UNIVERSITY OF CALIFORNIA

Los Angeles

Recurrent Aneuploidy Patterns Enable Fitness Gains in Tumor Metabolism

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biomedical Physics

by

Aspram Minasyan

2015

ABSTRACT OF THE DISSERTATION

Recurrent Aneuploidy Patterns

Enable Fitness Gains in Tumor Metabolism

by

Aspram Minasyan

Doctor of Philosophy in Biomedical Physics

University of California, Los Angeles, 2015

Professor Thomas G. Graeber, Chair

Copy number alteration (CNA) profiling of human tumors has revealed recurrent patterns of DNA amplifications and deletions. These patterns are indicative of conserved selection pressures, but cannot be fully explained by known oncogenes and tumor suppressor genes. Using integrative analysis of CNA data from patient tumors and experimental systems, we report that principal component analysis-defined CNA signatures are predictive of glycolytic phenotypes, including FDG-avidity of patient tumors, and increased proliferation. The primary glycolysis-linked CNA signature is associated with p53 mutation and shows coordinate amplification of glycolytic genes and other cancer-linked metabolic enzymes including TIGAR and RPIA. In contrast, alternative signatures involve both different mechanisms of tumor suppression loss (eg, MDM2 amplification) and different glycolysis enzyme isoforms. Furthermore, a cross-species CNA comparison identified 21 conserved CNA regions, containing 13 enzymes in the glycolysis and pentose phosphate pathways in addition to known cancer

driving genes. In validation experiments, exogenous expression of hexokinase and enolase enzymes resulted in reduced propensities for amplifications at the corresponding endogenous loci. Our findings support metabolic stress as a selective pressure underlying the recurrent CNAs observed in human tumors, and further cast genomic instability as an enabling event in tumorigenesis and metabolic evolution.

The dissertation of Aspram Minasyan is approved.

Arion-Xenofon Hadjioannou

Anna Wu Work

Heather Christofk

Thomas G. Graeber, Committee Chair

University of California, Los Angeles

2015

To my most patient collaborators and most charitable critics -

my husband Artur, my daughter Veronica, my son Arno and my lovely parents.

Without your support I wouldn't have gotten so far...

Table of Contents

List of Figures

List of Tables and Equations

Acknowledgments

I would like to thank the funding agencies that allowed this work to happen. A.M. was supported by UCLA Competitive Edge Fellowship (2010), UCLA Eugene V. Cota-Robles Fellowship (2010-2011; 2013-2014) and UCLA Dissertation Year Fellowship (2014-2015). N.A.G. was supported by the UCLA Scholars in Oncologic Molecular Imagining program (NCI/NIH grant R25T CA098010). T.G.G. is supported by the NCI/NIH (P01 CA168585, P50 CA086306, U19 AI067769), an American Cancer Society Research Scholar Award (RSG-12-257-01-TBE), a Melanoma Research Alliance Established Investigator Award (20120279), the Norton Simon Research Foundation, the UCLA Jonsson Cancer Center Foundation, the National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR000124, the UC Cancer Research Coordinating Committee, a Concern Foundation CONquer CanCER Now Award, and the UCLA Stein/Oppenheimer Endowment.

I would like to express my deepest gratitude to my advisor, Dr. Thomas Graeber, for his helpful and inspirational guidance, for his thoughtful and respectful mentorship, and for turning my graduate school experience into a breathtaking and joyful adventure. Special Thanks to Dr. Nicholas Graham who patiently dealt with my stubbornness and with whom we spent so many fun hours working on this project. Also, this project would have never been a success without the help of all of my wonderful labmates, especially Anastasia Lomova and Ashley Cass.

I received much advice and had many helpful discussions with my committee members Dr. Arion-Xenofon Hadjioannou, Dr. Heather Christofk and Dr. Anna Wu Work. I appreciate your time and dedication in helping me achieve my goals. Special thanks to Dr. Nzhde Agazaryan and Dr. Gregory Kivenzor for their continuous support through friendship and useful advices.

Thank you for the administrative support of Terry Moore, Reth Thach, Stacey Chiong, and Erika Corrin. With the special gratitude to Terry Moore for all of her inspirational and motivational help and for becoming an important part of my family's and mine lives. Many thanks to the director of Biomedical Physics program, Dr. Michael McNitt-Gray for making me feel home in the program.

I would like to thank my Biomedical Physics classmates Yanisley Valenciaga, Yazeed Hiyari, Robert Harris and Brendon Villegas, for the good times we spent together and for the support during my years of graduate school.

My road to graduate school was possible thanks to special people who surrounded me during my undergraduate studies. I am very thankful to my undergraduate advisor and one of the best mathematics teachers Dr. Katherine Stevenson. I also would like to thank Dr. Duane Doty and Dr. Say-Peng Lim, the two amazing physics professors who struggled with me during and after my undergraduate career. Also, I would like acknowledge my high school math teacher, Larisa Ivanovna, who first opened me the beauty of science.

Finally, I would like to thank my family - you are the ones who make me believe that everything is possible.

Vita

Education:

Sep 2010 – Present, University of California, Los Angeles, CA

Advanced to Ph.D Candidacy (Mar 2014) M.S, Biomedical Physics (Dec 2013)

Sep 2003 – Dec 2008, California State University, Northridge, CA

- B.A, Biology, *Cum Laude* - B.A, Physics, *Cum Laude* - B.S, Mathematics/Statistics, *Cum Laude*

Research Experience:

- Aug 2011 – Present and Mar 2010 – Sep 2010, Systems Biology of Cancer Signaling, Thomas Graeber Lab, Dept. of Molecular and Medical Pharmacology, UCLA, Crump Institute for Molecular Imaging.
- Mar 2011 – Jun 2011, Molecular Genetics of Ataxia-Telangiectasia, Richard Gatti, Dep. of Pathology and Laboratory Medicine, UCLA.
- Jan 2006 – Jun 2007, Learning Taxonomies of the Visual World. Mentor: Katherine F. Stevenson, Dep. of Mathematics, CSU, Northridge. Co-mentor: Pietro Perona, Dep. of Electrical Engineering and Computation and Neural Systems, California Institute of Technology.

Other Experience:

- 2015, graduate student representative at the Department of Mol. and Med. Pharmacology, UCLA.
- Jan 2014 – May 2014, part-time faculty position teaching “Computer Application in Health Sciences” course for the Department of Health Sciences at CSU, Northridge.
- Jul 2013, teaching assistant in AACR Annual Aspen Workshop on Mol. Biology in Clinical Oncology.
- Aug 2006 – Jan 2006 and Aug 2007 – Jan 2007, grading assistant for Dep. of Mathematics, CSUN.
- Jan 2007 – Jun 2007, tutor for Dep. of Mathematics, CSUN.

Grants and Fellowships:

- Dissertation Year Fellowship, UCLA, 2014-2015.
- Eugene V. Cota-Robles Fellowship, UCLA, 2010-2011 and 2013-2014.
- The Competitive Edge Program, UCLA, Aug 2010.
- PUMP Program, CSU, Northridge, Spring 2007.

Awards and Honors:

- 14th Annual Biomedical Physics Research Colloquium: best oral presentation award, May, 2013.
- A member of Sigma Pi Sigma Nat. Physics Honor Society. Received into the Chapter May 16, 2008.
- First Place Winner, Undergraduate Division, Student Research Competition, 2006.

Publications:

- Nicholas A Graham*, **Aspram Minasyan***, Anastasia Lomova, Ashley Cass, Nikolas G Balanis, Michael Friedman, Shawna Chan, Sophie Zhao, Adrian Delgado, James Go, Lillie Beck, Christian Hurtz, Carina Ng, Rong Qiao, Johanna ten Hoeve, Nicolaos Palaskas, Hong Wu, Markus Müschen, Elisa Port, Steven M Larson, Nikolaus Schultz, Daniel Braas, Heather R Christofk†, Ingo K Mellinghoff†, and Thomas G Graeber, et. al. Recurrent aneuploidy patterns enable fitness gains in tumor metabolism (In progress). * equal contribution; † equal contribution.

- Koya RC*, Mok S*, Otte N, Blacketor KJ, Comin-Anduix B, Tumeh PC, **Minasyan A**, Graham NA, Graeber TG, Chodon T, Ribas A. BRAF Inhibitor Vemurafenib Improves the Antitumor Activity of Adoptive Cell Immunotherapy. Cancer Res. 2012;72(16):3928–3937. * equal contribution.

Presentations:

- UCLA Annual Biomedical Physics Research Colloquium:
    - Talk, May 2015 and May 2013.
- UCLA Molecular and Medical Pharmacology Annual Retreat:
    - Talk, Nov 2014.
    - Poster, Nov 2013, Nov 2012, Nov 2011, Oct 2010.
- UCLA Crump Student and Post-doc Floor Talks:
    - Talk, Dec 2013 and Feb 2013.
- Undergraduate:
    - Talk and Poster (Spring 2007). PUMP fellowship requirements at CSUN.
    - Talk (Fall 2006). Undergraduate Student Research Competition at Caltech.
    - Talk (Spring 2006). Undergraduate Student Research Competition at CSUN.

CHAPTER 1

Introduction

"Healing is a matter of time, but it is sometimes also a matter of opportunity."

-Hippocrates

1.1 Background

A British historian and scientist June Goodfield has noted "Cancer begins and ends with people". It is one of the most ancient diseases known to human kind. However, despite our tremendous gain in knowledge, cancer continues to pose questions. The first evidence of a human tumor was found among fossilized bone in an Egyptian mummy, while the first documented description of the disease dates back to 3000 BC. Approximately 2700 years later, the Greek physician Hippocrates gave the disease of uncontrolled division of abnormal cells a name, *carcinos* and *carcinoma* (Greek for crab), for the finger-like spreading projections of tumors. Later the Roman physician, Celsus, translated the word into *cancer*. Another commonly used name for cancer was given by Greek physician Galen. He termed it *oncos*, which translates to "swelling". The first cancer surgery was documented in the 18th century, and by the 19th century, pathologists were able to determine whether tumor resection had completely removed cancer. With discoveries and breakthroughs in tumor diagnostics and non-surgical therapies, the 20th century was a time of major advancement in oncology. The various therapies that were developed during this time include but are not limited to: radiation therapy, which uses ionizing radiation to damage and kill cancer cells; chemotherapy, in which the growth of fast growing and quickly dividing cancer cells is stopped or slowed by drugs; hormone therapy, which affects

the ability of cancer cells to multiply by addition, blockage or removal of hormones; and immunotherapy, which uses the body's immune system to fight cancer.

Nevertheless, despite our progress, cancer remains the second leading cause of death in the United States and other parts of the world, currently yielding only to heart diseases. However, the number of deaths caused by cancer in the U.S. is expected to surpass the number of deaths due to heart disease within the next few years (1). Moreover, the lifetime probability of developing an invasive cancer is 43% for men and 38% for women (1). It is evident from these statistics that better cancer treatments, resulting from a more thorough understanding of cancer progression, are greatly needed. In order to address these gaps in knowledge, the research project presented in this work is aimed at elucidating the underlying processes that either trigger or occur during tumorigenesis.

Using an integrative analysis of copy number alterations (CNA) data from human tumors, cancer cell lines and an experimental immortalization system, in this work we show that the loci of metabolic genes impact the recurrent CNA changes observed in genomically unstable tumors. Our bioinformatic and experimental results support a tumorigenesis model in which copy number changes in metabolic genes contribute toward an enhanced glycolytic and proliferative state.

1.2 The Hallmarks of cancer

Cancer cells differ from normal cells in that they exhibit aberrant proliferation, resist apoptosis, and invade other tissues (2). Various mechanisms and properties of tumor cells have been discovered and described within the past few decades. By integrating this complex body of knowledge, Hanahan and Weinberg have proposed six hallmarks that are fundamental to most, if not all, neoplastic disease (2,3). First, all cancer cells need to find a way to sustain

proliferative signaling. Cancer cells can achieve this by overexpressing growth factors or by increasing growth factor receptor expression. Next, cancer cells also need to escape downregulated proliferation. Evasion of growth suppression can be accomplished by loss of tumor suppressor function such as retinoblastoma-associated, RB, or TP53. The third hallmark observed in oncological malignancies is activation of invasion and metastasis, which tumorigenic cells often achieve via the epithelial-mesenchymal transition (EMT). Another hallmark of cancer is replicative immortality. Malignant cells achieve this by maintaining long telomeric DNA. In addition, tumors must also induce angiogenesis in order to provide their quickly growing mass with necessary nutrients and oxygen. Finally, the above properties would not benefit the cancer cells in the event of cell death. Therefore, cancer cells find many mechanisms for resisting apoptosis, autophagy and necrosis.

Nearly a decade after their initial review, Hanahan and Weinberg revised their understanding of tumorigenesis and proposed two new categories of hallmarks in addition to the original six (2). They defined ‘enabling characteristics’, or functions that are acquired by cancer cells during tumorigenesis. An example of this category is tumor-promoting inflammation that is driven by the cells of the immune system and genomic instability and mutation in cells that lead to exceptionally abnormal cancer cells. The other category of hallmarks, the ‘emerging hallmarks’, defines the properties that facilitate development and progression of tumors. For instance avoiding destruction by the immune cells and reprogramming of energy metabolism. In the context of the hallmarks of cancer, the work in this dissertation touches primarily on the interplay that allows genomic instability to enable the evolution of tumor metabolism.

1.3 Reprogramming of energy metabolism in cancer

One of the fundamental and consequential differences between non-transformed and tumorigenic cells is the re-programming of cellular metabolism (2). Tumor cells consume glucose and glutamine at a surprisingly high rate and secrete most of the glucose-derived carbon as lactate rather than oxidizing it completely. First observed by Otto Warburg, this altered metabolism of tumors is thought to benefit transformed cells in several ways. Upregulation of glucose metabolism allows proliferating cells to meet their energy demand through synthesis of adenosine triphosphate (ATP), while increased flux through glycolysis branch pathways provides dividing cells with intermediates necessary for biosynthesis of nucleotides and fatty acids, as well as reducing agents such as NADPH (4,5). Moreover, in addition to glucose, cancer cells frequently upregulate consumption of other metabolites for energy and biomass generation, including glutamine, serine and glycine (6,7). Notably, several individual metabolic enzymes have been directly implicated in tumorigenesis (8–15) and/or immortalization (16–18), suggesting that altered metabolism is not a passive bystander, but rather a driving force of oncogenesis (19,20).

Additional evidence has demonstrated that many oncogenes drive the high metabolic demands of cancer cells. A prime example is activation of Akt which is likely the most important signaling event in terms of cell metabolism, because Akt is sufficient to drive glycolysis and lactate production and to suppress macromolecular degradation in cancer cells (21). Finally, decreased oxygen availability (hypoxia) stimulates cells to consume glucose and produce lactate (22). Although these individual events are clearly important in the regulation of tumor metabolism, our analysis suggest that coordinated alterations of DNA copy number should be added to the list of causal mechanisms – thus expanding our systems-view of this classic phenomenon.

1.4 Genomic instability

Modern cancer classification relies on molecular characterization, including examination of genomic DNA mutations and copy number alterations (CNA) (23). Although individual oncogenes and tumor suppressor genes are preferential targets of DNA amplifications and deletions, respectively, the recurrent CNA patterns in tumors cannot be fully explained by canonical cancer genes (24–26). Thus, the recurrent CNA patterns observed in human cancer subtypes are suggestive of additional, not yet fully defined, selective pressures that are conserved across patients and tumor types (27,28).

In this study, we have extended transcriptional metabolic pathway analysis (29) to the analysis of DNA copy number patterns that are selected for during genomic instability. Genomic instability is an 'enabling' hallmark of cancer, with the cause and effect relationship between genomic instability and tumorigenesis highly intertwined. Here we describe coordinated copy number changes in a set of co-functioning metabolic genes that have not been previously reported. Our results indicate that the pattern of DNA alterations selected for by genomic instability are influenced by the linked metabolic function of the genes involved, offering a new variation on how genomic instability can promote tumor-beneficial metabolism.

1.5 FDG-avid tumors

FDG-avid tumor types with highly recurrent CNA patterns often show bad prognosis in cancer patients. Cancer is a heterogeneous disease consisting of multiple subtypes defined by molecular signatures in addition to morphological characteristics. PET imaging with metabolic tracer 18F-deoxyglucose (18F-FDG) is widely used in many tumors for detecting, staging and characterizing malignancies across different cancer types (30). For example, the study of lung

cancers concluded that the presence of any tumor suppressor gene abnormality is associated with an expected augmentation of 18F-FDG accumulation (31). Similarly, higher 18F-FDG accumulation has been shown to reflect the metastatic propensity in the breast tumors as well as predict their poor response to neoadjuvant chemotherapy. On the other hand, the recent advent of microarray technology has revealed that many tumor types share similarity in the patterns of the DNA copy number alterations. Moreover, high genomic instability and high aneuploidy tumors are some of the most difficult to treat tumors that have high mortality rates. The work of our lab has shown that basal-like breast tumors (that are characterized by genomic instability, like BRCA1/2 mutation (32)) also generally have higher uptake of 18F-FDG compared to other breast-cancer subtype (29). In addition, basal-like breast cancers are often refractory to targeted therapeutics because they typically lack expression of hormone receptors and HER2. In general, metabolic alterations and genomic instability in tumor cells are common to neoplastic cells. Higher FDG-uptake and higher aneuploidy reflects tumors at the later stage and with worse prognosis.

1.6 Redox selective forces during tumorigenesis

Carcinogenesis is often described as Darwinian process where traits are thought to arise as adaptive mechanisms to environmental proliferative constraints depending on their contribution to individual fitness. We hypothesized that the genomic DNA copy number alterations recurrently observed in human cancers are a result of the oxidative stress-based selective pressures encountered as tumor cells undergo tumorigenesis, e.g. while overcoming senescence checkpoints. Cells undergoing high oxidative stress prior to substantial mutation in oncogenes and tumor suppressor genes are prone to undergo the high genomic instability associated with genomic crisis (**Fig. 1.6-1**). It was shown that hypoxic tumors, which require

increased glycolysis to survive, are often more invasive and metastatic than those with normal oxygen levels (22). Also, deriving cells from tumors often preserves their metabolic phenotypes in culture even under normoxic conditions (22). This indicates that aerobic glycolysis is constitutively upregulated through stable genetic or epigenetic changes.

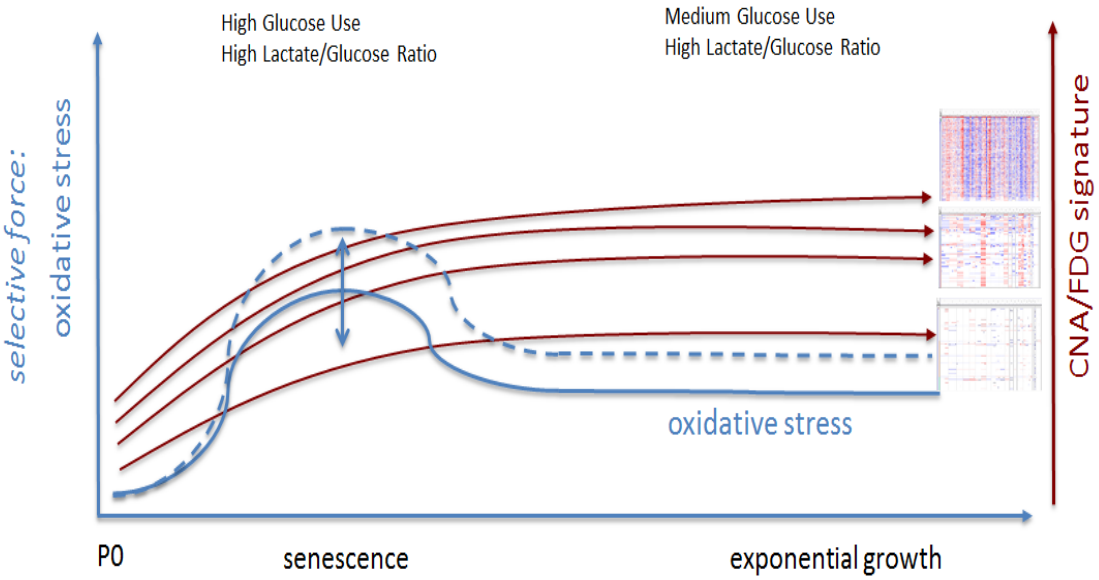


Figure 1.6-1. Model for redox selective forces selection. Genomic CNA patterns of a tumor reflect its past tumorigenesis path.

1.7 Significance

Our lab has implicated the amplification and deletion of glycolysis and redox metabolism genes as a driving force underlying the recurrent pattern of DNA copy number alterations observed in cancer patients using microarray-based genome-wide profiling (see **Fig. 1.7-1** for a schematic of our overall approach). Despite of the long appreciation of the role of each individual selective event during the tumorigenesis, we are addressing the question whether redox stress and a corresponding need for increased glycolysis (Warburg effect) are one of the major selective forces for the highly recurring patterns in human tumors that were only recently

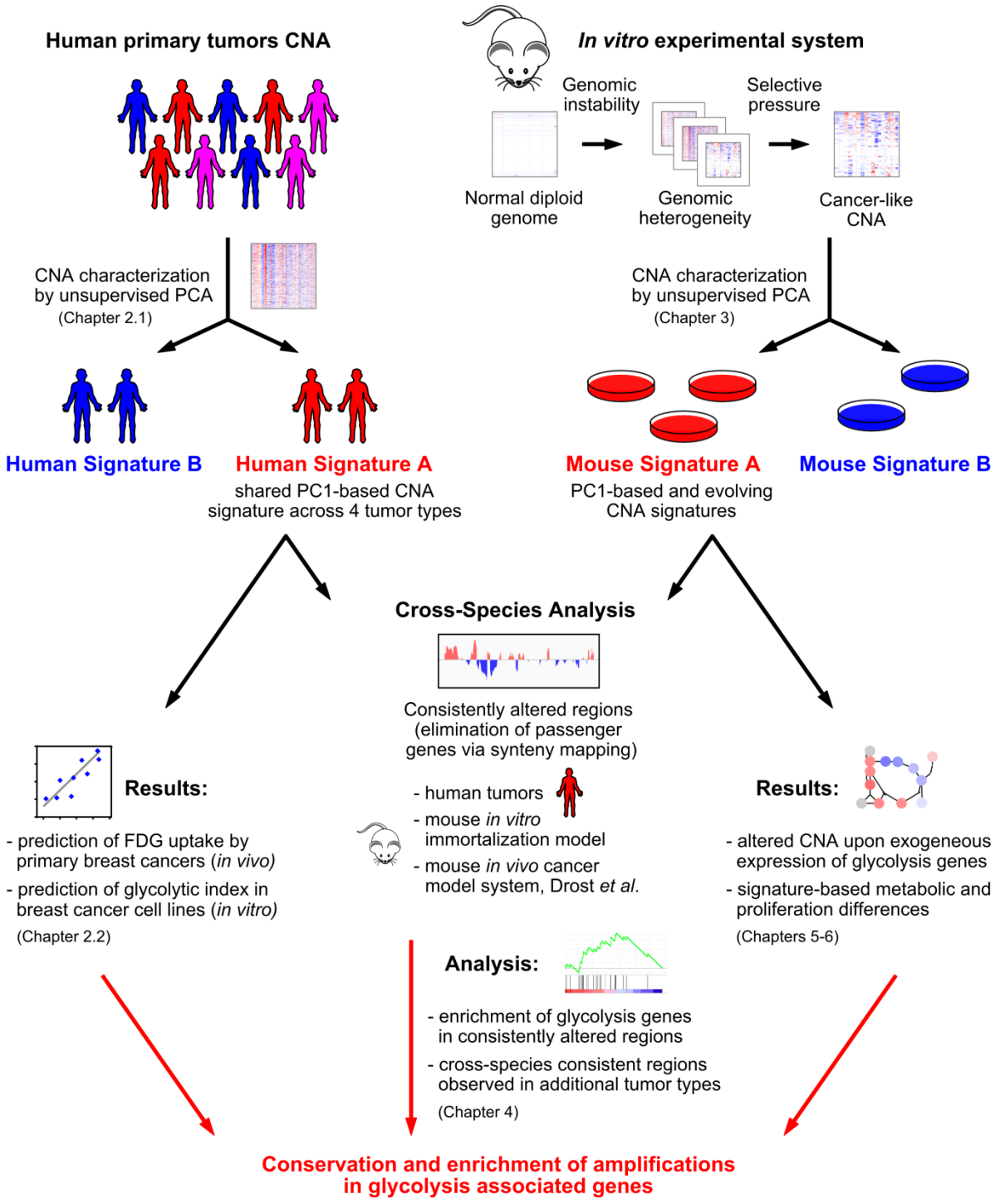


Figure 1.7-1. Schematic of bioinformatic and experimental analysis.

observed due to the advent in technology. We show that:

- Copy number alterations (CNA) of energy metabolism genes provide a selective advantage during tumorigenesis, namely by promoting glycolytic metabolism.
- Recurrent copy number patterns observed in patient tumors are in part determined by the coordinated selection of amplifications and deletions of metabolic genes due to oxidative stress.

Overall, our findings offer a new category of how oxidative stress-based genomic instability can enable tumor progression.

CHAPTER 2

Analysis of copy number alterations in primary tumors implicates glycolysis genes in shaping recurrent DNA amplification and deletion patterns.

2.1 PCA-defined CNA signatures in human cancers

To develop an unbiased understanding of DNA copy number alterations (CNA) in cancer, we performed principal component analysis (PCA) of gene-based CNA data from 16 tumor types available from The Cancer Genome Atlas (TCGA). This pan-cancer analysis revealed a high

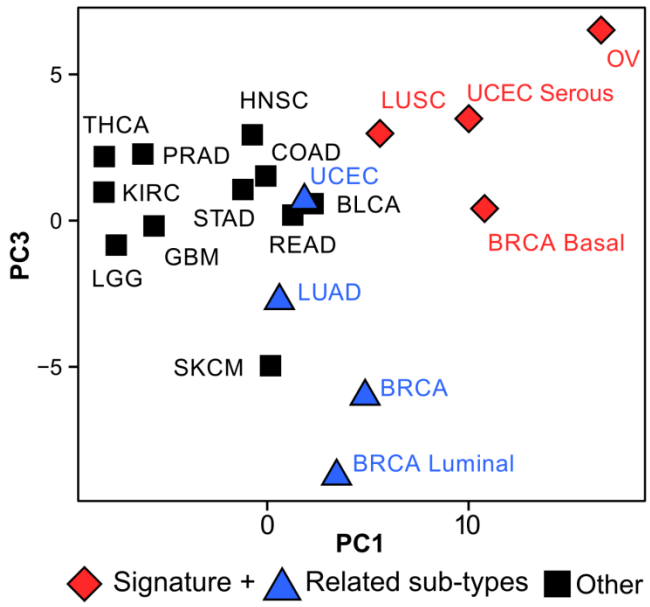


Figure 2.1-1. Principal component analysis (PCA) of 16 tumor types. PCA reveals a shared CNA signature in breast, lung, ovarian and uterine carcinomas. PCA of copy number data from a balanced, random sampling of tumors of 16 tumor types from The Cancer Genome Atlas (TCGA). The average tumor PC scores for each tumor subtype are shown. PC2 distinguished GBM from the other tumor types (not shown). Four tumor subtypes with similar PC scores are labeled as 'Signature +'. Tumor type abbreviations are as defined by TCGA and listed in **Chapter 8.2.6**.

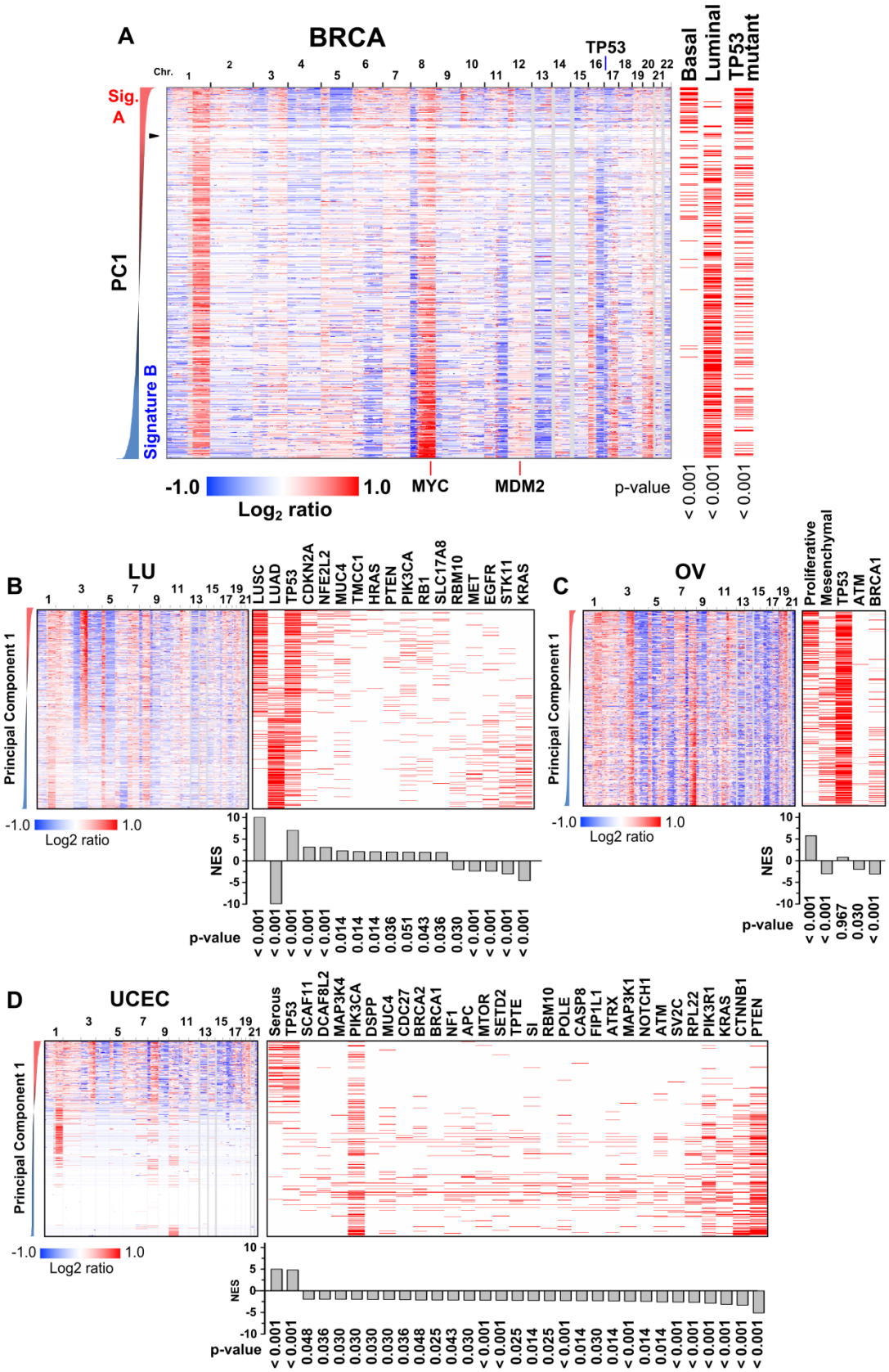

A
BRCA
Chr.
TP53
Basal
Luminal
TP53 mutant
Sig. A
PC1
Signature B
-1.0
1.0
Log2 ratio
MYC
MDM2
p-value
< 0.001
B
LU
Principal Component 1
LUSC
LUAD
TP53
CDKN2A
NFE2L2
MUC4
TMCC1
HRAS
PTEN
PIK3CA
RB1
SLC17A8
RBM10
MET
EGFR
STK11
KRAS
Log2 ratio
NES
C
OV
Proliferative
Mesenchymal
ATM
BRCA1
0.967
0.030
D
UCEC
Serous
SCAF11
DCAF8L2
MAP3K4
DSPP
CDC27
BRCA2
BRCA1
NF1
APC
MTOR
SETD2
TPTE
SI
POLE
CASP8
FIP1L1
ATRX
MAP3K1
NOTCH1
SV2C
RPL22
PIK3R1
CTNNB1
0.048
0.025
0.043

Figure 2.1-2. PCA-defined shared signature. PCA reveals a shared CNA signature in breast, lung, ovarian and uterine carcinomas. (A-D) Copy-number profiles of 873 breast invasive carcinomas (BRCA) (A), 359 lung squamous carcinomas (LUSC) and 368 lung adenocarcinomas (LUAD) (LU) (B), 583 ovarian serous cystadenocarcinoma (OV) (C), and 492 uterine corpus endometrial carcinomas (UCEC) (D). Tumors (rows) sorted by tumor-specific principal component 1 (PC1) score with genomic locations listed across the top. PCA identified and distinguished two signatures with similar degrees of variance in the first component. In panel A, The triangle marks the transition from signature A through diploid samples to signature B. PC1-ranked tumors were analyzed for enrichment of known mutations or clinical sub-types (eg, basal/luminal in breast, proliferative/mesenchymal in ovarian, serous/non-serous in uterine). Normalized enrichment score and permutation p-value for each significantly enriched mutation or clinical sub-type are shown below the tick plots.

