UNIVERSITY OF CALIFORNIA, SAN DIEGO

Insights from Reconstructing Cellular Networks in Transcription, Stress, and Cancer

A dissertation submitted in the partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Eugene Yunghung Ke

Committee in charge:

Professor Shankar Subramaniam, Chair
Professor Inder Verma, Co-Chair
Professor Web Cavenee
Professor Alexander Hoffmann
Professor Bing Ren

2012

The Dissertation of Eugene Yunghung Ke is approved, and it is acceptable in quality and form for the publication on microfilm and electronically

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2012

DEDICATION

To my parents, Victor and Tai-Lee Ke

EPIGRAPH

[T]here are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there some things we do not know. But there are also unknown unknowns; there are things we do not know we don't know.

Donald Rumsfeld

TABLE OF CONTENTS

SIGNATURE PAGE iii

DEDICATION iv

EPIGRAPH v

TABLE OF CONTENTS vi

LIST OF FIGURES xiv

LIST OF TABLES xix

ACKNOWLEDGEMENTS xx

VITA xxii

ABSTRACT OF THE DISSERTATION xxiii

INTRODUCTION 1

Bioinformatics and Systems Biology 1

Experimental Methods 2

Microarrays 2

Next-Gen Sequencing. 3

Computational Methods 4

Preprocessing. 4

Normalization. 5

Summarization. 5

Differentially expressed genes. 6

Multiple Testing, ... 7

Functional annotation. 7

Experimental Design 11

Curse of Dimensionality 12

CHAPTER 1 NF-kB BINDING 14

Introduction 14

Representations of Response Elements. 15

NF-kappaB family. 17

Secret Word Problem 19

Historical Perspective. 19

Predictive Power. 19

Conservation. 20

Expression 22

Target genes. 22

Expression. 23

Guilt By Association. 24

Cog in the machine. 25

Discussion..... 26
CHAPTER 2 OXIDATIVE STRESS..... 29
Introduction..... 29
Hydrogen Peroxide Induces Oxidative Stress in Primary Endothelial Cells..... 31
Results..... 33
Hemichannels and Connexins..... 33
Cellular Sources of ROS..... 34
ROS clearance..... 37
MAPK/P38 signaling..... 40
NRF Pathway..... 42
DNA repair pathways..... 43
BCL2 family..... 45
Loss of Mitochondria..... 46
Caspase Cascade..... 47
TP53, ROS clearance, and MTOR..... 49
Heme function, synthesis and degradation..... 51
Reproducibility Issues..... 53
Verification of microarray trends..... 53
Testing of H202 and HMVEC-L..... 56

Calibration of Assays. 57
Loss of Hydrogen Peroxide in Media. 59
Assay conditions. 60
Real Time-PCR data. 61
Discussion 63
Heme. 64
Methods 65
Cell culture. 65
Cell Staining. 65
Spectrophotometry. 66
Hydrogen Peroxide Colormetric kit. 66
LDH assay. 66
Caspase Activity. 66
Total RNA preparation. 66
RT-PCR. 66
Microarray analysis. 66
Data analysis. 67
Pathways. 67
Acknowledgements 67

CHAPTER 3 LENTIVIRAL MEDIATED MOUSE CANCER MODELS 68
Introduction 68
Oncogenes and Tumor Suppressors. 69
RAS. 69
P53. 71
Mouse models of cancers. 72
CRE-LOX. 73
Tumor Progression. 73
Lentiviral mediated models. 74
Biomarkers and Molecular Signatures. 75
KRASLA2 Lung Adenocarcinoma Model 76
Molecular signatures. 76
Effect of NF-kappaB. 79
GSEA. 80
Glioblastoma Multiforme Model 81
Lentiviral Mediated Mouse Model. 81
Clustering. 82
Verhaak Molecular Signatures. 83
Phillips signature. 85

Cell of Origin. 87
Combined signatures. 88
Discussion 90
Methods 96
Transcriptome Analysis. 96
Mouse to human mapping. 97
Z-scaling and rescaled estimates. 97
Clustering. 97
Differential Gene Expression. 97
GSEA. 97
Single sample gene set enrichment analysis. 98
Verhaak signatures. 98
Phillips signatures. 98
Cahoy signatures. 99
Combined signatures. 99
Acknowledgements 99
APPENDIX 100
Real TIme PCR Primers 100
Sweet-Cordero 103

KRAS up signature 103
KRAS down signature 105
Adenocarcinoma signature 107
KRAS signature 108
NF-kappaB Targets 109
Lung Tumor GSEA 111
Terms higher in Lung compared to primary mouse IKK2 KO tumors 111
Terms higher in Lung compared to primary mouse IKK2 WT tumors 124
IKK2 KO/WT GSEA 135
Primary Tumors, all terms enriched in WT 135
Cell Lines, all terms enriched in WT 136
Verhaak 140
Classical_Down 140
Classical_Up 140
Mesenchymal_Down 141
Mesenchymal_Up 141
Neural_Down 141
Neural_Up 142
Proneural_Down 142

Proneural_Up 143
Phillips 144
35 Signature Genes 144
Proneural full signature 144
Proliferative full signature 145
Mesenchymal full signature 146
Cahoy 147
OPC 147
Oligo 147
Neuron 147
Astrocyte 148
Cultured Astrocyte 148
REFERENCES 149

LIST OF FIGURES

Figure Introduction. 1: Distribution of genes with more than one Pubmed citation and the number of associated citations. Genes with one or few citations were ignored. 9

Figure Introduction. 2: Percentage of genes distributed by the number of functional annotations. Majority of genes have unknown functions. 10

Figure 1. 1: Degenerate sequences of transcription factor binding sites can be represented as consensus, or winner take all, sequences. For more complex specificities, an alternate alphabet IUPAC was created. 16

Figure 1. 2: Position specific scoring matrices (PSSMs) can be calculated based on nucleotide frequencies in a set of binding sets. 17

Figure 1. 3: NF-kB PSSMs from TRANSFAC predict multiple binding sites in the NFKBIA, a key target gene of NF-kB. 20

Figure 1. 4: Conservation of binding sites between organisms falsely appears to be a reasonable approach to filter putative binding sites. In the case of ACT, the only high scoring site can be rejected on the basis of conservation. 21

Figure 1. 5: Conservation is not a reliable indicator for the likelihood of NF-kB binding. TNF alpha is a critical gene in both mouse and human, but real binding sites can be lost if only conservation is considered. 21

Figure 2. 1: LDH release (units in raw fluorescence) rises in response to increasing amounts of hydrogen peroxide (μm)..... 31

Figure 2. 2: LDH release (units in raw fluorescence) increases over time in response to 100 μm H202; near maximal release after 6 hours. Values not adjust for LDH half life, which is 9 hours. 32

Figure 2. 3: Live Dead assay with Calcelin AM (Green) measuring live cells and ethidium bromide (RED) measuring dead cells. Majority of HMVEC-L cells are dead after 12 hours induction of 100 μm H202. 32

Figure 2. 4: Mitochondrial sources of hydrogen peroxide are not uniformly up- or down- regulated. 36

Figure 2. 5: Electron transport chain shows an increase in expression, indicating an increase in metabolic ability like due to an increase in energy demands of the cell. 36

Figure 2. 6: NOX family of proteins is not statistically different, but de-regulation of accessory proteins NOXO1 and NOXA1 in response to hydrogen peroxide reduces ability of NOX1 to generate ROS. 37

Figure 2. 7: Major reactive species in the cell are Superoxide, Hydrogen Peroxide, and the hydroxl radical. 38

Figure 2. 8: Antioxidant proteins involved in ROS clearance become up-regulated in response to hydrogen peroxide 39

Figure 2. 9: Western of MAPK signaling cascade shows kinase activity. Cells were induced with H202 and harvested over a time course. p indicates against phosphorylated form. 40

Figure 2. 10: Kinase specific inhibitors reveal caspase 3/7 activity is regulated by ERK, JNK, and p38. HMVEC-L cells were induced with 100 μm. 41

Figure 2. 11: Gene expression profiles for the MAPK signaling pathway show an increase in expression correlating with increased activity. 41

Figure 2. 12: ROS cause NRF2 to be released from KEAP1. NRF2 translocates to the nucleus and increases expression of antioxidant target genes such as HMOX1. 42

Figure 2. 13: Non-homologous end-joining follows pattern of strong late induction. 43

Figure 2. 14: Members of the homologous recombination repair pathway exhibit strong patterns of expression. 44

Figure 2. 15: Nucleotide excision repair pathways show a late pattern of up-regulation for some members. 44

Figure 2. 16: Mismatch repair genes show weak expression profiles. 44

Figure 2. 17: Base excision repair members follow an inconsistent profile. 45

Figure 2. 18: Bad gene expression profiles. 46

Figure 2. 19: Gene expression profiles of the caspase cascade indicate strong up-regulation 48

Figure 2. 20: Combined caspase 3 and 7 activity is greatly increased in response to hydrogen peroxide as measured by a peptide whose cleavage by caspases results in luminescence. 49

Figure 2. 21: Protein abundance against pro-caspase forms shows a lack of abundance as caspases are cleaved 49

Figure 2. 22: Relationship between genotoxic stress, p53, sestrins, and MTOR indicates a dynamic and active network in response to hydrogen peroxide. 51

Figure 2. 23: Critical interacting partners of heme are up-regulated in response to hydrogen peroxide 53

Figure 2. 24: LDH release of HMVEC-L cells in response to increasing concentrations of H202 after 6 hours. Cells from two separate donor lots, 5F1577 and GF3497, were tested. 54

Figure 2. 25: White light images of HMVEC-L cells after 6 hours of (A) 0 µm H202, (B) 100 µm H202, (C) 1 mM H202, and (D) Staurosporine. No evidence of cell death until 1000 µm H202. Cells under complete apoptosis in 6 hours with induction of Staurosporine (D). 55

Figure 2. 26: Microarray data from HMVEC-L cells induced with 100 H202 µm compared to real time quantitative PCR with increasing concentrations of H202. 56

Figure 2. 27: Staining of (A) CD31, (B) VWF, (C) Acetylated DL uptake confirms cells are likely endothelial cells. Dapi is in blue. 57

Figure 2. 28: Testing of H202 stock demonstrated performance to manufacture's specifications. 57

Figure 2. 29: Calibration of Cytox-One LDH release assay. 58

Figure 2. 30: Calibration of caspase 3/7- glo assay. (A) Titration of staurosporine concentration. (B) Caspase activity across various seeding densities using 10 µm Staurosporine. 58

Figure 2. 31: 100 µm hydrogen peroxide is rapidly dissipated in culture media but not PBS. Approximately half of H202 concentration is lost after 10 minutes. Detection performed using a h202 colorimetric detection kit from Enzo Life Sciences. 59

Figure 2. 32: Rapid of addition of H202 increases degree of cell membrane leakage. 700 µm H202 killed approximately half of all cells at 9 hours. As half life of LDH is 9 hours, all cells were likely dead at post 18 hours given the robust response at 12hr and 24hr. 60

Figure 2. 33: Caspase 3/7 activity was observed for concentrations as low as 500 µm after 12 hours. At 24 hours, activity had rapidly decreased which suggested only initial small population of cells had undergone apoptosis. 61

Figure 2. 34: IER2 is a generic stress response gene that only shows a short burst of transcription. (A) Microarray gene expression profile. (B) Real-Time PCR. 300 μm and 500 μm of hydrogen peroxide induce an early burst of transcription as seen in the array data. 700 μm induces a delayed response. ... 62

Figure 2. 35: FOS is an early immediate transcription factor that responds to a wide variety of stressors. (A) Microarray gene expression profile. (B) Real-Time PCR. Higher levels of H202 induce a significantly different expression profile... 62

Figure 2. 36: ATF3. (A) Microarray gene expression profile. (B) Real-Time PCR 63

Figure 2. 37: HMOX1 is an antioxidant gene that degrades free heme. (A) Microarray gene expression profile. (B) Real-Time PCR. HMOX induction was greatly increased compared to microarray data..... 63

Figure 2. 38: Expression ratios on all pathways are rescaled for comparative and visualization purposes. 67

Figure 3. 1: Lentiviral mediated KRASLA2, IKK2 wildtype (WT) and IKK2 knockout (KO) tumors display similar KRAS expression signatures as genetically engineered mouse tumors (Sweet-Cordero).. 77

Figure 3. 2: Dendrogram of tumors reveal two clusters that exhibit opposite patterns of expression. Samples are joined based upon the correlation coefficient, with higher correlation on the bottom. Tumors do not cleanly cluster based on the CRE promoter or the injection site. 83

Figure 3. 3: Tumors score highly for mesenchymal signatures or present double neural/proneural signatures. Tumors do not score similarly based upon their promoter or injection site, but follow closely with clusters in figure 3.2. 84

Figure 3. 4: Normal tissues present a double neural, proneural signature. 84

Figure 3. 5: Comparison of TCGA normal samples, neural tumors, and proneural tumors to mouse neural/proneural tumors.. 85

Figure 3. 6: Samples hierarchically clustered based upon 35 key signature genes from Phillips et al. 2006 86

Figure 3. 7: SSGSEA scores for the full gene sets of molecular subtype signatures from Phillips et al 2006. 87

Figure 3. 8: Cell type specific signatures as calculated from SSGSEA do not correspond to the expected cell of origins. 88

Figure 3. 9: Cell type specific signatures for normal samples display only a strong positive signature for neurons..... 88

Figure 3. 10: Cell type specific signatures follow closely with molecular subtype. Mesenchymal tumors display a strong signature cultured astroglia. Other tumors typically display signatures for the remaining cell types. 89

Figure 3. 11: Signatures for the average of all tumors in a subtype were calculated using TCGA data. While the cell type specific signatures are much weaker than compared to mouse, they still show preferences to particular subtypes. 89

Figure 3. 12: Correlation coefficients for each tumor compared to the signatures of the TCGA subtypes. The highest correlation coefficient for each tumor was on average 0.77, which indicates a high degree of correlation. 90

LIST OF TABLES

Table Introduction. 1: Number of genes distributed by the number of Pubmed citations according to the NCBI gene database. Large percentage of genes have no associated publications while a significant fraction have more than thirty. 8

Table 3. 1: KRAS signatures, as defined by Sweet-Cordero, are enriched in lentiviral mediated tumors as called by GSEA.. 78

Table 3. 2: Glioblastomas exhibit enrichment of lung adenocarcinoma specific KRAS signatures when compared to normal brain. 79

Table 3. 3: NF-kappaB target genes are enriched in normal tissue when compared to IKK2 WT and KO tumors. 79

Table 3. 4: Common functional annotations enriched in both mouse tumors and derived cell lines when comparing IKK2 wild type to IKK2 knockout. Majority of terms are cell cycle related. 80

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Shankar Subramaniam for his support as the chair of my committee and as my primary advisor through these many, many years.

I would also like to acknowledge Professor Bing Ren, a co-advisor, for providing an excellent training regime in molecular biology and the opportunity to learn high throughput experiment techniques during my time in his lab.

I would also like to acknowledge Professor Inder Verma, a co-advisor, and his lab for their intellectual training and support, and to thank him for the opportunity to perform my postdoctoral training in his lab.

I would like to acknowledge the rest of my committee, Alexander Hoffman and Webster Cavenee, for their patience and continued support.

I would like to acknowledge Professor Ratnesh Lal for providing space and resources in his lab for the research involving oxidative stress. I would also like to acknowledge Srini Ramachandran, a postdoc in his lab for his time and effort in our collaboration. Chapter 2 uses with permission Srini Ramachandran's experimental results encompassing figures 2.1, 2.2, 2.9, 2.10 and 2.20.

Chapter 3, in part, uses microarray data and figure 3.1 from *Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model*. Xia, Y; Yeddula, N; Leblanc, M; Ke, E; Zhang, Y; Oldfield, E; Shaw, RJ; Verma, IM. Nat Cell Biol. 2012 Feb 12;14(3):257-6. The dissertation author was a co-author on this paper. Chapter 3 also uses, in part, microarray data and figures 3.10 and 3.12 from

Dedifferentiation of astrocytes and neurons by oncogenes can induce glioblastomas. Friedmann-Morvinski, D; Bushong, E; Ke, E; Soda, Y; Marumoto, T; Singer, O; Ellisman, M; Verma, IM. In review. The dissertation author was a co-author on this paper.

VITA

2001	Bachelor of Science, Texas A&M University
2005	Master of Science, University of California San Diego
2012	Doctor of Philosophy, University of California, San Diego

Publications

Friedmann-Morvinski D, Bushong E, Ke E, Soda Y, Marumoto T, Singer O, Ellisman M, Verma IM. *Dedifferentiation of astrocytes and neurons by oncogenes can induce glioblastomas.* In preparation for submission.

Xia Y, Yeddula N, Leblanc M, Ke E, Zhang Y, Oldfield E, Shaw RJ, Verma IM. *Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model.* Nat Cell Biol. 2012 Feb 12;14(3):257-6

Zhu X, Hart R, Chang MS, Kim JW, Lee SY, Cao YA, Mock D, Ke E, Saunders B, Alexander A, Grossoehme J, Lin KM, Yan Z, Hsueh R, Lee J, Scheuermann RH, Fruman DA, Seaman W, Subramaniam S, Sternweis P, Simon MI, Choi S. *Analysis of the major patterns of B cell gene expression changes in response to short-term stimulation with 33 single ligands.* J Immunol. 2004 Dec 15;173(12):7141-9.

Liang Y, Ke EY, Zhou ZH. IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database. J Struct Biol. 2002 Mar;137(3):292-304.

ABSTRACT OF THE DISSERTATION

Insights from Reconstructing Cellular Networks in Transcription, Stress, and Cancer

by

Eugene Yunghung Ke

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego, 2012

Professor Shankar Subramaniam, Chair

Professor Inder Verma, Co-Chair

The cell is a complex biological network that is capable of transitioning to a wide variety of states. Enumerating, defining, and understanding the mechanisms behind cellular states are important problems of Systems Biology. This document contains insights gleaned from the study of three systems wide problems: transcription regulation by NF-kB, oxidative stress in response to reactive oxidative species, and gene expression changes caused by creation of

lentiviral mediated cancer models. A consideration by literature review is provided of the historical problem formulations for studying mechanisms of NF-kB target gene regulation. Previous formulations of regulation are useful as frameworks for experimental design of future experiments when considered without bias towards prior assumptions. A description of the construction of a network bridging the multitude of cell responses to hydrogen peroxide is provided along with failed attempts to validate that network. Potential regulation by heme in response to oxidative stress reveals an ever tighter relationship between ROS, metabolism, and cell death. Application of molecular signatures defined from human primary cancers is used for determining the suitability of mouse cancer models generated from lentiviral constructs for the study of human primary cancers. Mouse tumors generated artificially display a surprising degree of concordance with primary cancers. The ability of high throughput technologies to query nearly the entire state of the cell can lead to undesirable complexity. Application of simplifying assumptions derived from the consideration of the biological fundamental problem as opposed from technical limitations allows a reduction of in complexity that elucidates areas for future study.

INTRODUCTION

Bioinformatics and Systems Biology

Bioinformatics is a broad interdisciplinary field associated with large biological data sets. High throughput experimental techniques allow the generation of data sets that span thousands of genes across multiple conditions. This size exceeds the ability of a human being to manually organize and analyze, and has posed novel statistical problems complicating analysis. Systems biology arises from the need to understand these large understand data sets and to study relations or interactions that are not readily apparent such as emergent properties[1]. In a sense, systems biology is the study of cellular complexity, or the cell as a system.

This dissertation contains the study of three independent problems: searching for mechanistic insight of NF-kappaB binding, reconstructing cellular networks involved in response to oxidative stress, and relating lentiviral mediated mouse tumors to existing genetically engineered mouse models or primary human tumors. Data was generated on thousands of genes, across multiple time points, conditions, or mutations. The major issues associated with high throughput data are: intrinsic noise, either biological or technical; low number of replicates; multiple testing; undefined and unannotated genes. The bioinformatics aspect of this work is involved in fitting experimental data to known models. The systems biology aspect is an attempt to glean additional insights beyond the

concerns of the original experimental designs, and to understand the interplay between cellular responses and pathways.

Experimental Methods

High throughput technologies are based on extensions of standard experimental protocols. High throughput assays are capable of querying on the order of thousands of genes at a time, but this breadth comes at the expense of cost, specificity, and sensitivity. The fundamental techniques having been utilized in various combinations for all high throughput assays are hybridization, sequencing, molecule based detection, and amplification with low bias. While the dizzying pace of technology may complicate proper experimental design, many limitations of high throughput assays are directly related to the fundamental technologies upon which the assays are based.

Microarrays. Hybridization is the key process for all array based technologies and is an extension of Northern and Southern blotting[2-4]. Synthesized oligonucleotide sequences, typically twenty five to fifty base pairs in length, are placed on defined coordinates, or in an array, very small distances apart on a substrate. Sequences, or probes, for each position in an array consist of one defined, complementary sequence to ideally one target. Samples are typically labeled with a fluorophore and thus the intensity of a probe correlates with abundance of the target sequence. Transcriptome analysis can be performed by microarrays[5] with corresponding coding DNA (cDNA) probes

mechanically spotted[6], or placed using inkjet printer technology (Agilent). De novo oligonucleotide synthesis can also be performed in place using silicon lithography based techniques (Affymetrix), or even technology developed for televisions (Nimblegen). Microarrays have also been successfully applied to other assays such as comparative genomic hybridization[7] (CGH), chromatin immunoprecipitation[8] (ChIP), micro RNA (miRNA), single nucleotide polymorphism (SNP), and protein-DNA binding (PBM) detection[9]. Antibody arrays for detection of protein abundance are conceptually similar, with antibodies used as detectors instead of oligonucleotides sequences. While array technologies are mature and robust, a key disadvantage is the requirement of detectors to be selected a priori.

Next-Gen Sequencing. Sequencing is the method used to determine the order of nucleic acids in a DNA molecule: guanine (G), cytosine (C), adenine (A), and thymine (T). Rapid sequencing was made possible by Sanger sequencing[10]. Sequencing was critical for the generation of the human reference genome by the Human Genome Project[11], and the mouse reference genome[12]. Current "next-gen" high throughput methods rely on pyrosequencing[13], or massively parallel sequencing by synthesis. This methodology provides many more sequences, or reads, at a lower price point, but at the cost of sequenced length or read-through. This is not necessarily disadvantageous as most experiments rely on a resequencing strategy[14]. Reads are mapped to references genomes as opposed to requiring sufficient coverage to fully assemble a new genome sequence. Sequencing has been successfully applied to transcriptome analysis[15]

(RNA-seq), ChIP[16] (Chip-seq), nuclear run on[17] (Gro-seq), and chromosome or genome conformation capture[18] (Hi-C). As a base technology, sequencing excels where the space of expected sequences is unknown or highly complex.

Computational Methods

While high-throughput technology has shifted towards a greater usage of sequencing and away from array based technologies, many problems are universal to high throughput data sets and previous solutions are directly applicable[14]. Beyond simple parsing, typical bioinformatics tasks are data transformation into useful metrics, statistical testing accounting for low replicates and multiple testing, and projection of biological knowledge.

Preprocessing. Data transformation comprises of removal of poorly performing data points, normalization, reduction of noise, and generation of useful metrics. All high throughput experimental techniques are designed to generate data, and often return miscellaneous or nonsense data. For example, gene expression microarrays will return intensity values even if a target transcript is absent due to nonspecific cross hybridization. Therefore an important aspect of analysis is identification of systematic errors inherent to a particular experimental technique, and to eliminate affected data so that it will not skew the entire data set. A common step is setting a low intensity threshold, to automatically reject probes below the threshold and from which values always remain low[19]. It should be stressed that this step is done without consideration to the gene identity or any other information that would introduce bias.

Normalization. Normalization in high throughput experiments is performed to eliminate technical sources of noise from within (intra-) and across (inter-) arrays. Loess, a popular method of intra-array normalization, is required for two color microarrays as dye bias and print deposition artifacts affect the distribution of probe intensity[20]. Inter-array normalization centers the distribution of separate samples so that they become comparable. Normalization is required to adjust changes in distribution due to variability arising from technical sources of noise such as the amount of starting material, labeling efficiency, hybridization efficiency, and scanning efficiency. Two common normalization methods are rescaling to the global median and quantile-quantile normalization[21, 22].

Summarization. Arrays and even sequencing often provide multiple expression levels for one gene. Affymetrix arrays have multiple probes, typically twenty to forty, spread throughout exons of a gene, the collection of which is termed a transcript. The task of generating one unified value for a transcript is known as probe set summarization. The most popular method, RMA, is based on linear regression[23]. While having multiple transcripts per gene is an indication of splicing, such information may not be truly useful as there is not enough useful exon information to reconstruct specific splice variants; often information on a transcript level is contradictory. For simplicity, a common procedure is to average the values for all splice variants or transcripts of a gene to reach a consensus value.

Early microarray work concerned itself with variance transforming metrics due to the fact that low intensities have higher intrinsic noise than higher

intensities[24]. Ratios or fold changes are thus more susceptible to noise if the denominator is of low intensity, creating a highly noisy spread of ratios. However, these transformations have fallen out of favor with the application of the *logarithmic base 2* function to reduce noise[22]. An important benefit is the calculation of base 2 fold changes by simple subtraction, which is favored for visual interpretation.

Differentially Expressed Genes. The challenges of identifying statistically significant differentially expressed genes are caused by the low number of replicates and multiple testing issues. P-values are random variables and are a product of the number of replicates and the true significance[25]. By the central limit theorem, p-values will trend towards the true significance given sufficient replicates. Simulations suggest sixteen to thirty two replicates are recommended for robust determination of a distribution. Due to cost constraints, however, samples are typically replicated only two to three times. This poses a severe mathematical issue as the standard deviation cannot be reliably calculated from so few replicates and many statistical tests perform poorly as a result.

Significance calling by arbitrary fold cutoff is not ideal[26]. Instead, many approaches are based on variants of the student's t-test, which is a common method to assess the statistical difference between two populations. Alternative computational methods based on the t-test have been developed that circumvent the low replicates per gene by inferring information from other probes[19, 27]. The most common are Significance Analysis of Microarrays[28] (SAM), Limma[29], and

Cyber-T[30]**.** The main difference between these methods is estimation of the underlying distribution via permutation versus calculation the standard deviation using information from neighboring probes. When the stronger assumptions used for Bayesian models are correct, Cyber-T and related approaches such as Vampire[19] identify fewer false positives at lower fold changes.

Multiple Testing, A fairly unique problem to high throughput biological experiments is the sheer number of statistical tests. Counter intuitively, each gene or probe represents a separate instance and statistical test as opposed to each array. Statistical significance is based on the p-value, which represents the probability of an observation resulting from chance given a null hypothesis. For one hundred tests with a p-value cutoff of 0.05, the rough expectation is that five positive tests incorrectly reject the null hypothesis. As array technology is capable of surveying tens of thousands of genes, correction for multiple testing is required. The Bonferroni correction is an over adjustment by dividing the p-vale cutoff by the number of tests to be performed. A more practical alternative is the false discovery rate[31] (FDR), which is an estimate of the number of false positives given the total number of positive calls. FDR is typically calculated by permutation[28], although model based approaches are in use[32].

Functional Annotation. There are several strategies to extract biological meaning from high throughput data sets. The most straightforward procedure is to focus on selected sets of known, biologically important genes for a given condition or treatment. A major hindrance to this strategy is knowledge regarding

genes is uneven; roughly one- of all genes have at most one published citation associated with them as shown in Table Introduction.1.

Table Introduction. 1: Number of genes distributed by the number of Pubmed citations according to the NCBI gene database. Large percentages of genes have no associated publications while a significant fraction has more than thirty.

Number of citation(s)	Mouse	Human
0	8,050	12,190
1	1,281	3,509
2-5	2,020	4,387
6-10	3,978	3,903
11-15	4,828	2,568
16-20	3,750	1,734
21-25	1,844	1,161
26-30	9,91	8,35
30+	4,121	4,716
Total Genes	**30,863**	**35,003**

In many cases, few citations indicate sequencing papers which simply list or describe the discovery of previously unknown transcripts without indicating function. Disregarding genes with fewer than two citations, the majority of genes have between two and fifteen papers published studying them as shown in Figure Introduction.1. In contrast, roughly 20% of the remaining genes have greater than thirty papers studying them. Only a subset of genes is well studied, with tenuous information or connections for other genes. By focusing on known

genes, there is less likelihood for studying or understanding novel functions and connections, but such an approach does provide confidence in the validity of the data.

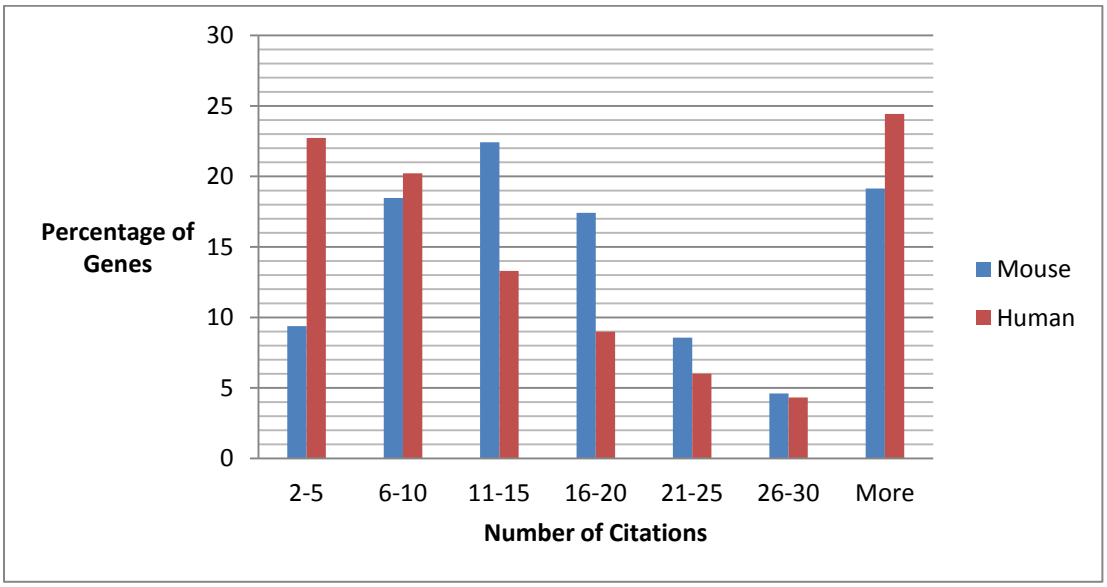


Figure Introduction. 1: Distribution of genes with more than one Pubmed citation and the number of associated citations. Genes with one or few citations were ignored.

In theory, automated procedures are attractive as they are less likely to introduce bias. The most basic is cluster analysis, to group genes by their expression[33]. The assumption is that genes will co-express and co-cluster due to commonality in regulation or function[34]. Significant functional terms can be identified through the application of the hyper geometric function[35]. Terms can be derived from assignment by Gene Ontology[36], or from curated gene sets derived from primary literature. Unfortunately, that vast majority of genes have no associated annotation function as shown in figure Introduction.2. Assignment of function inferred microarray experiments or homology is often problematic as there is a strong tendency to propagate any errors. This can be understood

easily when considering a gene cluster of poorly annotated genes. The few annotations that are known may be incorrectly assigned to the cluster as a whole; however, there may be an unknown true function which in actuality is causing the genes to cluster.

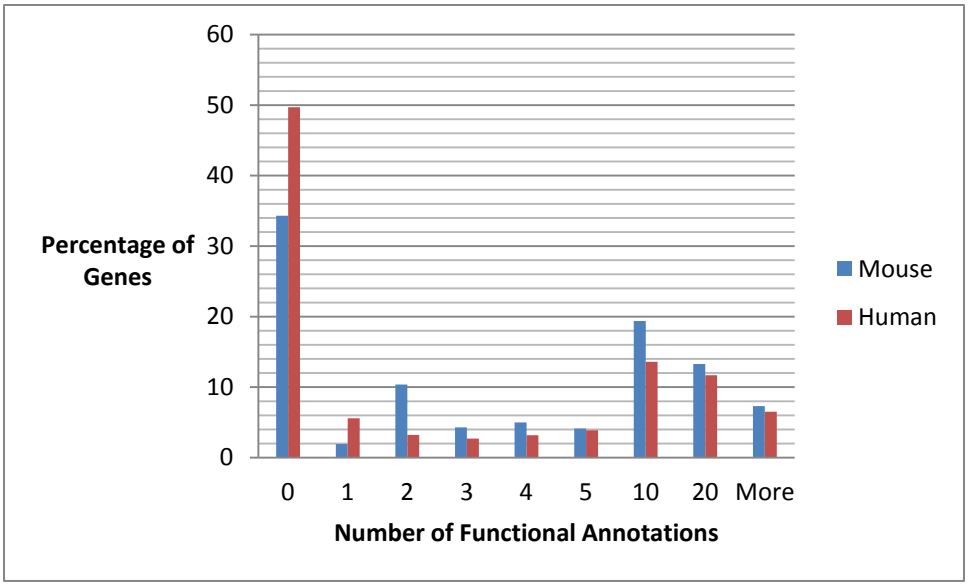


Figure Introduction. 2: Percentage of genes distributed by the number of functional annotations. Majority of genes have unknown functions.

As with differential gene expression, significance testing of multiple terms requires multiple testing correction. Gene Set Enrichment Analysis[37] (GSEA) is a commonly used alternative to the hyper geometric that controls for multiple testing. GSEA is based on the non-parametric Kolmogorov-Smirnov statistic, and unlike the hyper geometric, it compares the difference in ranked expression of a gene list between two conditions. The background distribution is calculated by permutating the phenotype or gene labels, which is used to estimate the FDR. A more thorough discussion of GSEA are covered by others[38, 39].

Beyond functional annotation, advanced computational methods attempt to reconstruct the transcription network controlling gene expression. Typically, these methods rely on additional network information such as transcription factor binding[40], protein-protein interactions[41], and sequence information. In terms of computational methods, most rely on the hyper geometric, with more complicated variants relying on regression trees[42]. However, this is an area of continuing research, as very few models have been built that adequately describe the biological complexity observed.

Experimental Design

The essential, critical step in high throughput experiments is the experimental design. The greatest barrier to successful design is a lack of experience, as many techniques are just reaching price points that are feasible for smaller scale iterative experiments. Many experimentalists rely too heavily on the global nature of high throughput experiments and fail to optimize their experimental conditions. Another common mistake is the exclusion of a universal baseline, such as a wild type or untreated condition in the mistaken assumption that no interesting information will be gained. High throughput assays are most reproducible across fold changes[43], and having a proper baseline will help ensure data is comparable to related, outside data sources.

Cost concerns may also affect experimental design. A beneficial approach is to first consider the design without regard to cost. This shifts the focus onto selection of conditions that will best fulfill experimental goals. An important

aspect of design is to consider the outcome if the overarching hypothesis is correct. Potentially, consideration should be made to modifications to the experimental design with regards to the kind and type of data that could be necessary for further analysis. Conversely, if the hypothesis is incorrect, modifications should be made to the experimental design incorporate aspects that could still allow the generation of useful data. When minimizing cost, the number of conditions as opposed to the number of replicates is more important; biological trends trump statistical significance at such low number of replicates. Replicates should be designed against greatest source of biological noise. As high throughput assays are notoriously noisy, a tendency exists to artificially reduce noise by replicating across more technically stable conditions. However, it is better to reject during analysis trends and hypothesis formed from noisy data than to experimentally verify many false ones.

Curse of Dimensionality

For multi-dimensional problems, the space of actual solutions is much smaller than the possible space. This is the basis on which principle component analysis and other such methods perform. However, when considering a multidimensional problem, it is not necessarily clear in which directions experimental observations should be collected. The problem of dealing with an extremely large variable space is known as the curse of dimensionality. For biological problems, this relates to finding the correct minimal set of observations

required to understand the maximal number of responses. A pertinent example is the exclusion of repeat regions from arrays ad sequencing as the length of oligonucleotide detectors or sequencing read through is insufficient to specifically identify an exact repetitive sequence. This excludes information that is potentially troublesome, but if the major effect involves repetitive sequence then no observations will be recorded.

The point being the use of assumptions forces the collapse of dimensionality, as certain portions of the solution space become unobservable. This is often done by design, as there are aspects of a system's behavior that are not of interest. At times assumptions are taken for practical reasons, in that the limitations of the assays involved impose deficiencies. However, if those assumptions are improper, the space of correct solutions may be unreachable. A key difficulty is when no correct solution is obtained, it is impossible to determine which assumptions may be incorrect without independently testing each assumption. The remaining solution space may still be intractable even with a correct set of assumptions.

CHAPTER 1 NF-kB BINDING

Introduction

The genome is the collection of inheritable traits passed from parent to offspring[2]. In mammals and higher organisms, the genome is encoded as a DNA double helix, and condensed into superstructures or chromosomes to fit within a cell nucleus. The Central Dogma, as formulated by Crick, highlights the major cellular actors and describes the flow of residue specific sequential information between them[44]. A broader view of this framework is that the information encoded in DNA is transcribed to messenger RNA (mRNA); mRNA is translated by the ribosome to a polypeptide chain of amino acids, or protein; proteins act as the functional units of the cell[2].

The genome is essentially the same for all cells within an multi-cellular organism; important exceptions being certain immunity and reproductive cells[2]. Therefore, mechanisms are required to control the activity of proteins as a specific cell type has a defined task with a requirement to react to specific extracellular responses. Besides controlling the activity of proteins through post-translational modifications, the activity of a protein can be controlled by its abundance. Sequences for proteins may not be transcribed and translated until specific regulatory conditions are met, or the rate of transcription may increase or decrease.

RNA polymerase II (POL II) is the enzyme that transcribes DNA into messenger RNA. Proteins that interact with POL II and affect the rate of transcription are known as transcription factors. The class of transcription factors that modulate transcription through direct binding of sequence based response elements are known as sequence specific transcription factors. A minority of sequence specific transcription factors have been well studied: the upstream signaling pathways and subsequent events required for transcription factor activation; typical interacting partners or additional events required for transcriptional regulation; collections of model or target genes; the three dimensional crystal structure with a bound response element; and predictive representations of the response element calculated from hundreds of sequences. Given the totality of knowledge regarding this select group, it is surprising that it remains difficult to accurately predict the expression patterns of target genes or to explain why a seemingly valid response elements is not active.

Representations of Response Elements. Protein-DNA interaction is best observed in its true three dimensional (3D) state. However, such crystal structures between transcription factor and response elements are rare, and do not adequately cover the possible interaction space. The *local* 3D structure of DNA is unknown and is often abstracted to a string a of simple sequences: A, C, G, T. For this reason response elements are typically simplified to sets of degenerate sequences, or collections of sequences that are specific but not absolutely so[45].

As response elements are degenerate, the collection of sequences can be represented as a motif or consensus binding site which is an average of nucleotide frequency[46] (figure 1.1). Nucleotides with the highest frequency, or consensus, at each position represent that position in a winner take all fashion. As some response elements are inadequately represented in such a fashion, an alternate alphabet was described, IUPAC. As additional sequences are collected, the IUPAC representation is still insufficient to represent response elements. A frequency matrix, position specific scoring matrix (PSSMs), based on the appearance of each nucleotide for each position over all sequences can be calculated. Shannon's information of the frequencies can be used for visual purposes, and graphically represent PSSMs sequence logos (figure 1.2). A major assumption is that each position is independent; independence is known to be incorrect, yet position independent effects often outweigh the positional dependent effects[47], such that PSSMs are adequate representations. In vitro protein-DNA binding data has been shown to be recapitulated by relatively simple models of binding, using positional independent models[48].

```
TACGAT
TATAAT
TATAAT
GATACT
TATGAT
TATGTT

TATAAT  consensus sequence
TATRNT  alternate consensus sequence
```

Figure 1. 1: Degenerate sequences of transcription factor binding sites can be represented as consensus, or winner take all, sequences. For more complex specificities, an alternate alphabet IUPAC was created.

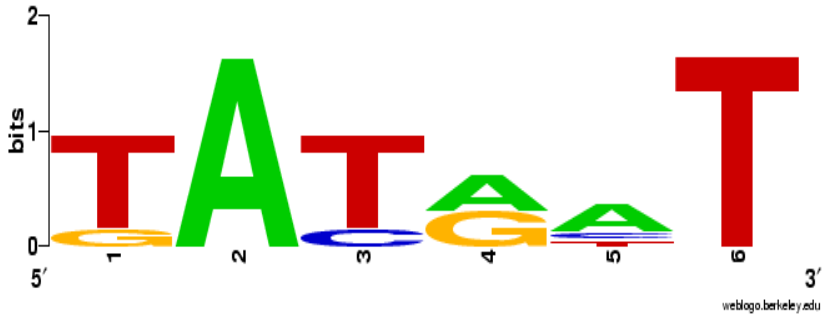


Figure 1. 2: Position specific scoring matrices (PSSMs) can be calculated based on nucleotide frequencies in a set of binding sets. PSSMs can be visualized using Shannon Information as sequence logos. Larger letters contribute more information and indicate a nucleotide-position specific requirement.

NF-kappaB Family. Nuclear factor kappa-light-chain-enhancer of activated B cells, abbreviated as NF-kappaB or NF-kB, is a family of transcription factors that was originally observed to bind to a enhancer sequence found in the light chain kappa immunoglobulin gene[49]. At first thought to be B-cell specific, latent activity of NF-kB was found to be ubiquitous, and highly inducible by a wide variety of ligands[50]. Signaling pathways[51], and the cross-talk between them[52], involved the proper activation of NF-kB for the recruitment of Pol II have been elucidated[53].

NF-kappaB family members RELA (p65), C-REL, and RELB have transcriptional activation domains that recruits POL II[54] . NFKB1 and NFKB2 lack transcriptional activation domains and require dimerization with other family remembers or cofactors such as BCL3 to activate transcription. NF-kappaB family members are known to form multiple hetero- and homo-dimers, but the RelA/p50 heterodimers is the most ubiquitous and is synonymous with name NF-kappaB or NF-kB. The crystal structure of DNA bound NF-kappaB has been

solved[55]. In vitro NF-kappaB dimer specific binding has been revealed to follow three general classes: homodimers of RelA or C-REL; heteodimers; and homodimers of p50 or p52[56],

Key NF-kB targets are of considerable interest and have provided broad biological insights. TNF-alpha is a critical component of inflammation and is associated with many disease states. NFKBIA is a direct negative regulator and target gene of NF-kB, and acts as a biological negative feedback system[57]. The interferon beta gene has been used in vitro transcription to study the effects of enhancer and transcription factor binding[58]. A multitude of other target genes have been identified and reveal NF-kB to be an important regulator in many biological processes[50, 59].

