_005754 Ras-GTPase-activating protein SH3-domain-binding protein
 NM_005757 muscleblind-like 2 (Drosophila)
 NM_005762 tripartite motif-containing 28
 NM_005787 asparagine-linked glycosylation 3 homolog (yeast, alpha-1,3-mannosyltransferase)
 NM_005834 translocase of inner mitochondrial membrane 17 homolog B (yeast)
 NM_005860 follistatin-like 3 (secreted glycoprotein)
 NM_005864 embryonal Fyn-associated substrate
 NM_005881 branched chain alpha-ketoacid dehydrogenase kinase
 NM_006051 amyloid beta (A4) precursor protein-binding, family B, member 3
 NM_006140 colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)
 NM_006244 protein phosphatase 2, regulatory subunit B (B56), beta isoform
 NM_006253 protein kinase, AMP-activated, beta 1 non-catalytic subunit
 NM_006271 S100 calcium binding protein A1
 NM_006285 testis-specific kinase 1
 NM_006298 zinc finger protein 192
 NM_006312 nuclear receptor co-repressor 2
 NM_006367 CAP, adenylate cyclase-associated protein 1 (yeast)
 NM_006415 serine palmitoyltransferase, long chain base subunit 1

NM_006435 interferon induced transmembrane protein 2 (1-8D)
 NM_006460 HMBA-inducible
 NM_006461 sperm associated antigen 5
 NM_006514 sodium channel, voltage-gated, type X, alpha
 NM_006561 CUG triplet repeat, RNA binding protein 2
 NM_006624 adenovirus 5 E1A binding protein
 NM_006672 solute carrier family 22 (organic anion transporter), member 7
 NM_006674 HLA complex P5
 NM_006709 HLA-B associated transcript 8
 NM_006738 A kinase (PRKA) anchor protein 13
 NM_006738 A kinase (PRKA) anchor protein 13
 NM_006771 keratin, hair, acidic, 8
 NM_006789 apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2
 NM_006848 hepatitis delta antigen-interacting protein A
 NM_006887 zinc finger protein 36, C3H type-like 2
 NM_006977 zinc finger protein 46 (KUP)
 NM_007079 protein tyrosine phosphatase type IVA, member 3
 NM_007101 sarcosine dehydrogenase
 NM_007161 leukocyte specific transcript 1
 NM_007172 nucleoporin 50kDa
 NM_007184 nischarin
 NM_007191 WNT inhibitory factor 1
 NM_007262 Parkinson disease (autosomal recessive, early onset) 7
 NM_007327 glutamate receptor, ionotropic, N-methyl D-aspartate 1
 NM_007350 pleckstrin homology-like domain, family A, member 1
 NM_007350 pleckstrin homology-like domain, family A, member 1
 NM_007350 pleckstrin homology-like domain, family A, member 1
 NM_007371 bromodomain containing 3
 NM_007375
 NM_009586 single-minded homolog 2 (Drosophila)
 NM_012068 activating transcription factor 5
 NM_012109 chromosome 19 open reading frame 4
 NM_012147
 NM_012148 double homeobox, 1
 NM_012172 F-box only protein 24
 NM_012183 forkhead box D3
 NM_012188 forkhead box I1
 NM_012189 calcium-binding tyrosine-(Y)-phosphorylation regulated (fibrousheathin 2)
 NM_012199 eukaryotic translation initiation factor 2C, 1
 NM_012203 glyoxylate reductase/hydroxypyruvate reductase
 NM_012217 tryptase delta 1
 NM_012226 purinergic receptor P2X, ligand-gated ion channel, 2
 NM_012237 sirtuin (silent mating type information regulation 2 homolog) 2 (S. cerevisiae)
 NM_012281 potassium voltage-gated channel, Shal-related subfamily, member 2
 NM_012445 spondin 2, extracellular matrix protein
 NM_012450 solute carrier family 13 (sodium/sulfate symporters), member 4
 NM_012481 zinc finger protein, subfamily 1A, 3 (Aiolos)
 NM_013233 serine threonine kinase 39 (STE20/SPS1 homolog, yeast)
 NM_013271 proprotein convertase subtilisin/kexin type 1 inhibitor
 NM_013279 chromosome 11 open reading frame 9

NM_013291 cleavage and polyadenylation specific factor 1, 160kDa
NM_013348 potassium inwardly-rectifying channel, subfamily J, member 14
NM_013355 protein kinase PKNbeta
NM_013382 protein-O-mannosyltransferase 2
NM_013389 NPC1 (Niemann-Pick disease, type C1, gene)-like 1
NM_013390 transmembrane protein 2
NM_014033 DKFZP586A0522 protein
NM_014045 low density lipoprotein receptor-related protein 10
NM_014183 dynein, cytoplasmic, light polypeptide 2A
NM_014183 dynein, cytoplasmic, light polypeptide 2A
NM_014268 microtubule-associated protein, RP/EB family, member 2
NM_014297 ethylmalonic encephalopathy 1
NM_014328 RUN and SH3 domain containing 1
NM_014336 aryl hydrocarbon receptor interacting protein-like 1
NM_014369 protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
NM_014404 calcium channel, voltage-dependent, gamma subunit 5
NM_014421 dickkopf homolog 2 (*Xenopus laevis*)
NM_014424 heat shock 27kDa protein family, member 7 (cardiovascular)
NM_014442 sialic acid binding Ig-like lectin 8
NM_014520 MYB binding protein (P160) 1a
NM_014556 Ellis van Creveld syndrome
NM_014665 synonym: KIAA0014; Homo sapiens leucine rich repeat containing 14 (LRRC14), mRNA.
NM_014678 KIAA0685
NM_014725 START domain containing 8
NM_014748 sorting nexin 17
NM_014877
NM_014891 PDGFA associated protein 1
NM_014901 synonym: KIAA1100; Homo sapiens ring finger protein 44 (RNF44), mRNA.
NM_015140 KIAA0153 protein
NM_015250 bicaudal D homolog 2 (*Drosophila*)
NM_015377
NM_015415 DKFZP564B167 protein
NM_015456 cofactor of BRCA1
NM_015839 ficolin (collagen/fibrinogen domain containing lectin) 2 (hucolin)
NM_015845 methyl-CpG binding domain protein 1
NM_015846 methyl-CpG binding domain protein 1
NM_015854 retinoic acid receptor, beta
NM_015854 retinoic acid receptor, beta
NM_015872 zinc finger protein 67 homolog (mouse)
NM_015901 nudix (nucleoside diphosphate linked moiety X)-type motif 13
NM_015917 glutathione S-transferase subunit 13 homolog
NM_015938 CGI-07 protein
NM_015958 CGI-30 protein
NM_015963 THAP domain containing 4
NM_015977
NM_016011 nuclear receptor binding factor 1
NM_016152 retinoic acid receptor, beta
NM_016183 chromosome 1 open reading frame 33
NM_016185 hematological and neurological expressed 1
NM_016223 protein kinase C and casein kinase substrate in neurons 3

NM_016240 scavenger receptor class A, member 3
 NM_016246 dehydrogenase/reductase (SDR family) member 10
 NM_016248 A kinase (PRKA) anchor protein 11
 NM_016302 protein x 0001
 NM_016305 synovial sarcoma translocation gene on chromosome 18-like 2
 NM_016318 purinergic receptor P2X, ligand-gated ion channel, 2
 NM_016341 phospholipase C, epsilon 1
 NM_016341 phospholipase C, epsilon 1
 NM_016364 dual specificity phosphatase 13
 NM_016404 hypothetical protein HSPC152
 NM_016407 chromosome 20 open reading frame 43
 NM_016425
 NM_016427 transcription elongation factor B polypeptide 3B (elongin A2)
 NM_016429 coatomer protein complex, subunit zeta 2
 NM_016458 brain protein 16
 synonyms: MUM-1, HSPC211, FLJ14868, FLJ22283; Homo sapiens melanoma associated
 NM_016473 antigen (mutated) 1 (MUM1), mRNA.
 NM_016569 T-box 3 (ulnar mammary syndrome)
 NM_016582 solute carrier family 15, member 3
 NM_016596 histone deacetylase 7A
 NM_016610 toll-like receptor 8
 NM_016643 mesenchymal stem cell protein DSC43
 NM_016653 sterile alpha motif and leucine zipper containing kinase AZK
 NM_016932 sine oculis homeobox homolog 2 (Drosophila)
 NM_017436 alpha 1,4-galactosyltransferase
 NM_017490 MAP/microtubule affinity-regulating kinase 2
 NM_017555 egl nine homolog 2 (C. elegans)
 NM_017622 hypothetical protein FLJ20014
 NM_017623 cyclin M3
 NM_017631 hypothetical protein FLJ20035
 NM_017664 ankyrin repeat domain 10
 NM_017703 F-box and leucine-rich repeat protein 12
 NM_017717 mucin and cadherin-like
 NM_017717 mucin and cadherin-like
 NM_017729 EPS8-like 1
 NM_017731 oxysterol binding protein-like 7
 NM_017784 oxysterol binding protein-like 10
 NM_017814 hypothetical protein FLJ20422
 NM_017895 DEAD (Asp-Glu-Ala-Asp) box polypeptide 27
 NM_017900 hypothetical protein MGC3047
 NM_017900 hypothetical protein MGC3047
 NM_017918 hypothetical protein FLJ20647
 NM_018014 B-cell CLL/lymphoma 11A (zinc finger protein)
 NM_018074 hypothetical protein FLJ10374
 NM_018081 hypothetical protein FLJ10385
 NM_018088 hypothetical protein FLJ10408
 NM_018106 zinc finger, DHHC domain containing 4
 NM_018129 pyridoxine-5'-phosphate oxidase
 NM_018174 VCY2 interacting protein 1
 NM_018205 hypothetical protein FLJ10751

NM_018212 enabled homolog (Drosophila)
 NM_018273 hypothetical protein FLJ10922
 NM_018358 ATP-binding cassette, sub-family F (GCN20), member 3
 NM_018391
 NM_018391
 NM_018487 hepatocellular carcinoma-associated antigen 112
 NM_018502 hypothetical protein PRO1580
 NM_018518 MCM10 minichromosome maintenance deficient 10 (*S. cerevisiae*)
 NM_018602 DnaJ (Hsp40) homolog, subfamily A, member 4
 synonyms: MBLL, MBLL39, PRO2032; isoform 2 is encoded by transcript variant 2;
 muscleblind-like protein MBLL39; Homo sapiens muscleblind-like 2 (*Drosophila*) (MBNL2),
 transcript variant 2, mRNA.
 NM_018615
 NM_018628 synonym: PRO2472; Homo sapiens hypothetical protein FLJ20014 (FLJ20014), mRNA.
 NM_018641 carbohydrate (chondroitin 4) sulfotransferase 12
 NM_018655 lens epithelial protein
 NM_018846 SBBI26 protein
 NM_018846 SBBI26 protein
 NM_018931 protocadherin beta 11
 NM_018933 protocadherin beta 13
 NM_018942 homeo box (H6 family) 1
 NM_018956 chromosome 9 open reading frame 9
 NM_018965 triggering receptor expressed on myeloid cells 2
 NM_018971 G protein-coupled receptor 27
 NM_019011 TRIAD3 protein
 NM_019033 hypothetical protein FLJ11235
 NM_020190 HNOEL-iso protein
 NM_020241 sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B
 NM_020246 thyroid hormone receptor interactor 6
 NM_020478 ankyrin 1, erythrocytic
 NM_020479 ankyrin 1, erythrocytic
 NM_020480 ankyrin 1, erythrocytic
 NM_020533 mucolipin 1
 NM_020638 fibroblast growth factor 23
 NM_020659 tweety homolog 1 (*Drosophila*)
 NM_020988 guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O
 NM_021092 pancreatic polypeptide 2
 NM_021096 calcium channel, voltage-dependent, alpha 11 subunit
 NM_021130 peptidylprolyl isomerase A (cyclophilin A)
 NM_021161 potassium channel, subfamily K, member 10
 NM_021175 hepcidin antimicrobial peptide
 NM_021198 CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase 1
 NM_021203 signal recognition particle receptor, B subunit
 NM_021226 Rho GTPase activating protein 22
 NM_021570 BarH-like homeobox 1
 NM_021628 arachidonate lipoxygenase 3
 NM_021730
 NM_021733 testis-specific kinase substrate
 NM_021733 testis-specific kinase substrate
 NM_021924 mucin and cadherin-like
 NM_021978 suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin)