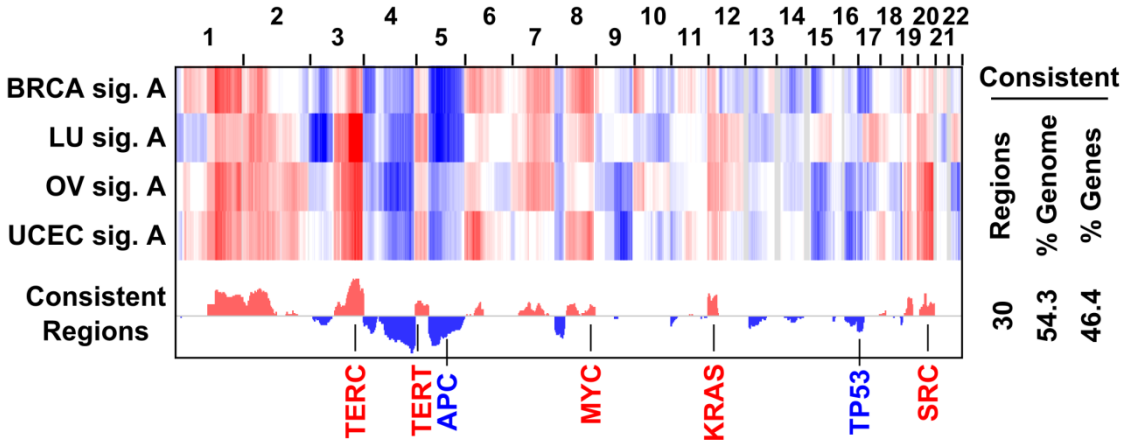
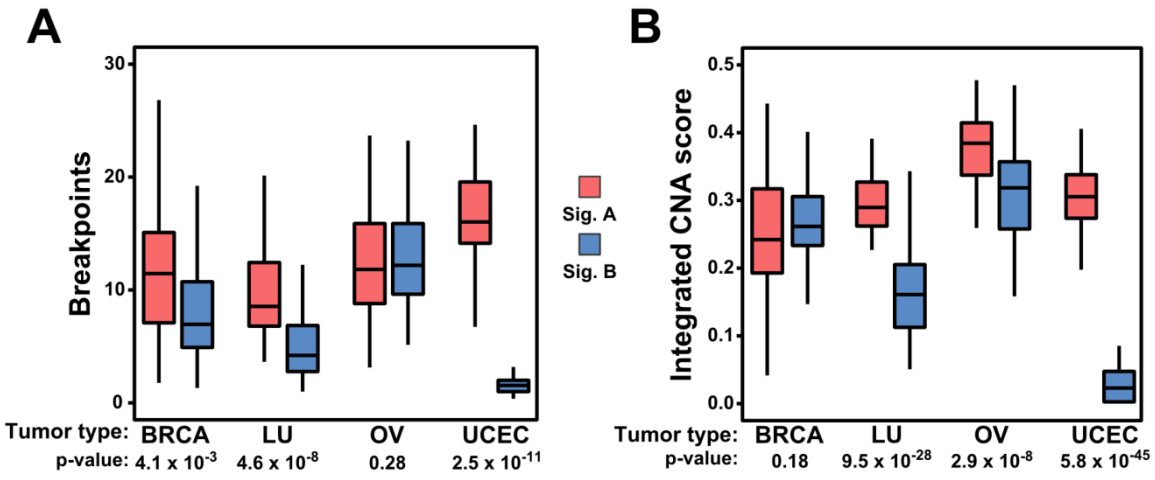


Figure 2.1-3. Consistency of PCA-defined shared signature. PCA reveals a shared CNA signature in breast, lung, ovarian and uterine carcinomas. Signature A summary profiles for breast (BRCA), lung (LU), ovarian (OV) and uterine (UCEC) tumor types. Summary signatures are normalized gene loci signal-to-noise ratios (SNRs) of the top 10% of PC1-based signature A tumors compared to patient-matched normal (non-tumor) samples. Consistency signatures of conserved amplification and deletions (consistent regions) are non-zero when all CNA summary signatures have the same sign across all tumor types, and are derived from the absolute value-based minimum summary metric, then re-signed positive for amplification or negative for deletion. In the bar graph, red and blue denote consistently amplified and deleted regions, respectively. The number or percentage of consistent regions, genome coverage, and percent gene loci are indicated.

degree of similarity in CNA profiles between basal breast invasive carcinoma (BRCA basal), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV) and serous uterine corpus endometrial carcinoma (UCEC serous) (**Fig. 2.1-1**).

Analyzing these four tumor subtypes revealed two PCA-based CNA signatures, termed signatures A and B, in each subtype. Signature A was highly consistent across all four tumor types (**Fig. 2.1-2** and **2.1-3**). In BRCA, signature A tumors were enriched for the basal subtype, p53 point mutations, high numbers of genomic breakpoints, and sub-chromosomal alterations (**Fig. 2.1-2A** and **2.1-4**). Signature B BRCA tumors, in contrast, were enriched for luminal type tumors and amplifications of the oncogenes MYC and MDM2. In the other tumor types, signature A tumors were enriched for lung squamous cell carcinomas, the proliferative subtype of ovarian cancer (33), and the serous subtype of uterine cancer. Overall, signature A tumors demonstrated enrichment of p53 mutations and more genomic breakpoints (BRCA, LU



F**igure 2.1-4. Characteristics of PCA-defined shared signatures.** PCA-defined signature A tumors exhibited more genomic breakpoints (D) and a higher degree of copy number alterations (integrated CNA score) (E) than signature B tumors. The exceptions are ovarian (OV), which exhibited similar numbers of genomic breakpoints in signature A and B tumors, and breast (BRCA), which exhibited similar levels of copy number alterations in signature A and B tumors. T-test p-values are shown for each Sig. A versus Sig. B comparison. Data is presented in box (median, first and third quartiles) and whisker (extreme value) plots.

and UCEC) and a higher degree of copy number alterations (LU, OV and UCEC) than signature B tumors (**Fig. 2.1-4**). Unlike signature A tumors, the signature B tumors identified in each tumor type were quite distinct, although some commonalities were observed including point mutations in oncogenes such as KRAS (LU and UCEC) and amplification of MYC (BRCA, LU and OV).

	Normalized Enrichment Scores				NES	p-value	FDR q-value
KEGG metabolism gene set	BRCA sig. A	LU sig. A	OV sig. A	UCEC sig. A	Consistent Sig. A (n = 4)		
Core Glycolysis	1.47	1.70	1.46	0.78	1.92	0.02	0.36
Propanoate metabolism	-0.65	1.11	1.11	0.95	1.70	0.04	0.36
Amino sugar and nucleotide sugar metabolism	1.10	0.63	-0.56	0.74	1.63	0.05	0.36
Glycolysis	1.03	1.35	1.08	-0.71	1.60	0.05	0.37
Glycolysis & Pentose Phosphate	1.06	1.11	1.34	-0.79	1.50	0.06	0.39
GPI-anchor biosynthesis	0.94	1.07	1.43	1.41	1.44	0.09	0.41
Valine, leucine and isoleucine degradation	1.33	1.29	1.20	0.90	1.30	0.12	0.41
Purine metabolism	0.76	0.72	-0.75	1.44	1.25	0.13	0.41
...	...						

Table 2.1-1. Metabolic pathway enrichment analysis of shared PCA-defined signature. Metabolic pathway enrichment analysis based on consistent CNA patterns in the signature A tumors of Fig. 2.1-3. The combined glycolysis-gluconeogenesis (glycolysis) and pentose phosphate pathway was included based on our prior mRNA work identifying the predictive value of this gene set (27). Core glycolysis is a KEGG-defined gene subset (M00001).

Since the four CNA-defined signature A tumor types are associated with high glycolysis and increased avidity for the positron emission tomography (PET) tracer 18F-fluorodeoxy-

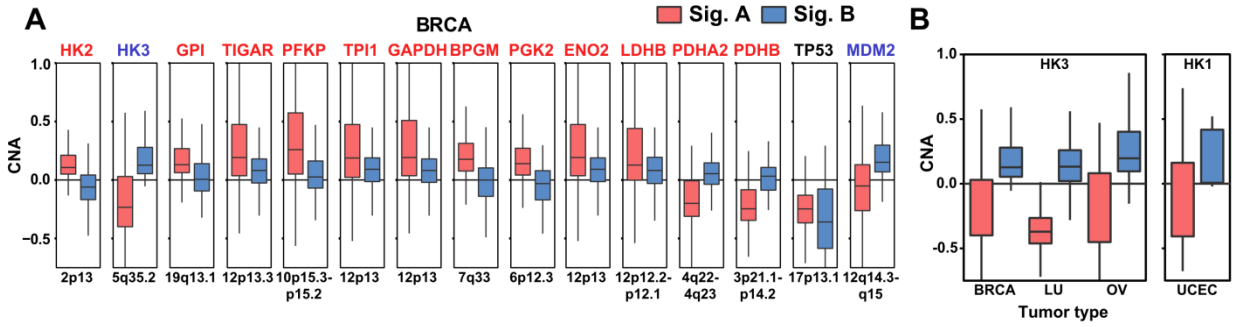


Figure 2.1-5. CNA differences in PCA-defined signatures. (A) Gene copy number alteration distributions of selected core glycolysis genes, TIGAR, and of the tumor suppressor p53 and its negative regulator MDM2 in BRCA tumors. (B) Signature B tumors amplify HK3 (BRCA, LU, OV) and HK1 (UCEC) showing alternative HK amplification as compared to signature A tumors (which amplify HK2).

glucose (FDG) (34–37), we tested whether the shared CNA signature A was enriched for genes from metabolic pathways. Using rank-based gene set enrichment analysis, we found that the signature A profile shared across breast, ovarian, uterine and lung tumors (**Fig. 2.1-3**) was significantly enriched for DNA amplifications of core glycolysis pathway genes (**Table 2.1-1**). For example, BRCA signature A tumors exhibited DNA amplification of most genes from the glycolytic pathway, as well as amplification of lactate dehydrogenase B, deletion of pyruvate dehydrogenase subunits A and B, and amplification of the glycolysis-regulating metabolic enzyme TIGAR (human gene C12orf5;) (**Fig. 2.1-5A**).

Notably, this shared CNA signature was defined by genome-wide patterns, rather than by single gene loci, which were not consistently altered in all tumors with a strong signature (**Figure 2.1-6**).

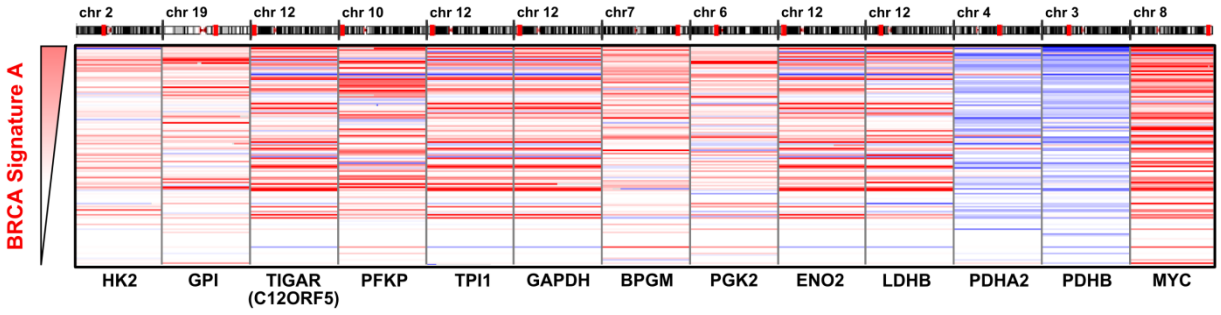


Figure 2.1-6. Genome-wide patterns of PCA-defined signatures. The PCA-identified signatures were defined by genome-wide patterns, as single gene loci were not consistently altered in all tumor cases with a strong signature score, here shown for glycolysis-related genes in the PC1-sorted signature A breast tumors from **Fig. 2.1-2A**.

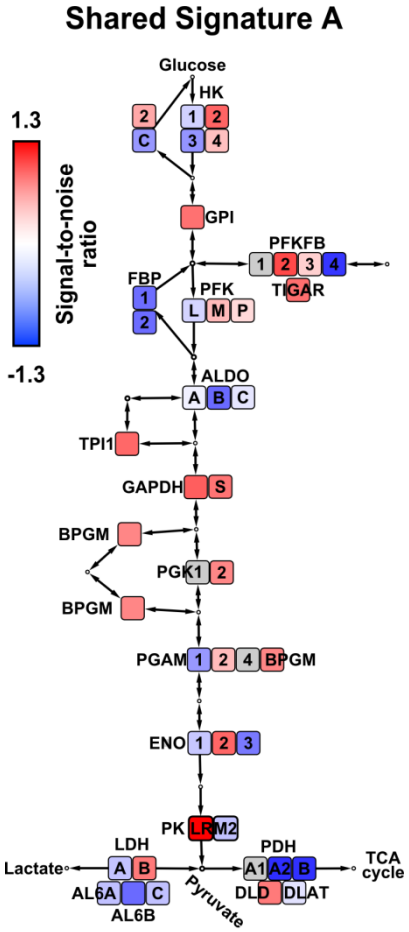


Figure 2.1-7. Signal-to-noise ratios of genes in core glycolysis pathway. Schematic showing average signal-to-noise (SNR) metrics of core glycolysis pathway genes plus 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase (PFKFB and TIGAR), lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH) in signature A tumors from **Figure 2.1-3** and **Table 2.1-1**. The top 10% of tumors from each indicated signature based on PC1 scores were used in the analysis.

Interestingly, Signature A summary profiles from breast, ovarian, uterine and lung tumors all exhibited hexokinase 2 (HK2) amplification (**Fig. 2.1-5A** and **2.1-7**), whereas signature B profiles had primarily either HK3 (BRCA, LU, OV) or HK1 (UCEC) amplification (**Fig. 2.1-5B)**. Thus, PCA identified a shared signature from breast, lung, ovarian and uterine carcinomas that was enriched for p53 mutations, higher numbers of genomic breakpoints, and CNA of genes from the core glycolysis pathway.

2.2 CNA signatures are predictive of glycolysis

In that the CNA-defined signature A is enriched for core glycolysis genes, we next tested whether signature A patient tumors were associated with functionally increased tumor glycolysis. To assign a signature A score to a set of FDG-PET-imaged tumors (37), we projected CNA data from these tumors onto a PCA analysis of the four Signature A tumor types. Indeed, we found a strong correlation between the strength of CNA signature A and the measured FDG-PET standardized uptake values (SUV) (**Fig. 2.2-1**). Thus, the CNA-defined signature A is associated with increased FDG uptake in human primary tumors *in vivo*.

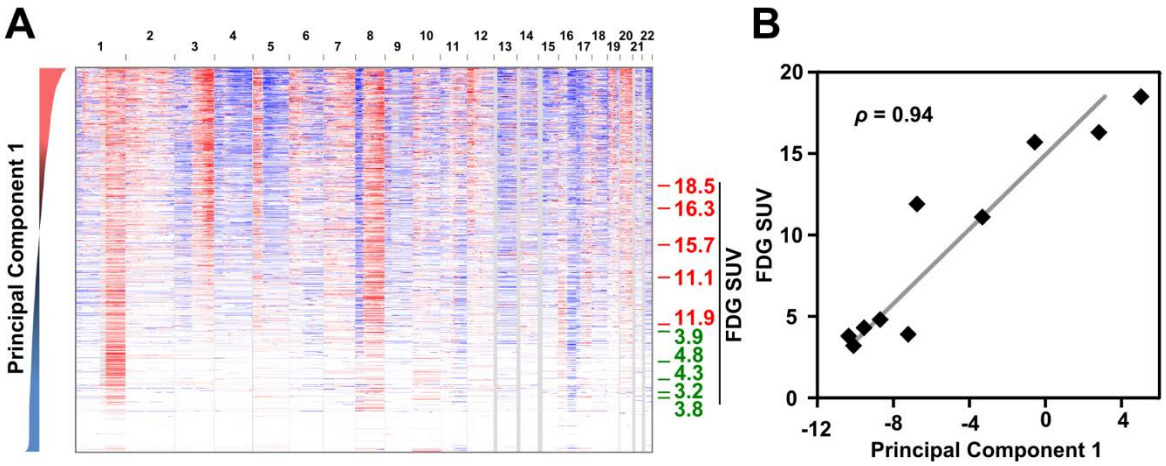


Figure 2.2-1. PCA-based signature is predictive of glycolysis in vivo. PCA-based CNA signatures are predictive of breast cancer glycolytic metabolism in vivo. (A) PC1-sorted copy-number profiles of a balanced, random sampling of tumors from 4 CNA-consistent tumor types (breast, lung, ovarian and uterine carcinomas, **Fig. 2.1-1**) along with copy-number profiles from primary breast carcinoma tumors with glycolytic levels imaged in vivo by 18F-fluorodeoxy-glucose-positron emission tomography (FDG-PET). On the right, red and green values indicate high and low FDG standardized uptake values (SUV), respectively, for FDG-imaged tumors. (B) FDG uptake values in FDG-PET measured tumors are highly correlated with CNA Signature A PC1 score (ρ, Pearson correlation).

KEGG metabolic pathway	Avg. rank	Rank FDG	Rank lactate	Num. genes	Test set:	Average WGV correlation with metabolic phenotypes Training set: BRCA, LU, OV or UCEC Sig. A tumors Tumor FDG SUV	Cell line lactate sec.	p-value	FDR q-value
Glycolysis & Pentose Phosphate Pathway	1	3	3	74		0.88	0.47	0.01	0.28
Valine, leucine & isoleucine degradation	2	1	11	42		0.95	0.41	0.01	0.28
Glycolysis	3	11	4	62		0.83	0.46	0.08	0.76
Purine metabolism	4	8	7	154		0.86	0.43	0.11	0.93
Lysine degradation	5	14	2	42		0.82	0.48	0.06	0.76
...	...								

Table 2.2-1. Predictive power of PCA-defined signatures. CNA values for genes from glycolysis and pentose phosphate pathways have stronger predictive power of FDG-PET SUV in breast tumors and of lactate secretion (sec.) in breast cancer cell lines than other metabolism pathway-based sets of genes. The table indicates the correlation between weighted gene voting (WGV)-based predictions on the test sets and the measured metabolic phenotypes. WGV was performed with individual training on signature A tumors from each of the four tumor types (breast (BRCA), lung (LU), ovarian (OV) and uterine (UCEC); top 10% signature A compared to patient-matched normal samples), the voting predictions were averaged, and compared to the measured metabolic phenotype.

Next, we asked which metabolic gene sets were most predictive of glycolytic phenotypes. We performed a CNA-based weighted gene voting (WGV) (38) analysis to predict the glycolytic phenotypes of breast tumors and breast cancer cell lines (39) using individual gene sets from the KEGG metabolic pathways database (40). Gene weights were calculated from each of our four tumor type CNA "training" signatures. Specifically, we tested the ability of individual metabolic pathways to predict i) FDG uptake in patient primary breast tumors and ii) the lactate secretion of a panel of 32 breast cancer cell lines. Averaging results across these two

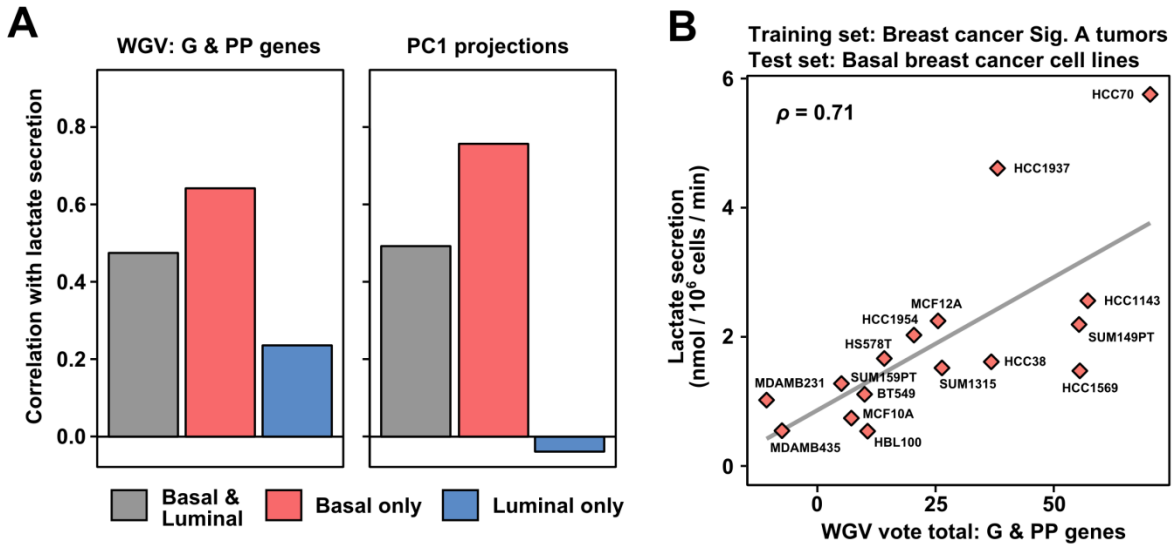


Figure 2.2-2. PCA-based signature is predictive of glycolysis in vitro. (A-B) Signature A-based WGV predictions and signature A-based PC1 projections (as in **Fig. 2.2-1B**) have stronger correlation with measured lactate secretion for basal-like breast cancer cell lines than for luminal-like lines (A). The breast cancer-based glycolysis and pentose phosphate pathway (G & PP) WGV predictions are shown as a representative case (A-B). Previously published basal and luminal designations are based on gene expression profiles (37).

test cases revealed that genes from the glycolysis-gluconeogenesis (henceforth called glycolysis) and pentose phosphate pathway were most predictive of these metabolic phenotypes (**Table 2.2-1**). Moreover, Signature A-based predictions were more predictive of lactate secretion for basal cell lines than for luminal lines (**Fig. 2.2-2**), consistent with the observed basal tumor enrichment in signature A samples (**Fig. 2.1-2B**). Thus, the glycolysis and pentose phosphate pathway DNA copy number alterations from signature A are predictive of glycolytic phenotypes of primary human breast tumors and cancer cell lines.

2.3 Methods summary of chapter 2

Copy-number profiles for 16 tumor types were obtained from The Cancer Genome Atlas (TCGA). The types and subtypes used for the principal component analysis, PCA, are: BLCA, LGG, BRCA (as well as BRCA basal and BRCA luminal), COAD, GBM, HNSC, KIRC, LUAD, LUSC, OV, PRAD, READ, SKCM, STAD, THCA, UCEC (as well as UCEC serous).

Principal component analysis (PCA) was performed using the mean-centered matrix of CNA values per gene locus. To test for enrichment of mutations within CNA-defined PC scores, tumors were queried for mutations in the most frequently mutated genes using the cBioPortal for Cancer Genomics (41). Tumors with known mutational status were then sorted by their PC1 score, and a Kolmogorov-Smirnov statistic against the expected distribution of mutations was calculated.

Consistency signatures of conserved amplification and deletions (consistent regions) were defined as non-zero when all CNA summary signatures had the same sign across all tumor types. Metabolic pathway enrichment analysis, and weighted gene voting (WGV) prediction analysis (38), were performed using 74 metabolic pathways defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (40). CNA changes were visualized in the context of metabolic pathway structure using Cytoscape (42). A genomic instability score termed 'integrated CNA' was calculated by integration of the absolute mean copy-number of segments multiplied by the length of each segment.

Ten patient breast cancer samples with imaged FDG-uptake within 4 weeks prior to surgery, excluding patients with secondary breast cancers and recurrent disease, were collected surgically and processed as previously described (37).

More detailed descriptions of these methods can be found in **Chapter 8**.

CHAPTER 3

Experimental mouse immortalization system recapitulates genomic instability patterns observed during tumorigenesis.

3.1 Experimental recapitulation of tumor CNA signatures

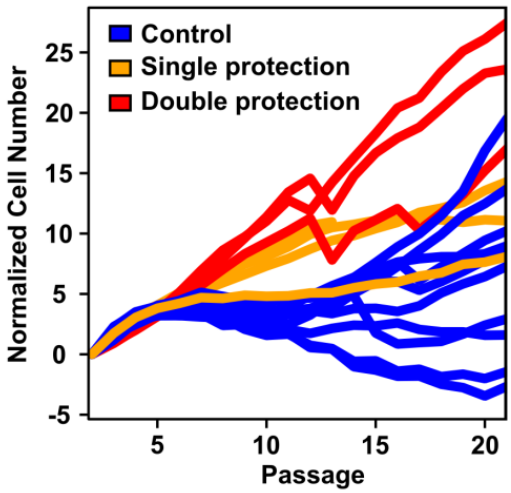


Figure 3.1-1. Representative growth curves of MEFs. Growth curves showing that protection of CD1 MEF lines from oxidative stress rescues cells from replicative senescence. Single protection: cells were cultured under physiological oxygen conditions (3% O2) or by media supplementation with 250 U/ml of the ROS scavenging enzyme catalase. Double protection: cells were cultured at 3% O2 and supplemented with catalase. Exogenous MYC expression also resulted in rescue from replicative senescence (not shown) and did not exclude evolution towards either signature A or B (**Fig. 3.1-2**).