Yet given the totality of this information, it is not possible to accurately predict *a priori* above random chance if a gene will be regulated by NF-kB. NF-kB is so well studied that the term "NF-kappa *BETA"* yields enough Pubmed citations to be considered a well studied gene. NF-kB binding appears too widespread; finding enrichment for the consensus site is simpler than understanding the mechanism by which seeming valid response elements are not regulated. Historical formulations for understanding transcription factor binding to their response elements arose due to limitations in technology and the state of knowledge. As technology and the state of the field have advanced, subsequent problem formulations have become more complex. Revisiting earlier formulations may still provide useful insights when considering an experimental design, especially in light of new technology.

Secret Word Problem

Historical Perspective. Before the sequencing of the reference genomes, potential regulatory sequence of many genes was unknown. A successful strategy employed for the initial discovery of some transcription factors was usage of the electron mobility shift assay[49]. Regulatory sequences could be screened and shown to be significant with the use of unlabeled cold oligo. Screening of the regulatory sequences for putative target genes also revealed the presence of consensus sites. This introduced the misconception that the sequences transcription factors bind to were relatively rare, akin to a secret word.

Predictive Power. With the publication of reference genomes such as human and mouse, the predictive power of motifs, consensus sequences, and PSSMs is easily testable. As the NF-kappaB family are sequence specific transcription factors, prediction of target gene expression should be strongly correlated with presence of a consensus site. A simple search using PSSMs across upstream, proximal regulatory sequences for genes (promoters) reveals that nearly every promoter has at least one reasonable facsimile for a NF-kB binding site. Even applying relatively strict scoring thresholds, it is difficult to discern a putative target from a false positive. That is not to say this approach cannot generate interesting results (figure 1.3), only that is generates too many targets to functionally test. The difficulty becomes not in finding target genes but in rejecting or ranking them.

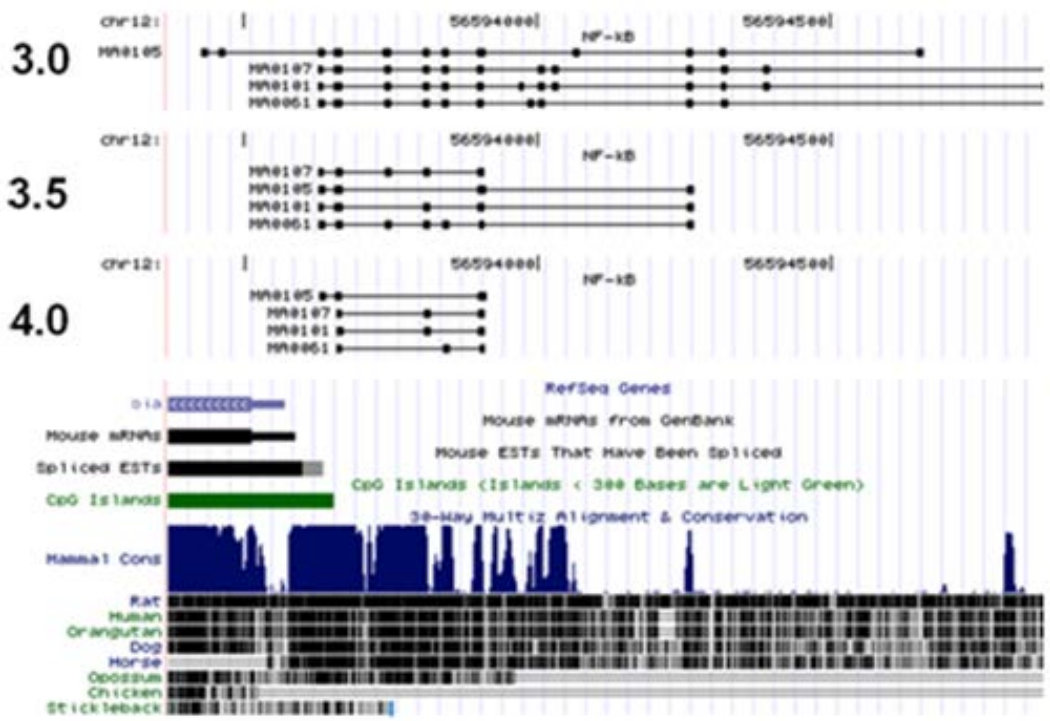


Figure 1. 3: NF-kB PSSMs from TRANSFAC predict multiple binding sites in the NFKBIA, a key target gene of NF-kB. Increasing of thresholds based on the standard deviation progressively filter hits. Hits 3.0 standard deviations above average the score represent the top 0.15%. Setting thresholds too highly quickly causes the rejection of true binding sites.

Conservation. As the number of reference genomes increased, an additional constraint of conservation was imposed. The argument for this approach is if a sequence is a true regulatory element, it is more likely conserved. It is known this assumption is grossly incorrect as species specific differences are purely genetic[60, 61]. Conservation as a filter is meant for convenience as opposed to being motivated by a biological underpinning (see figures 1.4 and 1.5). For example, microarray expression is filtered by a significance cutoff to reduce the number of false positive, not because the number of genes is cumbersome. Conservation would not be a bad assumption if

a putative gene is known to exhibit a similar expression pattern across multiple organisms, then filtering would be appropriate.

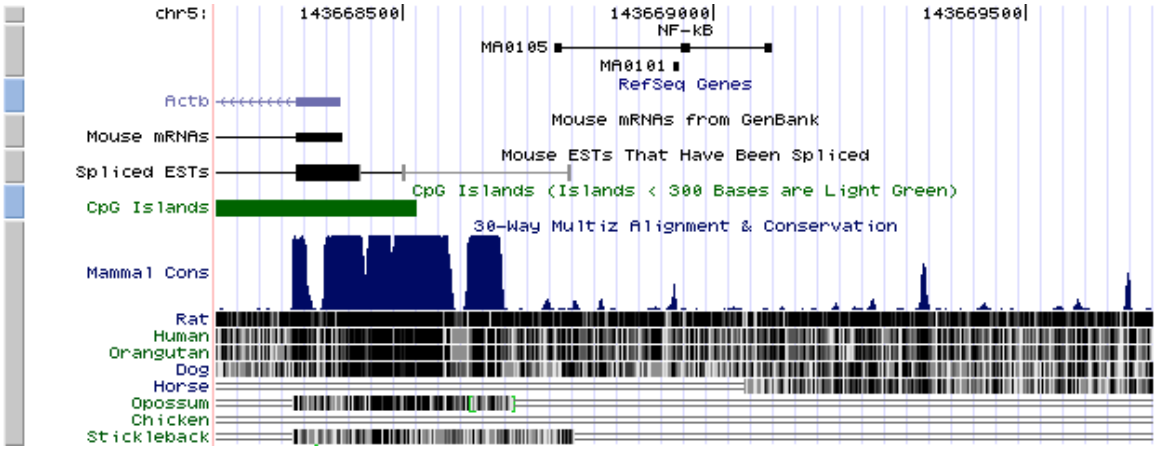


Figure 1. 4: Conservation of binding sites between organisms falsely appears to be a reasonable approach to filter putative binding sites. In the case of ACT, the only high scoring site can be rejected on the basis of conservation.

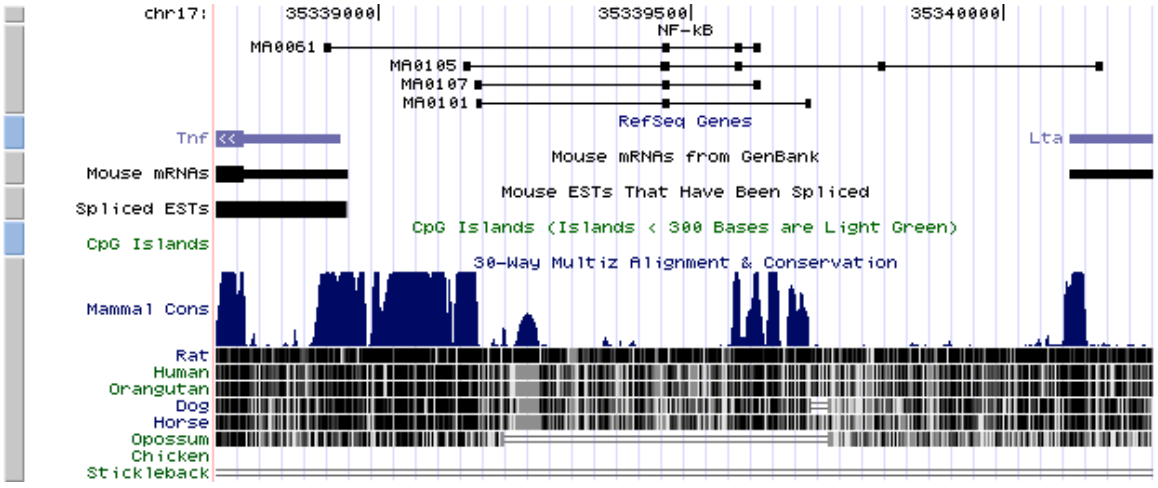


Figure 1. 5: Conservation is not a reliable indicator for the likelihood of NF-kB binding. TNF alpha is a critical gene in both mouse and human, but real binding sites can be lost if only conservation is considered.

Given the disconnect between the expected number of target genes and the abundance of putative targets from genomic scans, aspects of PSSMs themselves were called into question. PSSMs were typically generated from a small number of sequences from in vitro assays[62], and assumed positional independence. The true promoter length was unknown, and the length of average PSSMs were too short compared to the often arbitrarily chosen promoter sizes, creating a "twilight zone" in which subtle motifs are difficult to detect[63]. A legitimate complaint against genome wide searches is the lack of context. The cell controls gene expression not only by sequence specific transcription factors, and searching only sequence (admittedly not by choice) expands the search space too large. However, the possibility that NF-kB binds many promoters in a wide spread fashion seemed unlikely and was largely rejected.

Expression

Target Genes. As Pahl 1999 commented to be considered a "bonafide" target a gene must first be proven to have a bound regulatory element in a cell based system and said element must be mutated and proven to change expression[50]. Before sequence was readily available it seemed reasonable to define target genes solely on that presence of a consensus site. It is now known that expression and enrichment together are weak predictors of target gene regulation[64], and lists of NF-kB targets likely contain false positives.

Many model genes were identified as target genes due to their biological significance, and for being potently induced by NF-kB. These properties do not necessarily translate as representative of other NF-kB targets. This is not to downplay their significance, but their uniqueness may indicate their mechanisms of regulation may not be completely general.

Expression. Large scale microarray studies of inducers of NF-kB have indicated expression of many thousands of genes. Expression profiles in response to pathogens[65], and purified components of pathogens[66, 67], have indicated a common response with cell type specific and ligand specific components. A key assumption of gene expression analysis is that genes with similar function or regulation will co-express[34]. While true, this does not necessarily mean that clustered genes are co-regulated. For example, if two transcription factors with different direct targets behave in a similar fashion, their target genes will co-express. In addition, many transcription factor directly target other transcription factors[68], and separating the difference between target genes and "dependent" genes is difficult.

Often motif and term enrichment is performed on clusters to derive biological significance. As motifs are over represented, it is not an easy task to determine the true biological significance of enrichment or interpret additional binding events. Term enrichment, such as using a list of NF-kB gene may be flawed as not all genes are really targets. High throughput assays have the ability to query the expression status of thousand of genes, yet in some respects

studying gene expression alone comes little closer to understanding the underlying rules of NF-kB regulation than sequence searching alone.

Guilt by Association.

Chromatin immunoprecipitation (ChIP) allows the direct interrogation of DNA bound to proteins. The assumption is that binding of a transcription factor is a better predictor of regulation than sequence or expression. Saccani et al. 2003 showed by ChIP experiments differing heterodimers exchanging occupancy on target genes[69]. This exchange was suggested as a mechanism that explained differential expression, as heterodimers vary in their interactions with other proteins that would affect transcription. Dimer binding was shown to be rare and restricted which increased the likelihood dimer exchange causing the differences in expression.

High throughput versions of ChIP are capable of studying the binding of a transcription factor on a genome-wide basis[8]. Martone et al described the occupancy of NF-kB across chromosome 22; NF-kB binds to many non-canonical sequences and to sequences beyond the assumed promoter region[70]. Lim et al. 2007 observed similarly wide spread binding of NF-kB on a genome-wide level[71]. Schreiber et al. 2006 observed different NF-kB family member occupancy, noting that highly expressed genes appear to be bound by multiple family members[72]. Kasowski et al. 2010 demonstrated that for only a small subset of genes, loss of activity could be traced to mutated NF-kB binding[73]; although Leung et al. 2004 had previously demonstrated a more interesting case

when mutation in NF-kB binding sequence lead to differential, as opposed to lost, expression[74]. Antonaki et al. 2011 described a high degree of non-functional binding of NF-kB to Alu-repeats[75].

It is now apparent that NF-kB binds to many sequences outside expected regulatory regions, and many bound sequences do not match canonical motifs. NF-kB, while less than appearance of a consensus site, is an insufficient indicator of regulation.

Cog in the Machine.

Transcription regulation is now studied as a complex mosaic of histone modifications, additional co-factors, multiple transcription factors, and cell-type specific enhancers[76]. ChiP-chip and Chip-seq have made rapid and major advances in the study of histone modifications and general mechanisms the genome is organized; high throughput ChIP assays have been an unabashed success in this arena. However, NF-kB is somewhat of an abstraction for any sequence specific transcription factor. The goal is not to better understand NF-kB mechanism, but broad based mechanisms that play a role in the expression of all genes. As histone modifications vary greatly across cell types, the best examples related to NF-kB are cell type specificity of inducible of enchancers[77-79].

Discussion

Technology has constantly motivated the reformulation of an essential problem for understanding of NF-kB: how does NF-kB regulate target genes in such a way to generate specific gene expression patterns? Prior formulations were not completely incorrect due to faulty assumptions as once thought and a large amount of data suggest that true NF-kB binding is in fact wide spread and in many instances may play no *cis* regulatory function. The problem can now be rephrased as what are the mechanistic differences between response elements that lead a NF-kB bound sequence to become a regulatory element? At each turn, more and more of the state the cell is queryable; the problem has become successively more complex. Consensus binding sites are not restricted sequences; the accessible sequence of the genome while tightly regulated, is still quite large; a large degree of NF-kB binding appears to play no direct regulatory role. Previously, deficiencies in problem formulations were assumed to lead to a lack of clear cut observations, but widespread binding of NF-kB appears to be real. As such, a useful exercise when designing an experiment would be to reconsider prior formulations for insight or guidance.

The secret word problem essentially concerned with the proper representation of a complex 3D interaction projected onto a 2D sequence. While PSSMs and their like are drastic simplifications, additional sequence has not lead to drastically different PSSMs[56]. Thus increasing the number and resolution of bound sequences may not provide additional information. Rather, PSSMs suggest that complex protein-DNA interactions can be with successful

represented using simple models. Perhaps the local neighborhood of bound regulatory NF-kB resulting from increases in resolution to chromosomal conformation assays[18, 80] could be similarly simplified.

Proper experimental design by maximizing expression differences can lead to novel insights using expression arrays. Amit et al. 2009 indentified 24 co-regulators and 76 "fine-tuners" that explain specificity of pathogen sensing pathways by a combined approach of gene expression analysis, sequence search, and siRNA pertrubations[81]. Large previously published expression data sets could be used as guidelines to select ligands or inducers that appear to maximally differentiate NF-kB regulation.

NF-kB binding is too broad, but still much less than general transcription factors. Histone modifications have been powerful markers and tools for understanding general transcription because they efficiently divide the expression space to subclasses and types. In a similar fashion, putative cofactors of NF-kb that divide target genes could be screen. Lim et al. 2007 showed E2F1 as an important coactivator in LPS stimulation. Barish et al. 2010 demonstrated repression by Bcl-6 as a mechanism for controlling innate immunity.

Application of new technology has a tendency to increases the dimensionality and thus complexity of the observation space. For NF-kB regulation of target genes, this has still not lead to a reasonable solution. However, not all solutions demand more observations. Berman et al. 2002 demonstrated a simple filtering procedure using motif enrichment overlap

between all regulatory factors in a specific stage in fly development was sufficient to predict novel regulator elements and target genes[82]. Segal et al. 2008 using a more complex computation approach was able to correctly predict a large degree of gene expression in fly development[83]. Both these approaches were based solely on PSSM enrichment to predict expression. Although it is not entirely clear the reasons for their success. Was that fly development is well understood, and the major transcription factors and their regulatory combinations are known? Or is it because development is a potentially more controlled cellular and physiology process than say innate immunity, which is required to react rapidly to a vast assortment of potential insults. Or is the just nature of NF-kB binding that is unusually more complex?

CHAPTER 2 OXIDATIVE STRESS

Introduction

Oxidative stress occurs when reactive oxygen species (ROS) accumulate within the cell beyond the ability of the anti-oxidant defense systems to clear[84]. Oxidative stress can impair cell function, signaling, and stability[85] as cells maintain a predominately reducing state by damaging proteins, lipids, and DNA. Oxidative stress is negatively associated with such disease states as Alzheimer's disease, atherosclerosis, cardiovascular disease, diabetes, lung disease, and Parkinson's disease[85, 86] while conversely ROS are the main therapeutic agents of ionizing radiation and chemotherapy for the treatment of cancer[87].

Potential outcomes of oxidative stress are clearance, “immunity”, and cell death via apoptosis or necrosis. The most destructive ROS, free hydroxyl radicals, can be generated from less reactive ROS. As such the cell maintains an energy intensive network of enzymes each of which eliminates a specific ROS[88]. However, ROS are endogenously generated, and significantly low basal levels of ROS are important mediators for redox cellular signaling pathways[89, 90]. Gap junctions and hemichannels rapidly uptake extracellular ROS in a cell type specific manner[91]. This highlights the importance of intracellular mechanisms dealing with oxidative stress caused by exogenous sources of ROS.

A major detrimental effect of oxidative stress is genotoxic stress as ROS directly damages DNA[92]. In response, DNA repair pathways are activated and if

damage is not sufficiently reversed, this can lead to cell death. Cell death can be divided into two distinct processes, apoptosis and necrosis[85]. Apoptosis is a controlled method of cell death that is phenotypically characterized by shrinkage. The relevant apoptosis pathway for oxidative stress is the p53 dependent intrinsic pathway. Necrosis on the other hand appears to be a passive process due to acute injury that results in a phenotype of swelling. Necrosis is typically considered to be an uncontrolled cell death as it results in damage to nearby cells, unlike apoptosis. “Immunity” to oxidative stress has been speculated to be a major source of resistance to treatment of cancer[93]. Mechanisms of acquired immunity proposed so far include genomic instability leading to gene loss, mutational events leading to up-regulation of ROS clearance, DNA damage repair, or loss of p53 activity.

To study the biological processes and signaling events that lead from oxidative stress to cell death, a systems biology was undertaken approach to generate a global view of the cell to observe the changes to networks that deal with basal levels of ROS in addition to networks activated by oxidative stress. Oxidative stress led to widespread changes in cell signaling, DNA damage repair, pro- and anti-apoptotic pathways, and metabolic and energy related pathways, especially those dealing with the mitochondria.

Hydrogen Peroxide Induces Oxidative Stress in Primary Endothelial Cells

Hydrogen peroxide is a commonly used inducer of oxidative stress. Cultured Primary Human Lung Micro Vascular Endothelial Cells (HMVEC-L) grown in EGM2 media were treated with hydrogen peroxide. After exposing cells to increasing concentrations of H202 for 6 hours Lactate Dehydrogenase (LDH) release was measured to determine cell membrane leakage as a surrogate for cell viability and found the median lethal dose to be 100 µm (Figure 2.1). LDH release was assayed over time and median lethal concentration and time was found to be at 2 hours with 100 µm of hydrogen peroxide. Cells were assayed for viability with Calcein AM and ethidium bromide for live and dead cells respectively. Little ethidium bromide staining was observed at 6 hours, while strong staining at 12 hours and complete staining at 24 hours occurred.

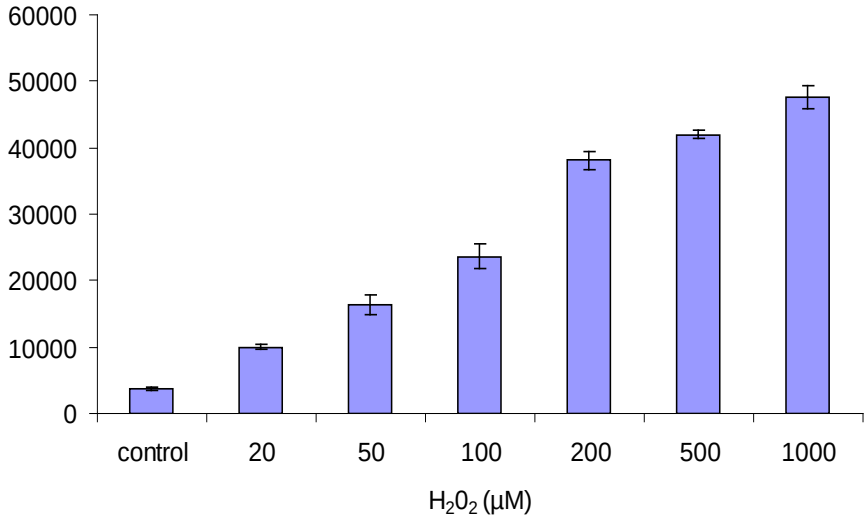


Figure 2. 1: LDH release (units in raw fluorescence) rises in response to increasing amounts of hydrogen peroxide (µm). LDH is sequestered in the cytosol in healthy cells. Median lethal dose occurs around 100 µm.

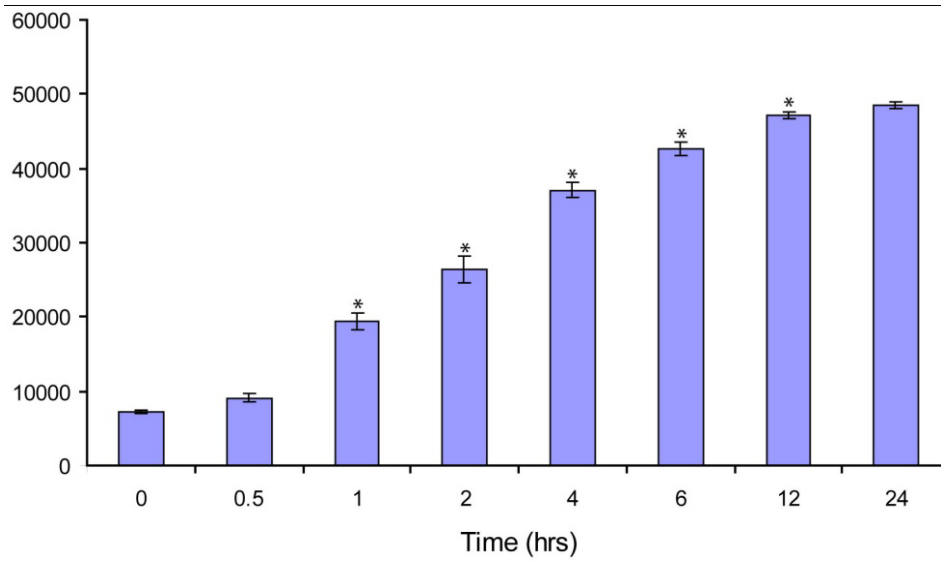


Figure 2. 2: LDH release (units in raw fluorescence) increases over time in response to 100 µm H202; near maximal release after 6 hours. Values not adjusted for LDH half life, which is 9 hours.

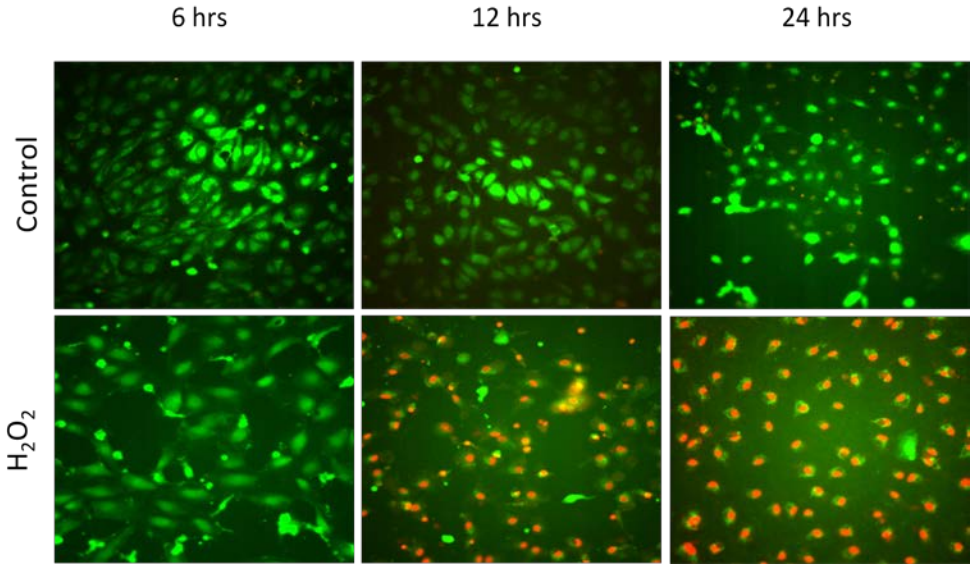


Figure 2. 3: Live Dead assay with Calcelin AM (Green) measuring live cells and ethidium bromide (RED) measuring dead cells. Majority of HMVEC-L cells are dead after 12 hours induction of 100 µm H202.

Cells were then treated and harvested at varying time points with 100 µm hydrogen peroxide for subsequent gene expression, mass spectrometry, and

phospho-protein high throughput experiments. High throughput expression data was collected for untreated, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours and 12 hours. Antibody array data was collected for untreated, 2 hours, 4 hours, 12 hours, and 24 hours. Data was analyzed using both unsupervised and supervised cluster and classification methods to identify pathways activated in response to oxidative stress. Supervised classification was derived from known ROS and oxidative stress related pathways using online databases and expanded with primary literature.

Results

Hemichannels and Connexins. Gap junctions have been implicated to have a role in the intercellular transmission of apoptotic signaling, characterized as the bystander effect[94]. Gap junctions are formed by the connexin family of proteins, which form hemichannels two of which comprise a gap junction. Hemichannels themselves have been shown independently of composition in gap junctions to mediate early apoptotic signaling, as they allow the direct extracellular entry of some apoptotic stimuli such as ROS. In general connexin expression (GJA4, GJA2, GJB1, GJB2, GJB3, GJD3, and GJD4) was reliably down-regulated consistent with the loss of gap junctions as apoptotic bodies form. The exception was GJA1 which showed a reliable up-regulation. Pannexin genes were observed to follow a similar trend, as PANX2 and PANX3 were sharply down-regulated while PANX1 was highly up-regulated (data not shown).

Cellular Sources of ROS. While excessive amounts of ROS are harmful, the cell tolerates a small basal level of hydrogen peroxide for cell signaling[90]. In response to extracellular hydrogen peroxide, changes in known endogenous sources of ROS were examined. The predominate sources of sub cellular ROS are generation of hydroxyl radicals by the Fenton reaction, generation of hydrogen peroxide by catalytic enzymes[90], and generation of superoxide ions by the electron transport chain[95].

In the presence of hydrogen peroxide, iron serves as a potent catalyst in the generation of hydroxyl radicals described by the Fenton reaction[96]. As such, iron is tightly sequestered as bound components of proteins, or sequestered by the ferritin family[97]. In response to hydrogen peroxide, only a slight increase in the transcriptional rate of FTH1 (ferritin heavy chain 1) was observed with no other discernible changes in other ferritin family members (data not shown). Transferritin is involved in the export of cellular iron, which is can then imported into other cells via the transferritin receptor[96]. Surprisingly, a marked decrease in transferritin expression, and a marked increase in transferritin receptor expression was observed. This suggest that the cell as a whole has a net deficit for iron, even given the excessive amounts of hydrogen peroxide and the increased potential to generate hydroxyl radicals.

The mitochondria have been implicated to be the major source of intracellular ROS, in particular superoxide[95]. Transcriptional response of enzymes that have been associated with ROS generation was examined (figure 2.4). In general, enzymes associated with the activity of the citric acid cycle were

up-regulated while enzymes for other pathways were down-regulated. The mono amine oxidases did not exhibit changes in expression. Alpha – glycerophosphate dehydrogenase 1 (GPD1), A-ketaglutarate dehydrogenase complex (OGDH), Dihydrooratate dehyrdogenase (DHOH) showed strong down-regulation at later time points. Conversely, the cytochrome b5 reductase, aconitase, and pyruvate dehydrogenase complexes were increasingly up-regulated over time. The electron transport chain has been suggested as the major source of sub cellular ROS with 0.1% to 1% of reactions generating a superoxide ion[95] (figure 2.5). Complex II, Cytochrome C, and enzymes responsible for biosynthesis of Coenzyme Q showed strong late expression. Complex I, III, and IV had members that were both strongly up and down regulated (data not shown). Taken as a whole, this suggests in response to oxidative stress mitochondrial activity is increased especially citric acid cycle activity. As many of these enzymes are associated with energy and in fact are up-regulated, it appears the cell has need for additional energy during oxidative stress beyond normal resting activity.

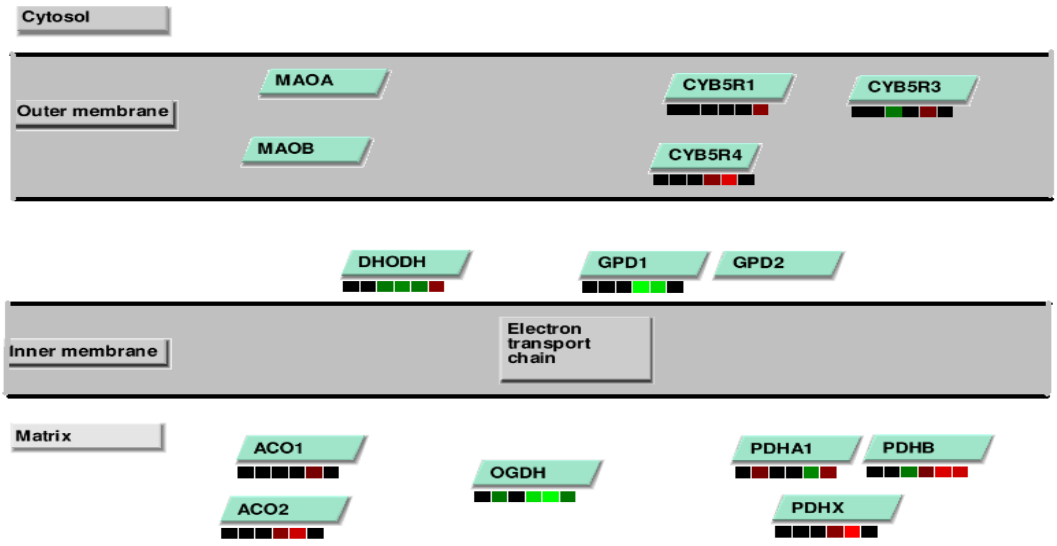


Figure 2. 4: Mitochondrial sources of hydrogen peroxide are not uniformly up- or down- regulated.

.

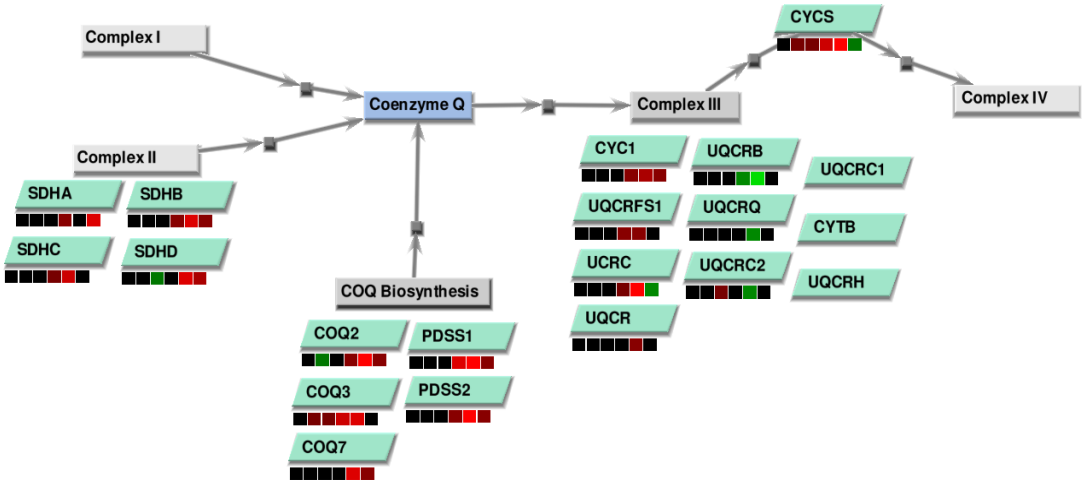


Figure 2. 5: Electron transport chain shows an increase in expression, indicating an increase in metabolic ability like due to an increase in energy demands of the cell.

NADPH oxidase (NOX) and dual oxidase (DUOX) family of proteins generate ROS species in a regulated manner[98]. They are structurally similar to phagocyte NADPH oxidase (PHOX), but generate low amounts of ROS to

mediate cell growth and signaling. No any noticeable expression changes in the NOX family were observed; however the two NOX accessory proteins, NOXO1, and NOXA1 showed a marked repression. NOXO1 and NOXA1 greatly increase the rate by which NOX1 generates hydrogen peroxide[98] (see figure 2.6). This suggests the cell is potentially limiting ROS generation of redox signaling via this the NOX family. However, as a whole the cell does not down-regulate enzymes or activities associated with ROS generation.

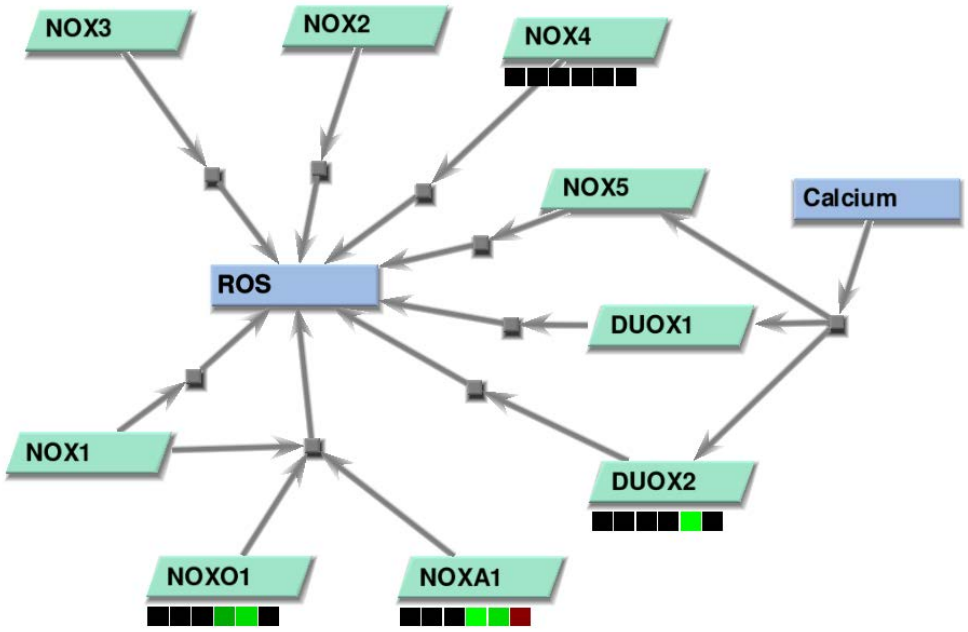


Figure 2. 6: NOX family of proteins is not statistically different, but de-regulation of accessory proteins NOXO1 and NOXA1 in response to hydrogen peroxide reduces ability of NOX1 to generate ROS.

ROS clearance. The three most common ROS are the hydroxyl radical (OH-), hydrogen peroxide (H202) and superoxide (O3-). Hydroxyl radicals are too reactive to clear before causing damage; hydroxyl radicals can be generated directly by the Fenton reaction[96], which describes a self sustaining oxidization of

ferrous iron by hydrogen peroxide. Hydrogen peroxide in turn is generated by the reaction of superoxide and water (figure 2.6). To compensate, the cell utilizes a large and ROS-specific system of enzymes to efficiently clear intracellular ROS. Enzymes are localized in either the cytoplasm or mitochondria on a protein specific basis as opposed to ROS specific basis.

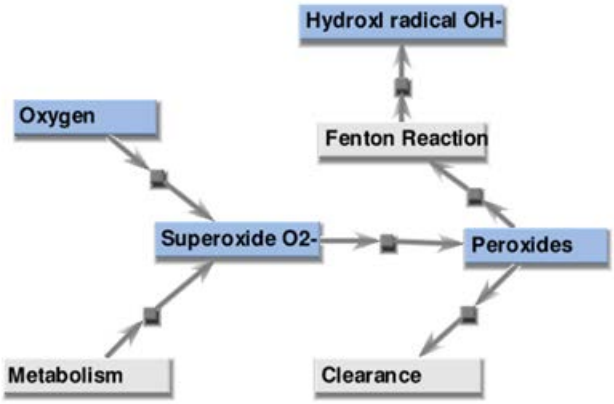


Figure 2. 7: Major reactive species in the cell are Superoxide, Hydrogen Peroxide, and the hydroxl radical. The cell maintains a network of antioxidant cells to clear ROS, which are endogenously generated.

The superoxide dismutases (SODs) convert superoxide into hydrogen peroxide[99]. There was strong up-regulation of the mitochondrial SOD2 and weak up-regulation of SOD3, which has been suggested to have extracellular activity. This is consistent with our previous observation that mitochondrial sources of ROS are up-regulated due to the cell's energy demand.

The three major protein families responsible for hydrogen peroxide reduction to water are the peroxiredoxins[100, 101] (PRDX), glutathione peroxidases[102](GPX), and catalse[103] (CAT). In addition, a large accessory network comprised of the sestrins (SESN) and thioredoxins (TXN) is responsible

for the recycling of PRDXs as they are inactivated after processing ROS[101]. This is considered the mechanism by which cells tolerate a basal level of hydrogen peroxide for cell signaling purposes, while still maintaining clearance capabilities. In general ROS clearance enzymes are strongly up-regulated in response to oxidative stress, regardless of their cellular localization (figure 2.8).

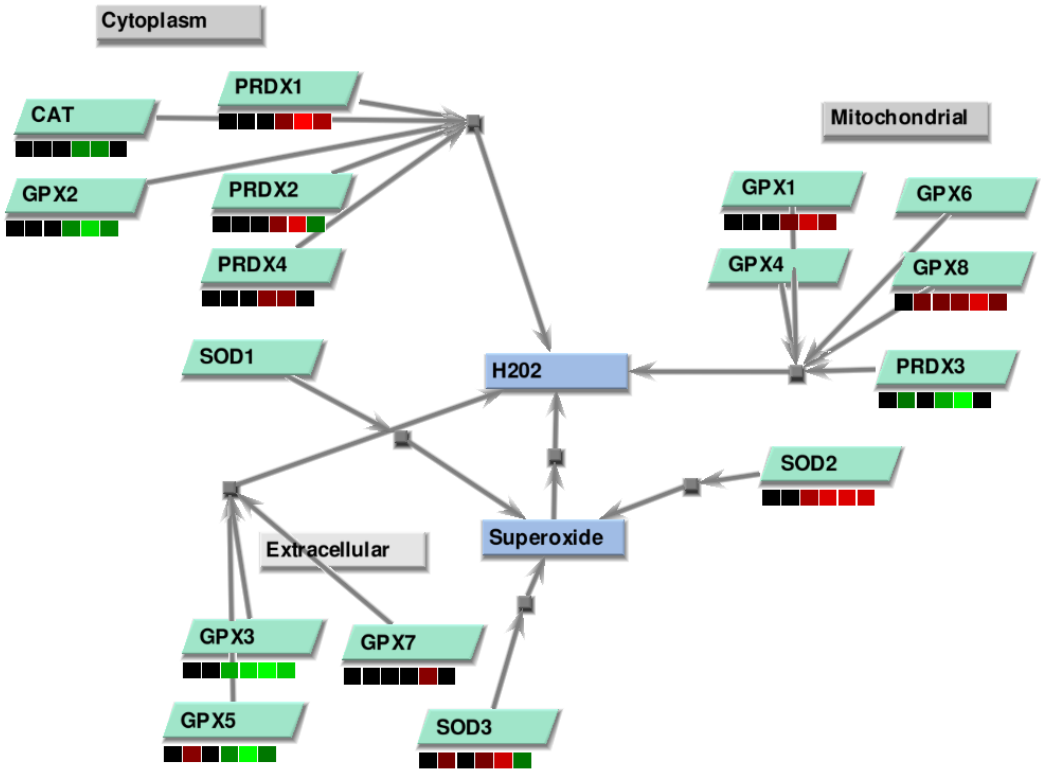


Figure 2. 8: Antioxidant proteins involved in ROS clearance become up-regulated in response to hydrogen peroxide

While it is not surprising that hydrogen peroxide and hydroxyl radical clearance mechanisms are up-regulated in response to oxidative stress, it is surprising to find SOD2 up-regulated to such a degree while SOD1 remains unchanged. This suggests mitochondrial activity is a specific concern with regard

to additional ROS burden experience by the cell. The ROS clearance network is itself energy intensive, and is a possible reason for increased cellular demand for energy.

MAPK/P38 signaling. The MAPK is a three tiered kinase cascade activated by cellular stress or extracellular ligands[104]. MAPK is a crucial pathway regulating proliferation, cell survival, differentiation and death. H202 has been reported to activate ERK1, ERK2, ERK5, JNK, and p38. JNK and p39 in particular are critical for stress induced cell death. Western blotting of MAPK phosphoproteins indicated increased activity of AKT, ERK, and JNK (figure 2.8). Specific kinase inhibitors revealed caspase activity is predominately through ERK, JNK, and p38 (figure 2.9) as expected[104]. In addition, many members of the MAPK cascade showed an up-regulation in gene expression (figure 2.10).