NM_022081 Hermansky-Pudlak syndrome 4
NM_022110 FK506 binding protein like
NM_022114 PR domain containing 16
NM_022123 neuronal PAS domain protein 3
NM_022161 baculoviral IAP repeat-containing 7 (livin)
NM_022304 histamine receptor H2
NM_022355 dipeptidase 2
NM_022363 LIM homeobox 5
NM_022368 praja 1
NM_022452 fibrosin 1
NM_022455 nuclear receptor binding SET domain protein 1
NM_022490 RNA polymerase I associated factor 53
NM_022719 DiGeorge syndrome critical region gene 14
NM_022750 zinc finger CCCH type domain containing 1
NM_022765 NEDD9 interacting protein with calponin homology and LIM domains
NM_022773 hypothetical protein FLJ12681
NM_022893 B-cell CLL/lymphoma 11A (zinc finger protein)
NM_022908 hypothetical protein FLJ12442
NM_023011 UPF3 regulator of nonsense transcripts homolog A (yeast)
NM_023073 synonym: FLJ21126; Homo sapiens hypothetical protein FLJ13231 (FLJ13231), mRNA.
NM_023933 hypothetical protein MGC2494
NM_023942 hypothetical protein MGC3036
NM_024029 hypothetical protein MGC3262
NM_024059 hypothetical protein MGC5356
NM_024078 hypothetical protein MGC3162
NM_024103 hypothetical protein MGC2615
NM_024327 zinc finger protein 576
NM_024333 fibronectin type 3 and SPRY (spla, ryanodine) domain containing (with coiled-coil motif) 1
NM_024426 Wilms tumor 1
NM_024509 leucine rich repeat and fibronectin type III domain containing 3
NM_024510 hypothetical protein MGC4368
NM_024546 chromosome 13 open reading frame 7
NM_024552 LAG1 longevity assurance homolog 4 (S. cerevisiae)
NM_024589 leucine zipper domain protein
NM_024618 NOD9 protein
NM_024682 TBC1 domain family, member 17
NM_024768 hypothetical protein FLJ12057
NM_024785 hypothetical protein FLJ22746
NM_024817 hypothetical protein FLJ13710
NM_024829 hypothetical protein FLJ22662
NM_024829 hypothetical protein FLJ22662
NM_024906 stearyl-CoA desaturase 4
NM_024919 FERM domain containing 1
NM_024988 hypothetical protein FLJ12355
NM_025024
NM_025067
NM_025215 pseudouridylate synthase 1
NM_025256 HLA-B associated transcript 8
NM_030576 hypothetical protein MGC10986
NM_030578 hypothetical protein MGC4093

NM_030594 cytoplasmic polyadenylation element binding protein 1
 NM_030615 kinesin family member 25
 NM_030774 olfactory receptor, family 51, subfamily E, member 2
 NM_030784 G protein-coupled receptor 63
 NM_030795 stathmin-like 4
 NM_030800 hypothetical protein DKFZp564O1664
 NM_031214
 NM_031220 PITPNM family member 3
 NM_031268
 NM_031286 SH3 domain binding glutamic acid-rich protein like 3
 NM_033029 leishmanolysin-like (metallopeptidase M8 family)
 NM_033104 stonin 2
 NM_052836 cadherin related 23
 NM_080869 WAP four-disulfide core domain 12
 NM_133490 potassium voltage-gated channel, subfamily G, member 4
 NM_138430 ADP-ribosylhydrolase like 1
 synonyms: NDPP1, FLJ10773; Homo sapiens enabled homolog (Drosophila) (ENAH), mRNA.
 NM_145240
 NM_145301 similar to CGI-148 protein
 NM_145653 transcription elongation factor B polypeptide 3C (elongin A3)
 BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)
 NM_145696
 NM_145811 calcium channel, voltage-dependent, gamma subunit 5
 NM_147133 nuclear transcription factor, X-box binding 1
 NM_147134 nuclear transcription factor, X-box binding 1
 NM_152843 Hermansky-Pudlak syndrome 4
 NM_170722 NOD9 protein
 R34841 superoxide dismutase 2, mitochondrial
 R41907 quaking homolog, KH domain RNA binding (mouse)
 yg18a05.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:32537 3', mRNA sequence.
 R43158
 R59977 transcriptional adaptor 3 (NGG1 homolog, yeast)-like
 R70029 hypothetical protein FLJ14775
 R76659 ubiquitin-conjugating enzyme E2, J2 (UBC6 homolog, yeast)
 R99562 forkhead box A3
 S64699 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)
 S64699 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)
 S75264 Wilms tumor 1
 S83390 nuclear receptor co-repressor 2
 T15657 IB1702 Infant brain, Bento Soares Homo sapiens cDNA 3'end, mRNA sequence.
 T30183 breakpoint cluster region
 yc66b09.s1 Stratagene liver (#937224) Homo sapiens cDNA clone IMAGE:85625 3', mRNA sequence.
 T62044
 U07000
 U07802 Human Tis11d gene, complete cds.
 U09609 nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
 U11701 Human LIM-homeobox domain protein (hLH-2) mRNA, complete cds.
 U14383 Human mucin (MUC8) mRNA, partial cds.

U14383 Human mucin (MUC8) mRNA, partial cds.
 U16120 solute carrier family 6 (neurotransmitter transporter, taurine), member 6
 U17040 kallikrein 3, (prostate specific antigen)
 U17074 cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)
 U17986 solute carrier family 6 (neurotransmitter transporter, creatine), member 8
 U19345 transcription factor 20 (AR1)
 U20428 Human SNC19 mRNA sequence.
 U20428 Human SNC19 mRNA sequence.
 U27331 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 U27332 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 U27335 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 U27336 fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
 U34846 Human mercurial-insensitive water channel mRNA, form 2, complete cds.
 U37012 cleavage and polyadenylation specific factor 1, 160kDa
 U38371 huntingtin-associated protein 1 (neuroan 1)
 U39361 ubiquitin-conjugating enzyme E2 variant 1
 U39573 lactoperoxidase
 U41163 Human creatine transporter (SLC6A10) gene, partial cds.
 U41813 homeo box A9
 U43342 nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
 U47674 N-acylsphingosine amidohydrolase (acid ceramidase) 1
 U47674 N-acylsphingosine amidohydrolase (acid ceramidase) 1
 U48734 Human non-muscle alpha-actinin mRNA, complete cds.
 U58658
 U58828 G protein-coupled receptor 30
 U59151 dyskeratosis congenita 1, dyskerin
 U59323 regulator of nonsense transcripts 1
 U61536 potassium large conductance calcium-activated channel, subfamily M, beta member 1
 U63622 aquaporin 4
 U63917 G protein-coupled receptor 30
 U65932 extracellular matrix protein 1
 U66691 ATP-binding cassette, sub-family A (ABC1), member 4
 U69546 CUG triplet repeat, RNA binding protein 2
 U69556 T-box 3 (ulnar mammary syndrome)
 U72763 tumor necrosis factor receptor superfamily, member 25
 U73844 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
 U79283 D site of albumin promoter (albumin D-box) binding protein
 U80756 polyglutamine rich; Homo sapiens CAGL114 mRNA, partial cds.
 U80761 cysteine rich; Homo sapiens CTG26 alternate open reading frame mRNA, complete cds.
 U84569 chromosome 21 open reading frame 2
 a disintegrin and metalloproteinase domain 17 (tumor necrosis factor, alpha, converting enzyme)
 U86755
 U88667 ATP-binding cassette, sub-family A (ABC1), member 4
 U94506 tumor necrosis factor receptor superfamily, member 25
 U94510 tumor necrosis factor receptor superfamily, member 25
 W01990 ankyrin repeat domain 10
 W46388 superoxide dismutase 2, mitochondrial
 W57731 forkhead box D3
 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)
 W60595

W60595	cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)
W79537	CUG triplet repeat, RNA binding protein 2
W80642	nudix (nucleoside diphosphate linked moiety X)-type motif 13
W93523	enhancer of polycomb homolog 1 (Drosophila)
X04014	ORF; Homo sapiens DNA for HBV integration sites.
X15132	superoxide dismutase 2, mitochondrial
X55666	upstream transcription factor 1
X60188	mitogen-activated protein kinase 3
X78262	H.sapiens mRNA for TRE5.
X79683	H.sapiens LAMB2 mRNA for beta2 laminin.
X79780	RAB11B, member RAS oncogene family
X89672	olfactory receptor, family 8, subfamily G, member 1 pseudogene
X89672	olfactory receptor, family 8, subfamily G, member 1 pseudogene
X90539	tripartite motif-containing 10
Y14330	jagged 2
Y15724	Homo sapiens SERCA3 gene, exons 1-7 (and joined CDS). cloned using degenerate PCR primers representing protein kinase subdomains VII and IX; encodes an open reading frame between subdomains VII and IX of protein kinase catalytic domain; H.sapiens protein-serine/threonine kinase gene, complete CDS.
Z25427	H.sapiens protein-serine/threonine kinase gene, complete CDS.
Z69744	H.sapiens ALL-1 gene exon 1 (and joined coding region).
Z78349	hypothetical protein MGC3262
Z93015	
Z94154	G protein-coupled receptor 17

Supplementary Table 4: Chromatin condensed by treatment

Genbank	Description
AB000114	Osteomodulin
AB002533	Karyopherin alpha 4 (importin alpha 3)
AB007619	Estrogen receptor binding site associated, antigen, 9
AB012610	Inositol 1,4,5-triphosphate receptor, type 2
AB013911	Hypothetical protein FLJ10134
AB014888	DnaJ (Hsp40) homolog, subfamily B, member 6
AB015320	Adaptor-related protein complex 1, sigma 2 subunit similar to Caenorhabditis elegans C42C1.9 gene sequence in GenBank Accession Number AF043695; Homo sapiens gene, complete cds, similar to Caenorhabditis elegans C42C1.9 gene sequence.
AB018790	BCL2-associated athanogene 5
AB020680	BCL2-associated athanogene 5
AB020686	Ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative function)
AB021643	Zinc finger protein 12 (KOX 3)
AB022663	Ring finger protein 14
AB023214	Zinc finger and BTB domain containing 1
AB026723	Pyrophosphatase (inorganic)
AB028995	Protein phosphatase 1E (PP2C domain containing)
AB029029	Myelin transcription factor 1-like
AB029037	Trophinin
AB033105	KIAA1279
AB033749	
AB034205	Homo sapiens mRNA for cisplatin resistance-associated overexpressed protein, complete cds.

AB037759 Eukaryotic translation initiation factor 2 alpha kinase 4
 AB037786 Leucine rich repeat containing 7
 AB055065 Tumor protein p73
 AB060691 UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 3
 AB060695 Toll-like receptor 5
 Homo sapiens RPS20, U54 genes for ribosomal protein S20 and U54
 AB061842 snoRNA, complete cds and sequence.
 AB063302 Chromosome 12 open reading frame 22
 AC005058
 AC005154 Homo sapiens PAC clone RP4-777O23 from 7, complete sequence.
 AC007032
 AF015593 Ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)
 AF019386 Heparan sulfate (glucosamine) 3-O-sulfotransferase 1
 AF023456 Protein phosphatase, EF hand calcium-binding domain 2
 AF027734 Deleted in bladder cancer 1
 AF038951 Hypothetical protein BC014148
 AF038966 Secretory carrier membrane protein 1
 Homo sapiens vascular endothelial cell growth inhibitor (VEGI) mRNA,
 AF039390 partial cds.
 AF039652 Ribonuclease H1
 AF041427 Homo sapiens ribosomal protein s4 Y isoform gene, complete cds.
 AF052573 Polymerase (DNA directed), theta
 Guanine nucleotide binding protein (G protein), alpha inhibiting activity
 polypeptide 1
 AF055013
 AF056334 Melanoma antigen family C, 1
 AF058317 Solute carrier family 23 (nucleobase transporters), member 1
 AF063596 Homo sapiens brain my038 protein mRNA, complete cds.
 AF070656 YME1-like 1 (*S. cerevisiae*)
 AF072693 G protein-coupled receptor 75
 AF074000 Lectin, galactoside-binding, soluble, 8 (galectin 8)
 AF074979 Regulator of G-protein signalling 20
 AF077031 Protein tyrosine phosphatase, non-receptor type 22 (lymphoid)
 AF081258 Chromodomain protein, Y-like
 AF087901 Reticulon 4
 AF087916 Homo sapiens olfactory receptor 17-2 (OR1E1) gene, complete cds.
 AF099082 Xenotropic and polytropic retrovirus receptor
 AF099935 Tumor necrosis factor, alpha-induced protein 8
 AF105267 Glypican 6
 AF107463 Survival motor neuron domain containing 1
 AF112201 Transgelin 3
 AF112345 Integrin, alpha 10
 AF112481 Fibrinogen silencer binding protein
 AF116456 Tumor necrosis factor (ligand) superfamily, member 13b
 AF116670 Nucleoside phosphorylase
 Cysteine and histidine-rich domain (CHORD)-containing, zinc binding protein
 1
 AF123249
 AF123320 Kelch-like 5 (*Drosophila*)
 AF123534 Nucleolar protein NOP5/NOP58
 transmembrane 4 superfamily member; Homo sapiens tetraspanin TM4-B
 AF133424 mRNA, complete cds.
 AF138303 Decorin