The conservation of CNA signature A across several tumor types suggests one or more shared selective pressures. To better define genomic regions with candidate driver genes within these recurrently altered regions, we reasoned that a cross-species comparison of CNA data from human tumors and murine tumorigenesis models would eliminate passenger genes via

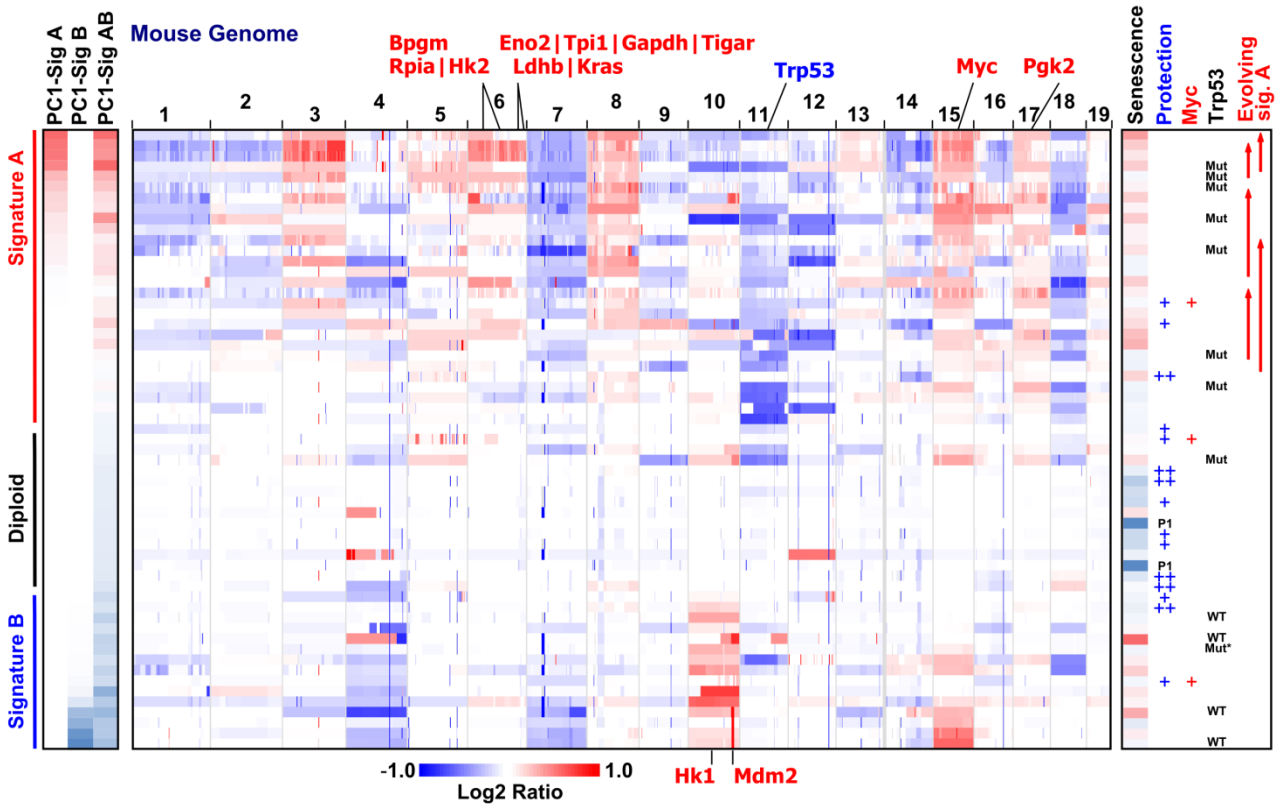


Figure 3.1-2. Copy-number profiles of MEFs. Copy-number profiles of 59 samples from 42 independent mouse embryonic fibroblasts (MEF) sublines before, during and after senescence recapitulate the two signature CNA patterns observed in human tumors. PC1 scores for analysis of only signature A MEFs (Sig A), only signature B MEFs (Sig B), or all MEFs (Sig AB) are indicated on the left. Also indicated are a metric of the degree of senescence observed during immortalization (senescence score), protection by 3% O2-based hypoxia or catalase (+), or both (++) during immortalization, CNA profiling at passage 1 (P1), exogenous MYC expression which enabled cells to bypass senescence, and Trp53 sequencing status (*indicates a non-severe mutation p.183D>E, blank indicates not sequenced, additional p53 sequencing information in **Table 3.1-1**). Signature A MEF lines profiled at more than one passage number are indicated by the start (earlier passage) and end points of the upward 'evolving signature A' arrows. The indicated chromosome 6 region includes loci for Bpgm, Hk2, Rpia, Tigar, Eno2, Tpi1, Gapdh, Ldhb, and Kras.

synteny mapping. We thus derived a panel of immortalized mouse embryo fibroblasts (MEFs) using the classical 3T9 protocol (43). In this experimental system, diploid cells undergo spontaneous genomic instability during senescence resulting in immortalized cells (**Fig. 3.1-1**)

Sample	Signature	Mutation in Mm p53	Homolog. mutation in Hs p53	Exons									
				2	3	4	5	6	7	8	9	10	11
H2 P25	B			wt	nd	wt	wt	wt	wt	wt	wt	wt	wt
E2 P54	B			wt	nd	wt	wt	wt	wt	nd	nd	wt	wt
A2 P25	B			wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
A1 P25	B	**p.183 D>E†‡**	**p.186 D>E**	wt	nd	wt	**m.549 T>A**	nd	nd	nd	nd	nd	nd
B7 P99	B			wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
D10 P36	mixed	**p.131 F>L**	**p.134 F>L**	nd	nd	wt	**m.393 C>A**	wt	wt	wt	wt	wt	nd
F1 P23	A	**p.R178***	**p.R181***	wt	nd	nd	**m.536 ***	nd	nd	nd	nd	nd	Nd
A6 P40	A	**p.156 A>P**	**p.159 A>P**	wt	nd	wt	**m.466 G>C**	wt	wt	wt	nd	nd	nd
C8 P40	A	**p.270 R>C**	**p.273 R>C**	wt	wt	wt	wt	nd	wt	**m.808 C>T**	wt	nd	nd
A3 P25	A	**p.277 R>T**	**p.280 R>T**	wt	wt	wt	wt	wt	wt	**m.830 G>C**	wt	wt	nd
H1 P25	A	**p.129 K>T**	**p.132 K>T**	wt	nd	d	**m.385 A>C**	wt	wt	nd	nd	nd	nd
C9 P42	A	**p.274 C>F**	**p.277 C>F**	nd	nd	wt	wt	wt	wt	**m.821 G>T**	wt	wt	nd

Table 3.1-1. Trp53 sequencing in immortalized CD1 MEFs. The cell lines and the passage number (P) at which p53 protein-coding exons were sequenced are indicated in the sample column. The homologous mutation in human p53 is indicated. †Non-severe mutation; ‡Heterozygous; *Nonsense mutation; wt, wild type sequence; nd, not determined.

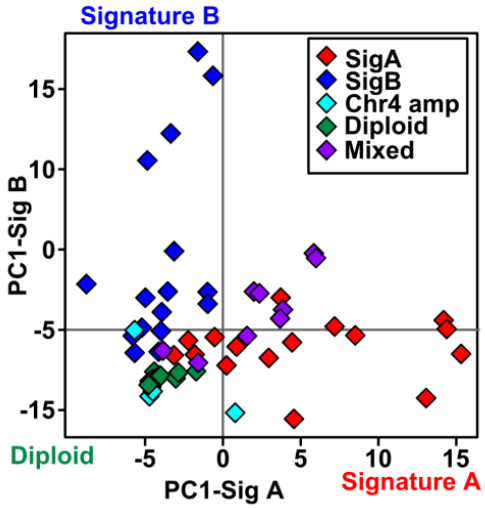


Figure 3.1-3. PCA signatures in MEFs. In order to obtain pure signatures for A and B lines, we ran PCA individually on either signature A only or signature B only MEF lines (PC1-Sig A and PC1-Sig B, respectively). Samples not run in a particular analysis were projected onto the rotated PCA axes to allow for a full comparison. Plotting Signature A against Signature B scores revealed that CNA patterns of these two groups were generally orthogonal, with the exception of a few 'mixed' signature samples that had CNA characteristics of both signature A and B. A less frequent CNA pattern involving chromosome 4 amplification was also observed.

with aneuploid genomes. This system has been used to study core cancer phenotypes such as proliferation, anti-apoptosis, and chromosomal instability (44–49) and is one of the few experimentally tractable cancer models involving spontaneous genomic instability (50). In addition, because this system is not driven by strong oncogenes (eg, KRAS mutation), it allows for complex CNA signatures to evolve from a combination of individual, presumably weaker, DNA alteration events.

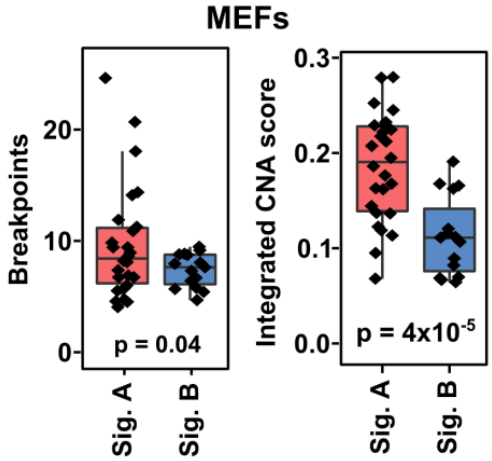


Figure 3.1-4. Genomic instability of PCA-defined MEF signatures. Signature A MEFs have more DNA breakpoints and a larger degree of copy number alteration (integrated CNA score) than signature B MEFs.

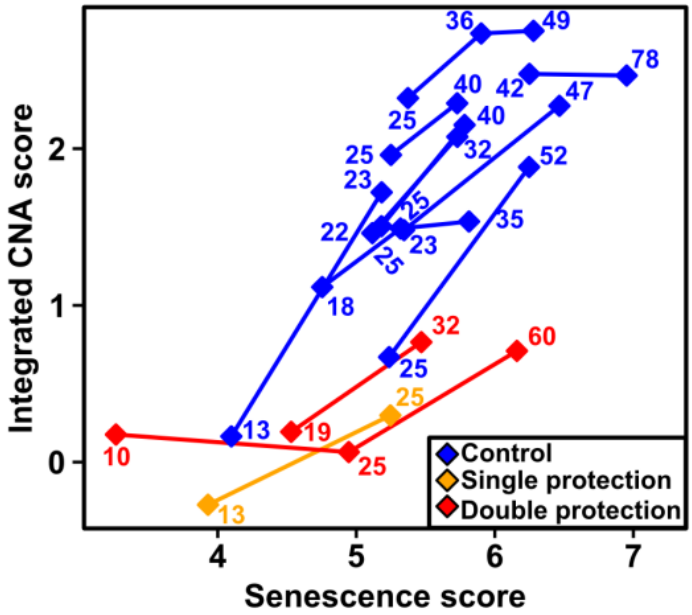


Figure 3.1-5. Genomic instability relation to senescence in evolving MEFs. The senescence score tended to increase during wild-type MEF line derivation (reflecting accumulated sub-optimal growth) and correlated with higher degrees of copy number alterations (integrated CNA score) observed at later passages. In comparison, ROS protected sublines experienced less senescence and obtained lower degree of copy number alterations at later passages. Single protection indicates protection by 3% O2 culture conditions or by media supplementation with 250 u/ml Catalase. Double protection indicates that cells were cultured at 3% O2 with catalase.

We profiled the genome-wide copy number of 42 independent MEF sublines (**Fig. 3.1-2**). Most samples were profiled after immortalization (post-senescence), with a few profiled before or during senescence. Analysis of this CNA data by PCA revealed that the MEF system recapitulated a two signature pattern (signatures A and B). These two signatures were generally orthogonal, with the exception of a few 'mixed' samples that had CNA characteristics of both signature A and B (**Fig. 3.1-3**). Importantly, the MEF-derived signature A resembled the shared signature A pattern derived from human breast, lung, ovarian and uterine tumors. In particular, the MEF-derived signature was characterized by p53 mutation and loss of

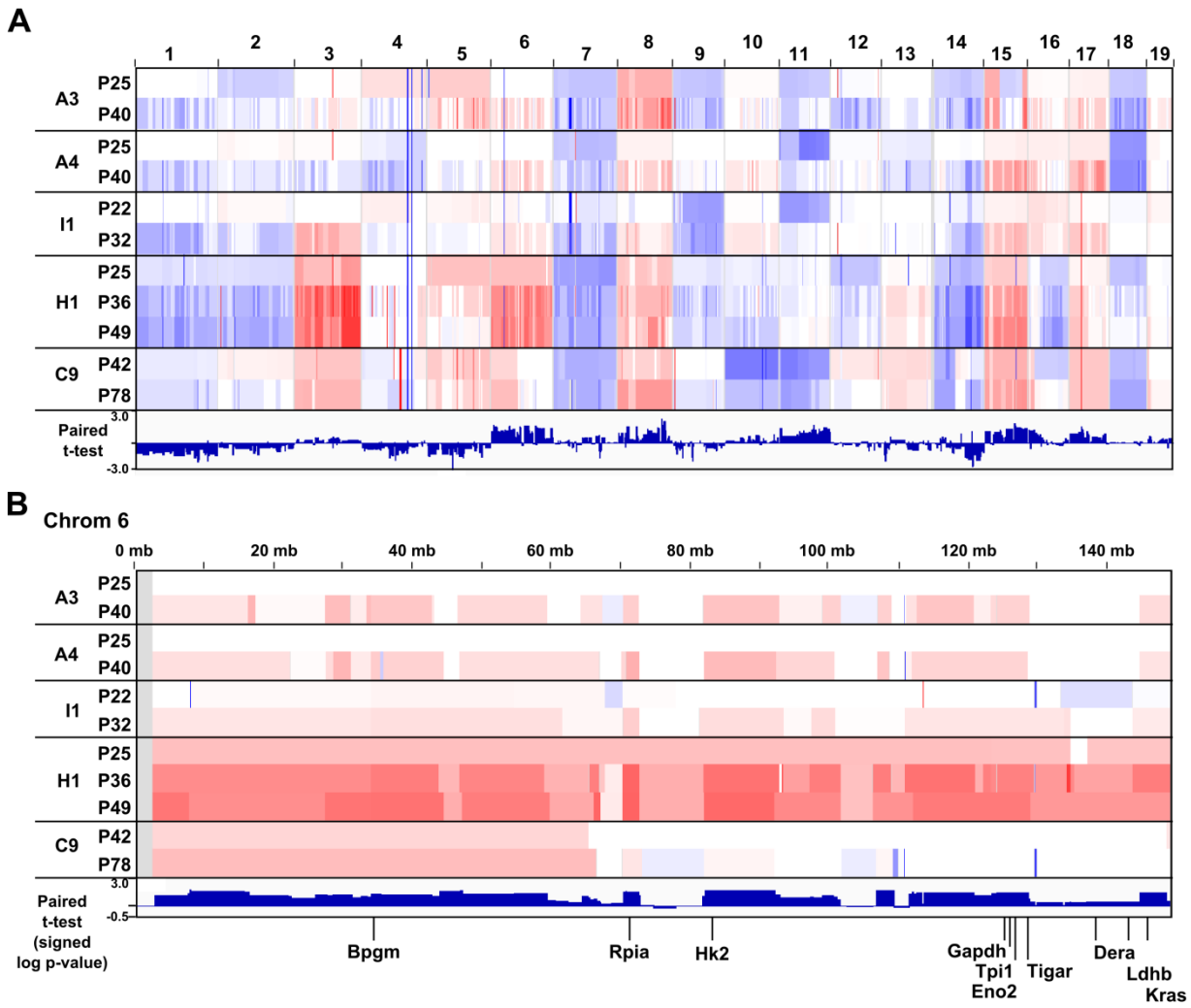


Figure 3.1-6. Paired analysis of evolving signature A MEF lines. (A-B) A paired t-test analysis of signature A MEF lines profiled at more than one passage number revealed genomic regions associated with mid- to late-passage CNA evolution (log10 t-test p-value signed positive for amplifications, negative for deletions; bottom row). The full genome (A) and chromosome 6 (B) are shown. As with the human tumor signatures A, this mid- to late-passage genomic signature was enriched for DNA amplifications of genes in the core glycolysis and glycolysis associated pathways (**Table 4.1-1**). We also observed a general co-evolution of higher growth rates and increased CNA signature strength of MEF lines profiled at more than one passage number (**Fig. 6.1-3B**).

heterozygosity, more genomic breakpoints, sub-chromosomal sized alterations, and a higher degree of copy number alteration (**Fig. 3.1-2**, **3.1-4** and **Table 3.1-1**). Additionally, this

experimental system allowed us to profile MEF lines at subsequent passages, and a paired statistical analysis of five evolving signature A lines revealed genomic regions changing from mid- to late-passage (**Fig. 3.1-2**, **3.1-5** and **3.1-6**).

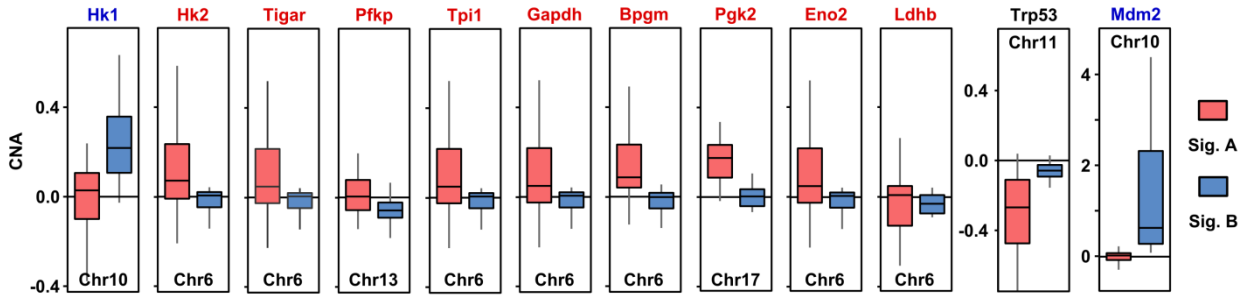


Figure 3.1-7. CNA of selected glycolysis associated genes in MEFs. Gene copy number alteration distributions of selected core glycolysis genes, Tigar, and of tumor suppressor p53 (Trp53) and its negative regulator Mdm2 in MEF lines.

Similar to the shared human tumor signature A, both the mouse signature A and the evolving MEF signature demonstrated DNA amplifications of glycolysis and glycolysis-related genes. In particular, both human and mouse CNA signatures A included amplification of Hk2, Bpgm, Rpia, Tigar (mouse gene 9630033F20Rik), Eno2, Tpi1, Gapdh, Ldhb, Kras (mouse chr. 6) and Pgk2 (mouse chr. 17) (**Fig. 3.1-2**, **2.1-5A** and **3.1-7**).

In contrast, the MEF-derived signature B was characterized by Mdm2 amplification and fewer overall copy number alterations. Mdm2 amplification is an alternative mechanism for inactivating p53 function in human tumors and in MEFs (50), but as found here results in a distinct CNA signature. Mdm2 amplification tends to co-occur with Hk1 amplification, both loci being located on mouse chr. 10. Thus, the signature B cases are associated with an alternate HK amplification, similar to our finding in human tumors (HK1 or HK3 rather than HK2) (**Fig. 2.1-5** and **3.1-7**). Immunoblotting confirmed that signature A MEF and signature B MEF lines generally had increased expression of Hk2 and Hk1 protein, respectively (**Fig. 3.1-8**).

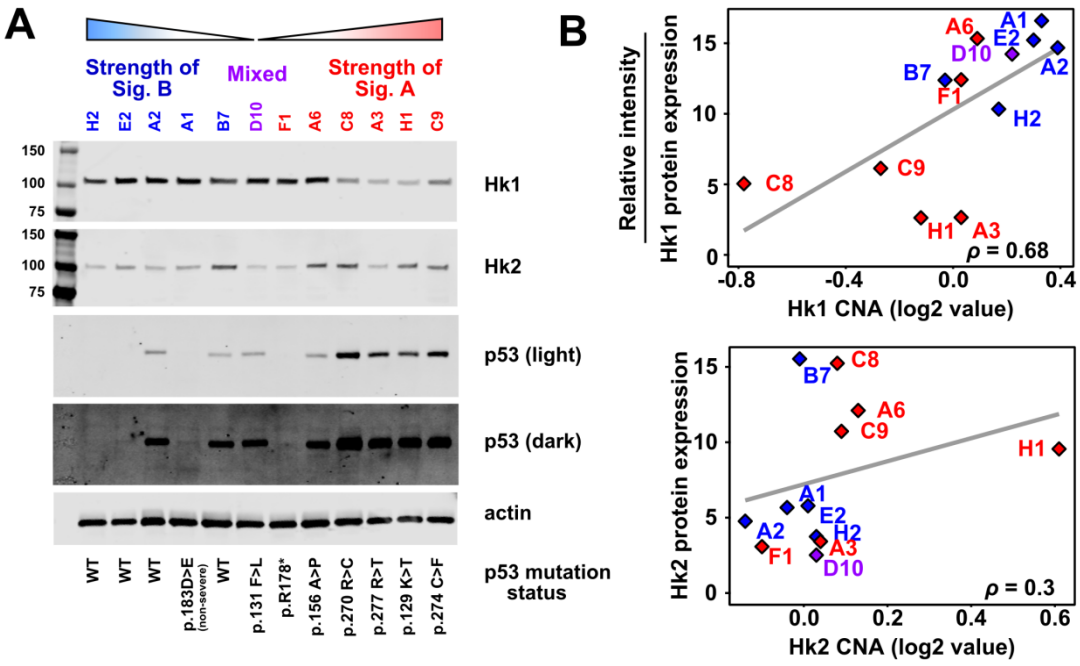


Figure 3.1-8. Immunoblot of representative MEF lines. (A) MEF lines representative of Signature A and Signature B (as defined by CNA profiles) were lysed and probed by immunoblotting for Hk1, Hk2 and p53. Signature A is associated with higher DNA copy number and higher protein expression of Hk2, whereas Signature B is associated with higher CNA and protein levels of Hk1. Signature A cells, which are characterized by p53 mutation (**Table 3.1-1**), show higher expression of p53 protein than Signature B cells, which typically have wild-type p53. Mutations in p53 commonly lead to elevated protein levels via decreased degradation. Cell lines are arranged by their CNA-defined PCA signature strength, and actin was included as an equal loading control. (B) Correlation between CNA values and protein expression for both Hk1 and Hk2. Western blots were quantified using ImageJ.

To further characterize the association between p53 loss and CNA signatures, we profiled the CNA patterns of 29 independent p53-/- MEF sublines. Comparison of the p53-/- MEFs CNA patterns to wild-type MEFs by PCA demonstrated that the CNA patterns of p53-/- MEFs resemble the wild-type MEF signature A pattern, with no signature B-like sublines observed (**Fig. 3.1-9** and **3.1-10A**). Thus, strong p53 functional loss (p53 mutation or genetic loss) tends to lead to the higher aneuploidy- and Hk2-associated CNA signature A pattern, while

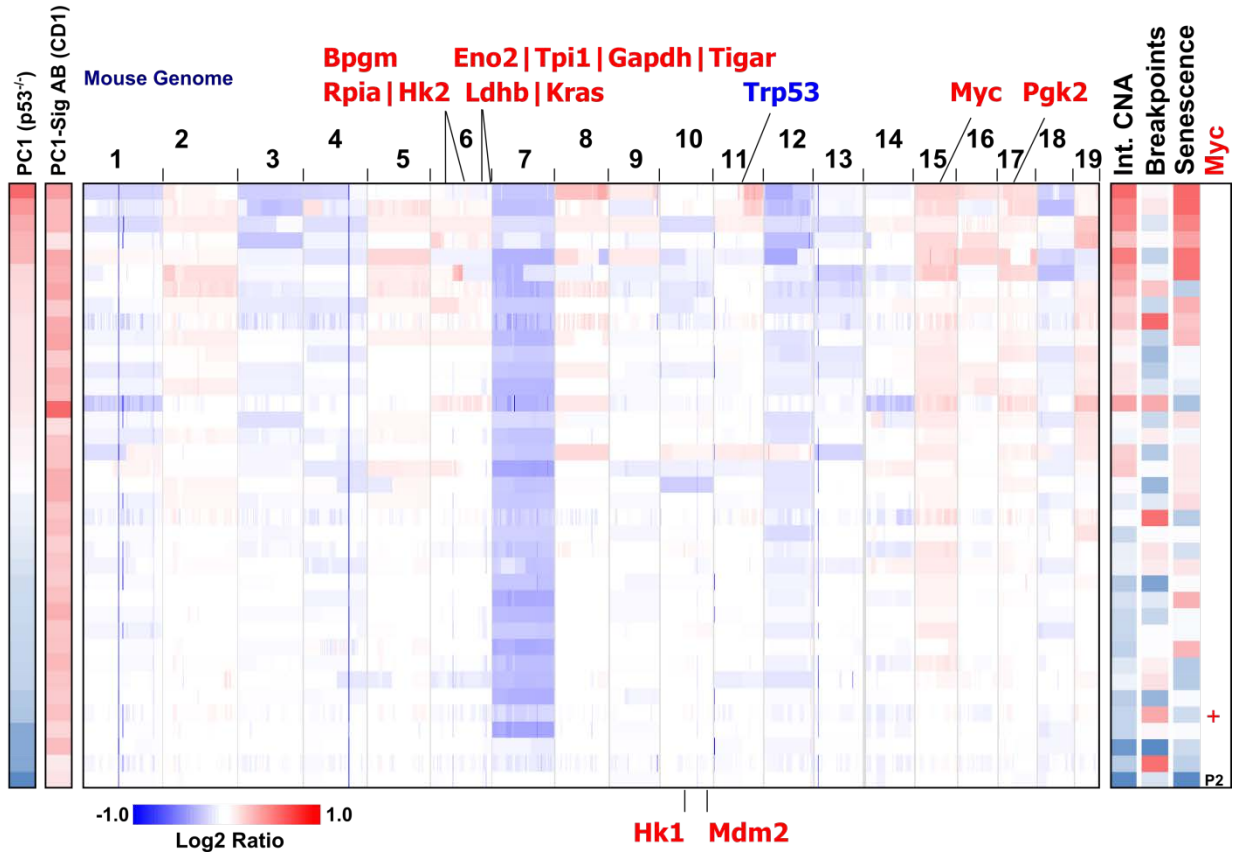


Figure 3.1-9. Copy-number profiles of p53-/- MEFs. Copy-number profiles of 36 samples from 28 independent p53-/- and 1 pre-senescence p53+/+ (passage 2) mouse embryonic fibroblasts (MEFs) sublines. MEF sublines, PC1 scores from analysis of p53-/- MEFs, PC1 (p53-/-), and PC1 scores projected on PC1 scores of CD1 MEFs, PC1-Sig AB (CD1), are indicated on the left. Metrics indicated on the right: copy-number alterations (integrated CNA score, number of breakpoints), degree of senescence observed during immortalization (senescence score), CNA profiling at passage 2 (P2), and exogenous Myc expression.

weaker or less complete p53 functional loss (e.g. mediated by Mdm2 amplification) is associated with an alternative signature (signature B) (**Fig. 3.1-10**).

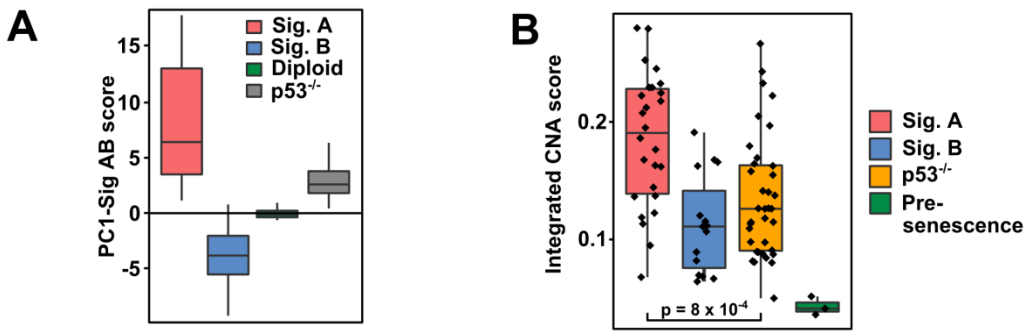


Figure 3.1-10. Characteristics of PCA-defined shared MEF signatures. (A) CNA patterns of 37 p53-/- MEF samples from 29 independent sub-lines resemble the wild-type MEF signature A CNA pattern, as demonstrated by positive PC1-Sig AB scores when projected onto the wild-type MEF PC1 axis. (B) p53-/- MEFs exhibit a lower degree of copy number alterations (integrated CNA score) than signature A MEFs. p53-/- cells do not undergo senescence and thus tend to have less strong copy number alterations.

3.2 Senescence-associated oxidative stress as a selective force for copy number alterations

The MEF system allowed us to investigate the selective pressures driving copy number changes during immortalization. Because MEFs cultured under physiological oxygen conditions (3% oxygen) undergo little to no senescence and exhibit less DNA damage (51), we tested whether oxidative stress-induced senescence is a selective pressure for copy number alterations. To protect cells from oxidative stress, we cultured MEFs in 3% oxygen in media supplemented with or without catalase, an enzymatic scavenger of reactive oxygen species (ROS) (52,53). When doubly protected from oxidative stress, MEFs did not undergo senescence and maintained relatively diploid genomes (**Fig. 3.1-2** and **3.1-1**). In addition, we observed a

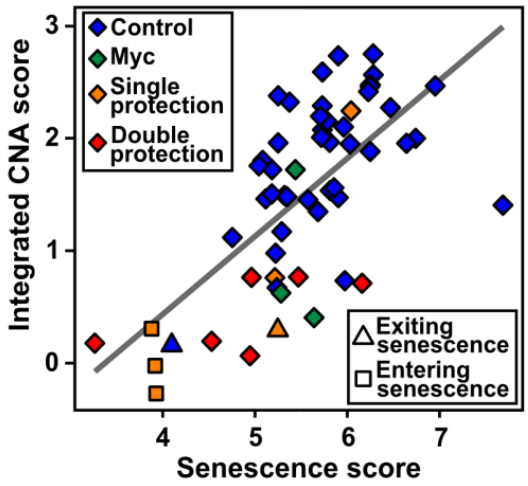


Figure 3.2-1. Genomic instability relation to senescence in MEFs. The amount of senescence demonstrated during MEF line derivation (senescence score) is highly correlated with the degree of copy number alterations obtained (integrated CNA score). Single protection indicates protection by 3% O2 culture conditions or media supplementation with 250 U/ml catalase. Double protection indicates that cells were cultured at 3% O2 with catalase.

variation in the degree of senescence experienced by MEFs derived in atmospheric oxygen concentrations (21%) (**Fig. 3.1-1**). Upon calculating the degree of senescence encountered by each subline, we found a correlation between senescence and the degree of copy number alterations (**Fig. 3.2-1** and **3.1-5**). Furthermore, $p53^{-/-}$ cells did not undergo senescence (54) and exhibited less strong copy number alterations than wild-type signature A MEFs (**Fig. 3.1-9** and **3.1-10**). Taken together, our results implicate senescence-associated redox stress as one of the selective forces driving the copy number alterations recurrently observed in human tumors.

3.3 Methods summary of chapter 3

CD1 mouse embryonic fibroblasts, E14.5, were purchased from Stem Cell Technologies. $p53^{fl/fl}$ MEFs were obtained at day E14.5 from $p53^{fl/fl}$ (FVB.129P2-Trp53tm1Brn/Nci) crossed with C57BL/6-129/SV mice. MEF cells were cultured using a 3T9 culturing protocol (43). To protect cells from reactive oxygen species (ROS), cells were cultured either at physiological oxygen concentrations (3% O2), with 250 U/ml catalase from bovine liver (Sigma-Aldrich), or with both 3% O2 and catalase.

Overexpression of the MYC oncogene or the control protein RFP in CD1 MEFs was accomplished by transduction of non-immortalized cells with pDS-FB-neo retrovirus, followed by selection in 600 µg/ml G418. Deletion of p53 in $p53^{fl/fl}$ MEFs was induced by infection of non-immortalized cells with either retroviral Cre-GFP or Cre-ERT2 plus treatment with 1 µM 4-OHT.

DNA from MEF lines and reference genomic DNA from C57BL/6J mouse tissue were hybridized to Agilent SurePrint G3 Mouse CGH 4x180k CGH microarray chips and processed using Bioconductor analysis tools.