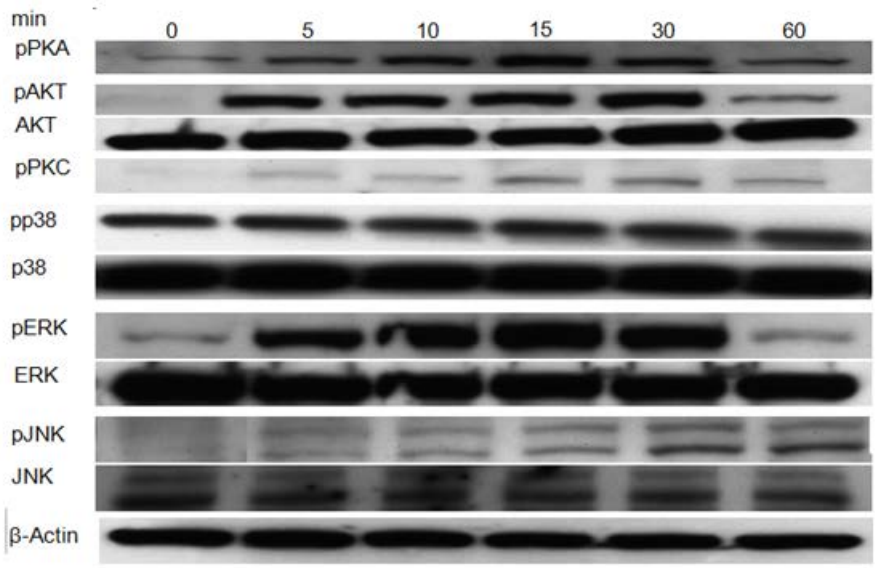


Figure 2. 9: Western of MAPK signaling cascade shows kinase activity. Cells were induced with H202 and harvested over a time course. p indicates against phosphorylated form.

H_2O_2	-	+	+	+	+	+	+	+	+
PKAi	-	-	+	-	-	-	-	-	-
AKTi	-	-	-	+	-	-	-	-	-
PKCi	-	-	-	-	+	-	-	-	-
ERKi	-	-	-	-	-	+	-	-	-
JNKi	-	-	-	-	-	-	+	-	+
P38i	-	-	-	-	-	-	-	+	+

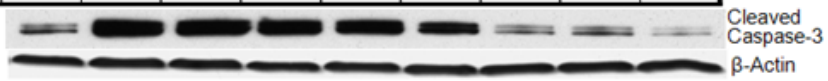


Figure 2.10: Kinase specific inhibitors reveal caspase 3/7 activity is regulated by ERK, JNK, and p38. HMVEC-L cells were induced with 100 μm.

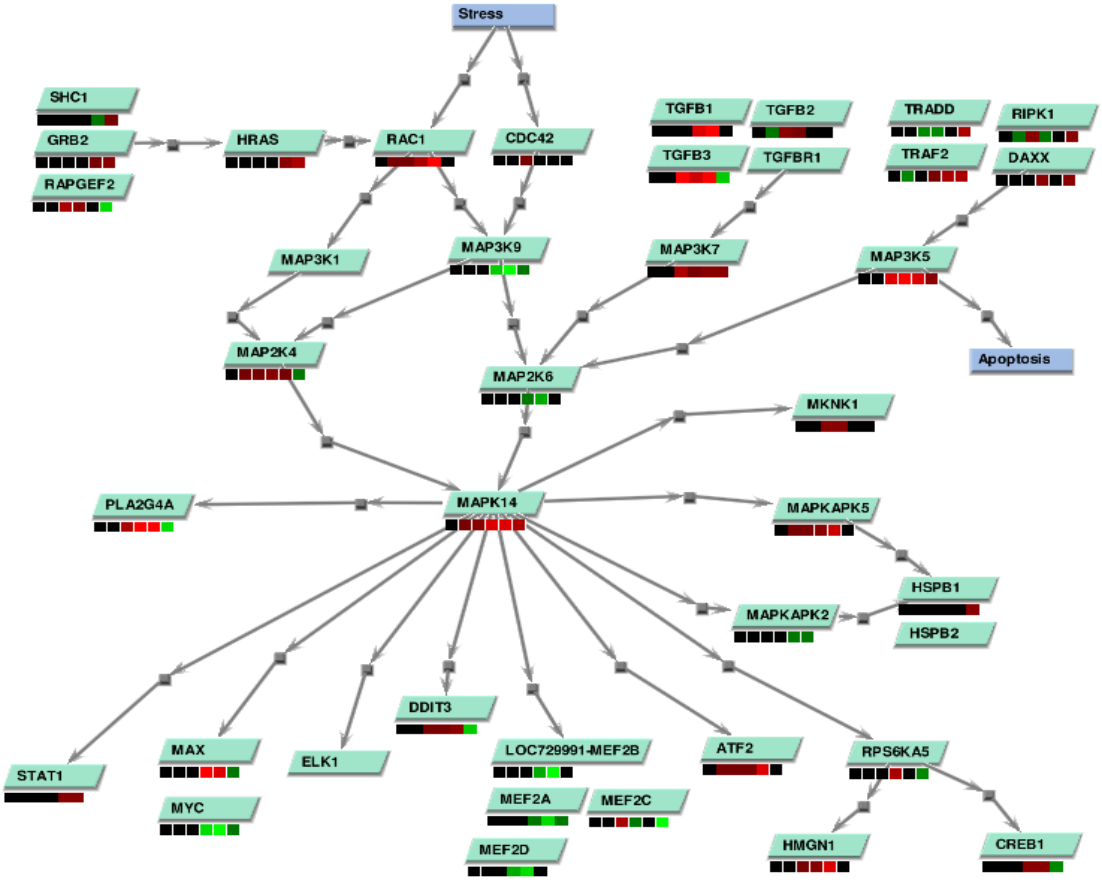


Figure 2. 11: Gene expression profiles for the MAPK signaling pathway show an increase in expression correlating with increased activity.

NRF Pathway. NFE2L2 (NRF2) is an important mediate of cellular response to oxidative stress[105]. Normally sequestered in the cytoplasm by KEAP1, ROS will oxidize residues on KEAP1 allowing NRF2 to translocate to the nucleus and activate target gene expression. In combination with the MAF family of proteins, NRF2 up-regulates the expression of significant antioxidant genes, including HMOX1 (see figure 2.12). BACH1 will terminate NRF2 activation by competition of response elements[106], thus nuclear export of BACH1 is required for sustained activation.

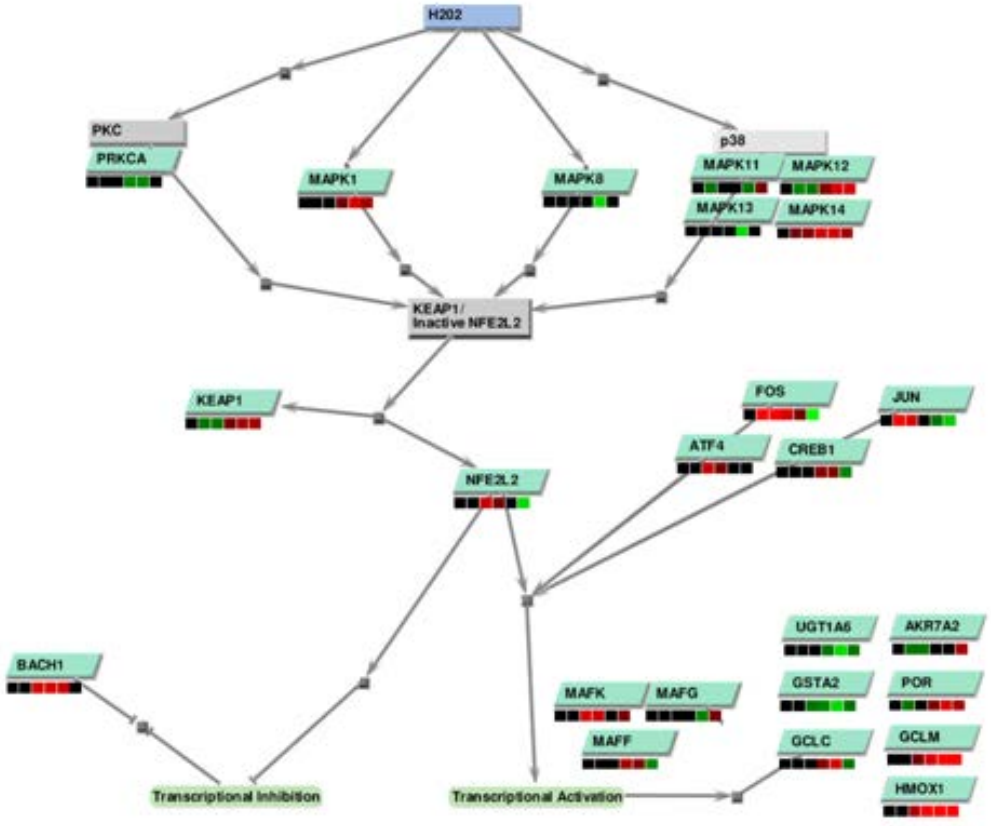


Figure 2. 12: ROS cause NRF2 to be released from KEAP1. NRF2 translocates to the nucleus and increases expression of antioxidant target genes such as HMOX1.

DNA Repair Pathways. ROS are genotoxic as they can directly damage DNA either by oxidizing nucleotides or introducing double stranded breaks[107]. ATM and ATR are critical sensors of DNA damage, and phosphorylate many proteins critical for DNA damage repair. Five major DNA repair pathways have been characterized[108]: homologous recombination repair (HRR), base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR),and non-homologous end-joining (NHEJ). NHEJ was the complex most consistently up-regulated in response to oxidative stress (figure 2.14). NHEJ involves the ligation bound by Ku (XRCC5/XRCC6) by DNA ligase 4 to double stranded breaks. The HRR and NER complexes also showed strong up-regulation, although several members in each were down-regulated. BER has been identified as the pathways involved in removing oxidized bases, but neither BER nor MMR exhibited a strong pattern of expression.

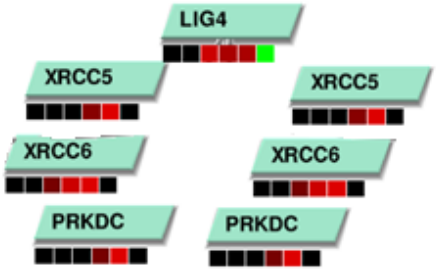


Figure 2. 13: Non-homologous end-joining follows pattern of strong late induction.

For HRR, NHEJ, and NER, follow an increasing, late time point profile suggesting that DNA repair pathways remain active even as apoptosis is underway. DNA repair also has been suggested to require large amounts of cellular energy, with inefficient amounts leading to necrosis[108].

Figure 2. 14: Members of the homologous recombination repair pathway exhibit strong patterns of expression.

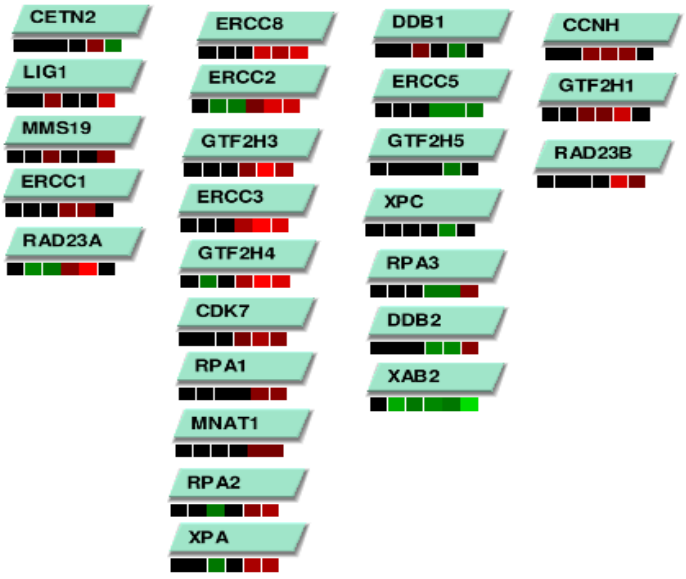


Figure 2. 15: Nucleotide excision repair pathways show a late pattern of up-regulation for some members.

Figure 2. 16: Mismatch repair genes show weak expression profiles.

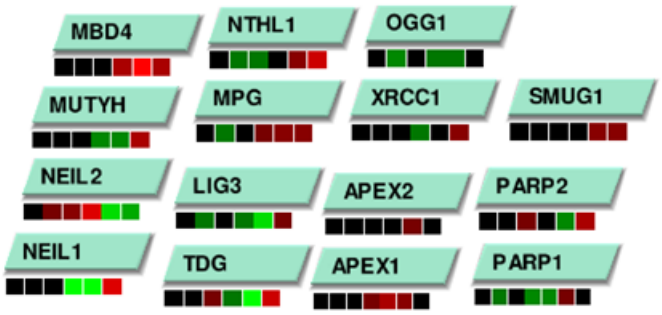


Figure 2. 17: Base excision repair members follow an inconsistent profile.

BCL2 Family. The mitochondrion is the primary energy powerhouse of the cell[2]. Apoptosis leads to loss of mitochondrial membrane potential, increased permeability, and eventual fission and fragmentation[109]. The mitochondria have two distinct membranes, an outer and inner membrane that allows the formation of an inter-membrane space and sequestered aqueous center called the matrix. p53 signals and mediates the BLC2 family of proteins, which contains both anti- and pro- apoptotic members[110] (figure 2.18). Pro-apoptotic members allow the release of mitochondrial matrix products and cytochrome C into the cytosol, either through activation of existing pores or formation of new ones. This release plays a critical role in apoptosis through the formation of the apoptosome and the inactivation of inhibitor of apoptosis proteins[111] (IAPs).

Inhibition of BCL2 (figure 2.18) has been shown to be regulated by NF-kB in response to chronic exposure to hydrogen promotes[112]. Additional repression of XIAP and up-regulation of TNF and FAS were also consistent with gene expression profiles (data not shown). NFKBIA was also consistently up-regulated

which indicates NF-kB activity (data not shown). However, NF-kB cell death was shown to act through caspase independent, but PARP1 dependent mechanism.

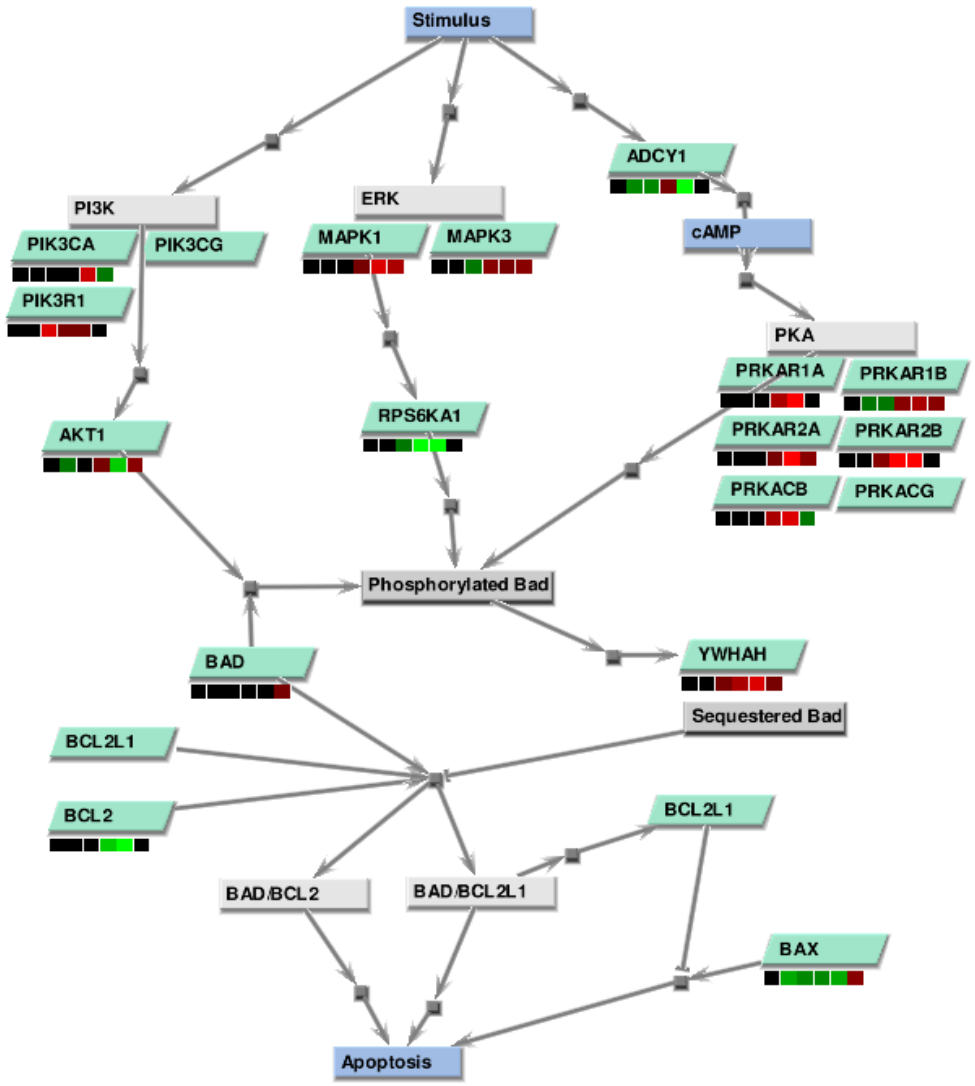


Figure 2. 18: Bad gene expression profiles.

Loss of Mitochondria. Cytochrome C is a member of the electron transport chain, and thus resides on the inner membrane wall to utilize the

electrical gradient present[109, 113]. While the only member of the electron transport chain to be soluble, it has been estimated that only a small fraction of cytochrome C is free in the inter-membrane space. The inner mitochondrial membrane forms folds called cristae to allow increased oxidation. OPA1 is responsible for cristae remodeling, where the inner and outer membranes become fused to allow the efficient release of cytochrome C[114]. OPA1 and other fusion proteins were strongly up-regulated in response to oxidative stress (data not shown). Mitochondrial membrane potential (MMP) was observed from JC-1 as a measure of healthy mitochondria. There was a distinct loss of MMP after 12 hours of 100 μm H202 induction using JC-1 (data not shown).

Caspase Cascade. As a result of cell signaling and DNA damage, p53 mediates the release of cytochrome C (CYCS) from the mitochondrial which forms the apoptosome with APAF1[111]. This complex signals through the caspase cascade which ultimately results in DNA fragmentation and cell death. Caspases are initially synthesized from genes as inactive pro-caspases. As a result of apoptotic signaling, the initiator caspases (CASP2, CASP8, CASP9, CASP10) are subsequently cleaved to their active forms[2]. These initiator caspases then cleave the executioner caspases (CASP3, CASP6, and CASP7); each subsequent cleaved protease amplifies caspase activity by cleaving additional caspases. Apoptosis is reversible however until executioner caspase activation. Execution caspases are responsible for cleaving a variety of protein substrates that lead to apoptosis. Both APAF1 and CYCS are transcriptionally up-regulated in response to hydrogen peroxide (figure 2.19). Caspase-Glo 3/7 showed a

strong induction of caspase 3/7 activity (figure 2.20) after 12 hours. According to expression profiling, the majority of caspases and in particular the executioners are strongly up-regulated (figure 2.19). Protein array data showed consistent loss of pro-caspase abundance indicating apoptosis at later time points (figure 2.21). Interestingly as caspase 8 typically initiates apoptosis through external stimuli, caspase 8 shows a marked down-regulation, and a loss of pro-caspase abundance.

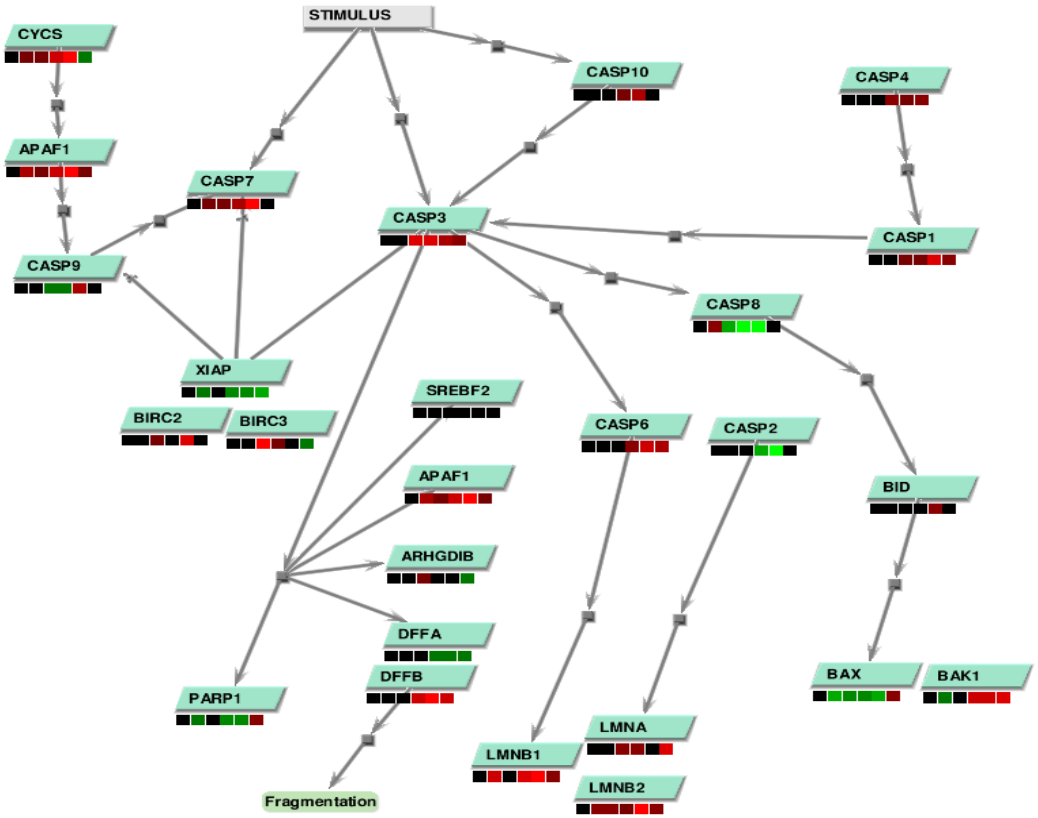


Figure 2. 19: Gene expression profiles of the caspase cascade indicate strong up-regulation

Figure 2. 20: Combined caspase 3 and 7 activity is greatly increased in response to hydrogen peroxide as measured by a peptide whose cleavage by caspases results in luminescence.

.

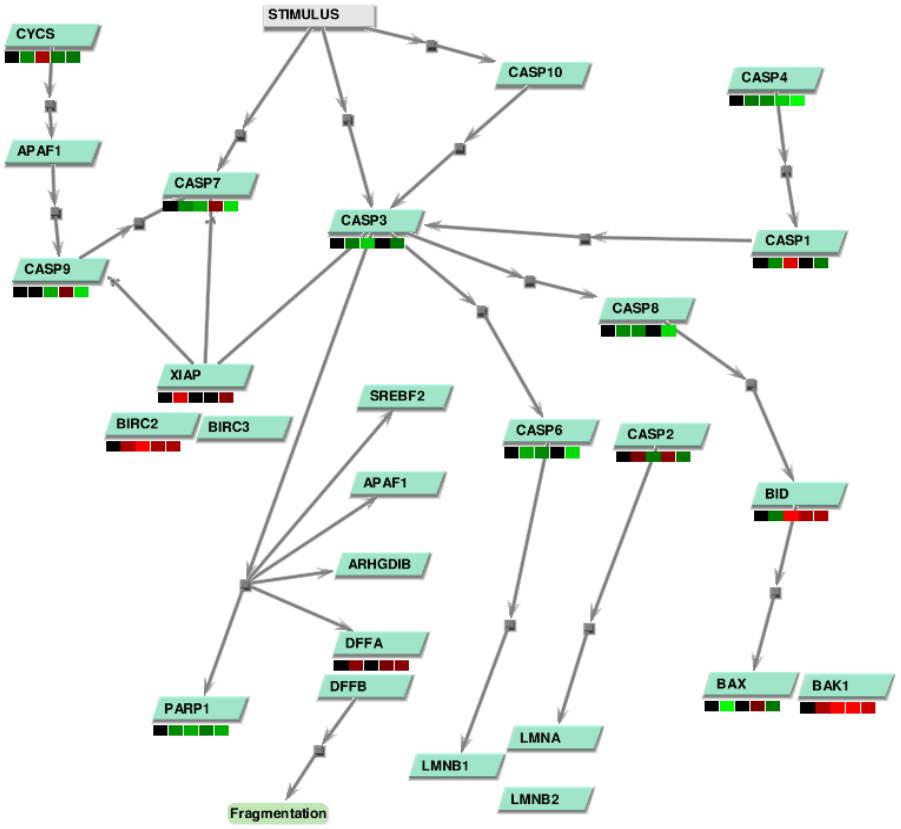


Figure 2. 21: Protein abundance against pro-caspase forms shows a lack of abundance as caspases are cleaved

TP53, ROS Clearance, and MTOR. Budanov et al. 2008 have a described a relationship between TP53, sestrins, and MTOR that leads to inhibition of cell growth[115]. SESN1 and SESN2 are target genes of TP53, and are

up-regulated upon genotoxic stress. The sestrins family of proteins is involved in the continual recycling of peroxiredoxins. As peroxiredoxins clear hydrogen peroxide they become inactivated; peroxiredoxins are recycled by the thioredoxins, which in turn are inactivated and require recycling by sestrins. Buadanov et al 2008 demonstrated that sestrins interact with the AMPK signaling. AMPK phosphorylates the TSC2 complex to reduce loading and activity of RHEB, a positive regulator of MTOR.

Gene expression profiles from microarray analysis indicate SESN1 and SESN2 are indeed activated with greater intensity in response to oxidative stress (figure 2.22). Corresponding proteins involved in peroxiredoxins recycling are also up-regulated. AMPK module shows an increase in expression, suggesting activation. However, the TSC complex is down regulated, and RHEB is up-regulated. Antibody array data showed a decrease in protein abundance of MTOR (data not shown), so transcriptional events might be regulatory feedback in response to decreased MTOR activity.

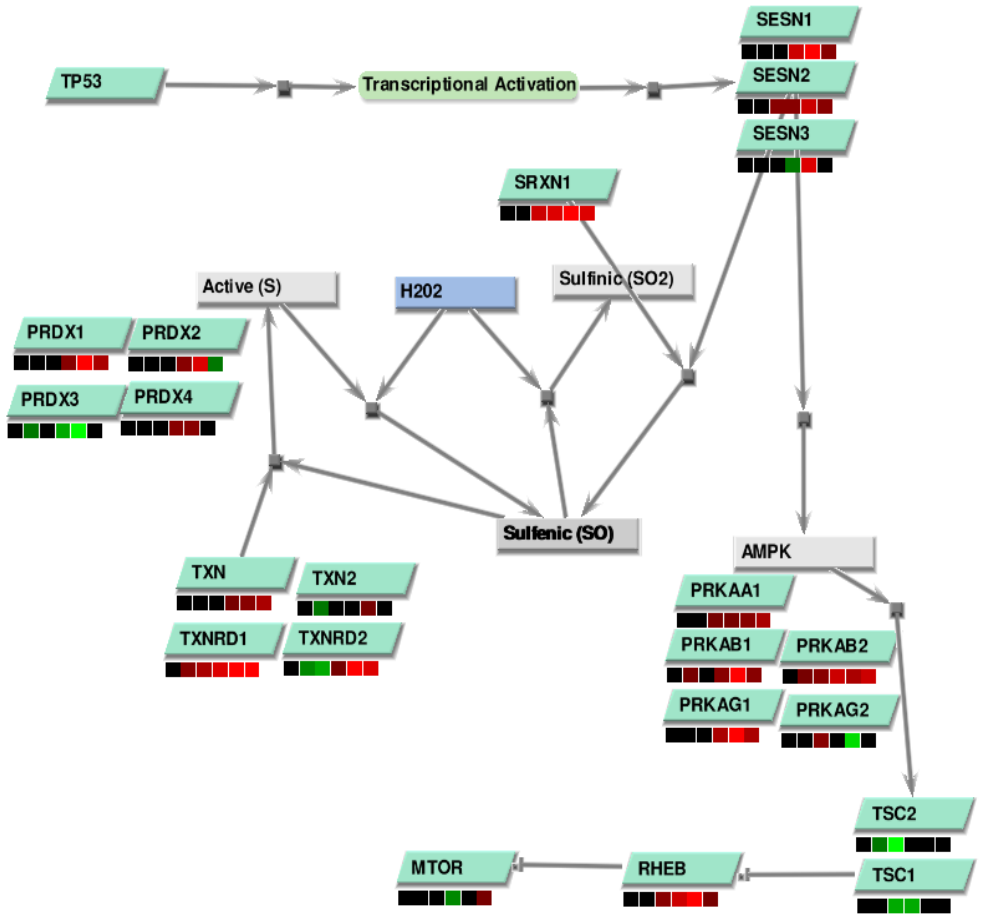


Figure 2. 22: Relationship between genotoxic stress, p53, sestrins, and MTOR indicates a dynamic and active network in response to hydrogen peroxide.

Heme Function, Synthesis and Degradation. Genes involved in the synthesis, function, and degradation are significant expressed in response to hydrogen peroxide (figure 2.23). Heme is a chemical porphyrin, or iron containing compound[116]. Synthesis of heme originates from mitochondrial succinyl-CoA by the enzyme ALAS1. This step is known to be rate limiting in non-erythroid cells. Synthesis continues from the mitochondrion to the cytoplasm back into the

mitochondrion where attachment of the iron ion occurs. Synthesized heme must then be exported back into the cytoplasm to be bound by cytochrome C.

Release of cytochrome C from the inner mitochondrial membrane is a critical step in the intrinsic apoptosis pathway. Only active cytochrome C can bind with APAF1 and drive apoptosis[117]. Conversion of cytochrome C from apo- (inactive) to holo- (active) form requires binding of heme which can only occur at on the cytosolic side of outer mitochondrial memberane[116]. Binding of heme, specifically heme C, to CYCS is mediated by cytochrome c heme-lyase (HCCS).

Unbound heme has been shown to have a cyto-protective and anti-oxidative effect by its degradation by the enzymes heme oxygenase 1 (HMOX1) and heme oxygenase 2 (HMOX2) [118]. The HMOXs cleave heme into biliverdin, which has cyto-protective effects, and iron which requires sequestration by ferritin. It has been shown in HMOX1 knockouts that fibroblasts exposed to hydrogen peroxide are more susceptible to apoptosis.

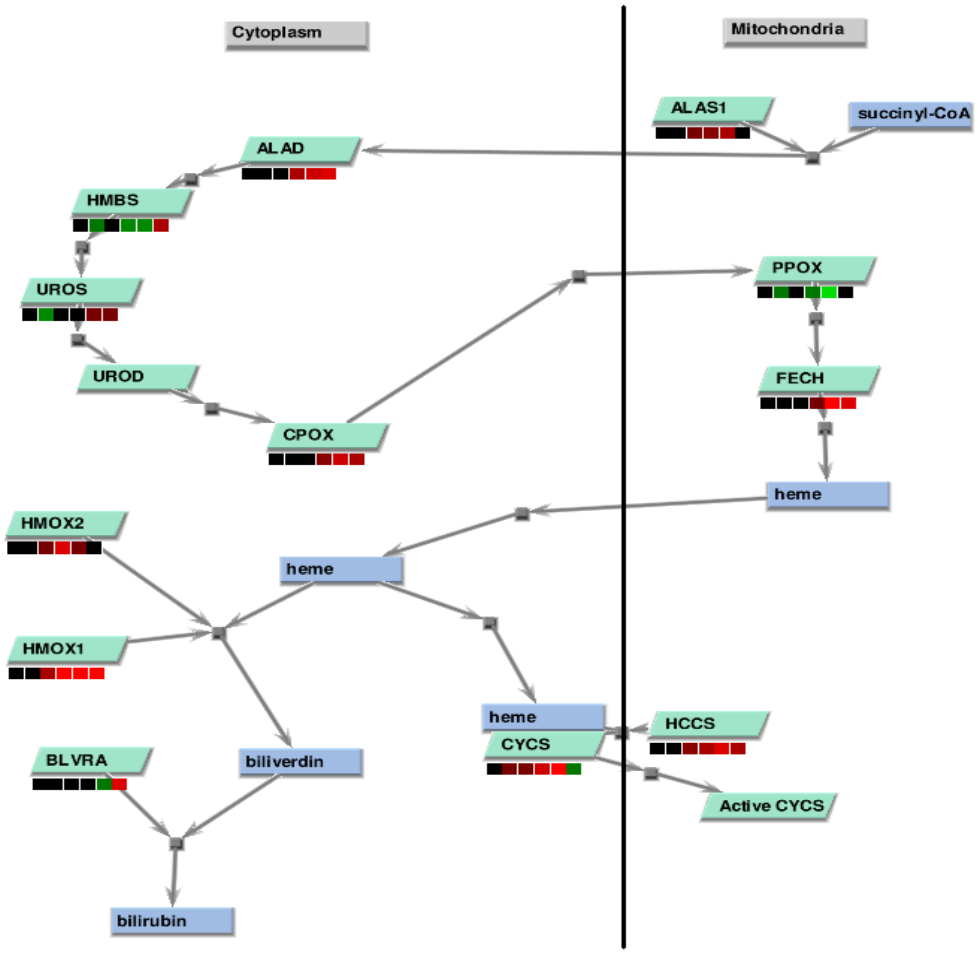


Figure 2. 23: Critical interacting partners of heme are up-regulated in response to hydrogen peroxide: ALAS1, the rate limiting enzyme for the synthesis of heme; HCCS, which is required for binding to CYCS; CYCS, which is a major effector of metabolism and apoptosis; and HMOX1, which degrades free heme into biliverdin and iron.

Reproducibility Issues

Verification of Microarray Trends. A validation strategy was undertaken to confirm the importance of several factors. However, during screening using LDH and caspase 3/7 following hydrogen peroxide induction positive control wells failed and showed no change in response. Repeated experiments with multiple lots of HMVEC-L showed no significant LDH activity at 6 hours after 100

μm H202 induction (figure 2.24). Cells showed no caspase 3/7 activity, except when induced by staurosporine. Cells required a 10 X fold increase in hydrogen peroxide to detect LDH leakage/ Visual inspect of induced cells also showed no phenotypic difference until 1mM H202 (see figure 2.25) after 6 hours. Real time quantitative PCR was performed on a select number of target genes over varying concentrations of hydrogen peroxide to compare current experimental conditions with previous microarray data (figure 2.26). No concentration adequately validated trends from the microarray data, although H202 expression was markedly distinct from staurosporine. Ratios from 500 μm H202 were the most robust, although cells exhibited no signs of cell death.

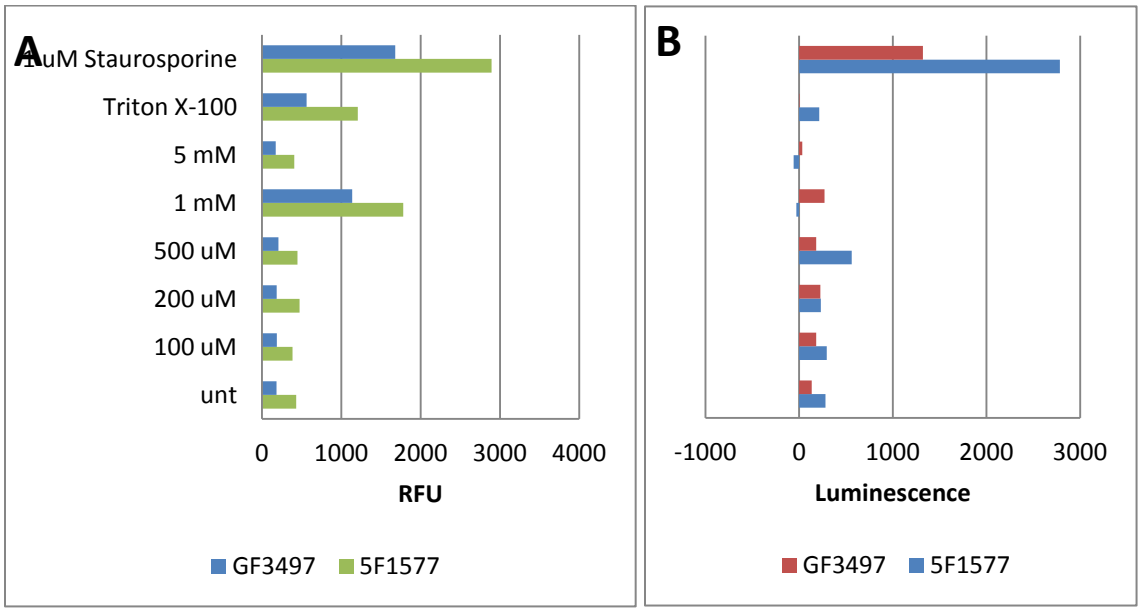


Figure 2. 24: LDH release of HMVEC-L cells in response to increasing concentrations of H202 after 6 hours. Cells from two separate donor lots, 5F1577 and GF3497, were tested.

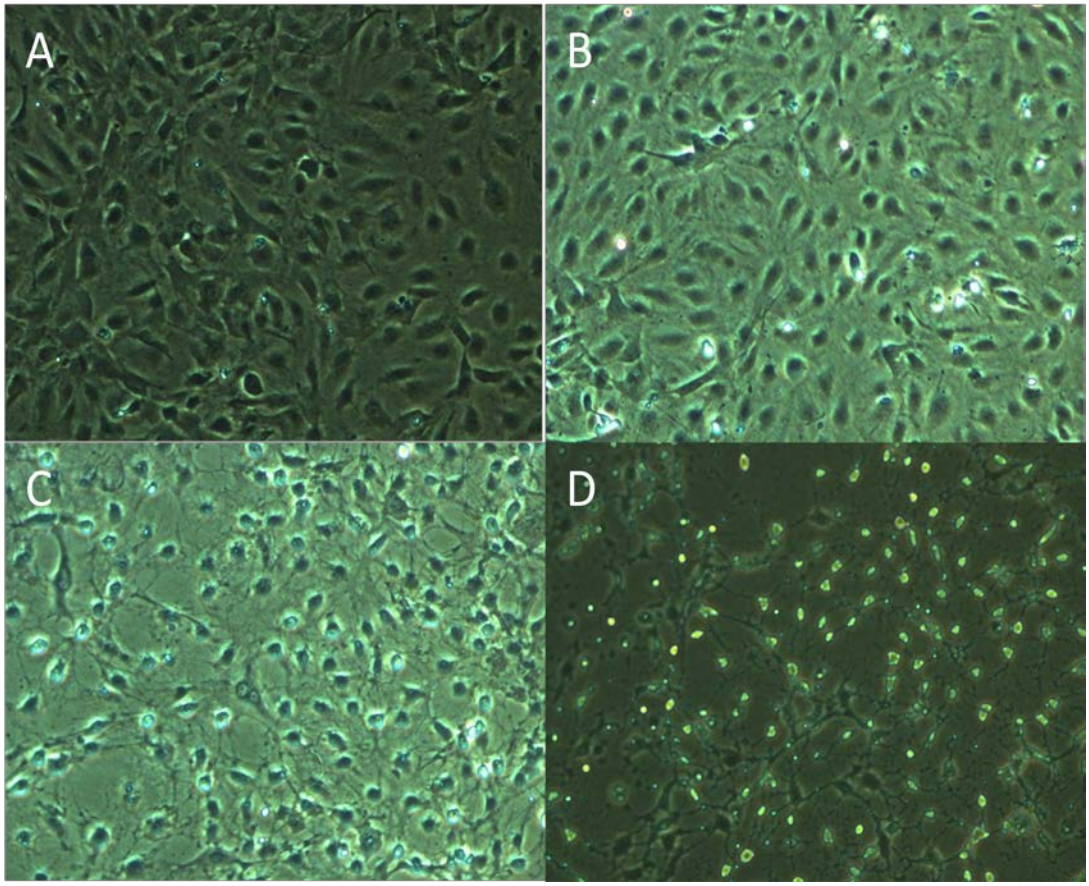


Figure 2. 25: White light images of HMVEC-L cells after 6 hours of (A) 0 μm H202, (B) 100 μm H202, (C) 1 mM H202, and (D) Staurosporine. No evidence of cell death until 1000 μm H202. Cells completely undergoing apoptosis in 6 hours with induction of Staurosporine (D).

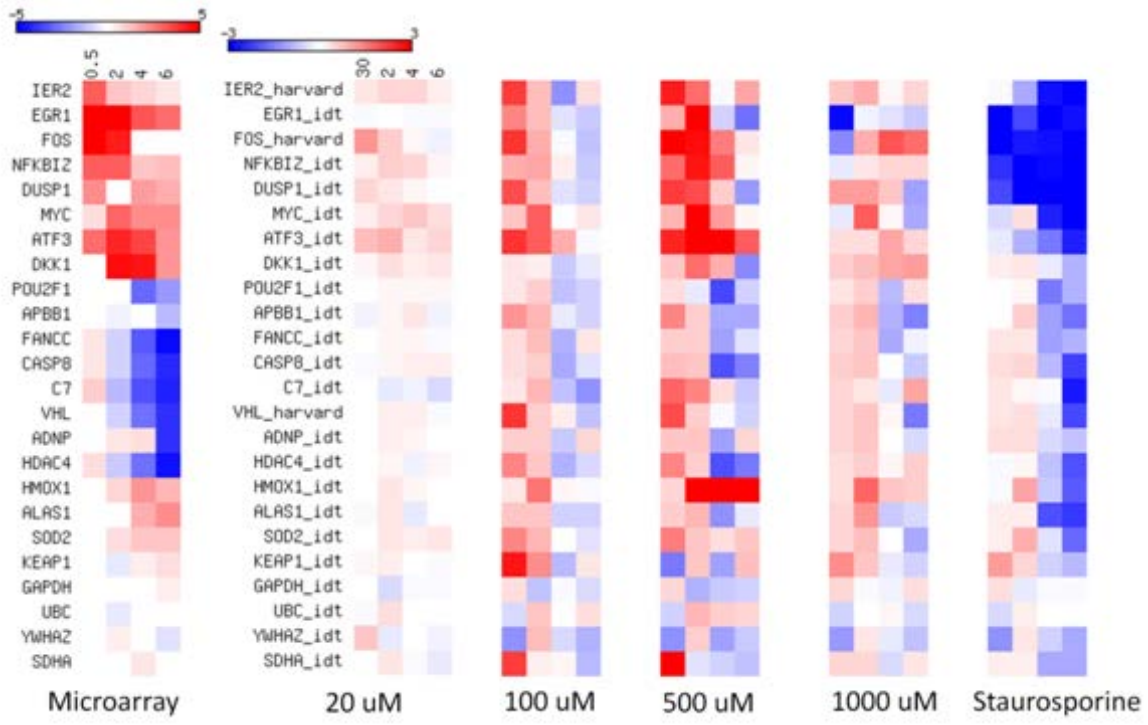


Figure 2. 26: Microarray data from HMVEC-L cells induced with 100 H202 µm compared to real time quantitative PCR with increasing concentrations of H202. Cells were harvested after 30 minutes, 2 hours, 4 hours and 6 hours. No concentration completely recapitulated microarray data. All values are fold changes calculated from time matched untreated.