AF146074 ATP-binding cassette, sub-family C (CFTR/MRP), member 5
 AF152319 Protocadherin gamma subfamily C, 3
 AF152493 Protocadherin beta 14
 AF161357 RAB GTPase activating protein 1
 AF161545 Homo sapiens HSPC060 mRNA, complete cds.
 AF168717 X 009 protein
 AF170025 Zinc finger protein 95 homolog (mouse)

 AF184110 RS-cyclophilin; Homo sapiens cyclophilin-related protein (NKTR) gene,
 complete cds.
 AF187858 Angiotensin II
 AF189269 Prostate and breast cancer overexpressed 1
 AF189723 ATPase, Ca⁺⁺ transporting, type 2C, member 1

 AF191047 intermediate filament; Homo sapiens beaded filament structural protein 1
 AF195969 filensin (BFSP1) gene, exon 3 and partial cds.
 AF217227 Rho GTPase activating protein 8
 AF218011 Zinc finger protein 287
 AF220532 Mortality factor 4 like 1
 AF221098 Nuclear receptor subfamily 2, group E, member 1
 AF229253 Ral GEF with PH domain and SH3 binding motif 1
 AF231916 Apoptosis inhibitor 5
 AF242772 Hydroxyacid oxidase (glycolate oxidase) 1
 AF247661 Fas (TNFRSF6) associated factor 1
 AF251296 Putative S1 RNA binding domain protein
 AF251700 OCIA domain containing 1
 AF255310 Dendritic cell-derived ubiquitin-like protein
 AF274026 TFIIA-alpha/beta-like factor
 AF279784 SAR1a gene homolog 1 (*S. cerevisiae*)
 AF282619 Solute carrier organic anion transporter family, member 1A2
 AF283769 Lysyl oxidase-like 3
 AF293366 Mediator of RNA polymerase II transcription, subunit 25 homolog (yeast)
 AF311304 Similar to hypothetical protein FLJ10883
 AF314544 TAR DNA binding protein
 AF316824 Transducin (beta)-like 1X-linked receptor 1
 AF318326 Asporin (LRR class 1)
 AF332234 APG10 autophagy 10-like (*S. cerevisiae*)
 AF332892 Testis-specific transcript, Y-linked 8
 AF333487 Phospholipase A2, group XIA
 AF345909 Kazal-type serine protease inhibitor domain 1
 AF346966 Spermatogenesis associated 16

 AF348207 PHD finger protein 20
 AF353720 DEAD (Asp-Glu-Ala-As) box polypeptide 19
 AF383173 Ubiquitin specific protease 33
 AF397424 O-acetyltransferase
 AF420474 Mitogen-activated protein kinase 6
 AF491780
 AF498040 Killer cell lectin-like receptor subfamily D, member 1
 AF498094 Amphiphysin (Stiff-Man syndrome with breast cancer 128kDa autoantigen)
 AF503166 Homo sapiens xeroderma pigmentosum, complementation group A (XPA)
 gene, complete cds.

AF504543	Homo sapiens AMPK beta-2 subunit (PRKAB2) gene, exon 8 and complete cds.
AJ238044	Bradykinin receptor B1
AJ271790	Smith-Magenis syndrome chromosome region, candidate 6
AJ277275	Transcriptional regulating factor 1
AJ278982	5-hydroxytryptamine (serotonin) receptor 4
AJ302552	Homo sapiens 6M1-3*02 gene for olfactory receptor, cell line BM28.7 olfactory receptor.
AJ315767	
AK000058	ADP-ribosylation factor related protein 2
AK000305	FLJ20298 protein
AK000889	Zinc finger CCCH type domain containing 7
AK001210	COP9 constitutive photomorphogenic homolog subunit 4 (Arabidopsis)
AK001377	Selenocysteine lyase
AK001419	Feline leukemia virus subgroup C cellular receptor
AK001514	Hypothetical protein FLJ20696
AK001670	Hypothetical protein FLJ10808
AK001894	Programmed cell death 1 ligand 1
AK001973	Hypothetical protein FLJ10884
AK002145	Pleckstrin homology, Sec7 and coiled-coil domains 4
AK022408	RAB GTPase activating protein 1
AK022509	DKFZP434B168 protein
AK022713	unnamed protein product; Homo sapiens cDNA FLJ12651 fis, clone NT2RM4002062, moderately similar to ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12).
AK023303	Coatmer protein complex, subunit epsilon
AK023385	Tropomyosin 4
AK023408	Chromosome 20 open reading frame 172
AK023676	Hypothetical protein FLJ13614
AK023732	Hypothetical protein FLJ11016
AK023921	Tetratricopeptide repeat domain 12
AK024093	Hypothetical protein FLJ14031
AK024259	KIAA0998
AK025287	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)
AK026660	Similar to hypothetical protein MGC17347
AK026810	TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor, 50kDa
AK027671	RAB3A interacting protein (rabin3)
AK027856	C-terminal tensin-like
AK056293	Interferon induced with helicase C domain 1
AK074520	PTX1 protein
AK074548	Chromosome 19 open reading frame 27
AK074611	Dymeclin
AK075192	Eukaryotic translation initiation factor 2-alpha kinase 1
AK075320	Solute carrier family 17 (anion/sugar transporter), member 5
AK090967	Adenylate kinase 5
AK092181	Malic enzyme 3, NADP(+)-dependent, mitochondrial
AK093069	Immunoglobulin superfamily, member 10
AK094883	Dystonin
AK094906	Hypothetical protein FLJ37587

AK095492	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)
AK097725	Phospholipase C, zeta 1
AK098486	Hypothetical protein MGC11061
AL035369	Hypothetical protein MGC9084
AL133544	
AL135939	
AL137067	
AL590524	PEPP subfamily gene 2
AY026351	Similar to ARP3BETA protein
AY035377	Peptidoglycan recognition protein 4
AY130859	Homo sapiens cyclin-dependent kinase 7 (CDK7) gene, complete cds.
BC000453	Pericentriolar material 1
BC000578	Hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)
BC000814	TGFB-induced factor (TALE family homeobox)
BC000934	Eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa
BC001062	Potassium channel tetramerisation domain containing 14
BC001139	5-azacytidine induced 2
BC001187	COP9 constitutive photomorphogenic homolog subunit 5 (Arabidopsis)
BC001426	Ubiquinol-cytochrome c reductase hinge protein
BC001916	Chromosome 14 open reading frame 94
BC002390	Etoposide induced 2.4 mRNA
BC002642	Cathepsin S
BC003049	PAI-1 mRNA-binding protein
BC004105	Melanoma antigen family A, 10
BC004885	Hypothetical protein RP4-622L5
BC004958	Leucine rich repeat (in FLII) interacting protein 1
BC004974	Secreted protein, acidic, cysteine-rich (osteonectin)
BC005359	Glia maturation factor, beta
BC005922	Islet cell autoantigen 1, 69kDa
BC005940	Guanine nucleotide binding protein (G protein), gamma 12
BC006268	Peroxisomal biogenesis factor 7
BC006293	Chromosome 6 open reading frame 153
BC008063	Signal peptidase complex subunit 2 homolog (S. cerevisiae)
BC008244	Protein tyrosine phosphatase, non-receptor type 2
BC009279	Mutated in colorectal cancers
BC009475	Mitochondrial ribosomal protein L33
BC009748	Dopamine receptor D5
BC009808	Membrane component, chromosome 17, surface marker 2 (ovarian carcinoma antigen CA125)
BC009850	Cell division cycle 34
BC011595	Glycoprotein (transmembrane) nmb
BC011837	DnaJ (Hsp40) homolog, subfamily C, member 7
BC012043	Hydroxysteroid (17-beta) dehydrogenase 12
BC013135	
BC014411	Surfeit 2
BC014441	NOL1/NOP2/Sun domain family, member 4
BC015212	Zinc finger protein 291
BC015877	Cadherin 19, type 2
BC015930	Ubiquitin specific protease 25
BC016372	Myosin regulatory light chain MRCL3

BC016775	SMT3 suppressor of mif two 3 homolog 2 (yeast)
BC017029	Farnesyltransferase, CAAX box, alpha
BC017309	KIN, antigenic determinant of recA protein homolog (mouse)
BC017700	Testis zinc finger protein
BC017701	Hypothetical protein DKFZp434L142 synonyms: SPCH1, CAGH44, TNRC10; Homo sapiens forkhead box P2, mRNA (cDNA clone IMAGE:4285527), complete cds.
BC018016	Leucine zipper, down-regulated in cancer 1-like
BC018713	Putative acyl-CoA dehydrogenase
BC019607	Deafness, autosomal dominant 5
BC019689	Hypothetical protein LOC283507
BC020814	
BC020951	
BC021090	Scinderin
BC022188	Hypothetical gene CG018
BC022251	Hypothetical protein FLJ14007
BC022344	PTK9 protein tyrosine kinase 9
BC024218	Vacuolar protein sorting 35 (yeast)
BC024238	MKI67 (FHA domain) interacting nucleolar phosphoprotein
BC024781	Dishevelled associated activator of morphogenesis 1
BC028120	Ring finger protein 32
BC028412	Elongation factor, RNA polymerase II, 2
BC029378	Telomeric repeat binding factor (NIMA-interacting) 1
BC029390	
BC031262	Catenin (cadherin-associated protein), alpha 1, 102kDa
BC031549	CDC-like kinase 1
BC032244	Transcription elongation factor B polypeptide 3 binding protein 1
BC032517	Nerve growth factor, beta polypeptide
BC032851	Cas-Br-M (murine) ecotropic retroviral transforming sequence b
BC033656	Polyamine-modulated factor 1
BC033826	Purinergic receptor P2X, ligand-gated ion channel, 4
BC034393	Endothelin 2
BC036768	Hypothetical protein FLJ13220
D10583	Membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide)
D13645	KIAA0020
D28472	Prostaglandin E receptor 4 (subtype EP4)
D87120	Family with sequence similarity 3, member C
D88532	Phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma)
D90040	N-acetyltransferase 2 (arylamine N-acetyltransferase)
empty080	
J03810	Solute carrier family 2 (facilitated glucose transporter), member 2
L01457	Exosome component 10
L02752	Potassium voltage-gated channel, shaker-related subfamily, member 2 Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3
L13744	
L16558	Ribosomal protein L7
L40411	Jumonji domain containing 1C
M13599	Human c-ros-1 proto-oncogene, exon 10, clone HYuros7C.
M18078	Statherin
M22616	Sucrase-isomaltase (alpha-glucosidase)

M23254	Calpain 2, (m/II) large subunit
M24486	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide I
M27024	Homo sapiens heat shock protein (HSP89-alpha) gene, complete cds. precursor; Homo sapiens eosinophil peroxidase (EPP) gene, exon 12 and complete cds.
M29913	Carbonic anhydrase I
M33987	son3 protein; Human son3 protein gene, partial cds.
M36428	5-hydroxytryptamine (serotonin) receptor 2A
M86841	Sodium channel, voltage-gated, type VII, alpha
M91556	Receptor tyrosine kinase-like orphan receptor 1
M97675	Collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)
NM_000493	Aldehyde dehydrogenase 9 family, member A1
NM_000696	Guanylate cyclase 1, soluble, alpha 3
NM_000856	Sucrase-isomaltase (alpha-glucosidase)
NM_001041	Signal sequence receptor, alpha (translocon-associated protein alpha)
NM_003144	Zinc finger protein 189
NM_003452	
NM_003832	Sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A
NM_003966	Carnitine acetyltransferase
NM_004003	Dystrophin (muscular dystrophy, Duchenne and Becker types)
NM_004007	Cadherin 17, LI cadherin (liver-intestine)
NM_004063	G protein pathway suppressor 1
NM_004127	Chondroitin sulfate proteoglycan 2 (versican)
NM_004385	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N- acetylgalactosaminyltransferase 3 (GalNAc-T3)
NM_004482	Interleukin enhancer binding factor 2, 45kDa
NM_004515	Low density lipoprotein-related protein 2
NM_004525	Phosphoinositide-3-kinase, class 2, gamma polypeptide
NM_004570	Matrix metalloproteinase 20 (enamelysin)
NM_004771	RAN binding protein 2-like 1
NM_005054	GA binding protein transcription factor, beta subunit 2, 47kDa
NM_005254	Integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)
NM_005501	Purinergic receptor P2Y, G-protein coupled, 5
NM_005767	Pre-B-cell leukemia transcription factor 3
NM_006195	Ribonuclease P/MRP 38kDa subunit
NM_006414	Regulatory solute carrier protein, family 1, member 1
NM_006511	Male-specific lethal 3-like 1 (Drosophila)
NM_006800	CD86 antigen (CD28 antigen ligand 2, B7-2 antigen)
NM_006889	Diaphanous homolog 2 (Drosophila)
NM_007309	F-box and leucine-rich repeat protein 21
NM_012159	synonyms: HSPC064, DKFZP564O0463; HSPC064 protein; Homo sapiens DKFZP564O0463 protein (Gm83), mRNA.
NM_014156	Influenza virus NS1A binding protein
NM_016389	Transmembrane protein 9
NM_016456	Zinc finger protein 571
NM_016536	Family with sequence similarity 49, member B
NM_016623	
NM_018572	
NM_018578	