A summary senescence score for each MEF subline was calculated by subtracting the area under the growth curve from the area under an ideal growth curve (i.e. consistent growth at the fastest observed rate).

More detailed descriptions of these methods can be found in **Chapter 8**.

CHAPTER 4

Cross-species analysis further implicates an impact of metabolic gene loci on the recurrent copy number alterations observed in aneuploid tumors.

4.1 Cross-species conserved CNA regions across human tumors and mouse *in vivo* and *in vitro* cancer models

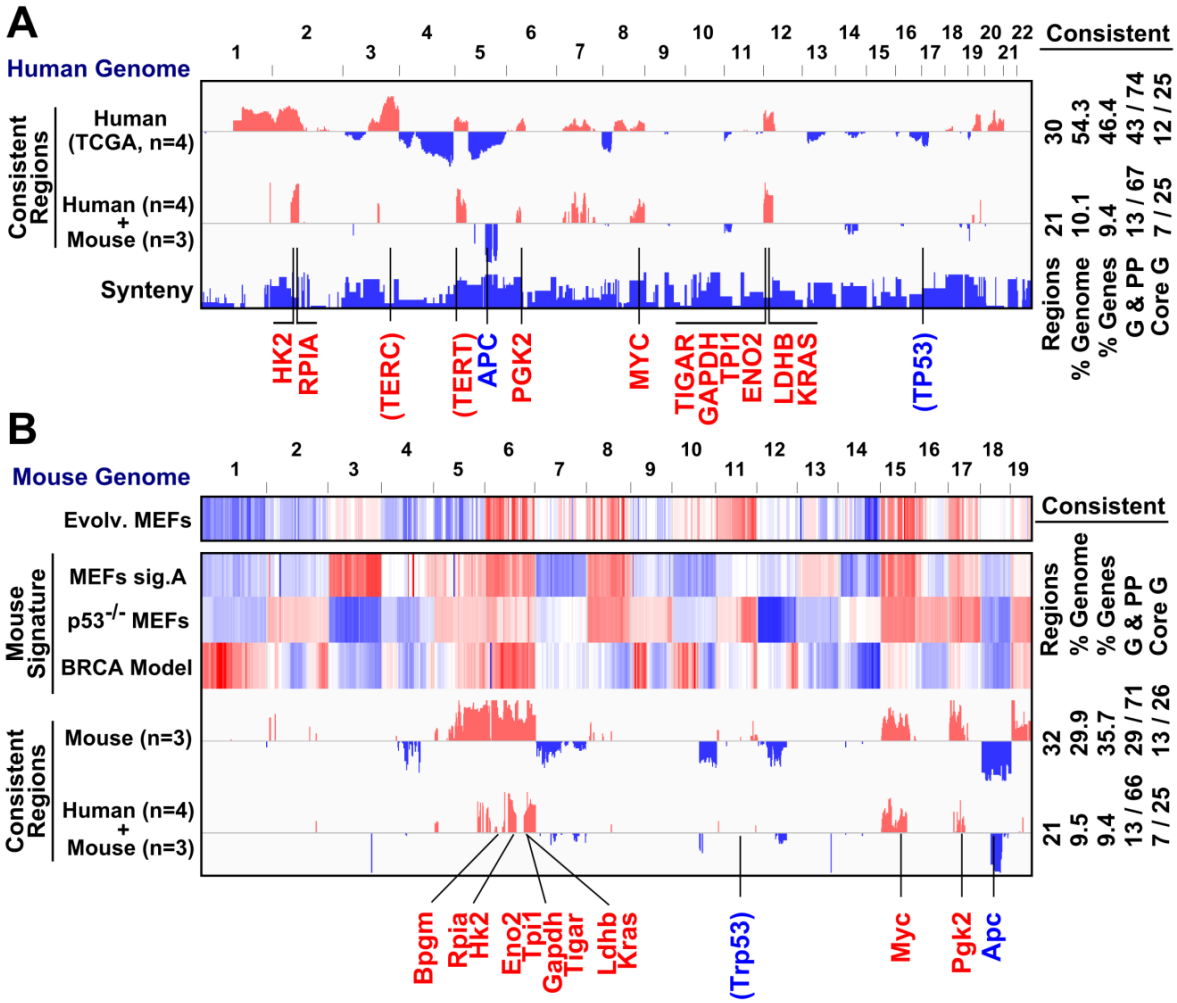


Figure 4.1-1. Cross-species consistency of PCA-defined shared signature A. (A-B) Conserved amplification and deletion loci in copy-number profiles across four human tumor types with consistent signature A patterns, and three mouse cancer model signatures. The four human signatures are based on the top 10% of PC1-based signature A tumors for breast, lung, ovarian and uterine, individually, as in **Fig. 2.1-3** (labeled as 'TCGA, n=4'). In the human consistent regions, the 54.3% consistently altered gene loci include bona-fide tumor suppressor genes (p53), oncogenes (MYC, KRAS), telomere components (TERC, TERT), and core glycolysis genes (eg, HK2) (A, displayed on the human genome). Mouse CNA signatures are normalized PC1 loadings for signature A wild-type MEFs, p53-/- MEFs, and a mouse breast carcinoma (BRCA) model involving disruption of p53 and Brca1 genes (51) (B, displayed on the mouse genome). A cross-species analysis of conserved amplification and deletion loci (human and mouse consistent regions, defined as in **Fig. 2.1-3**) greatly reduces (4-fold) the percentage of consistent copy number alterations, thereby reducing the percentage of the genome implicated as candidate driver regions (displayed in both A and B). Genes listed in parenthesis (TERC, TERT, and TP53) are not copy number altered in the p53 genetic knockout models but are consistent otherwise. The synteny graph indicates the syntenic mouse chromosome number and thus the broken synteny regions between human and mouse genomes. The conserved regions have a median size of 8.4 megabase pairs (sub-chromosomal arm resolution) and involve a median of 61 genes. A paired t-test analysis of signature A MEF lines profiled at more than one passage number revealed genomic regions associated with mid- to late-passage CNA evolution (log10 t-test p-value signed positive for amplifications, negative for deletions; panel B, labeled 'Evolv. MEFs').

While canonical oncogenes and tumor suppressor genes drive some recurrent DNA copy number alterations, many recurrent CNA regions cannot be fully explained by the presence of known cancer genes (24–26). To identify candidate driver genes for the cancer phenotypes shared by human tumors and mouse models, we used a cross-species comparison, and the presence of broken synteny between human and mouse genomes, to define conserved amplification and deletion loci. For human signatures, we used breast, lung, ovarian and uterine signature A tumors (**Fig. 2.1-3**). These four CNA signatures demonstrated consistent amplification or deletion in 54% of the genome (**Fig. 2.1-3** and **4.1-1A**). For mouse signatures, we used signatures defined from our wild-type MEFs (signature A), our $p53^{-/-}$ MEFs,

and a mouse model of p53-deficient and Brca1-deficent breast cancer (55). Together, these mouse signatures demonstrated consistent amplification or deletion in 30% of the genome (**Fig. 4.1-1B**). When the human and mouse signatures were combined via syntenic mapping, consistent genome regions dropped to covering only 10% of the genome, spread over 21 conserved regions (**Fig. 4.1-1**). Thus, cross-species analysis strongly reduced the number of candidate driver gene loci that play a role in shaping a core set of recurrent tumor CNA patterns.

Examination of the cross-species conserved genomic regions revealed consistent deletion of the regions containing the *TP53* and *APC* tumor suppressors and consistent amplification of the regions containing the *MYC* and *KRAS* oncogenes. The telomerase components *TERC* and *TERT* were also consistently amplified, except in the *p53* genetic knockout models. Notably, four of the cross-species conserved CNA genomic regions included genes from the glycolysis and pentose phosphate pathways: *HK2* and *RPIA* (chr. 2),

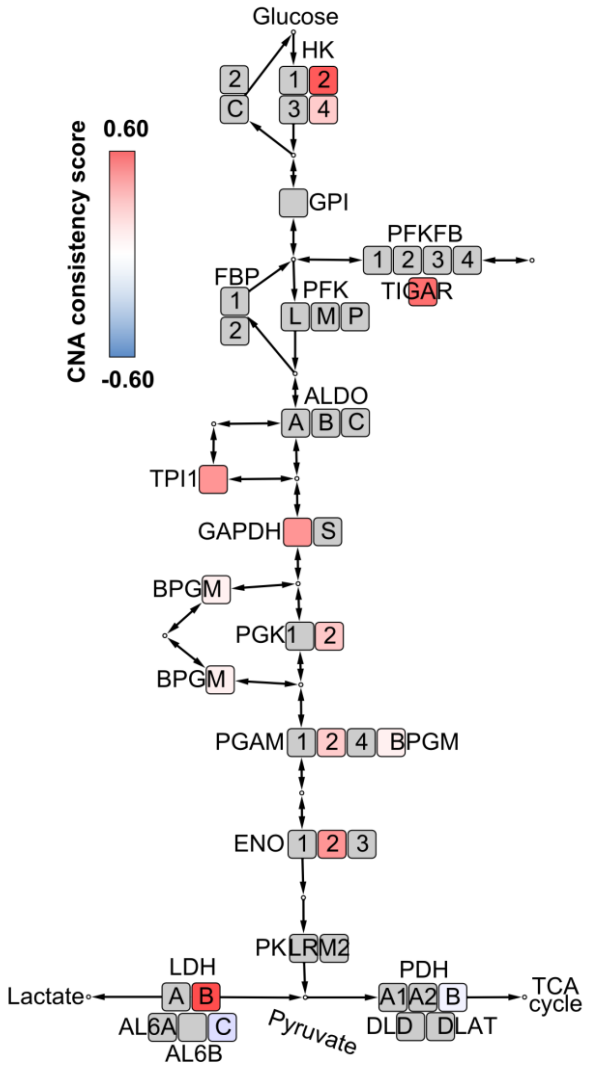


Figure 4.1-2. Schematic of the consistently altered genes in the core glycolysis pathway. Schematic of the consistently altered genes from **Fig. 4.1-1** in the core glycolysis pathway plus PFKFB, TIGAR, LDH and PDH. Genes are colored according to their strength of CNA-based consistency across the 4 human TCGA tumor types and 3 mouse cancer models.

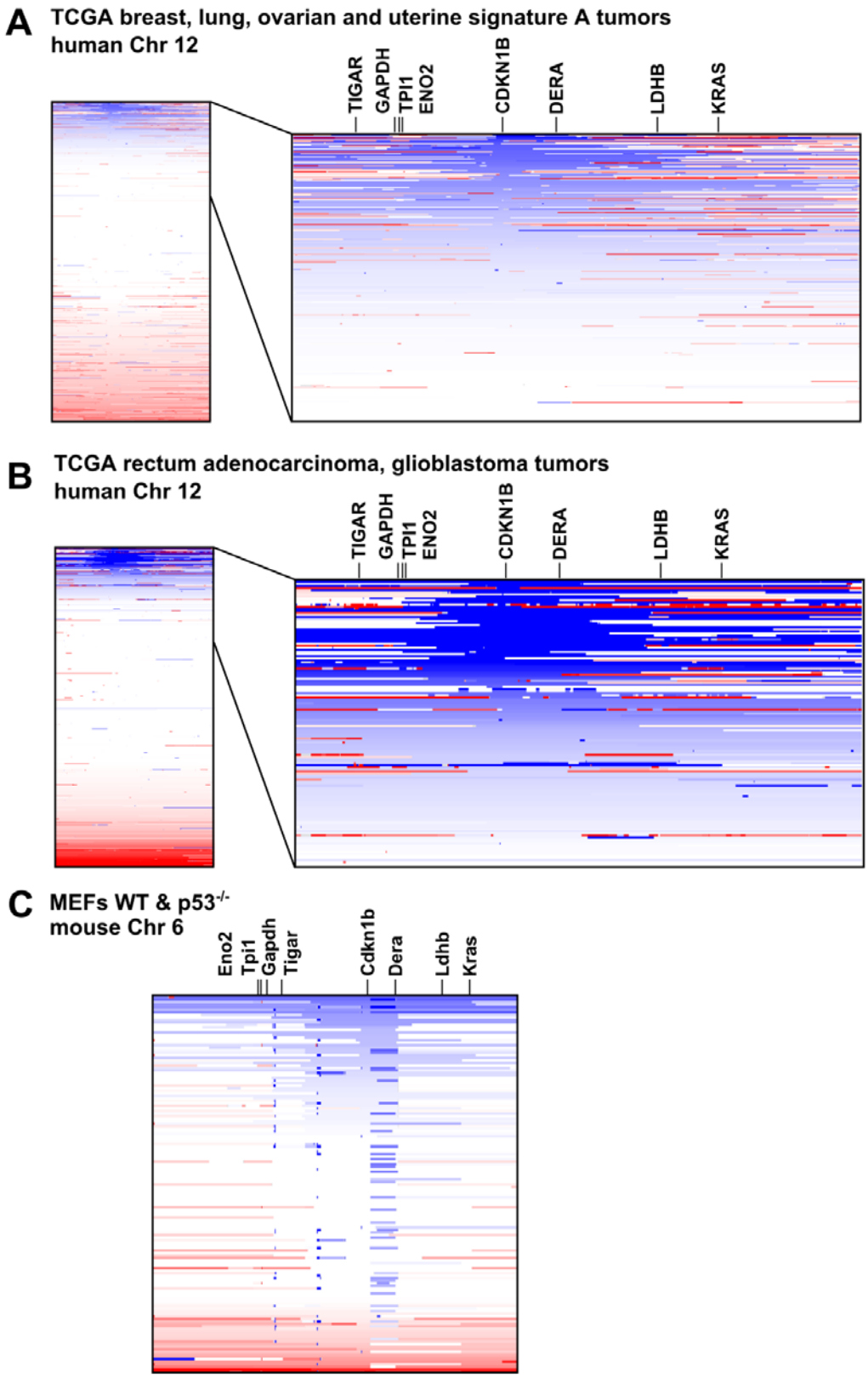

A
TCGA breast, lung, ovarian and uterine signature A tumors
human Chr 12
TIGAR
GAPDH
TPI1
ENO2
CDKN1B
DERA
LDHB
KRAS
B
TCGA rectum adenocarcinoma, glioblastoma tumors
human Chr 12
TIGAR
GAPDH
TPI1
ENO2
CDKN1B
DERA
LDHB
KRAS
C
MEFs WT & p53-/-
mouse Chr 6
Eno2
Tpi1
Gapdh
Tigar
Cdkn1b
Dera
Ldhb
Kras

Figure 4.1-3. A positive-negative-positive selection pattern around CDKN1B. (A-C) A positive-negative-positive selection pattern commonly observed around the tumor suppressor CDKN1B locus. In subsets of both human tumors and mouse models of cancer, the genomic region harboring TIGAR-GAPDH-TPI-ENO2 was separated from the LDHB-KRAS region by a deletion region that includes the tumor suppressor CDKN1B. (A) TCGA breast, lung, ovarian and uterine signature A tumors, human Chr12p region. (B) TCGA rectum adenocarcinoma and glioblastoma tumors. (C) MEFs, wild-type and p53-/-, mouse Chr6 region. Tumor subsets are indicated by the expansion lines.

ENO2, TPI1 and *GAPDH* (chr. 12, region 1, which includes *TIGAR*), *LDHB* (chr. 12, region 2, which also includes *KRAS*) and *PGK2* (chr. 6) (**Fig. 4.1-1** and **4.1-2**). In both human tumors and mouse models of cancer, the genomic region harboring *TIGAR-ENO2-TPI1-GAPDH* was separated from the *LDHB-KRAS* region by a deletion-prone region that includes the tumor suppressor *CDKN1B* (**Fig. 4.1-3**).

We next repeated the metabolic pathway enrichment analysis using the cross-species conserved regions. The cross-species approach substantially increased the enrichment scores for glycolysis-related pathways due to the synteny-based elimination of passenger genes

Signatures Included								
	Human (n=4)	-	-	+	+	+	+	+
	Evolv.MEFs	+	-	-	+	-	-	-
	MEFs sig. A	-	+	-	-	+	+	+
	p53 -/- MEFs	-	+	-	-	+	+	+
	BRCA Model	-	+	-	-	+	+	+
KEGG Metabolism Gene Set		**Normalized Enrichment Scores**				**NES**	**p-value**	**FDR**
Core Glycolysis		1.77	1.63	1.92	3.92	4.70	0.0001	0.01
Glycolysis & Pentose Phosphate		0.95	0.93	1.50	2.29	3.05	0.002	0.06
Glycolysis		1.04	1.13	1.60	2.30	2.82	0.004	0.09
Glycosphingolipid biosynthesis-ganglio series		1.21	0.90	0.97	1.86	2.26	0.03	0.48
Alanine, Aspartate & Glutamate Metabolism		0.79	1.06	1.06	0.51	1.94	0.05	0.48
Amino Sugar & Nucleotide Sugar Metabolism		0.96	1.19	1.63	1.16	1.86	0.04	0.48
Fructose & Mannose Metabolism		1.28	-0.66	0.66	0.91	1.81	0.06	0.51
...								

Table 4.1-1. Metabolic pathway enrichment analysis based on consistent CNA patterns. Metabolic pathway enrichment analysis based on consistent CNA patterns in the mouse and human signatures from **Fig. 4.1-1**, including the indicated cross-species combinations of signatures.

(**Table 4.1-1** and **Fig. 4.1-4A**). Improved enrichment results were obtained using human tumor-defined signatures combined with either i) individual mouse *in vitro*-derived signatures, or ii) the *in vivo*-derived signature from a mouse model of p53- and Brca1-deficent breast cancer (**Fig. 4.1-1** and **4.1-4B**). The strongest enrichment scores were observed when all human and mouse signatures were combined (**Fig. 4.1-4C**). Taken together, these results support a model in which the selection pressures shared during immortalization and tumorigenesis result in cross-species conservation of the glycolysis-related copy number alterations (**Fig. 4.1-1**).

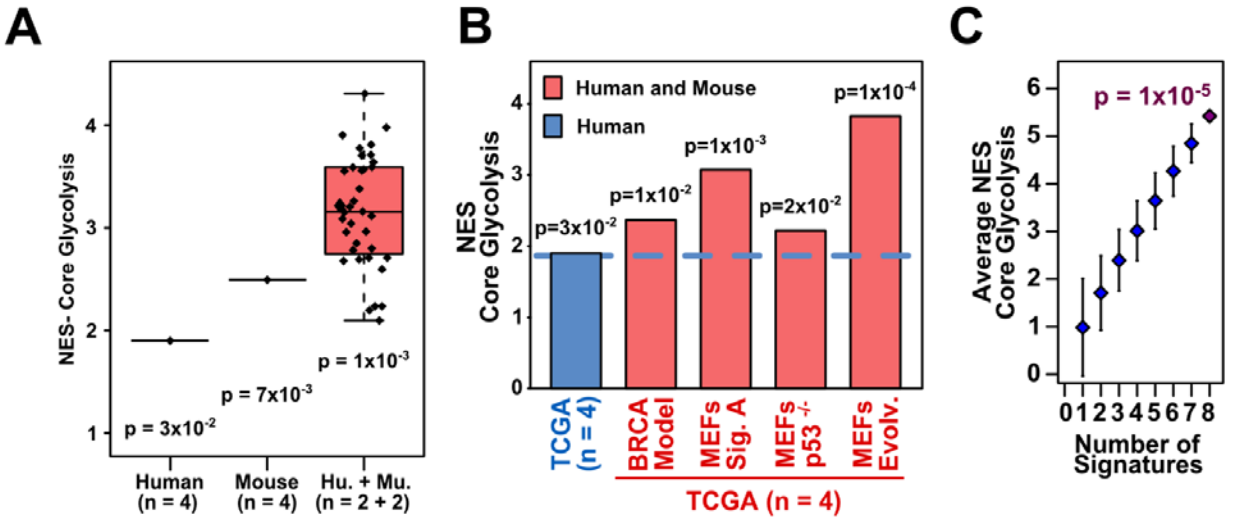


Figure 4.1-4. NES of combined consistency signatures. (A-B) Consistency-signature based normalized enrichment scores (NES) for the core glycolysis gene set improve when CNA signatures from mouse models are combined with the 4 TCGA-based human signatures (breast, lung, ovarian, uterine). For the case when multiple combinations are possible, the box plot representation is shown, (A). Nominal permutation p-values corresponding to the mean NES values are shown. (C) Consistency-signature based normalized enrichment scores (NES) for the core glycolysis gene set improve as additional human and/or mouse CNA signatures are included. Signatures used in this analysis are the 4 human and 4 mouse signatures from **Fig. 4.1-1B**. For each indicated number of signatures (n), averages and standard deviations of all possible combinations using n signatures are shown. The nominal permutation p-value corresponding to the NES value using 8 signatures is shown.

4.2 Cross-species regions are conserved across many tumor types

Although the signature A subtypes of breast, lung, ovarian, and uterine tumors were most similar in their genome-wide CNA patterns (**Fig. 2.1-1**), almost all tumor types analyzed demonstrated consistent amplification of the cross species-defined conserved alteration regions from **Fig. 4.1-1** (**Fig. 4.2-1**). In regards to glycolytic genes, 15 of 17 tumor types shared the cross-species conserved region containing *HK2* and *RPIA*, and 16 of 17 tumor types shared the two genomic amplification regions containing *TIGAR-ENO2-TPI1-GAPDH* and *LDHB-KRAS*. Thus, our cross-species and pan-cancer CNA analysis revealed conserved amplification regions shared by the majority of tumor types studied that are enriched for genes involved in core glycolysis. To aid others in pan-cancer and cross-species CNA signature comparisons, we have created an interactive web-interface resource available at http://systems.crump.ucla.edu/cna_conservation/.

4.3 Methods summary of chapter 4

Cross-species consistency signatures of conserved amplification and deletions (consistent regions) were defined as non-zero when all CNA summary signatures had the same sign across top 10% from the four human signature A tumor types (BRCA, LUSC, OV and UCEC) and three mouse signatures (CD1 MEF signature A, $p53^{-/-}$ MEFs and published data of mouse model of breast cancer). Consistent regions defined by cross-species analysis of the base seven signatures (four human tumor types and three mouse models) were further analyzed for consistency with the CNA summary from other tumor types/subtypes (HNSC, BLCA, LUAD, COAD, READ, SKCM, LGG, GBM, KIRC, STAD, PRAD, THCA, BRCA luminal).

More detailed descriptions of these methods can be found in **Chapter 8**.

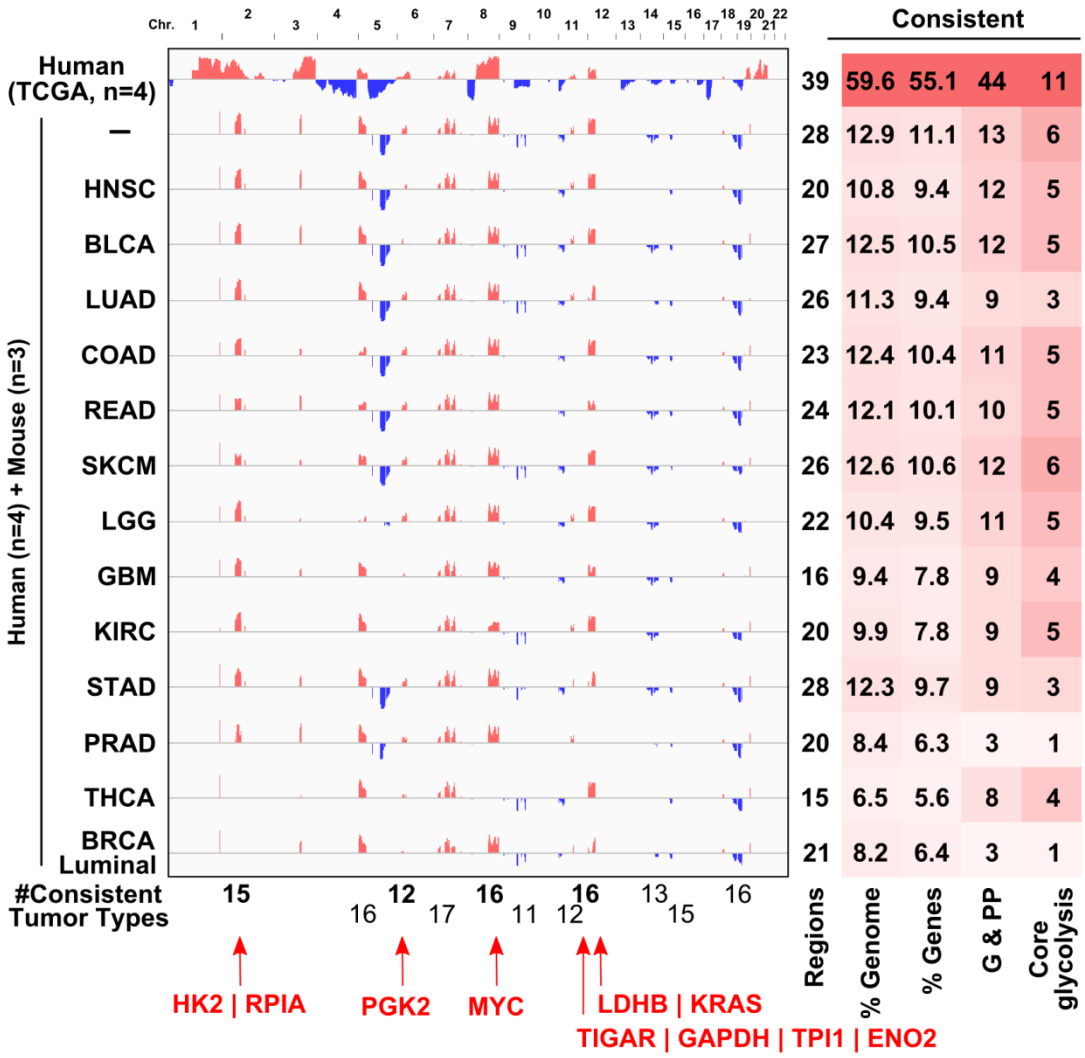


Figure 4.2-1. Cross-species regions are conserved across many tumor types. The consistently altered regions defined by cross-species analysis across 4 human tumor types (breast, lung, ovarian and uterine) with consistent signature A and three mouse cancer model signatures (**Fig. 4.1-1**) are also amplified or deleted in many additional tumor types. After the first two rows, each subsequent row shows the consistent regions profile upon inclusion of the additional tumor type indicated. Tumor type abbreviations are as defined by TCGA and listed in **Chapter 8.2.6**. Here basal and luminal breast cancer are treated as separate tumor types. The numbers of human tumor types consistently amplifying or deleting each region are indicated.

CHAPTER 5

Exogenous expression of metabolic enzymes reduces the propensities for amplification of the corresponding endogenous loci.

5.1 Alteration of CNA signatures by exogenous expression of metabolic enzymes

The presence of core glycolysis genes in cross-species conserved amplification regions suggests that these metabolic gene loci drive the amplification of these regions. To test this hypothesis, we transduced pre-senescent MEFs with either wild-type HK2 or HK1, kinase-dead HK2 (D209A/D657A) (56) or wild-type ENO2 and allowed the cells to senesce and immortalize in the presence of these exogenously expressed proteins (**Fig. 5.1-1**). Analyzing the signature A set of sublines, we found that the endogenous *Hk2* locus (chr. 6) was less amplified in cells expressing exogenous wild-type hexokinase (p = 0.048) (**Fig. 5.1-2** and **5.1-3A**).

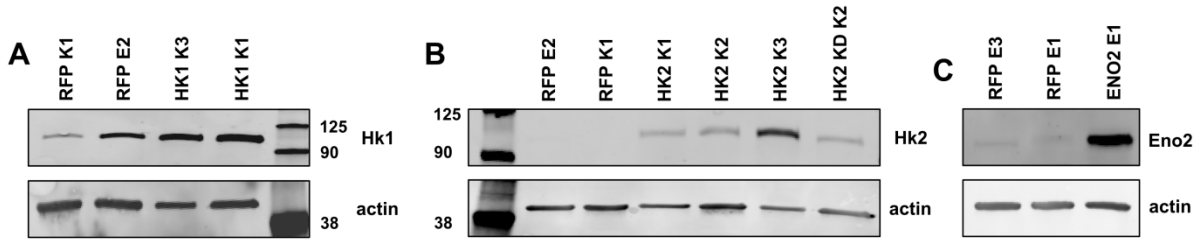


Figure 5.1-1. Exogenous HK1, HK2 and ENO2 expression of representative MEFs. Exogenous HK1, HK2 and ENO2 expression. MEF lines expressing exogenous RFP, wild-type HK1, wild-type HK2, kinase-dead HK2 (D209A/D657A), or wild-type ENO2 were lysed and probed by immunoblotting for Hk1 (A), Hk2 (B) and Eno2 (C). Actin was included as an equal loading control.