Testing of H202 and HMVEC-L. As series of experiments were undertaken to isolate the cause of experimental variance. HMVEC-L cells stained for the endothelial markers CD31 and VWF, and were capable of uptake of acetylated LDL (figure 2.27). Cells tested negative for mycoplasma contamination. As hydrogen peroxide is a notoriously unstable compound, experiments were performed to test the activity of H202 using KMNO4 titration experiments (data not shown) and spectrophotometer data (figure 2.28). All experiments showed concentrations of h202 within manufacture specifications.

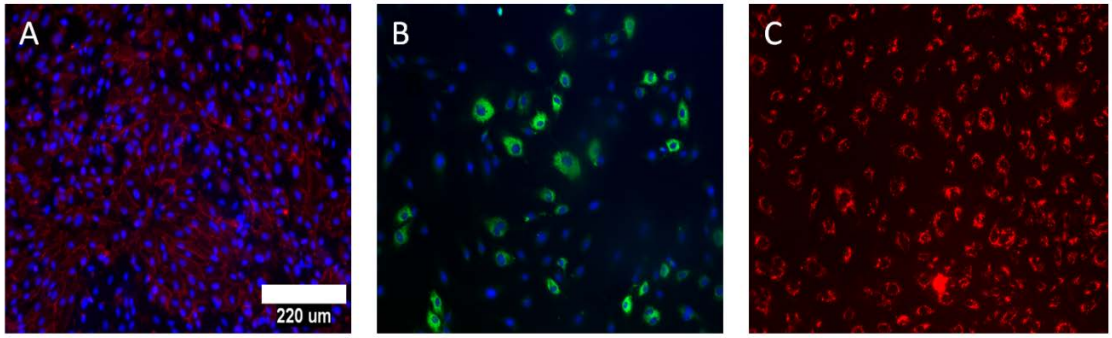
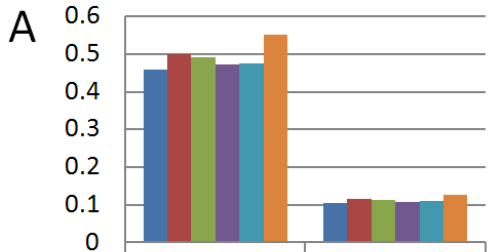


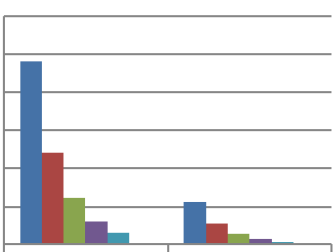
Figure 2. 27: Staining of (A) CD31, (B) VWF, (C) Acetylated DL uptake confirms cells are likely endothelial cells. Dapi is in blue.

A



	absorbance	concentration (M)
1	0.458333333	0.105122324
2	0.499666667	0.114602446
3	0.489	0.112155963
4	0.471	0.108027523
5	0.474666667	0.108868502
1 Heated	0.551	0.126376147

B



	absorbance	concentration (M)
9.8	47.986	11.0059633
4.9	24.077	5.522247706
2.45	12.237	2.806651376
1.225	5.938	1.361926606
0.6125	2.938	0.673853211
0.105122324	0.472	0.108256881

Figure 2. 28: Testing of H202 stock demonstrated performance to manufacture's specifications. Spectrophotometer data of 5 separate H202 lots at stock concentration of 9.8 M diluted to 100 mM. Absorbance values and calculated concentrations are shown. B. Dilution series of H202 from Lot 5. Absorbance and calculated concentrations are shown.

Calibration of Assays. The source of experimental issues was not isolated, the sensitivity and specificity of the assays was tested (figure 2.29 and figure 2.30). Positive controls for Cytox-One and Caspase 3/7-Glo were titrated for optimal signal. Sensitivity was tested using positive controls against serial

dilutions of cells. Assays performed up to manufacturers specifications and repeatedly indicated no cell death at low concentrations of hydrogen peroxide (data not shown).

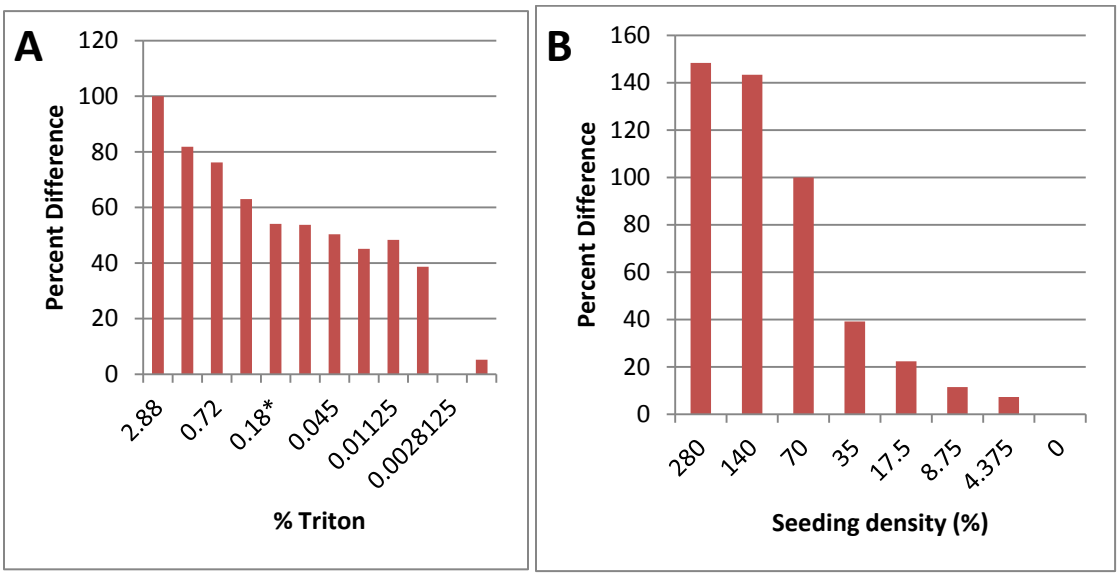


Figure 2. 29: Calibration of Cytox-One LDH release assay. (A) Titration of Triton X-100 concentrations reveal high concentrations of triton x-100 do not interfere with the detection of LDH and provide a robust signal. Manufacturer's recommendation is addition of 0.18% Triton X-100 as a positive control. (B) Addition of 1.44% Triton on decreasing confluence of cells (%) shows a generally linear response except for highly confluent wells. As 100 % confluence is 140 % of 70% confluence, maximum signal is limited by the space available to cell attachment.

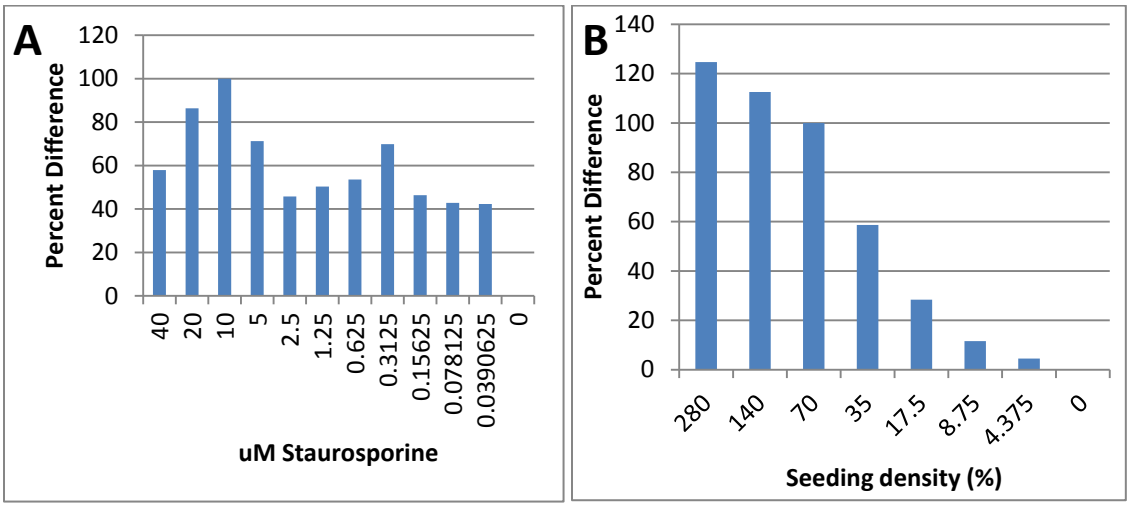


Figure 2. 30: Calibration of caspase 3/7- glo assay. (A) Titration of staurosporine concentration. (B) Caspase activity across various seeding densities using 10 µm Staurosporine.

Loss of Hydrogen Peroxide in Media. Given the high instability of H202, the rate of disippation of H202 using a colorimetric detection kit was measured, which found h202 activity was compeltely lost by two hours (data not shown). Further testing revealed that H202 likely reacts with basal EGM2 media alone, losing over half the added concentration after 10 minutes (figure 2.31). Addition of serum or growth factors, including ascorbic acid, did not change this effect. Testing of hydrogen peroxide in DMEM yielded similar results. However, H202 does not react with PBS and remained stable past 1 hour indicating H202 was likely reacting with media components as opposed to disipatting.

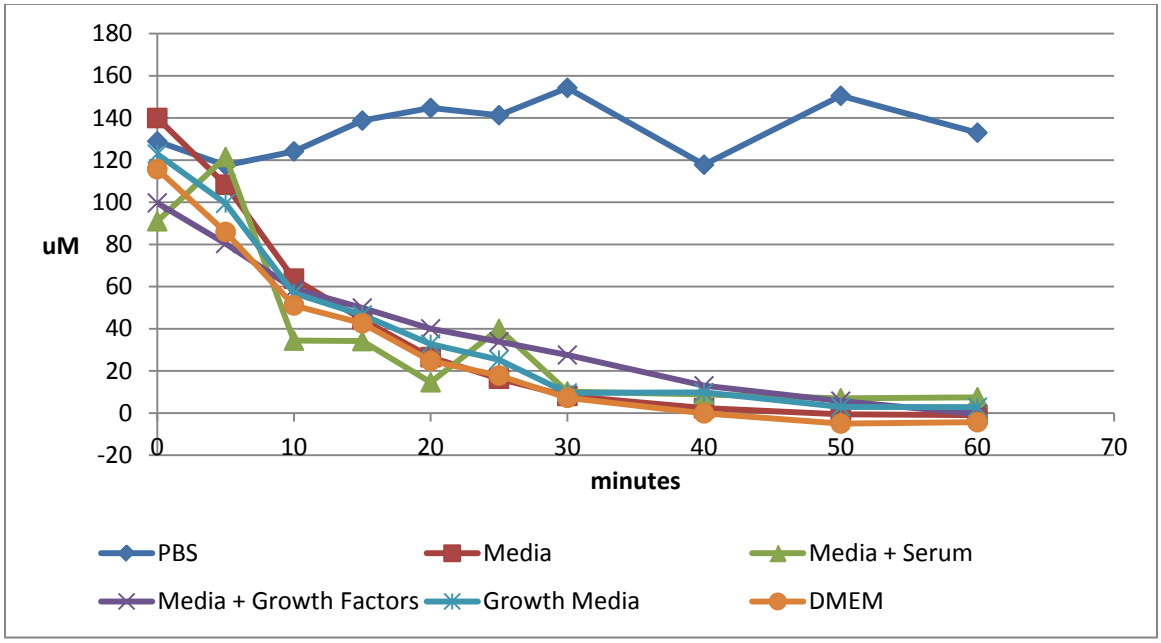


Figure 2. 31: 100 µm hydrogen peroxide is rapidly dissipated in cell-free culture media but not PBS. Approximately half of H202 concentration is lost after 10 minutes. Detection performed using an h202 colorimetric detection kit from Enzo Life Sciences.

Assay Conditions. Protocols were adjusted to compensate for H202 loss, such that h202 was stored in a working solution of PBS, and added within in two minutes of addition to media. While cells did exhibit some degree of cell death at lower concentrations, hydrogen peroxide was still required at greater than 700 µm to observe adequate LDH release. Caspase 3/7 activity was detected though for concentrations as low as 500 µm. As assay half lives are 9 hours, data presented is an under representation of the true signal however. Visually, cells at 700 µm rapidly died in response to hydrogen peroxide while cells exposed to 500 µm died at time points past 6 hours.

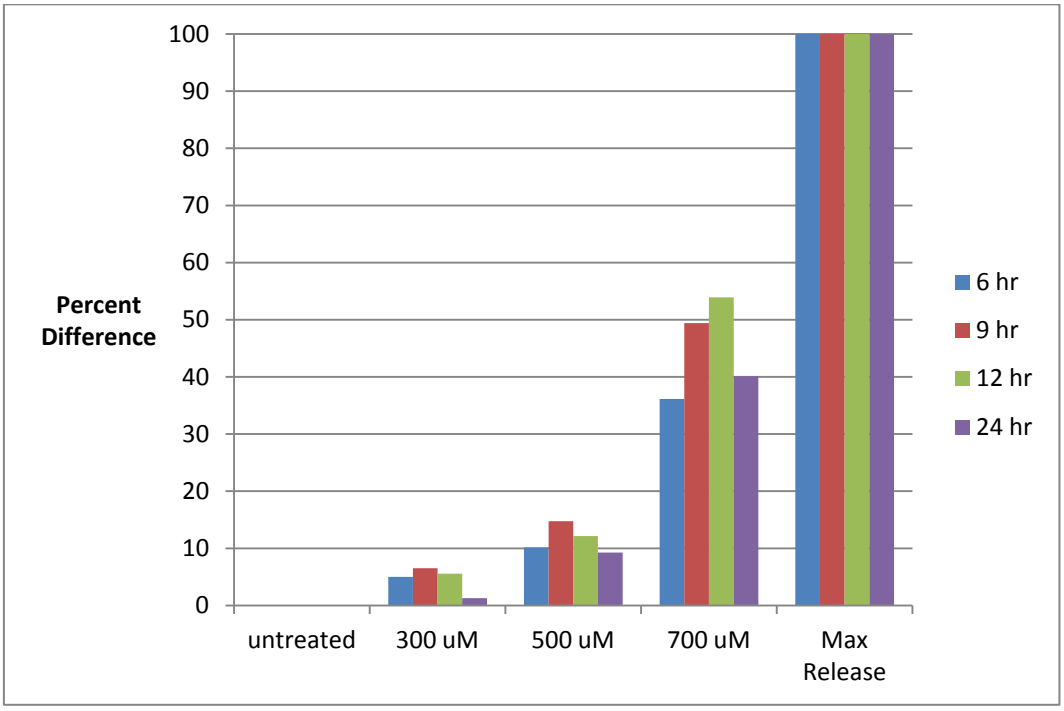


Figure 2. 32: Rapid of addition of H202 increases degree of cell membrane leakage. 700 µm H202 killed approximately half of all cells at 9 hours. As half life of LDH is 9 hours, all cells were likely dead at post 18 hours given the robust response at 12hr and 24hr.

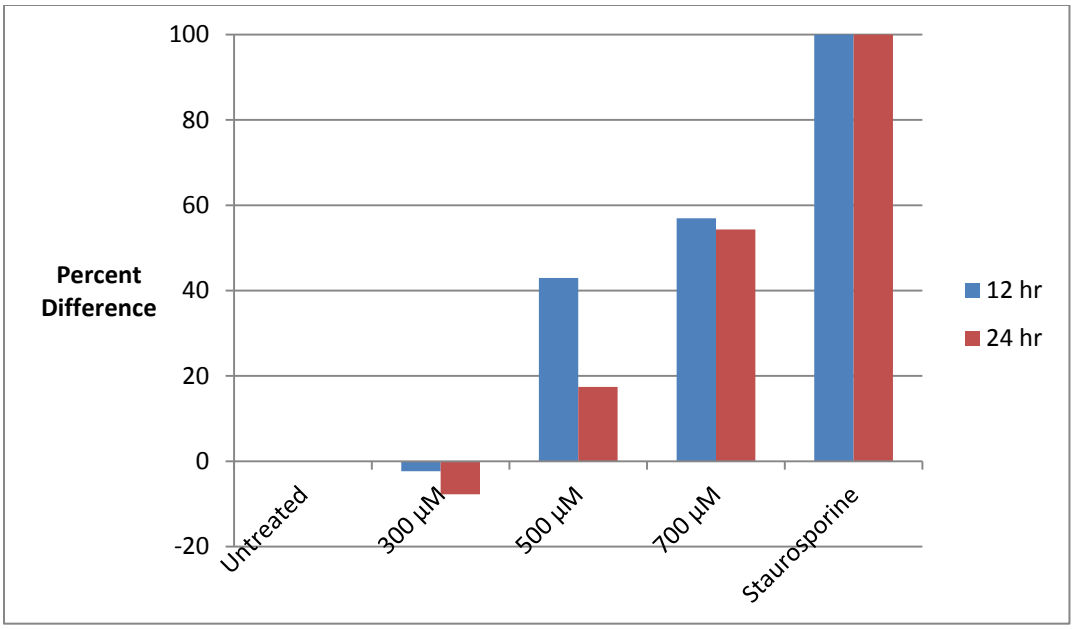


Figure 2. 33: Caspase 3/7 activity was observed for concentrations as low as 500 µm after 12 hours. At 24 hours, activity had rapidly decreased which suggested only initial small population of cells had undergone apoptosis.

Real Time-PCR data. Given a more robust response to hydrogen peroxide, a select number of target genes were tested by real time PCR. The majority of target genes did not show similar trends to microarray data (data not shown). Only general stress and HMOX1 mimicked microarray data (figures 2.34-2.39). Induction was seen to be dependent on concentration, as 700 µm of H202 reliably produced a pronounced, delayed pattern. Induction with 500 µm h202 appeared to most resemble microarray data for this limited set of genes. However, viability assays of cells done alongside of the microarray data had indicated nearly complete cell death. 500 µm H202 currently shows a minimal amount (figure 2.23).

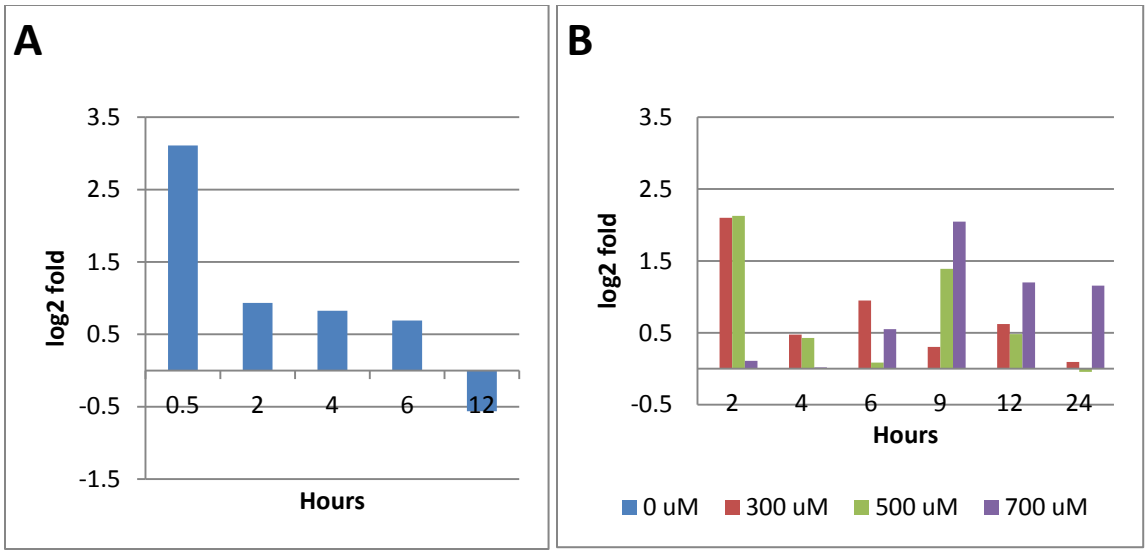


Figure 2. 34: IER2 is a generic stress response gene that only shows a short burst of transcription. (A) Microarray gene expression profile. (B) Real-Time PCR. 300 µm and 500 µm of hydrogen peroxide induce an early burst of transcription as seen in the array data. 700 µm induces a delayed response.

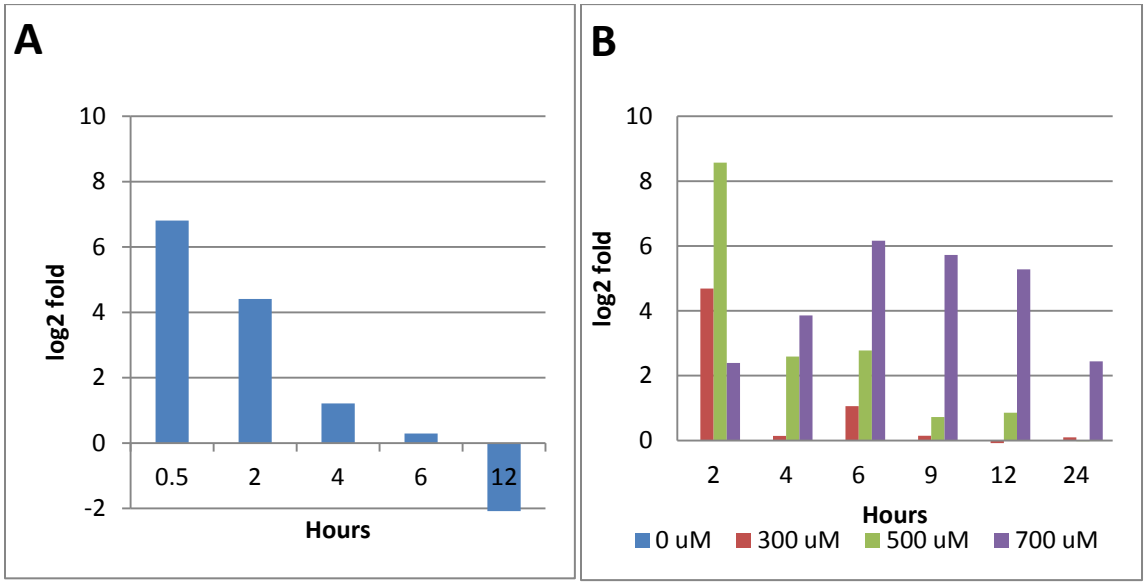


Figure 2. 35: FOS is an early immediate transcription factor that responds to a wide variety of stressors. (A) Microarray gene expression profile. (B) Real-Time PCR. Higher levels of H202 induce a significantly different expression profile.

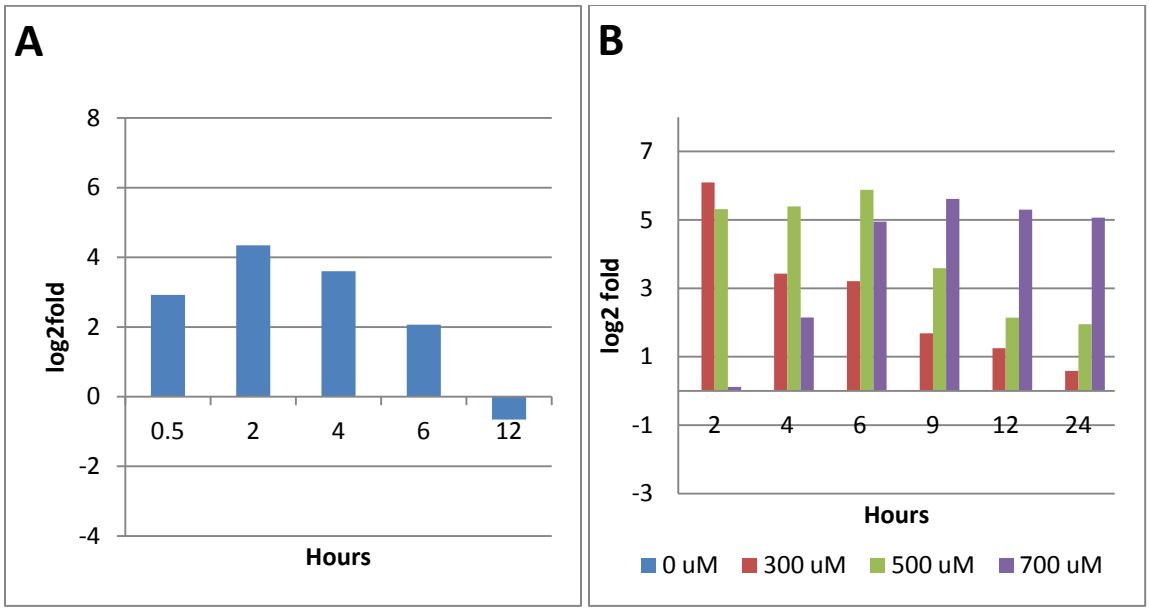


Figure 2. 36: ATF3 is an oxidative stress specific response gene. (A) Microarray gene expression profile. (B) Real-Time PCR

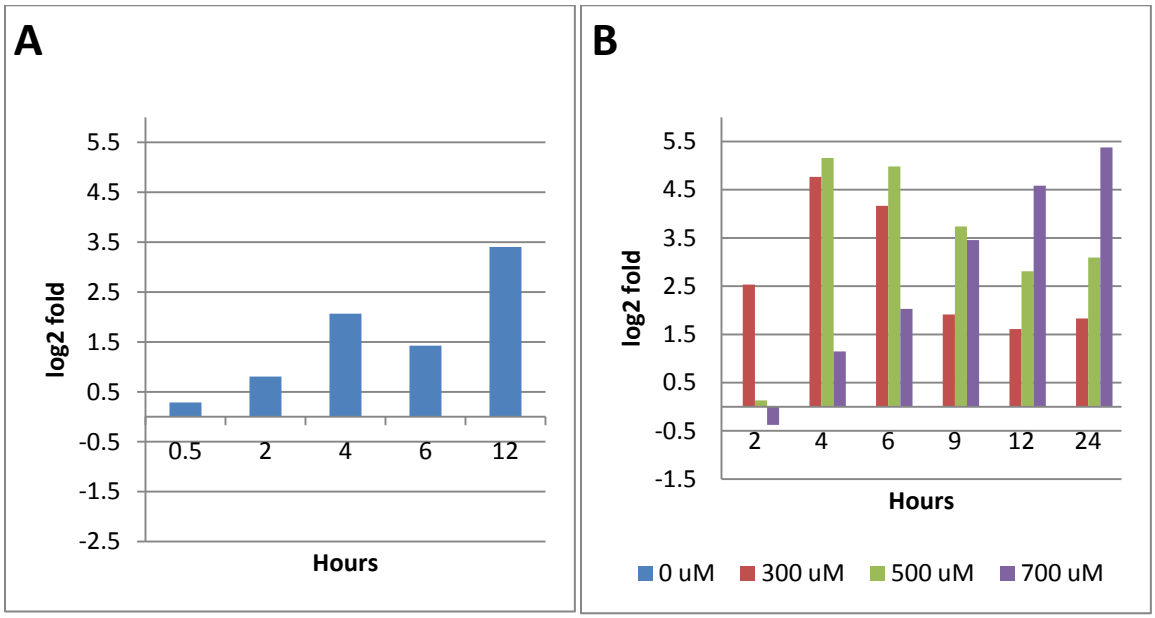


Figure 2. 37: HMOX1 is an antioxidant gene that degrades free heme. (A) Microarray gene expression profile. (B) Real-Time PCR. HMOX induction was greatly increased compared to microarray data.

Discussion

High throughput data generated in this instance allowed a global view of oxidative stress and reconstruction of the cellular network involved in cellular response to oxidative stress. Oxidative stress is well studied, but from a fragmented perspective that leads to difficult understanding the overall importance of a gene or pathway. Gene expression profiles had a large degree of concordance with published and expected trends. The data seems sensible as in response to oxidative stress, the cell increases ROS clearance mechanism and DNA repair pathways. Independently, apoptotic processes proceed resulting in the loss of mitochondria and cell death. Many known oxidative stress specific exceptions to the canonical activity or mechanism for some processes or pathways were recapitulated in the microarray data. As such, it would have been an ideal platform to perturb such actors to quantify the global effect they have on apoptosis due to oxidative stress.

However, the network failed verification, and a satisfactory explanation for the discrepancy was not obtained. Low throughput assays were contradictory and inconsistent from previous results, casting doubt onto the high throughput data. While verification of microarray trends by RT-PCR was in general discouraging, stress response and genes related to heme activity were consistent.

Heme. The possibility of heme as a small molecular regulator of oxidative stress is intriguing; as a bound partner of cytochrome C, heme plays a vital role in the metabolism and apoptosis. Free heme is a significant cyto-protective molecule. HMOX1 nonreversible degrades heme to bilirubin and free iron as

opposed to succinyl-CoA; the cell must divert additional energy from the citric acid cycle to synthesize new heme molecules. Iron must be sequestered by ferritin to avoid the oft repeated Fenton reaction. Excess intracellular iron has also been identified by Dixon et al 2012 as an important factor for ferroptosis: a non-necrotic, non-apoptotic form of cell death[119]. As ferroptosis is mediate by ROS, perhaps antioxidant response involving heme is playing a role, although the mechanism of cell death in ferroptosis is non-apoptotic and unrelated to the mitochondria.

Takahashi and Masudada 2009 have described a methodology to easily determine the abundance of free heme[120]. Cytosolic lysates in conjunction with a reconstituted apo- form of horseradish peroxidase reacts with common western reagent for sensitive chemiluminescent detection. Perturbations could be made not only against heme related proteins, but heme abundance could be tested against perturbations of pathways important to oxidative stress induced apoptosis.

Methods

Cell Culture. Human lung microvascular endothelial cells (HMVEC-L) cells from Lonza were cultured according to manufacturer instructions in EGM2 media bullet kits. Mycoplasma detection kit was also obtained from Lonza.

Cell Staining. Cells were stained with antibodies from Abcam according to manufacture instructions.

Inducers. Hydrogen peroxide was acquired from Fischer scientific. To obtain 100 uM of H202 in media, stock 8M solution was first diluted to 100 mM.

Spectrophotometry. H202 concentration was calculated using the published millimolar extinct coefficient for hydrogen peroxide on a Nanodrop 2000.

Hydrogen Peroxide Colormetric kit. To detect low H202 peroxide concentrations used for cell experiments, a kit was acquired from ENZO Life Sciences and used according to manufacture instructions.

LDH assay. Cytox-ONE LDH detection kit was acquired from Promega and used according to manufacturer's instructions.

Caspase Activity. Caspase 3/7-Glo kit was acquired from Promega and used according to manufacturer's instructions.

Total RNA Preparation. After induction, cells were harvested using Qiashredders. Total RNA was harvested using RNAeasy kits from Qiagen according to manufacturer's instructions and stored at -80C.

RT-PCR. Total RNA was subjected DNAse treatment by using Ambion DNAfree Turbo. Reverse transcription using Superscript VILO acquired from Invitrogen. Primers were added to 10 ng of cDNA per reaction and real time PCR perform on an ABI 4000 machine using powersybr mix. PCR primer sequences are listed in appendix Real Time PCR Primers.

Microarray Analysis. Total RNA was labeled and hybridized using Agilent two color Human Gene Expression V2.0 arrays. Treated samples were hybridized against time matched untreated.

Data analysis. Microarray data was preprocessed from Agilent spot and intensity calls. Data was normalized using LOESS, and median normalized across arrays. Normalized data was used with VAMPIRE[19] for significant gene detection.

Pathways. Initial pathways were derived from BioCarta. Pathways were drawn using Pathway editor[121]. Only genes that were called significantly significant in at least one time point were considered. Ratios were rescaled for comparative purposes.

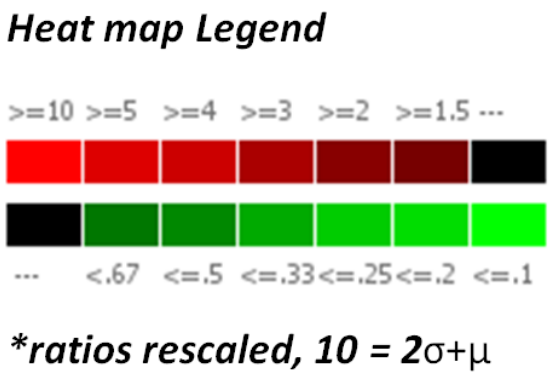


Figure 2. 38: Expression ratios on all pathways are rescaled for comparative and visualization proposes.

Acknowledgements

I would like to acknowledge Professor Ratnesh Lal for providing space and resources in his lab for the research involving oxidative stress. I would also like to acknowledge Srini Ramachandran, a postdoc in his lab for his time and effort in our collaboration. Chapter 2 uses with permission Srini Ramachandran's experimental results: figures 2.1, 2.2, 2.9, 2.10 and 2.20.

CHAPTER 3 LENTIVIRAL MEDIATED MOUSE CANCER MODELS

Introduction

Cancer is an aberrant cellular state in which individual cells cease cooperating for the benefit of a multi-cellular organism[2]. A conceptual model has been proposed and refined that defines significant characteristics or hallmarks of cancerous cells[122, 123]. While initially meant as a simplifying intellectual framework, these hallmarks have been expanded to include: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, reprogramming energy metabolism, and evading immune detection. In addition, two enabling characteristics are highlighted: genome instability and mutation, and reprogramming energy metabolism. Beyond the characterization of genetic mutations that lead to cancer, the intellectual goal of any cancer model would be to extend and illuminate the mechanisms by which a cancer achieves these hallmarks, or any other defined set of central features.

Practically, a more immediate milestone for cancer models is general acceptance of their applicability to human cancers; a goal that is based upon current knowledge and thus a continually moving target. The end result is a requirement for increasingly complex models that better mimic the progression and nuance of human cancers. This chapter focuses on an emerging aspect of characterization of cancer models by comparison to molecular signatures. Mouse

cancer models generated by lentiviral mediated approaches in combination with genetically engineering mice (GEM) are compared with previously defined molecular signatures for both human primary glioblastoma and GEM lung adenocarcinoma.

Oncogenes and Tumor Suppressors. Genes involved in tumorigenesis are mainly divided into two categories, oncogenes and tumor suppressors[124]. Oncogenes are genes in which over expression or constitutive activation promote tumor formation. Conversely, tumor suppressors are genes in which their reduced or lost activity no longer prevents tumorigenesis. In both cases, point mutations within genes can lead to changes in activity, or larger scale chromosomal defects leading to duplications or deletions of the entire gene can occur. An additional category of genes, caretaker genes[124], affect the global mutation rate of the genome and are often associated with DNA repair pathways. Caretaker genes do not actively prevent or promote tumorigenesis, but are responsible to prevent or correct the mutation of all genes, and most importantly in this instance oncogenes or tumor suppressors. This is an important function because as a general rule tumorigenesis requires mutations in multiple genes[124]. Both cancer models studied in this chapter involve expression of a mutated RAS oncogene in combination with loss of the tumor suppressor TP53.

RAS. RAS is a family of proteins involved in the signal transduction between cell surface growth receptors and downstream effectors pathways[125]. RAS proteins revolve between a guanosine triphosphate (GTP) -bound active state, and guanosine diphosphate (GDP) -bound inactive state. GTPase-

activating proteins (GAPs) regulate RAS inactivation by GTP hydrolysis. Inactivation of GAPs activities are accomplished by somatic mutations to specific RAS residues[126], particularly Q61, G12, and G13. The mutated RAS proteins remain constitutively active and lead to sustained induction of downstream transcription factors associated with cell growth or survival: FOS, SRF, JUN, EK1, ATF2, and NF-kappaB . While mutations appear to be interchangeable, individual RAS oncogenes have displayed tissue specificity[125]. KRAS is frequently mutated in colorectal tumors, lung carcinomas (non-small-cell lung cancer), and in pancreatic carcinomas. Mutated HRAS tumors are found in the skin, head and neck. NRAS mutations are typically observed in hematopoietic malignancies.

RAS oncogenes induce hyper proliferation and enhanced survival, but at the cost of replicative stress which leads to DNA damage and activation of DNA damage response. As such tumorigenesis requires additional mutations to escape either senescence or apoptosis. A critical nexus is CCND1 (Cyclin D1), which acts in concert with CDK4 and CDK6 to override RB mediated cell cycle arrest[127]. Important upstream mediators of CCND1 activated by oncogenic RAS are the PI3K and RAF pathways. In addition to regulation of CCND1, both pathways increase cell survival[125]. PI3K down-regulates prop-apoptotic proteins such as BAK1, and up-regulates anti-apoptotic proteins through NF-kappaB. RAF suppresses apoptosis by down-regulation of PAWR (PAR4), and up-regulation of BCL2. RAF and PI3K mediate BAD phosphorlyation leading to an inactive complex with 14-3-3. Oncogenic RAS induces sustained proliferation by

activation of growth factors such as VEGFA, FGF2, and PDGF and down regulating anti-angiogenesis factors THBS1 and THBS2. RAS also plays a role in remodeling of the tumor micro environment and tumor metastasis. HRAS is known to up-regulate MMP2, MMP9, and PLAU which are important enzymes in the removal of the neighboring extracellular matrix. RAS up-regulates SNAI1 (SNAIL) and SNAI2 (SLUG) which degrade E-cadherin and allow cell mobility. The oncogenic effects of RAS manifest in a context-dependent manner with sub-cellular, cellular and tissue environments determining its functional output.

P53. TP53 (p53) is a shorted-lived transcription factors that regulates cellular tumor suppression[128]. An important step of the pathway is the release of TP53 from the negative regulators MDM2 and MDM4 which leads to accumulation of stable TP53. TP53 induces a wide range of genes, leading to DNA damage repair, growth arrest, or apoptosis[129]. Loss or mutation of TP53 predisposes the cell to a range of spontaneous and induced tumors; TP53 is disabled during the pathogenesis of most human cancers. However, TP53 does not influence the rate of tumor initiation or mutation, but prevents malignant progression of tumor cells. Further, restoration of TP53 expression can promote tumor regression and clearance in vivo[130-132]. Initiation of apoptosis by TP53 depends on the type and intensity of stress, cell type, and genetic background along with other pathways such as RB. The common principle is protection by maintaining integrity of the cell and its genome or preventing proliferation of nascent cancer cells. DNA damage response pathways potentially activate TP53, but classical induction is through the ARF tumor suppression pathway.

Mouse Models of Cancers. The creation of transgenic mice has allowed the study of inheritable "traits" that lead to the formation of tumors. According to a perspective by Hanahan et al. 2007,

> Before these developments, cancer was largely modeled by tissue culture of cell lines established from human and animal tumors, and by inoculation (transplantation) of such cell lines under the skin of immunodeficient mice, where lump-like solid tumors would form. While of clear utility in studying parameters of tumor growth, such models did not necessarily recapitulate the subtleties observed in human tumors arising in different organs, in terms of polymorphic genetic susceptibility, histological characteristics, and progression from benign premalignant lesions to tumors of increasing aggressiveness[133].

Generation of a tumor is therefore insufficient, i.e. a tumor in the brain is not in and by itself necessarily a useful model of glioblastoma. As expressed by Dyke and Jacks (2002),

> ... there is a common (and not unreasonable) expectation that these mouse models will "model" human cancer; that is, cancer in the mouse should look and act like that disease. Mouse tumors should have the same or similar histological features of comparable human tumors; they should progress through the same stages and cause the same physiological and systemic effects on the host; the same genes and/or pathways should be affected in tumor initiation and progression; the response of a given tumor to a particular therapy in the mouse should accurately reflect the response in human patients; and the results from preclinical testing of experimental therapies in mouse models should ultimately predict the efficacy of such therapies in clinical trials in humans[134].

While there are no concrete guidelines, the general criteria for acceptability of cancer models to human cancers can be broadly considered as expressing or acquiring a corresponding set of genetic mutations, displaying a similar

phenotype, showing an increase in tumor progression from the cell of origin or tumor initiating cell, and responding in a therapeutically similar manner.

CRE-LOX. Special consideration has been given to the controlled expression of oncogenes or tumor suppressors. Earlier models had sustained increased or reduced expression of the gene of interest, while latter models have increasing specificity with regard to tissue specificity or with respect to time. A powerful tool for controlled expression is the CRE-LOX system[135]. Cre is a P1 phage derived recombinase that act by restricting and ligating adjacent *loxP* sites. Efficient excision of specific genomic sequences is possible as *loxP* sites are not normally found in the mouse genome. Strategic placement of *loxP* site enables generation of conditional knockouts, knockins, and other variants. Expression of the Cre enzyme can be controlled via tissue specific promoters or drug inducible elements such as tetracycline[136]. To date, many hundreds of cre-transgenic mice have been generated[137]. The rate limiting step in generating mouse cancer models is breeding of mice with appropriate *loxP* flanked constructs to study a gene of interest.

Tumor Progression. Many cancers require several mutational events to proceed from benign to malignant tumors. The classical model for such tumor progression is colorectal cancer[138] as it arises over many decades from successive acquired mutations. However, the cancer initiating cell may not be the cell stem cell which propagates tumor growth[139]. For cancers that exhibit tumor heterogeneity, differing cell of origins could be an alternative source of cancer subtypes as opposed to different sets of oncogenic mutations. Lineage tracing[140]

is an important aspect to identify the cancer initiating cell, however cell-lineage specific promoters and markers are not commonly available for all tissues and organs.

Lentiviral Mediated Models. The lentivirus is a engineered form of HIV that is capable integrating a delivered sequence into the genome of almost all cells, including non-dividing ones[141, 142]. Originally a major tool of gene therapy[143], significant issues have arisen in this application such as those highlighted by the X-SCID trials[144, 145]. In contrast with gene therapy, viral vectors can be used to deliberately induce tumorigenesis. Use of viruses to generate cancers is hardly novel[133], but more sophisticated viral vectors coupled with Cre-*Loxp* technology has allowed generation of more sophisticated, tissue specific models[146]. The main advantages of lentivirus according to Xia, Y. et al. 2011 are:

> (1) lentiviruses infect almost any type of cell, and transgene expression can be controlled by a tissue-specific promoter, which enables a more precise tracing of the origin of the cancer cell; (2) lentiviruses integrate into genomic DNA so that it is possible to stably deliver oncogenes and short hairpin RNAs (shRNAs) against tumour suppressors, and bypass the requirement of numerous conventional genetic crossings; (3) viral titres can be controlled to infect only a few cells, to more faithfully recapitulate human cancer initiation[147].