NM_018579	Mitochondrial solute carrier protein
NM_018630	synonyms: DER1, DER-1, MGC3067, PRO2577, FLJ13784; derlin-1;
NM_020221	
NM_021023	Complement factor H-related 3
NM_021213	Phosphatidylcholine transfer protein
NM_022482	Zinc finger protein 336
NM_030893	CD1E antigen, e polypeptide
NM_030959	Olfactory receptor, family 12, subfamily D, member 3
NM_031362	Collagen, type IV, alpha 3 (Goodpasture antigen)
NM_032891	
NM_138634	Microseminoprotein, beta-
NM_145695	Diacylglycerol kinase, beta 90kDa
S52784	Cystathionase (cystathionine gamma-lyase)
	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3
S64671	
U10990	Human nuclear receptor hTAK1 (hTAK1) mRNA, complete cds.
U17034	Phospholipase A2 receptor 1, 180kDa
U23435	Abl interactor 2
U31345	Calpastatin
U49844	Ataxia telangiectasia and Rad3 related
U49973	
U57843	Phosphoinositide-3-kinase, catalytic, delta polypeptide
U58111	Vascular endothelial growth factor C
U83463	Syndecan binding protein (syntenin)
XM_009363	
XM_016672	
XM_016674	
XM_017557	
XM_027783	
XM_027798	
XM_028958	
XM_032201	
XM_035825	Homo sapiens KIAA0143 protein (KIAA0143), mRNA.
XM_039114	synonyms: ACO3, IRP2, IRP2AD; iron regulatory protein 2;
XM_039551	
XM_042168	
XM_043814	
XM_058676	
XM_060365	
XM_065856	
XM_066953	
XM_070805	
XM_070916	
XM_085634	Hypothetical protein LOC146909
XM_085741	
XM_086240	
XM_086343	
XM_087254	ATPase, Class VI, type 11B
XM_087813	
XM_088514	

XM_088535	
XM_092986	
XM_096733	Chromosome 14 open reading frame 72
XM_096908	
XM_097719	
XM_098290	
XM_113540	
XM_114099	
XM_115290	
XM_115457	
XM_116684	
XM_166037	
XM_166160	Chromosome 10 open reading frame 37
XM_166685	
XM_167159	
XM_167269	
XM_170463	
XM_170519	
XM_170709	
XM_170863	
XM_171028	
XM_171032	Succinate dehydrogenase complex, subunit A, flavoprotein-like 2
XM_171863	
XM_172106	
XM_172341	Hypothetical protein FLJ35036
XM_172806	
XM_173039	

Supplementary table 5: Genes with multiple probes

Genbank	Description
BC033692	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
AF033026	3'-phosphoadenosine 5'-phosphosulfate synthase 1
AB044805	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2
BC018029	6-pyruvoyltetrahydropterin synthase
AF009615	A disintegrin and metalloproteinase domain 10
D14665	A disintegrin and metalloproteinase domain 9 (meltrin gamma)
BC006082	Acheron
S62884	Acid phosphatase 1, soluble
BC007422	Acid phosphatase 1, soluble
AK021784	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member A
AJ311392	Acidic repeat containing
NM_002313	Actin binding LIM protein 1
D31883	Actin binding LIM protein 1
U50523	Actin related protein 2/3 complex, subunit 2, 34kDa
BC000018	Actin related protein 2/3 complex, subunit 5-like
NM_004024	Activating transcription factor 3
AB018327	Activity-dependent neuroprotector
AF116690	Acyl-CoA synthetase long-chain family member 3
NM_003664	Adaptor-related protein complex 3, beta 1 subunit
AB051472	Additional sex combs like 2 (Drosophila)

AK001553	Adenylate kinase 3 like 1
NM_001122	Adipose differentiation-related protein
AF052179	ADP-ribosylation factor 1
AF493882	ADP-ribosylation factor 3
NM_014447	ADP-ribosylation factor interacting protein 1 (arfaptin 1)
AF124489	ADP-ribosylation factor interacting protein 1 (arfaptin 1)
NM_000690	Aldehyde dehydrogenase 2 family (mitochondrial)
NM_000695	Aldehyde dehydrogenase 3 family, member B2
NM_000696	Aldehyde dehydrogenase 9 family, member A1
S99468	Aminolevulinate, delta-, dehydratase
AJ132583	Aminopeptidase puromycin sensitive
AB011109	AMP-activated protein kinase family member 5
M15533	Amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)
D86981	Amyloid beta precursor protein (cytoplasmic tail) binding protein 2
NM_014885	Anaphase promoting complex subunit 10
BC006301	Anaphase promoting complex subunit 5
NM_015928	Androgen-induced proliferation inhibitor
AB023196	Androgen-induced proliferation inhibitor
NM_016201	Angiomotin like 2
AK092527	Ankyrin 3, node of Ranvier (ankyrin G)
D86982	Ankyrin repeat and sterile alpha motif domain containing 1
AJ278463	Annexin A11
BC000182	Annexin A4
BC005830	Annexin A9
AF293841	APG5 autophagy 5-like (<i>S. cerevisiae</i>)
U83857	Apoptosis inhibitor 5
AF229254	Apoptosis inhibitor 5
AF229253	Apoptosis inhibitor 5
BC007973	Apoptosis-related protein PNAS-1
BC013566	Aquaporin 3
AB011132	Aquarius homolog (mouse)
AJ296161	Arginyl aminopeptidase (aminopeptidase B)
AJ130978	Ariadne homolog 2 (<i>Drosophila</i>)
AF072832	Ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1 (<i>Drosophila</i>)
AF006082	ARP2 actin-related protein 2 homolog (yeast)
AB038229	ARP6 actin-related protein 6 homolog (yeast)
NM_004317	ArsA arsenite transporter, ATP-binding, homolog 1 (bacterial)
NM_016374	AT rich interactive domain 4B (RBP1- like)
AF083249	AT rich interactive domain 4B (RBP1- like)
L32832	AT-binding transcription factor 1
D10250	AT-binding transcription factor 1
BC029653	AT-binding transcription factor 1
M23115	ATPase, Ca ⁺⁺ transporting, cardiac muscle, slow twitch 2
BC000988	ATPase, Ca ⁺⁺ transporting, plasma membrane 4
NM_005173	ATPase, Ca ⁺⁺ transporting, ubiquitous
XM_030577	ATPase, Class II, type 9A
NM_000053	ATPase, Cu ⁺⁺ transporting, beta polypeptide (Wilson disease)
BC008861	ATPase, H ⁺ transporting, lysosomal 38kDa, V0 subunit d isoform 1
Y17975	ATPase, H ⁺ transporting, lysosomal accessory protein 2
AF245517	ATPase, H ⁺ transporting, lysosomal V0 subunit a isoform 4

AY048757 ATP-binding cassette, sub-family G (WHITE), member 1
 AF038563 Atrophin-1 interacting protein 1
 AB014605 Atrophin-1 interacting protein 1
 NM_139321 Attractin
 NM_012070 Attractin
 NM_001187 B melanoma antigen
 AB010894 BAI1-associated protein 1
 D13630 Basic leucine zipper and W2 domains 1
 BC008873 BBP-like protein 2
 BC018906 B-cell linker
 NM_005745 B-cell receptor-associated protein 31
 AF536326 BCL2/adenovirus E1B 19kDa interacting protein 3-like
 AB071199 BCL2-like 11 (apoptosis facilitator)
 AB071198 BCL2-like 11 (apoptosis facilitator)
 NM_020926 BCL6 co-repressor
 AF053470 Bladder cancer associated protein
 BC004248 Bone morphogenetic protein 7 (osteogenic protein 1)
 NM_007300 Breast cancer 1, early onset
 NM_007299 Breast cancer 1, early onset
 NM_007298 Breast cancer 1, early onset
 AF221130 Bromodomain adjacent to zinc finger domain, 1A
 NM_023005 Bromodomain adjacent to zinc finger domain, 1B
 AK074613 Bromodomain containing 7
 AK027308 Bromodomain containing 7
 AF152604 Bromodomain containing 7
 AB025106 Cadherin 1, type 1, E-cadherin (epithelial)
 L08599 Cadherin 1, type 1, E-cadherin (epithelial)
 BC028357 Calmegin
 M19311 Calmodulin 2 (phosphorylase kinase, delta)
 M23254 Calpain 2, (m/II) large subunit
 U31345 Calpastatin
 AF013759 Calumenin
 BC013963 CAP, adenylate cyclase-associated protein 1 (yeast)
 NM_004003 Carnitine acetyltransferase
 L39211 Carnitine palmitoyltransferase 1A (liver)
 BC000185 Carnitine palmitoyltransferase 1A (liver)
 AC004973 carrier protein-like; similar to Q01888 (PID:g266574);
 U26710 Cas-Br-M (murine) ecotropic retroviral transforming sequence b
 BC032851 Cas-Br-M (murine) ecotropic retroviral transforming sequence b
 AK026762 Cas-Br-M (murine) ecotropic retroviral transforming sequence-like 1
 L37042 Casein kinase 1, alpha 1
 AF447582 Casein kinase 1, alpha 1
 AB024597 Casein kinase 1, epsilon
 NM_033307 Caspase 4, apoptosis-related cysteine protease
 NM_033306 Caspase 4, apoptosis-related cysteine protease
 BC017839 Caspase 4, apoptosis-related cysteine protease
 NM_001227 Caspase 7, apoptosis-related cysteine protease
 AF110376 Caspase 9, apoptosis-related cysteine protease
 L23805 Catenin (cadherin-associated protein), alpha 1, 102kDa
 BC031262 Catenin (cadherin-associated protein), alpha 1, 102kDa

AF135021	Catenin (cadherin-associated protein), alpha-like 1
M14221	Cathepsin B
Y18462	Cathepsin L
AF109161	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2
NM_005194	CCAAT/enhancer binding protein (C/EBP), beta
AF263279	CD164 antigen, sialomucin
BC011988	CD9 antigen (p24)
BC031549	CDC-like kinase 1
NM_021873	Cell division cycle 25B
AJ297709	Cell division cycle 2-like 5 (cholinesterase-related cell division controller)
AK027507	Cereblon
AF015599	Ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)
AF015593	Ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)
AF161526	CGI-119 protein
AF026292	Chaperonin containing TCP1, subunit 7 (eta)
NM_001829	Chloride channel 3
AF102166	Chloride intracellular channel 3
AK001342	Chromodomain helicase DNA binding protein 1-like
AY034072	Chromodomain helicase DNA binding protein 6
NM_014837	Chromosome 1 open reading frame 16
AB050854	Chromosome 1 open reading frame 38
AF044896	Chromosome 1 open reading frame 38
AL035291	Chromosome 1 open reading frame 9
AF097535	Chromosome 1 open reading frame 9
BC015994	Chromosome 10 open reading frame 32
AK023325	Chromosome 10 open reading frame 57
AK026630	Chromosome 10 open reading frame 84
AF226047	Chromosome 11 open reading frame 30
NM_007211	Chromosome 12 open reading frame 2
AK000399	Chromosome 14 open reading frame 101
NM_016472	Chromosome 14 open reading frame 129
AK093969	Chromosome 14 open reading frame 65
AK027369	Chromosome 19 open reading frame 13
BC016842	Chromosome 19 open reading frame 13
AK074548	Chromosome 19 open reading frame 27
AK000562	Chromosome 2 open reading frame 18
AK023408	Chromosome 20 open reading frame 172
NM_031229	Chromosome 20 open reading frame 18
NM_031228	Chromosome 20 open reading frame 18
XM_045421	Chromosome 20 open reading frame 194
AK001745	Chromosome 20 open reading frame 36
AF490768	Chromosome 21 open reading frame 106
AF282851	Chromosome 21 open reading frame 59
AF157316	Chromosome 5 open reading frame 5
AK026774	Chromosome 6 open reading frame 103
AK000946	Chromosome 6 open reading frame 109
AK000972	Chromosome 9 open reading frame 40
AK027675	Cirrhosis, autosomal recessive 1A (cirhin)
BC006332	Clathrin, light polypeptide (Lcb)
BC016047	Claudin 23