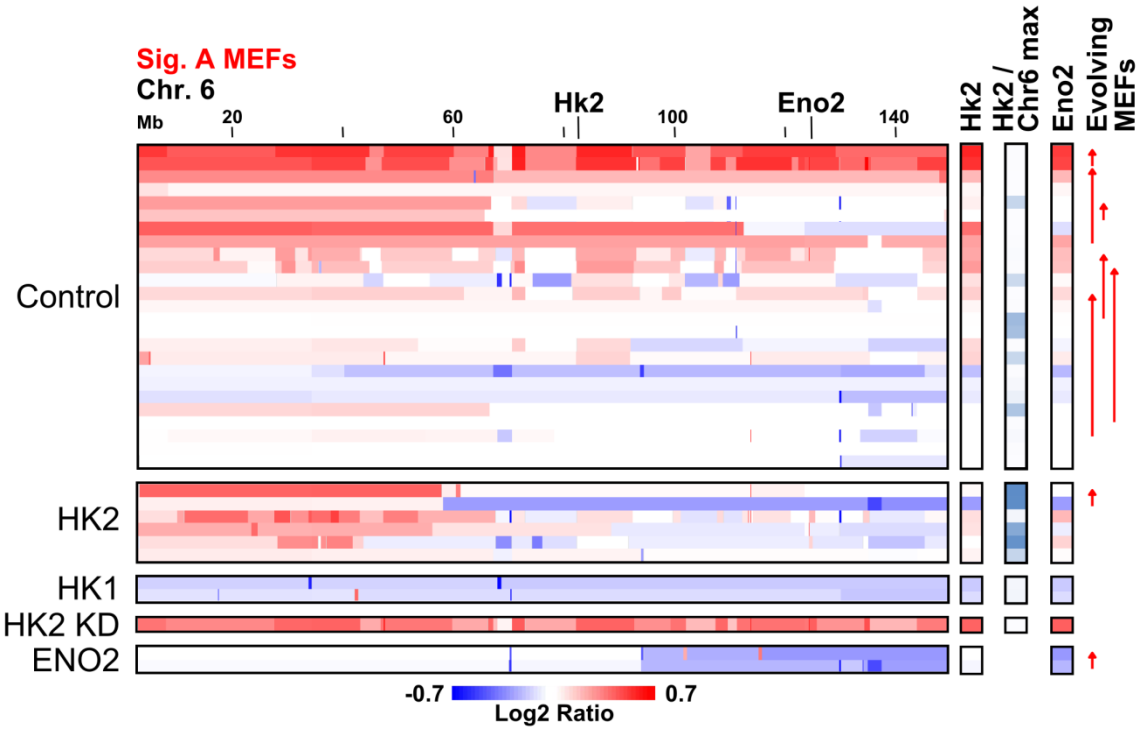


Figure 5.1-2. Signature A MEFs expressing exogenous HK1/HK2 or ENO2. CD1 MEFs expressing exogenous HK1/HK2 or ENO2 exhibit reduced amplification of the endogenous Hk2 or Eno2 loci, respectively. Chr. 6 copy-number profiles from control signature A MEFs (untransduced or expressing red fluorescent protein (RFP)) compared to signature A MEFs expressing either wild-type HK1 or HK2, kinase-dead HK2 (HK2 KD, D209A/D657A), or wild-type ENO2. The positions of endogenous Hk2 and Eno2 are indicated. MEF lines profiled at more than one passage number are indicated by the start (earlier passage) and end points of the arrows under 'Evolving MEFs'.

In contrast, a signature A MEF cell line expressing kinase-dead hexokinase did not show reduced amplification of the *Hk2* locus ($p = 2 \times 10^{-4}$). In that MEF lines exogenously expressing hexokinase still demonstrated positive selection for the centromere-proximal half of chr. 6 (**Fig. 5.1-2**), we examined the ratio of *Hk2* gene locus copy number to the maximal amplification on chromosome 6 (Hk2:Chr6 max). In this analysis, we found that cells expressing exogenous hexokinase demonstrated significantly reduced Hk2:Chr6 max ratios ($p = 3 \times 10^{-3}$) (**Fig. 5.1-2**

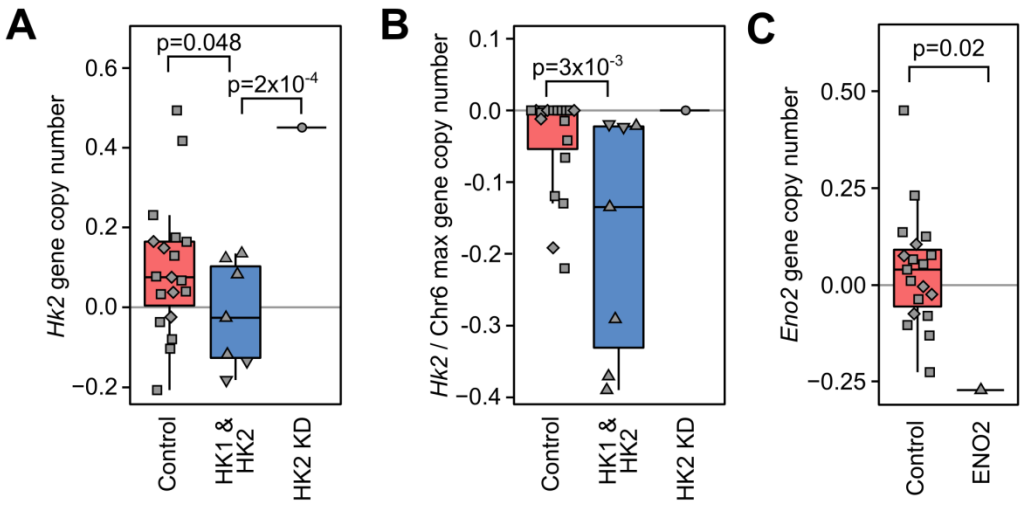


Figure 5.1-3. CNA of signature A MEFs expressing exogenous Hk1/Hk2 or Eno2. (A-C) Boxplots of Hk2 copy number (A), ratio of Hk2 copy number to maximum amplification of chromosome 6 (defined operationally as the 5th percentile CNA value across chr. 6 to avoid outlier effects) (B), and Eno2 copy number (C). (Control MEFs: untransduced (squares) or expressing RFP (diamonds); test MEFs expressing wild-type HK1 (upside-down triangles), HK2 (triangles in A-B), kinase-dead HK2 (HK2 KD, D209A / D657A, circles), or wild-type ENO2 (triangle in C).) Indicated p-values are one-tailed Student's t-test for normally distributed data, Mann-Whitney otherwise, and z-score for data with a single data point in one comparison group). Copy number data from samples profiled at multiple passages were averaged to prevent overrepresentation of these cell lines.

and **5.1-3B**). Additionally, a MEF subline expressing exogenous ENO2 exhibited deletion rather than amplification of the *Eno2* locus on chr. 6 (p = 0.02) (**Fig. 5.1-2** and **5.1-3C**). Analyzing the signature B set of sublines, we found that the endogenous *Hk1* locus (chr. 10) was copy number neutral, rather than amplified, in a cell line expressing exogenous hexokinase (p = 0.17), whereas a signature B MEF cell line expressing kinase-dead hexokinase did not show reduced amplification of the *Hk1* locus (**Fig. 5.1-4**). Taken together, these results demonstrate that exogenous expression of metabolic enzymes can alter the copy number status of the endogenous genomic loci, supporting these metabolic genes as drivers within the conserved amplification regions observed in human tumors and mouse models.

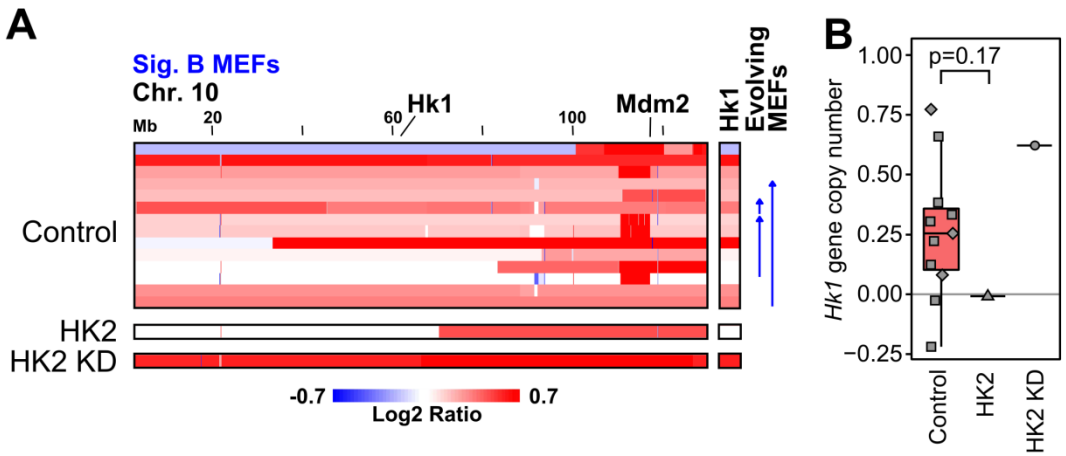


Figure 5.1-4. Signature B MEFs expressing exogenous HK2. (A) Signature B MEFs expressing exogenous HK2 exhibit less amplification of the Hk1 locus Chr. 10 copy-number profiles from control signature B MEFs (untransduced or expressing RFP) compared to signature B MEFs expressing wild-type HK2 or kinase-dead HK2 (HK2 KD). The positions of endogenous Hk1 and Mdm2 are indicated. (B) Boxplot of the Hk1 copy number with indicated p-value (z-score).

For all comparisons in this figure and in **Figure 5.1-3**, no significant differences were observed between untransduced and RFP-expressing MEFs. The Fisher's combined p-values for the alteration of endogenous loci CNA by exogenous expression of metabolic enzymes is 3 × 10-5 (**Fig. 5.1-3A,C** and **panel B**) or 7 × 10-4 (**Fig. 5.1-3B,D** and **panel B**).

5.2 Methods summary of chapter 5

Overexpression of HK1, HK2, kinase-dead HK2 (D209A/D657A) or ENO2 glycolysis enzymes or the control protein RFP in CD1 MEFs was accomplished by transduction of non-immortalized cells with pDS-FB-neo retrovirus, followed by selection in 600 µg/ml G418.

Primary antibodies used for western blot analysis included Hexokinase 1 (2024, Cell Signaling Technology), Hexokinase 2 (2867, Cell Signaling Technology), p53 (NB200-103, Novus Biologicals), and Enolase 2 (8171, Cell Signaling Technology).

More detailed descriptions of these methods can be found in **Chapter 8**.

CHAPTER 6

Metabolism and growth phenotypes scale with CNA signatures

6.1 Characterization of metabolic and growth changes associated with copy number alteration signatures

To test whether there exist phenotypic differences between signature A and signature B MEFs, we characterize ed 11 wild-type MEF lines representative of either signatures A or B and one mixed signature line. We found that signature A MEFs generally had higher rates of glucose consumption and lactate production than signature B MEF lines (**Fig. 6.1-1**). Plotting the PC1 score versus glucose consumption, we found that CNA-based signature A was highly predictive of glucose consumption in the MEF signature A lines (**Fig. 6.1-2A**). In contrast, signature B

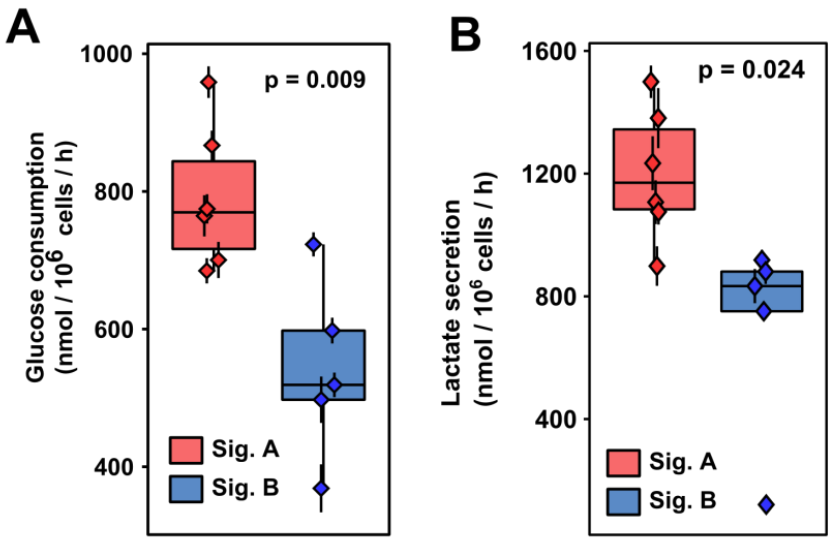


Figure 6.1-1. Metabolic differences of PCA-defined signatures. PCA-defined signature A MEFs exhibit higher glucose consumption (A) and lactate production (B) rates than signature B MEFs. Glucose consumption and lactate secretion as measured by a bioanalyzer. Error bars indicate standard deviations of biological replicates.

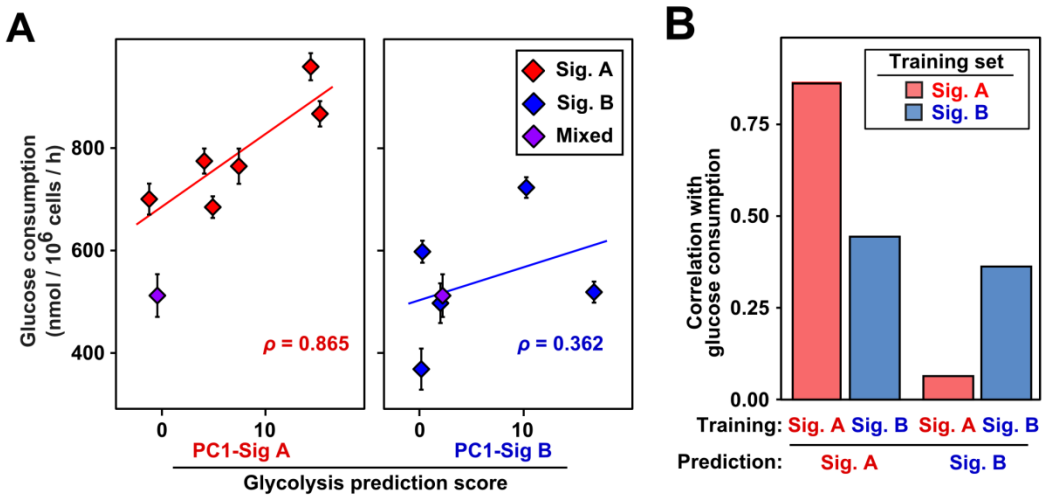


Figure 6.1-2. PCA-based MEFs signature A is predictive of glycolysis. (A) PC1 scores of signature A MEFs are predictive of glucose consumption rates in signature A MEFs, while scores from signature B MEFs have weaker predictive power in signature B MEFs. Error bars indicate standard deviations of biological replicates. (B) In cross-prediction tests, signature A-based predictions of signature A MEF line metabolic phenotypes perform the best. Glucose consumption as measured by a bioanalyzer.

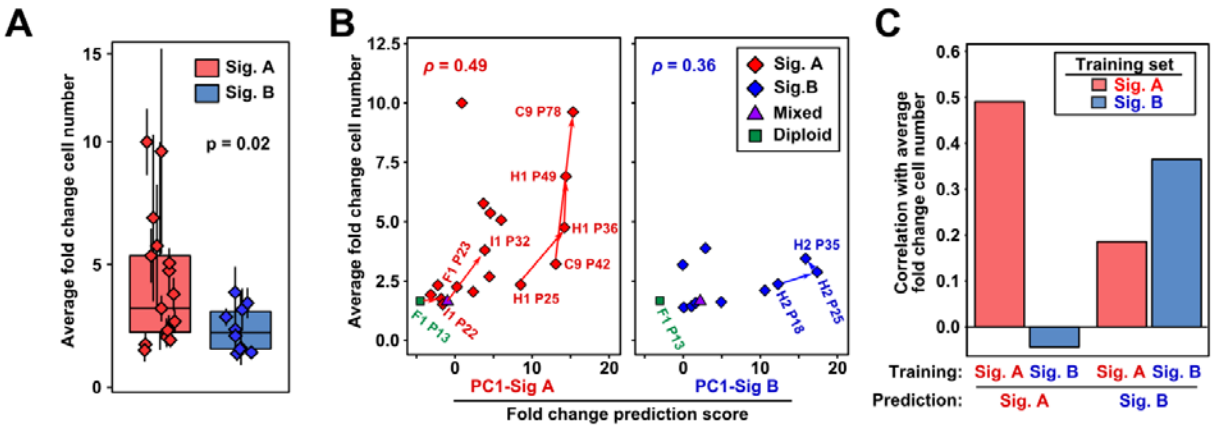


Figure 6.1-3. Growth properties of PCA-defined MEFs signatures. (A) PCA-defined signature A MEFs exhibit higher average growth fold change, as observed in 3T9 culture, compared to signature B MEFs. Error bars indicate standard deviations of biological replicates. (B) Correlation and a general co-evolution of higher growth rates and increased CNA signature strength. Arrowed lines indicate progressing MEF lines profiled at more than one passage number (**Fig. 3.1-5** and **3.1-6**). (C) In cross-prediction tests, PC1 scores of signature A MEFs are more predictive of average growth fold change in signature A MEFs, while scores from signature B MEFs are more predictive in signature B MEFs.

was only moderately predictive of the glucose consumption of signature B lines, and was not as accurate as signature A in predicting the glucose consumption of signature A lines (**Fig. 6.1-2**). In addition, we noted that the signature A cell lines generally exhibited higher rates of proliferation than the cell lines in signature B (**Fig. 6.1-3A**).

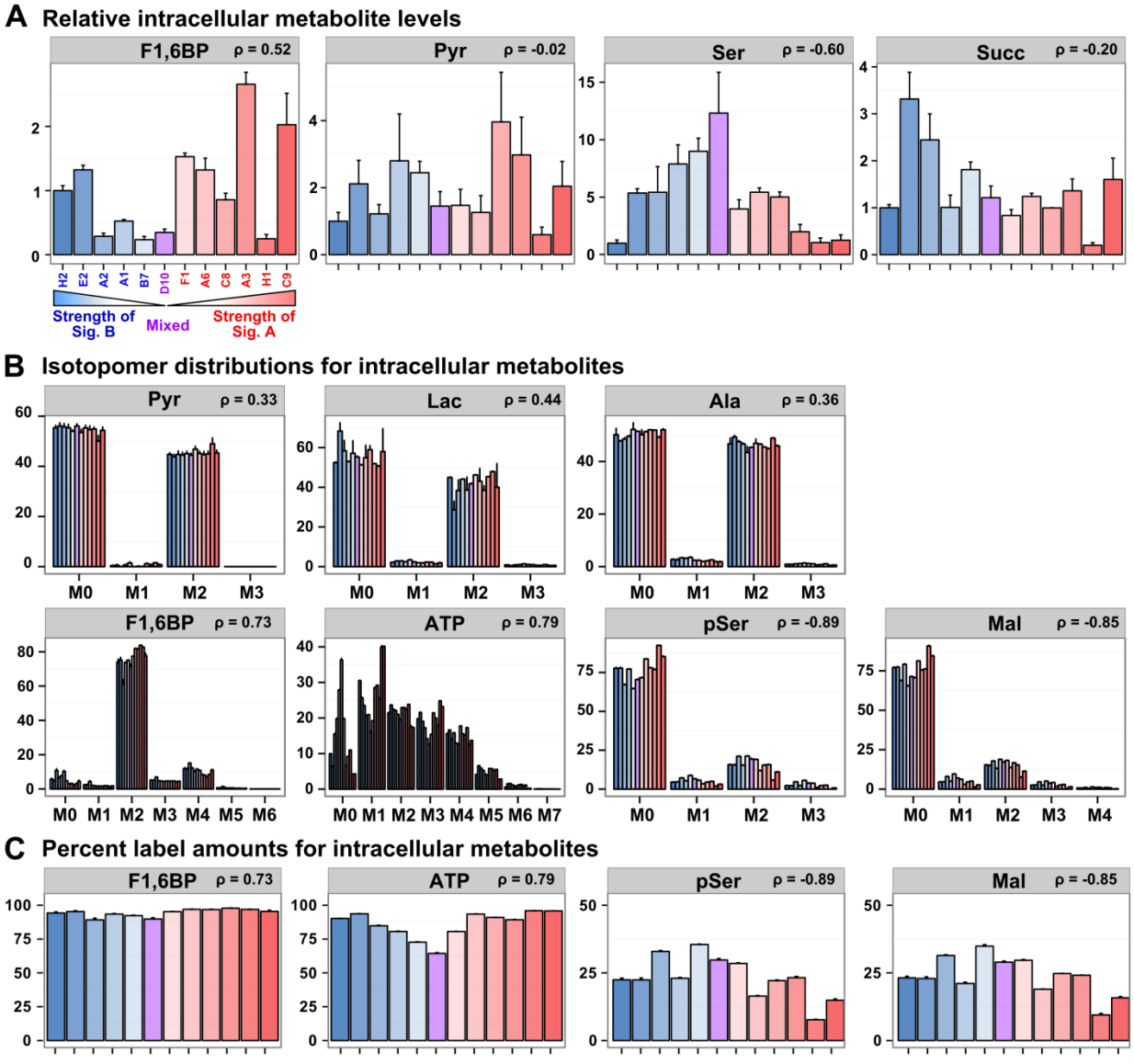


Figure 6.1-4. Mass spectrometry-based intracellular metabolite measurements. Samples are arranged left to right from the strongest signature B (blue) PC1 scores to the strongest signature A PC1 (red) scores. Sample D10 (purple) is a mixed signature A/B line. (A) Relative intracellular metabolite levels. (B) Isotopomer distributions for intracellular metabolites (M0, monoisotopic molecular weight). (C) Percent label for intracellular metabolites defined as percent of metabolite molecules with isotopomer mass greater than the monoisotopic molecular weight. Note, the high to low range for percent metabolite molecules with incorporated label varies for each metabolite and generally does not extend from 0 to 100%. In panels B-C, the correlation of each metabolic parameter with glucose consumption (as measured by a bioanalyzer, and as a positive variable, **Fig. 6.1-1A**) is indicated (Pearson correlation, ρ). In the isotopomer cases (B), the summary metric of percent label is used for the correlation determination. Notably, cells with high glucose consumption rates tend to exhibit high consumption of serine, glycine and glutamine; and high secretion of lactate (with concordant lactate secretion results obtained by bioanalyzer-based measurement of the same spent cell culture media). Error bars indicate standard deviations of biological replicates.

Similar to glucose consumption, signature A was predictive of the growth rates of signature A MEFs and signature B was predictive of the growth rates of signature B MEFs, while cross-signature predictions had less power (**Fig. 6.1-3B,C**). Furthermore, we observed a general co-evolution of higher growth rates and increased CNA signature strength in MEF lines that were profiled at different passage numbers (**Fig. 6.1-3B**). As noted above, the evolving MEF CNA signature pattern was enriched for DNA amplifications of genes in the core glycolysis and glycolysis-associated pathways (**Table 4.1-1**), particularly due to amplification of chromosome 6, which contains multiple metabolic gene loci including *Hk2* and *Rpia* (**Fig. 3.1-2**, **4.1-1B**, **3.1-6** and **3.1-7**).

To test whether signature A and signature B MEFs differentially use glucose, we cultured MEF cells with [1,2-^{13}C]-labeled glucose and conducted metabolomic profiling by mass spectrometry (**Fig. 6.1-4** and **6.1-6**) (57). In all MEF lines tested, we observed a low percentage of single heavy labeled carbon [M1 isotopomer compared to M2] in pyruvate,

lactate, and alanine, indicating that the contribution of glucose-derived carbon from the oxidative arm of the pentose phosphate pathway to these metabolites was relatively low (**Fig. 6.1-4B**). Nonetheless, the patterns of heavy isotope labeling revealed differences in nutrient utilization between signature A and B MEF cell lines. On average, metabolites of early glycolysis, the pentose phosphate and nucleotide synthesis pathways showed a higher percentage of glucose-derived heavy carbon labeling in signature A MEFs (**Fig. 6.1-5** and **6.1-4C**).

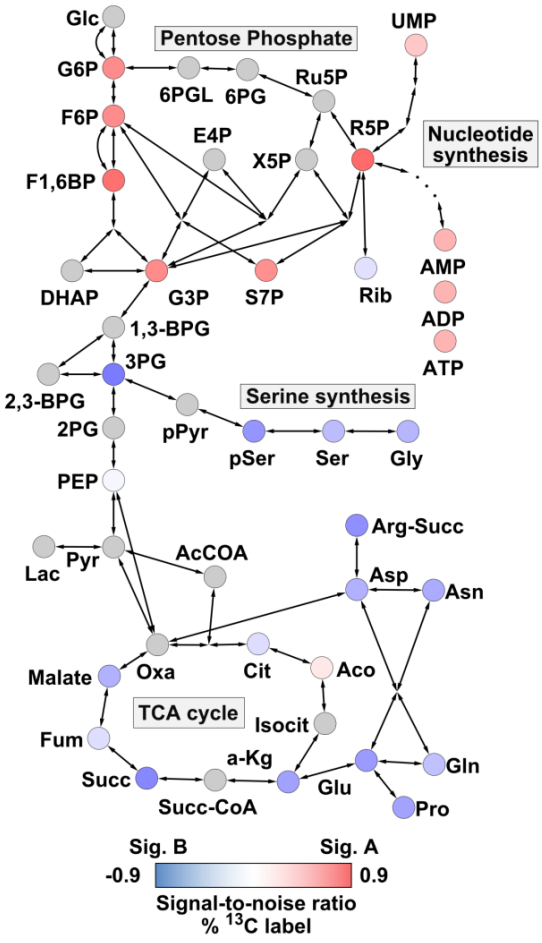


Figure 6.1-5. Signature-based metabolic pathway schematic of heavy labeled glucose in MEFs. Metabolomic profiling of 11 wild-type MEF lines representative of either signatures A or B cultured for 24 h with [1,2-13C]-labeled glucose. The metabolic pathway schematic is colored based on differences observed in the percent heavy label for intracellular metabolites (defined as percent of metabolite molecules with isotopomer mass greater than the monoisotopic molecular weight, M0) between signature A and B MEFs. Red indicates a higher heavy carbon-labeling percentage in Signature A MEFs, and blue indicates a higher heavy carbon-labeling percentage in signature B MEFs. Signature A MEF lines incorporate more glucose-derived carbon in metabolites from the early glycolysis steps, the pentose phosphate pathway, and nucleotide synthesis; and a smaller fraction of glucose-derived carbon per molecule in metabolites from the later glycolysis steps, the serine synthesis pathway, and the TCA cycle.

In contrast, signature A cells had a lower percentage of glucose-derived heavy carbon labeling in metabolites of the serine synthesis pathway and the TCA cycle. When compared to signature B MEFs, signature A cells also tended to exhibit increased consumption of serine,

glutamine, and other amino acids (**Fig. 6.1-6**). In general, the percentage of metabolites from the early glycolysis, the pentose phosphate and nucleotide synthesis pathways that incorporated heavy, glucose-derived carbon metabolites was positively correlated with glucose consumption (eg, fructose-1,6-bisphosphate and ATP, **Fig. 6.1-7**).

Conversely, metabolites from the serine synthesis pathway and the TCA cycle showed a negative correlation between glucose consumption and the percentage of each molecule containing a heavy, glucose-derived carbon (eg, 3-phosphoserine and malate, **Fig. 6.1-7**).

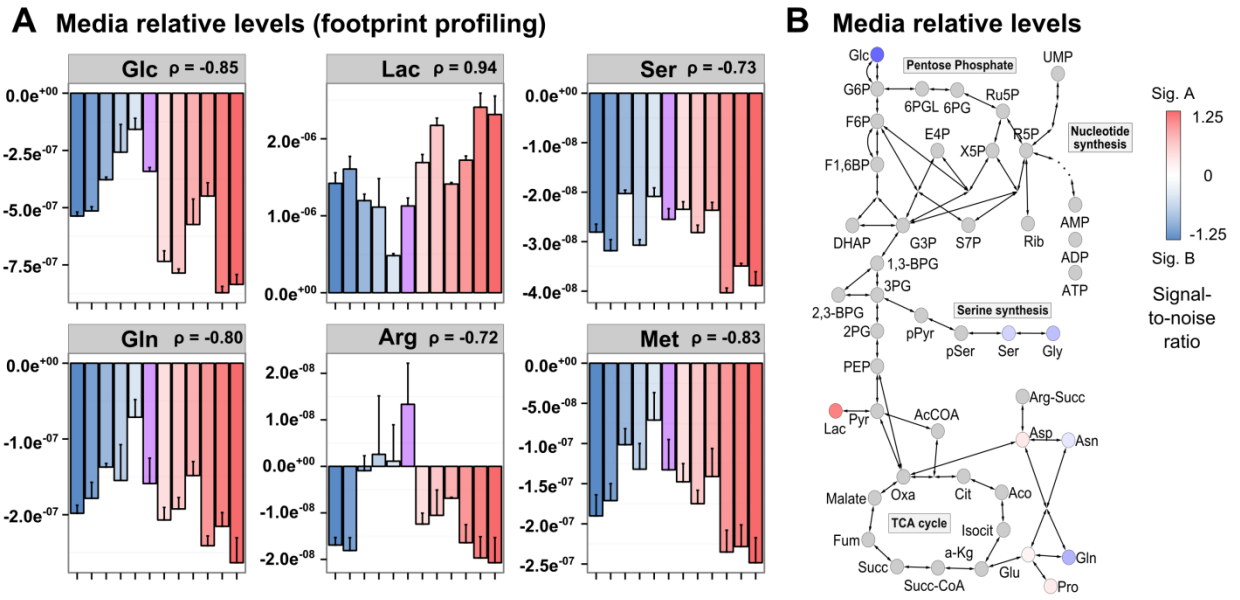


Figure 6.1-6. Mass spectrometry-based metabolite measurements of media relative levels. Samples are arranged left to right from the strongest signature B (blue) PC1 scores to the strongest signature A PC1 (red) scores. Sample D10 (purple) is a mixed signature A/B line. (A) Relative metabolite consumption from and secretion into media normalized to initial media levels (ie, cellular metabolite footprint). Here, positive and negative values indicate metabolite secretion into and consumption from the media, respectively. In panel A, the correlation of each metabolic parameter with glucose consumption (as measured by a bioanalyzer, and as a positive variable, **Fig. 6.1-1A**) is indicated (Pearson correlation, ρ). (B) Pathway schematics of differences in relative media levels (footprint, from panel A) between signature A and signature B MEFs. Each node is colored according the signal-to-noise ratio. Red indicates higher post-culture media levels (lower consumption or higher secretion rates) in signature A MEFs, and blue indicates higher post-culture media levels in signature B MEFs. Error bars indicate standard deviations of biological replicates.

Of note, differences between signature A and B lines were mainly in regards to scale, with the strongest signature B lines demonstrating similar glycolysis and proliferation rates, as well as similar metabolic profiles, as the weakest signature A lines.