The issues with generating cancer models using a lentiviral system are the selection of appropriate oncogenes or tumor suppressors and the selection of a tissue specific promoter that is not too "leaky." The chosen promoter may in reality be expressed at low levels in other cell types or tissues. These issues may not be easily solvable as the mutations required to initiate a tumor may differ

from the mutations acquired to sustain one. Closely related cell types within a tissue may also share common regulatory transcription factors, or cell type specific promoters may not be known. As such it is necessary to prove that any tumors generated are in fact generally applicable to human cancers as opposed to a random cancerous cell.

Biomarkers and Molecular Signatures. Classification of human tumors in a clinical setting is predominately done by pathohistological and morphological characteristics. High throughput image analysis of tumors is not yet common[148, 149], but the study of somatic mutations[150] and associated gene expression patterns and their relation to clinical outcomes is a well worn path[151-154]. Many cancers exhibit gene expression patterns that naturally cluster into multiple subtypes for a given cancer; however, when comparing mutations, often only a fraction of a subclass will exhibit the same sets. This suggests that differing sets of mutations can lead to the same subclass or phenotype. When a pattern has been reduced to a smaller set of genes with the ability to classify, it is often referred to as a molecular signature. Molecular signatures differ from traditional markers in that the signature as a whole will be enriched while individual genes will not be statistically significant across all samples of subtype. Given the increasing availability, breadth, and resolution of high throughput data for primary human cancers[150, 155-159], classification and prediction of clinical outcomes by molecular signatures will like intensify.

KRASLA2 Lung Adenocarcinoma Model

KRAS is frequently mutated in lung adenocarcinomas[125]; nearly 60% of lung adenocarcinomas have mutually exclusive mutations in either KRAS or EGFR[157]. A key disadvantage of transgenic KRAS genetically engineered mice is that they express the oncogene in all cells of a tissue type. A latent, oncogenic KRAS model has been developed that employs "hit-and-run" strategy in that one (KRASLA1) or both alleles (KRASLA2) are capable of spontaneous activation[160]. All KRASLA2 mice were observed to develop lung carcinomas, which followed normal carcinoma progression similar to non-small cell lung cancer. NF-kappaB was implicated in this model to play a significant role in tumor progression[161, 162], especially upon p53 restoration[162].

Lentiviral constructs containing CA2 (carbonic anhydrase 2) promoter driven CRE and shTP53 specifically transduce alveolar epithelial cells[147]. KRASLA2 mice were infected intratracheally and were crossed with floxed IKBKB (IKK2) mice to study the loss of NF-kB activity. Phenotypically, IKK2 knockout (KO) mice early on exhibited reduced tumor burden, but end point tumor load was similar to IKK2 wild type (WT). To understand which significant pathways were differentially expressed, microarray analysis was performed on primary mouse tumors and cell lines derived from these tumors.

Molecular Signatures. KRASLA2 mice have been previously compared to primary human lung adenocarcinomas and other mouse models using gene expression arrays. Sweet-Cordero et al. 2005 developed a strategy to identify sequentially smaller gene sets[163] that contained: genes from tumors that were

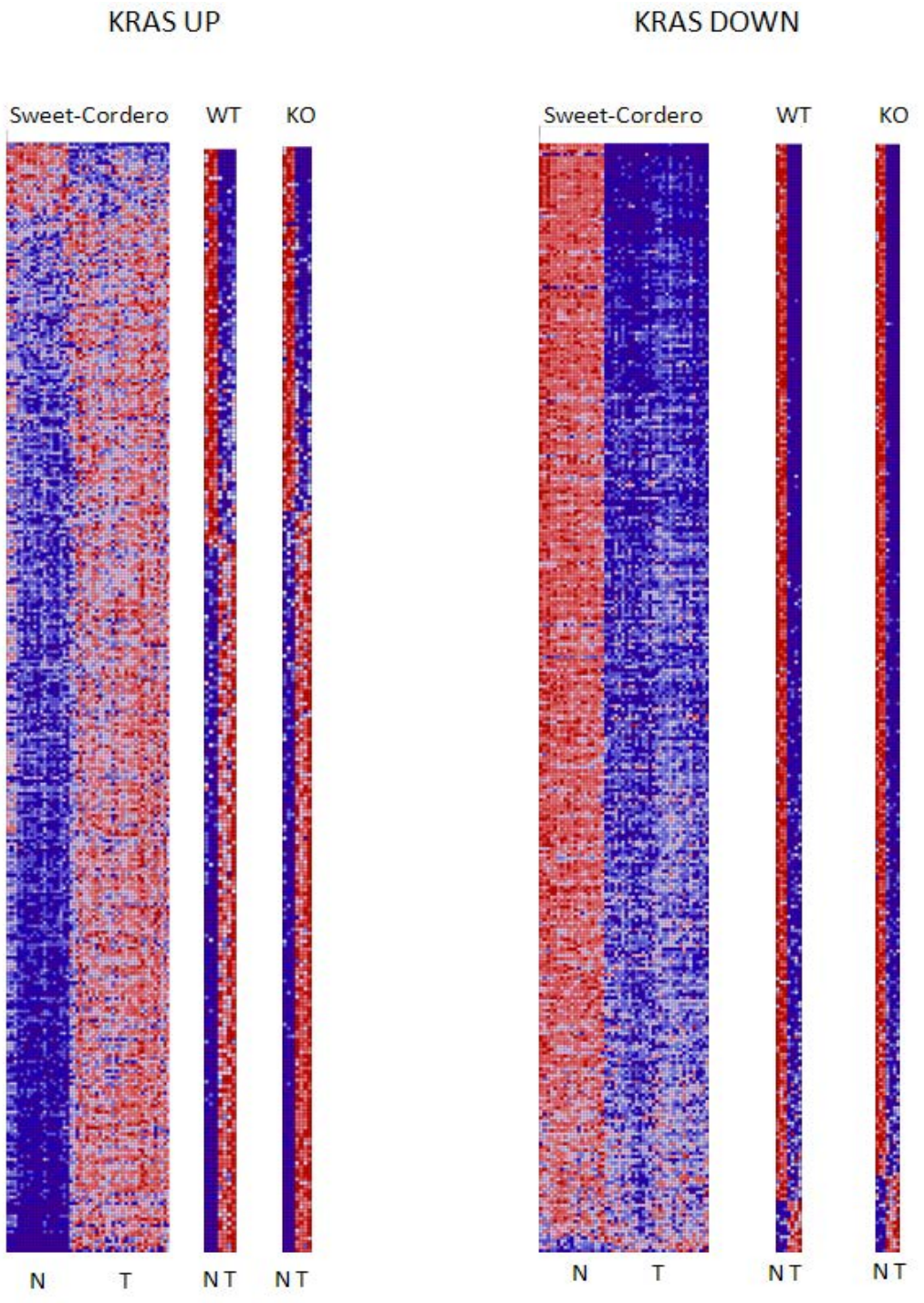


Figure 3. 1: Lentiviral mediated KRASLA2, IKK2 wildtype (WT) and IKK2 knockout (KO) tumors display similar KRAS expression signatures as genetically engineered mouse tumors (Sweet-Cordero). Absolute expression of Normal (N) and Tumor (T) samples, where UP markers were found to be significantly expressed in tumors while DOWN markers were higher in normal tissue[163]. Red indicates higher expression levels than average, blue below average, and values close to the average are in white.

significant up regulated when compared to normal lung, KRAS up signature; genes that were significantly down regulated, KRAS down signature; up regulated genes that were in common with human adenocarcinomas, adenocarcinoma signature; and common genes up regulated in pancreatic adenocarcinomas, KRAS signature (see Appendix Sweet-Cordero for gene sets). To validate the applicability of the lentiviral mediated lung cancer model, we applied these molecular signatures to tumor data compared against normal lung tissue from the Affymetrix test platform data set (Figure 3.1). Tumors showed statistically significant enrichment for all signatures using GSEA (table 3.1). On this basis, the lung adenocarincomas mediated by lentivirus are qualitatively identical on a gene expression level to tumors arising solely from genetically engineered mice.

Table 3. 1: KRAS signatures, as defined by Sweet-Cordero, are enriched in lentiviral mediated tumors as called by GSEA. Normalized Enrichment Scores (NES) are provided; higher positive numbers indicate enrichment of gene set in normal lung tissue while negative numbers indicate enrichment in tumor. All signatures pass a FDR < 10% cutoff.

Signature	Sweet-Cordero	IKK2 WT	IKK2 KO
KRAS up	-5.4	-2.8	-3.0
KRAS down	4.4	4.7	4.7
Adenocarcinoma	-4.6	-2.5	-2.6
KRAS	-4.0	-2.0	-2.0

For comparison, KRASLA2 signatures were applied to array data from putative glioblastomas generated from mutated HRASV12, shTP53 mouse tumors[164, 165]. Tumors showed significant enrichment for all signatures when glioblastomas were compared to brain, but not in comparison to lung (Table 3.2). This result calls into question the specificity of the molecular signatures to lung

adenocarcinomas as was originally claimed[163]. At the same time, the general approach is likely still valid as these glioblastomas have a RAS family mutation that likely leads to the same gene expression patterns as the KRAS lung adenocarcinomas, and signatures were only significant when compared to the correct tissue of origin.

Table 3. 2: Glioblastomas exhibit enrichment of lung adenocarcinoma specific KRAS signatures when compared to normal brain. Negative values are enriched in tumor over normal tissue. Loss of proper directionality for up-regulated signatures when compared to normal lung. Values with a FDR < 10% are in bold.

Signature	Brain	Lung
KRAS up	**-3.2**	**1.8**
KRAS down	**1.5**	**3.55**
Adenocarcinoma	**-3.8**	**1.5**
KRAS	**-2.9**	1.5

Effect of NF-kappaB. To test the effect of NF-kappaB, a gene set of NF-kappaB targets from nf-kb.org was derived[59] (see appendix NF-kappaB targets). Using GSEA, NF-kappaB target genes were consistently enriched in normal lung as opposed to tumors (Table 3.3). No difference in enrichment was observed in IKK2 KO versus IKK2 WT tumors, but IKK2 WT cell lines derived from primary tumors showed enrichment of NF-kappaB targets over IKK2 KO cells.

Table 3. 3: NF-kappaB target genes are enriched in normal tissue when compared to IKK2 WT and KO tumors. No statistically significant difference in expression between WT and KO primary tumors, but enrichment of NF-kappaB targets in IKK2 WT derived cell lines compared to IKK2 KO cell lines. Positives vales are enriched in condition 2 over condition 1. Values with a FDR < 10% are in bold.

Condition 1	Condition 2	NF-kappaB
Mice IKK2 WT	Normal Lung	**2.7**
Mice IKK2 KO	Normal Lung	**2.7**
Mice IKK2 KO	Mice IKK2 WT	1.1
Cell IKK2 KO	Cell IKK2 WT	**2.4**

GSEA. Automated functional analysis by GSEA using gene sets derived from gene ontology and pathway databases revealed every significant term as enriched in lung over mouse tumors (see Appendix Lung Tumors GSEA). When comparing IKK2 KO and IKK2 WT primary mouse tumors and cell lines, the majority of statistically significant terms are cell cycle related (Table 3.4; see

Table 3. 4: Common functional annotations enriched in both mouse tumors and derived cell lines when comparing IKK2 wild type to IKK2 knockout. Majority of terms are cell cycle related.

Term	Mouse	Cell
MITOTIC_CELL_CYCLE	3.7	1.7
CELL_CYCLE_PROCESS	3.6	1.8
REACTOME_CELL_CYCLE_MITOTIC	3.6	2.7
CELL_CYCLE_PHASE	3.5	1.8
M_PHASE_OF_MITOTIC_CELL_CYCLE	3.3	2.1
M_PHASE	3.3	2.1
MITOSIS	3.3	2
REACTOME_MITOTIC_PROMETAPHASE	3.2	2.6
KEGG_CELL_CYCLE	3	1.7
REACTOME_G1_S_TRANSITION	2.8	2.5
DNA_REPLICATION	2.6	2.3
RESPONSE_TO_DNA_DAMAGE_STIMULUS	2.4	1.8
DNA_REPAIR	2	1.8
KEGG_PYRIMIDINE_METABOLISM	1.9	1.9
RESPONSE_TO_ENDOGENOUS_STIMULUS	1.8	1.8
CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	1.7	1.7

Appendix IKK2 KO/WT GSEA). This is in agreement with other experimental observations that NF-kappaB is not acting through anti-apoptotic pathways, but encouraging cell proliferation by sustained activation of ERK[147].

Glioblastoma Multiforme Model

Glioblastoma multiforme (GBM) is a highly aggressive and lethal intracranial brain cancer[166, 167]. GBMs display a wide degree of heterogeneity in terms of pathology, genomic mutations, and gene expression. Genetically engineered mice are potentially useful tools to identify the cell of origin[168], the potential order and combination of mutations, and as a therapeutic test bed for novel treatments. GBMs have been subjected to intensive query by a variety of high throughput technologies: copy number alterations[169], somatic mutations[158], and transcriptional analysis[155, 158, 170]. Mutations were observed to overwhelmingly occur to members of the TP53, RB, and RTK/RAS/PI3K pathways[158, 167].

Lentiviral Mediated Mouse Model. Marumoto et al 2009 developed a "proof of principle" lentiviral mediated model of glioma[164]. Within the lentiviral construct, red fluorescent protein (RFP) is floxed and placed between a CMV promoter and tumorigenic payload. When injected into a transgenic CRE mouse and transduced into a CRE expressing cell, RFP is excised and the CMV promoter is then able to drive oncogene expression. For this study, the tumorigenic payload is HRASV12, a constitutively active mutant of HRAS, and a

small hairpin targeting TP53. GFAP-, SYN-, and NES- CRE transgenic mice were used to target astrocytes, neurons, and neuronal stem cells respectively. Constructs were injected into: the cortex of SYN-CRE mice (SYN); the hippocampus of NES-CRE mice (NES); and the cortex (CTX), hippocampus (HP), and subventricular zone (SVZ) of GFAP-CRE mice. Along with tumors that arose from these injections, normal cortex (NCTX) and hippocampus (NHP) tissue were also harvested and microarray analysis was performed.

Clustering. Hierarchical clustering of highly variable genes revealed two completely opposite clusters (Figure 3.2), excluding genes on the X and Y chromosome (data not shown). Tumors from the same injection site and construct do not reliably cluster together. In contrast, normal tissues exhibited very tight correlations, even between hippocampus (NHP) and cortex (NCTX) samples. For this reason, the lack of cohesion between promoters and injection sites is suggestive of tumor heterogeneity rather than noisy data

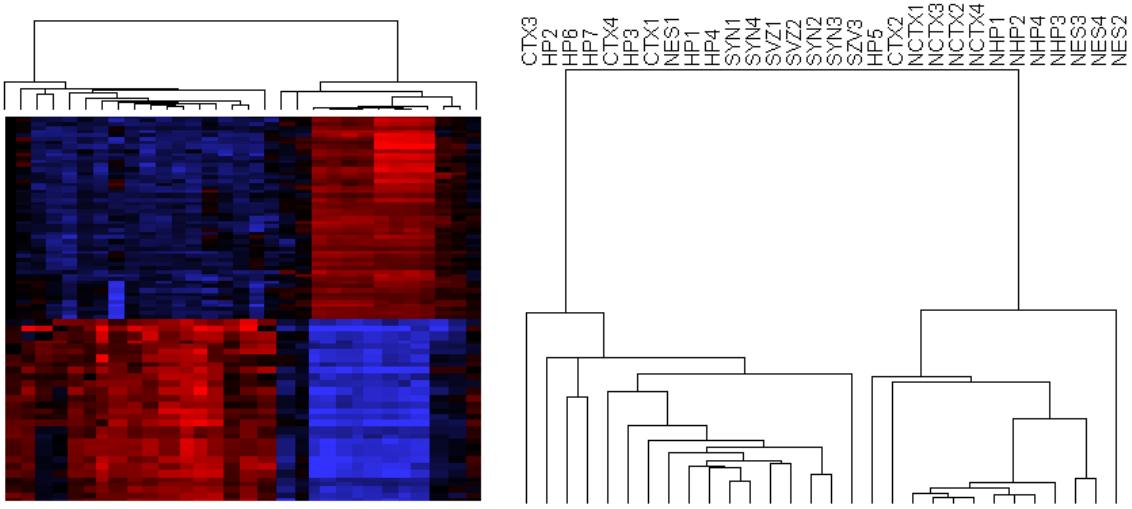


Figure 3. 2: Dendrogram of tumors reveal two clusters that exhibit opposite patterns of expression. Samples are joined based upon the correlation coefficient, with higher correlation on the bottom. Tumors do not cleanly cluster based on the CRE promoter or the injection site.

Verhaak Molecular Signatures. The Cancer Gene Atlas (TCGA) surveyed 200 glioblastomas and identified four major clusters, or molecular subtypes, based on gene expression: Classical (CL), Mesenchymal (MES), Neural (NL), and Proneural (NL)[159]. Verhaak et al 2010 also introduced a computational method for calculating GSEA for single samples (SSGSEA). Signatures for the four subtypes using a modified SSGSEA procedure were scored (see Methods and Appendix Verhaak for details). Corresponding to clustering results of our data, tumors aligned into two groups (see figures 3.3 and 3.4). One group of arrays, including normal tissue samples, scored highly for the proneural and neural signatures. The second group scored highly for mesenchymal signatures with weak classical scores. Others have shown results in GEM for the proneural class[168, 171], but to our knowledge the mesenchymal or neural subtypes have not been previously generated in mice.

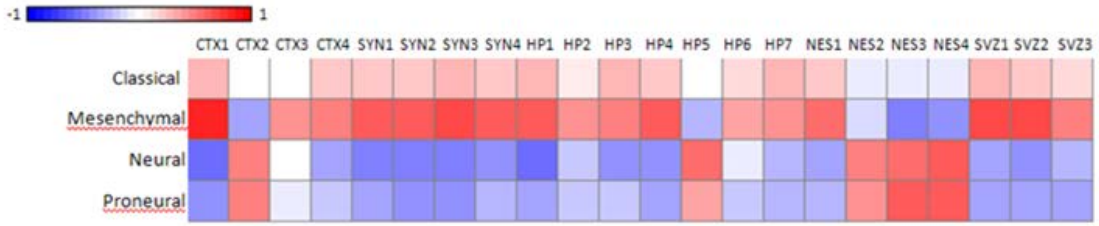


Figure 3. 3: Tumors score highly for mesenchymal signatures or present double neural/proneural signatures. Tumors do not score similarly based upon their promoter or injection site, but follow closely with clusters in figure 3.2.

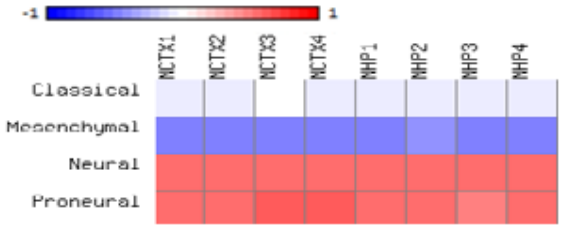


Figure 3. 4: Normal tissues present a double neural, proneural signature.

Normal samples in the TCGA data set were mentioned to have high neural scores according to Verhaak et al (2010)[159]. When samples were scored according to our modified procedure, a double signature similar to our mouse normal samples (NCTX, NHP) was observed. In addition, normal samples in TCGA on average have higher neural subtype scores than neural tumors. However, TCGA Neural subtype tumors on average only score highly for the neural signature (see figure 3.5). Based only on molecular signatures, mouse tumors appear to belong to either mesenchymal or neural subtypes. It is not

definitive, however, that the neural tumors are not proneural based upon molecular signatures alone.

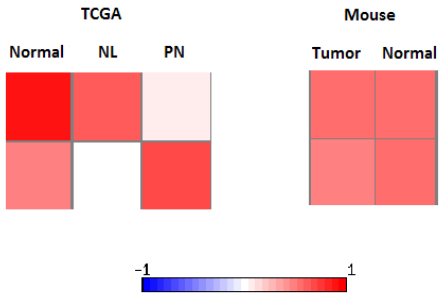


Figure 3. 5: Comparison of TCGA normal samples, neural tumors, and proneural tumors to mouse neural/proneural tumors. For TCGA, normal samples score highest in neural, but also present a proneural signature which is similar to mouse normal samples.

Phillips Signature. Phillips et al. 2006 defined three subtypes based on genes in their data set that positively or negatively correlated with survival[170]. It should be noted that these subtypes do not perfectly match Verhaak subtypes. The overlap between the two classifications using TCGA data is: Verhaak proneural is largely Phillips proneural with double proliferative signatures; Verhaak neural is split between Phillips proneural and proliferative; Verhaak classical is predominately mesenchymal with some proliferative; and Verhaak mesenchymal is predominately Phillips mesenchymal (data not show). Clustering on the 35 Phillips signatures genes (figure 3.6 and Appendix Phillips) divides a distinctly proneural subtype and a cluster with mixed high mesenchymal and proliferative signatures.

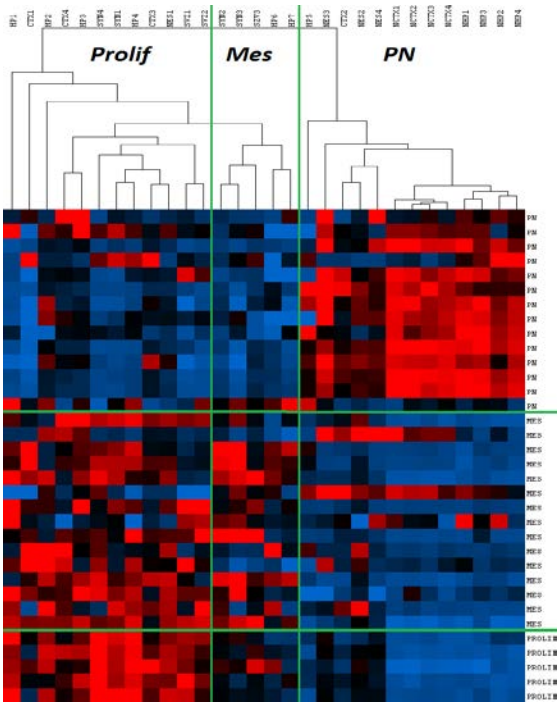


Figure 3. 6: Samples hierarchically clustered based upon 35 key signature genes from Phillips et al. 2006

Using full gene sets from Phillips for the three molecular subtypes, samples were also scored using SSGSEA (see Appendix Phillips). High scores for all three subtypes were present in the data (Figure 3.7), although there was not complete agreement between SSGSEA and clustering results. It may be likely that the tumors belonging to the proliferative and mesenchymal clusters are in fact one subtype, as they both contain highly expressed mesenchymal signature genes. As Phillips et al. 2006 (and Verhaak et al. 2010) rely on various forms of z-scaling; it probably that the lack of a true proliferative or mesenchymal tumor is skewing results. Again, tumors do not classify solely based on promoter or injection site.

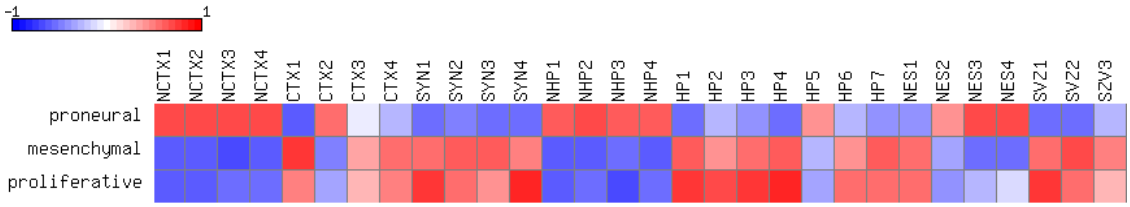


Figure 3. 7: SSGSEA scores for the full gene sets of molecular subtype signatures from Phillips et al 2006.

Cell of Origin. A significant area of interest in the study of glioblastoma is the cell of origin[167]. Verhaak et al. (2010) and others applied cell type specific signatures for the brain[172] to glioblastomas[159, 168, 171]. For comparative purposes, tumors were scored using molecular signatures from Lei et al (2011, but derived from Cahoy et al. (2008) which contained signatures for oligodendrocyte precursor cells (OPC), oligodendrocytes, astrocytes, neurons, and cultured astroglia (see Appendix Cahoy). SSGSEA results showed little correlation with the expected cell of origin (figure 3.8); GFPA- and SYN- Cre are not astrocytic or neuronal and instead predominately cultured astroglia. Obviously, there are no cultured astroglia in the brain. Nes-CRE tumors are predominately score highly for all three of the differentiated cell signatures along with OPC. These results are in sharp contrast compared to normal samples, which display extremely strong neuronal cell signatures (figure 3.9). Given these results, cell type specific signatures of terminal tumors do not necessarily indicate the cell of origin.

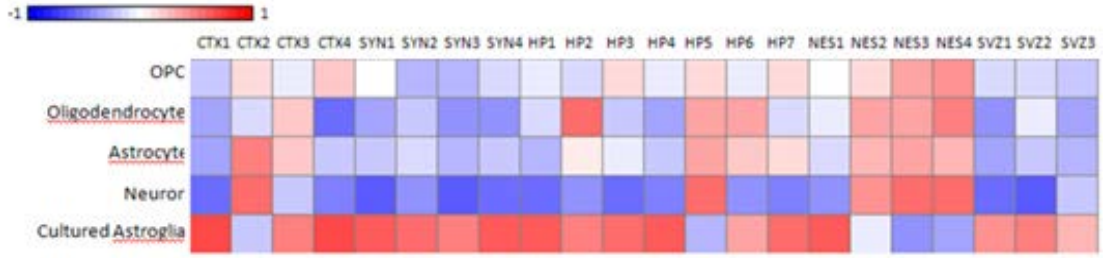


Figure 3. 8: Cell type specific signatures as calculated from SSGSEA do not correspond to the expected cell of origins.

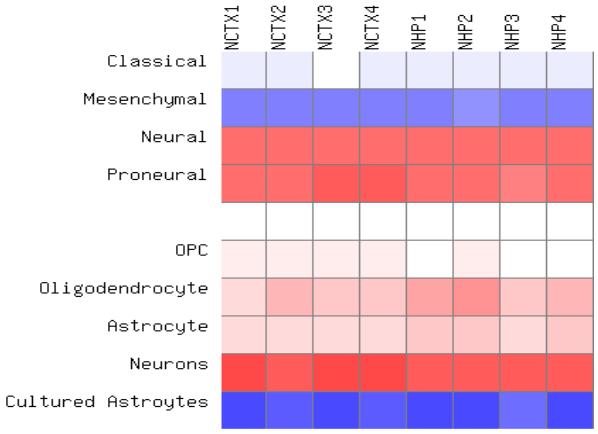


Figure 3. 9: Cell type specific signatures for normal samples display only a strong positive signature for neurons.

Combined Signatures. While cell type specific signatures do not match the expected cell of origin, they do correlate well with the molecular subclasses[159, 168, 171] (figure 3.10). Signatures for the average of all tumors in a subclass were calculated from the TCGA data set (figure 3.11). As the cell type specific signatures in human samples were very weak, reanalysis of those signatures for use in humans might be recommended. In general, CL tumors exhibit OPC, astrocytic and cultured astroglia cell type specific signatures. Mesenchymal solely score high for cultured astroglia. Neural subtype exhibits high oligodendrocytic, astrocytic, and neuronal signatures. Finally, proneural

tumors have high OPC and oligodendrocytic signatures. These average signatures were then correlated with scores from the signatures for each tumor (figure 3.12). Correlation coefficients were very strong for the final subtype, typically R2 = 0.77.

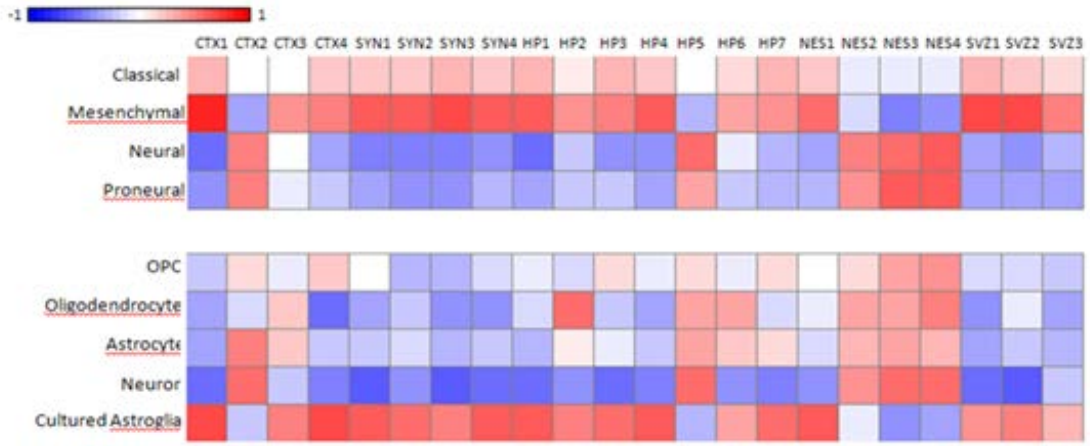


Figure 3. 10: Cell type specific signatures follow closely with molecular subtype. Mesenchymal tumors display a strong signature cultured astroglia. Other tumors typically display signatures for the remaining cell types.

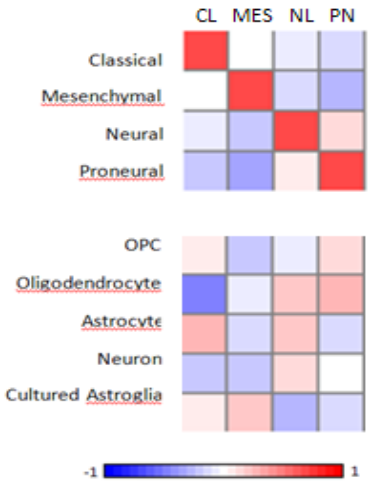


Figure 3. 11: Signatures for the average of all tumors in a subtype were calculated using TCGA data. While the cell type specific signatures are much weaker than compared to mouse, they still show preferences to particular subtypes.

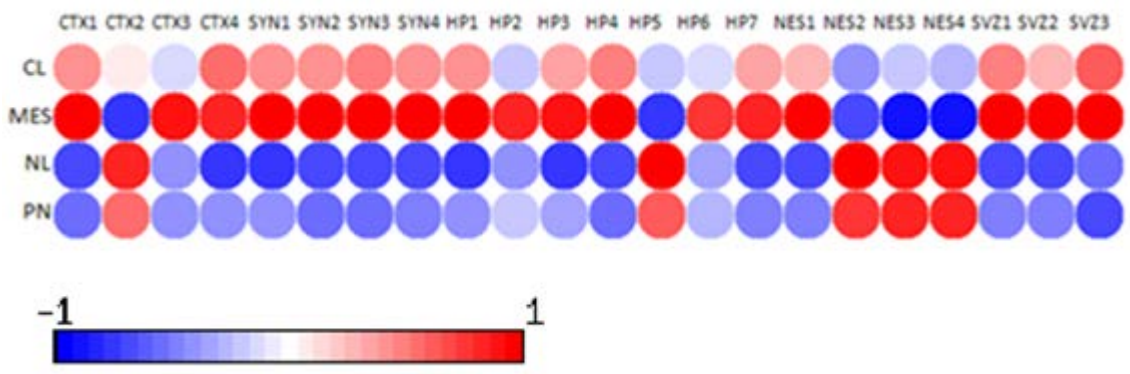


Figure 3. 12: Correlation coefficients for each tumor compared to the signatures of the TCGA subtypes. The highest correlation coefficient for each tumor was on average 0.77, which indicates a high degree of correlation.

Discussion

As with any set of tools, progress can be considered in terms of invention or innovation. The technological and computational methods utilized in this chapter are not unique, novel nor inventive or innovative. The original and novel aspect of this work is the use of lentiviral vectors to express oncogenes or to knockdown tumor repressors to create mouse cancer gene models in combination with genetically engineered mice. According to results presented here, mouse cancer models generated by lentiviral constructs have significant molecular signatures identifiable to both a mouse genetically engineered lung cancer model, and primary human glioblastomas. Importantly, the computational methods utilized are nearly identical to the original characterizations and incorporate molecular signatures as correspondingly defined. This is not to say that existing methods are perfect, only it is more intellectually direct and less

distracting to prove applicability by using previous methodology. As high throughput data sets for primary cancers increase in scope and abundance, the demand for characterization via molecular signature is rapidly becoming mandatory.

Molecular signatures are of great utility as a bioinformatics approach to proving applicability of mouse cancer models. Primary tumors can exhibit heterogeneity in terms of mutations and chromosomal defects. However, different mutations may lead to a common set of repeating patterns of expression across multiple tumors. Given the lack of defined criteria, proving the applicability of a cancer model is a difficult endeavor, and molecular characterization is a powerful tool to prove the internal mechanisms may be similar. For the lung cancer mouse model, both the IKKWT and IKKKO tumors exhibited substantial enrichment of both KRAS and adenocarcinoma gene sets. For the glioblastoma mouse models, brain tumors matched substantially to some of the subtypes defined by Verhaak et al. 2010 and all subtypes defined by Phillips et al. 2006.

From a clinical standpoint, segregation of tumors into subtypes highly is advantageous if those subtypes could direct diagnosis and treatment for better outcomes[167]. For glioblastoma, multiple subtype classifications have been presented based on pathology, critical pathways, molecular signatures, and image analysis. Verhaak et al. 2010 started from high variable genes and ended with subtypes that displayed clinical significance[159]. Phillips et al. 2006 clustered based on genes correlated with survival[170]. Disappointedly, these classification

schemes appear to share significant overlap between them or to known pathways that are highly mutated[158], yet they all claim clinical significance.

Given the known and marked heterogeneity of glioblastomas[166], it should not have been surprisingly that tumors display a wide degree of transcriptional heterogeneity. In fact, the TCGA data set yields 70% of all genes as differentially expressed (data not shown). This in and by itself does not pose a drastic difficulty as long as it is not a high indication of noise. Rather, the largest challenge to application of molecular signatures is the lack of orthogonality between signatures[173], meaning the signatures are not independent of each other. For a thorough analysis of these issues please see Marko et al 2011.

Lack of orthogonality is further impacted by the use of rescaling procedures required in Phillips et al. 2006 and Verhaak et al. 2010. If samples do not sufficiently cover the tumor space, rescaled estimates could falsely score high signatures if there are not enough extreme values in the data set. Use of a standard or uniform baseline is desirable, but it is not surprising that matching normal brain tissue was not collected as this would likely be harmful to the donor. A major advantage of mouse models is the ability to compare against genetically identical normal tissue, and molecular signatures derived from such comparison are more likely to be stable. However, this is not necessarily the best solution, as the sensitivity of the brain prevented the harvesting of matching human tissue to begin with.

As such, it should be stressed that characterization by molecular signature is not sufficient alone. As demonstrated from the IKK2 WT/KO lung cancer

models, defining a unique gene set for a specific cancer is not yet a simple task. In that case, Sweet-Cordero et al. 2005 defined gene sets specific to adenocarcinomas. Yet adenocarcinomas specific gene sets were found to be statistically significant in mouse glioblastomas, yet the signatures were explicitly filtered against human primary GBMs when created. However, the molecular signatures were only statistically significant when compared to the matching normal tissue (see table 3.2). The fundamental approach is valid, but signatures should if possible be linked to underlying oncogenes and tumor suppressors in relation to originating tissue or organ as opposed to be considered as static, definitive, and stand-alone.

A particularly troubling aspect of molecular signatures is the lack of agreement on methodology, for both the generation of gene lists and application as signatures. Cahoy et al. 2008 was subsequently reanalyzed by multiple groups, who used different selection criteria, size of the gene list, and even differing number of cell types. Complicating this problem is most of these lists have not been published. As GSEA does not penalize scores if a gene is not present in a list (not on the array or filtered from further analysis), the trend towards ever increasing gene sets is deeply troubling; care must be taken to do extend the net too far. In reality, only a very small fraction of genes contribute significantly to the final scores. Unfortunately, it is not always the same set of genes across data sets.

As bioinformatics has been dominated by statisticians, it is not surprising the majority methods are based on statistical tests. However, for molecular

signatures only one replicate typically exists per tumor, calling into doubt the wisdom of relying statistics. This in many respects is a product of disagreement in how to measure functional enrichment in general though, and not cancers in particular. On the surface, classification and feature selection are standard engineering practices, and methodology from engineering or computer science would likely be beneficial. However, the greatest impediment to this approach is a lack of expert assignment of cancer subtypes which also leads to a disagreement on the meaning of the classifications. Clinical significance may simply be a result of false correlation, instead of a mechanistic underpinning to molecular subtypes.

Differential expression detected from gene expression microarray may not correspond to the initial oncogenes or tumor suppressors. In the mouse KRASLA2 mutated model, KRAS maintains a statistically significant over expressed state (data not shown). However, no statistically significant change in TP53 gene expression has been detected, which suggests silencing of the shRNA against TP53. In the GBMs, neither HRAS nor TP53 show detectable changes in expression (data not shown). This may indicate the mutations required to sustain a cancer is not the same as the mutations required to initiate a cancer[139], the entire construct has been silenced. Another possibility exists that TP53 and HRAS have no relation to tumor initiating mutations but allow the transduced cell to escape cell death and senescence until actual glioblastoma initiating mutations can occur. This seems unlikely however, given the number of tumors with mutations in RTK/RAS/PI3K and p53 pathways[158, 167].

While cell type specific markers do not always indicate the cell of origin when applied to tumors, they can still provide clues as to the progression of tumors. Tumors display cell type specific signatures matching the TCGA tumors as opposed to the expected cell of origin (see figure 3.10). Constructs targeting differentiated cell types (GFAP- and SYN-CRE) typically generate mesenchymal subtypes (CTX, HP, SYN) with cultured astroglial signatures. This strongly suggests that dedifferentiation is required for the generation of mesenchymal subtype tumors.

In contrast, NES-CRE, which should target neuronal stem cells, generates neural subtype tumors with signatures for all cell types except cultured astroglial. This does seem to be a coordinated phenomenon, as it differs from normal samples which display only a neuronal cell type signature (see figure 3.9). In addition, there is a strong negative correlation of cultured astroglia signature, which suggests neuronal subtypes do not simply randomly express all genes. Intriguingly, this suggests mutations required for neural subtypes preferentially maintain cancer cells in a semi-differentiated state.

A small number of constructs generate subtypes opposite of the majority. It is unclear if there are additional, but rare mutations that can lead the cell of origin to different subtype tumor cancer cells as in the genetic mutation model[139], or if this is simply a product of promoter leakiness where the CRE transgene is expressing or at a low level of expression in the wrong cell type. This is in sharp contrast with results published from others regarding proneural subtype tumors, which showed consistent oligodendrocytic and OPC signatures and only arose

from those lineages[168, 171]. However, as the cell of origins and set of all initiating mutations for human primary glioblastomas are unknown, care should be taken when defining the cell of origin for a particular subtype. All that can be definitely stated is it is possible to generate Mesenchymal and Neural subtypes given initiating mutations to TP53 and HRAS in neurons, astrocytes, and neuronal precursor cells.

Given the large quantity of data, a systems biology approach to reconstructing the underlying oncogenic network is of significant value[174]. Carro et al. 2010 applied the ARACNe[175] (algorithm for the reconstruction of accurate cellular networks) to discover critical transcription factors driving the mesenchymal subtype[176]. In the end, a focus on a causative approach to cancer subtypes may be stronger than relying on correlative characterization by molecular signatures. From a systems biology perspective, lentiviral mediated models are a boon to rapid testing of hypotheses for cancer networks. Combinations of putative driver mutations could be tested and compared, with tumorigenesis or chemotherapy resistance as strong end measurements. More complicated experimental designs may address the cell of origin issue for the different GBM subtypes, or possibly the order of critical mutations through inducible constructs.

Methods

Transcriptome Analysis. Samples were prepared from Trizol and hybridized against Affymetrix Mouse Gene ST 1.0 arrays.

Mouse to Human Mapping. Gene annotations were updated to the latest identifiers using data from NCBI and ENSEMBL. Mapping from mouse to human and vice versa was performed through matching of official gene symbols and homology data.

Z-scaling and Rescaled Estimates. The z-score, or standard score, is a method of rescaling each gene to have identical distributions across samples and allows them to be directly comparable. The z-score is calculated by subtracting the mean across all samples and dividing by the standard deviation. Verhaak et al. 2010 reweighted factor analysis gene estimates[177] by multiplying by the median absolute difference (MAD), which is the median of the median subtract from all samples.

Clustering. Initial clustering of gene expression data was performed to test the replication of similar samples. Absolute expression measures were filtered for genes that had deviations greater than typically 1. Hierarchical clustering was performed using the Cluster 3.0 program[178] and visualized with java treeview[179].

Differential Gene Expression. Significantly expressed genes were identified using Cyber-T[30], with a p-value cutoff < 0.05 and a PPDE > 0.95 (FDR $< 5\%$).

GSEA. Functional term enrichment was performed using Gene Set Enrichment analysis[37] by permuting across gene sets. A FDR cutoff of $< 10\%$ was arbitrarily chosen.

Single Sample Gene Set Enrichment Analysis. Verhaak et al. 2010 defined a modification to the GSEA procedure to provide enrichment for single samples. Briefly, all genes in a gene list are ranked according to absolute expression. For a gene set, the empirical cumulative distribution function is calculated using a decreasing hit weighted, or a static miss score. In practice usage of this weight score is problematic as the distribution of gene expression was not truly Gaussian. The CalNC clusters defined by Verhaak et al. 2005 provided a better partitioning of the subtypes, but contained both positive and negative correlated genes. We modified SSGSEA by ranking by signed expression and scoring up and down components separately. A final score of a signature was calculated by subtracting the down component from the up component score. For normalization purposes, scores were divided by the maximum possible score which is closely related to the number of genes in the gene list. Testing of modified SSGSEA on the TCGA data set demonstrated an improvement in resulting subtype scores and a lower misclassification rate (data not shown).

Verhaak Signatures. Subtype signatures were derived from assignments of the CalNC clusters from Verhaak et al. 2010. Genes in all signatures are provided in Appendix Verhaak.

Phillips Signatures. Clustering was performed on the 35 signature genes were provided for that purpose in Phillips et al. 2006. SSGSEA scores were calculated using the full set of differentially expressed genes used to determine the subtypes. Genes are provided in Appendix Phillips.