AB000714	Claudin 3
AB000712	Claudin 4
AK055521	C-Mpl binding protein
BC010039	Coactosin-like 1 (Dictyostelium)
BC036757	Coiled-coil domain containing 6
NM_032040	Coiled-coil domain containing 8
NM_000088	Collagen, type I, alpha 1
BC008760	Collagen, type V, alpha 1
NM_001848	Collagen, type VI, alpha 1
NM_016094	COMM domain containing 2
BC000628	COMM domain containing 7
AF041207	contains ring finger domain; contains 61B3-R marker sequence;
BC003090	COP9 constitutive photomorphogenic homolog subunit 8 (Arabidopsis)
NM_001823	Creatine kinase, brain
AF091555	C-terminal binding protein 1
AF222711	C-terminal binding protein 2
AK075512	CTL2 gene
BC025422	Cut-like 1, CCAAT displacement protein (Drosophila)
AF084555	Cyclic AMP phosphoprotein, 19 kD
AF367477	Cyclin L1
AF208843	Cyclin L1
AF216965	Cyclin M3
AK022833	Cyclin M4
NM_003885	Cyclin-dependent kinase 5, regulatory subunit 1 (p35)
NM_001260	Cyclin-dependent kinase 8
AY008263	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)
L47232	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)
BC001935	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)
AY008263	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)
AF192466	Cysteine and histidine-rich domain (CHORD)-containing, zinc binding protein 1
AF123249	Cysteine and histidine-rich domain (CHORD)-containing, zinc binding protein 1
NM_001915	Cytochrome b-561
NM_001352	D site of albumin promoter (albumin D-box) binding protein
AB018268	DDHD domain containing 2
AK001467	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18
NM_001356	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
AF083255	DEAD (Asp-Glu-Ala-Asp) box polypeptide 42
NM_020162	DEAH (Asp-Glu-Ala-His) box polypeptide 33
BC036035	DEAH (Asp-Glu-Ala-His) box polypeptide 36
L13848	DEAH (Asp-Glu-Ala-His) box polypeptide 9
BC016392	Dedicator of cytokinesis 7
BC018148	Delta sleep inducing peptide, immunoreactor
AB023154	Deltex 4 homolog (Drosophila)
AF251700	Dendritic cell-derived ubiquitin-like protein
AB033075	Development and differentiation enhancing factor 1
AB007860	Development and differentiation enhancing factor 2
AY113704	DIP13 beta
BC001446	Dipeptidylpeptidase 3
AB017970	Dipeptidylpeptidase 3
NM_130434	Dipeptidylpeptidase 8

NM_017743 Dipeptidylpeptidase 8
 NM_014992 Dishevelled associated activator of morphogenesis 1
 BC024781 Dishevelled associated activator of morphogenesis 1
 BC013635 DKFZP434C212 protein
 AK023383 DKFZP564J0863 protein
 NM_014043 DKFZP564O123 protein
 NM_001379 DNA (cytosine-5-)-methyltransferase 1
 AF480163 DNA (cytosine-5-)-methyltransferase 3 alpha
 AF331856 DNA (cytosine-5-)-methyltransferase 3 alpha
 AF176228 DNA (cytosine-5-)-methyltransferase 3 beta
 AF156487 DNA (cytosine-5-)-methyltransferase 3 beta
 NM_145325 DNA directed RNA polymerase II polypeptide J-related gene
 AK002064 DNA polymerase-transactivated protein 6
 BC001620 DNA segment on chromosome X and Y (unique) 155 expressed sequence
 BC002352 DnaJ (Hsp40) homolog, subfamily B, member 1
 AB014888 DnaJ (Hsp40) homolog, subfamily B, member 6
 AF427142 DnaJ (Hsp40) homolog, subfamily C, member 1
 AB007942 DnaJ (Hsp40) homolog, subfamily C, member 6
 AF126743 DnaJ (Hsp40) homolog, subfamily D, member 1
 BC027860 Downregulated in ovarian cancer 1
 NM_017613 Downstream neighbor of SON
 AK026326 DRE1 protein
 AB052156 Dual specificity phosphatase 16
 U53530 Dynein, cytoplasmic, heavy polypeptide 1
 BC021297 Dynein, cytoplasmic, heavy polypeptide 1
 NM_003746 Dynein, cytoplasmic, light polypeptide 1
 BC007223 Dynein, cytoplasmic, light polypeptide 2A
 NM_032980 Dystrobrevin, alpha
 NM_001390 Dystrobrevin, alpha
 NM_004022 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004021 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004014 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004013 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004011 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004009 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004007 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 NM_004006 Dystrophin (muscular dystrophy, Duchenne and Becker types)
 U73844 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
 AF517841 E74-like factor 3 (ets domain transcription factor, epithelial-specific)
 AF115403 E74-like factor 5 (ets domain transcription factor)
 BC032630 Echinoderm microtubule associated protein like 2
 BC003033 EF hand domain family, member D2
 AF229245 Egl nine homolog 1 (C. elegans)
 AF323924 EH-domain containing 4
 AK023886 Engulfment and cell motility 3 (ced-12 homolog, C. elegans)
 L38734 Ephrin-B2
 NM_001981 Epidermal growth factor receptor pathway substrate 15
 AK000372 Epithelial protein lost in neoplasm beta
 AF198455 Epithelial protein lost in neoplasm beta
 AF030455 Epithelial V-like antigen 1

NM_133180 EPS8-like 1
 NM_017729 EPS8-like 1
 NM_139204 EPS8-like 1
 AF276423 ErbB2 interacting protein
 AK000388 Erythrocyte membrane protein band 4.1 like 4B
 U47678 Estrogen receptor 1
 M12674 Estrogen receptor 1
 AF258451 Estrogen receptor 1
 AF258449 Estrogen receptor 1
 AF124790 Estrogen receptor 2 (ER beta)
 AF060555 Estrogen receptor 2 (ER beta)
 AB006590 Estrogen receptor 2 (ER beta)
 AB006589 Estrogen receptor 2 (ER beta)
 AB007619 Estrogen receptor binding site associated, antigen, 9
 BC002513 Eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa
 AF110146 Eukaryotic translation initiation factor 2-alpha kinase 3
 BC000494 Eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa
 BC018727 Eukaryotic translation initiation factor 2C, 2
 AF121255 Eukaryotic translation initiation factor 2C, 2
 AK022827 Eukaryotic translation initiation factor 2C, 3
 AK023388 Eukaryotic translation initiation factor 3, subunit 1 alpha, 35kDa
 AF090923 Eukaryotic translation initiation factor 3, subunit 1 alpha, 35kDa
 BC001173 Eukaryotic translation initiation factor 3, subunit 9 eta, 116kDa
 BC014930 Eukaryotic translation initiation factor 4 gamma, 2
 AB076839 Eukaryotic translation initiation factor 4B
 U49436 Eukaryotic translation initiation factor 5
 NM_015904 Eukaryotic translation initiation factor 5B
 NM_013986 Ewing sarcoma breakpoint region 1
 AF327066 Ewing sarcoma breakpoint region 1
 NM_144502 F11 receptor
 AK097391 Family with sequence similarity 11, member A
 D87120 Family with sequence similarity 3, member C
 AK001444 Fas apoptotic inhibitory molecule
 AK001933 F-box and WD-40 domain protein 7 (archipelago homolog, Drosophila)
 NM_033624 F-box protein 21
 AY005144 F-box protein 22
 AK024048 F-box protein 22
 NM_033406 F-box protein 3
 NM_012177 F-box protein 5
 AF233223 F-box protein 6
 AL050254 F-box protein 7
 AK027381 F-box protein, helicase, 18
 AB007856 Fem-1 homolog b (C. elegans)
 AF159615 FGF receptor activating protein 1
 AF112481 Fibrinogen silencer binding protein
 AF007866 Fibrinogen silencer binding protein
 NM_023029 Fibroblast growth factor receptor 2 (bacteria-expressed kinase,
 NM_022976 Fibroblast growth factor receptor 2 (bacteria-expressed kinase,
 NM_022973 Fibroblast growth factor receptor 2 (bacteria-expressed kinase,
 AF238374 Fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)

NM_001456	Filamin A, alpha (actin binding protein 280)
AF238609	Filamin B, beta (actin binding protein 278)
AK092764	FLJ10378 protein
NM_021942	FLJ12716 protein
AF041336	Forkhead box O3A
AF265550	Formin binding protein 1
AF049524	Formin binding protein 3
AK022987	Formin binding protein 4
AF063002	Four and a half LIM domains 1
BC003060	Fucosidase, alpha-L- 2, plasma
AF419331	FUS interacting protein (serine-arginine rich) 1
AF047448	FUS interacting protein (serine-arginine rich) 1
AB040799	G protein-coupled receptor 27
AJ012188	G protein-coupled receptor 51
AF237763	G protein-coupled receptor 87
NM_057170	G protein-coupled receptor kinase interactor 2
NM_002040	GA binding protein transcription factor, alpha subunit 60kDa
NM_005254	GA binding protein transcription factor, beta subunit 2, 47kDa
M27508	Galactosidase, beta 1
L23116	Galactosylceramidase (Krabbe disease)
BC015613	GATA binding protein 2
BC020597	General transcription factor IIB
NM_005513	General transcription factor IIE, polypeptide 1, alpha 56kDa
BC030572	General transcription factor IIE, polypeptide 2, beta 34kDa
NM_004128	General transcription factor IIF, polypeptide 2, 30kDa
BC005345	General transcription factor IIH, polypeptide 2, 44kDa
AF291673	Giant axonal neuropathy (gigaxonin)
AF058922	GLE1 RNA export mediator-like (yeast)
BC005359	Glia maturation factor, beta
AF159851	Glucocorticoid receptor DNA binding factor 1
AB051509	Glucocorticoid receptor DNA binding factor 1
D83485	Glucose regulated protein, 58kDa
BC016634	Glutaminyl-tRNA synthetase
AF069668	Glutaredoxin (thioltransferase)
BC019874	Glycerol-3-phosphate dehydrogenase 2 (mitochondrial)
L33801	Glycogen synthase kinase 3 beta
BC009273	GM2 ganglioside activator
NM_080426	GNAS complex locus
NM_080425	GNAS complex locus
BC022875	GNAS complex locus
AF105253	GNAS complex locus
XM_093895	go_function: calcium ion binding [goid 0005509]
NM_007016	go_function: FAD binding [goid 0050660] [evidence IEA];
U31906	Golgi autoantigen, golgin subfamily a, 4
AK056277	Golgi autoantigen, golgin subfamily a, 4
AF273047	Golgi autoantigen, golgin subfamily a, 4
AF085199	Golgi autoantigen, golgin subfamily a, 5
BC008928	Golgi reassembly stacking protein 1, 65kDa
AK091168	Golgi reassembly stacking protein 1, 65kDa
BC001408	Golgi reassembly stacking protein 2, 55kDa

AK027349	Golgi reassembly stacking protein 2, 55kDa
AB020662	Golgin-67
M75161	Granulin
BC000324	Granulin
AK022142	GRB2-associated binding protein 1
NM_004864	Growth differentiation factor 15
BC000529	Growth differentiation factor 15
BC025415	GTP cyclohydrolase 1 (dopa-responsive dystonia)
AY137464	GTP cyclohydrolase 1 (dopa-responsive dystonia)
M16538	Guanine nucleotide binding protein (G protein), beta polypeptide 2
BC005940	Guanine nucleotide binding protein (G protein), gamma 12
BC001103	GULP, engulfment adaptor PTB domain containing 1
NM_138609	H2A histone family, member Y
BC018002	H2A histone family, member Z
AF218029	H3 histone, family 3B (H3.3B)
AK098265	H63 breast cancer expressed gene
AF039942	HCF-binding transcription factor Zhangfei
AF068284	HDCMA18P protein
AB033492	Headcase homolog (Drosophila)
AF133207	Heat shock 22kDa protein 8
M30627	Heat shock 90kDa protein 1, alpha-like 3
M65217	Heat shock transcription factor 2
AF177862	Hematological and neurological expressed 1
AF060925	Hematological and neurological expressed 1
BC002396	Heme oxygenase (decycling) 2
NM_147175	Heparan sulfate 6-O-sulfotransferase 2
M64283	Heparan sulfate proteoglycan 2 (perlecan)
AF116631	Hepatocellular carcinoma-associated antigen 66
AL109689	Hepatoma-derived growth factor, related protein 3
M29063	Heterogeneous nuclear ribonucleoprotein C (C1/C2)
BC008423	Heterogeneous nuclear ribonucleoprotein C (C1/C2)
BC003394	Heterogeneous nuclear ribonucleoprotein C (C1/C2)
D89678	Heterogeneous nuclear ribonucleoprotein D-like
NM_031262	Heterogeneous nuclear ribonucleoprotein K
NM_002140	Heterogeneous nuclear ribonucleoprotein K
BC014305	High density lipoprotein binding protein (vigilin)
AF146222	High-mobility group 20A
NM_005342	High-mobility group box 3
AL079310	High-mobility group protein 2-like 1
BC001287	Histidine triad nucleotide binding protein 1
U75696	Histone deacetylase 3
BC000614	Histone deacetylase 3
AF130111	Histone deacetylase 3
AF005482	Histone deacetylase 3
AB020708	Histone deacetylase 6
AF255717	HLA-A*0206; Homo sapiens MHC class I antigen (HLA-A) gene
NM_004639	HLA-B associated transcript 3
AB021179	HMBA-inducible
M16937	Homeo box B7
AF326592	Homeodomain interacting protein kinase 2