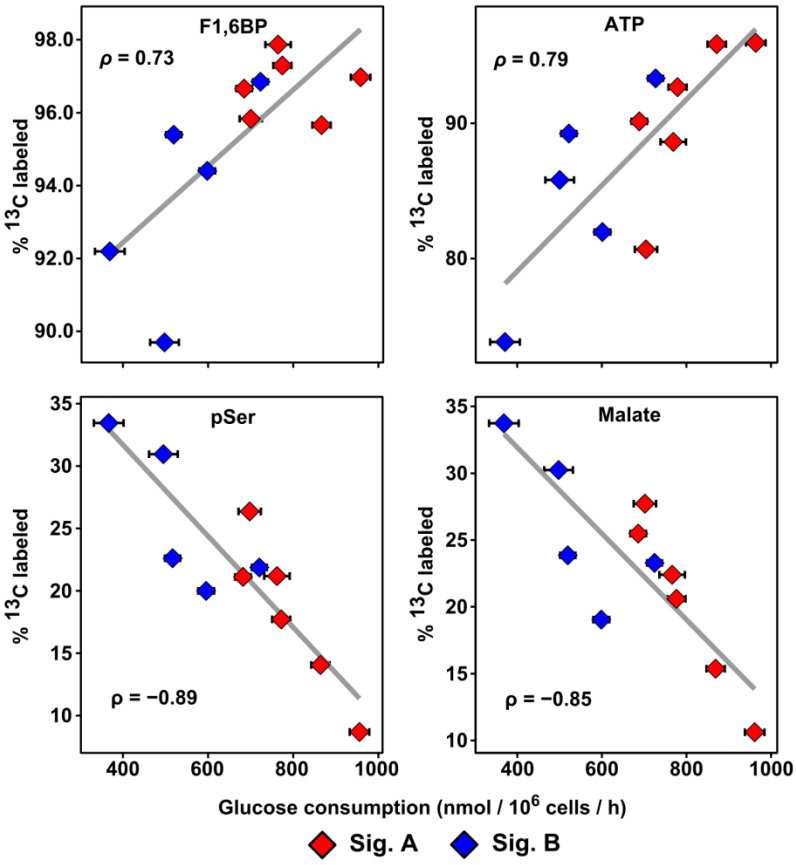


Figure 6.1-7. Correlation of selected metabolites with glucose consumption. Correlation of fructose-1,6-bisphosphate (F1,6BP), adenosine triphosphate (ATP), 3-phosphoserine (pSer) and malate with glucose consumption. Note, the high to low range for percent metabolite molecules with incorporated label varies for each metabolite, but generally does not extend from 0 to 100%. Signature A and B MEFs are colored red and blue, respectively. Error bars indicate standard deviations of biological replicates.

Taken together, these results demonstrate that signature A MEFs, which resemble signature A human tumors, exhibit increased glycolysis and proliferation, and have an increased

relative proportion of glucose-derived carbon in metabolites of pentose phosphate-associated biosynthetic pathways such as nucleotide synthesis. These findings are consistent with published mouse model studies in which tumor cells that are channeling glucose towards nucleotide biogenesis achieve faster rates of proliferation (58,59).

6.2 Methods summary of chapter 6

Lactate secretion rates of breast cancer cell lines were measured from the culture media using a colorimetric assay kit (BioVision). Glucose consumption and lactate secretion rates of MEFs were measured using a BioProfile Basic bioanalyzer (NOVA Biomedical). All samples were run as biological triplicates and in multiple independent experiments.

For the mass spectrometry-based metabolomic analyses, MEF sublines were cultured in media containing 4.5 g/L [1,2-^{13}C]-labeled glucose for 24 h. Cell culture media for footprint profiling was collected and metabolites were extracted with ice-cold 80% methanol. For intracellular metabolite analysis, cells were washed with cold 150 mM ammonium acetate pH 7.3, and metabolites extracted in cold 80% methanol. Samples were run on a Q-Exactive mass spectrometer (MS) coupled to an UltiMate 3000RSLC UHPLC system (Thermo Scientific). The MS was run in polarity switching mode. Metabolites were separated on a Luna 3 µm NH2 100 Å (150 x 2.0 mm) (Phenomenex) column. Relative amounts of metabolites between various conditions, percentage of metabolite isotopomers (relative to all isotopomers of that metabolite), and percentage of labeled metabolite molecules (isotopomer M1 and greater, relative to all isotopomers) were calculated, and corrected for naturally occurring ^{13}C abundance. All samples were run as biological triplicates and subjected to statistical tests across independent experiments as described in the extended experimental procedures.

More detailed descriptions of these methods can be found in **Chapter 8**.

CHAPTER 7

Discussion

7.1 Overview

Highly aneuploid and highly glycolytic tumors are some of the most aggressive cancers, and the complexity and plasticity of their genomes can hinder molecularly targeted therapies (35,60,61). While the glycolytic changes associated with tumorigenesis were one of the early defining phenotypes of cancer cells (62), they have not previously been linked to recurrent DNA copy number patterns. Taken together, our data support a model in which metabolism-linked selective pressures encountered during tumorigenesis (eg, redox stress and senescence) shape the highly recurrent DNA copy number alterations found in aneuploid human tumors (**Fig. 7.1-1**). We found that CNAs in core glycolysis enzymes (eg, HK2) and other cancer-linked metabolic enzymes such as TIGAR and RPIA are coordinately enriched in tumors with distinct CNA signatures. These CNA signatures are predictive of glycolysis, including patient FDG-PET activity, and proliferation phenotypes. In that exogenous expression of hexokinase and enolase enzymes reduced the propensities for amplifications of the corresponding endogenous hexokinase and enolase loci, these metabolic genes empirically score as driver loci within the conserved amplification regions. Combined with the observation that metabolic genes can facilitate cellular immortalization (16–18), our results implicate tumor metabolism as an additional fitness measure linked to how genomic instability can enable tumorigenesis.

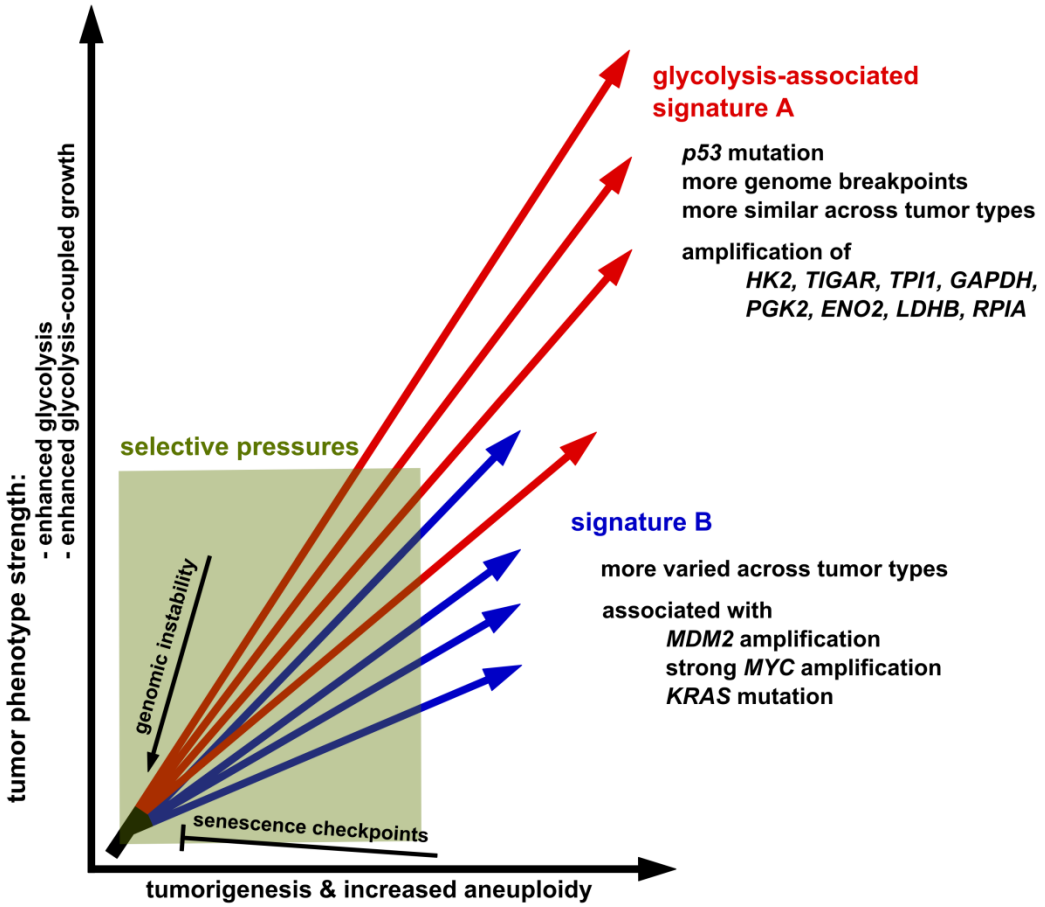


Figure 7.1-1. Genomic instability enables fitness gains in tumor metabolism. In human tumors, cancer cell lines, and an experimental MEF immortalization system, immortalization and tumorigenesis lead to multiple CNA signatures that are predictive of the tumor phenotypes of metabolism and proliferation. Senescence-associated redox stress and other tumorigenesis-related constraints select for stronger CNA signatures. A shared high glycolysis-associated signature A is observed in breast, lung, ovarian and uterine tumors and the MEF model system, and is linked with a stronger range of glycolysis and proliferation phenotypes. Genetic manipulation of glycolysis enzymes leads to alteration of corresponding CNA signature propensities. Signature A and B genomes reflect two distinct trajectories from diploidy to tumor aneuploidy. Signature A tumors are enriched for mutations in p53 and have smaller sized amplification and deletion genomic regions (ie, have a higher number of genomic breakpoints), potentially providing increased alternative genome options. Signature A involves amplification of several genes in glycolysis-related pathways (such as HK2, TIGAR, TPI1, GAPDH, PGK2, ENO2, LDHB and RPIA). Signature B CNA patterns occur in generally less glycolytic and proliferative samples and show more variation across different tumor types. In particular, signature B is enriched for MDM2 amplification, strong MYC amplification, and KRAS mutation; and involves alternate hexokinase isoforms (HK1, HK3).

7.2 Redox stress, biomass accumulation and associated glycolytic changes in tumorigenesis

Tumorigenesis is a complex, multi-stage process during which cells must acquire the capability to maintain redox balance while accumulating the macromolecular precursors required for proliferation (2,4). Numerous stimuli, including RAS mutations, matrix detachment, altered metabolism and hypoxia, induce the accumulation of intracellular ROS (63–66). Because increased ROS levels can trigger replicative senescence and subsequent cell cycle arrest (66,67), tumors must maintain pools of reduced glutathione using NADPH in part produced via the pentose phosphate pathway. Additionally, increased levels of ROS can divert glycolytic flux into the pentose phosphate pathway through, for example, oxidation and inhibition of the glycolytic enzyme PKM2, thereby supplying cells with the reducing power and precursors for anabolic processes (63,59). Consistent with this published knowledge on the role of metabolism in tumorigenesis, our study suggests that the metabolic stress associated with senescence (**Fig. 3.1-2** and **3.2-1**) and the metabolic demands of rapid proliferation (**Chapter 6**) are components of the selective pressures underlying recurrent CNA changes.

7.3 Experimentally and computationally deciphering CNA patterns

Our experimental and bioinformatic approaches complement existing approaches for testing hypotheses for the selection pressures underlying recurrent CNA patterns observed in human tumors. Other CNA-analysis approaches target different resolutions of the genome. Statistical algorithms such as GISTIC (Genomic Identification of Significant Targets in Cancer) have identified many strong individual driver genes and candidate regions (68). Integrating CNA

data with RNA knockdown screens and gene expression data has further identified driver genes missed by statistical analysis of CNA data alone (69). RNA knockdown-based analyses have also been used to support a more systems-level model in which the selection for amplification and deletion of particular DNA regions is based on the cumulative effects of many positive and negative fitness gains from multiple genes within a genomic region (70). Subsequent computational extensions have incorporated somatic mutation patterns to infer the cumulative impact of co-localized genes on fitness, and to successfully predict whole chromosome and chromosome-arm resolution level CNAs (24).

Our experimental and bioinformatic approach using phenotypic data, functional gene sets, and cross-species syntenic mapping has yielded additional insight into the selective pressures shaping tumor CNA patterns, namely coordinated alteration of genes involved in glycolytic metabolism. To our knowledge, our approach is the first reported experimental system in which CNA and associated phenotypes have been followed and repeatedly sampled as non-immortalized cells undergo spontaneous genomic instability and proceed from a diploid state to an immortalized aneuploid state. Using this approach, we identified genomic regions that are (i) associated with increased glycolysis and proliferation and enriched for genes from the core glycolysis pathway, and that are (ii) conserved in both mouse models of cancer and human tumors. Metabolic pathways are known to be coordinately regulated by modest changes in mRNA expression of functionally related genes (37,71). The coordinated alterations of metabolic genes at the DNA level adds an additional mechanism, namely conserved sets of CNA changes, by which glycolysis is dysregulated to promote tumorigenesis.

A strength of our PCA-based approach is the ability to unbiasedly reveal distinct CNA sub-signatures within a tumor type. Observed sub-signatures were found to be associated with previously known pathology- or profiling-defined tumor subtypes (eg, basal/luminal BRCA) (**Fig.**

2.1-2). In our experimental follow-up, the MEF system recapitulated a two signature pattern. Notably, the two signatures and their associated phenotype strengths were defined by the initiating loss of tumor suppression event, namely *Trp53* mutation versus *Mdm2* amplification. Thus, while the consequences of *TP53* mutation and amplification of *MDM2* are considered functionally similar and therefore mutually exclusive (72), our findings indicate they are not fully equivalent in terms of genomic instability and subsequent metabolic evolution. The tolerance of more highly disrupted and rearranged genomes upon p53 mutation appears to allow more flexibility in the evolution of aneuploid cancer genomes, thereby resulting in stronger glycolysis and enhanced proliferation. The specific combinations of CNA changes occurring in enzyme isoforms defining a metabolic pathway may be considered "onco-metabolic isoenzyme configurations" with differential potency, and the sets of combinations possible may be limited in part by the degree or specifics of tolerance to genomic re-arrangements.

Our work illustrates the value of cross-species comparisons in the analysis of DNA copy number data, much as recent cross-tissue and integrated genomic approaches have uncovered novel cancer subtypes and driver genes (73,74). The broken chromosome synteny between human and mouse genomes reduced the size of potential driver regions by over 4-fold on average, and identified a relatively small number (21) of amplification and deletion regions highly conserved between mouse and human. These conserved regions are consistently found across 15 of 17 tumor types examined, have sub-chromosomal scale with a median size of 10 megabase pairs, and include the glycolysis and pentose phosphate pathway enzyme CNAs observed in our experimental work.

7.4 Therapeutic and diagnostic implications

The most copy number aberrant tumors tend to have fewer point mutations in canonical oncogenes (eg, *KRAS*, **Fig. 2.1-2B – 2.1-2D**) and less canonical oncogene amplification (eg, *MYC*, **Fig. 2.1-2A**). Hence, genomic instability and subsequent coordinate alterations in multiple genes within a functional pathway may provide an alternate, more complex, pathway to acquisition of more aggressive tumor phenotypes – with tumor evolution and selection guiding the trajectory (75). Future models of the most aggressive cases of cancer will need to incorporate aspects of spontaneous genomic instability (mediated by distinct instability mechanisms) and resulting copy number alterations. Translationally, understanding how CNA patterns alter cancer genomes and impact cancer phenotypes will aid in the identification of metabolic or other 'hard-wired' vulnerabilities that can be therapeutically targeted. Furthermore, the relative stability of DNA samples, combined with the growing linkage between highly recurrent copy number changes and phenotypes, support the potential for molecular classification and diagnostic tests based on DNA copy number patterns (76).

7.5 Future directions

7.5.1 Impact of oxidative stress on the genomic CNA patterns recurrently observed in human cancers. To further test our model that genomic CNA patterns of a tumor are a reflection of its past tumorigenesis path (**Chapter 1.5**) we can genetically perturb genes involved in the cellular oxidative stress response regulators. This will determine whether the altered ROS selective forces can lead to altered resulting CNA patterns.

Among our prioritized targets would be Nrf2, a master antioxidant transcription factor that upregulates the expression of several antioxidant and detoxifying molecules (77) or inhibition of signaling through mTOR, the hypoxia target and a major regulator of multiple mechanisms contributing to the altered metabolic phenotype (77). After the selection period, cells with the perturbed ROS regulation will be microarray profiled and analyzed using our in-house downstream analysis pipeline.

7.5.2 CNA of energy metabolism genes provide a selective advantage during tumorigenesis by promoting glycolytic metabolism. Our CNA analysis implicates glycolysis genes as drivers for the recurrent chromosome region amplifications and deletions observed in human tumors (**Fig 7.1-1**). Guided by our bioinformatic analysis, we will continue to genetically manipulate our top candidate glycolysis genes and test their contribution to the observed CNA patterns. We will repeat the results with overexpression of metabolic enzymes enolase 2 and hexokinase 2. Hexokinase 2 is an enzyme that is considered to be responsible for trapping glucose in cells. Certain cancers are known to be specifically dependent on the hexokinase 2 (HK2) isoform of the hexokinase enzyme (78). Our other targets include phosphoglycerate mutate (PGM-M, also known as PGAM2) which has been shown to be capable of immortalizing mouse embryonic fibroblasts and enhancing their glycolytic flux (79). Similarly bisphosphoglycerate mutase (BPGM) also has a mutase and a phosphatase function like its glycolitic cousin, PGM. Additionally BPGM is located physically close to HK2 enzyme on chromosome 6 in the mouse genome. Overexpression of HK2 leads to the corresponding locus deletion in MEFs (**Fig 5.1-2 and 5.1-3)**. Hence BPGM would be an interesting target both because of its impact on glycolysis and as a control for the “hot-spot” for chromosomal aberration on chromosome 6. Another gene of interest is pyruvate dehydrogenase kinase (PDK) which acts to inactivate the enzyme pyruvate dehydrogenase. Therefore, overexpressing PDK

can prove effective in pushing cellular metabolism through the lactate arm promoting the Warburg effect in cells. We will also test whether combination of top 2-3 driver candidate genes can promote stronger CNA patterns.

7.5.3 Incorporation of genomic instability into a mouse model to further define the role of redox stress and energy metabolism genes in tumorigenesis using an in vivo system. *In vivo* mouse tumorigenesis models provide better understanding of the disease process and can serve as a pre-clinical system for testing therapeutic options against cancer. The lab of Sanaz Memarzadeh at UCLA has successfully generated tumors with pathology phenotypes matching type I endometrial adenocarcinomas (80) using the use tissue recombination and subrenal capsule grafting technique. This *in vivo* regeneration and transformation system can allow us to incorporate genomic instability at various stages before (e.g. implanting modified cells) or after (e.g. using tet-inducible gene regulation after implantation) introduction into mouse. This system results in tumors within 6-8 weeks providing us with much faster results compared to transgenic models. Cross-species conservation and enrichment of amplifications in glycolysis and glycolysis associated genes, has demonstrated high concordance between CNA results with the MEF immortalization system and CNA patterns observed in uterine carcinoma patients (**Chapter 4.1**). Therefore, genetical manipulation of metabolism and redox regulator genes prioritized in **Chapters 7.5.1** and **7.5.2** is expected to provide rich and useful *in vivo* mouse model for evaluation of resulting phenotypes.

7.5.4 Co-targeting DNA repair together with glycolysis and redox metabolism in genomically unstable tumors. Co-targeting DNA repair together with glycolysis and redox metabolism in genomically unstable tumors may be more effective than either inhibitor alone. We will test the susceptibility of genomically unstable and highly glycolytic breast cancer cell lines, mouse embryonic fibroblasts as well as mouse *in vivo* model to i) DNA-repair targeted

therapies, ii) anti-metabolism drugs, iii) reactive oxygen species protection or augmentation (altering homeostasis), and iv) combinations of the above therapeutic modalities. Metabolism in tumors can be targeted by different inhibitors. For example, both dichloroacetate (DCA) and metformin are targeting mitochondrial physiology, yet DCA decreases lactate production and is used to treat lactic acidosis, whereas metformin increases lactate production and lactic acidosis is an important side effect of metformin treatment (78,81). Inhibition of hexokinase 2 (HK2), which is generally expressed in skeletal muscle and adipose tissue, may provide a therapeutic window for treating cancers that are specifically dependent on the HK2. Other metabolic inhibitors to be tested include GLUT1 inhibitors (82) and PKM2 inhibitors (83). Targeting DNA-repair can be attained via inhibition of PARP in tumors with BRCA1/2 or PALB2 mutations (84). Different ways of neutralizing reactive oxygen species exist including the genetic manipulations of **Chapter 7.5.1**, such as targeting genes in the Nrf2 axis. We will test both a panel of breast cancer cell lines with published CNA data (86), cells from our mouse immortalization system, including the newly emerging sublines from **Chapter 7.5.1** and **7.5.2**, as well as the mouse model described in **Chapter 7.5.3** for the evidence of additive or synergistic effects of the suggested therapeutics.

CHAPTER 8

Methods

8.1 Cell culture and mouse strains

8.1.1 Cell culture: CD1 mouse embryonic fibroblasts, E14.5, were purchased from Stem Cell Technologies. $p53^{fl/fl}$ MEFs were obtained at day E14.5 from $p53^{fl/fl}$ (FVB.129P2-$Trp53^{tm1Brn}$/Nci) crossed with C57BL/6-129/SV mice. MEF cells were maintained in Dulbecco's Modified Eagle's Medium without pyruvate and supplemented with 10% FBS and 1% SPF. Cells were lifted and re-plated at a density of 9 x 10^5 viable cells / 60 mm dish every 3-4 days (ie, 3T9 protocol (43)) or in few cases at sub-confluent density (**Table 8.1-1**).

TC plate		Equivalent Cell No.		
Plate size	**Surface Area (cm^2)**	**3T9 Density**	**Sub-confluent Density**	**Media per plate (ml)**
150 mm	152	6.5E+06	1.4E+06	25
100 mm	55	2.4E+06	5.0E+05	10
60 mm	21	9.0E+05	1.9E+05	3.0
6-well plate	9	3.9E+05	8.2E+04	2.0
12-well plate	3.8	1.6E+05	3.5E+04	1.5
24-well plate	2	8.6E+04	1.8E+04	1.0
96-well plate	0.34	1.5E+04	3.1E+03	0.2

Table 8.1-1. Density of seeding cells chart. Density of seeded cells and media volume for mouse embryonic fibroblast culture according to TC, tissue culture, plate size.

Growth curves are presented by looking at the passage, **i**, versus the normalized cell number, $\mathbf{NCellNo_i}$ (**Fig. 3.1-1** and *Eq. 8.1.1-8.1.2*).

Eq. 8.1.1: $$NCellNo_i = FC_i \times NCellNo_{i-1},\ where$$

Eq. 8.1.2: $$FC_i = \frac{Total\ Number\ of\ Cells_{\ i}}{Cells\ Plated_{\ i-1}} \quad and \quad NCellNo_1 = 1$$

8.1.2 Genetic manipulation: Overexpression of HK1, HK2, kinase-dead HK2 (D209A/D657A) or ENO2 glycolysis enzymes and the MYC oncogene or the control protein RFP in CD1 MEFs was accomplished by transduction of non-immortalized cells with pDS-FB-neo retrovirus, followed by selection in 600 µg/ml G418. Deletion of p53 in $p53^{fl/fl}$ MEFs was induced by infection of non-immortalized cells with either retroviral Cre-GFP or Cre-ERT2 plus treatment with 1 µM 4-OHT.

8.1.3 ROS protection: To protect cells from reactive oxygen species (ROS), cells were cultured either at physiological oxygen concentrations (3% O_2), with 250 U/ml catalase from bovine liver (Sigma-Aldrich), or with both 3% O_2 and catalase. 3% oxygen conditions were achieved using a generic incubator and BioSpherix Proox Model 110 oxygen controller. Catalase was used as an enzymatic scavenger of reactive oxygen species that catalyzes the decomposition of hydrogen peroxide generated in the culture medium to water and oxygen.

8.2 CNA profiling

8.2.1 Background: To determine genome-wide copy number alteration we used Array Comparative Genomic Hybridization (aCGH). Copy number changes at a level of 5-10 kilobases can be detected by comparing the genomic DNA (gDNA) from a test sample to the gDNA from a reference sample. Fluorophores Cyanine 5 (Cy5) and Cyanine 3 (Cy3) are used to label test and reference samples respectively.

The Agilent SurePrint CGH Microarray, used in this study, is a glass slide with printed triplicates of 60-mer oligonucleotides that act as probes for the labeled gDNA. Fluoresently labled equal amounts of gDNA targets from test and reference samples are competitively cohybridized onto the oligonucleotide probes and scanned for the relative fluorescence intensities. The ratio Cy5 and Cy3 is proportional to the ratio of copy numbers of DNA sequences. This ratio of quantified intensities is subjected to background correction, normalization and segmentation algorithm to determine the copy number ratio for the whole genome.

8.2.2 Extract protocol: Genomic DNA was harvested using the DNeasy kit (Qiagen) following manufacturer's instructions. DNA from MEF lines and reference genomic DNA from C57BL/6J mouse tissue were hybridized to Agilent SurePrint G3 Mouse CGH 4x180k CGH microarray chips at the UCLA Pathology Clinical Microarray Core.

8.2.2 Label protocol: The Agilent SureTag DNA Labeling Kit was used to fragment, amplify, label, purify, and prepare the gDNA for hybridization. Only samples with greater than 1.6 260/280 ratio and with validated DNA integrity (by running a 1% agarose gel for 30 mins at 85V) were used. Restriction digestion using Alu I (10 µg/µl) and RSA I (10 µg/µl) and stabilizing acetylated BSA (10 µg/µl) was run on samples for 2 hrs at 37°C, 20 mins at 65°C to fragment isolated gDNA. 5 µl Random Primers were added to the 24 µl of the restriction digest and sample mix for 3 mins at 95°C, 10 mins on ice to anneal to the ssDNA fragments. To amplify the gDNA while generating labeled targets for array hybridization 1 µl of Exo-Klenow fragment of DNA polymerase, 5 µl 10XdNTPs, and 3 µl Cy5-dUTP (Cy3-dUTP) were added to the restriction digest and test (reference) sample mix and ran for 2 hrs at 37°C, 10 mins at 65°C. Unreacted dye and unincorporated nucleotides were removed from the labeled gDNA using

Amicon 30 kDa column purification. The degree of labeling and concentration was then determined, and the test and reference gDNA mixed at a 1:1 ratio.

8.2.3 Hybridization protocol: 55 µl 2X Hi-RPM hybridization buffer, 11 µl 10X aCGH blocking agent, 5 µl Cot-1 DNA (1.0 mg/ml) were added to the test and reference gDNA mix and ran 3 mins at 95°C, 30 mins at 37°C to promote binding of the gDNA targets to the probes and to prevent nonspecific binding. The mix was then allowed to hybridize to the microarray slide at 65°C for 24 hrs. After hybridization, slides were washed sequentially using Agilent Oligo aCGH/ChIP-on-Chip Washer Buffers.

8.2.4 Scan protocol: Microarrays were scanned using an Agilent Technologies Scanner G2505C. The raw data was extracted using Agilent Feature Extraction 11.5.1.1.

8.2.5 Data processing: Bioconductor analysis tools were used for data processing and quality control. Moving minimum background correction and print-tip loess normalization were performed in snapCGH package (87). More details on pre-processing aCGH data are specified in **Chapter 8.3**.

Moving minimum background correction was used to remove the effects of non-specific binding or spatial heterogeneity across the microarray. In this method, the background estimates are replaced with the minimums of the backgrounds of the spot and its eight neighbors, i.e., the background is replaced by a moving minimum of 3x3 grids of spots (**Fig. 8.2-1**).

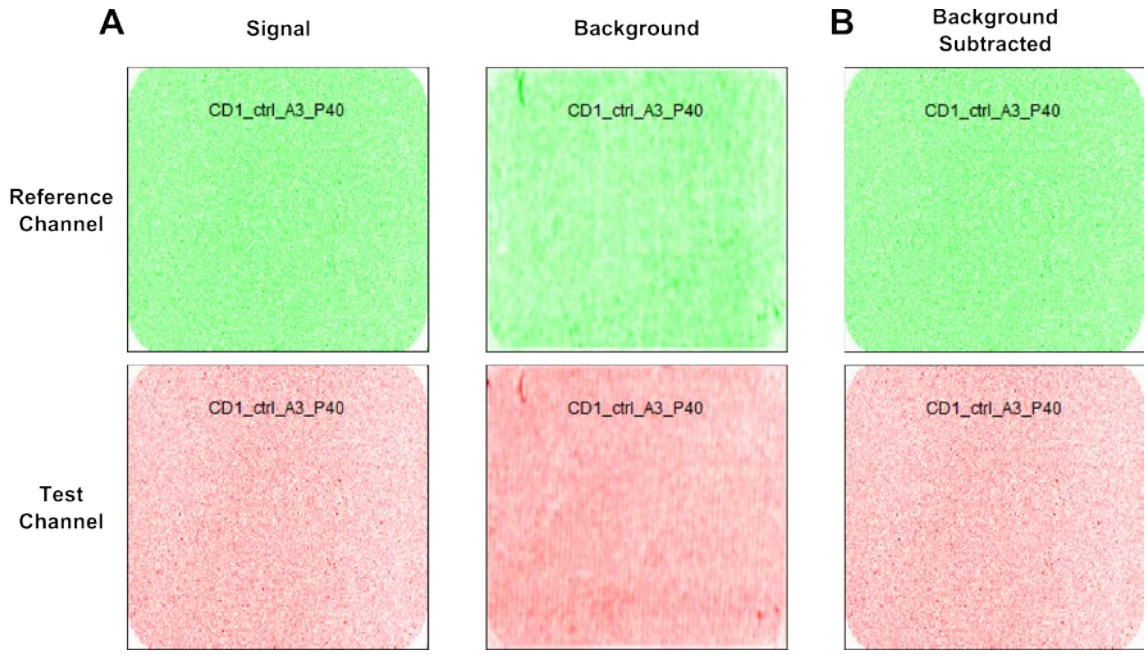


Figure 8.2-1. aCGH background correction. Moving minimum background subtraction in a representative MEF sample, CD1 wild-type A3 at passage 40. Reference or green channel, Cy3, is the reference genomic DNA from C57BL/6J mouse tissue. Test or red channel, Cy5, is the test MEF sample. (A) Foreground intensities from the reference and test channels along with the background intensities. (B) Intensities of test and reference samples after background correction.