Cahoy Signatures. While Cahoy et al 2008 provided a top 25 list of potential marker genes; the underlying data has been subsequently reanalyzed for better coverage. Cell type specific signatures were directly taken as defined by Lei et al. 2011 as they provided their cell type specific list. Genes are listed in Appendix Cahoy.

Combined Signatures. To calculate the final classification of a tumor, the highest correlation coefficient between all signatures of a tumor and the average signature profile of tumors from the TCGA data set was chosen. The average of the highest coefficient values was 0.77, representing a high degree of correlation.

Acknowledgements

Chapter 3, in part, uses microarray data and figure 3.1 from *Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model*. Xia, Y; Yeddula, N; Leblanc, M; Ke, E; Zhang, Y; Oldfield, E; Shaw, RJ; Verma, IM. Nat Cell Biol. 2012 Feb 12;14(3):257-6. The dissertation author was a co-author on this paper. Chapter 3 also uses, in part, microarray data and figures 3.10 and 3.12 from *Dedifferentiation of astrocytes and neurons by oncogenes can induce glioblastomas*. Friedmann-Morvinski, D; Bushong, E; Ke, E; Soda, Y; Marumoto, T; Singer, O; Ellisman, M; Verma, IM. In review. The dissertation author was a co-author on this paper.

APPENDIX

Real Time PCR Primers

Name	Forward	Reverse
ACTB_idt1	ACCTTCTACAATGAGCTGCG	CCTGGATAGCAACGTACATGG
ACTB_idt2	GTCTTCCCCTCCATCGTG	GTACTTCAGGGTGAGGATGC
ADNP_harvard	AGCAGGGTAGTCACACTAACA	GGGCATCCCTCAGATTGTATGT
ADNP_idt	GAGGATGTAGGACTGTGGGA	TCTTCACTATGGACATTGCGG
ALAS1_harvard	TGCCGTTAAGAAAGAGGGTGC	TGGTCTCTGCTTTTGCATGAT
ALAS1_idt	TCTGCAAAGCCAGTCTTGAG	CCTCCATCGGTTTTCACACTA
APBB1_harvard	GGACAGTGGAAGGGACTTTGC	GGTGGCGATGTTCTTGGCA
APBB1_idt	CCCTGGACCACTCTAAACTTG	GCCCCATTAATCACATCTACCC
ATF3_harvard	GAGGATTTTGCTAACCTGACGC	GGCTACCTCGGCTTTTGTGAT
ATF3_idt	AGAAGGAACATTGCAGAGCTAAG	GGATTCTAGAGGTACACAGGAAG
ATP5B_harvard	CTATGCGGCGCAAACATCTC	GGTGGTAGTCCCTCATCAAACT
ATP5B_idt	GATCCTCTAGACTCCACCTCTC	AGAAAGTTCATCCATACCCAGG
B2M_harvard	GTGGCCTTAGCTGTGCTCG	ACCTGAATGCTGGATAGCCTC
B2M_idt	GGCATTCCTGAAGCTGACAG	TGGATGACGTGAGTAAACCTG
BAX_IDT	ACGGCAACTTCAACTGGG	CCAATGTCCAGCCCATGAT
BIRC3_HARVARD	AAGCTACCTCTCAGCCTACTTT	CCACTGTTTTCTGTACCCGGA
BRCA1_IDT	CAGAAACCGTGCCAAAAGAC	TGCTTTGTCCTCAGAGTTCTC
C7_harvard	GGCGGTCAGTTGCTGTGTAT	TCCTCTGTTGGACATCCTCTTG
C7_idt	CAGCCTTGTGTTGGAAATGC	TCACAGTCAGAATCCCCATTG
CASP7_HARVARD	GGGACCGAGTGCCTACATATC	CGCCCATACCTGTCACTTTATCA
CASP8_harvard	AGAGCCAGGGTGGTTATTGAA	GCAGTCTCCGAGTCCCCTA
CASP8_idt	AAATGAAAAGCAAACCTCGGG	CTTCAAAGGTCGTGGTCAAAG
CASP8_idt	GGAAATCTCCAAATGCAAACTGG	TGGGCACAGACTCTTTTCAG
CCR3_harvard	ATACAGGAGGCTCCGAATTATGA	ATGCCCCCTGACATAGTGGAT
CCR3_idt	TGCTGAGTTGTATTGGAGAAGTG	CCATCAGTGCTCTGGTATCAG
CD86_harvard	GAACTGTCAGTGCTTGCTAACT	ACCGTGTATAGATGAGCAGGTC
CD86_idt	TCCCTGATGTTACGAGCAATATG	ATCCAAGGAATGTGGTCTGG
CYC1_harvard	CCAGGGAAGCTGTTCGACTA	GCACGATGTAGCTGAGGTCA

CYC1_idt	CTTCAACCCCTACTTTCCTGG	CCTTGGCTATCTGGGACATG
CYCS_IDT	AATCTCCATGGTCTCTTTGGG	TCCATCAGTGTATCCTCTCCC
CYCS_idt1	GCGAGTTTGGTTGCACTTAC	TGCCTTTCTCAACATCACCC
CYCS_idt2	GTGCCACACCGTTGAAAAG	AGTGTATCCTCTCCCCAGATG
DKK1_harvard	ATAGCACCTTGGATGGGTATTCC	CACAGTCTGATGACCGGAGA
DKK1_idt	GTTACTGTGGAGAAGGTCTGTC	GTTCACTGCATTTGGATAGCTG
DUSP1_harvard	CCAGTACAAGAGCATCCCTGT	AGTGGACAAACACCCTTCCTC
DUSP1_idt	ACCACAAGGCAGACATCAG	AAGGTAAGCAAGGCAGATGG
EGR1_harvard	ACCTGACCGCAGAGTCTTTTC	GCCAGTATAGGTGATGGGGG
EGR1_idt	CAGCACCTTCAACCCTCAG	AGTCGAGTGGTTTGGCTG
EIF4A2_harvard	TGGAGATTGAGTTCAAGGAGACC	CAAGTGCCAGAATTACCTTTTGG
EIF4A2_idt	CCACATTTGCTATTTCCATCCTG	GCTCCCATATAGTCTCCAAGTG
FANCC_harvard	AAGGTCTTGGGTATGCACCTA	TCGCCTTTGAGTGTTAAATCCAT
FANCC_idt	ACTTCTCCATCTCTTGCCATG	TGGTCTTCAACTGCTTCTCTG
FOS_harvard	GGGCAAGGTGGAACAGTTATC	CCGCTTGGAGTGTATCAGTCA
FOS_idt	TTGTGAAGACCATGACAGGAG	CCATCTTATTCCTTTCCCTTCGG
GAPDH_harvard	CATGAGAAGTATGACAACAGCC	AGTCCTTCCACGATACCAAAGT
GAPDH_idt	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG
GCSH_IDT	TCACAGAGAAACACGAATGGG	CCAACTTCAGGGAGACTACAAT
GPX8_IDT	CTGAGACTTCCCTCTAGAATCC	GCATAGAACTATAGACAGCAAAACTG
HDAC4_harvard	GGCCCACCGGAATCTGAAC	GCTGCGTTTTCCCGTACCA
HDAC4_idt	ACAAGGAGAAGGGCAAAGAG	GCGTTTTCCCGTACCAGTAG
HMOX1_harvard	CAGTGCCACCAAGTTCAAGC	GTTGAGCAGGAACGCAGTCTT
HMOX1_idt	TCAGGCAGAGGGTGATAGAAG	TTGGTGTCATGGGTCAGC
HPRT1_IDT	TGCTGAGGATTTGGAAAGGG	ACAGAGGGCTACAATGTGATG
IER2_harvard	CAAAGTCAGCCGCAAACGAC	CAGACGGGCTTTCTTGCTC
IER2_idt	AACTTCAGTTTCCCTTCCAGG	CACCCAGGCCCTTTTATACC
KEAP1_harvard	CTGGAGGATCATACCAAGCAGG	GAACATGGCCTTGAAGACAGG
KEAP1_idt	AACAGAGACGTGGACTTTCG	GTGTCTGTATCTGGGTCGTAAC
MBD4_IDT	GTGGCTCTGAAATGGACAAC	TCTTTCTATCTGTGTTCGTGGG
MYC_harvard	CCACAGCAAACCTCCTCACAG	GCAGGATAGTCCTTCCGAGTG
MYC_idt	TTCGGGTAGTGGAAAACCAG	AGTAGAAATACGGCTGCACC
NFKBIZ_harvard	AGCCACACTACACCCACAAAC	GGCAAAACTGTGATTCTGGACC

NFKBIZ_idt	AAGGATGCAGATGGTGACAC	CAAGAACATAGGAAAGTGCCC
POU2F1_harvard	CCCTGTCTCAGCCCATACAGA	GCTGCAAATTGGTGGTTGGAT
POU2F1_idt	CAGCATAGAGACCAACATCCG	GAACCAAACACGAATCACCTC
PRDX1_HARVARD	CATTCCTTTGGTATCAGACCCG	CCCTGAACGAGATGCCTTCAT
PSMB4_HARVARD	GAAGCGTTTTTGGGGTCGC	GAGTGGACGGAATGCGGTA
PTGS2_harvard	ATATGTTCTCCTGCCTACTGGAA	GCCCTTCACGTTATTGCAGATG
PTGS2_idt	ACAGGCTTCCATTGACCAG	TCACCATAGAGTGCTTCCAAC
RHEB_idt1	AAGACCTGCATATGGAAAGGG	CTGCCTCCAAAATTATCCTTCG
RHEB_idt2	GCGGTTGATGTGGTTGGG	TCGTAGGAGTCCACAAATTGG
RRAD_harvard	TGCACGGCAAACAGATGATGT	GCCGCTGATGTCTCAATGAAC
RRAD_idt	GAAACCCTAAAGTCCGAGTCC	GTTCAGGGTCATCGCGTC
SDHA_harvard	TCGCTATTGCACACCTTATATGG	GCACAGTGCGATGACACCA
SDHA_idt	TGGTTGTCTTTGGTCGGG	GCGTTTGGTTTAATTGGAGGG
SESN1_HARVARD	TCAAATACCGAGTCTTCGGATGG	AGGGACACCTCTTAGAAAGCA
SF3A1_harvard	AGCCCAAGTAATCCAAGAGACC	CTTCACCACATCCAAGTCGAA
SF3A1_idt	TTGACTTTCTCCGCCCAC	ACCTGATCCAAAACTTCTCGG
SIRT6_harvard	TGTAAGACGCAGTACGTCCGA	CCCCTGCAATGAGGAAGCTG
SIRT6_idt	AGGATGTCGGTGAATTACGC	GAAGACTGCCAGACCAGC
SOD2_harvard	CTGCTGGGGATTGATGTGTGG	TGCAAGCCATGTATCTTTCAGT
SOD2_idt	CCTGGAACCTCACATCAACG	GCTATCTGGGCTGTAACATCTC
SRXN1_harvard	ACAACTCCACGAAGGTAGGG	CGCAGACATGATTCTTGGGGATA
SRXN1_idt	AGCATCCACACCAGACTTG	ACCCCTGCTATCCCTTCTG
TOP1_harvard	TACTTGGCTGGTTTCCTGGAC	GCCGAGCAGTCTCGTATTTC
TOP1_idt	CTGTAGCCCTGTACTTCATCG	CTACCACATATTCCTGACCATCC
tsc1_HARVARD	CCATGCTACCAATGATTCCACA	TGATGACAGACGGCCAAAAATG
UBC_idt	GATTTGGGTCGCAGTTCTTG	CCTTATCTTGGATCTTTGCCTTG
VHL_harvard	TTGTGCCATCTCTCAATGTTGAC	TCTCAGGCTTGACTAGGCTCC
VHL_idt	TGCCAATATCACACTGCCAG	GTCTTTCTGCACATTTGGGTG
YWHAQ_IDT	AATAACCCAGAGCTTGCCTG	TGAGGGTGCTGTCTTTGTATG
YWHAZ_harvard	CCTGCATGAAGTCTGTAACTGAG	GACCTACGGGCTCCTACAACA
YWHAZ_idt	CTACCGTTACTTGGCTGAGG	CCAGTCTGATAGGATGTGTTGG

Sweet-Cordero

KRAS Up Signature

0610010K14RIK, 1100001G20RIK, 1110008P14RIK, 1190002H23RIK, 1600029D21RIK, 1810046J19RIK, 2410004N09RIK, 2610005L07RIK, 4933407C03RIK, 5730469M10RIK, 6330416G13RIK, A630007B06RIK, AA516738, AASS, ACADL, ACE2, ACLY, ACSL4, ACSL5, ACTN1, ACTN4 /// CAPN12, ADAM19, ADCY7, ADIPOR2, ADSSL1, AES, AI596198, AK1, ALDOA, ALDOC, Ank3, ANXA4, APEX1 /// TMEM55B, APOC1, APRT, ARCN1, AREG, Arf6, ARG1, ARG2 /// VTI1B, Arglu1, ARL8B, ATOX1, ATP11A, ATP1A1, ATP1B1, ATP5C1, ATP5F1, ATP5G1, ATP6V0A1, ATP6V0C, ATP6V0D1, ATP6V1C1, ATXN10, AU021092, AVPI1, AXIN1, AXL, AZIN1, B3GAT3, B4GALNT1, B4GALT3, BASP1, BCL2A1A /// BCL2A1D /// BCL2A1B /// BCL2A1C, BCL2A1B /// BCL2A1D /// BCL2A1A, BCL2A1D /// BCL2A1A /// BCL2A1B /// BCL2A1C, BEX1, BEX4, Bhlhe40, BMP4, BOP1, BRAF, BRD7, Brix1, BSG, BST1, BTG1, BTG3 /// GM7334, C1QB, C1QC, C330027C09RIK, C77080, CAMSAP1, CAPZA3, CAR8, CASK, CCL6, CCL9, CCND1, CCR5 /// CCR2, CCT3, CD14, CD44, CD63, CD68, CD74, Cd8b1, CD9, CDK2AP2, CDKN1A, CEACAM1, CEBPA, CES3, CH25H, CHCHD7, CHD4, CHI3L1, CHI3L3, CHIA, CHL1, CHRNB1, Cisd1, CITED2, CKMT1, CKS2, CLCN5 /// LOC100045272, CLDN3, CLDN7, CLDND1, CLIC1, CLTC, CLU, CNDP2, CNIH2, COL15A1, COL18A1, COTL1, CPOX, CRB3, CRLF1, CRYGD, CSF2, CSF2RB2, CSRP2, CSTB, CTAGE5, CTNND2, Ctsa, CTSB /// FDFT1, CTSC, CTSD, CTSE, CTSH, CTSK, CTSS, CTSZ, CXCL1, CYB5R1, CYB5R3, CYBA, CYHR1, D15ERTD50E, D17WSU104E, DAP, DLK1, DSC2, DUSP6, EDEM1, EEF1B2, EEF1D, EEF2, EHD1, EHMT2, EIF1Ax, EIF2AK4, Eif3e, EIF4B, EIF4G1, ELF5, ELL2, ELOVL1, ENO1 /// GM5506, ENTPD1, Epcam, EPHA7, ERH, ERRFI1, Esrp1, ETV2, F10, F3, F7, FABP1, Fam117a, Fam162a, Fam3c, Fam49b, FASN, FBP2, FCER1G, FCGR2B, FKBP2, FKBP4, FMR1, FNTA, FPR3 /// FPR2, FUCA1, FV1, FV4 /// AI506816 /// LOC622147 /// LOC628577 /// ENV /// LOC664798 /// LOC666404 /// LOC668227 /// LOC668269 /// LOC669098 /// LOC669176 /// LOC669658 /// LOC669821 /// LOC670510 /// LOC671760 /// LOC672016 /// LOC676636, G6PDX, GABPAP, GADD45A, GALNT3, gapdh, GARS, GAS5,

GCH1, GGCX, GJA1, GJA3, GJB2 /// GM10871, GJB3, GLRX, Gm13611, GNB2L1, GNE, GNL3, GNS, Golm1, GP49A /// LILRB4, GPI1, GPR56, GRHPR, GRINA, GSTT1, H13, H19, H2-AA, H2-AB1, H2-DMA, H2-DMB2 /// H2-DMB1, H2-EA, H2-EB1, H2-M1, H2-M9, HAP1, HDC, HDLBP, HEXA, HHEX, HIBADH, HIF1A, HIST1H2AO /// GM11276 /// HIST1H2AC /// HIST1H2AN /// HIST1H2AF /// HIST1H2AH /// HIST1H2AD /// HIST1H2AG /// HIST1H2AI /// HIST1H2AK, HIST1H3F, HK1, HK2, HMGB3, HMGN1, Hnf1b, HNRNPA1L2 /// LOC634350, HOXD1, HPN, HRSP12, HSPA1A /// HSPA1B, HSPA5, HSPA8 /// LOC624853, HSPA9, Hsph1, IBSP, ID2, IFI30, IGFBP3, IGH /// IGH-1A /// AI324046 /// LOC544903 /// LOC628614 /// LOC629871 /// LOC629884 /// LOC634081 /// LOC634100 /// LOC634136 /// LOC634206 /// LOC634222 /// LOC634338 /// LOC634572 /// LOC637000 /// LOC640522, IGH-1A, IGH-4, IGH-6, IGHG, IGHG /// IGH-VJ558 /// AI324046 /// LOC544903 /// IGH-6 /// IGH-5 /// IGHG1 /// IGH-3 /// IGH-VS107 /// LOC380804 /// LOC630565 /// IGH-VX24 /// IGHV14-2 /// IGH-2 /// IGG2A, Ighv1-54, IGH-V7183 /// IGK-V1 /// IGKV1-117 /// CR1 /// IGKV1-110, IGK-V1, IGKV10-95, IL11, IL13RA2, IL18, IL4ra, INHBB, IQGAP1, ITGA4, ITGA8, ITGAX, ITGB2, ITIH4, ITM2C, ITPR2, KCNJ15, KCNK1, KDELR1, KLF5, KLHDC2, KNG1, KRAS, Krt18, Krt7, Krt8, KRT83 /// KRT81 /// 5430421N21RIK, LAMB3, LAMC1, LAMC2, LAP3, LAPTM5, LAS1L, LBP, LCN2, LCP1, LDHA, LGALS3, LITAF, Lpcat1, Lpcat3, LRG1, LRP2, LRRFIP1, LSR, LY6D, LY75, MAN1A, Manf, MAP3K11, MAPK1, MAPRE1, MARCH5, MARCO, MATN4, MBTD1, MCPT8, MDFI, MDFIC, MEF2B /// 2310045N01RIK, MEG3, MMP12, MMP2, MPEG1, MRC1, MRPS18B, MRPS34, MSR1, Mt2, MTIF2, MTX1, MUC1, MX1, MYCN, NAGK, NAPSA, NCL, Ndufaf4, NEK1, NEK4, NFIL3, NME2, NMT1, NNT, Nop56, NPC2, NPDC1, NR1D1 /// THRA, NR2F1, NR2F6, NRP2, NUCB2, NUDT4, NUP88, Obfc2a, ORM1, ORM2, OSBPL1A, OSTF1, PABPC1, PAFAH1B3, Pan2, PAPOLA, PCBD1, PCYT1A, PDIA6, PDK3, PFDN1, PFDN2, PFKL, PFN1, PGK1, PGLS, PGLYRP1, PHB2, PHLDA1, PHLDA2, PIGA, PIGQ, Pip4k2c, PIP5K1A, PISD-PS1 /// PISD-PS3, PKHD1, PKM2, PLA2G5, PLA2G7, Plbd1, PLD3, Plin2, PLP2, PLXNB2, POLG, Polr1c, POLR2E, POLRMT, PON2, PPARG, PPP1R14B, PPP2R5C, PRAMEL4 /// GM13102 /// PRAMEL5, PRB1 /// GM8882 /// PRPMP5 /// GM4736 /// LOC100044247 ///

LOC100044541, PRCC, PRDX4, Prelid1, PRNP, PSAP, PSAT1, PSCA, PSEN1, PSMB5, PSMD4, PSMD5, PSME1, Ptgr1, PTGS1, PTPRF, RABGGTB, RAP1GAP, RBP4, RDH11, REEP6, RELB, RFK, RNASET2A /// RNASET2B, RNF149, Rnf181, RNF4, ROS1, RPL10A, RPL14, RPL28, RPL3, RPL37, RPL6 /// GM6807, RPL8, RPRM, RPS18, RPS2, RPS8, RRBP1, S100A1, S100G, SAT1, SCAMP1, Scd1, SCG3, SDC1, SEC23B, SERPINE1, SERPINE2, SFRS6, SFTPB, SH3RF1, SHC1, SHMT1, SIRPA, Siva1, Slain1, SLC12A2, SLC15A2, SLC16A1, Slc25a39, SLC25A5, SLC2A4, SLC31A1, SLC34A2, SLC38A2, SLC4A4, SLCO6C1, SLPI, Snd1, SNRPE, SNX10, SOAT1, SOCS2, SOCS3, SPECC1, SPG21, SPINT1, SPP1, SSR4, ST13, ST3GAL4, ST6GAL1, ST7, STARD10, STXBP2, TACSTD2, TAF15, TANK, TAOK3, TBC1D24, TCFCP2L1, TCRB-J /// LOC665506, TES, TGFBI, Tgif1, TGOLN1 /// TGOLN2, THBS1, TM2D2, TMEM30A, TMEM30B, TMEM49 /// MIR21, TMEM50B, TMEM56, TMEM62, TNFAIP1, TNFSF9, TNNT1, TOB1, TOM1L1 /// COX11, TPI1, TPM4, TRMT112 /// PRDX5, TROVE2, TSPAN8, TSR1 /// SRR, TSTA3, TSTD2 /// TMOD1, TULP2, TYROBP, UBQLN2, Ubxn1, UOX, VAMP2, VAMP8, VASP, VEGFB, VIL1, WBP5, Wls, XBP1, ZDHHC3, ZDHHC6, ZFP1, ZFP143, ZFP282, ZFP42, ZNHIT3 /// MYO19

KRAS Down Signature

1190002N15RIK, 2210023G05RIK, 2610042L04Rik, 2900062L11RIK, ABCA1, ABCC1, ABLIM1, ACE, ACSL1, ACTA1, ACTA2, ACTC1, ACVR2A, ACVRL1, ADARB1, ADCY8, ADH1A, ADIPOQ, ADRB2, ADRB3, AHNAK, AHR, AKAP12, AKAP8, ALAS2, ALDH1A1, ALDH1A7, ALDH7A1 /// PHAX, ALDOB, ANGPT1, ANKRD1, ANKRD10, Ankrd33b, Ankrd40, ANTXR2, AQP1, ARFGEF1, ARHGEF3, ARMCX2, ARRB1, ART3, ATOH7, ATP1A2, ATP2A2, Atp5k, AW112010, BC026782, BCL6B, BDNF, BMP6, BMYC, BNC1, BNIP2, BPGM, BPHL, BUB1, C3, CACNB2, Cadm1, CALCR, CALCRL, Car2, Car3, CAV1, CAV3, CCBP2, CCKAR, CCL21A /// CCL21B /// CCL21C, CCRL1 /// ACAD11, CCRL2, CD3D, CD47, CD93, CD97, CDC25B /// SPEF1, CDH11, CDH5, CDK14, CDKN1C, CDO1, CDR2, CES3, CFD, CFH, CFHR1, CIDEC, CKMT2, CLDN5, CLEC3B, CLIC4, CNTN1, COL13A1, COL1A1, COL1A2, COL3A1, COL6A2, COX7A1, CP, CPE, CRHR1, CRIP1, CTLA2A, CTLA2B, CXCL12, CXCL14, CXCR4, CXCR7,

CYP2A4 /// CYP2A5, CYP2B10, CYP2B9, CYP2E1, CYP2F2, CYP2S1, CYP4B1, CYR61, Cyth3, D16H22S680E, DCN, Dennd4c, DHX15, DNM1, DPEP1, DPT, DUSP1, EDN1, EDNRB, EFNB2, EMCN, EMP2, ENAH, ENG, ENPP2, EPAS1, EPB4.1, EPB4.1L2, EPHA5, ETS1, ETS2, Fam65b, FANCC, FAS, Fermt2, FEZ2, FGF7, FGF9, FGFR2, FGL2, FHL1, FIGF, FKBP9, FLT1, FMO1, FMO3, FMR1, FOXF1A, FOXF2, FXYD1, FYN, G0S2, GADD45B, GBP2, GBP4, Gbp7, Gcom1, GFRA2, GGH, GIMAP4, GLUD1, GLUL, GM10334 /// PRSS3 /// EG436523 /// PRSS1, Gm4788, GMFG, GNB4, GNG11, GNG2, GPAM, GPC3, GPM6B, GPR182, GPX3, GREM2, GRK5, GSN, GSTA3, GSTM2, GUCY1B3, Gyg, H2AFX, Hamp, HBA-A1 /// HBA-A2, HBA-A2 /// HBA-A1, HBB-B1 /// HBB-B2, HCK, HEPH, HEY1, HIST1H2BC, HNRNPA1L2 /// LOC634350, Hopx, HOXA5, HOXA6, HOXB5, HP, HSD11B1, ICAM2, ID3, IFI203, IFI204 /// MNDA, IFIH1, IFIT3, IFITM3, IGFBP2, IGFBP5, IGFBP6, IGH-1A, IGSF4, IL11RA1 /// IL11RA2 /// GM13305 /// GM2002, IL1B, IL27RA, IL6ST, INMT, INPP5A, ITPKB, JUN, KALRN, Kank3, KCTD12, KDR, KIT, Kitl, KLF2, KLF4, KLF7, KLF9 /// GM9971, KLRA3, Krt13, Krt4, Krt85, LAMA2, LAMB1-1, LATS2, LEPR, Lifr, Limch1, LIN9, LMO2, LOC670044, LOR, LOX, LOXL1, LTB, LTBP4, LY6A, LYL1, LYSMD2, Lyve1, MACF1, MAPT, MATN2, MEF2C, MEIS1, MEOX2, METAP1, MFAP2, MFAP4 /// MAPK7, MFAP5, MFHAS1, MGP, MPDZ, MS4A1, MS4A4D, MS4A6B, MSLN, MTAP4, Mtap7d1, Mtss1l, MYB, MYH1, MYH11, MYH6 /// MYH7, MYL3, MYL4, MYL7, MYL9, MYO1B, MYO6, NDN, NDRG2, NDST1, NESPAS, NFIB, NFKBIA, NID1, NOTCH4, NPNT, NPR3, NR2F2, NT5DC2 /// STAB1, NTN1, NUMB, OGN, OMD, PAM, PAPSS2, PCDHA10 /// PCDHA4 /// PCDHA11 /// PCDHA12 /// PCDHA2 /// PCDHA5 /// PCDHA6 /// PCDHA7 /// PCDHA8 /// PCDHA9 /// PCDHAC1 /// PCDHAC2 /// PCDHA1 /// PCDHA3 /// PCDHA4-G, PCK1, PDGFRA, PDGFRB, PDLIM1, PECAM1, PEG3, PGM1, PGRMC1, Pip4k2c, PKD2, PKIA, PLAC9, PLEKHA1, PLTP /// CTSA, PMP22, POLRMT, PON1, POSTN /// A630052E07RIK, PPAP2A, PPAP2B, PPP1CB, ppp2r3c, PRDX6, PRDX6-RS1, PRG1, PRKCDBP, PRKCE, PROM1, PRSS29, PRX, PSIP1, PSMB10, PTCH1, PTGES, PTGIS, PTPRB, PTPRD, PTRF, Qk, RAB12, RAB28, RABGGTA, RAG1 /// B230118H07RIK, RAMP2, RARRES2 /// LRRC61, RASIP1, RBP1, RCN1, RDH11, RECK, REG3G, RGS2, RHOB, RHOJ,

ROBO1, RPTN, S100A14, S100A8, S100A9, S1PR1, SASH1, SATB1, Sc5d, SCEL, SCGB1A1, SCN7A, SDPR, SEMA3C, SEMA7A, SEPT4, SERPINA3C, SERPING1, SESN1, SH3BP5, SHE, SHOX2, SIAH1A, SLC10A2, Slc4a5, SLC7A5, SLCO3A1, SMARCA2, SNCA, SOD3, SORBS1, SOX11, SOX17, SOX2, SPA17, SPARC, SPARCL1, SPIB, SPNA2, SPNB2, SPOCK2, SSPN, ST5, ST8SIA4, STAB1, STMN2, STMN3, SULT1A1, SULT1D1, SURF2, SUV420H2, TAGLN, TBX2, TBX3, TCF21, Tcf3, TCF4, TCRB-J /// LOC665506, TEK, TEKT1, TENC1, TFPI, TFRC, TGTP1 /// TGTP2 /// GM12185, THBD, TIAM1, TIE1, TIMP3, TJP1, TM2D3 /// TARSL2, TMEFF1, TMEM45A, TMEM71, TNFRSF19, TNNC1, TNNI3, TNNT2, TNNT3, TNXB, TOP2A, TPH1, Tprgl, TPST2, TSPAN13, TSPAN3, TSPAN6, TSPAN7, Tuba1a, TWSG1, U46068, UPK3B, USP18, VAMP3, VAX1, VCL, Vegfa, VPREB3, VWF, WISP2, WNK1, WWTR1, XIST, ZBTB16, ZBTB20, ZBTB46, ZCCHC3, Zeb1, ZMYND11

Adenocarcinoma Signature

ACADL, ACLY, ACTN1, ADCY7, ADSS, AES, AK1, ANK3, ANXA4, APEX1 /// TMEM55B, APOC1, APRT, AREG, ARF6, ARG1, ARG2 /// VTI1B, ATOX1, ATP1B1, ATP5C1, ATP5F1, ATP6V0A1, ATP6V0C, ATP6V0D1, ATP6V1C1, AXL, Azin1, BMP4, BOP1, BSG, BTG1, BTG3 /// GM7334, C1QB, CCL8, CCND1, CD14, CD44, CD63, CD68, CD9, CDKN1A, CEACAM1, CEBPA, CHD4, CHRNB1, CITED2, CKMT1, CKS2, CLIC1, CLU, COL15A1, COL18A1, CSF2, CSRP2, CST3, CSTB, Ctr9, Ctsa, CTSC, CTSD, CTSH, CTSK, CTSS, DAP, Diap1, DLK1, DSC2, DUSP6, EEF1B2, EEF1D, EEF2, Ehmt2, Eif3e, EIF4B, EIF4G1, ELL2, ENO1, ENTPD1, EPHA7, ERH, F10, F3, F7, FABP1, FASN, FBP2, FCER1G, FCGR2B, FEZ2, FKBP2, FKBP4, FMR1, FNTA, GADD45A, GALNT3, GGCX, GJA1, GNAT2, GNS, GSTT1, HDC, HDLBP, HHEX, HIF1A, HMGN1, Hnf1b, HNRNPA1L2 /// LOC634350, HPN, HSPA5, HSPA8 /// LOC624853, IBSP, IFI30, IGFBP3, IL11, IL13RA2, IL18, INHBB, ITGA2, ITGA4, ITGA8, ITGAM, ITGAX, ITGB2, ITPR2, KCNJ15, KDELR1, KLF5, Kras, LAMB3, LAMC1, LAMC2, LAPTM5, LCN2, LGALS3, LITAF, LRP2, LSR, Manf, MAP2K1, MAPK1, MDFI, MMP12, MRC1, MSR1, MTIF2, MUC1, MYH7, NCL, NDN, NEK4, NFIL3, NME2, NMT1, NNT, NPC2, NR2F1, NUCB2, NUP88, ORM1, OSTF1, PABPC1, PAFAH1B3, PAPOLA, PCBD1, PCYT1A, PDK3, PFKL, PFN1, PGK1,

PHLDA1, PHLDA2, PIGA, PIP5K1B, PKM2, PLA2G5, PLA2G7, PLD3, Plin2, POLR2E, PON2, PRCC, PRDX4, PRNP, PSEN1, PSMB5, PSMD4, PSMD5, PTGS1, RABGGTB, Rap1gap, RBP4, RELB, RNF4, ROS1, RPL10A, RPL3, RPL37, RPL6 /// GM6807, RPL8, RPS2, S100A1, S100g, SCAMP1, SEC23B, SERPINE1, SFTPB, SHC1, SHMT1, Sirpa, SLC12A2, SLC15A2, SLC16A1, SLC25A5, SLPI, SPP1, SSR4, ST13, St3gal4, STXBP2, TACSTD2, TANK, TGFBI, Tgif1, THBS1, TNFAIP1, TNFSF9, TNNT1, TOB1, Trove2, Tspan8, TSTA3, TYR, VASP, XBP1

KRAS Signature

ACLY, ATP1B1, ATP5F1, AXIN1, BOP1, BSG, BTG1, CCND1, CD68, CDKN1A, CEACAM1, CITED2, CSF2, CST3, CTSS, DUSP6, EEF1B2, EEF1D, EEF2, Eif3e, FAM3C, FCGR2B, FEZ2, FKBP2, GABPB2, GADD45A, GLRX, GNB2L1, GNS, H2-Aa, HIF1A, Hnf1b, HNRNPA1L2 /// LOC634350, HPN, ID2, IL18, ITGAM, ITGB2, KDELR1, KRT18, KRT8, LAMC1, LCN2, LGALS3, Man1a, Manf, MAPK1, MDFI, MRC1, Mt2, MTIF2, MYCN, MYH7, NMT1, NPC2, NPR2, NUP88, PABPC1, PAFAH1B3, Pcbd1, PDK3, Phb2, PHLDA1, PHLDA2, Plin2, PON2, PSEN1, RABGGTB, RPL14, RPL3, RPL6 /// GM6807, RPL8, RPS2, S100A1, SFTPB, SHC1, Sirpa, SLC12A2, SLC25A5, SND1, TANK, TGFBI, TGOLN1 /// TGOLN2, TNNT1, Trove2, TSTA3, TYR, Ubxn1, VASP

NF-kappaB Targets

ABCA1, ABCB1B, ABCB4, ABCB9, abcc6, abcg5, abcg8, adam19, ADORA1, ADORA2A, ADRA2B, AFP, AGER, AGT, AHCTF1, AICDA, akr1c21, ALOX12, ALOX5, AMACR, AMH, ANGPT1, APOBEC2, APOC3, APOD, APOE, App, AQP4, AR, ARFRP1, ART1 /// CHRNA10, ASPH, ASS1 /// GM5424, ATP1A2, B2M, BACE1, BAX, BCL2 /// D630008O14RIK, BCL2L1, BCL2L11, BCL3, BDKRB1, BDNF, BGN, BCL2A1A /// BCL2A1D /// BCL2A1B /// BCL2A1C, BLNK, BMI1, BMP2, BMP4, BNIP3, BRCA2, BTK, C2CD4A, C3, C4A, C4BP, CALCB, CASP4, CAV1, CCL1, CCL17, CCL19 /// LOC100043921 /// LOC100043918, CCL2, CCL20, CCL22, CCL28, CCL3, CCL4, CCL5, CCND1, CCND2, CCND3, CCR5 /// CCR2, CCR7, CD209a, CD274, CD38, CD3G, CD40, CD40LG, CD44, CD48, cd69, CD80, CD83, CD86, CDK6, CDKN1A, CDX1, CEBPD, CFB, CFLAR, CHI3L1, CIDEA, CLDN2, COL1A2, CR2, CREB3 /// GBA2, CRP, CSF1, CSF2, CSF3, CTSB /// FDFT1, CTSL1, CXCL1, CXCL1, CXCL1, CXCL10, CXCL11, CXCL2, CXCL3, CXCL5, CXCL9, cxcr2, cxcr2, CXCR5, CYP19A1, CYP27B1, CYP2E1, CYP7B1, DCTN4, DEFB2, DIO2, DMP1, DNASE1L2, DPYD, DUSP1, E2F3, EBI3, EDN1, EGFR, EGR1, ELF3, ENG, ENO2, EPHA1, EPO, ERBB2, F11R, F3, F8, FABP6, FAS, fasl, fcer2a, FCGRT, FGF8, FN1, FOS, FSTL3 /// PRSSL1, FTH1, G6PC, G6PDX, GAD1, GADD45B, GATA3, GBP1 /// GBP5, GCLC, GCLM, GCNT1, GJB1, GNAI2, GNB2L1, GRIN1, GRIN2A, GRM2, GSTP1, GUCY1A2, GZMB, H28, H2-K2, H2-M3, H2-Q2 /// H2-Q1, H47, HAMP, HAS1, hba-x, HBB-Y, HGF, hif1a, HMGN1, HMOX1, HOXA9, HPSE, HSD11B2, H2-Ke6, HSP90AA1, ICAM1, ICOS, IDO1, IER2, IER3, IFNB1, IFNG, IGFBP1, IGFBP2, IGHG /// IGH-VJ558 /// AI324046 /// LOC544903 /// IGH-6 /// IGH-5 /// IGHG1 /// IGH-3 /// IGH-VS107 /// LOC380804 /// LOC630565 /// IGH-VX24 /// IGHV14-2 /// IGH-2 /// IGG2A, GM1419 /// GM8760 /// GM1524 /// GM1499 /// IGK-C /// GM10880, Iigp1, IL10, IL11, IL12A, IL12B, IL13, IL15, IL17A, IL1A, IL1B, IL1RN, IL2, IL23A, IL27, IL2RA, IL6, IL9, INHBA, IRF1, IRF2, IRF4, IRF7, JUNB, KCNK5, KCNN2, kdm6b, KISS1 /// GOLT1A, KITL, KLF10, KRT15, KRT5, KRT6B, LAMB2, LBP, LCN2, LEF1, LGALS3, LIPG, LTA, LTB, LTF, LYZ1, MADCAM1, MAP4K1, MBP, MDK, MMP13, MMP3, MMP9, MT3, MTHFR, MUC2, MX1, MYB, MYC, MYLK, MYOZ1, ncam1,

NFKB1, NFKB2, NFKBIA, NFKBIE, NFKBIZ, ngf, NLRP2, NOD2, NOS1, nos2, NOX1, NPY1R, NQO1, NR3C1, NR4A2, NRG1, NUAK2, OLR1, OPN1SW, OPRD1, OPRM1, ORM1, OXTR, PAFAH2, PAX8, PDE7A, PDGFB, PDYN, PENK, PGK1, PGLYRP1, PGR, PIGF /// RHOQ, pIgR, PIK3CA, PIM1, PLA2G4C, PLAU, PLCD1, PLK3, POMC, PPP5C, PRDM1, PRF1, PRKACA /// SAMD1, PRKCD, PRL, PSMB9, PSME1, PSME2, PTAFR, PTEN, PTGDS, PTGES, PTGIS, PTGS2, PTHLH, PTPN1, PTPN13, PTS, PTX3, PYCARD, RAG1 /// B230118H07RIK, RAG2, SYNC /// RBBP4, Rdh1, Rdh7, REL, RELB, REV3L, RIPK2, S100A10, S100A4, S100A6, SAA1 /// SAA2, SAA2, SAA3, SAT1, SCNN1A, SDC4, SELE, SELP, SENP2, SERPINA1B /// SERPINA1A, SERPINA1A /// SERPINA1C /// SERPINA1B /// SERPINA1D /// SERPINA1E, SERPINA1c, SERPINA1d, SERPINA3n, SERPINB1a, SERPINE1, SERPINE2, sfpi1, SH3BGRL, SKP2, SLC11A2, SLC16A1, SLC3A2, SLC6A6, SLFN2, SNAI1, SOD1, SOD2, SOX9, sp7, SPATA19, SPP1, ST6GAL1, ST8SIA1, STAT5A, SUPV3L1 /// 4930507D05RIK, TACR1, TAP1, TAPBP, Tcfec, TCRB, TERT, trf, TFF3, TFPI2, TGM1, TGM2, THBS1, THBS2, TICAM1, TIFA, TLR2, TLR9, TNC, TNF, TNFAIP2, TNFAIP3, TNFRSF1B, TNFRSF4, TNFRSF9, TNFSF10, TNFSF13B, TNFSF15, TNIP1, TNIP3, TRAF1, TRAF2, TREM1, trp53, TRPC1, TWIST1, UBE2M, UCP2, uggt1, UPK1B, UPP1, VCAM1, VEGFC, VIM, VPS53, WNT10B, WT1, XDH, XIAP, YY1, ZFP366