AF207702 Homeodomain interacting protein kinase 2
AF004849 Homeodomain interacting protein kinase 3
AF349467 Homo sapiens amplified in breast cancer 2 (ABC2) mRNA, complete cds.
U15174 Homo sapiens BCL2/adenovirus E1B 19kD-interacting protein 3 (BNIP3) mRNA,
AF549495 Homo sapiens cyclin G2 (CCNG2) gene, complete cds.
AF159621 Homo sapiens FGF receptor activating protein 1 (FRAG1) gene, exon 6
XM_088328 Homo sapiens hypothetical protein LOC157567 (LOC157567), mRNA.
XM_035825 Homo sapiens KIAA0143 protein (KIAA0143), mRNA.
XM_042685 Homo sapiens KIAA1414 protein (KIAA1414), mRNA.
AJ005577 Homo sapiens pfkfb2 gene, exons 1 to 15.
AF022382 Homo sapiens UDP-galactose 4' epimerase (GALE) gene, complete cds.
BC017763 HSPB (heat shock 27kDa) associated protein 1
AF125099 HSPC038 protein
NM_014181 HSPC159 protein
J05011 Human immunodeficiency virus type I enhancer binding protein 1
NM_014159 Huntingtin interacting protein B
BC007882 Hypothetical protein BC007882
NM_018982 Hypothetical protein DJ167A19.1
NM_031476 Hypothetical protein DKFZp434B044
BC007689 Hypothetical protein DKFZp434B044
NM_031306 Hypothetical protein DKFZp564B1023
AK001138 Hypothetical protein FLJ10276
AK001162 Hypothetical protein FLJ10300
AK022618 Hypothetical protein FLJ10407
AK023532 Hypothetical protein FLJ10706
AK001670 Hypothetical protein FLJ10808
AK001763 Hypothetical protein FLJ10901
AK001842 Hypothetical protein FLJ10980
BC010221 Hypothetical protein FLJ11046
AK000289 Hypothetical protein FLJ11171
AK002121 Hypothetical protein FLJ11259
BC006205 Hypothetical protein FLJ12895
AK023518 Hypothetical protein FLJ13456
AK022815 Hypothetical protein FLJ13910
AK027659 Hypothetical protein FLJ14753
AF255650 Hypothetical protein FLJ14753
AK000287 Hypothetical protein FLJ20280
AK000320 Hypothetical protein FLJ20313
AF153417 Hypothetical protein FLJ20457
AK025561 Hypothetical protein FLJ21908
AK026110 Hypothetical protein FLJ22457
AK026184 Hypothetical protein FLJ22531
AK026315 Hypothetical protein FLJ22662
AK026486 Hypothetical protein FLJ22833
AK074370 Hypothetical protein FLJ23790
AF523265 Hypothetical protein FLJ30435
BC012198 Hypothetical protein FLJ30525
XM_172341 Hypothetical protein FLJ35036
AK096044 Hypothetical protein FLJ38725
AF161527 Hypothetical protein HSPC138

AF208694	Hypothetical protein IMPACT
BC008502	Hypothetical protein LOC201895
AB009282	Hypothetical protein LOC283852
NM_016627	Hypothetical protein LOC51321
AF217995	Hypothetical protein LOC92497
BC007072	Hypothetical protein MGC12538
BC006110	Hypothetical protein MGC12966
BC009283	Hypothetical protein MGC14327
BC032461	Hypothetical protein MGC40405
BC004208	Hypothetical protein MGC4268
NM_144736	Hypothetical protein PRO1853
AF182645	IK cytokine, down-regulator of HLA II
AB007935	Immunoglobulin superfamily, member 3
NM_016389	Influenza virus NS1A binding protein
M13144	Inhibin, alpha
BC001864	Inositol hexaphosphate kinase 2
M10051	Insulin receptor
M29929	insulin receptor (AA at 78); Human insulin receptor (allele 1) gene
NM_000876	Insulin-like growth factor 2 receptor
NM_002205	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
M13918	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
NM_033666	Integrin, beta 1 (fibronectin receptor, beta polypeptide)
NM_002211	Integrin, beta 1 (fibronectin receptor, beta polypeptide)
NM_000213	Integrin, beta 4
AF011375	Integrin, beta 4
U88964	Interferon stimulated gene 20kDa
M13755	Interferon, alpha-inducible protein (clone IFI-15K)
NM_004515	Interleukin enhancer binding factor 2, 45kDa
NM_004516	Interleukin enhancer binding factor 3, 90kDa
AK000018	Interleukin enhancer binding factor 3, 90kDa
AK000302	Intersectin 2
AF248540	Intersectin 2
AF116615	JTV1 gene
BC002853	JTV1 gene
L40411	Jumonji domain containing 1C
NM_005056	Jumonji, AT rich interactive domain 1A (RBBP2-like)
AF087481	Jumonji, AT rich interactive domain 1B (RBP2-like)
NM_002264	Karyopherin alpha 1 (importin alpha 5)
AB002533	Karyopherin alpha 4 (importin alpha 3)
AK001698	Kelch-like 5 (Drosophila)
AF123320	Kelch-like 5 (Drosophila)
AF111113	Kelch-like 7 (Drosophila)
BC034697	Keratin 10 (epidermolytic hyperkeratosis; keratosis palmaris et plantaris)
NM_005556	Keratin 7
NM_002284	Keratin, hair, basic, 6 (monilethrix)
BC028604	KIAA0052
D43949	KIAA0082
D86971	KIAA0217
BC003381	KIAA0217
D87438	KIAA0251 protein

AB002308	KIAA0310
XM_036708	KIAA0368
XM_051017	KIAA0657 protein
AF201291	KIAA0669 gene product
AB018345	KIAA0802
NM_015070	KIAA0853
AB020684	KIAA0877 protein
NM_015072	KIAA0998
AK024259	KIAA0998
AB023229	KIAA1012
AB033022	KIAA1196 protein
AL132998	KIAA1219 protein
BC016987	KIAA1327 protein
BC006215	KIAA1387 protein
AK000573	KIAA1404 protein
AB037835	KIAA1414 protein
BC007671	KIAA1463 protein
AB046861	KIAA1641
AK091423	KIAA1946
BC008881	Kinesin 2 60/70kDa
BC017705	Kinesin family member 23
NM_004520	Kinesin heavy chain member 2
AB030824	Kruppel-like factor 5 (intestinal)
NM_003937	Kynureninase (L-kynurenine hydrolase)
AF151841	Lactamase, beta 2
AF189062	LAG1 longevity assurance homolog 2 (<i>S. cerevisiae</i>)
NM_002292	Laminin, beta 2 (laminin S)
AK094050	Laminin, beta 2 (laminin S)
AK022001	Latrophilin 1
AB028019	LATS, large tumor suppressor, homolog 2 (<i>Drosophila</i>)
BC015761	Lectin, galactoside-binding, soluble, 3 binding protein
L13210	Lectin, galactoside-binding, soluble, 3 binding protein
BC006503	Lethal giant larvae homolog 2 (<i>Drosophila</i>)
NM_004524	Lethal giant larvae homolog 2 (<i>Drosophila</i>)
U69609	Leucine rich repeat (in FLII) interacting protein 1
BC004958	Leucine rich repeat (in FLII) interacting protein 1
AF241787	Likely ortholog of mouse acyl-Coenzyme A thioesterase 2, mitochondrial
AF132950	Likely ortholog of mouse acyl-Coenzyme A thioesterase 2, mitochondrial
AB018290	Likely ortholog of mouse membrane bound C2 domain containing protein
NM_030952	Likely ortholog of rat SNF1/AMP-activated protein kinase
BC007560	LIM and SH3 protein 1
BC003600	LIM domain only 4
AK094511	LIM homeobox 2
AB015292	Limkain beta 2
AF087693	Lin-7 homolog A (<i>C. elegans</i>)
NM_145693	Lipin 1
AY005111	LUC7-like (<i>S. cerevisiae</i>)
J04183	Lysosomal-associated membrane protein 2
AK001088	Macrophage erythroblast attacher
D64146	Major histocompatibility complex, class I, C

AK025266 MAK10 homolog, amino-acid N-acetyltransferase subunit, (*S. cerevisiae*)
BC037400 Makorin, ring finger protein, 1
AF027156 Mannosidase, alpha, class 1A, member 2
U31520 Mannosidase, alpha, class 2A, member 1
AY048774 Mannosidase, endo-alpha
AB000409 MAP kinase interacting serine/threonine kinase 1
AK001730 MAP kinase interacting serine/threonine kinase 2
M80359 MAP/microtubule affinity-regulating kinase 3
AF221759 Mastermind-like 1 (*Drosophila*)
XM_031689 MAX gene associated
AF275813 McKusick-Kaufman syndrome
AJ010089 MCM3 minichromosome maintenance deficient 3 (*S. cerevisiae*) associated protein
AF007130 Mdm4, transformed 3T3 cell double minute 1, p53 binding protein (mouse)
NM_002364 Melanoma antigen family B, 2
NM_006986 Melanoma antigen family D, 1
M28882 Melanoma cell adhesion molecule
AK027042 Membrane associated DNA binding protein
AK000720 Membrane associated DNA binding protein
BC030594 Membrane cofactor protein (CD46, trophoblast-lymphocyte cross-reactive antigen)
BC009808 Membrane component, chromosome 17, surface marker 2
AB014579 Meningioma expressed antigen 5 (hyaluronidase)
BC001373 Mesoderm development candidate 1
BC009443 Metastasis associated 1
BC005218 Methionine adenosyltransferase II, beta
NM_015847 Methyl-CpG binding domain protein 1
NM_015846 Methyl-CpG binding domain protein 1
NM_015845 Methyl-CpG binding domain protein 1
NM_015832 Methyl-CpG binding domain protein 2
AF072242 Methyl-CpG binding domain protein 2
BC026244 Microfibrillar-associated protein 3
NM_138634 Microseminoprotein, beta-
AF230977 Midline 1 (Opitz/BBB syndrome)
AF161556 Mitochondrial ribosomal protein L18
AY061855 Mitochondrial ribosomal protein S6
AF329637 Mitofusin 1
L35263 Mitogen-activated protein kinase 14
AF100544 Mitogen-activated protein kinase 14
AB074150 Mitogen-activated protein kinase 14
AF420474 Mitogen-activated protein kinase 6
NM_139046 Mitogen-activated protein kinase 8
M37191 Mitogen-activated protein kinase associated protein 1
L05624 Mitogen-activated protein kinase kinase 1
BC026245 Mitogen-activated protein kinase kinase 1 interacting protein 1
NM_145110 Mitogen-activated protein kinase kinase 3
NM_145687 Mitogen-activated protein kinase kinase kinase kinase 4
NM_139078 Mitogen-activated protein kinase-activated protein kinase 5
BC024238 MKI67 (FHA domain) interacting nucleolar phosphoprotein
AK024029 Modulator of apoptosis 1
AJ270950 Monoglyceride lipase
AF218011 Mortality factor 4 like 1

AL137163	Motile sperm domain containing 1
AK023297	Mov10, Moloney leukemia virus 10, homolog (mouse)
AF293076	MRS2-like, magnesium homeostasis factor (<i>S. cerevisiae</i>)
AK027314	Mucin 20
AK023527	Multimerin 2
AK093888	Musashi homolog 2 (<i>Drosophila</i>)
NM_144778	Muscleblind-like 2 (<i>Drosophila</i>)
AF047489	Muskelin 1, intracellular mediator containing kelch motifs
AF106685	Myelin expression factor 2
AF118124	Myeloid cell leukemia sequence 1 (BCL2-related)
BC013589	Myeloid differentiation primary response gene (88)
S78570	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
L04284	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
L01986	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
D14540	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF492830	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF272382	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF272381	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF272375	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF232001	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF231999	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
NM_004641	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
AF487905	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)
BC007237	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>);
AF187016	Myosin regulatory light chain interacting protein
BC016372	Myosin regulatory light chain MRCL3
U90236	Myosin VI
NM_002473	Myosin, heavy polypeptide 9, non-muscle
M81105	Myosin, heavy polypeptide 9, non-muscle
M69180	Myosin, heavy polypeptide 9, non-muscle
BC032947	Myotubularin related protein 1
AK055513	Myotubularin related protein 1
AJ224979	Myotubularin related protein 1
D10522	Myristoylated alanine-rich protein kinase C substrate
AF085355	N-acetyltransferase 5 (ARD1 homolog, <i>S. cerevisiae</i>)
AK023114	NAD kinase
BC007659	NAD(P)H dehydrogenase, quinone 1
AF087661	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 10, 42kDa
AF077028	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa
NM_002525	Nardilysin (N-arginine dibasic convertase)
AF043119	NCK adaptor protein 2
NM_006393	Nebulette
NM_003826	N-ethylmaleimide-sensitive factor attachment protein, gamma
D42055	Neural precursor cell expressed, developmentally down-regulated 4
D23662	Neural precursor cell expressed, developmentally down-regulated 8
AF493919	Neuroblastoma RAS viral (v-ras) oncogene homolog
M38106	Neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)
AF145712	Neuropilin 1
NM_012343	Nicotinamide nucleotide transhydrogenase
BC032370	Nicotinamide nucleotide transhydrogenase