Print-tip loess normalization was used to remove any systematic trends in copy number data which arise from the microarray technology rather than from differences between the probes or between the test DNA samples hybridized to the arrays. The copy number log-ratios were normalized separately to remove dye-bias within each array, so that the log-ratios averaged to zero within each array (within-array normalization). The distribution of the log red/green intensity ratios, M (***Eq. 8.2.1***), plotted by the average intensity, A (***Eq. 8.2.2***), is called MA-plot (**Fig. 8.2-2**) and is used to visualize raw microarray data to determine the need of normalization. A typical bias determined is when |M| is higher with higher A, that is the

Eq. 8.2.1: $$M = log_2\left(\frac{Intensity_{Red}}{Intensity_{Green}}\right)$$

Eq. 8.2.2: $$A = \frac{1}{2}log_2(Intensity_{Red} \times Intensity_{Green})$$

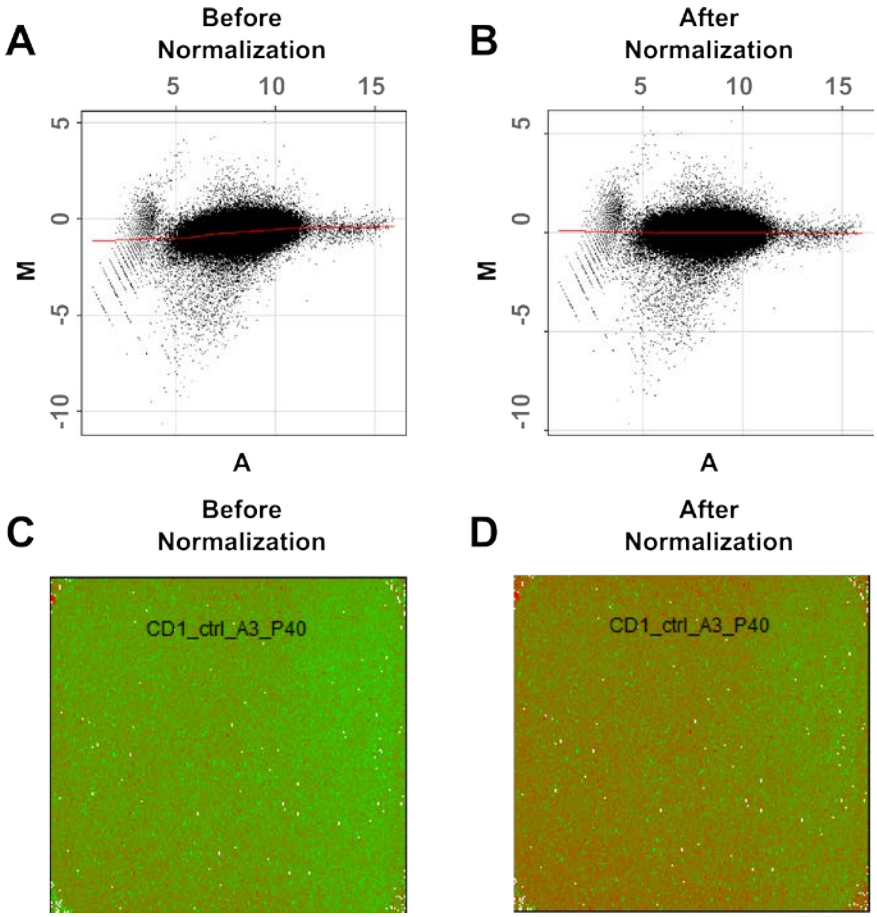


Figure 8.2-2. aCGH normalization. Print-tip loess normalization of a representative MEF sample, CD1 wild-type A3 at passage 40. Plots of log-intensity ratios (M-values) versus log-intensity averages (A-values). (A-B) A coplot of MA-plot with loess regression curve before (A) and after (B) normalization. The underlying assumption is that majority of genes are unchanged or there is symmetry between deletion and amplification. Therefore the majority of the points on the y-axis would be located at 0, since M = Log2(Intensity Red/Intensity Green) = Log2(1) = 0. If this is not the case, then intensity-dependent trends are modeled using regression loess curve, shown in red in panels A-B. (C-D) MA-plots before (C) and after (D) normalization.

dimmer the spot the less likely is an observed difference between the test and the reference samples.

Circular binary segmentation (with a minimum of 3 markers per segment and using permutation method) was performed on smoothed and log2-transformed copy-number profiles using DNAcopy package (88). Segmented data was converted into a matrix by genes for downstream analyses using *mus musculus 9* RefSeq reference genome from 2011.08.11 in CNTools (89). Copy-number profiles are presented using the Integrative Genomics Viewer (IGV) (90).

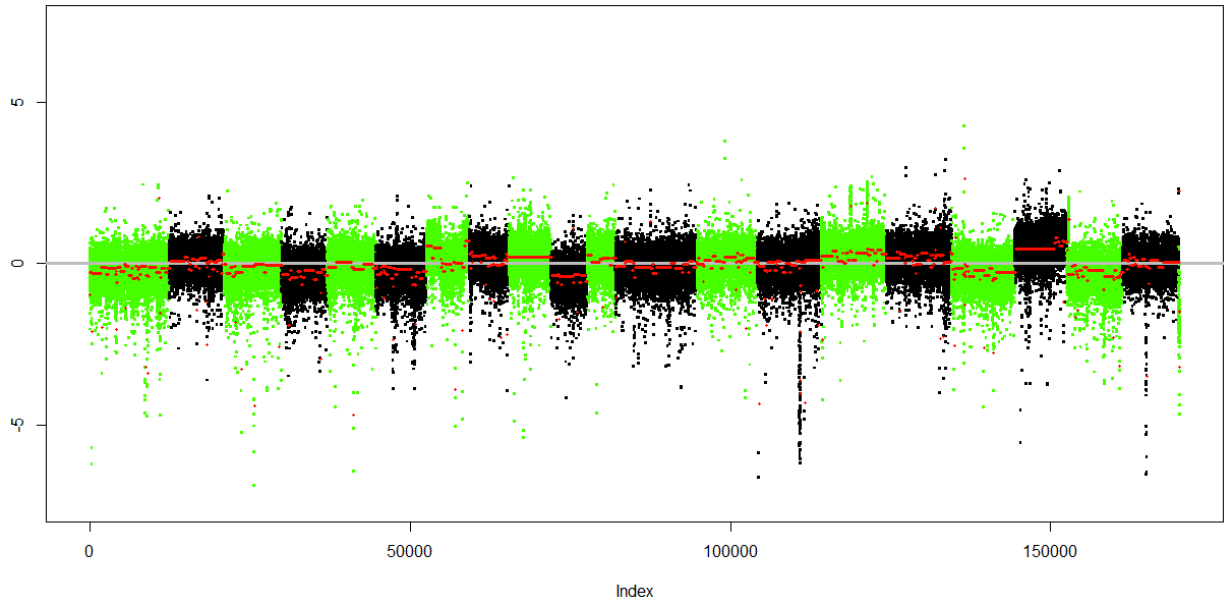


Figure 8.2-3. aCGH segmentation. Whole genome plot of raw probe log-ratios (y-axis). Data ordered by chromosome and map positions. Green – odd numbered chromosomes and chromosome Y; black – even numbered chromosomes and chromosome X. The red lines correspond to mean values in segment.

8.2.6 TCGA data: The Cancer Genome Atlas (TCGA) samples were downloaded from the TCGA portal in September 2012 (https://tcga-data.nci.nih.gov/tcga/). Copy-number profiles obtained were pre-processed level 3 data based on human genome 19, with copy number variations (CNVs) removed. TCGA tumor type abbreviations: bladder urothelial carcinoma

(BLCA), brain lower grade glioma (LGG), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), uterine corpus endometrial carcinoma (UCEC).

8.3 Pre-processing and quality assessment of two-color array CGH

8.3.1 Background: One of the most popular high-throughput tools used in the modern research is the array-based comparative genomic hybridization (aCGH) technology. This high-resolution assay is used to detect and compare the copy numbers of DNA sequences along the genome. It has been observed that copy number alterations are not only involved in phenotypic or gene expressional variation but also play important role during disease development, e.g. during tumorigenesis (32,91,92). Therefore with the constantly growing demand for the copy number state information and the constantly developing microarray technologies and their supporting software, it becomes extremely important to know how to utilize these developments correctly. In this chapter, our goal was to learn how to pre-process the two-color aCGH data and assess quality and validity of result using a freely available R package snapSGH: Segmentation, Normalization and Processing of aCGH Data (93).

The analysis of the two-color aCGH data generally involves three steps: 1) pre-processing and normalization of the raw intensity values; 2) aligning the data to the genomic locations and applying segmentation algorithm; 3) downstream analysis to explore biological

meaning of the differential copy number state. There exist many commercial and freely available software that provide tools for one or all of the analysis steps (94). However, the implementation of the existing algorithms is complicated because many of them are expanded from analysis performed on expression data. The main differences between expression data and copy number data analysis is that analysis of the later should account for the fact that aCGH hybridization signals are typically smaller and with higher background due to the excess of non-target labeling products. Also, it is important to remember that small changes in copy number may mark biologically significant events, such as single-copy gains or losses that may accompany unbalanced translocations or chromosomal rearrangements. Lastly, log2 ratios are expected to be highly correlated between probes targeting adjacent regions in the genome.

Acknowledging described above complications we tested the snapCGH algorithm on two publicly available human lung squamous cell carcinoma (LUSC) samples and compared processed results to the published results obtained from three different platforms. Upon confirming general comparability to the published results we established a pipeline for the in-house mouse cell lines.

8.3.2 Methods: *Data Sets.* The raw Level 1 data and processed Level 3 data for two patient samples were downloaded from The Cancer Genome Atlas (TCGA) (95). The samples were analyzed using Agilent SurePrint G3 Human CGH Microarray Kit 1x1M and processed using standard Agilent feature extraction algorithm followed by the in-house algorithm (96). The data then was segmented using circular binary segmentation algorithm (CBS) (97). The processed Level 3 data from Affymetrix platform was analyzed using Genome-Wide SNP6 platform and processed using genepattern pipeline (98). Finally the third processed Level 3 data was analyzed using Agilent HG-CGH-415K_G4124A platform and processed using Loess

Normalization and CBS segmentation. The reason for the chosen small sample size for the control test set was memory limitation while using snapCGH R package.

The 12 in-house mouse embryonic fibroblasts (MEFs) were analyzed using Agilent SurePrint G3 Mouse CGH Microarray Kit.

Background Correction Method. The method for background correction was chosen empirically from the data sets. Even though it has been shown that convolution-based models perform the best for expression data (99), these methods tend to weaken the aCGH signal at low intencities. Since the traditional background subtraction method is also not recommended due to the produced negative intensities, which in log-space become "missing values", a moving minimum background correction method was chosen. In this method the background estimates are replaced with the minimums of the backgrounds of the spot and its eight neighbors, i.e., the background is replaced by a moving minimum of 3x3 grids of spots.

Normalization Method. Normalization is intended to remove systematic trends that arise from the microarray technology rather than from the actual copy number variations in samples. Since different normalization methods performed equally well for the tested data sets, the final normalization method was selected based on the literature review (100) – print-tip loess normalization. In the absence of print-tip groups on Agilent arrays, this method performed similar to the global loess normalization method.

Processing Data. The last step in pre-processing the microarray data is probe summarization in which duplicates are removed following the averaging leaving only one occurrence of each clone set.

Segmentation. After the pre-processing step a segmentation algorithm is applied to aCGH data, which is used to translate noisy intensity measurements into regions of equal copy number. The

circular binary segmentation (CBS) algorithm was chosen to perform analysis as it has been shown that this algorithm has smallest FDR (94). In addition, the two publicly available processed data sets, that we used as control sets, also applied CBS segmentation algorithm.

8.3.3 Results: *Background Correction.* Different background correction to the raw intensities methods were applied to two human LUSC samples (**Fig. 8.3-1A – 8.3-1D**) that were used as a control in this study. The intensity density plots reveal that background subtraction methods which were designed to produce positive corrected intensities performed better than convolution-based methods for aCGH data (**Fig. 8.3-1C – 8.3-1D**). **Fig. 8.3-1E,F** assess whether the chosen moving minimum background correction method fits the mouse embryonic fibroblasts test set.

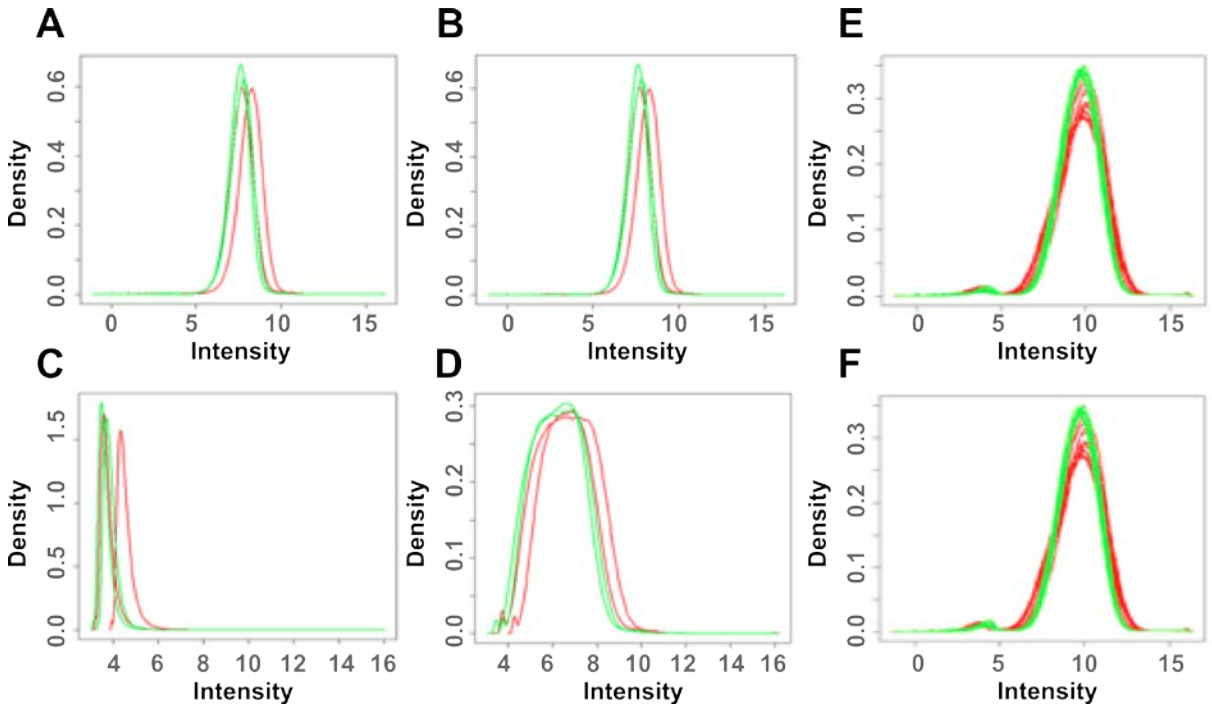


Figure 8.3-1. Comparison of background correction methods. Selected density plots comparing individual-channel intensities for two-color microarray data before and after background correction. Human LUSC samples, (A-D). Mouse embryonic fibroblasts, (E-F). x-axis, intensity; y-axis, density. (A) No background correction applied. (B) Moving minimum background correction method. (C) Convolution-based NormExp RMA method. (D) Convolution-based NormExp saddle method. (E) No background correction applied. (F) Moving minimum background correction method.

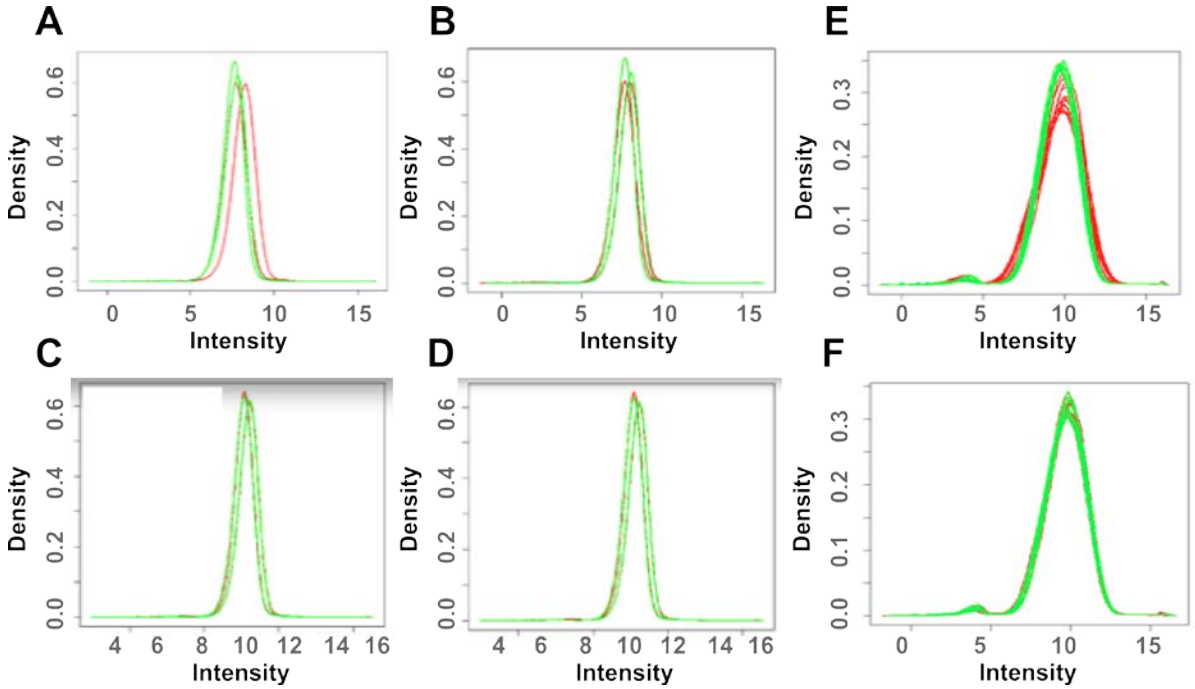


Figure 8.3-2. Comparison of normalization methods. Density plots comparing individual-channel intensities for two-color microarray data before and after normalization. Human LUSC samples, (A-D). Mouse embryonic fibroblasts, (E-F). x-axis, intensity; y-axis, density. (A) No normalization applied. (B) Median normalization method applied. (C) Print tip loess normalization method. (D) Global loess normalization method. (E) No normalization applied. (F) Print tip loess normalization method.

Normalization. Three normalization methods were tested on the human patient samples that are used as a control in this study. Median normalization method involves subtraction of the weighted median from the M-values (log ratios) for each array. The idea behind print-tip loess normalization is that from each M-value the corresponding value of each tip group curve is subtracted. A simpler form of loess normalization is global loess normalization in which the global loess curve is subtracted. The density plot of normalized intensities (**Fig. 8.3-2A – 8.3-2D**) reveals that loess normalizations has slightly better performance for the human patient samples. The same trend is observed in print tip loess normalized mouse aCGH samples (**Fig. 8.3-2E – 8.3-2F**). Normalization is intended to remove systematic errors, therefore looking at the log ratio, M (***Eq. 8.2.1***), versus log sum, A (***Eq. 8.2.2***), plots of the normalized intensities,

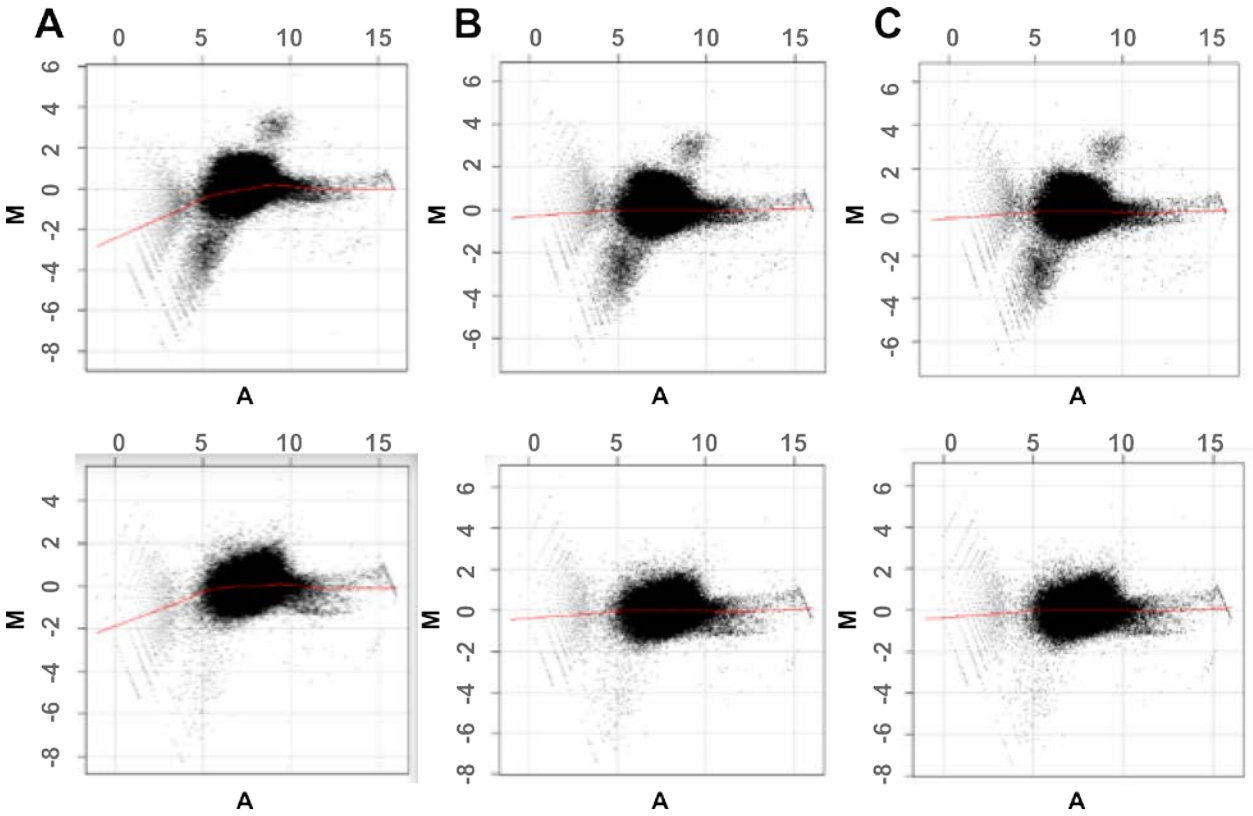


Figure 8.3-3. Log ratio vs log sum of normalized intensities in control samples. y-axis, log ratio M; x-axis, log sum A. Human LUSC samples: top row sample "1070", bottom row sample "1071". (A) Median method normalization applied. (B) Print tip loess normalization method. (C) Global loess normalization method.

MA-plots, can aid in determining which normalization method to apply. MA-plots in **Fig. 8.3-3** one more time show that loess methods are performing slightly better for the given data set. **Fig. 8.3-4** is the confirmation that loess normalization works well on the mouse embryonic fibroblasts set.

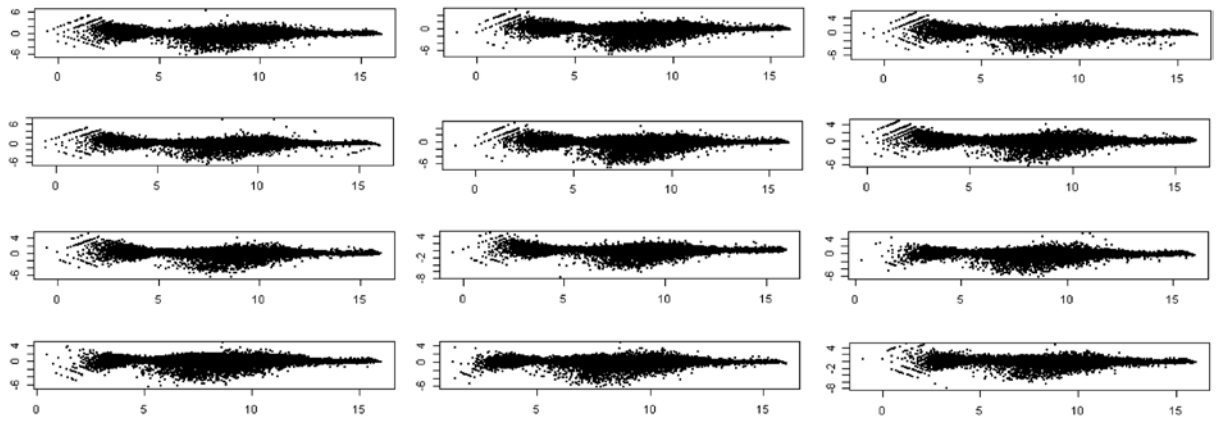


Figure 8.3-4. Log ratio vs log sum of normalized intensities in MEFs. y-axis, log ratio M; x-axis, log sum A. Twelve mouse embryonic fibroblasts samples.

Segmentation. Plotting the segmented data into genomic viewer (**Fig. 8.3-5**) reveals that pre-processing and segmentation of the two-color data using snapCGH methods gives slightly noisier result then were obtained in the published processed data (Affy SNP6, Agilent CGH415, Agilent Sureprint). Nevertheless, all the copy number amplification and deletion trends

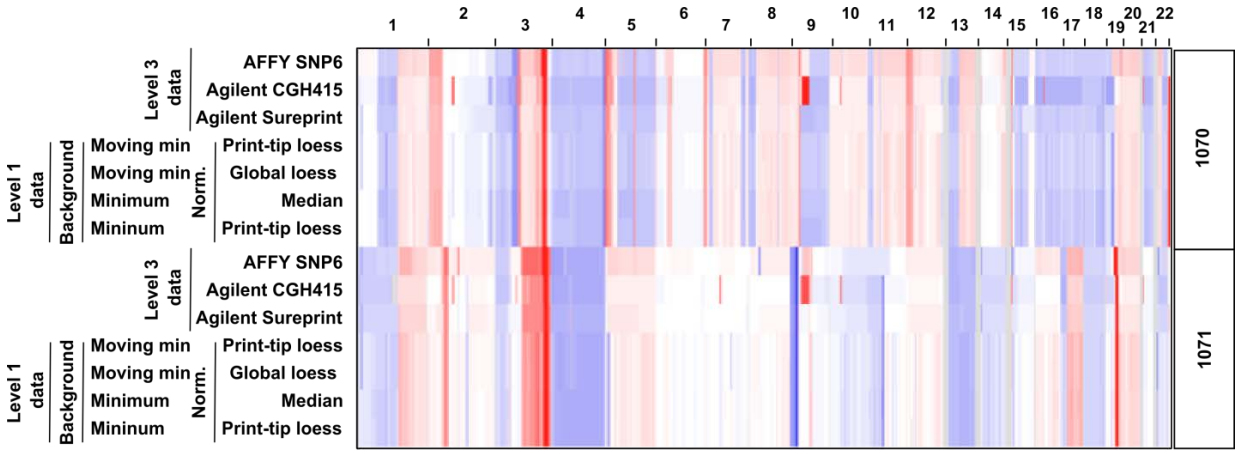


Figure 8.3-5. IGV view of CNA in control samples. Integretive Genomic view of the chromosomal aberrations in two control human patient samples: “1070” and “1071”. Red stands for genomic regions that are amplified; blue stands for genomic regions that are deleted. Columns represent chromosomes while each row is different preprocessing and segmentation method for the two samples. TCGA Level1 (raw) and Level 3 (processed) data are shown. In this figure, Level 1 data were subject to ‘moving minimum’ and ‘minimum’ background correction methods from snapCGH R package and ‘Print-tip loess’, ‘Global loess’, and ‘Median’ normalization (Norm.) methods.

are very consistent between snapCGH and published processed samples. Moreover, different background correction and normalization methods used for pre-processing the data in snapCGH gave very consistent results between each other. Finally, **Fig. 8.3-6** shows the segmentation patterns of the mouse embryonic fibroblasts data which was pre-processed by moving minimum background subtraction and print tip loess normalization.

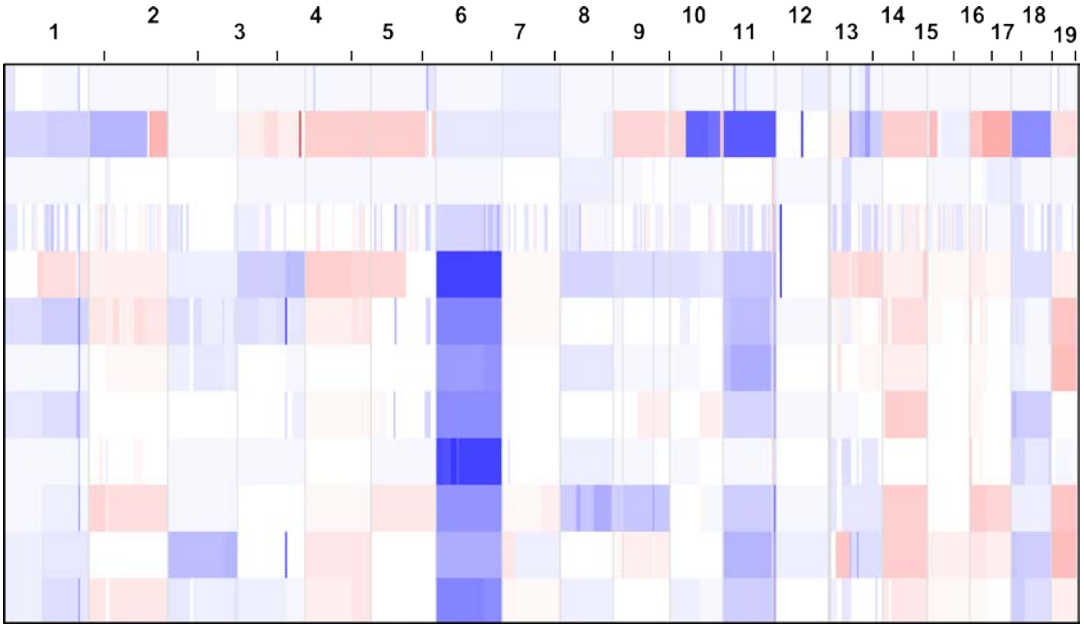


Figure 8.3-6. IGV view of CNA in MEF samples. Integretive Genomic view of the chromosomal aberrations in twelve mouse embryonic fibroblast samples. Red stands for genomic regions that are amplified; blue stands for genomic regions that are deleted. Columns represent chromosomes while each row represents a different sample.