Lung Tumor GSEA

Terms Higher in Lung Compared to Primary Mouse IKK2 KO Tumors

IMMUNE_SYSTEM_PROCESS

IMMUNE_RESPONSE

REACTOME_SIGNALING_IN_IMMUNE_SYSTEM

REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_N

ON_LYMPHOID_CELL

KEGG_PRIMARY_IMMUNODEFICIENCY

KEGG_HEMATOPOIETIC_CELL_LINEAGE

KEGG_CHEMOKINE_SIGNALING_PATHWAY

DEFENSE_RESPONSE

REACTOME_TCR_SIGNALING

CELLULAR_DEFENSE_RESPONSE

RESPONSE_TO_WOUNDING

REACTOME_MUSCLE_CONTRACTION

KEGG_DILATED_CARDIOMYOPATHY

REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY

REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS

REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES

KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION

BIOCARTA_CSK_PATHWAY

REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES

INFLAMMATORY_RESPONSE

REACTOME_HEMOSTASIS

KEGG_FOCAL_ADHESION

ST_T_CELL_SIGNAL_TRANSDUCTION

T_CELL_ACTIVATION

BEHAVIOR

REACTOME_CELL_SURFACE_INTERACTIONS_AT_THE_VASCULAR_WALL

KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION

CELL_ACTIVATION

KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION

KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS

LEUKOCYTE_ACTIVATION

BIOCARTA_TCR_PATHWAY

KEGG_CELL_ADHESION_MOLECULES_CAMS

KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY

KEGG_COMPLEMENT_AND_COAGULATION_CASCADES

KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY

LOCOMOTORY_BEHAVIOR

REACTOME_G_ALPHA_I_SIGNALLING_EVENTS

REACTOME_FORMATION_OF_PLATELET_PLUG

KEGG_LEISHMANIA_INFECTION

LYMPHOCYTE_ACTIVATION

IMMUNE_SYSTEM_DEVELOPMENT

KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM

HEMOPOIESIS

RESPONSE_TO_EXTERNAL_STIMULUS

HEMOPOIETIC_OR_LYMPHOID_ORGAN_DEVELOPMENT

REACTOME_STRIATED_MUSCLE_CONTRACTION

REACTOME_SEMAPHORIN_INTERACTIONS

REACTOME_PLATELET_ACTIVATION

CATION_HOMEOSTASIS

KEGG_ECM_RECEPTOR_INTERACTION

KEGG_REGULATION_OF_ACTIN_CYTOSKELETON

REACTOME_APOPTOTIC_EXECUTION_PHASE

REACTOME_DOWNSTREAM_TCR_SIGNALING

KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC

REACTOME_AXON_GUIDANCE

KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION

RESPONSE_TO_BIOTIC_STIMULUS

REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS

REACTOME_RHO_GTPASE_CYCLE

REACTOME_MITOTIC_PROMETAPHASE

CELL_SURFACE_RECEPTOR_LINKED_SIGNAL_TRANSDUCTION_GO_0007166

REACTOME_DOWNSTREAM_EVENTS_IN_GPCR_SIGNALING

CELLULAR_CATION_HOMEOSTASIS

REGULATION_OF_IMMUNE_SYSTEM_PROCESS

CELL_MIGRATION

REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS

REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS

CHEMICAL_HOMEOSTASIS

KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY

KEGG_GAP_JUNCTION

BIOCARTA_NO1_PATHWAY

REACTOME_MITOTIC_M_M_G1_PHASES

REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS

REACTOME_APOPTOSIS

LEUKOCYTE_DIFFERENTIATION

KEGG_VIRAL_MYOCARDITIS

REACTOME_GPCR_LIGAND_BINDING

REACTOME_THE_ROLE_OF_NEF_IN_HIV1_REPLICATION_AND_DISEASE_PATHOGENESIS

REACTOME_INNATE_IMMUNITY_SIGNALING

BIOCARTA_MCALPAIN_PATHWAY

REACTOME_PLATELET_DEGRANULATION

REGULATION_OF_T_CELL_ACTIVATION

BIOCARTA_KERATINOCYTE_PATHWAY

REGULATION_OF_LYMPHOCYTE_ACTIVATION

REACTOME_TRANSLATION_INITIATION_COMPLEX_FORMATION

POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS

DI___TRI_VALENT_INORGANIC_CATION_TRANSPORT

ST_TUMOR_NECROSIS_FACTOR_PATHWAY

RAS_PROTEIN_SIGNAL_TRANSDUCTION

CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

ION_HOMEOSTASIS

RESPONSE_TO_OTHER_ORGANISM

REACTOME_SEMA4D_IN_SEMAPHORIN_SIGNALING

KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY

ORGAN_DEVELOPMENT

REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS

SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION

REACTOME_PHASE_1_FUNCTIONALIZATION_OF_COMPOUNDS

RESPONSE_TO_VIRUS

KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY

G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

BIOCARTA_GCR_PATHWAY

REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43S_COMPLEX

KEGG_LONG_TERM_DEPRESSION

BIOCARTA_NKT_PATHWAY

ST_B_CELL_ANTIGEN_RECEPTOR

KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION

REACTOME_CD28_CO_STIMULATION

REGULATION_OF_CELL_MIGRATION

MICROTUBULE_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

REACTOME_TELOMERE_MAINTENANCE

REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS

MICROTUBULE_BASED_PROCESS

MESODERM_DEVELOPMENT

INTRACELLULAR_SIGNALING_CASCADE

REACTOME_SMOOTH_MUSCLE_CONTRACTION

BIOCARTA_PDGF_PATHWAY

CALCIUM_ION_TRANSPORT

HOMEOSTATIC_PROCESS

KEGG_PRION_DISEASES

KEGG_CALCIUM_SIGNALING_PATHWAY

SECOND_MESSENGER_MEDIATED_SIGNALING

BIOCARTA_EGF_PATHWAY

HUMORAL_IMMUNE_RESPONSE

ANATOMICAL_STRUCTURE_MORPHOGENESIS

BIOCARTA_IL2RB_PATHWAY

PHOSPHOINOSITIDE_MEDIATED_SIGNALING

REACTOME_SIGNALING_BY_PDGF

CELL_PROLIFERATION_GO_0008283

BIOCARTA_IL12_PATHWAY

REACTOME_CELL_CYCLE_MITOTIC

CELL_SUBSTRATE_ADHESION

REACTOME_NCAM1_INTERACTIONS

POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS

REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE

BIOCARTA_ECM_PATHWAY

BIOCARTA_MET_PATHWAY

AXONOGENESIS

BIOCARTA_IL2_PATHWAY

BIOCARTA_AMI_PATHWAY

BIOCARTA_CASPASE_PATHWAY

RESPONSE_TO_STRESS

REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

SYSTEM_DEVELOPMENT

REACTOME_PLATELET_ACTIVATION_TRIGGERS

REGULATION_OF_CELL_PROLIFERATION

BIOCARTA_MYOSIN_PATHWAY

REACTOME_RNA_POLYMERASE_I_PROMOTER_OPENING

KEGG_MAPK_SIGNALING_PATHWAY

REACTOME_GENES_INVOLVED_IN_APOPTOTIC_CLEAVAGE_OF_CELLULAR_PROTEINS

CELLULAR_HOMEOSTASIS

REACTOME_INTEGRIN_ALPHAIIBBETA3_SIGNALING

ACTIN_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

ACTIN_FILAMENT_BASED_PROCESS

CELL_MATRIX_ADHESION

CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION

REGULATION_OF_HYDROLASE_ACTIVITY

REACTOME_PLATELET_AGGREGATION_PLUG_FORMATION

RESPONSE_TO_CHEMICAL_STIMULUS

SIG_BCR_SIGNALING_PATHWAY

REACTOME_SYNTHESIS_OF_DNA

NEGATIVE_REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS

REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX

REACTOME_NEURORANSMITTER_RECEPTOR_BINDING_AND_DOWNSTREAM_TRANSMI

SSION_IN_THE_POSTSYNAPTIC_CELL

REGULATION_OF_IMMUNE_RESPONSE

ST_FAS_SIGNALING_PATHWAY

BIOCARTA_ERK_PATHWAY

BIOCARTA_BIOPEPTIDES_PATHWAY

REACTOME_COLLAGEN_MEDIATED_ACTIVATION_CASCADE

REGULATION_OF_BIOLOGICAL_QUALITY

BIOCARTA_STATHMIN_PATHWAY

REACTOME_NCAM_SIGNALING_FOR_NEURITE_OUT_GROWTH

POSITIVE_REGULATION_OF_CELL_PROLIFERATION

POSITIVE_REGULATION_OF_HYDROLASE_ACTIVITY

KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY

REACTOME_G1_S_TRANSITION

NEURON_DEVELOPMENT

NEGATIVE_REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENE

SIS

SYSTEM_PROCESS

EXOCYTOSIS

REACTOME_CELL_CYCLE_CHECKPOINTS

REACTOME_GLUCAGON_SIGNALING_IN_METABOLIC_REGULATION

CELL_CYCLE_PROCESS

KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS

POSITIVE_REGULATION_OF_CATALYTIC_ACTIVITY

BIOCARTA_INTEGRIN_PATHWAY

REACTOME_PLC_BETA_MEDIATED_EVENTS

REACTOME_G_ALPHA_12_13_SIGNALLING_EVENTS

APOPTOTIC_PROGRAM

TRANSLATION

REACTOME_GS_ALPHA_MEDIATED_EVENTS_IN_GLUCAGON_SIGNALLING

BIOCARTA_BCR_PATHWAY

REGULATION_OF_CELL_ADHESION

IMMUNE_EFFECTOR_PROCESS

NEGATIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS

VASCULATURE_DEVELOPMENT

REGULATION_OF_CATALYTIC_ACTIVITY

REACTOME_DNA_STRAND_ELONGATION

POSITIVE_REGULATION_OF_IMMUNE_RESPONSE

KEGG_CELL_CYCLE

ACTIVATION_OF_PROTEIN_KINASE_ACTIVITY

CELLULAR_BIOSYNTHETIC_PROCESS

BIOCARTA_PTDINS_PATHWAY

KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY

MITOSIS

CELL_DEVELOPMENT

REGULATION_OF_SIGNAL_TRANSDUCTION

MULTI_ORGANISM_PROCESS

REACTOME_G2_M_CHECKPOINTS

BIOCARTA_CHEMICAL_PATHWAY

BIOCARTA_VEGF_PATHWAY

REACTOME_ORC1_REMOVAL_FROM_CHROMATIN

BIOCARTA_ETS_PATHWAY

KEGG_ENDOCYTOSIS

REACTOME_REGULATION_OF_INSULIN_SECRETION_BY_GLUCAGON_LIKE_PEPTIDE_1

G_PROTEIN_SIGNALING_COUPLED_TO_IP3_SECOND_MESSENGERPHOSPHOLIPASE_C

_ACTIVATING

REGULATION_OF_DEVELOPMENTAL_PROCESS

NITROGEN_COMPOUND_BIOSYNTHETIC_PROCESS

I_KAPPAB_KINASE_NF_KAPPAB_CASCADE

MACROMOLECULE_LOCALIZATION

BIOCARTA_AGR_PATHWAY

G_PROTEIN_SIGNALING_COUPLED_TO_CAMP_NUCLEOTIDE_SECOND_MESSENGER

PEPTIDYL_TYROSINE_MODIFICATION

REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE

KEGG_PPAR_SIGNALING_PATHWAY

REGULATION_OF_MOLECULAR_FUNCTION

NEURITE_DEVELOPMENT

G_PROTEIN_SIGNALING_COUPLED_TO_CYCLIC_NUCLEOTIDE_SECOND_MESSENGER

REACTOME_HIV_INFECTION

REACTOME_DNA_REPLICATION_PRE_INITIATION

REACTOME_G_ALPHA_Q_SIGNALLING_EVENTS

PROTEIN_KINASE_CASCADE

BIOCARTA_FCER1_PATHWAY

REACTOME_REGULATION_OF_APC_ACTIVATORS_BETWEEN_G1_S_AND_EARLY_ANAP
HASE

REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS

POSITIVE_REGULATION_OF_SIGNAL_TRANSDUCTION

REACTOME_PACKAGING_OF_TELOMERE_ENDS

BIOCARTA_FAS_PATHWAY

REACTOME_E2F_TRANSCRIPTIONAL_TARGETS_AT_G1_S

PROTEIN_LOCALIZATION

BIOCARTA_PAR1_PATHWAY

REACTOME_CAM_PATHWAY

REACTOME_NRAGE_SIGNALS_DEATH_THROUGH_JNK

ESTABLISHMENT_OF_PROTEIN_LOCALIZATION

REACTOME_CYTOCHROME_P450_ARRANGED_BY_SUBSTRATE_TYPE

REGULATION_OF_HEART_CONTRACTION

KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY

APOPTOSIS_GO

REGULATION_OF_RESPONSE_TO_STIMULUS

NEGATIVE_REGULATION_OF_BIOLOGICAL_PROCESS

REACTOME_OPIOID_SIGNALLING

REACTOME_GTP_HYDROLYSIS_AND_JOINING_OF_THE_60S_RIBOSOMAL_SUBUNIT

POSITIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS

PROGRAMMED_CELL_DEATH

BIOCARTA_NFAT_PATHWAY

POSITIVE_REGULATION_OF_BIOLOGICAL_PROCESS

REACTOME_M_G1_TRANSITION

REACTOME_DIABETES_PATHWAYS

KEGG_DNA_REPLICATION

BIOCARTA_P38MAPK_PATHWAY

REACTOME_EXTENSION_OF_TELOMERES

REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS

CYCLIC_NUCLEOTIDE_MEDIATED_SIGNALING

CAMP_MEDIATED_SIGNALING

REACTOME_S_PHASE

KEGG_AXON_GUIDANCE

REACTOME_VIRAL_MRNA_TRANSLATION

REACTOME_ACTIVATION_OF_KAINATE_RECEPTORS_UPON_GLUTAMATE_BINDING

ORGANELLE_ORGANIZATION_AND_BIOGENESIS

NEUROGENESIS

PROTEIN_AMINO_ACID_PHOSPHORYLATION

KEGG_JAK_STAT_SIGNALING_PATHWAY

ANGIOGENESIS

GENERATION_OF_NEURONS

NEGATIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS

JAK_STAT_CASCADE

BIOCARTA_HIVNEF_PATHWAY

NEGATIVE_REGULATION_OF_SIGNAL_TRANSDUCTION

M_PHASE_OF_MITOTIC_CELL_CYCLE

REGULATION_OF_G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

REACTOME_CDC20_PHOSPHO_APC_MEDIATED_DEGRADATION_OF_CYCLIN_A

CELL_CYCLE_GO_0007049

REACTOME_INSULIN_SYNTHESIS_AND_SECRETION

KEGG_TYPE_I_DIABETES_MELLITUS

POSITIVE_REGULATION_OF_CELLULAR_PROCESS

TISSUE_DEVELOPMENT

POSITIVE_REGULATION_OF_RESPONSE_TO_STIMULUS

POSITIVE_REGULATION_OF_TRANSLATION

REGULATION_OF_ORGANELLE_ORGANIZATION_AND_BIOGENESIS

NEGATIVE_REGULATION_OF_CELLULAR_PROCESS

PHOSPHORYLATION

TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY

REACTOME_ACTIVATION_OF_NMDA_RECEPTOR_UPON_GLUTAMATE_BINDING_AND_P

OSTSYNAPTIC_EVENTS

REGULATION_OF_PROTEIN_METABOLIC_PROCESS

PROTEIN_SECRETION

REACTOME_BIOLOGICAL_OXIDATIONS

CELL_CYCLE_PHASE

KEGG_PATHWAYS_IN_CANCER

REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE

ORGAN_MORPHOGENESIS

REGULATION_OF_BODY_FLUID_LEVELS

ANATOMICAL_STRUCTURE_FORMATION

REACTOME_FURTHER_PLATELET_RELEASATE

KEGG_SMALL_CELL_LUNG_CANCER

KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION

POSITIVE_REGULATION_OF_CASPASE_ACTIVITY

POSITIVE_REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE

REACTOME_PEPTIDE_CHAIN_ELONGATION

HEMOSTASIS

REACTOME_FORMATION_OF_A_POOL_OF_FREE_40S_SUBUNITS

CASPASE_ACTIVATION

ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

MYELOID_CELL_DIFFERENTIATION

MITOTIC_CELL_CYCLE

NEURON_DIFFERENTIATION

REACTOME_TRANSLATION

REGULATION_OF_ANGIOGENESIS

BIOCARTA_MPR_PATHWAY

REACTOME_METABOLISM_OF_AMINO_ACIDS

REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS

NITROGEN_COMPOUND_METABOLIC_PROCESS

REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION

MACROMOLECULE_BIOSYNTHETIC_PROCESS

CELLULAR_COMPONENT_ASSEMBLY

CELL_CYCLE_CHECKPOINT_GO_0000075

KEGG_PROTEASOME

REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES

BLOOD_COAGULATION

BIOCARTA_TPO_PATHWAY

REACTOME_G_PROTEIN_BETA_GAMMA_SIGNALLING

REACTOME_INFLUENZA_LIFE_CYCLE

CELLULAR_COMPONENT_DISASSEMBLY

NEGATIVE_REGULATION_OF_CELL_PROLIFERATION

BIOSYNTHETIC_PROCESS

KEGG_MELANOGENESIS

KEGG_RIBOSOME

SKELETAL_DEVELOPMENT

COAGULATION

REGULATION_OF_PROGRAMMED_CELL_DEATH

DNA_PACKAGING

REGULATION_OF_APOPTOSIS

KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY

REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION

Terms Higher in Lung Compared to Primary Mouse IKK2 WT Tumors

IMMUNE_RESPONSE

IMMUNE_SYSTEM_PROCESS

REACTOME_SIGNALING_IN_IMMUNE_SYSTEM

REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_N

ON_LYMPHOID_CELL

KEGG_CHEMOKINE_SIGNALING_PATHWAY

DEFENSE_RESPONSE

KEGG_HEMATOPOIETIC_CELL_LINEAGE

KEGG_FOCAL_ADHESION

KEGG_DILATED_CARDIOMYOPATHY

CELLULAR_DEFENSE_RESPONSE

KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY

INFLAMMATORY_RESPONSE

REACTOME_SEMAPHORIN_INTERACTIONS

REACTOME_TCR_SIGNALING

KEGG_LEISHMANIA_INFECTION

KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY

REACTOME_MUSCLE_CONTRACTION

KEGG_PRIMARY_IMMUNODEFICIENCY

REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS

KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM

CATION_HOMEOSTASIS

KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION

KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION

REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES

CELLULAR_CATION_HOMEOSTASIS

REACTOME_STRIATED_MUSCLE_CONTRACTION

RESPONSE_TO_WOUNDING

KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION

REACTOME_G_ALPHA_I_SIGNALLING_EVENTS

ION_HOMEOSTASIS

NF-KB TARGETS

REACTOME_DOWNSTREAM_EVENTS_IN_GPCR_SIGNALING

REACTOME_HEMOSTASIS

REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY

CHEMICAL_HOMEOSTASIS

ST_T_CELL_SIGNAL_TRANSDUCTION

CELL_SURFACE_RECEPTOR_LINKED_SIGNAL_TRANSDUCTION_GO_0007166

REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES

LOCOMOTORY_BEHAVIOR

BIOCARTA_TCR_PATHWAY

REACTOME_INNATE_IMMUNITY_SIGNALING

BIOCARTA_KERATINOCYTE_PATHWAY

KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION

KEGG_GAP_JUNCTION

T_CELL_ACTIVATION

KEGG_REGULATION_OF_ACTIN_CYTOSKELETON

REACTOME_AXON_GUIDANCE

BEHAVIOR

KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY

KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC

KEGG_CELL_ADHESION_MOLECULES_CAMS

REACTOME_FORMATION_OF_PLATELET_PLUG

REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS

KEGG_COMPLEMENT_AND_COAGULATION_CASCADES

REACTOME_GPCR_LIGAND_BINDING

REACTOME_CELL_SURFACE_INTERACTIONS_AT_THE_VASCULAR_WALL

RESPONSE_TO_EXTERNAL_STIMULUS

REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS

REACTOME_RHO_GTPASE_CYCLE

REACTOME_THE_ROLE_OF_NEF_IN_HIV1_REPLICATION_AND_DISEASE_PATHOGENESI

S

KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY

REACTOME_DOWNSTREAM_TCR_SIGNALING

CELL_ACTIVATION

LEUKOCYTE_ACTIVATION

KEGG_CALCIUM_SIGNALING_PATHWAY

RESPONSE_TO_VIRUS

HOMEOSTATIC_PROCESS

REACTOME_PLATELET_ACTIVATION

HEMOPOIETIC_OR_LYMPHOID_ORGAN_DEVELOPMENT

KEGG_ECM_RECEPTOR_INTERACTION

SECOND_MESSENGER_MEDIATED_SIGNALING

SYSTEM_PROCESS

PHOSPHOINOSITIDE_MEDIATED_SIGNALING

IMMUNE_SYSTEM_DEVELOPMENT

CELL_MIGRATION

KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS

CELLULAR_HOMEOSTASIS

G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS

HEMOPOIESIS

LYMPHOCYTE_ACTIVATION

REACTOME_PHASE_1_FUNCTIONALIZATION_OF_COMPOUNDS

REACTOME_COLLAGEN_MEDIATED_ACTIVATION_CASCADE

KEGG_MAPK_SIGNALING_PATHWAY

KEGG_VIRAL_MYOCARDITIS

G_PROTEIN_SIGNALING_COUPLED_TO_IP3_SECOND_MESSENGERPHOSPHOLIPASE_C

_ACTIVATING

RESPONSE_TO_OTHER_ORGANISM

KEGG_ENDOCYTOSIS

BIOCARTA_INTEGRIN_PATHWAY

KEGG_AXON_GUIDANCE

REACTOME_NCAM1_INTERACTIONS

REGULATION_OF_CELLULAR_COMPONENT_ORGANIZATION_AND_BIOGENESIS

ST_TUMOR_NECROSIS_FACTOR_PATHWAY

REACTOME_NCAM_SIGNALING_FOR_NEURITE_OUT_GROWTH

BIOCARTA_PDGF_PATHWAY

REACTOME_SEMA4D_IN_SEMAPHORIN_SIGNALING

KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION

KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY

BIOCARTA_MYOSIN_PATHWAY

BIOCARTA_MET_PATHWAY

REACTOME_APOPTOSIS

BIOCARTA_EGF_PATHWAY

BIOCARTA_IL2RB_PATHWAY

KEGG_PRION_DISEASES

ORGAN_DEVELOPMENT

INTRACELLULAR_SIGNALING_CASCADE

BIOCARTA_NO1_PATHWAY

REGULATION_OF_IMMUNE_SYSTEM_PROCESS

REGULATION_OF_BIOLOGICAL_QUALITY

AXONOGENESIS

RAS_PROTEIN_SIGNAL_TRANSDUCTION

ANATOMICAL_STRUCTURE_MORPHOGENESIS

G_PROTEIN_SIGNALING_COUPLED_TO_CAMP_NUCLEOTIDE_SECOND_MESSENGER

POSITIVE_REGULATION_OF_CATALYTIC_ACTIVITY

REACTOME_SIGNALING_BY_PDGF

REACTOME_NEURORANSMITTER_RECEPTOR_BINDING_AND_DOWNSTREAM_TRANSMI

SSION_IN_THE_POSTSYNAPTIC_CELL

BIOCARTA_AGR_PATHWAY

REGULATION_OF_T_CELL_ACTIVATION

RESPONSE_TO_BIOTIC_STIMULUS

REACTOME_G_ALPHA_Q_SIGNALLING_EVENTS

CELL_SUBSTRATE_ADHESION

NEURITE_DEVELOPMENT

G_PROTEIN_SIGNALING_COUPLED_TO_CYCLIC_NUCLEOTIDE_SECOND_MESSENGER

REACTOME_CD28_CO_STIMULATION

KEGG_LONG_TERM_DEPRESSION

POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS

RESPONSE_TO_CHEMICAL_STIMULUS

CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION

KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION

REGULATION_OF_IMMUNE_RESPONSE

BIOCARTA_BIOPEPTIDES_PATHWAY

REGULATION_OF_CELL_MIGRATION

REACTOME_REGULATION_OF_INSULIN_SECRETION_BY_GLUCAGON_LIKE_PEPTIDE_1

POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS

SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDUCTION

CYCLIC_NUCLEOTIDE_MEDIATED_SIGNALING

MULTI_ORGANISM_PROCESS

KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY

CAMP_MEDIATED_SIGNALING

LEUKOCYTE_DIFFERENTIATION

REACTOME_PLATELET_ACTIVATION_TRIGGERS

REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS

KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY

BIOCARTA_BCR_PATHWAY

SIG_BCR_SIGNALING_PATHWAY

KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY

REACTOME_SMOOTH_MUSCLE_CONTRACTION

DI___TRI_VALENT_INORGANIC_CATION_TRANSPORT

REACTOME_PLATELET_DEGRANULATION

SYSTEM_DEVELOPMENT

RESPONSE_TO_STRESS

VASCULATURE_DEVELOPMENT

ST_B_CELL_ANTIGEN_RECEPTOR

BIOCARTA_NKT_PATHWAY

NEURON_DEVELOPMENT

REGULATION_OF_CATALYTIC_ACTIVITY

PROTEIN_AMINO_ACID_PHOSPHORYLATION

REACTOME_APOPTOTIC_EXECUTION_PHASE

REGULATION_OF_CELL_PROLIFERATION

REACTOME_PLC_BETA_MEDIATED_EVENTS

REACTOME_TOLL_RECEPTOR_CASCADES

CELL_MATRIX_ADHESION

REACTOME_INTEGRIN_ALPHAIIBBETA3_SIGNALING

HUMORAL_IMMUNE_RESPONSE

KEGG_PATHWAYS_IN_CANCER

ENZYME_LINKED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

ELECTRON_TRANSPORT_GO_0006118

CATION_TRANSPORT

BIOCARTA_ERK_PATHWAY

REACTOME_GLUCAGON_SIGNALING_IN_METABOLIC_REGULATION

POSITIVE_REGULATION_OF_IMMUNE_RESPONSE

ORGAN_MORPHOGENESIS

REGULATION_OF_G_PROTEIN_COUPLED_RECEPTOR_PROTEIN_SIGNALING_PATHWAY

REACTOME_CYTOCHROME_P450_ARRANGED_BY_SUBSTRATE_TYPE

POSITIVE_REGULATION_OF_TRANSFERASE_ACTIVITY

ION_TRANSPORT

REACTOME_GS_ALPHA_MEDIATED_EVENTS_IN_GLUCAGON_SIGNALLING

BIOCARTA_CCR5_PATHWAY

REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE

REACTOME_ACTIVATION_OF_NMDA_RECEPTOR_UPON_GLUTAMATE_BINDING_AND_P

OSTSYNAPTIC_EVENTS

PROTEIN_KINASE_CASCADE

REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43

S_COMPLEX

POSITIVE_REGULATION_OF_MAP_KINASE_ACTIVITY

KEGG_PPAR_SIGNALING_PATHWAY

REGULATION_OF_HEART_CONTRACTION

REGULATION_OF_MOLECULAR_FUNCTION

REACTOME_G_ALPHA_S_SIGNALLING_EVENTS

BIOCARTA_PAR1_PATHWAY

REGULATION_OF_RESPONSE_TO_STIMULUS

KEGG_PANCREATIC_CANCER

REACTOME_G_ALPHA_12_13_SIGNALLING_EVENTS

REGULATION_OF_ANGIOGENESIS

G_PROTEIN_SIGNALING_ADENYLATE_CYCLASE_ACTIVATING_PATHWAY

POSITIVE_REGULATION_OF_CYTOKINE_BIOSYNTHETIC_PROCESS

BIOCARTA_MAPK_PATHWAY

REGULATION_OF_LYMPHOCYTE_ACTIVATION

BIOCARTA_FCER1_PATHWAY

NEGATIVE_REGULATION_OF_BIOLOGICAL_PROCESS

PHOSPHORYLATION

PEPTIDYL_TYROSINE_MODIFICATION

REACTOME_PLATELET_AGGREGATION_PLUG_FORMATION

BIOCARTA_IL1R_PATHWAY

GENERATION_OF_NEURONS

CELL_DEVELOPMENT

TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_SIGNALING_PATHWAY

KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS

REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES

KEGG_BUTANOATE_METABOLISM

PROTEIN_SECRETION

NEGATIVE_REGULATION_OF_CELL_PROLIFERATION

NEGATIVE_REGULATION_OF_CELLULAR_PROCESS

REACTOME_PI3K_CASCADE

KEGG_CARDIAC_MUSCLE_CONTRACTION

CELLULAR_BIOSYNTHETIC_PROCESS

REGULATION_OF_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

CELL_PROLIFERATION_GO_0008283

ST_INTEGRIN_SIGNALING_PATHWAY

REACTOME_TOLL_LIKE_RECEPTOR_3_CASCADE

REACTOME_OPIOID_SIGNALLING

REACTOME_RNA_POLYMERASE_I_PROMOTER_OPENING

REGULATION_OF_DEVELOPMENTAL_PROCESS

POSITIVE_REGULATION_OF_HYDROLASE_ACTIVITY

REGULATION_OF_HYDROLASE_ACTIVITY

ANGIOGENESIS

BIOCARTA_IL12_PATHWAY

BIOCARTA_FMLP_PATHWAY

REACTOME_BIOLOGICAL_OXIDATIONS

MAPKKK_CASCADE_GO_0000165

ANATOMICAL_STRUCTURE_FORMATION

BIOCARTA_HIVNEF_PATHWAY

POSITIVE_REGULATION_OF_RESPONSE_TO_STIMULUS

APOPTOTIC_PROGRAM

NEUROGENESIS

NEURON_DIFFERENTIATION

REACTOME_TRANSLATION_INITIATION_COMPLEX_FORMATION

APOPTOSIS_GO

PROGRAMMED_CELL_DEATH

REACTOME_CLASS_B2_SECRETIN_FAMILY_RECEPTORS

KEGG_MELANOGENESIS

KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY

BIOCARTA_PTDINS_PATHWAY

GENERATION_OF_PRECURSOR_METABOLITES_AND_ENERGY

REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS

BIOCARTA_GPCR_PATHWAY

METAL_ION_TRANSPORT

CALCIUM_ION_TRANSPORT

KEGG_ACUTE_MYELOID_LEUKEMIA

TISSUE_DEVELOPMENT

REACTOME_SIGNALING_BY_EGFR

POSITIVE_REGULATION_OF_CELL_PROLIFERATION

REACTOME_ACTIVATION_OF_KAINATE_RECEPTORS_UPON_GLUTAMATE_BINDING

CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

BIOCARTA_VEGF_PATHWAY

REACTOME_SIGNALLING_BY_NGF

REGULATION_OF_PROTEIN_METABOLIC_PROCESS

ACTIN_FILAMENT_BASED_PROCESS

ACTIN_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS

KEGG_RETINOL_METABOLISM

REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE

REACTOME_GLUCAGON_TYPE_LIGAND_RECEPTORS

KEGG_LONG_TERM_POTENTIATION

POSITIVE_REGULATION_OF_TRANSLATION

REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS

ACTIVATION_OF_MAPK_ACTIVITY

TRANSLATION

BIOCARTA_NFAT_PATHWAY

MEMBRANE_ORGANIZATION_AND_BIOGENESIS

KEGG_PYRUVATE_METABOLISM

REGULATION_OF_SIGNAL_TRANSDUCTION

REGULATION_OF_TRANSFERASE_ACTIVITY

REACTOME_FURTHER_PLATELET_RELEASATE

REACTOME_POST_NMDA_RECEPTOR_ACTIVATION_EVENTS

KEGG_APOPTOSIS

REGULATION_OF_CELLULAR_PROTEIN_METABOLIC_PROCESS

IKK2 KO/WT GSEA

Primary Tumors, All Terms Enriched in WT Over KO

MITOTIC_CELL_CYCLE

CELL_CYCLE_PROCESS

CELL_CYCLE_GO_0007049

REACTOME_CELL_CYCLE_MITOTIC

CELL_CYCLE_PHASE

REACTOME_MITOTIC_M_M_G1_PHASES

M_PHASE_OF_MITOTIC_CELL_CYCLE

M_PHASE

MITOSIS

REACTOME_MITOTIC_PROMETAPHASE

KEGG_CELL_CYCLE

REACTOME_G1_S_TRANSITION

REGULATION_OF_CELL_CYCLE

DNA_REPLICATION

REACTOME_CELL_CYCLE_CHECKPOINTS

REACTOME_S_PHASE

REACTOME_DNA_REPLICATION_PRE_INITIATION

INTERPHASE_OF_MITOTIC_CELL_CYCLE

REACTOME_TELOMERE_MAINTENANCE

RESPONSE_TO_DNA_DAMAGE_STIMULUS

REACTOME_SYNTHESIS_OF_DNA

INTERPHASE

REACTOME_G2_M_CHECKPOINTS

DNA_METABOLIC_PROCESS

REACTOME_HEMOSTASIS

KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION

CELL_PROLIFERATION_GO_0008283

DNA_REPAIR

REACTOME_PLATELET_DEGRANULATION

REACTOME_FORMATION_OF_PLATELET_PLUG

KEGG_PYRIMIDINE_METABOLISM

REACTOME_PLATELET_ACTIVATION

PROTEIN_LOCALIZATION

RESPONSE_TO_ENDOGENOUS_STIMULUS

MACROMOLECULE_LOCALIZATION

KEGG_OOCYTE_MEIOSIS

CHROMOSOME_ORGANIZATION_AND_BIOGENESIS

Cell Lines, All Terms Enriched in WT Over KO

REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX

REACTOME_DNA_REPLICATION_PRE_INITIATION

REACTOME_MITOTIC_M_M_G1_PHASES

REACTOME_SYNTHESIS_OF_DNA

REACTOME_G2_M_CHECKPOINTS

REACTOME_DNA_STRAND_ELONGATION

KEGG_DNA_REPLICATION

REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS

REACTOME_S_PHASE

REACTOME_CELL_CYCLE_MITOTIC

REACTOME_CELL_CYCLE_CHECKPOINTS

REACTOME_MITOTIC_PROMETAPHASE

REACTOME_EXTENSION_OF_TELOMERES

REACTOME_G1_S_TRANSITION

REACTOME_LAGGING_STRAND_SYNTHESIS

REACTOME_TELOMERE_MAINTENANCE

REACTOME_M_G1_TRANSITION

DNA_DEPENDENT_DNA_REPLICATION

DNA_REPLICATION

DNA_METABOLIC_PROCESS

REACTOME_METABOLISM_OF_CARBOHYDRATES

REACTOME_TRANSPORT_OF_MATURE_MRNA_DERIVED_FROM_AN_INTRON_CONTAINI

NG_TRANSCRIPT

REACTOME_ORC1_REMOVAL_FROM_CHROMATIN

KEGG_HEMATOPOIETIC_CELL_LINEAGE

REACTOME_DOUBLE_STRAND_BREAK_REPAIR

REACTOME_METABOLISM_OF_RNA

REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS

REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS

REACTOME_TRANSPORT_OF_THE_SLBP_INDEPENDENT_MATURE_MRNA

M_PHASE

M_PHASE_OF_MITOTIC_CELL_CYCLE

REACTOME_SNRNP_ASSEMBLY

KEGG_HOMOLOGOUS_RECOMBINATION

KEGG_MISMATCH_REPAIR

REACTOME_DNA_REPAIR

REACTOME_GLUCOSE_TRANSPORT

REACTOME_STEROID_METABOLISM

MITOSIS

DNA_RECOMBINATION

KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY

KEGG_PYRIMIDINE_METABOLISM

REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX

KEGG_CYSTEINE_AND_METHIONINE_METABOLISM

CELL_CELL_ADHESION

LOCOMOTORY_BEHAVIOR

REACTOME_PHASE_II_CONJUGATION

KEGG_SPLICEOSOME

RESPONSE_TO_DNA_DAMAGE_STIMULUS

CHROMOSOME_SEGREGATION

REACTOME_GLOBAL_GENOMIC_NER

RESPONSE_TO_ENDOGENOUS_STIMULUS

CELL_CYCLE_PROCESS

DNA_REPAIR

CELL_CYCLE_PHASE

REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA

REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION

KEGG_COMPLEMENT_AND_COAGULATION_CASCADES

CELL_CYCLE_CHECKPOINT_GO_0000075

BEHAVIOR

REACTOME_INNATE_IMMUNITY_SIGNALING

KEGG_CHEMOKINE_SIGNALING_PATHWAY

REACTOME_SIGNALING_IN_IMMUNE_SYSTEM

MITOTIC_CELL_CYCLE

REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS

KEGG_CELL_CYCLE

REACTOME_GLUCOSE_METABOLISM

DNA_DAMAGE_RESPONSESIGNAL_TRANSDUCTION

CHROMOSOME_ORGANIZATION_AND_BIOGENESIS

KEGG_NUCLEOTIDE_EXCISION_REPAIR

REACTOME_BIOLOGICAL_OXIDATIONS

RESPONSE_TO_EXTERNAL_STIMULUS

REACTOME_REGULATION_OF_APC_ACTIVATORS_BETWEEN_G1_S_AND_EARLY_ANAP

HASE

POSITIVE_REGULATION_OF_CELL_PROLIFERATION

REACTOME_HIV_LIFE_CYCLE

REACTOME_NUCLEOTIDE_EXCISION_REPAIR

REACTOME_INFLUENZA_LIFE_CYCLE

KEGG_RNA_DEGRADATION

RESPONSE_TO_STRESS

ANATOMICAL_STRUCTURE_MORPHOGENESIS

REACTOME_LATE_PHASE_OF_HIV_LIFE_CYCLE

KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION

APOPTOTIC_PROGRAM

REACTOME_REGULATION_OF_BETA_CELL_DEVELOPMENT

REACTOME_GENE_EXPRESSION

RNA_SPLICING

MEIOTIC_CELL_CYCLE

REACTOME_REGULATION_OF_GENE_EXPRESSION_IN_BETA_CELLS

REACTOME_GPCR_LIGAND_BINDING

Verhaak

Classical Down

4930506M07RIK, ACSL1, ACSL4, AGTPBP1, ANKS1B, ARRB1, ATRNL1, BASP1, BCAS1, BEST1, CDC42, CDR1, CUTC, CYTH1, DYNC1I1, EDIL3, ENPP2, ENPP4, EPB4.1, EPB4.1L3, EVI2A, FAM49B, FHIT, FOLR2, FUT9, GNAI1, HPRT, MAGEH1, MBP, MMD, MORF4L2, MS4A4A, MSRB2, NANOS1, PARP8, PGBD5, PIGP, PLCL1, POPDC3, PPA1, PPFIA2, RABGAP1L, REPS2, SAR1A, SCPEP1, SGK3, SH3GL2, SH3GL3, SLC16A7, SLC31A2, SYNGR2, TEC, TLR4, TPM3, UCP2

Classical Up

2510012J08RIK, 3110056O03RIK, 4931406P16RIK, ABCD2, ACSBG1, ACSL3, ADAM19, AKAP8L, AKT2, APBA3, ARAP2, ARAP3, ARHGEF18, B3GALT1, BLM, BTBD2, C030046I01RIK, CALM1, CAMK2B, CC2D1A, CD151, CD3EAP, CDH2, CDH4, CDH6, CDK6, CHERP, CLIP2, CREB5, DAG1, DENND2A, DMWD, DOCK6, EGFR, ELOVL2, ERCC2, EXTL3, EYA2, FBXO17, FGFR3, FZD3, FZR1, GAS1, GLG1, GLI2, GM10991, GM9847, GNA11, GNAS, GNG7, GPR56, GRIK1, GRIK5, GTF2F1, HMG20B, HS3ST3B1, HSPBP1, IRF3, IRS2, ITGA7, ITGB8, JAG1, JUND, KCNF1, KEAP1, KLHDC8A, KLHL25, KLHL4, LAMA5, LAMB2, LFNG, LHFP, LMO2, LRFN3, LRP5, MAB21L1, MAU2, MCC, MEGF8, MEIS1, MEOX2, MLC1, MYO10, MYO5C, NCLN, NES, NOS2, NOTCH3, NPAS3, NPEPL1, NR2F6, ORF61, PDGFA, PEPD, PLCG1, PLEKHA4, POFUT1, POLRMT, POMT2, PRKD2, PRPF31, PTPRA, QTRT1, RASGRP1, RBCK1, RBM42, RFX2, RFXANK, RGS12, RGS6, SARS2, SCAMP4, SEMA6A, SEMA6D, SEPT11, SHOX2, SIPA1L1, SLC12A4, SLC4A4, SLC6A11, SLC6A9, SMO, SOX9, SPRY2, STK11, TBX2, TGIF2, TLE2, TMED1, TMEM147, TMEM161A, TRIB2, TYK2, UNC45A, UPF1, VAV3, VPS16, WSCD1, ZFHX4, ZFP111, ZFP112, ZFP128, ZFP235, ZFP446, ZFP94, ZFP954, ZYX

Mesenchymal Down

ABAT, ANKRD46, ASCL1, BAI3, BCAN, BEX1, CDK5R1, CDKN1B, CKB, CLASP2, CRB1, CSPG5, DLL3, DPF1, DPP6, FXYD6, GPM6A, GRIA2, GSTA4, MAPT, MARCKSL1, MPPED2, MTAP2, MYST2, NCALD, NLGN3, NRXN1, OLIG2, PAFAH1B3, PHLPP1, PRPSAP2, PURG, REEP1, RUFY3, SCG3, SCHIP1, SEZ6L, SOX2, SPAST, SRGAP3, TSPAN3, TTYH1, VEZF1, ZFP606, ZFP821

Mesenchymal Up

A230050P20Rik, ACPP, ADAM12, AIM1, ALDH3B1, ALOX5, AMPD3, ANXA2, ARPC1B, BATF, BC013712, BDKRB2, BNC2, C5AR1, CASP4, CASP4, CAST, CCR5, CD14, CD2AP, CD4, CDCP1, CEBPB, CHPF2, CLCF1, CNN2, COL1A1, COL1A2, COL5A1, COL8A2, CSTA, CTSA, CTSB, CTSC, CTSZ, CYTH4, CYTIP, DAB2, DCBLD2, DOK3, DSC2, DSE, ELF4, ENG, FCGR2B, FCGR3, FES, FHL2, FHOD1, FMNL1, FNDC3B, FPR3, FURIN, FXYD5, GCNT1, GLT25D1, GM7665, GNA15, GRN, HEXA, HEXB, HK3, IFI30, IGFBP6, IL15RA, IL1R1, IL4RA, IQGAP1, ITGA4, ITGA5, ITGAM, ITGB2, KYNU, LAIR1, LAMB1, LAPTM5, LCP1, LCP2, LHFPL2, LILRA6, LILRB3, LOX, LRRFIP1, LTBP2, LY75, LY96, MAFB, MAN1A, MAN2A1, MAN2B1, MAPK13, MFSD1, MGAT1, MSR1, MVP, MYH9, MYO1F, MYOF, NCF2, NCF4, NOD2, NPC2, NRP1, P4HA2, PLA2G15, PLAU, PLAUR, PLBD1, PLK3, POLD4, PROCR, PTGER4, PTPN22, PTPN6, PTPRC, RAB11FIP1, RAB27A, RAC2, RBMS1, RELB, RHOG, RRAS, RUNX2, S100A4, SAT1, SEC24D, SERPINA1E, SERPINE1, SFT2D2, SH2B3, SHC1, SIGLECE, SIGLECE, SLAMF8, SLC10A3, SLC11A1, SLC16A3, SQRDL, SRPX2, ST14, STAB1, STAT6, STXBP2, TCIRG1, TES, TGFBI, TGFBR2, TGOLN2, THBD, THBS1, TIMP1, TLR2, TNFAIP3, TNFAIP8, TNFRSF11A, TNFRSF1B, TRADD, TRPM2, TYMP, UAP1, VDR, WIPF1, WWTR1

Neural Down

ABL1, ACTN4, ADCY9, AFAP1, AFF4, AKAP13, ANKRD11, AP3D1, BICD2, BMS1, BOP1, BPTF, BRD4, BRPF1, CASP2, CDV3, CHD4, CHST3, CIZ1, CKAP4, COL4A2, D19BWG1357E, DCP1A, DDX42, DIAP1, DNAJC13, DNMT1, DOT1L, DPP3, DROSHA, EEF2, ELAVL1, EP400,