BC002532	Niemann-Pick disease, type C2
AF283538	NIF3 NGG1 interacting factor 3-like 1 (S. pombe)
AF058696	Nijmegen breakage syndrome 1 (nibrin)
BC014991	N-methylpurine-DNA glycosylase
NM_004688	N-myc (and STAT) interactor
NM_006096	N-myc downstream regulated gene 1
M73980	Notch homolog 1, translocation-associated (Drosophila)
AK027869	Notch homolog 2 (Drosophila)
NM_000435	Notch homolog 3 (Drosophila)
NM_003204	Nuclear factor (erythroid-derived 2)-like 1
AF133059	Nuclear factor (erythroid-derived 2)-like 3
U59302	Nuclear receptor coactivator 1
NM_003743	Nuclear receptor coactivator 1
AF010227	Nuclear receptor coactivator 3
BC001562	Nuclear receptor coactivator 4
NM_020967	Nuclear receptor coactivator 5
AF044209	Nuclear receptor co-repressor 1
M29959	Nuclear receptor subfamily 2, group C, member 1
M96824	Nucleobindin 1
NM_139132	Nucleoporin 98kDa
NM_016320	Nucleoporin 98kDa
AB040538	Nucleoporin 98kDa
AK098544	Nucleoside phosphorylase
AF116670	Nucleoside phosphorylase
NM_031438	Nudix (nucleoside diphosphate linked moiety X)-type motif 12
AF191651	Nudix (nucleoside diphosphate linked moiety X)-type motif 4
BC020788	Numb homolog (Drosophila)
AF171940	Numb homolog (Drosophila)
AF223393	O-linked N-acetylglucosamine (GlcNAc) transferase
AF420371	Optineurin
AF392447	Oxysterol binding protein-like 2
BC003049	PAI-1 mRNA-binding protein
AF151813	PAI-1 mRNA-binding protein
AF089897	PAP associated domain containing 5
AF320308	Paraneoplastic antigen MA1
BC003647	Paxillin
AB029018	PDZ domain containing RING finger 3
AF104012	Peptidylprolyl isomerase E (cyclophilin E)
L27841	Pericentriolar material 1
BC000453	Pericentriolar material 1
AF068293	Pericentriolar material 1
BC007063	Peroxiredoxin 1
AF072864	Peroxisomal membrane protein 4, 24kDa
NM_018288	PHD finger protein 10
AF017786	Phosphatidic acid phosphatase type 2B
AB000889	Phosphatidic acid phosphatase type 2B
NM_021213	Phosphatidylcholine transfer protein
NM_007166	Phosphatidylinositol binding clathrin assembly protein
NM_020472	Phosphatidylinositol glycan, class A (paroxysmal nocturnal hemoglobinuria)
BC038236	Phosphatidylinositol glycan, class A (paroxysmal nocturnal hemoglobinuria)

AF332652	Phosphodiesterase 7A
D88532	Phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma)
NM_000293	Phosphorylase kinase, beta
NM_015092	PI-3-kinase-related kinase SMG-1
AF186377	PI-3-kinase-related kinase SMG-1
NM_002687	Pinin, desmosome associated protein
AF149785	Pituitary tumor-transforming 1 interacting protein
NM_003628	Plakophilin 4
M22299	Plastin 3 (T isoform)
BC001136	Pleckstrin homology domain containing
AK000790	Pleckstrin homology domain containing, family B (evectins) member 2
AB016533	Polyglutamine binding protein 1
BC029269	Polyhomeotic-like 2 (Drosophila)
AF089896	Polymerase (DNA directed) sigma
BC006980	Postsynaptic protein CRIPT
AF275817	PR domain containing 10
AF426259	PR domain containing 15
NM_015866	PR domain containing 2, with ZNF domain
NM_012231	PR domain containing 2, with ZNF domain
D45132	PR domain containing 2, with ZNF domain
M94890	Pregnancy specific beta-1-glycoprotein 9
NM_015387	Preimplantation protein 3
NM_007319	Presenilin 1 (Alzheimer disease 3)
U40379	Presenilin 1 (Alzheimer disease 3)
NM_007319	Presenilin 1 (Alzheimer disease 3)
BC002475	Profilin 1
AK026325	Progesterone and adipoQ receptor family member VI
BC002506	Programmed cell death 10
AF035606	Programmed cell death 6
AK002122	Programmed cell death 6 interacting protein
AB037796	Programmed cell death 6 interacting protein
BC035395	Proline-serine-threonine phosphatase interacting protein 2
NM_000532	Propionyl Coenzyme A carboxylase, beta polypeptide
AF217984	Propionyl Coenzyme A carboxylase, beta polypeptide
AF054185	Proteasome (prosome, macropain) subunit, alpha type, 7
AF128536	Protein kinase C and casein kinase substrate in neurons 2
BC001000	Protein kinase C, eta
NM_002709	Protein phosphatase 1, catalytic subunit, beta isoform
AK027650	Protein phosphatase 1, regulatory (inhibitor) subunit 15B
D26445	Protein phosphatase 2, regulatory subunit B (B56), gamma isoform
NM_000944	Protein phosphatase 3 (formerly 2B), catalytic subunit, Protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)
M29550	
L39000	Protein tyrosine phosphatase type IVA, member 2
NM_014369	Protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
NM_007039	Protein tyrosine phosphatase, non-receptor type 21
NM_002829	Protein tyrosine phosphatase, non-receptor type 3
AJ430580	Protein tyrosine phosphatase, receptor type, E
NM_003626	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF)
D49354	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF)
D25545	Protein-L-isoaspartate (D-aspartate) O-methyltransferase

BC007424	PRP4 pre-mRNA processing factor 4 homolog (yeast)
BC035404	PTK2 protein tyrosine kinase 2
AK074520	PTX1 protein
M96684	Purine-rich element binding protein A
AF042384	Putative breast adenocarcinoma marker (32kD)
BC000842	Putative homeodomain transcription factor 1
AJ420896	Putative intramembrane cleaving protease
AF070626	Putative membrane protein
AF100737	Putative translation initiation factor
AB026723	Pyrophosphatase (inorganic)
J03503	Pyruvate dehydrogenase (lipoamide) alpha 1
BC020790	Rab geranylgeranyltransferase, beta subunit
AF095352	RAB, member of RAS oncogene family-like 2B
AF070629	RAB2, member RAS oncogene family
BC010006	RAB24, member RAS oncogene family
AF258583	RAB24, member RAS oncogene family
BC009831	RAB25, member RAS oncogene family
AK027671	RAB3A interacting protein (rabin3)
AF498937	RAB5B, member RAS oncogene family
AF267863	RAB5B, member RAS oncogene family
BC013728	RAB7, member RAS oncogene family
BC007120	Rad50-interacting protein 1
L02320	Radixin
BC013126	RalA binding protein 1
NM_005054	RAN binding protein 2-like 1
D21239	Rap guanine nucleotide exchange factor (GEF) 1
AB002311	Rap guanine nucleotide exchange factor (GEF) 2
NM_002886	RAP2B, member of RAS oncogene family
BC033015	RAS p21 protein activator (GTPase activating protein) 1
U26914	Ras responsive element binding protein 1
AF178983	Ras-associated protein Rap1
AF498964	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)
BC004159	REC8-like 1 (yeast)
AF113538	Receptor associated protein 80
L07872	Recombining binding protein suppressor of hairless (Drosophila)
BC005103	RecQ protein-like 5
NM_002928	Regulator of G-protein signalling 16
M63488	Replication protein A1, 70kDa
AF219990	responsible for macular corneal dystrophy; C-GlcNAc6ST
AF087901	Reticulon 4
AF063601	Reticulon 4
NM_000328	Retinitis pigmentosa GTPase regulator
M19701	Retinoblastoma 1 (including osteosarcoma)
AF352051	Retinoblastoma binding protein 6
AF116625	Retinoblastoma binding protein 6
NM_005611	Retinoblastoma-like 2 (p130)
NM_000965	Retinoic acid receptor, beta
M62303	Retinoic acid receptor, beta
AF157483	Retinoic acid receptor, beta
NM_002957	Retinoid X receptor, alpha

AF498969 Rho family GTPase 3
BC022931 Rho GTPase activating protein 21
NM_003899 Rho guanine nucleotide exchange factor (GEF) 7
AB018283 Rho-related BTB domain containing 1
U49083 Ribosomal protein L29
BC008926 Ribosomal protein L29
AB062291 Ribosomal protein L3
M60725 Ribosomal protein S6 kinase, 70kDa, polypeptide 1
AJ010346 Ring finger protein (C3H2C3 type) 6
NM_014868 Ring finger protein 10
AF155650 Ring finger protein 130
BC005084 Ring finger protein 135
BC018107 Ring finger protein 138
BC018104 Ring finger protein 141
AK090864 Ring finger protein 170
AK024996 Ring finger protein 38
AF013591 RIO kinase 3 (yeast)
NM_002896 RNA binding motif protein 4
AF262323 RNA-binding protein
AB023967 ROD1 regulator of differentiation 1 (*S. pombe*)
D89788 Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
D10570 Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
BC010936 RWD domain containing 3
S59184 RYK receptor-like tyrosine kinase
NM_002958 RYK receptor-like tyrosine kinase
BC005928 S100 calcium binding protein A8 (calgranulin A)
AF304450 Sarcolemma associated protein
NM_002967 Scaffold attachment factor B
NM_016106 Sec1 family domain containing 1
AF118083 SEC10-like 1 (*S. cerevisiae*)
BC005032 Sec23 homolog B (*S. cerevisiae*)
AF119883 SEC63-like (*S. cerevisiae*)
AF100141 SEC63-like (*S. cerevisiae*)
BC001511 SEC6-like 1 (*S. cerevisiae*)
AK023078 SECIS binding protein 2
BC001376 Secretory carrier membrane protein 2
AF052059 Sel-1 suppressor of lin-12-like (*C. elegans*)
AB020335 Sel-1 suppressor of lin-12-like (*C. elegans*)
AF328864 Selenoprotein S
M17783 Serine (or cysteine) proteinase inhibitor
Y14485 Serine hydroxymethyltransferase 1 (soluble)
AF048977 Serine/arginine repetitive matrix 1
BC023508 Serine/threonine kinase 17a (apoptosis-inducing)
AF039692 Serologically defined colon cancer antigen 10
BC028671 SET domain, bifurcated 1
BC009362 SET domain, bifurcated 1
AF448510 SET domain-containing protein 7
AF042081 SH3 domain binding glutamic acid-rich protein like
AF247790 SH3 domain binding glutamic acid-rich protein like 3
AF147747 SH3-domain binding protein 4