8.3.4 Discussion:

As was discussed in **chapter 8.3.1**, the array comparative hybridization experiments are among most popular in modern research. There exist a number of specialized algorithms, which accomplish the pre-processing and segmentation steps for detecting truly DNA amplified or deleted regions. One of such algorithms is the bioconductor R package snapCGH that can be used to pre-process the aCGH data and obtain good quality signals.

Overall, based on the analysis in **Chapter 8.3,** snapCGH perform well on pre-processing the data. The observed slight variations between samples processed in snapCGH and TCGA Level 3 samples (**Fig. 8.3-5**) appear to be due to differential segmentation parameters (data not shown) used by different processing centers rather than by pre-processing the data. A fairly consistent segmentation results (**Fig. 8.3-5**) is observed when different normalization and background correction techniques are applied to the two control human LUSC samples.

8.4 Bioinformatic and statistical analysis

8.4.1 PCA: Principal component analysis (PCA) was performed using the mean-centered matrix of CNA values per gene locus. Genes with identical profiles across samples were collapsed to a single representative gene.

8.4.2 Enrichment of mutations: To test for enrichment of mutations within CNA-defined PC scores, tumors were queried for mutations in the most frequently mutated genes using the cBioPortal for Cancer Genomics (101). Tumors with known mutational status were then sorted by their PC1 score, and we calculated a Kolmogorov-Smirnov statistic against the expected distribution of mutations. The statistical significance of enrichment was determined by permutation analysis.

8.4.3 GSEA and WGV: Metabolic pathway enrichment analysis (gene set enrichment analysis, GSEA (102)) and weighted gene voting (WGV) prediction analysis (38) were performed using 74 metabolic pathways defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (40), using pathways with 7 or more measured genes. For KEGG-based metabolic pathway enrichment analysis, we collapsed metabolic isoenzyme loci (genes with the same enzyme activity (Enzyme Commission [EC] numbers)) that were within 100 kilobases from each other into a single representative locus.

8.4.4 Consistency analysis: Consistency signatures of conserved amplification and deletions (consistent regions) were defined as non-zero when all CNA summary signatures have the same sign across all tumor types, and is derived from the absolute value-based minimum summary metric, then re-signed positive for amplification or negative for deletion. In GSEA using genome consistency signatures, enrichment scores were calculated through the ranked set of consistently amplified genes since after this point the many genes that are not consistent across signatures have a consistency value of zero and accordingly have tied ranks. For WGV predictions of metabolic phenotypes, t-scores were used as gene weights. To calculate p-values, we calculated the fraction of 1,000 randomly-chosen gene sets of equal size that gave better average rankings than the true gene set. False discovery rate (FDR) q-values were calculated using the Benjamini-Hochberg procedure. CNA and metabolite changes were visualized in the context of metabolic pathway structure using Cytoscape (42).

8.4.5 Senescence score: A summary senescence score for each MEF subline was calculated by subtracting the area under the MEF's growth curve, **Z(x)**, from the area under an ideal growth curve, **Y(x)**, i.e. consistent growth at the fastest observed rate. The area difference was then averaged by dividing by the passage number, **x**, and log2-transformed, resulting in a normally distributed score (**Fig. 8.4-1** and ***Eq. 8.4.1***).

Eq. 8.4.1: $$Senescence\ Score = log_2 \frac{\sum_{x=2}(Y_{(x)} + Y_{(x-1)}) - \sum_{x=2}(Z_{(x)} + Z_{(x-1)})}{2x}$$

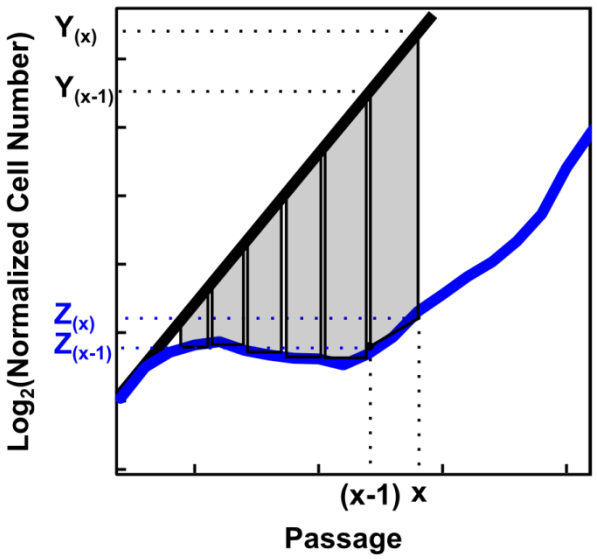


Figure 8.4-1. Senescence score calculation. Blue curve, Z(x), representative growth of MEF sample with initial expansion followed by decrease due to senescence stage and culminated by expansion due to immortalization of sub-population of cells. Black curve, Y(x), 'ideal growth curve' of cells that grow at a constant rate without entering the senescence. The shaded shapes used for calculating area between curves using Riemann sum are shown.

8.4.6 Genomic instability scores: A genomic instability score 'breakpoints per chromosome' was calculated by counting the number of segments in each sample, **# segments**, and dividing by the number of chromosomes, **# chr** (***Eq. 8.4.2***). The segments were inferred by the Circular Binary Segmentation algorithm (**Chapter 8.2.5**) and chromosomes X and Y were excluded from the analysis.

Eq. 8.4.2: $$Breakpoints\ per\ chromosome = \frac{\#\ segments - \#\ chr}{\#\ chr}$$

A genomic instability score termed 'integrated CNA' was calculated by integration of the Circular Binary Segmentation algorithm-inferred absolute mean copy-number of segments, **|segment**

mean|, multiplied by the length of each segment, **|segment end – segment start|**, and normalized by the total number of base pairs in a sample, **# base pairs in sample** (***Eq. 8.4.3***). The integrated value was normalized to the average integrated CNA of diploid samples (***Eq. 8.4.4***) and log2-transformed resulting in a normally distributed score.

Eq. 8.4.3:

$$Int.CNA_{sample} = \log_2 \frac{\sum_{segments} |segment\ end - segment\ start| \times |segment\ mean|}{\#\ base\ pairs\ in\ sample}$$

Eq. 8.4.4:

$$Integrated\ CNA_{sample} = Int.CNA_{sample} - \frac{1}{N} \sum_{\substack{diploid \\ sample}}^{N} Int.CNA_{sample},$$

where N = number of diploid samples.

8.4.7 Box and whisker plots: In box and whisker plots, the box represents the median, as well as the first and third quartiles, and the whisker indicates the extreme values within 1.5 times the inter-quartile range. In cases where the number of samples permitted, individual values are superimposed as jitter plots.

8.5 Glucose consumption and lactate secretion measurements

8.5.1 BRCA cellines: Lactate secretion rates of breast cancer cell lines were measured from the culture media using a colorimetric assay kit (BioVision).

8.5.2 MEFs: Glucose consumption and lactate secretion rates of MEFs were measured using a BioProfile Basic bioanalyzer (NOVA Biomedical). Data were normalized to the integrated

cell number, which was calculated based on cell counts at the start and end of the time course and an exponential growth equation (***Eq. 8.5.1 - 8.5.3***).

Eq. 8.5.1: $$\int_0^t N = \int_0^t N_o e^{\kappa t}, where$$

$$N = number\ of\ cells\ at\ time\ t,$$

$$N_o = number\ of\ cells\ at\ t = 0,$$

$$and\ \kappa = \frac{1}{t} ln \frac{N}{N_o} \quad is\ a\ rate\ constant.$$

Eq. 8.5.2: $$Nt = N_o \left[\frac{1}{\kappa} e^{\kappa t}\right] \frac{t}{0} = N_o \left(\frac{1}{\kappa} e^{\kappa t} - \frac{1}{\kappa} e^{0}\right) = \frac{N_o}{\kappa} (e^{\kappa t} - 1)$$

Eq. 8.5.3: $$N = \frac{N_o}{\kappa t} (e^{\kappa t} - 1)$$

Because the proliferation rates of MEF sublines vary, each cell line was seeded at the appropriate density so as to give an integrated cell number of approximately 6.5 x 10^5 cells in a 6 well plate. All samples were run as biological triplicates, and consistent results were seen in multiple independent experiments.

8.6 Mass spectrometry-based metabolomic analyses

8.6.1 Cell culture: MEF sublines were seeded onto 6-well plates, and after 24 hours media was replaced with media containing 4.5 g/L [1,2-^{13}C]-labeled glucose. Sample collection occurred after 24 hours of culture in the labeled glucose media. For cell culture media metabolite analysis (footprint profiling), 20 µl of cell-free media samples were collected.

8.6.2 Metabolite extraction: Metabolites were extracted by adding 300 µl ice-cold 80% methanol, followed by vortexing three times over 15 min, and centrifugation for 10 min at 13k

rpm at 4 °C. The supernatant was transferred to a fresh tube, and the solvent was reduced using a vacuum evaporator. For intracellular metabolite analysis, cells were washed with ice-cold 150 mM ammonium acetate (NH_4AcO) pH 7.3, and metabolites extracted in 1 ml ice-cold 80% MeOH. The cells were quickly transferred into a microfuge tube, and 10 nmol norvaline was added to the cell suspension for use as an internal standard. The suspension was subsequently vortexed three times over 15 min and then spun down at 4 °C for 5 min. The supernatant was transferred into a glass vial, the cell pellet re-extracted with 200 µl ice-cold 80% MeOH, spun down and the supernatants combined. Metabolites were dried at 30°C under vacuum and resuspended in 50 µl of 70% acetonitrile (ACN). 5 µl were used for mass spectrometer-based analysis.

8.6.3 Running protocol: Samples were run on a Q-Exactive mass spectrometer coupled to an UltiMate 3000RSLC UHPLC system (Thermo Scientific). The mass spectrometer was run in polarity switching mode (+3.00 kV / -2.25 kV) with an m/z window ranging from 65 to 975. Mobile phase A was 5 mM NH4AcO, pH 9.9 and mobile phase B was ACN. Metabolites were separated on a Luna 3 µm NH2 100 Å (150 x 2.0 mm) (Phenomenex) column. The flow was kept at 200 µl / min, and the gradient was from 15% A to 95% A in 18 min, followed by an isocratic step for 9 min and re-equilibration for 7 min.

8.6.4 Data processing: Metabolites were detected and quantified as area under the curve (AUC) based on retention time and accurate mass (≤ 3ppm) using the TraceFinder 3.1 (Thermo Scientific) software. Relative amounts of metabolites between various conditions, percentage of metabolite isotopomers (relative to all isotopomers of that metabolite), and percentage of labeled metabolite molecules (isotopomer M1 and greater, relative to all isotopomers) were calculated and corrected for naturally occurring ^{13}C abundance (103). Footprinting data were normalized to the integrated cell number as described above, and

intracellular metabolite concentrations were normalized to the number of cells present at the time of extraction. All samples were run as biological triplicates, and consistent results were seen in multiple independent experiments. Our analysis focused on metabolite level, percent isotopomer, and percent labeled-metabolite measurements with ANOVA p-values across the sample panel of less than 0.05 in individual experiments, and Pearson correlation coefficients across all samples and between independent experiments of greater than 0.5.

8.7 Patient tumor samples and quantitative FDG-PET imaging

Ten patient breast cancer samples with imaged FDG-uptake within 4 weeks prior to surgery, excluding patients with secondary breast cancers and recurrent disease, were collected surgically and processed as previously described (37). None of the patients received systemic therapy or radiation prior to imaging. ^{18}FDG-tumor uptake was quantified as standardized uptake values (SUV) and showed the expected wide dynamic range (3.8-18.5). There was no significant difference in patient age, tumor volume, and lymph node involvement between the groups of FDG-high and FDG-low breast cancers. Breast cancers with high ^{18}FDG-PET SUVs frequently lacked expression of the estrogen receptor (ER) and the progesterone receptor (PR), but hormone receptor-negative tumors were also represented amongst the tumors with the lowest FDG uptake(37). We excluded lobular breast carcinomas, because they have been shown to take up less FDG than ductal carcinomas (104) We excluded large breast carcinomas (>5-cm) and breast carcinomas with multifocal FDG uptake because our protocol did not include tissue autoradiography to direct the molecular tissue analysis to areas of distinct radiotracer retention. This study was approved by the Institutional Review Board (IRB) of Memorial Sloan-Kettering Cancer Center, and all participating patients signed the informed consent.

8.8 Immunoblot Analysis

Cells were lysed in modified RIPA buffer (50 mM Tris-HCl (pH 7.5), 150 NaCl, 10 mM β-glycerophosphate, 1% NP-40, 0.25% sodium deoxycholate, 10 mM sodium pyrophosphate, 30 mM sodium fluoride, 1 mM EDTA, 1 mM vanadate, 20 µg/ml aprotinin, 20 µg/ml leupeptin, and 1 mM phenylmethylsulfonyl fluoride). Whole cell lysates were resolved by SDS-PAGE on 4-15% gradient gels and blotted onto nitrocellulose membranes (Bio-Rad). Membranes were blocked overnight and then incubated sequentially with primary and either HRP-conjugated (Pierce) or IRDye-conjugated secondary antibodies (Li-Cor). Blots were imaged using the Odyssey Infrared Imaging System (Li-Cor). Protein levels were quantitated using ImageJ (http://imagej.nih.gov/ij/). Primary antibodies used for western blot analysis included Hexokinase 1 (2024, Cell Signaling Technology), Hexokinase 2 (2867, Cell Signaling Technology), p53 (NB200-103, Novus Biologicals), and Enolase 2 (8171, Cell Signaling Technology).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015 Jan 1;65(1):5–29.

2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646–74.

3. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell. 2000 Jan 7;100(1):57–70.

4. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. Cell Metab. 2008 Jan;7(1):11–20.

5. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011 Feb;11(2):85–95.

6. Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, et al. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science. 2012;336(6084):1040–4.

7. Maddocks ODK, Berkers CR, Mason SM, Zheng L, Blyth K, Gottlieb E, et al. Serine starvation induces stress and p53-dependent metabolic remodelling in cancer cells. Nature. 2013 Jan 24;493(7433):542–6.

8. Patra KC, Wang Q, Bhaskar PT, Miller L, Wang Z, Wheaton W, et al. Hexokinase 2 Is Required for Tumor Initiation and Maintenance and Its Systemic Deletion Is Therapeutic in Mouse Models of Cancer. Cancer Cell. 2013 Aug 12;24(2):213–28.

9. Xie H, Hanai J, Ren J-G, Kats L, Burgess K, Bhargava P, et al. Targeting Lactate Dehydrogenase-A Inhibits Tumorigenesis and Tumor Progression in Mouse Models of Lung Cancer and Impacts Tumor-Initiating Cells. Cell Metab. 2014 May 6;19(5):795–809.

10. Li B, Qiu B, Lee DSM, Walton ZE, Ochocki JD, Mathew LK, et al. Fructose-1,6-bisphosphatase opposes renal carcinoma progression. Nature. 2014 Sep 11;513(7517):251–5.

11. Kim J, Gao P, Liu Y-C, Semenza GL, Dang CV. Hypoxia-Inducible Factor 1 and Dysregulated c-Myc Cooperatively Induce Vascular Endothelial Growth Factor and Metabolic Switches Hexokinase 2 and Pyruvate Dehydrogenase Kinase 1. Mol Cell Biol. 2007 Nov 1;27(21):7381–93.

12. Wang L, Xiong H, Wu F, Zhang Y, Wang J, Zhao L, et al. Hexokinase 2-Mediated Warburg Effect Is Required for PTEN- and p53-Deficiency-Driven Prostate Cancer Growth. Cell Rep. 2014 Nov 9;8(5):1461–74.

13. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature. 2011 Aug 18;476(7360):346–50.

14. Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nat Genet. 2011 Sep;43(9):869–74.

15. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009 Dec 10;462(7274):739–44.

16. Kondoh H, Lleonart ME, Gil J, Wang J, Degan P, Peters G, et al. Glycolytic Enzymes Can Modulate Cellular Life Span. Cancer Res. 2005 Jan 1;65(1):177–85.

17. Kondoh H. The Role of Glycolysis in Cellular Immortalization. In: Sarangarajan R, Apte S, editors. Cellular Respiration and Carcinogenesis [Internet]. Humana Press; 2009 [cited 2014 Oct 21]. p. 91–102. Available from: http://link.springer.com/chapter/10.1007/978-1-59745-435-3_7

18. Kaplon J, Zheng L, Meissl K, Chaneton B, Selivanov VA, Mackay G, et al. A key role for mitochondrial gatekeeper pyruvate dehydrogenase in oncogene-induced senescence. Nature. 2013 Jun 6;498(7452):109–12.

19. Yun J, Rago C, Cheong I, Pagliarini R, Angenendt P, Rajagopalan H, et al. Glucose Deprivation Contributes to the Development of KRAS Pathway Mutations in Tumor Cells. Science. 2009 Sep 18;325(5947):1555–9.

20. Zhang WC, Shyh-Chang N, Yang H, Rai A, Umashankar S, Ma S, et al. Glycine Decarboxylase Activity Drives Non-Small Cell Lung Cancer Tumor-Initiating Cells and Tumorigenesis. Cell. 2012 Jan 20;148(1–2):259–72.

21. Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt stimulates aerobic glycolysis in cancer cells. Cancer Res. 2004 Jun 1;64(11):3892–9.

22. Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? Nat Rev Cancer. 2004 Nov;4(11):891–9.

23. Stuart D, Sellers WR. Linking somatic genetic alterations in cancer to therapeutics. Curr Opin Cell Biol. 2009 Apr;21(2):304–10.

24. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, et al. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Cell. 2013 Jul 11;155(4):948–62.

25. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010 Feb 18;463(7283):899–905.

26. Muller FL, Colla S, Aquilanti E, Manzo VE, Genovese G, Lee J, et al. Passenger deletions generate therapeutic vulnerabilities in cancer. Nature. 2012 Aug 16;488(7411):337–42.

27. Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. Genetic instability and darwinian selection in tumours. Trends Cell Biol. 1999 Dec 1;9(12):M57–60.

28. Sheltzer JM. A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis. Cancer Res. 2013 Nov 1;73(21):6401–12.

29. Palaskas N, Larson SM, Schultz N, Komisopoulou E, Wong J, Rohle D, et al. 18F-fluorodeoxy-glucose positron emission tomography marks MYC-overexpressing human basal-like breast cancers. Cancer Res. 2011 Aug 1;71(15):5164–74.

30. Czernin J, Phelps ME. POSITRON EMISSION TOMOGRAPHY SCANNING: Current and Future Applications. Annu Rev Med. 2002;53(1):89–112.

31. Jadvar H, Alavi A, Gambhir SS. 18F-FDG Uptake in Lung, Breast, and Colon Cancers: Molecular Biology Correlates and Disease Characterization. J Nucl Med. 2009 Nov 1;50(11):1820–7.

32. Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR. Genomic instability in breast cancer: pathogenesis and clinical implications. Mol Oncol. 2010 Jun;4(3):255–66.

33. TCGA Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011 Jun 30;474(7353):609–15.

34. Jadvar H, Alavi A, Gambhir SS. 18F-FDG Uptake in Lung, Breast, and Colon Cancers: Molecular Biology Correlates and Disease Characterization. J Nucl Med. 2009 Nov 1;50(11):1820–7.

35. Nakamura K, Hongo A, Kodama J, Hiramatsu Y. The measurement of SUVmax of the primary tumor is predictive of prognosis for patients with endometrial cancer. Gynecol Oncol. 2011 Oct;123(1):82–7.

36. Nam EJ, Yun MJ, Oh YT, Kim JW, Kim JH, Kim S, et al. Diagnosis and staging of primary ovarian cancer: Correlation between PET/CT, Doppler US, and CT or MRI. Gynecol Oncol. 2010 Mar;116(3):389–94.

37. Palaskas N, Larson SM, Schultz N, Komisopoulou E, Wong J, Rohle D, et al. 18F-Fluorodeoxy-glucose Positron Emission Tomography Marks MYC-Overexpressing Human Basal-Like Breast Cancers. Cancer Res. 2011 Aug 1;71(15):5164–74.

38. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science. 1999 Oct 15;286(5439):531–7.

39. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006 Dec;10(6):515–27.

40. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014 Jan 1;42(D1):D199–205.

41. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1–pl1.

42. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011 Feb 1;27(3):431–2.

43. Todaro GJ, Green H. Quantitative Studies of the Growth of Mouse Embryo Cells in Culture and Their Development into Established Lines. J Cell Biol. 1963 May 1;17(2):299–313.

44. Gupta S, Ramjaun AR, Haiko P, Wang Y, Warne PH, Nicke B, et al. Binding of Ras to Phosphoinositide 3-Kinase p110α Is Required for Ras- Driven Tumorigenesis in Mice. Cell. 2007 Jan 6;129(5):957–68.

45. Lowe SW, Ruley HE, Jacks T, Housman DE. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. Cell. 1993 Sep 24;74(6):957–67.

46. Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, et al. Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. Cancer Cell. 2007 Jan;11(1):9–23.

47. Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW. Aneuploidy Acts Both Oncogenically and as a Tumor Suppressor. Cancer Cell. 2007 Jan;11(1):25–36.

48. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, et al. A microRNA component of the p53 tumour suppressor network. Nature. 2007 Jun 28;447(7148):1130–4.

49. Sun H, Gulbagci NT, Taneja R. Analysis of growth properties and cell cycle regulation using mouse embryonic fibroblast cells. Methods Mol Biol Clifton NJ. 2007;383:311–9.

50. Sherr CJ, DePinho RA. Cellular Senescence: Minireview Mitotic Clock or Culture Shock? Cell. 2000 Aug 18;102(4):407–10.

51. Parrinello S, Samper E, Krtolica A, Goldstein J, Melov S, Campisi J. Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. Nat Cell Biol. 2003 Aug;5(8):741–7.

52. Halliwell B. Oxidative stress in cell culture: an under-appreciated problem? FEBS Lett. 2003 Apr 10;540(1–3):3–6.

53. Graham NA, Tahmasian M, Kohli B, Komisopoulou E, Zhu M, Vivanco I, et al. Glucose deprivation activates a metabolic and signaling amplification loop leading to cell death. Mol Syst Biol [Internet]. 2012 Jan 1 [cited 2014 Apr 11];8(1). Available from: http://msb.embopress.org/content/8/1/589

54. Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, et al. Mutant p53 Gain of Function in Two Mouse Models of Li-Fraumeni Syndrome. Cell. 2004 Dec 17;119(6):847–60.

55. Drost R, Bouwman P, Rottenberg S, Boon U, Schut E, Klarenbeek S, et al. BRCA1 RING Function Is Essential for Tumor Suppression but Dispensable for Therapy Resistance. Cancer Cell. 2011 Dec 13;20(6):797–809.

56. McCoy MK, Kaganovich A, Rudenko IN, Ding J, Cookson MR. Hexokinase activity is required for recruitment of parkin to depolarized mitochondria. Hum Mol Genet. 2013 Aug 19;ddt407.

57. Metallo CM, Walther JL, Stephanopoulos G. Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. J Biotechnol. 2009 Nov;144(3):167–74.

58. Ying H, Kimmelman AC, Lyssiotis CA, Hua S, Chu GC, Fletcher-Sananikone E, et al. Oncogenic Kras Maintains Pancreatic Tumors through Regulation of Anabolic Glucose Metabolism. Cell. 2012 Apr 27;149(3):656–70.

59. Boros LG, Lee PWN, Brandes JL, Cascante M, Muscarella P, Schirmer WJ, et al. Nonoxidative pentose phosphate pathways and their direct role in ribose synthesis in tumors: is cancer a disease of cellular glucose metabolism? Med Hypotheses. 1998 Jan;50(1):55–9.

60. McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. EMBO Rep. 2012 Jun 1;13(6):528–38.

61. Shi H, Moriceau G, Kong X, Lee M-K, Lee H, Koya RC, et al. Melanoma whole-exome sequencing identifies (V600E)B-RAF amplification-mediated acquired B-RAF inhibitor resistance. Nat Commun. 2012;3:724.

62. Warburg O. On the Origin of Cancer Cells. Nature. 1956 Feb 24;123(3191):309–14.

63. Anastasiou D, Poulogiannis G, Asara JM, Boxer MB, Jiang J, Shen M, et al. Inhibition of Pyruvate Kinase M2 by Reactive Oxygen Species Contributes to Cellular Antioxidant Responses. Science. 2011 Dec 2;334(6060):1278–83.

64. Schafer ZT, Grassian AR, Song L, Jiang Z, Gerhart-Hines Z, Irie HY, et al. Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. Nature. 2009 Sep 3;461(7260):109–13.

65. Weinberg F, Hamanaka R, Wheaton WW, Weinberg S, Joseph J, Lopez M, et al. Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. Proc Natl Acad Sci U S A. 2010 May 11;107(19):8788–93.

66. Lee AC, Fenster BE, Ito H, Takeda K, Bae NS, Hirai T, et al. Ras Proteins Induce Senescence by Altering the Intracellular Levels of Reactive Oxygen Species. J Biol Chem. 1999 Mar 19;274(12):7936–40.

67. Takahashi A, Ohtani N, Yamakoshi K, Iida S, Tahara H, Nakayama K, et al. Mitogenic signalling and the p16INK4a–Rb pathway cooperate to enforce irreversible cellular senescence. Nat Cell Biol. 2006 Nov;8(11):1291–7.

68. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011 Apr 28;12(4):R41.

69. Sanchez-Garcia F, Villagrasa P, Matsui J, Kotliar D, Castro V, Akavia U-D, et al. Integration of Genomic Data Enables Selective Discovery of Breast Cancer Drivers. Cell. 2014 Apr 12;159(6):1461–75.

70. Solimini NL, Xu Q, Mermel CH, Liang AC, Schlabach MR, Luo J, et al. Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential. Science. 2012 Jul 6;337(6090):104–9.

71. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003 Jul;34(3):267–73.

72. Wade M, Li Y-C, Wahl GM. MDM2, MDMX and p53 in oncogenesis and cancer therapy. Nat Rev Cancer. 2013 Feb;13(2):83–96.

73. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. Cell. 2014 Aug 14;158(4):929–44.

74. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nat Genet. 2014 Oct;46(10):1051–9.

75. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013 Oct;45(10):1127–33.

76. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, et al. Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci U S A. 2014 Jul 29;111(30):11139–44.

77. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011 Feb;11(2):85–95.

78. Vander Heiden MG. Targeting cancer metabolism: a therapeutic window opens. Nat Rev Drug Discov. 2011 Sep;10(9):671–84.

79. Kondoh H, Lleonart ME, Gil J, Wang J, Degan P, Peters G, et al. Glycolytic enzymes can modulate cellular life span. Cancer Res. 2005 Jan 1;65(1):177–85.

80. Memarzadeh S, Zong Y, Janzen DM, Goldstein AS, Cheng D, Kurita T, et al. Cell-autonomous activation of the PI3-kinase pathway initiates endometrial cancer from adult uterine epithelium. Proc Natl Acad Sci U S A. 2010 Oct 5;107(40):17298–303.

81. Bonnet S, Archer SL, Allalunis-Turner J, Haromy A, Beaulieu C, Thompson R, et al. A mitochondria-K+ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. Cancer Cell. 2007 Jan;11(1):37–51.

82. Chan DA, Sutphin PD, Nguyen P, Turcotte S, Lai EW, Banh A, et al. Targeting GLUT1 and the Warburg Effect in Renal Cell Carcinoma by Chemical Synthetic Lethality. Sci Transl Med. 2011 Aug 3;3(94):94ra70–94ra70.

83. Matthew G Vander Heiden HRC. Identification of small molecule inhibitors of pyruvate kinase M2. Biochem Pharmacol. 2009;79(8):1118–24.

84. Buisson R, Dion-Côté A-M, Coulombe Y, Launay H, Cai H, Stasiak AZ, et al. Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nat Struct Mol Biol. 2010 Oct;17(10):1247–54.

85. Watson J. Oxidants, antioxidants and the current incurability of metastatic cancers. Open Biol. 2013 Jan 1;3(1):120144.

86. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006 Dec;10(6):515–27.

87. Smith M, Marioni J, McKinney S, Hardcastle T, Thorne N. snapCGH: Segmentation, normalisation and processing of aCGH data. R Package Version 1360.

88. Seshan V, Olshen A. DNAcopy: DNA copy number data analysis. R Package Version 1400.

89. Zhang J. CNTools: Convert segment data into a region by sample matrix to allow for other high level computational analyses. R Package Version 1220.

90. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013 Mar 1;14(2):178–92.

91. Heidenblad M, Lindgren D, Veltman JA, Jonson T, Mahlamäki EH, Gorunova L, et al. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. Oncogene. 2005 Jan 31;24(10):1794–801.

92. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010 Feb 18;463(7283):899–905.

93. Smith ML, Marioni JC, Hardcastle TJ, Thorne NP. snapCGH: Segmentation, Normalization and Processing of aCGH Data Users' Guide. Bioinformatics. 2010;22:1144–6.

94. Karimpour-Fard A, Dumas L, Phang T, Sikela JM, Hunter LE. A survey of analysis software for array-comparative genomic hybridisation studies to detect copy number variation. Hum Genomics. 2010 Aug;4(6):421–7.

95. Data Matrix - Data Portal [Internet]. [cited 2012 Dec 14]. Available from: https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm?mode=ApplyFilter

96. CGCC Project (TCGA) @MSKCC [Internet]. [cited 2012 Dec 14]. Available from: http://cbio.mskcc.org/CGCC/

97. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostat Oxf Engl. 2004 Oct;5(4):557–72.

98. GenePattern: Home [Internet]. [cited 2012 Dec 14]. Available from: http://www.broadinstitute.org/cancer/software/genepattern/

99. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. Bioinforma Oxf Engl. 2007 Oct 15;23(20):2700–7.

100. Smyth GK, Speed T. Normalization of cDNA microarray data. Methods San Diego Calif. 2003 Dec;31(4):265–73.

101. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1–pl1.

102. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005 Oct 25;102(43):15545–50.

103. Yuan J, Bennett BD, Rabinowitz JD. Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc. 2008 Jul;3(8):1328–40.

104. Avril N, Menzel M, Dose J, Schelling M, Weber W, Jänicke F, et al. Glucose Metabolism of Breast Cancer Assessed by 18F-FDG PET: Histologic and Immunohistochemical Tissue Analysis. J Nucl Med. 2001 Jan 1;42(1):9–16.