EXT1, F630110N24RIK, FAM38A, FAM46A, FERT2, FLNA, GANAB, GATAD2A, GCN1L1, GNL1, GNL2, GOLGA2, GOLGA3, GPR161, GPR172B, HCFC1, HELZ, HNRNPA3, HNRNPAB, HNRNPM, HNRNPUL2, HSP90B1, ILF3, KDM2A, KDM4B, KDM5A, KHSRP, KIRREL, KPNB1, LAMC1, LARP1, LEPREL2, LMAN1, LMNB2, MAML1, MBTPS1, MC1R, MED12, MIER2, MLEC, MLXIP, MORC2A, MSL2, MYO9B, MYST3, NCL, NCOR2, NFATC3, NIPBL, NUP188, P4HB, PABPC1, PCSK7, PHC2, PLOD3, PLXNA1, PPM1G, PRKDC, PRRC2C, PTBP1, PXN, QTRTD1, RAD54L2, RBBP6, RBM10, RBM15B, RRP1B, SAFB, SEC61A1, SERPINH1, SMARCA4, SNTB2, SP1, SRF, SRRM2, SSRP1, STK10, TARS, TCF3, THOC2, TMEM43, TOP1, TPM4, TPR, TRAM2, TRIO, TRRAP, TSPAN9, TTC28, UBN1, WIZ, XPO6, ZBTB43, ZDHHC18, ZFP146, ZFP629

Neural Up

1810012P15Rik, ACYP2, ADD3, AGXT2L1, AI747448, AKR7A5, ANXA3, ANXA7, ATP5F1, CALM2, CAMK2G, CAR4, CASQ1, CCDC121, CCK, CHN1, COX5B, CPNE6, CRBN, CRYL1, CRYM, CRYZL1, DHRS9, FBXO3, FEZF2, FXYD1, GABARAPL2, GABRB2, GM6822, GPR22, GRM1, GRM3, GUK1, HPCA, HPCAL4, IMPA1, KCNJ3, KCNK1, LYRM1, MAT2B, MDH1, MGST3, MRPL49, MYBPC1, NDP, NDRG2, NDUFS3, NSL1, NTSR2, ORC4, PDE6D, PEX11B, PEX19, PPP1R1A, PPP2R5A, RBKS, RERGL, RND1, ROGDI, S1PR1, SEPP1, SEPW1, SERPINI1, SIRT5, SLC30A10, SLCO1A4, SNCG, SNTA1, SNX11, TCEAL1, THTPA, TMEM144, TSNAX, TTC1, TTPA, UROS, USP33, VIP, VSX1, YPEL5

Proneural Down

ACSS3, ANXA1, ANXA4, ANXA5, ARHGAP29, ARNTL, ARSJ, ASL, BC028528, BLVRB, CASP1, CASP8, CCDC109B, CD97, CHI3L1, CLIC1, COPZ2, CYBRD1, DLC1, DRAM1, EFEMP2, EHD2, EMP3, EPHB4, FZD7, GALNT4, GJA1, GM6907, GSTK1, HFE, ICAM5, ILK, LGALS1, LGALS3, LRP10, LRRC16A, LTBP1, MGST2, MRC2, MYO1E, NR2E1, OSBPL3, PCSK5, PDPN, PGCP, PIPOX, PLA2G5, PLIN3, PLS3, PMP22, PTPN14, PTRF, PYGL, RAB32, RIN1, RREB1, S100A13, SLC2A10, SP100, SSH3, SWAP70, SYPL, TEAD3, TGFB3, TMBIM1, TNFRSF1A, TRIM38, TRIP6, VAMP5, YAP1, ZFP217

Proneural Up

2610020H08Rik, 5730559C18RIK, ACTR1A, ALCAM, AMOTL2, ARHGAP33, ARHGEF9, ATAD5, ATAT1, ATP1A3, BCL7A, BCOR, C1QL1, CAMSAP2, CAR10, CASK, CBX1, CDC25A, CDC7, CELF3, CHD7, CLGN, CNTN1, CRMP1, CSNK1E, CXXC4, DBN1, DCAF7, DCX, DGKI, DNM3, DPYSL4, DUSP26, E130309F12Rik, E2F3, EPHB1, ERBB3, FAM110B, FAM125B, FBXO21, FERMT1, FGF9, FHOD3, FLRT1, GABRA3, GADD45G, GM11223, GNG4, GPR17, GRID2, GSK3B, HDAC2, HMGB3, HN1, HNRNPH3, HOXD3, HRASLS, ICK, IL1RAPL1, KDM1A, KIF21B, KLRC1, KLRC3, KLRK1, LPAR4, LPHN3, LRP6, LRRTM4, MARCKS, MAST1, MATR3, MCM10, MLLT11, MMP15, MMP16, MTSS1, MYB, MYT1, NCAM1, NKAIN1, NKX2-2, NOL4, NR0B1, NRXN2, P2RX7, PAK3, PAK7, PCDH11X, PCDH11X, PDE10A, PELI1, PFN2, PHF16, PLCB4, PODXL2, PPM1D, PPM1E, RAB33A, RAD21, RALGPS1, RALGPS2, RAP2A, RBPJ, RNFT2, SATB1, SCN3A, SEC61A2, SLC1A1, SLCO5A1, SORCS3, SOX10, SOX11, SOX4, SPNB3, STMN4, TAF5, TMCC1, TMEFF1, TMEM35, TMSB15A, TOP2B, TOPBP1, TOX3, TTC3, UGT8A, VAX2, WASF1, YPEL1, ZC4H2, ZEB2, ZFP184, ZFP248, ZFP286, ZFP300, ZFP711, ZFP804A

Phillips

35 Signature Genes

DLL3 (PN), SRRM2 (PN), SOX8 (PN), FERMT1 (PN), CSDC2 (PN), GALNT13 (PN), NDRG2 (PN), NCAM1 (PN), RASL10A (PN), GABBR1, (PN), SCG3 (PN), SNAP91 (PN), ATP6V1G2 (PN), KLRC3 (PN), PDLIM4 (MES), PLA2G5 (MES), COL4A2 (MES), COL4A1 (MES), PDPN (MES), FAM20C (MES), ANGPTL4 (MES), SPOCD1 (MES), SERPINE1 (MES), TAGLN (MES), MYL9 (MES), LIF (MES), FOSL2 (MES), CHI3L1 (MES), TIMP1 (MES), E2F7 (PROLIF), DTL (PROLIF), IQGAP3 (PROLIF), HMMR (PROLIF), CENPK (PROLIF)

Proneural Full Signature

ABHD6, ABLIM1, ABLIM3, ACSM5, ADAM22, ADCY2, AJAP1, AKR1C21, ALDH5A1, ALDOC, ANKS1B, AP2B1, APOE, ARHGAP22, ARL3, ARPP21, ASB13, ASCL1, ATP6V1G2, ATRNL1, B3GAT1, BCAN, BEND7, BMP2, 1110014N23RIK, C530028O21RIK, 1190002H23RIK, C1QL1, CADM2, CALCRL, CALN1, CBX7, CCNK, CDR1, CECR6, CMTM5, CNTN1, CNTN3, , CRTC1, CRYAB, CSDC2, CSMD3, CYFIP2, DLGAP1, DLL1, DLL3, DNAJC12, DNM3, DOK6, DPP10, DSCAM, DSCAML1, DTX4, DUSP26, EFHA2, EHD3, ELMO1, ENHO, EPB4.1L2, EPHB1, F2, FAIM2, FAM110B, FAM13C, FAM155A, FAM19A5, FBXL15, FBXO2, FERMT1, FGF12, FGF13, FGF14, FLRT1, FRY, FSD1, FSTL5, FUT9, FXYD6, GAB2, GABBR1, GABBR2, GABRA3, GABRB3, GAD1, GALNT13, GFRA1, GLUD1, GLUD2, GNAL, GNAO1, GPR158, GPR27, GPRC5B, GRIA1, GRIA2, GRIA4, GRID1, GRIK4, HDAC5, HEY2, HIP1R, HLF, HS3ST4, HSPA12A, ID4, IKZF5, IL17D, JPH3, JPH4, KCNB1, KCNN3, KCNQ5, KCTD4, D10BWG1379E, KIF1A, KIF21B, KIF5A, KLRC1, KLRC2, KLRC3, KSR2, LGR5, LMF1, E130309F12RIK, LRRC4, LUZP2, MAF, MAPK8IP2, MAPT, MCF2, MMP16, MN1, NALCN, NAP1L3, NCAM1, NDRG2, NET1, NEU4, NKAIN4, NOG, NRG3, NRSN1, NTN4, NTRK2, NUMA1, OLIG1, OLIG2, OMG, OPCML, OVOL1, P2RX7, P2RY13, PARD3, PCSK1N, PCSK6, PDE2A, PDK2, PDK4, PDZD8, PHACTR3, PHYHIPL, PID1, PKNOX2, PKP4, PLCB1, PLEKHB1, PLK1S1, PRKCZ, PSD, PTGDS, RAB11FIP4, RAB6B, RAC3, RAP2A, RAP2B, RASGEF1C, RASL10A, RASSF4,

REPS2, RGS9, RIMS2, RIPPLY2, RPL13-PS3, RPL37, RPL5, RPRM, RTN1, RUNDC3A, SATB1, SCD1, SCG3, SCN3A, SEC31B, SEPT4, SERINC5, SEZ6L, SGCG, SGSM1, SH3GL2, SHD, SLC1A1, SLC1A4, SLIT1, SLITRK2, SLITRK5, SMAD9, SMOC1, SNAP91, SNRPN, SORBS1, SORCS3, SOX6, SOX8, SPHKAP, SSTR1, SSTR2, STOX1, SUSD5, TAL1, THRA, TIMP4, TMEM100, TMEM59L, TMLHE, TMOD2, TNKS2, TPCN2, TPM1, TRIM31, TTYH1, USH1C, WDR86, WNT7B, ZC3H12B, ZCCHC24, ZDHHC22, ZFP488, ZFP804A

Proliferative Full Signature

ABCA5, ABHD3, ACN9, ACYP1, ANKRD32, ANKRD5, ARNTL2, ASPM, ATG12, AURKA, BARD1, BRCA1, BRIP1, BUB1, 4930547N16RIK, 5730455P16RIK, BC055324, , 3830406C13RIK, 4932425I24RIK, 1500031L02RIK, 1700029J07RIK, CACYBP, CBWD1, CCDC34, CCNA2, CCNB1, CCNE2, CDC25C, CDC6, CDCA7, CDK1, CDK2, CDKN2A, CDKN2C, CENPA, CENPE, CENPF, CENPI, CENPK, CENPL, CENPN, CENPW, CEP152, CHAF1A, CHAF1B, CHEK1, CKS2, CREBZF, DBF4, DEK, DHFR, DLGAP5, DONSON, DSN1, DTL, E2F1, E2F7, E2F8, ECT2, EFCAB2, EIF1AX, EMP2, ERCC6L, EVC2, EXOSC9, EZH2, FANCD2, FANCI, FBXO11, FBXO5, GCLM, GINS1, GINS2, GJC1, GMPS, HAUS1, HAUS6, HELLS, HJURP, HMGB2, HMMR, HSPB11, IFT74, IL13RA2, ITGA2, ITGB3BP, 2810417H13RIK, 9530077C05RIK, KIF14, KIF18A, KIF23, KIF4, KNTC1, LMNB1, LRIG3, LSM5, MAD2L1, MAGOH, MAGOHB, MCM2, MCM6, MDFIC, MELK, MIRLET7D, MLF1, MLF1IP, MND1, MNS1, MTF2, NASP, NCAPG, NCAPH, NDC80, NEK2, NPHP1, NUCB2, NUF2, NUSAP1, ORC6, PABPC4L, PAIP1, PAWR, PCNA, PDK1, PEG10, PIN4, PLK4, PPIC, PPIG, LRR1, PRIM2, PRPS2, RAD51, RAD51AP1, RBBP8, RBM24, RECQL, RFC4, RPA3, RRM1, RRM2, SGOL2, SHFM1, SHOX2, SIP1, SLC25A24, SMC2, SMC4, SPC24, SPIN4, STIL, TCF19, TEX9, TIFA, TIMELESS, TMEM106C, TMEM38B, TMEM79, TMPO, TOM1L1, TOP2A, TRIM36, TRMT6, TTC12, TTC26, TTK, TYMS, USP1, WDHD1, WDR34, WDR76, WEE1, XRCC4, YEATS4, ZC3HAV1L, ZFP367, ZWILCH, ZWINT

Mesenchymal Full Signature

ACTA2, ACTN1, ALDH16A1, ANGPT2, ANGPTL4, ANPEP, B4GALT1, BACE2, BCL3, BHLHE40, A430105I19RIK, C1QTNF1, C1RA, C1RL, , CAR12, CD151, CD248, CD274, CD97, CECR2, CHI3L1, COL4A1, COL4A2, DEF6, DLC1, ECE1, EFEMP2, EFNB2, EHD2, EMP1, EMP3, EPAS1, ESM1, FAM20C, FAM38A, FBN1, FES, FGFRL1, FLNA, FLT1, FOSL2, FPR2, GALNT4, GGN, GPR116, HK3, HOMER3, HRH1, HSD3B7, ICAM1, IFITM2, IFITM3, ITGA1, ITGA3, ITGA5, ITGA7, JUNB, KLF16, LIF, LPAR1, LRRC29, LRRC32, LZTS1, MAP2K3, METRNL, METTL7B, MMP14, MVP, MYH9, MYL12A, NCLN, NDUFA10, NEURL2, NRP1, NRP2, OSBPL3, OSMR, PAPPA, PARP10, PDGFA, PDGFRL, PDLIM4, PDLIM7, PDPN, SERPINA1C, PLA2G5, PLAU, PLAUR, PLEKHF1, PMEPA1, PML, POC1B, PRR24, PTRF, PVRL2, RAB34, RHOJ, RRAS, RRBP1, RUNX1, RYR3, GM7665, SALL4, SBNO2, SERPINA1E, SERPINE1, SERPINH1, SGSH, SHC1, SHROOM3, SLC12A9, SLC16A3, SLC22A18, SLC25A37, SLC39A8, SOCS3, SPOCD1, STEAP3, TAGLN, THBD, TIMP1, TNC, TPP1, TRABD, TRIM47, TRIM56, TTC38, TWF1, UNC93B1, VWA1, ZYX

Cahoy

OPC

1190002H23RIK, 3830612M24, PID1, FAM70A, A530047J11RIK, A730017C20RIK, FAM19A2, CACNG4, CALCRL, CAR8, CCND1, CDO1, CHRNA4, CHST11, CNTN6, COL11A1, CSPG4, CSPG5, CXADR, D3BWG0562E, DDAH1, DPYSL3, E130114P18RIK, E130309F12RIK, EMID1, ENC1, ETV5, F2R, FSTL5, GFRA2, GRIA3, HES5, KCND2, KCND3, KLF12, LNX1, LPHN3, LRP1, LRRTM3, MAP3K1, MATN4, MKI67, NETO1, NR2E1, NXPH1, OLFM2, OPRL1, PBK, PCDH20, PDGFRA, PDZRN4, PRKG2, PRRX1, PTGFRN, PTPRZ1, RLBP1, RNF180, RPRM, RRM2, SDC3, SLC35F1, SLC7A3, SLITRK1, SOX11, SPON1, SULF1, TACC2, THSD7B, TMEM100, TOP2A, VCAM1, VSTM2A, DCAF12L1, ZBED4, ZFP36L1

Oligo

1700047M11RIK, 2810468N07RIK, PRR5L, 4930452G13RIK, 5730559C18RIK, NIPAL4, 9630013A20RIK, TMEM88B, ERMN, ADAMTS4, ADSSL1, AI314604, ANLN, PRIMA1, BCAS1, CHN2, CLDN11, CPM, CPOX, CYP27A1, DDC, DOCK10, E130308A19RIK, LPAR1, S1PR5, ELOVL7, ENPP6, ERBB3, EVI2A, FA2H, GAL3ST1, GJC2, GJB1, GJE1, GM98, GPR17, GPR62, GSN, HAPLN2, IL1RAP, IL23A, KNDC1, LGI3, MAG, MAL, MBP, MOBP, MOG, MYO1D, NKX6-2, PDLIM2, PHLDB1, PLA2G4A, PLEKHH1, PLLP, PLP1, PLXNB3, PPAP2C, PPP1R14A, PRKCQ, RFFL, SEMA3D, SGK2, SLC45A3, SOX10, SRPK3, ST18, OPALIN, TNNI1, TRF, TRIM59, TSPAN2, UGT8A, UNC5B

Neuron

SPHKAP, 6330527O06RIK, 9130024F11RIK, IPCEF1, A930009L07RIK, SNHG11, ASPH, C030017B01RIK, CACNA1B, CALB1, CAMK2B, CAMK4, CCK, CDH8, CLSTN2, CRH, CYB561, RBFOX3, DLX1, NECAB1, EPHA7, GABRA1, GABRA5, GABRG2, GAP43, GDA, GLRA2, GPR88, HS3ST2, HTR2C, ICA1L, ICAM5, KCNC2, KCNF1, L1CAM, LPL, MAL2, MEF2C, MYO5B, MYT1L, NAPB, NEFL, NEFM, NELL1, NEUROD6, NOV, NPAS4, NRG3, NTS, ODZ2, OLFR1344, PCSK2, PENK, PGM2L1, PLCXD3, PRDM8, RGS4, SATB2, SCG2, SCN2A1, SLA,

SLC12A5, SLC17A6, SLC6A7, SNAP25, SSTR2, STMN2, SV2B, SYT1, SYT4, TMEM130, TRHDE, TTC9, TTR, VGF, VIP, VSNL1

Astrocyte

BTBD17, SLC1A2, A730056I06RIK, ACOT11, ACSBG1, ADHFE1, AGT, AI464131, ALDOC, ATP1A2, FAM20A, FAM107A, BMPR1B, PREX2, CBS, CCDC80, CHRDL1, CLDN10, CTH, CYBRD1, CYP4F14, CYP4F15, DIO2, S1PR1, EGFR, EMP2, ENTPD2, F3, FGFR3, FMO1, FZD2, GJA1, GJB6, GLDC, GLI3, GM266, GRIN2C, HAPLN1, HTRA1, ID4, KCNE1L, LONRF3, MERTK, MGST1, NTSR2, PAPSS2, PDK4, PLCD4, PPP1R3C, PPP1R3G, PRODH, RFX4, SLC14A1, SLC15A2, SLC1A2, SLC1A3, SLC25A18, SLC4A4, SLC7A10, SLC7A2, SLC9A3R1, SOX9, THRSP, TLCD1, TLR3, FAM176A, TMEM47, TNC, TTPA, AQP4, GFAP, MLC1, PLA2G7, SLC39A12

Cultured Astrocyte

1500015O10RIK, 2810417H13RIK, 6330512M04RIK, 9930013L23RIK, AKAP12, AKR1C14, ANXA1, ANXA2, ANXA3, ASNS, AURKB, BACE2, BMP6, C1QL3, CASQ1, CCDC109B, CCNB2, CD24A, CDK1, CEP55, CKAP2, CNN2, COL3A1, COL5A2, COL8A1, CP, CRABP1, CRLF1, ECM1, ECT2, EMP1, EPHB2, FBLN5, FMOD, GAS2L3, GPR126, GRB10, HSPB1, IFI35, IFITM1, IGF2BP2, IGFBP3, KLHDC8A, LGALS1, LOX, MATN2, MELK, MMP2, NDRG1, NPR3, NUAK1, OCIAD2, OGN, PMAIP1, PRSS23, PTGS2, S100A11, SAMD9L, SEMA3C, SHROOM3, SOSTDC1, SPP1, ST8SIA2, TAGLN, TFPI, TGFBI, TGM2, TNFRSF12A, UBE2

REFERENCES

1. Arkin, A.P. & Schaffer, D.V. Network News: Innovations in 21st Century Systems Biology. *Cell* **144**, 844-849 (2011).

2. Alberts, B. et al. Molecular Biology of the Cell (Garland Science, New York, 2002).

3. Southern, E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517 (1975).

4. Alwine, J., Kemp, D. & Stark, G. Method of detection of specific RNAs in agrose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5350-5354 (1977).

5. MAQC Consortium et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproduciblity of gene expression measurements. *Nature Biotechnology* **24**, 1151-1161 (2006).

6. Shalon, D., Smith, S. & Brown, P. A DNA microarray system for analzying complex DNA samples sing two-color flourescent probe hybridization. *Genome Research* **6**, 639-645 (1996).

7. Snijders, A. et al. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genetics* **29**, 263-264 (2001).

8. Ren, B. et al. Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309 (2000).

9. Bulyk, M., Gentalen, E., Lockhart, D. & Church, G. Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nature Biotechnology* **17**, 573-577 (1999).

10. Sanger, F. & Coulson, A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 441-448 (1975).

11. Lander, E. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

12. Mouse Genome Sequencing Consortium et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).

13. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Research* **11**, 3-11 (2001).

14. Garber, M., Grabherr, M.G., Guttman, M. & Trapnell, C. Computational methods for transciptome annotation and quantification using RNA-seq. *Nature Methods* **8**, 469-477 (2011).

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

16. Johnson, D., Mortazavi, A., Myers, R. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-502 (2007).

17. Core, J., Waterfall, J. & JT, L. Nascent RNA sequencing reveals wide spread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848 (2008).

18. Lieberman-Aiden, E. et al. Comphrehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

19. Hsiao, A., Worrall, D., Olefsky, J. & Subramaniam, S. Variance-modeled posterior inference of microarray data: detecting gene-expression changes in 3T3-L1 adiopocytes. *Bioinformatics* **20**, 3108-3127 (2004).

20. Yang, Y. et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* **30**, e15 (2002).

21. Shippy, R. et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nature Biotechnology* **24**, 1123-1131 (2006).

22. Boldstad, B., Irizarry, R., Astrand, M. & Speed, T. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).

23. Bolstad, B., Irizarry, R., Astrand, M. & Speed, T. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).

24. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to quantification of differential expression. *Bioinformatics* **18**, S96-S104 (2002).

25. McLeod, I. in Wolfram Demonstrations Project.

26. Klebanov, L., Qui, X., Welle, S. & Yakovlev, A. Statistical methods and microarray data. *Nature Biotechnology* **24**, 1151-1161 (2007).

27. Subramaniam, S. & Hsiao, G. Gene-expression measurement: variance-modeling considerations for robust data analysis. *Nature Immunology* **13**, 199-203 (2012).

28. Tusher, V., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116-5121 (2001).

29. Smyth, G. Linear models and emperical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article 3 (2004).

30. Baldi, P. & Long, A. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17**, 509-519 (2001).

31. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* **57**, 289-300 (1995).

32. Aliison, D. et al. A mixture model approach for the analysis of microarray gene expression dasta. *Computational Statistics and Data Analysis* **39**, 1-20 (2002).

33. Eisen, M., Spellman, P., Brown, P. & Botsein, D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14863-14868 (1998).

34. Beer, M. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117**, 185-198 (2004).

35. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M. Systematic determination of genetic network architecture. *Nature Genetics* **22**, 281-285 (1999).

36. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29 (2000).

37. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interperting genome-wide expression profiles. *Proceedings*

of the National Academy of Sciences of the United States of America **102**, 15545-15550 (2005).

38. Irizarry, R., Wang, C., Zhou, Y. & Speed, T. Gene set enrichment analysis made simple. *Stat Methods Med Res* **18**, 565-575 (2009).

39. Tamayo, P.S., G, Liberzon, A. & Mesirov, J. Gene Set Enrichment Analysis Made Right. *Stat Methods Med Res* (In submission).

40. Bar-Joseph, Z. et al. Computational discovery of gene modules and regulatory networks. *Nature Biotechnology* **21**, 1337-1342 (2003).

41. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498-2508 (2003).

42. Segal, E. et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**, 166-176 (2003).

43. Canales, R. et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nature Biotechnology* **24**, 1115-1122 (2006).

44. Crick, F. Cetral dogma of molecular biology. *Nature* **227**, 561-563 (1970).

45. Edelman, G. & Gally, J. Degeneracy and complexity in biological systems. *Proc Natl Acad Sci* **98**, 13763-1768 (2001).

46. Stormo, G. DNA binding sites: representaiton and discovery. *Bioinformatics* **16**, 16-23 (2000).

47. Benos, P., Bulyk, M. & Stormo, G. Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Res* **30**, 4442-4451 (2002).

48. Zhao, Y. & Stormo, G. Quantitaive analysis demonstrates most transcription factors required only simple models of specificity. *Nat Biotechnol* **29**, 483-484 (2011).

49. Sen, R. & Baltimore, D. Multilpe nuclear factors interact with the immunoglobulin enchancer sequences. *Cell* **46**, 705-716 (1986).

50. Pahl, H. Activators and target genes of Rel/NF-kappaB transcription factors. *Oncogene* **18**, 6853-6866 (1999).

51. Hayden, M. & Ghosh, S. Signaling to NF-kappaB. *Genes Dev* **18**, 2195-2224 (2004).

52. Oeckinghaus, A., Hayden, M. & Ghosh, S. Crosstalk in NF-kB signaling pathways. *Nat Immunol* **12**, 695-708 (2011).

53. Huang, B., Yang, X., Lamb, A. & CHen, L. Posttranslational modificaitons of NF-kappaB: another layer of regulation for NF-kappaB signaling pathway. *Cell Signal* **22**, 1282-1290 (2010).

54. Li, Q. & Verma, I. NF-kappaB regulation in the immune system. *Nat Rev Immunol* **2**, 725-734 (2002).

55. Chen, F., Huang, D., Chen, Y. & Ghosh, G. Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature* **391**, 410-413 (1998).

56. Siggers, T. et al. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-kB family DNA binding. *Nat Immunol* **13**, 95-102 (2011).

57. Hoffmann, A., Levchenko, A., Scott, M. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245 (2002).

58. Thanos, D. & Maniatis, T. Identification of the rel family members required for virus induction of the human beta interferon gene. *Mol Cell Biol* **15**, 152-164 (1995).

59. The Gilmore Lab. (2012).

60. Pointing, C. The functional repertoires of metazoan genomes. *Nat Rev Genet* **9**, 689-698 (2008).

61. Wilson, M. et al. Species-specific transcription in mice carrying human chromsome 21. *Science* **322**, 434-438 (2008).

62. Kunsch, C., Ruben, S. & Rosen, C. Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. *Mol Cell Biol* **12**, 4412-4421 (1992).

63. Keich, U. & Pevzner, P. Subtle motifs: defining the limits of motif finding algorithms. *Bioinformatics* **18**, 1382-1390 (2002).

64. Kim, H., Shay, T., O'Shea, E. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325**, 429-432 (2009).

65. Jenner, R. & Young, R.A. Insights into host reponses against pathogens from transcriptional profiling. *Nature Reviews Microbiology* **3**, 281-294 (2005).

66. Nau, G. et al. Human macrophage activaiton programs induced by baterial pathogens. *Proc Natl Acad Sci* **99**, 1503-1508 (2002).

67. Zhu, X. et al. Analysis of the major patterns of B cell gene expression changes in response to short-term stimulation with 33 single ligands. *J Immunol* **173**, 7141-7149 (2004).

68. Lee, T. et al. Deciphering gene expression regulatory networks. *Curr Opin Genet Dev* **12**, 130-136 (2002).

69. Saccani, S., Pantano, S. & Natoli, G. Modulation of NF-kappaB activity by exchange of diers. *Mol Cell* **11** (2003).

70. Martone, R. et al. Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci* **100**, 12247-12252 (2003).

71. Lim, C. et al. Genome-wide mapping of RELA(p65) binding identifies E2F1 as a transcriptional activator recruited by NF-kappaB upon TLR4 activation. *Mol Cell* **27** (2007).

72. Schreiber, J. et al. Coordinated binding of NF-kappaB members in the response of human cells to lipopolysaccharide. *Proc Natl Acad Sci* **103**, 5899-5904 (2006).

73. Kasowski, M. et al. Variation in transcription factor binding among humans. *Science* **328**, 232-235 (2010).

74. Leung, T., Hoffmann, A. & Baltimore, D. One Nucleotide in a κB Site Can Determine Cofactor Specificity for NF-κB Dimers. *Cell* **118**, 453-464 (2004).

75. Antonaki, A. et al. Genomic analysis reveals a novel nuclear factor-κB (NF-κB)-binding site in Alu-repetitive elements. *J Biol Chem* **286**, 38768-38782 (2011).

76. Smale, S. Hierachies of NF-kB target-gene regulation. *Nat Immunol* **12**, 689-694 (2011).

77. Heintzman, N. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enchancers in the human genome. *Nat Genet* **39**, 311-318 (2007).

78. Ghisletti, S. et al. Identification and characterization of enchancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**, 317-328 (2010).

79. Heinz, S. et al. Simple combinations of lineage-determining transcripton factors primer cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).

80. Dixon, J. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

81. Amit, I. et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257-263 (2009).

82. Berman, B. et al. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci* **99**, 757-762 (2002).

83. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in Drosophila segementation. *Nature* **451**, 535-540 (2008).

84. Halliwell, B. Free radicals and antioxidants: a personal view. *Nutr Rev* **8**, 253-265 (1994).

85. Chandra, J., Samali, A. & Oreenius, S. Triggering and moldulation of apoptosis by oxidative stress. *Free Radical Biology and Medicine* **3-4**, 323-333 (2000).

86. Curtin, J., Donovan, M. & Cotter, T. Regulation and measurement of oxidative stress in apoptosis. *J Immunol Methods* **265**, 49-72 (2002).

87. Ozben, T. Oxidative stress and apoptosis: impact on cancer threapy. *J Pharm Sci* **96**, 2181-2196 (2007).

88. Rojkind, M., Domínguez-Rosales, J., Nieto, N. & Greenwel, P. Role of hydrogen peroxide and oxidative stress in healing response. *Cell Mol Life Sci* **59**, 1872-1891 (2002).

89. Rhee, S., Bae, Y., Lee, S. & Kwon, J. Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci STKE* **2000**, pe1 (2000).

90. Rhee, S. Cell signaling: H202, a neccessary evil for cell signaling. *Science* **312**, 1882-1883 (2006).

91. Ramachandran, S., Xie, L., John, S., Subramaniam, S. & R, L. A novel role for connexin hemichannel in oxidative stress and smoking-induced cell injury. *PLoS ONE* **2**, e712 (2007).

92. Barzilai, A. & Yamamoto, K. DNA damage responses to oxidative stress. *DNA Repair (Amst)* **3**, 1109-1115 (2004).

93. Bao, S. et al. Glioma stem cells promote radioresitance by preferential activaiton fo DNA damage response. *Nature* **444**, 756-760 (2006).

94. Azzam, E., de Toldeo, S. & Little, J. Oxidative metabolism, gap junctions and the ionizing radiation-induced bystander effect. *Oncogene* **22**, 7050-7057 (2003).

95. Starklov, A. The role of mitochondria in reactive oxygen species metabolism and signaling. *Ann N Y Acad Sci* **1147**, 37-52 (2008).

96. Imlay, J. & Linn, S. DNA damage and oxygen radical toxicity. *Science* **240**, 1302-1309 (1988).

97. Hentze, M., Muckenthaler, M., Galy, B. & Camaschella, C. Two to Tango: regulation of mammalian iron metabolism. *Cell* **142**, 24-38 (2010).

98. Lambeth, J. NOX enzymes and the biology of reactive oxygen. *Nat Rev Immunol* **4**, 181-189 (2004).

99. Fukai, T. & Ushio-Fukai, M. Superoxide dismutases: role in redox signaling, vascular funciton, and diseases. *Antoxid Redox Signal* **15**, 1583-1606 (2011).

100. Immenschuh, S. & Baumgart-Vogt, E. Peroxiredoxins, oxidative stress, and cell proliferaiton. *Antoxid Redox Signal* **7**, 768-777 (2005).

101. Neumann, C. & Fang, Q. Are peroxiredoxins tumor suppressors? *Curr Opin Pharmacol* **7**, 375-380 (2007).

102. Lu, J. & Holmgren, A. Selenoproteins. *J Biol Chem* **284**, 723-7 (2009).

103. Putnam, C., Arvai, A., Bourne, Y. & Tainer, J. Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism. *J Mol Biol* **296**, 295 (2000).

104. Runchel, C., Matsuzawa, A. & Ichijo, H. Mitogen-activated protein kineases in mammalian oxidative stress reponses. *Antoxid Redox Signal* **15**, 205-218 (2011).

105. Ray, P., Huang, B. & Tsuji, Y. Reactive oxygen species (ROS) homeostatis and redox regulation in cellular signaling. *Cell Signal* **24**, 981-990 (2012).

106. Dhakshinamoorthy, S., Jain, A., Bloom, D. & Jaiswal, A. Bach1 competes with Nrf2 leading to negative regulation of the antioxidant response element (ARE)-mediated NAD(P)H:quinone oxidoreductase 1 gene expression and induction in response to antioxidants. *J Biol Chem* **280**, 16891-16900 (2005).

107. Matsuoka, S. et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160-1166 (2007).

108. Bernstein, C., Bernstein, H., Payne, C. & Garewal, H. DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat Res* **511**, 145-178 (2002).

109. Scorrano, L. Opening the doors to cytochrome c: changes in mitochondrial shape and apoptosis. *Int J Biochem Cell Biol* **41**, 1875-1883 (2009).

110. Susnow, N., Zeng, L., Margineantu, D. & Hockenbery, D. BCL-2 family proteins as regulators of oxidative stress. *Semin Cancer Biol* **19**, 42-49 (2009).

111. Ow, Y., Green, D., Hao, Z. & TW, M. Cytochrome c: functions beyond respiration. *Nat Rev Mol Cell Biol* **9**, 532-542 (2008).

112. Ho, J., Asagiri, M., Hoffmann, A. & G, G. NF-κB potentiates caspase independent hydrogen peroxide induced cell death. *PLoS ONE* **6**, e16815 (2011).

113. Cereghetti, G. & Scorrano, L. The many shapes of mitochondrial death. *Oncogene* **25**, 4717-4724 (2006).

114. Suen, D.-F., Norris, K. & Youle, R. Mitochondrial dynamicsl and apoptosis. *Genes & Development* **22**, 1577-1590 (2008).

115. Budanov, A. & Karin, M. p53 targets genes sestrin1 and sestrin2 connection genotoxic stress and mtor signaling. *Cell* **134**, 451-460 (2008).

116. Hamel, P., Corvest, V., Giegé, P. & Bonnard, G. Biochemical requirements for the maturation of mitochondrial c-type cytochromes. *Biochimica et Biophysica Acta* **1793**, 125-138 (2009).

117. Liu, X., Kim, C., Yang, J., Jemmerson, R. & Wang, X. Induction of apoptosis program in cell-free extracts; requirement of dATP and cytochrome C. *Cell* **86**, 147-157 (1996).

118. Morse, D., Lin, L., Choi, A. & Ryter, S. Heme oxygenase-1, a critical arbitrator of cell death pathways in lung injury and disease. *Free Radical Biology and Medicine* **47**, 1-12 (2009).

119. Dixon, S. et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* **149**, 1060-1072 (2012).

120. Takahashi, S. & Masuda, T. High throughput heme assay by detection of chemiluminescence of reconstituted horseradish peroxidase. *Comb Chem High Throughput Screen* **12**, 532-535 (2009).

121. Byrnes RW, C.D., Maer A, Li J, Nadeau D, Subramaniam S. An editor for pathway drawing and data visualization in the Biopathways Workbench. *BMC Syst Biol* **3**, 99 (2009).

122. Hanahan, D. & Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-74 (2011).

123. Hanahan, D. & Weinberg, R.A. The Hallmarks of Cancer. *Cell* **100**, 57-70 (2000).

124. Vogelstein, B. & Kinzler, K.W. Cancer genes and the pathways they control. *Nature Medicine* **10**, 789-799 (2004).

125. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. Ras oncogenes: weaving a tumorigenic web. *Nature Reviews Cancer* **11**, 761-774 (2011).

126. Scheffzek, K. et al. The Ras-RasGAP complex: structural basis for GTPase actibaiton and its loss in oncogenic Ras mutants. *Science* **277**, 333-338 (1997).

127. Musgrove, E.A., Caldon, C.E., Barraclough, J., Stone, A. & Sutherland, R.L. Cyclin D as a threapeutic target in cancer. *Nature Reviews Cancer* **11**, 558-572 (2011).

128. Meek, D.W. Tumor suppression by p53: a role for the DNA damage response? *Nature Reviews Cancer* **9**, 714-723 (2009).

129. Brady, C.A. et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor supression. *Cell* **145**, 571-583 (2011).

130. Martins, C., Brown-Swigart, L. & GI, E. Modeling the threapeutic efficacy of p53 restoration in tumors. *Cell* **127** (2006).

131. Xue, W. et al. Senescence and tumor clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656-660 (2007).

132. Ventura, A. et al. Restoration of p53 function leads to tumor regression in vivo. *Nature* **445**, 661-665 (2007).

133. Hanahan, D., Wagner, E.F. & Palmiter, R.D. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. *Genes & Development* **21**, 2258-2270 (2007).

134. Dyke, T.V. & Jacks, T. Cancer Modeling in the Modern Era: Progress and Challenges. *Cell* **108**, 135-144 (2002).

135. Lasko, M. et al. Targeted oncogene activation by site-specific reombination in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6232-6236 (1992).

136. St-Onge, L., Furth, P. & Gruss, P. Temporal control of the Cre recombinase in transgenic mice by a tetracycline reponsive promoter. *Nucleic Acids Research* **24**, 3875-3877 (1996).

137. Nagy, A., Mar, L. & Watts, G. Creation and Use of a Cre Recombinase Transgenic Database. *Methods in Molecular Biology* **530**, 365-378 (2009).

138. Kinzler, K. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-170 (1996).

139. Visvader, J. Cells of origin in cancer. *Nature* **469**, 314-322 (2011).

140. Zong, H., Epspinosa, J., Su, H., Muzumdar, M. & Liqun, L. Mosiac Analysis with Double Markers in Mice. *Cell* **121**, 479-492 (2005).

141. Naldini, L. et al. In vivo gene delivery and stable transduction of nondiving cells by a lentiviral vector. *Science* **272**, 263-267 (1996).

142. Naldini, L., Blomer, U., Gage, F., Trono, D. & Verma, I. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proc Natl Acad Sci* **93**, 11382-11388 (1996).

143. Cavazzana-Calvo, M. et al. Gene therapy of human severe combined immunodeificiency (SCID)-X1 disease. *Science* **288**, 627-629 (2000).

144. Hacein-Bey, S. et al. A serious adverse event after successful gene threapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**, 255-256 (2003).

145. Woods, N., Bottero, V., Schmidt, M., Von Kalle, C. & Verma, I. Gene therapy: therapeutic gene causing lymphoma. *Nature* **440**, 1123 (2006).

146. DuPage, M., Dooley, A. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nature Protocols* **4**, 1064-1072 (2009).

147. Xia, Y. et al. Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model. *Nature Cell Biology* **14**, 257-265 (2012).

148. Kong, J. et al. Integrative, Multi-modal Analysis of Glioblastoma Using TCGA Molecular Data, Pathology Images, and Clinical Outcomes. *IEEE Trans Biomed Eng* **58**, 3469-3474 (2012).

149. Chang, H. et al. Morphometic analysis of TCGA glioblastoma multiforme. *BMC Bioinformatics* **12**, 484 (2011).

150. Beroukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905 (2010).

151. Golub, T. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).

152. Alizadeh, A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-5011 (2000).

153. van 't Veer, L. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-536 (2002).

154. Ramaswamy, S. et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci* **98**, 15149-15154 (2001).

155. Madhavan, S. et al. Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* **7**, 157-167 (2009).

156. Cancer Genome Atlas Research Network. Integrated genomic analysis of ovarian carcinoma. *Nature* **474**, 609-615 (2011).

157. Ding , L. et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069-1075 (2008).

158. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068 (2008).

159. Verhaak, R. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98-110 (2010).

160. Johnson, L. et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410**, 1111-1116 (2001).

161. Basseres, D., Ebbs, A., Levantini, E. & Baldwin, A. Requirement of the NF-kB subunit p65/RelA for K-RAS-Induced Lung Tumorigenesis. *Molecular and Cellular Pathobiology* **70**, 3537-3546 (2010).

162. Meylan, E. et al. Requirement for NF-kB signalling in a mouse model of lung adenocarcinoma. *Nature* **462**, 104-108 (2009).

163. Sweet-Cordero, A. et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nature Genetics* **37**, 48-55 (2005).

164. Marumoto, T. et al. Development of a novel mouse glioma model using lentiviral vectors. *Nature Medicine* **15**, 110-116 (2009).

165. Friedmann-Morvinski, D. et al. Dedifferentiation of astrocytes and neurons by oncogenes can induce glioblastomas. (In submission).

166. Furnari, F. et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* **21**, 2683-2710 (2007).

167. Dunn, G. et al. Emerging insights into the molecular and cellular basis of glioblastoma. *Genes Dev* **26**, 756-784 (2012).

168. Liu, C. et al. Mosiac Analysis with Double Markers Reveals Tumor Cell of Origin in Glioma. *Cell* **146**, 209-221 (2011).

169. Beroukhim, R. et al. Assessing the significance of chromosomal abberations in cancer: methodology and application to glioma. *Proc Natl Acad Sci* **104**, 20007-20012 (2007).

170. Phillips, H. et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157-173 (2006).

171. Lei, L. et al. Glioblastoma models reveal the conneciton between adult glial progenitors and the proneural phenotype. *PLoS ONE* **6**, 1-13 (2011).

172. Cahoy, D. et al. A Transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *The Journal of Neuronscience* **28**, 264-278 (2008).

173. Marko, N., Quackenbush, J. & Weil, R. Why is there a lack of consensus on molecular subgroups of glioblastoma? Understanding the nature of biological and statistical variability in glioblastoma expression data. *PLoS ONE* **6**, 1-19 (2011).

174. Pe'er, D. & Hacohen, N. Principles and strategies for developing network models in cancer. *Cell* **144**, 864-873 (2011).

175. Basso, K. et al. Reverse engineering of regulatory networks in B cells. *Nature Genetics* **37**, 382-90 (2005).

176. Carro, M. et al. The transcriptional network for mesenchymal transformation of brain tumors. *Nature* **463**, 318-325 (2010).

177. Wang, X., Verhaak, R., Purdom, E., Spellman, P. & Speed, T. Unifying gene expression measures from multiple platforms using factor analysis. *PLoS ONE* **11**, 3 (2011).

178. de Hoon, M., Imoto, S., Nolan, J. & Miyano, S. Open sources clustering software. *Bioinformatics* **20**, 1453-1454 (2004).

179. Saldanha, A. Java Treeview - extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248 (2004).