AB055660	Shroom-related protein
NM_006278	Sialyltransferase 4C (beta-galactoside alpha-2,3-sialyltransferase)
NM_003135	Signal recognition particle 19kDa
BC000652	Signal recognition particle 54kDa
AF077019	Signal recognition particle 72kDa
NM_003139	Signal recognition particle receptor ('docking protein')
NM_003144	Signal sequence receptor, alpha (translocon-associated protein alpha)
BC002704	Signal transducer and activator of transcription 1, 91kDa
AB037810	Signal-induced proliferation-associated 1 like 2
AK026660	Similar to hypothetical protein MGC17347
BC024033	Sine oculis homeobox homolog 2 (Drosophila)
U70730	SKI-like
NM_015523	Small fragment nuclease
NM_004757	Small inducible cytokine subfamily E, member 1 (endothelial monocyte-activating)
NM_005496	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)
AB062294	SMT3 suppressor of mif two 3 homolog 1 (yeast)
BC016775	SMT3 suppressor of mif two 3 homolog 2 (yeast)
AF400432	SNRPN upstream reading frame
AF101044	SNRPN upstream reading frame
AF105365	Solute carrier family 12 (potassium/chloride transporters), member 7
AK075320	Solute carrier family 17 (anion/sugar transporter), member 5
NM_003054	Solute carrier family 18 (vesicular monoamine), member 2
BC018514	Solute carrier family 19 (thiamine transporter), member 2
AY007135	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator)
BC014416	Solute carrier family 33 (acetyl-CoA transporter), member 1
D87969	Solute carrier family 35 (CMP-sialic acid transporter), member A1
BC035747	Solute carrier family 35 (UDP-galactose transporter), member A2
BC011888	Solute carrier family 35, member B1
AB007916	Solute carrier family 35, member E2
AK023080	Solute carrier family 35, member F2
NM_032295	Solute carrier family 37 (glycerol-3-phosphate transporter), member 3
BC029379	Solute carrier family 38, member 2
AB037803	Solute carrier family 38, member 2
AK098651	Solute carrier family 39 (zinc transporter), member 13
U17986	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8
AF030409	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 6
NM_138926	SON DNA binding protein
NM_138925	SON DNA binding protein
NM_058183	SON DNA binding protein
L13857	Son of sevenless homolog 1 (Drosophila)
L12387	Sorcin
NM_003105	Sortilin-related receptor, L(DLR class) A repeats-containing
AF171229	Sorting nexin 12
AY044865	Sorting nexin 14
BC003382	Sorting nexin 2
AF130078	Sorting nexin 4
AF060509	Sorting nexin family member 27
D87465	Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2
AB029006	Spastic paraplegia 4 (autosomal dominant; spastin)
S54769	Spastic paraplegia 7, paraplegin (pure and complicated autosomal recessive)

BC007692	Spastic paraplegia 7, paraplegin (pure and complicated autosomal recessive)
BC001269	Speckle-type POZ protein
J05243	Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)
AF311312	Sperm associated antigen 1
NM_003971	Sperm associated antigen 9
AF200328	Sphingosine kinase 1
D63879	Squamous cell carcinoma antigen recognised by T cells 3
AK075528	SREBP cleavage-activating protein
AF116571	SRY (sex determining region Y)-box 13
AF083105	SRY (sex determining region Y)-box 13
AB007891	Start codon is not identified; Homo sapiens KIAA0431 mRNA, partial cds.
AB029007	Start codon is not identified; Homo sapiens mRNA for KIAA1084 protein, partial cds.
BC004936	Stearoyl-CoA desaturase 4
AK024685	Stearoyl-CoA desaturase 4
AF187981	Sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like
BC010703	Stomatin
AF380833	Stonin 2
AF109126	Stromal cell derived factor receptor 1
BC007716	Succinate-CoA ligase, GDP-forming, beta subunit
BC028583	SUMO1/sentrin specific protease 6
AF217504	SUMO1/sentrin specific protease 7
AF151697	SUMO1/sentrin/SMT3 specific protease 2
NM_003932	Suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)
NM_007192	Suppressor of Ty 16 homolog (S. cerevisiae)
AF027150	Survival of motor neuron protein interacting protein 1
AK026431	Sushi domain containing 2
AF210818	SWAP-70 protein
NM_003069	SWI/SNF related, matrix associated, actin dependent regulator of chromatin
NM_003601	SWI/SNF related, matrix associated, actin dependent regulator of chromatin
AJ295938	Synaptobrevin-like 1
BC000407	Synaptogyrin 2
BC014087	Synaptogyrin 3
AB028952	Synaptopodin
U55936	Synaptosomal-associated protein, 23kDa
NM_130798	Synaptosomal-associated protein, 23kDa
BC015540	Synaptotagmin-like 2
J05392	Syndecan 1
XM_049683	synonym: KIAA1203; ubiquitin specific proteinase 31;
NM_018459	synonyms: BM045, HIBDL; 3-hydroxyisobutyrate dehydrogenase-like;
NM_005673	synonyms: GDA, GDC, ML7, hML7, HGT.1, D10S105E, MGC39851;
NM_016837	synonyms: YC1, MSSP, SCR2, MSSP-1, MSSP-2, MSSP-3, MGC3331, MGC15146;
BC019042	Syntaxin 16
NM_003184	TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa
NM_003194	TATA box binding protein
AB024057	TBC1 domain family, member 8 (with GRAM domain)
BC026031	T-box 6
AF212247	T-cell immunomodulatory protein
AF018167	Telomerase reverse transcriptase
AF015950	Telomerase reverse transcriptase
AB086950	Telomerase reverse transcriptase

NM_017489 Telomeric repeat binding factor (NIMA-interacting) 1
 BC029378 Telomeric repeat binding factor (NIMA-interacting) 1
 AF289599 Telomeric repeat binding factor 2, interacting protein
 AF262988 Telomeric repeat binding factor 2, interacting protein
 NM_003217 Testis enhanced gene transcript (BAX inhibitor 1)
 XM_027236 Tetratricopeptide repeat domain 9
 AF255310 TFIIA-alpha/beta-like factor
 NM_016397 TH1-like (Drosophila)
 AJ238379 TH1-like (Drosophila)
 AF161479 TH1-like (Drosophila)
 BC001312 Thioredoxin domain containing 7 (protein disulfide isomerase)
 NM_006472 Thioredoxin interacting protein
 D88687 Thioredoxin reductase 1
 AF441770 THO complex 2
 NM_004239 Thyroid hormone receptor interactor 11
 BC002656 Thyroid hormone receptor interactor 11
 AF011368 Thyroid hormone receptor interactor 11
 BC000866 Tissue inhibitor of metalloproteinase 1
 NM_003313 Tissue specific transplantation antigen P35B
 J03250 Topoisomerase (DNA) I
 NM_001068 Topoisomerase (DNA) II beta 180kDa
 AF246219 Tousled-like kinase 1
 AB023508 TP53TG3 protein
 AF223469 TRAF and TNF receptor associated protein
 NM_004180 TRAF family member-associated NFkB activator
 NM_006756 Transcription elongation factor A (SII), 1
 M62810 Transcription factor A, mitochondrial
 NM_006365 Transcriptional activator of the c-fos promoter
 AJ277275 Transcriptional regulating factor 1
 M99435 Transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)
 AF076974 Transformation/transcription domain-associated protein
 AK027527 Transforming growth factor beta regulator 1
 NM_003238 Transforming growth factor, beta 2
 AY139007 Transforming, acidic coiled-coil containing protein 1
 AB029026 Transforming, acidic coiled-coil containing protein 1
 AF027516 Trans-golgi network protein 2
 AF030162 Translocase of inner mitochondrial membrane 23 homolog (yeast)
 BC000687 Translocation associated membrane protein 1
 AF007118 Transmembrane phosphatase with tensin homology
 AF090890 Transmembrane protein 23
 AK001338 Transmembrane protein vezatin
 AY128643 Transmembrane, prostate androgen induced RNA
 AF250310 Tribbles homolog 1 (Drosophila)
 AK001722 Trinucleotide repeat containing 17
 AB029036 Tripartite motif-containing 33
 BC013166 Tripartite motif-containing 44
 NM_007118 Triple functional domain (PTPRF interacting)
 AK023385 Tropomyosin 4
 AK024096 Tubulin, delta 1
 AF201333 Tubulin, delta 1

AF119121	Tudor and KH domain containing
AK075001	Tuftelin 1
BC028607	Tumor differentially expressed 2
M92357	Tumor necrosis factor, alpha-induced protein 2
AF099935	Tumor necrosis factor, alpha-induced protein 8
BC018117	Tumor protein D52
AF004430	Tumor protein D52-like 2
AF004429	Tumor protein D52-like 2
AB017926	Tumor protein p53 inducible nuclear protein 1
NM_006765	Tumor suppressor candidate 3
AF089814	Tumor suppressor deleted in oral cancer-related 1
AF222340	Type 1 tumor necrosis factor receptor shedding aminopeptidase regulator
X57346	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
AF107406	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
BC022094	Ubiquitin specific protease 30
AF383173	Ubiquitin specific protease 33
AF383172	Ubiquitin specific protease 33
AK091712	Ubiquitin specific protease 38
AK023092	Ubiquitin specific protease 49
NM_003470	Ubiquitin specific protease 7 (herpes virus-associated)
D29956	Ubiquitin specific protease 8
AB044550	Ubiquitin-conjugating enzyme E2 variant 1
BC007656	Ubiquitin-conjugating enzyme E2C
BC002775	Ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, <i>C. elegans</i>)
BC006277	Ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)
U31882	Ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)
AK001930	Ubiquitin-conjugating enzyme E2-like
BC000848	Ubiquitin-conjugating enzyme E2Q (putative)
AF305827	Ubiquitin-fold modifier 1
AF044221	Ubiquitin-like 3
D87684	UBX domain containing 2
NM_021139	UDP glycosyltransferase 2 family, polypeptide B4
AB060691	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 3
D50840	UDP-glucose ceramide glucosyltransferase
NM_006759	UDP-glucose pyrophosphorylase 2
AK075170	UDP-glucuronate decarboxylase 1
NM_017540	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
AK025287	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
BC010659	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
AK022753	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase
AY026825	UL16 binding protein 2
AF259961	UMP-CMP kinase
AK074938	unnamed protein product; Homo sapiens cDNA FLJ90457 fis, clone NT2RP3001685
AK096191	UPF2 regulator of nonsense transcripts homolog (yeast)
NM_080687	UPF3 regulator of nonsense transcripts homolog A (yeast)
BC004304	Uroplakin 3B
M35296	V-abl Abelson murine leukemia viral oncogene homolog 2 (arg, Abelson-related gene)
BC022505	Vacuolar protein sorting 26 (yeast)
AF175266	Vacuolar protein sorting 26 (yeast)
BC024218	Vacuolar protein sorting 35 (yeast)

AF135593 Vacuolar protein sorting 41 (yeast)
 AF085234 V-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)
 NM_003371 Vav 2 oncogene
 NM_005206 V-crk sarcoma virus CT10 oncogene homolog (avian)
 M11730 V-erb-b2 erythroblastic leukemia viral oncogene homolog 2,
 S61953 V-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
 NM_006370 Vesicle transport through interaction with t-SNAREs homolog 1B (yeast)
 AF053233 Vesicle-associated membrane protein 8 (endobrevin)
 J05021 Villin 2 (ezrin)
 AF059194 V-maf musculoaponeurotic fibrosarcoma oncogene homolog K (avian)
 BC011603 V-rel reticuloendotheliosis viral oncogene homolog A,
 M15990 V-yes-1 Yamaguchi sarcoma viral oncogene homolog 1
 M79321 V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
 NM_006103 WAP four-disulfide core domain 2
 AF330262 WAP four-disulfide core domain 2
 AF330261 WAP four-disulfide core domain 2
 AF106684 WD repeat and SOCS box-containing 1
 AB010427 WD repeat domain 1
 NM_018117 WD repeat domain 11
 BC001635 WD repeat domain 5
 BC001264 WD-repeat protein
 AB060283 Wingless-type MMTV integration site family, member 9A
 AJ295991 Wolf-Hirschhorn syndrome candidate 1-like 1
 AB023214 Zinc finger and BTB domain containing 1
 AK022814 Zinc finger and BTB domain containing 10
 AF323460 Zinc finger and BTB domain containing 26
 AK000889 Zinc finger CCCH type domain containing 7
 D28118 Zinc finger protein 161
 NM_003452 Zinc finger protein 189
 AF507946 Zinc finger protein 19 (KOX 12)
 BC002372 Zinc finger protein 207
 AF220492 Zinc finger protein 267
 AK027482 Zinc finger protein 289, ID1 regulated
 BC015212 Zinc finger protein 291
 D89928 Zinc finger protein 354A
 NM_006887 Zinc finger protein 36, C3H type-like 2
 M27877 Zinc finger protein 83 (HPF1)
 AB051498 Zinc finger, CCHC domain containing 6
 NM_017612 Zinc finger, CCHC domain containing 8
 AF247703 Zinc finger, DHHC domain containing 3
 BC034784 Zinc finger, MYND domain containing 11
 NM_006624 Zinc finger, MYND domain containing 11

VITAE

Michael Ryan Weil was born in Falls Church Virginia on February 11, 1977, the son of Linda Haffner Weil and Michael Max Weil. After completing his work at Coronado High School, El Paso, Texas he entered Texas A&M University, College Station in January of 1996. During the fall of 1995 and summers of 1996 and 1997 he attended the University of Texas at El Paso. He received the degree of Bachelor of Science with a major in microbiology from Texas A&M University in May 1999. In July, 1999 he entered the Division of Cellular and Molecular Biology (now the Division of Basic Sciences) at the University of Texas Southwestern Medical Center in Dallas, Texas. After qualifying for Ph.D. candidacy he joined the Department of Molecular Biophysics.

Permanent Address: 1209 Cerrito Perdido
El Paso, TX 79912